A novel approach to the detection of unusual mitochondrial protein change suggests hypometabolism of ancestral simians

Bala Anı Akpınar[1], Vivek Sharma[1,2], Cory D. Dunn[1]

[1] Institute of Biotechnology, University of Helsinki, Helsinki, 00014, Finland
[2] Department of Physics, University of Helsinki, Helsinki, 00014, Finland

Corresponding author:

Cory Dunn, Ph.D.
P.O. Box 56
University of Helsinki
00014 Finland
Email: cory.dunn@helsinki.fi
Phone: +358 50 311 9307

Abstract:

The mitochondrial genome encodes core subunits of the oxidative phosphorylation machinery, and it is expected that these mitochondria-encoded polypeptides would be shaped by bioenergetic needs corresponding to diverse diets and environments. Here, we have developed a robust and effective method for highlighting phylogenetic tree edges with unexpectedly rapid, and likely efficacious, mitochondrial protein evolution. Further, our approach allows detection of those discrete protein substitutions likely to alter enzyme performance. A survey of mammalian taxonomic groups performed using our method supports the idea that widely conserved residues in mitochondria-encoded proteins are more likely to rapidly mutate within specific clades. Intriguingly, primates share a substitution profile with a number of mammals characterized by low mass-specific rates of metabolism. Our data suggest low metabolic performance and activity of ancestral simians, as well as reduced cellular metabolism across many extant primates.

Introduction:

Mitochondria are derived from an ancestral endosymbiont related to alpha-proteobacteria (Bonen *et al.* 1977; Roger *et al.* 2017). While most of the genes encoded by the pre-mitochondrial endosymbiont were lost or transferred to the (proto-)nucleus, a handful of protein-coding genes continue to reside upon a mitochondrial genome that is inherited maternally in vertebrates (Hoekstra 2000; Christie and Beekman 2017). The mitochondrial DNA (mtDNA) of mammals typically encodes 22 transfer RNAs, 2 ribosomal RNAs, and 13 proteins that are crucial for the process of oxidative phosphorylation (OXPHOS) (Gustafsson *et al.* 2016). Mitochondria-encoded proteins, along with numerous proteins encoded by nuclear DNA, meet at the mitochondrial inner membrane to form the complexes and super-complexes that drive OXPHOS (Pfanner *et al.* 2019). In cells engaged in mitochondrial ATP production, Complexes I through IV convert energy derived from food into a proton gradient across the mitochondrial inner membrane. Complex V takes advantage of this proton gradient to synthesize ATP (Sousa *et al.* 2018).

It is not surprising that many amino acid positions found within core OXPHOS complex components are well conserved (Nachman *et al.* 1996; Nachman 1998; Stewart *et al.* 2008; da Fonseca *et al.* 2008), given the fundamental role of these proteins in cellular bioenergetics. However, against this background of high protein conservation, changes have been identified within particular clades that may be linked to their bioenergetic needs. Unexpected substitutions to mtDNA-encoded proteins that might be associated with transitions in environment or physiology have been identified within multiple vertebrates (Garvin *et al.* 2015), including bats (Shen *et al.* 2010), primates (Pierron *et al.* 2012), hummingbirds (Dunn *et al.* 2020), 'high-performance' fish (Dalziel *et al.* 2006), snakes (Castoe *et al.* 2008), and aquatic mammals (McClellan *et al.* 2005).

Here, we have developed a straightforward and computationally undemanding method that can highlight phylogenetic tree edges and polypeptides characterized by greater-than-expected, potentially efficacious change to conserved amino acid positions. We have applied our approach to the full set of mammalian reference mtDNAs at multiple taxonomic levels, and we found that specific groups tend to be enriched for mutations at highly conserved sites. Primates clearly possess an elevated propensity for substitution at key locations within mitochondria-encoded proteins, and our analysis also indicates that ancestral simians were likely to be characterized by low mass-independent metabolic rates.

Results:

Toward a comprehensive analysis of mitochondrial protein evolution in mammals, we recovered full GenBank records for mammalian mtDNAs stored in the RefSeq database (O'Leary *et al.* 2016). The mtDNA record for the reptile *Anolis punctatus* was also retrieved for the purpose of rooting a mammalian phylogenetic tree. Coding sequences were extracted and concatenated, a practice consistent with uniparental inheritance and co-evolution of vertebrate mtDNA-encoded polypeptides (Sato and Sato 2017). Concatenated protein-coding sequences were aligned using MAFFT (Katoh and Standley 2013), and phylogenetic tree inference was performed using RAxML-NG (Kozlov *et al.* 2019). Next, we used the PAGAN package (Löytynoja *et al.* 2012) to reconstruct the amino acid character values of each position at all bifurcating nodes within our maximum likelihood mtDNA tree. Output protein alignments from PAGAN were ungapped using *Bos taurus* protein sequences as a reference. Using the tree and node character values emerging from PAGAN, we inferred all amino acid substitutions at positions for which less than 2% of extant and ancestral sequences contained gaps (Supplemental File 1). Taxonomic information for each input sequence was recovered from the National Center for Biotechnology Information (NCBI) (Schoch *et al.* 2020), and taxonomic information at internal nodes was assigned by concordant annotation of all descendent external tree nodes at each selected taxonomic level. The resulting taxonomic dataset was then linked to each analyzed tree edge and amino acid substitution.

Using our table of mitochondrial protein substitutions, we determined the relative conservation of each analyzed amino acid position by calculating a total substitution score (TSS, Figure 1a), which is simply the sum of substitution events mapped to a given alignment position across all tree edges. This 'tree-aware' metric may surpass measures of conservation that consider positional character frequencies extracted from multiple sequence alignments, since these character frequencies are sensitive to sampling biases. The dynamic range of TSSs is potentially unlimited and can take advantage of the continuous accretion of new sequence information. When plotting the TSSs of each analyzed position across mtDNA-encoded proteins, it is clear that the vast majority of these mitochondria-encoded amino acids are conserved and seldom change over roughly 200 million years of mammal evolution (Figure 1b). Indeed, the median TSS of analyzed positions is nine, the TSS at the 25th percentile is one, and the TSS at the 75th percentile is 56. However, a smaller subset of aligned, sparsely gapped positions were quite accepting of substitutions, and the maximum TSS across all analyzed mtDNA-encoded amino acid positions is 361. TSS values for each mitochondrial protein position, calculated from the mammalian dataset and indexed using the *Bos taurus* reference sequence, are provided as Supplemental File 2.

We then extended our analysis to highlight those specific tree edges where more substantive change to mammalian mtDNA-encoded proteins had occurred than would be expected based upon the overall level of nucleotide divergence (Figure 1c). Substitutions at a position with a high TSS (low conservation) and those occurring at a position with a low TSS (high conservation) might be expected to have effects of differing magnitude, so any approach

highlighting unexpected and efficacious protein evolution along an edge should take into account both the number and the potential effect of substitutions. Toward this goal, we recovered the TSS for each substituted position along a tree edge, added 1, calculated the reciprocal, and summed these reciprocal values in order to generate a substantive protein change sum (SPCS). A subset of tree edges were clearly associated with an elevated SPCS when considering all analyzed positions within mtDNA-encoded proteins. We then compared the SPCS of an edge to the nucleotide divergence along that edge, as represented by the tree branch length (BL) emerging from our PAGAN analysis. Only edges with BLs averaging less than 0.2 substitutions per site were included within this regression analysis (Figure 1d), and longer edges were discarded. When plotting SPCSs against BLs, we encountered a linear relationship (R-squared of 0.32) between SPCS and BL, with most tree edges (69%) falling below the regression line. When plotting SPCS residuals for each tree edge, residual magnitude appeared to scale with BL (Figure 1e), as might be expected if longer divergence times provide a greater opportunity for the accumulation of functionally neutral substitutions or of substitutions with potential effects on OXPHOS. Consequently, we calculated the arctangent (expressed in radians) of the SPCS residual divided by the BL, so that we might more meaningfully compare SPCSs from edges of low BL with edges of high BL. The resulting angle (θ_{evo}) is a measure of protein change suggestive of functional relevance.

Clearly, a number of edges exhibit a substantially elevated value of θ_{evo} when considering all mtDNA-encoded amino acid positions analyzed together (Figure 1f), or when analyzing individual respiratory complexes (Supplemental Figure S1). Demonstrating the utility of our approach, we encountered edges leading to several domesticated species among those with the highest θ_{evo} scores calculated across all subunits: *Bos taurus* (cattle), *Sus scrofa* (pig), and *Ovis aries* (sheep) were ranked first, seventh, and eighth of 2329 edges for which a θ_{evo} value was calculated (Supplemental File 3). These findings likely correspond with adaptation to the needs of their human captors or with potential population bottlenecks encountered during domestication (Larson and Fuller 2014; Bortoluzzi *et al.* 2020). Edges leading to several vulnerable or endangered species (IUCN 2008a; b, 2016a; b) are also commonly found within the top-ten ranked edges with respect to θ_{evo} calculated across all subunits, including *Gorilla gorilla* (western gorilla), *Ursus thibetanus mupinensis* (Asian black bear, Sichuan subspecies), *Budorcas taxicolor* (takin), and *Panthera tigris amoyensis* (South China tiger). These data are consistent with population bottlenecks that allow persistence of unusual and potentially deleterious substitutions (Ohta 1973). Unusual substitutions at polypeptide positions that are, with the exception of the indicated edge, totally conserved among mammals contribute greatly to the high θ_{evo} rank of these edges (Supplemental File 1). Our full θ_{evo} dataset, as well as the matching inferred substitutions, should serve as a valuable resource for those studying the evolution of mitochondrial proteins within specific mammalian groups, as well as their ancestors and kin.

After calculating the θ_{evo} values along each edge, we asked whether particular clades might harbor an unusual number of edges along which a considerable measure of potentially efficacious protein change had occurred. Toward this goal, we calculated the median θ_{evo} for

orders assigned five or more edges. Although orders typically encompassed edges with a wide diversity of θ_{evo} values (Figure 2a), several orders were clearly characterized by greater-than-expected median θ_{evo} values. The three orders with the highest median θ_{evo} values were Proboscidea (encompassing elephants and mammoths), Pilosa (sloths and anteaters), and Primates (including monkeys, apes, lemurs, and lorises), while the three orders with the lowest median θ_{evo} values were Perissodactyla (odd-toed ungulates), Artiodactyla (even-toed ungulates), and Lagomorpha (rabbits, hares, and pika). Lower boundaries of the 90% confidence interval of the median θ_{evo} (Figure 2b) further indicated that rapid and substantive protein evolution was occuring within the mitochondria of Primates and Pilosa, while the upper boundaries of the 90% confidence interval of the median θ_{evo} (Figure 2c) supported constrained evolution of mitochondrial proteins in Artiodactyla, Lagomorpha, and Perissodactyla. Similar outcomes emerged during analysis of individual OXPHOS complexes (Supplemental Figures S2 and S3)

A comparable evaluation was performed for mammalian families. As seen during our analysis of mammalian orders, some families exhibited widespread, rapid protein evolution at more conserved positions, while the opposite held true for other families. Here, Bradypodidae (three-toed sloths), Cercopithecidae (Old World monkeys), and Octodontidae (encompassing degus, rock rats, and viscacha rats) were associated with the three highest median θ_{evo} values (Figure 3a), while Equidae (horses and related animals), Leporidae (rabbits and hares), and Cervidae (deer) exhibited the three lowest median θ_{evo} values. Calculation of 90% confidence intervals supported the ranking of families obtained by analysis of median θ_{evo} values (Figures 3b and 3c). Analysis of individual OXPHOS complexes signalled similar divergence in the magnitude of protein evolution among families (Supplemental Figures S4, S5, and S6).

Given the abundance of mitochondrial protein evolution found among primates, we focused additional attention upon this clade by examining the trend among primate suborders, infraorders, subfamilies, and genera. At the rank of suborder, Haplorrhini (which includes Old World monkeys, New World monkeys, and apes) manifested a higher median θ_{evo} value than Strepsirrhini (encompassing lemurs, lorises, and galagos) (Figure 4a), a result supported by examining the 90% confidence limits of the median (Figures 4b). At the level of infraorder, Simiiformes (Old World monkeys, New World monkeys, and apes) exhibited a substantially higher median θ_{evo} value than other groups (Figure 4a), and this finding was also supported by examination of 90% confidence intervals (Figure 4b). When considering subfamilies, Cercopithecinae (encompassing most Old World monkeys, including baboons, macaques, and vervet monkeys) provided the clearest evidence for abundant and likely efficacious mitochondrial protein evolution, although the θ_{evo} values for several other subfamilies, such as Cebinae (New World capuchin monkeys) and Colobinae (multiple genera of Old World monkeys) were also elevated (Figures 4c and 4d). Finally, when genera were analyzed, substantive mitochondrial protein change was clearly evident within *Macaca* (macaques) (Figures 4e and 4f). Median θ_{evo} values were also elevated for the genera *Chlorocebus* (encompassing vervet and green monkeys), *Gorilla* (gorillas), *Papio* (baboons), and *Trachypithecus* (lutungs).

We hoped to better understand the potential outcome of the bulk of specific primate substitutions, defined as a given ancestral character linked to a given descendent character along the same edge, occurring at the most highly conserved protein positions. Toward this goal, we located matches for specific substitutions from selected groups at different TSS cut-off values across mammalian orders. Intriguingly, as the TSS cut-off was decreased toward a value of five or lower, an enrichment of analyzed primate substitutions within specific clades became clear. Specifically, the five orders Hyracoidea (hyraxes), Proboscidea, Microbiotheria (the monito del monte), Monotremata (includes platypuses and spiny anteaters), and Pilosa were most enriched with primate substitutions of TSS less than or equal to five when compared to their share of substitutions at all TSSs (Figure 5a). When focusing upon substitutions at highly conserved sites found within haplorhines (Figure 5b) and simians (Figure 5c), the results were similar, with Diprotodontia (which includes multiple marsupial families) replacing Proboscidea within the five top-ranked orders.

The analyses described above report upon the trends detected across selected taxonomic levels. Since simians, as well as multiple descendent lineages, seem to be replete with substitutions at positions highly conserved across mammals, we looked closely at the substitutions inferred along the most ancestral simian tree edges at amino acid positions of TSS five or below, since the identification of those ancestral mutations may be informative regarding the lifestyle of early anthropoids and may allow speculation regarding the effects of later substitutions found among descendants. When focusing upon the branch leading to Old World monkeys, apes, and New World monkeys (edge **i**, Figure 6), a number of changes to highly conserved positions with a TSS of five or less were predicted. Intriguingly, all of these mutations were found within Complex IV, consistent with previous analyses highlighting substantial change to primate cytochrome *c* oxidase components (Gissi *et al.* 2000; Grossman *et al.* 2004; Pierron *et al.* 2011, 2012). Additional changes likely to be functionally relevant, mostly located within Complex IV, were found along the descendent edges leading toward Old World monkeys and apes (edge **ii**, Figure 6), or toward New World monkeys (edge **iii**, Figure 6).

We then sought mutations outside of primates that would precisely match those with a TSS of five or less that are inferred at these three ancestral simian edges. Along the edge leading to Old World monkeys, apes, and New World monkeys (edge **i**, Figure 6), a Q52H change in *Antilope cervicapra* (blackbuck), which has a very high running speed, was predicted (Table 1). However, when considering the descendent edges leading to Old World monkeys and apes (edge **ii**, Figure 6) or to New World monkeys (edge **iii**, Figure 6), most of the substitutions at highly conserved sites matched those found within, or in edges leading to, groups quite conspicuous for their reduced metabolism, such as tuco-tucos, koalas, naked mole-rats, three-toed sloths, vampire bats, anteaters, and the monito del monte (Table 1).

224 **Discussion:**
225
226 We mapped predicted amino acid substitutions to the edges of a tree inferred from the
227 coding regions of mammalian mitochondrial genomes. Using this substitution map, we
228 calculated the TSS, a measure of conservation robust against sampling biases and benefitting
229 from a high dynamic range, at positions within mitochondria-encoded proteins. Using those
230 TSSs, our substitution map, and branch lengths determined from protein coding mtDNA
231 sequences, we calculated a score (θ_{evo}) that reports upon whether an unusual amount of
232 functional change may have occurred along a phylogenetic tree edge. Specific substitutions
233 along high-ranking tree edges that may be most impactful, based upon their location at
234 conserved positions, are easily retrieved for further analysis. To our knowledge, we have
235 carried out the most comprehensive analysis of mammalian mtDNA-encoded proteins ever
236 performed, and our findings serve as a resource for those investigating the structure and
237 function of mitochondria-encoded OXPHOS components. In addition, our dataset is highly
238 valuable for those interested in selective forces that may have helped shape mtDNAs during
239 the establishment and expansion of diverse mammalian clades.
240
241 Other approaches have previously been deployed toward the detection of substitutions
242 linked to functional change within mitochondrial proteins (Garvin *et al.* 2015). For example, the
243 dN/dS ratio, where the number of non-synonymous nucleotide changes at non-synonymous
244 sites is compared to the number of synonymous changes at synonymous sites, serves as the
245 basis for several popular methods seeking unusually rapid evolution of proteins (Suzuki and
246 Gojobori 1999; Yang *et al.* 2000; Kosakovsky Pond and Frost 2005). However, analysis of
247 large datasets by this approach can be computationally intensive (Jeffares *et al.* 2015), and the
248 ability of this metric to detect positive or relaxed selection on a protein may be diminished
249 when many sites are highly conserved (Sharp 1997; Hughes 2007; Echave *et al.* 2016). Other
250 weaknesses further limit the utility of dN/dS-based approaches (Kryazhimskiy and Plotkin
251 2008; Venkat *et al.* 2018; Wisotsky *et al.* 2020), including a difficulty in identifying instances of
252 fixed adaptive substitutions subject to stabilizing selection, even though these changes are
253 often linked to habitation of a new niche or to an interesting evolutionary transition (Hansen
254 1997; Bloom 2017). Alternative methodologies can take advantage of the chemical properties
255 of amino acids to identify adaptive change in polypeptides (Woolley *et al.* 2003; McClellan *et*
256 *al.* 2005). However, generalized measures of amino acid side behavior may not always allow
257 detection of functional change, since the site-specific acceptability of a given substitution is
258 ultimately decided within the context of its local protein environment (Zuckerkandl and Pauling
259 1965). Our rapid and computationally inexpensive approach will complement these existing
260 and established methods that seek notable change to mitochondrial polypeptides, and we
261 certainly anticipate the application of our method outside of the study of bioenergetics.
262
263 Orders and families can encompass members encountering a wide scope of ecological
264 factors and characterized by widely divergent life histories, and we indeed encounter many
265 outliers when examining the θ_{evo} values found within each taxonomic group. However, as
266 might be expected based on generations of selection for survival within differing ecological

267 niches, metabolic characteristics differ among taxonomic groups (McNab 2008; Griebeler and
268 Werner 2016), and we have noted a tendency for athleticism and high metabolic performance
269 to correspond with the conservation of mtDNA-encoded OXPHOS complex members among
270 land mammals. Indeed, artiodactyls and lagomorphs exhibit the greatest residual basal
271 metabolic rates among mammals, consistent with their often exceptional running speeds
272 (Hayssen and Lacy 1985; Lovegrove 1996, 2000; Garland 2009), and we detect the strongest
273 conservation of mtDNA-encoded proteins among these orders. Similarly, horses can exhibit an
274 exceptionally high mass-specific maximal rate of oxygen consumption (Hoppeler *et al.* 1987;
275 Poole and Erickson 2011), and we also encounter strong conservation of mitochondrial
276 proteins within this group. In contrast, many changes rapidly accumulate at conserved
277 mitochondrial protein positions when considering tree edges assigned to the order Pilosa and
278 family Bradypodidae. Accordingly, sloths are also characterized by very low metabolic rates,
279 poor thermoregulation, and low levels of activity (Britton and Atkinson 1938; Pauli *et al.* 2016;
280 Cliffe *et al.* 2018). Substantial change to positions conserved among mammals is also
281 detected across primates, and we discuss the evolution of primate mitochondrial proteins
282 below. Cladistic differences in the substitution profiles of mitochondrial proteins support the
283 idea that metabolism can be subject to phylogenetic inertia that may limit adaptation to new
284 diets and climes (Westoby *et al.* 1995; Hansen 1997).
285
286 ***Primates are enriched for mutations likely to affect oxidative phosphorylation***
287
288 Consistent with the results described above, previous studies of primates, necessarily
289 more limited in scale due to contemporary limitations in sequence availability, have also
290 revealed accelerated evolution and unusual substitutions within mtDNA- and nucleus-encoded
291 OXPHOS complex subunits (Gissi *et al.* 2000; Grossman *et al.* 2004; Pierron *et al.* 2011,
292 2012). At lower taxonomic levels, we found the clearest evidence of accelerated change of
293 mtDNA-encoded proteins within the suborder Haplorrhini, the infraorder Simiiformes, and
294 several families, subfamilies, and genera spread distributed throughout the primate tree. When
295 those specific substitutions located at the most conserved positions across primates,
296 haplorhines, or simian tree edges were sought outside of these groups, there was a substantial
297 enrichment within the orders Hyracoidea, Proboscidea, Microbiotheria, Monotremata, Pilosa,
298 and Diprotodontia. Low metabolism and difficulties in thermoregulation appear to typify the
299 order Hyracoidea (Rübsamen *et al.* 1982). The related order, Proboscidea, includes elephants
300 and their close relatives, which are also characterized by a reduced mass-specific basal and
301 field metabolism relative to other mammals (Christiansen 2004). The order Microbiotheria
302 contains one extant marsupial species, the monito del monte, which presents a low mass-
303 specific metabolic rate and can enter a prolonged torpor state in order to conserve energy
304 (Bozinovic *et al.* 2004). The metabolic rate of monotremes is considerably lower than most
305 placental mammals (Nicol 2017), with monotremes displaying the lowest body temperature
306 among groups of mammals (Clarke and O'Connor 2014). The enfeebled metabolism of Pilosa
307 members has been described above, and several members of the marsupial order
308 Diprotodontia, such as the koala bear (Degabriele and Dawson 1979; Nagy and Martin 1985)

309 and the wombat (Barboza *et al.* 1993; Evans *et al.* 2003), are low-activity mammals exhibiting
310 reduced metabolism.
311
312 Next, we looked specifically at mutations occuring along three selected edges
313 representing the origin of simians. The majority of those mutations at the most conserved
314 positions were found in Complex IV, and a subset of these were previously placed at the
315 interaction surface between cytochrome *c* and cytochrome *c* oxidase (Schmidt *et al.* 2005).
316 Complex IV is a key control point for OXPHOS activity (Piccoli *et al.* 2006; Dalmonte *et al.*
317 2009; Arnold 2012; Hill 2016), and given the dearth of these selected mutations among other
318 clades, compensation allowing functional neutrality (Starr and Thornton 2016) would, in our
319 view, be highly unlikely. Instead, alteration of Complex IV activity by these substitutions is a
320 more plausible scenario. For example, a change similar to the ancestral simian substitution
321 COII D119N, where a negatively charged side chain is replaced by a neutral side chain,
322 reduces cytochrome *c* oxidase activity in the alpha-proteobacterium *Rhodobacter sphaeroides*
323 (Wang *et al.* 1999; Zhen *et al.* 1999), and there is no clear evidence that this change might be
324 compensated in the apposed cytochrome *c* protein (Schmidt *et al.* 2005). At COI, D50N
325 (ancestral to Old World monkeys and apes) and G49E (ancestral to New World monkeys)
326 substitutions are found near residue D51, which is subject to redox-dependent conformational
327 and protonation changes (Yoshikawa *et al.* 1998; Okuno *et al.* 2003; Maréchal *et al.* 2012;
328 Sharma *et al.* 2017) and forms a key component of the 'H-channel' that serves as either a
329 proton transfer pathway or as a dielectric well (Yoshikawa *et al.* 1998; Rich and Maréchal
330 2013; Sharma *et al.* 2017). The Q52H (ancestral to simians) substitution adds a titratable
331 histidine to a region important for proton and water release (Son *et al.* 2017; Cai *et al.* 2018).
332 All three of these COI substitutions, quite unusual among mammals, should be expected to
333 alter enzyme kinetics.
334
335 ***Primates substitutions are likely to compromise OXPHOS***
336
337 If those mutations in stem simians do lead to changes in OXPHOS activity, might they
338 boost or reduce cellular metabolism? Equivalent substitutions to those found along these three
339 edges, in findings that mirror those obtained by our analysis of all primate, haplorhine, or
340 simian edges, were found among clades dominated by species with poor temperature
341 regulation and/or low metabolic rates (Britton and Atkinson 1938; McNab 1966, 2003, 2009,
342 2012; Degabriele and Dawson 1979; Nagy and Martin 1985; Busch 1989; Christiansen 2004;
343 Bozinovic *et al.* 2004; Stahl *et al.* 2012; Pauli *et al.* 2016; Cliffe *et al.* 2018). Specifically,
344 inferred substitutions found along the tree edge leading to Old World monkeys could be
345 matched in Myrmecophagidae (anteaters), *Dromiciops* (monito del montes), *Bradypus* (three-
346 toed sloths), and *Elephas* (elephants). Inferred mutations found along the tree edge leading to
347 New World monkeys were found in Desmodontinae (vampire bats), Bathyergidae (mole rats),
348 *Phascolarctos* (koalas), and Ctenomys (tuco-tucos).
349
350 Taken together, our analysis of mammalian mtDNAs suggests a general reduction in
351 the activity of OXPHOS complexes encoded by ancestral, and many extant, simians. These

352 findings are consilient with recent data demonstrating that the energy expenditure of primates
353 falls significantly below what would be anticipated when considering similarly sized mammals
354 (Pontzer *et al.* 2014; Pontzer 2015) and with the extremely slow life history of many primates
355 (Charnov 1993; Charnov and Berrigan 1993). In contrast, evidence that the changes to
356 mtDNA-encoded proteins in primates would generally lead an increase in OXPHOS activity is
357 scant, although it is certainly possible that mtDNA changes in select lineages, including
358 reversions of substitutions that limit metabolic activity, may bypass the effects of ancestral
359 OXPHOS component degradation.
360
361 If the mtDNA-encoded proteins of primates have indeed been adapted to a frugal
362 lifestyle, poor diet, and reduced cellular metabolism, how might the expansion of an
363 energetically expensive primate brain (Mink *et al.* 1981; Harris *et al.* 2012) be rationalized?
364 The mass-specific metabolic rates of whole organisms cannot easily be directly linked to
365 encephalization (Roth and Dicke 2005), or the extent to which the brain size corresponds to
366 expectations based upon size (Jerison 2012). As an example, monotreme encephalization
367 generally appears to fall in line with that of placental mammals (Macrini *et al.* 2006; Ashwell
368 and Ashwell 2013; Nicol 2017), in spite of their comparatively reduced metabolism. Elephants
369 also exhibit lower-than-expected metabolic rates (Christiansen 2004), yet elephants are more
370 highly encephalized than many mammals (Hofman 1982; Roth and Dicke 2005). Lagomorphs,
371 which exhibit a relatively high mass-specific metabolic rate (Hayssen and Lacy 1985;
372 Lovegrove 2000) tend not to be highly encephalized (Kruska 2005; Roth and Dicke 2005), and
373 artiodactyls are typically less encephalized than primates, although artiodactyls tend to exhibit
374 an elevated mass-specific metabolic rate (Lovegrove 2000; Snodgrass *et al.* 2007; Pontzer *et*
375 *al.* 2014; Pontzer 2015). If high metabolic performance is not directly linked to brain expansion,
376 then one may speculate instead (Pontzer 2015), consistent with the expensive tissue
377 hypothesis (Aiello and Wheeler 1995; Aiello *et al.* 2001), that OXPHOS components, and a
378 cellular physiology selected for energetic parsimony, can reduce the basic metabolic
379 requirements of mammals and license later evolution of a larger brain upon a sustained
380 improvement in diet. Further support for this idea will require additional comparison of the
381 mitochondrial activities and the metabolic properties of primates with those of other groups of
382 mammals.
383
384

385 **Methods:**
386
387 *Sequence acquisition and concatenation*
388
389 Record accessions for mammalian reference mtDNAs were obtained using the Organelle
390 Genome Resources provided by the National Center for Biotechnology Information Reference
391 Sequence project (NCBI RefSeq, Release 204,
392 https://www.ncbi.nlm.nih.gov/genome/organelle/) (O'Leary *et al.* 2016). All records not
393 containing 'NC_' at the beginning of their accession name were removed, then the accession
394 for the reptile *Anolis punctatus* was added. Using this list of accessions, full GenBank records
395 for these mtDNAs were downloaded by use of the NCBI Batch Entrez server
396 (https://www.ncbi.nlm.nih.gov/sites/batchentrez). This file (Supplemental File 4) was used as
397 input for the 'vert_mtDNA_genbank_grab.py' script
398 (https://github.com/corydunnlab/mammal_mitoprot_evolve).
399
400 *Sequence alignment*
401
402 The amino acid sequences of individual mitochondrial proteins extracted from the GenBank
403 input file, as well as a concatenate of protein coding sequences, was aligned using the FFT-
404 NS-2 method in MAFFT v7.475 (Katoh and Standley 2013).
405
406 *Phylogenetic tree inference*
407
408 Our alignment of mtDNA sequences encoding mitochondrial proteins was used as input for
409 RAxML-NG v0.9.0 (Kozlov *et al.* 2019) under a single partition, a GTR+FO+G4m model of
410 DNA change, and a seed set as 200. 10 random and 10 parsimony-based starting trees were
411 used to initiate tree inference, and the average relative Robinson-Foulds distance (Robinson
412 and Foulds 1981) for the resulting set of trees was 0.01. The best scoring tree is provided as
413 Supplemental File 5. 600 bootstrap replicates were generated using RAxML-NG, and a
414 weighted Robinson-Foulds distance converged below the 1% cutoff value. Both Felsenstein's
415 Bootstrap Proportions (Supplemental File 6) (Felsenstein 1985) and, perhaps more
416 appropriate for our large dataset (Soltis and Soltis 2003), the Transfer Bootstrap Expectations
417 (Supplemental File 7) (Lemoine *et al.* 2018) were calculated and used to label our best scoring
418 tree. The best maximum likelihood tree was used for downstream analyses after using FigTree
419 1.4.4 (https://github.com/rambaut/figtree/releases) to place the root upon the branch leading to
420 *Anolis punctatus*.
421
422

423 *Calculation and classification of edge θ_{evo} values*
424
425 GenBank records, sequence alignments of mitochondrial proteins, and our best maximum
426 likelihood tree were used as input for the script 'mtDNA_protein_evolution_calculator.py'
427 (https://github.com/corydunnlab/mammal_mitoprot_evolve). Within this script, PAGAN
428 v0.61(Löytynoja *et al.* 2012) is called to perform ancestral prediction of protein sequences at
429 internal tree nodes, and the resulting PAGAN alignments were ungapped using the *Bos taurus*
430 reference sequence (NC_006853.1). A PAGAN output tree is provided as Supplemental File 8.
431 Taxonomy of input accessions was recovered from the NCBI Taxonomy Database (access
432 date: January 29, 2021). Taxonomy of internal edges at a given taxonomic level was
433 determined by unanimous agreement among descendent edges. If multiple taxons were found
434 downstream of a descendent edge, 'MIXED' was assigned to that edge at that taxonomic level.
435 Edge θ_{evo} values were further examined and plotted in GraphPad Prism v9.0.2, and only
436 taxonomic groups containing five or more edges (allowing calculation of confidence intervals of
437 greater than or equal to 90%) are reported in our median-based analyses of specific taxonomic
438 groups.
439
440 *Enrichment of specific substitutions from a chosen clade within other clades, assessed at*
441 *different levels of conservation*
442
443 Specific substitutions found within the order Primates, the suborder Haplorrhini, and the
444 infraorder Simiiformes were extracted from our table (Supplemental File 1) containing all amino
445 acid substitutions at analyzed positions and their associated taxonomic assignments. These
446 same substitutions were then sought among edges found outside of these selected groups.
447 The recovered dataset was filtered by TSS to determine the enrichment of simian substitutions
448 in other clades at different levels of conservation. These analysis were performed using the
449 scripts 'primates_specific_substitutions_outside_of_primates_with_TSS_enrichment.py',
450 'haplorrhini_specific_substitutions_outside_of_haplorrhini_with_TSS_enrichment.py', and
451 'simiiformes_specific_substitutions_outside_of_simiiformes_with_TSS_enrichment.py'
452 (https://github.com/corydunnlab/mammal_mitoprot_evolve).
453
454 *Detection of specific substitutions found along a selected edge across the phylogenetic tree*
455
456 Provided a table containing all amino acid substitutions at analyzed positions and their
457 associated taxonomic data (Supplemental File 1), a selected edge, and a TSS threshold, the
458 script 'TSS_five_or_below_inside_and_outside_selected_edge.py'
459 (https://github.com/corydunnlab/mammal_mitoprot_evolve) finds substitutions along the
460 selected edge at alignment positions below the provided TSS value, as well as those specific
461 substitutions located in any other edge besides the selected edge.
462
463

Author contributions:

C.D.D. conceived of the methodology. C.D.D. and B.A.A. wrote software. C.D.D. and V.S. analyzed data. B.A.A., V.S., and C.D.D. wrote the manuscript.

Acknowledgements:

Funding for this project was obtained from the Sigrid Jusélius Foundation (Senior Researcher and Project Grants to C.D.D and V.S.), the European Research Council (ERC Starting Grant RevMito 637649 to C.D.D.), the Academy of Finland (to C.D.D. and V.S.), the University of Helsinki (to V.S.) and the Magnus Ehrnrooth Foundation (to V.S). C.D.D. and V.S. also thank the Center for Scientific Computing, Finland (Puhti and Mahti supercomputers), the Information Technology for Science Group, University of Helsinki (Ukko2 cluster) and PRACE (Marconi100 hosted at CINECA, Italy) for essential computational support. We thank Gülayşe İnce Dunn, Svetlana Konovalova, and Tamara Somborac for comments on the manuscript.

References:

Aiello L. C., and P. Wheeler, 1995 The Expensive-Tissue Hypothesis: The Brain and the Digestive System in Human and Primate Evolution. Current Anthropology 36: 199–221.

Aiello L. C., N. Bates, and T. Joffe, 2001 In defense of the Expensive Tissue Hypothesis, pp. 57–78 in *Evolutionary Anatomy of the Primate Cerebral Cortex*, edited by Falk D., Gibson K. R. Cambridge University Press, Cambridge.

Arnold S., 2012 The power of life—Cytochrome c oxidase takes center stage in metabolic control, cell signalling and survival. Mitochondrion 12: 46–56.

Ashwell K. W. S., and K. Ashwell, 2013 Cerebral cortex and claustrum/endopiriform complex. Neurobiology of Monotremes: Brain Evolution in Our Distant Mammalian Cousins, ed KWS Ashwell (Collingwood, Vic: CSIRO Publishing) 131–160.

Barboza P. S., I. D. Hume, and J. V. Nolan, 1993 Nitrogen Metabolism and Requirements of Nitrogen and Energy in the Wombats (Marsupialia: Vombatidae). Physiol. Zool. 66: 807–828.

Bloom J. D., 2017 Identification of positive selection in genes is greatly improved by using experimentally informed site-specific models. Biol. Direct 12: 1.

Bonen L., R. S. Cunningham, M. W. Gray, and W. F. Doolittle, 1977 Wheat embryo mitochondrial 18S ribosomal RNA: evidence for its prokaryotic nature. Nucleic Acids Res. 4: 663–671.

Bortoluzzi C., M. Bosse, M. F. L. Derks, R. P. M. A. Crooijmans, M. A. M. Groenen, *et al.*, 2020 The type of bottleneck matters: Insights into the deleterious variation landscape of small

501 managed populations. Evol. Appl. 13: 330–341.

502 Bozinovic F., G. Ruiz, and M. Rosenmann, 2004 Energetics and torpor of a South American
503 "living fossil", the microbiotheriid Dromiciops gliroides. J. Comp. Physiol. B 174: 293–297.

504 Britton S. W., and W. E. Atkinson, 1938 Poikilothermism in the Sloth. J. Mammal. 19: 94–99.

505 Busch C., 1989 Metabolic rate and thermoregulation in two species of tuco-tuco, Ctenomys
506 talarum and Ctenomys australis (Caviomorpha, Octodontidae). Comp. Biochem. Physiol.
507 A Comp. Physiol. 93: 345–347.

508 Cai X., K. Haider, J. Lu, S. Radic, C. Y. Son, *et al.*, 2018 Network analysis of a proposed exit
509 pathway for protons to the P-side of cytochrome c oxidase. Biochim. Biophys. Acta
510 Bioenerg. 1859: 997–1005.

511 Castoe T. A., Z. J. Jiang, W. Gu, Z. O. Wang, and D. D. Pollock, 2008 Adaptive evolution and
512 functional redesign of core metabolic proteins in snakes. PLoS One 3: e2201.

513 Charnov E. L., 1993 *Life History Invariants: Some Explorations of Symmetry in Evolutionary*
514 *Ecology*. Oxford University Press.

515 Charnov E. L., and D. Berrigan, 1993 Why do female primates have such long lifespans and
516 so few babies? Or life in the slow lane. Evolutionary Anthropology: Issues, News, and
517 Reviews 1: 191–194.

518 Christiansen P., 2004 Body size in proboscideans, with notes on elephant metabolism. Zool. J.
519 Linn. Soc. 140: 523–549.

520 Christie J. R., and M. Beekman, 2017 Uniparental Inheritance Promotes Adaptive Evolution in
521 Cytoplasmic Genomes. Mol. Biol. Evol. 34: 677–691.

522 Clarke A., and M. I. O'Connor, 2014 Diet and body temperature in mammals and birds: Diet
523 and body temperature. Glob. Ecol. Biogeogr. 23: 1000–1008.

524 Cliffe R. N., D. M. Scantlebury, S. J. Kennedy, J. Avey-Arroyo, D. Mindich, *et al.*, 2018 The
525 metabolic response of the Bradypus sloth to temperature. PeerJ 6: e5600.

526 Dalmonte M. E., E. Forte, M. L. Genova, A. Giuffrè, P. Sarti, *et al.*, 2009 Control of Respiration
527 by Cytochrome c Oxidase in Intact Cells. J. Biol. Chem. 284: 32331–32335.

528 Dalziel A. C., C. D. Moyes, E. Fredriksson, and S. C. Lougheed, 2006 Molecular evolution of
529 cytochrome c oxidase in high-performance fish (teleostei: Scombroidei). J. Mol. Evol. 62:
530 319–331.

531 Degabriele R., and T. J. Dawson, 1979 Metabolism and heat balance in an arboreal marsupial,
532 the koala (Phascolarctos cinereus). J. Comp. Physiol. 134: 293–301.

533 Dunn C. D., B. A. Akpınar, and V. Sharma, 2020 An Unusual Amino Acid Substitution Within
534 Hummingbird Cytochrome c Oxidase Alters a Key Proton-Conducting Channel. G3 10:
535 2477–2485.

536 Echave J., S. J. Spielman, and C. O. Wilke, 2016 Causes of evolutionary rate variation among
537 protein sites. Nat. Rev. Genet. 17: 109–121.

538 Evans M., B. Green, and K. Newgrain, 2003 The field energetics and water fluxes of free-living
539 wombats (Marsupialia: Vombatidae). Oecologia 137: 171–180.

540 Felsenstein J., 1985 CONFIDENCE LIMITS ON PHYLOGENIES: AN APPROACH USING
541 THE BOOTSTRAP. Evolution 39: 783–791.

542 Fonseca R. R. da, W. E. Johnson, S. J. O'Brien, M. J. Ramos, and A. Antunes, 2008 The
543 adaptive evolution of the mammalian mitochondrial genome. BMC Genomics 9: 119.

544 Garland T., 2009 The relation between maximal running speed and body mass in terrestrial
545 mammals. J. Zool. 199: 157–170.

546 Garvin M. R., J. P. Bielawski, L. A. Sazanov, and A. J. Gharrett, 2015 Review and meta-
547 analysis of natural selection in mitochondrial complex I in metazoans. J. Zoolog. Syst.
548 Evol. Res. 53: 1–17.

549 Geiser F., 1994 Hibernation and Daily Torpor in Marsupials - a Review. Aust. J. Zool. 42: 1–16.

550 Gissi C., A. Reyes, G. Pesole, and C. Saccone, 2000 Lineage-specific evolutionary rate in
551 mammalian mtDNA. Mol. Biol. Evol. 17: 1022–1031.

552 Griebeler E. M., and J. Werner, 2016 Mass, phylogeny, and temperature are sufficient to
553 explain differences in metabolic scaling across mammalian orders? Ecol. Evol. 6: 8352–
554 8365.

555 Grossman L. I., D. E. Wildman, T. R. Schmidt, and M. Goodman, 2004 Accelerated evolution
556 of the electron transport chain in anthropoid primates. Trends Genet. 20: 578–585.

557 Gustafsson C. M., M. Falkenberg, and N.-G. Larsson, 2016 Maintenance and Expression of
558 Mammalian Mitochondrial DNA. Annu. Rev. Biochem. 85: 133–160.

559 Hansen T. F., 1997 STABILIZING SELECTION AND THE COMPARATIVE ANALYSIS OF
560 ADAPTATION. Evolution 51: 1341–1351.

561 Harris J. J., R. Jolivet, and D. Attwell, 2012 Synaptic energy use and supply. Neuron 75: 762–
562 777.

563 Hayssen V., and R. C. Lacy, 1985 Basal metabolic rates in mammals: taxonomic differences in
564 the allometry of BMR and body mass. Comp. Biochem. Physiol. A Comp. Physiol. 81:

565 741–754.

566 Hill G. E., 2016 Mitonuclear coevolution as the genesis of speciation and the mitochondrial
567 DNA barcode gap. Ecol. Evol. 6: 5831–5842.

568 Hirakawa H., and Y. Nagasaka, 2018 Evidence for Ussurian tube-nosed bats (Murina
569 ussuriensis) hibernating in snow. Sci. Rep. 8: 12047.

570 Hoekstra R. F., 2000 Evolutionary origin and consequences of uniparental mitochondrial
571 inheritance. Hum. Reprod. 15 Suppl 2: 102–111.

572 Hofman M. A., 1982 Encephalization in mammals in relation to the size of the cerebral cortex.
573 Brain Behav. Evol. 20: 84–96.

574 Hoppeler H., J. H. Jones, S. L. Lindstedt, H. Claassen, K. E. Longworth, *et al.*, 1987 Relating
575 maximal oxygen consumption to skeletal muscle mitochondria in horses. Equine exercise
576 physiology 2: 278–289.

577 Hughes A. L., 2007 Looking for Darwin in all the wrong places: the misguided quest for positive
578 selection at the nucleotide sequence level. Heredity 99: 364–373.

579 IUCN, 2008a Panthera tigris ssp. amoyensis: Nyhus, P: The IUCN Red List of Threatened
580 Species 2008: e.T15965A5334628. IUCN Red List of Threatened Species.

581 IUCN, 2008b Budorcas taxicolor: Song, Y.-L., Smith, A.T. & MacKinnon, J: The IUCN Red List
582 of Threatened Species 2008: e.T3160A9643719. IUCN Red List of Threatened Species.

583 IUCN, 2016a Ursus thibetanus: Garshelis, D. & Steinmetz, R: The IUCN Red List of
584 Threatened Species 2020: e.T22824A166528664. IUCN Red List of Threatened Species.

585 IUCN, 2016b Gorilla gorilla: Maisels, F., bergl, R.a. & Williamson, E.a: The IUCN red list of
586 threatened species 2018: E.T9404A136250858. IUCN Red List of Threatened Species.

587 Jeffares D. C., B. Tomiczek, V. Sojo, and M. dos Reis, 2015 A beginners guide to estimating
588 the non-synonymous to synonymous rate ratio of all protein-coding genes in a genome.
589 Methods Mol. Biol. 1201: 65–90.

590 Jerison H., 2012 *Evolution of The Brain and Intelligence*. Elsevier.

591 Katoh K., and D. M. Standley, 2013 MAFFT multiple sequence alignment software version 7:
592 improvements in performance and usability. Mol. Biol. Evol. 30: 772–780.

593 Kosakovsky Pond S. L., and S. D. W. Frost, 2005 Not so different after all: a comparison of
594 methods for detecting amino acid sites under selection. Mol. Biol. Evol. 22: 1208–1222.

595 Kozlov A. M., D. Darriba, T. Flouri, B. Morel, and A. Stamatakis, 2019 RAxML-NG: a fast,

596 scalable and user-friendly tool for maximum likelihood phylogenetic inference.
597 Bioinformatics 35: 4453–4455.

598 Kruska D. C. T., 2005 On the evolutionary significance of encephalization in some eutherian
599 mammals: effects of adaptive radiation, domestication, and feralization. Brain Behav. Evol.
600 65: 73–108.

601 Kryazhimskiy S., and J. B. Plotkin, 2008 The population genetics of dN/dS. PLoS Genet. 4:
602 e1000304.

603 Larson G., and D. Q. Fuller, 2014 The Evolution of Animal Domestication. Annu. Rev. Ecol.
604 Evol. Syst. 45: 115–136.

605 Lemoine F., J.-B. Domelevo Entfellner, E. Wilkinson, D. Correia, M. Dávila Felipe, *et al.*, 2018
606 Renewing Felsenstein's phylogenetic bootstrap in the era of big data. Nature 556: 452–
607 456.

608 Lovegrove B. G., 1996 The low basal metabolic rates of marsupials: the influence of torpor and
609 zoogeography, pp. 141–151 in *Adaptations to the Cold (F. Geiser, AJ Hulbert, and SC*
610 *Nicol, Eds.). Tenth International Hibernation Symposium, University of New England*
611 *Press, Armidale, Australia*,.

612 Lovegrove B. G., 2000 The Zoogeography of Mammalian Basal Metabolic Rate. Am. Nat. 156:
613 201–219.

614 Lovegrove B. G., 2004 Locomotor mode, maximum running speed, and basal metabolic rate in
615 placental mammals. Physiol. Biochem. Zool. 77: 916–928.

616 Löytynoja A., A. J. Vilella, and N. Goldman, 2012 Accurate extension of multiple sequence
617 alignments using a phylogeny-aware graph algorithm. Bioinformatics 28: 1684–1691.

618 MacMillen R. E., and J. E. Nelson, 1969 Bioenergetics and body size in dasyruid marsupials.
619 Am. J. Physiol. 217: 1246–1251.

620 Macrini T. E., T. Rowe, and M. Archer, 2006 Description of a cranial endocast from a fossil
621 platypus, Obdurodon dicksoni (Monotremata, Ornithorhynchidae), and the relevance of
622 endocranial characters to monotreme monophyly. J. Morphol. 267: 1000–1015.

623 Maréchal A., B. Meunier, and P. R. Rich, 2012 Assignment of the CO-sensitive carboxyl group
624 in mitochondrial forms of cytochrome c oxidase using yeast mutants. Biochim. Biophys.
625 Acta 1817: 1921–1924.

626 McClellan D. A., E. J. Palfreyman, M. J. Smith, J. L. Moss, R. G. Christensen, *et al.*, 2005
627 Physicochemical evolution and molecular adaptation of the cetacean and artiodactyl
628 cytochrome b proteins. Mol. Biol. Evol. 22: 437–455.

McNab B. K., 1966 The metabolism of fossorial rodents: A study of convergence. Ecology 47: 712–733.

McNab B. K., 1992 The comparative energetics of rigid endothermy: the Arvicolidae. J. Zool. 227: 585–606.

McNab B. K., 2003 Standard energetics of phyllostomid bats: the inadequacies of phylogenetic-contrast analyses. Comp. Biochem. Physiol. A Mol. Integr. Physiol. 135: 357–368.

McNab B. K., 2008 An analysis of the factors that influence the level and scaling of mammalian BMR. Comp. Biochem. Physiol. A Mol. Integr. Physiol. 151: 5–28.

McNab B. K., 2009 Physiological convergence amongst ant-eating and termite-eating mammals. J. Zool. 203: 485–510.

McNab B. K., 2012 *Extreme Measures: The Ecological Energetics of Birds and Mammals*. University of Chicago Press.

Mink J. W., R. J. Blumenschine, and D. B. Adams, 1981 Ratio of central nervous system to body metabolism in vertebrates: its constancy and functional basis. Am. J. Physiol. 241: R203–12.

Nachman M. W., W. M. Brown, M. Stoneking, and C. F. Aquadro, 1996 Nonneutral mitochondrial DNA variation in humans and chimpanzees. Genetics 142: 953–963.

Nachman M. W., 1998 Deleterious mutations in animal mitochondrial DNA, pp. 61–69 in *Mutation and Evolution*, edited by Woodruff R. C., Thompson J. N. Springer Netherlands, Dordrecht.

Nagy K. A., and R. W. Martin, 1985 Field Metabolic Rate, Water Flux, Food Consumption and Time Budget of Koalas, Phascolarctos Cinereus (Marsupialia: Phascolarctidae) in Victoria. Aust. J. Zool. 33: 655–665.

Nicol S. C., 2017 Energy Homeostasis in Monotremes. Front. Neurosci. 11: 195.

Ohta T., 1973 Slightly deleterious mutant substitutions in evolution. Nature 246: 96–98.

Okuno D., T. Iwase, K. Shinzawa-Itoh, S. Yoshikawa, and T. Kitagawa, 2003 FTIR detection of protonation/deprotonation of key carboxyl side chains caused by redox change of the Cu(A)-heme a moiety and ligand dissociation from the heme a3-Cu(B) center of bovine heart cytochrome c oxidase. J. Am. Chem. Soc. 125: 7209–7218.

O'Leary N. A., M. W. Wright, J. R. Brister, S. Ciufo, D. Haddad, *et al.*, 2016 Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 44: D733–45.

662 Park T. J., J. Reznick, B. L. Peterson, G. Blass, D. Omerbašić, *et al.*, 2017 Fructose-driven
663 glycolysis supports anoxia resistance in the naked mole-rat. Science 356: 307–311.

664 Pauli J. N., M. Z. Peery, E. D. Fountain, and W. H. Karasov, 2016 Arboreal Folivores Limit
665 Their Energetic Output, All the Way to Slothfulness. Am. Nat. 188: 196–204.

666 Pfanner N., B. Warscheid, and N. Wiedemann, 2019 Mitochondrial proteins: from biogenesis to
667 functional networks. Nat. Rev. Mol. Cell Biol. 20: 267–284.

668 Piccoli C., R. Scrima, D. Boffoli, and N. Capitanio, 2006 Control by cytochrome c oxidase of
669 the cellular oxidative phosphorylation system depends on the mitochondrial energy state.
670 Biochem. J 396: 573–583.

671 Pierron D., J. C. Opazo, M. Heiske, Z. Papper, M. Uddin, *et al.*, 2011 Silencing, positive
672 selection and parallel evolution: busy history of primate cytochromes C. PLoS One 6:
673 e26269.

674 Pierron D., D. E. Wildman, M. Hüttemann, T. Letellier, and L. I. Grossman, 2012 Evolution of
675 the couple cytochrome c and cytochrome c oxidase in primates. Adv. Exp. Med. Biol. 748:
676 185–213.

677 Pontzer H., D. A. Raichlen, A. D. Gordon, K. K. Schroepfer-Walker, B. Hare, *et al.*, 2014
678 Primate energy expenditure and life history. Proc. Natl. Acad. Sci. U. S. A. 111: 1433–
679 1437.

680 Pontzer H., 2015 Energy Expenditure in Humans and Other Primates: A New Synthesis. Annu.
681 Rev. Anthropol. 44: 169–187.

682 Poole D. C., and H. H. Erickson, 2011 Highly athletic terrestrial mammals: horses and dogs.
683 Compr. Physiol. 1: 1–37.

684 Rich P. R., and A. Maréchal, 2013 Functions of the hydrophilic channels in protonmotive
685 cytochrome c oxidase. J. R. Soc. Interface 10: 20130183.

686 Robinson D. F., and L. R. Foulds, 1981 Comparison of phylogenetic trees. Math. Biosci. 53:
687 131–147.

688 Roger A. J., S. A. Muñoz-Gómez, and R. Kamikawa, 2017 The Origin and Diversification of
689 Mitochondria. Curr. Biol. 27: R1177–R1192.

690 Roth G., and U. Dicke, 2005 Evolution of the brain and intelligence. Trends Cogn. Sci. 9: 250–
691 257.

692 Rübsamen K., I. D. Hume, and W. von Engelhardt, 1982 Physiology of the rock hyrax. Comp.
693 Biochem. Physiol. A Comp. Physiol. 72: 271–277.

694 Sato K., and M. Sato, 2017 Multiple ways to prevent transmission of paternal mitochondrial
695 DNA for maternal inheritance in animals. J. Biochem. 162: 247–253.

696 Schmidt T. R., D. E. Wildman, M. Uddin, J. C. Opazo, M. Goodman, *et al.*, 2005 Rapid
697 electrostatic evolution at the binding site for cytochrome c on cytochrome c oxidase in
698 anthropoid primates. Proc. Natl. Acad. Sci. U. S. A. 102: 6379–6384.

699 Schoch C. L., S. Ciufo, M. Domrachev, C. L. Hotton, S. Kannan, *et al.*, 2020 NCBI Taxonomy:
700 a comprehensive update on curation, resources and tools. Database 2020.
701 https://doi.org/10.1093/database/baaa062

702 Sharma V., P. G. Jambrina, M. Kaukonen, E. Rosta, and P. R. Rich, 2017 Insights into
703 functions of the H channel of cytochrome c oxidase from atomistic molecular dynamics
704 simulations. Proc. Natl. Acad. Sci. U. S. A. 114: E10339–E10348.

705 Sharp P. M., 1997 In search of molecular darwinism. Nature 385: 111–112.

706 Shen Y.-Y., L. Liang, Z.-H. Zhu, W.-P. Zhou, D. M. Irwin, *et al.*, 2010 Adaptive evolution of
707 energy metabolism genes and the origin of flight in bats. Proc. Natl. Acad. Sci. U. S. A.
708 107: 8666–8671.

709 Snodgrass J. J., W. R. Leonard, and M. L. Robertson, 2007 Primate Bioenergetics: An
710 Evolutionary Perspective, pp. 703–737 in *PRIMATE ORIGINS: Adaptations and Evolution*,
711 edited by Ravosa M. J., Dagosto M. Springer US, Boston, MA.

712 Soltis P. S., and D. E. Soltis, 2003 Applying the Bootstrap in Phylogeny Reconstruction. Stat.
713 Sci. 18: 256–267.

714 Son C. Y., A. Yethiraj, and Q. Cui, 2017 Cavity hydration dynamics in cytochrome c oxidase
715 and functional implications. Proc. Natl. Acad. Sci. U. S. A. 114: E8830–E8836.

716 Sousa J. S., E. D'Imprima, and J. Vonck, 2018 Mitochondrial Respiratory Chain Complexes,
717 pp. 167–227 in *Membrane Protein Complexes: Structure and Function*, edited by Harris J.
718 R., Boekema E. J. Springer Singapore, Singapore.

719 Stahl M., C. Osmann, S. Ortmann, M. Kreuzer, J.-M. Hatt, *et al.*, 2012 Energy intake for
720 maintenance in a mammal with a low basal metabolism, the giant anteater
721 (Myrmecophaga tridactyla). J. Anim. Physiol. Anim. Nutr. 96: 818–824.

722 Starr T. N., and J. W. Thornton, 2016 Epistasis in protein evolution. Protein Sci. 25: 1204–
723 1218.

724 Stewart J. B., C. Freyer, J. L. Elson, A. Wredenberg, Z. Cansu, *et al.*, 2008 Strong purifying
725 selection in transmission of mammalian mitochondrial DNA. PLoS Biol. 6: e10.

726 Suzuki Y., and T. Gojobori, 1999 A method for detecting positive selection at single amino acid

727 sites. Mol. Biol. Evol. 16: 1315–1328.

728 Venkat A., M. W. Hahn, and J. W. Thornton, 2018 Multinucleotide mutations cause false
729 inferences of lineage-specific positive selection. Nat Ecol Evol 2: 1280–1288.

730 Wang K., Y. Zhen, R. Sadoski, S. Grinnell, L. Geren, *et al.*, 1999 Definition of the interaction
731 domain for cytochrome c on cytochrome c oxidase. Ii. Rapid kinetic analysis of electron
732 transfer from cytochrome c to Rhodobacter sphaeroides cytochrome oxidase surface
733 mutants. J. Biol. Chem. 274: 38042–38050.

734 Westoby M., M. R. Leishman, and J. M. Lord, 1995 On Misinterpreting the `Phylogenetic
735 Correction'. J. Ecol. 83: 531.

736 Wisotsky S. R., S. L. Kosakovsky Pond, S. D. Shank, and S. V. Muse, 2020 Synonymous Site-
737 to-Site Substitution Rate Variation Dramatically Inflates False Positive Rates of Selection
738 Analyses: Ignore at Your Own Peril. Mol. Biol. Evol. 37: 2430–2439.

739 Woolley S., J. Johnson, M. J. Smith, K. A. Crandall, and D. A. McClellan, 2003 TreeSAAP:
740 selection on amino acid properties using phylogenetic trees. Bioinformatics 19: 671–672.

741 Yang Z., R. Nielsen, N. Goldman, and A. M. Pedersen, 2000 Codon-substitution models for
742 heterogeneous selection pressure at amino acid sites. Genetics 155: 431–449.

743 Yoshikawa S., K. Shinzawa-Itoh, R. Nakashima, R. Yaono, E. Yamashita, *et al.*, 1998 Redox-
744 coupled crystal structural changes in bovine heart cytochrome c oxidase. Science 280:
745 1723–1729.

746 Zhen Y., C. W. Hoganson, G. T. Babcock, and S. Ferguson-Miller, 1999 Definition of the
747 interaction domain for cytochrome c on cytochrome c oxidase. I. Biochemical, spectral,
748 and kinetic characterization of surface mutants in subunit ii of Rhodobacter sphaeroides
749 cytochrome aa(3). J. Biol. Chem. 274: 38032–38041.

750 Zuckerkandl E., and L. Pauling, 1965 Evolutionary Divergence and Convergence in Proteins.
751 Evolving Genes and Proteins 97–166.

Edge analyzed	Edge description	Complex	Substitution	Matching edge	Comment on taxonomy
i	Edge leading to Old World monkeys, apes, and New World monkeys	IV	COX1 Q52H	to *Antilope cervicapra*	*Antilope cervicapra* has a very high maximal running speed (Lovegrove 2004).
ii	Edge leading to Old World monkeys and apes	I	ND1 C301L	Myrmecophagidae internal edge	Daughter edges lead to *Tamandua tetradactyla*, *Tamandua mexicana*, *Myrmecophaga tridactyla*. Anteaters typically exhibit a reduced basal metabolic rate and body temperature relative to other mammals (Stahl et al. 2012; McNab 2009).
ii	Edge leading to Old World monkeys and apes	IV	COX2 I42L	Dasyuromorphia internal edge	Daughter edges encompass a number of marsupial families. Marsupials typically have mass-specific metabolic rates that are lower than placental mammals. Several species descending from this edge undergo torpor, low-metabolism, low-temperature, deceased-activity state. (MacMillen and Nelson, 1969; Geiser 1994).
ii	Edge leading to Old World monkeys and apes	IV	COX2 I42L	to *Dromiciops gliroides*	*Dromiciops gliroides* is a marsupial with a basal metabolic rate below that predicted of a placental mammal of similar size. This organism can enter a prolonged torpor state induced by low ambient temperature and reduced food intake (Bozinovic et al. 2004).
ii	Edge leading to Old World monkeys and apes	IV	COX2 I42L	to *Murina ussuriensis*	*Murina ussuriensis* is the only bat known to hibernate in snowbanks, a location may provide greater temperature stability during hibernation (Hirakawa and Nagasaka 2018).
ii	Edge leading to Old World monkeys and apes	IV	COX2 M45T	*Bradypus* internal edge	Members of the *Bradypus* genus manifest greatly reduced metabolic rates and low activity relative to other mammals, and the body temperatures of these sloths can fluctuate wildly (Cliffe et al. 2018; Pauli et al. 2016; Britton and Atkinson 1938).
ii	Edge leading to Old World monkeys and apes	IV	COX2 M45T	to *Elephas antiquus*	*Elephas antiquus* (or *Palaeoloxodon antiquus*) is an extinct elephant of greater size than extant elephants. Larger mammals tend to have lower mass-specific metabolic rates than smaller mammals, and elephants have a lower energy expenditure and basal metabolic rate than its size would predict (McNab 2012; Christiansen 2004)
iii	Edge leading to New World monkeys	I	ND4 S80T	Desmodontinae internal edge	Daughter edges lead to *Diphylla ecaudata*, *Diaemus youngi*, and *Desmodus rotundus*. Vampire bats typically have lower mass-independent basal metabolic rates than other bats (McNab 2003).
iii	Edge leading to New World monkeys	I	ND5 S343A	Arvicolinae internal edge	Many arvicolids manifest a high mass-independent basal metabolic rate and high body temperature, consistent with a high rate of energy expenditure (McNab 1992).
iii	Edge leading to New World monkeys	I	ND5 S343A	Bathyergidae internal edge	Daughter edges lead to the fossorial rodents *Heterocephalus glaber* (naked mole-rat) and *Fukomys damarensis* (Damaraland mole-rat). *Heterocephalus glaber* has a very low mass-independent metabolic rate, is a poor thermoregulator, and is highly tolerant of hypoxia (McNab 1966; Park et al. 2017).
iii	Edge leading to New World monkeys	IV	COX1 G49E	to *Phascolarctos cinereus*	The mass-independent metabolic rate of koalas are only ~74% of expectations for marsupials of its size, and koalas are not characterized by high activity (Degabriele and Dawson 1979; Nagy and Martin 1985).
iii	Edge leading to New World monkeys	IV	COX1 G49E	to *Ctenomys leucodon*	Tuco-tucos are fossorial rodents. Typically, tuco-tucos have low mass-independent metabolic rates (Busch 1989).

Table 1: Substitutions found outside of primates matching those in selected primate ancestral edges. Those substitutions matching mutations inferred at the ancestral edges (**i**), (**ii**), and (**iii**) shown in Figure 6 are listed, along with comments regarding the metabolic proficiency of the associated clade or species.

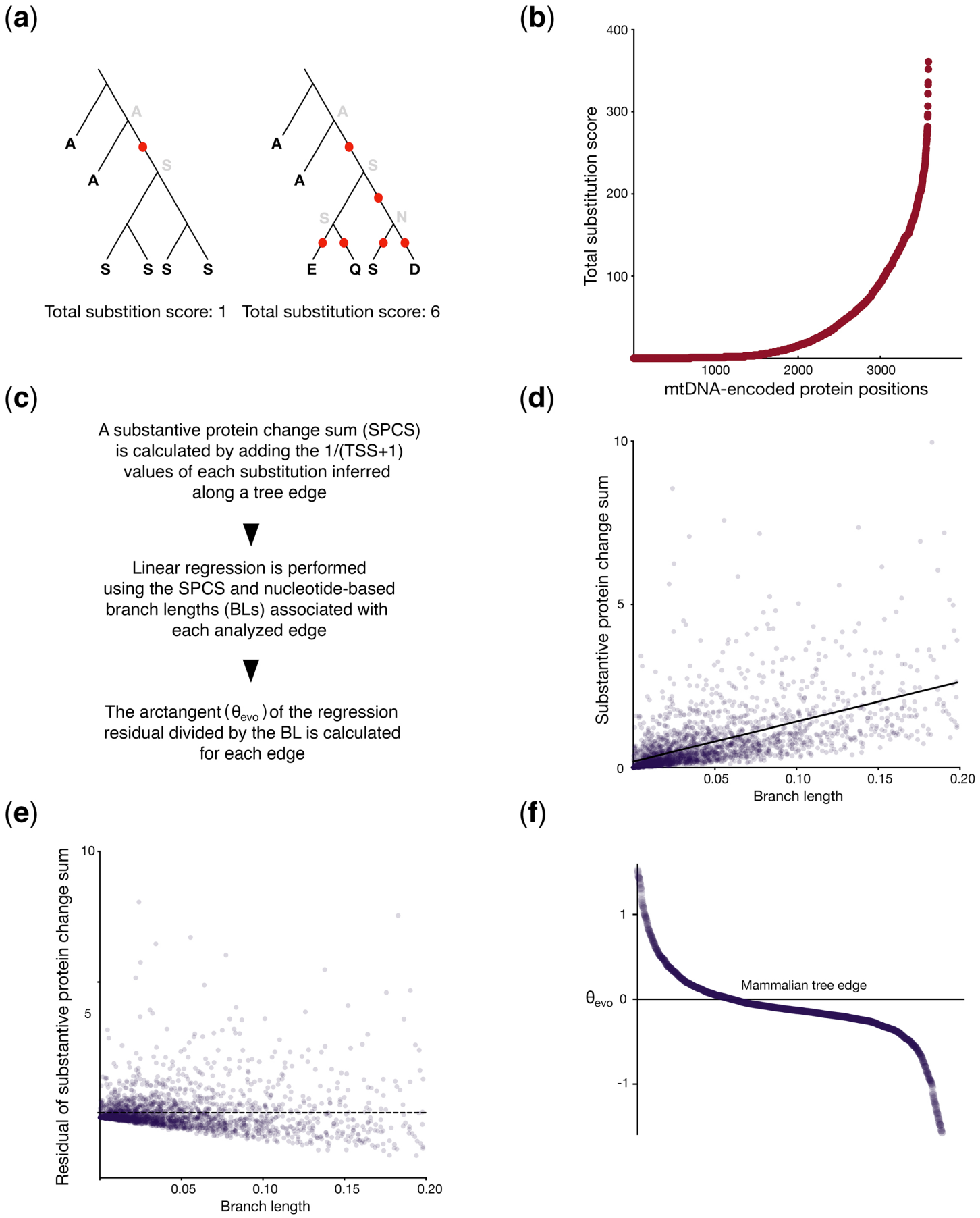


Figure 1: Detection of accelerated and potentially functional protein evolution along phylogenetic tree edges. (**a**) The total substitution score (TSS) is a measure of alignment position conservation. Character values at internal nodes are predicted, and inferred substitutions along an edge are summed. (**b**) The TSS values of all alignment positions gapped within less than 2% of inferred ancestral and input sequences are plotted for all mitochondria-encoded proteins. Alignment positions are ordered by increasing TSS value. (**c**) The methodology for detecting when an edge is characterized by greater-than-expected protein evolution is outlined: (**d**) The alignment position TSS for each substitution detected along an edge was added to 1, and the reciprocals of these values were summed to generate a substantive protein change sum (SPCS) for each edge found within a tree of mammalian mtDNA sequences. Edges with a branch length greater than or equal to 0.2 were removed, and linear regression was performed (black line). (**e**) Residuals for each sample were calculated, and (**f**) the arctangent of the residual divided by the branch length, θ_{evo} , for each analyzed edge across all analyzed mitochondria-encoded amino acid positions is computed.

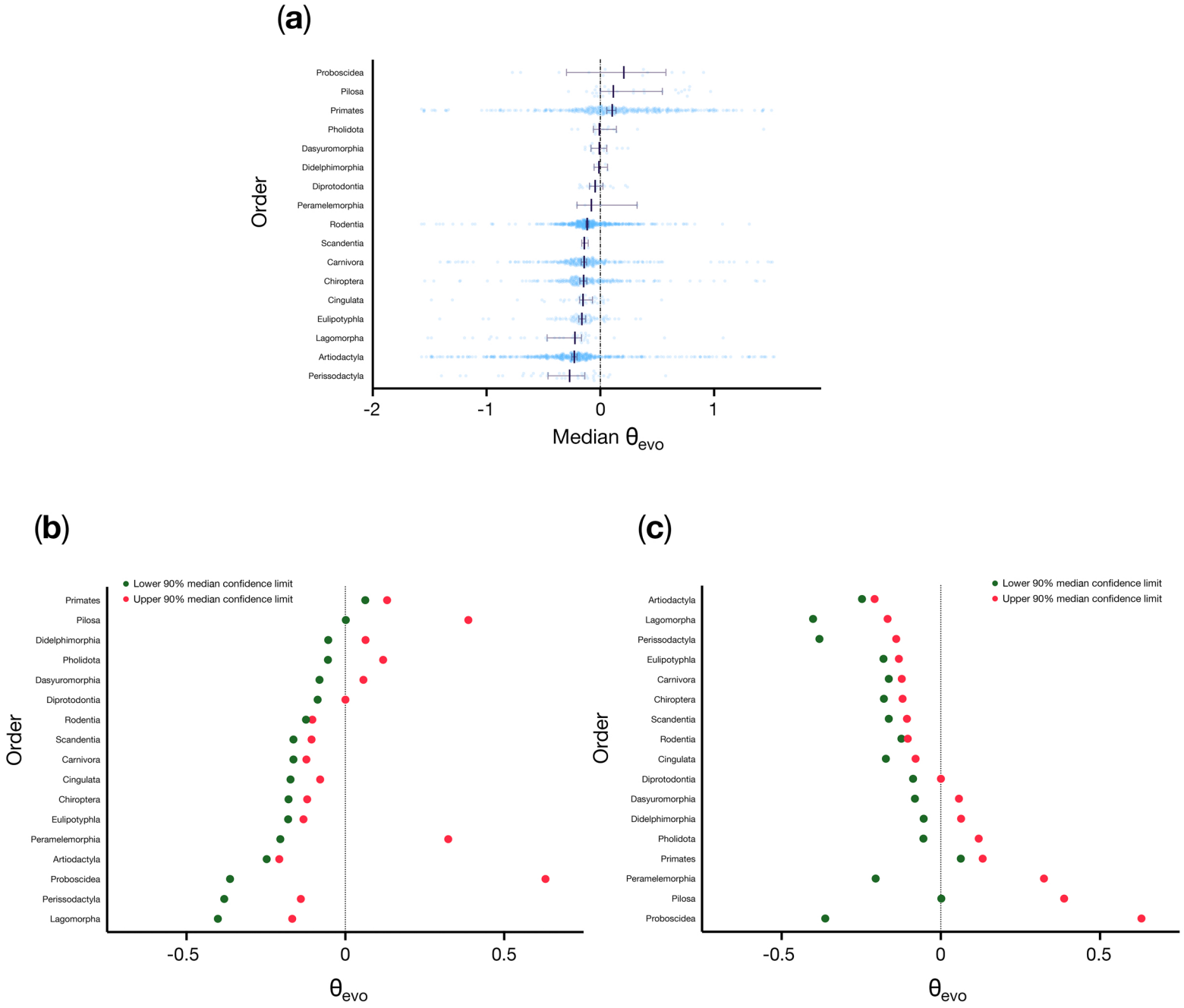


Figure 2: Mammalian orders differ in their propensity for potentially efficacious mitochondrial protein substitutions. (**a**) The median θ_{evo} calculated for each order assigned five or more edges is shown (vertical bar), along with the interquartile range (error bars). Lower and upper 90% confidence intervals of the median are plotted, with the orders arranged from either the highest 90% median confidence limit to the lowest (**b**) or arranged from the lowest upper 90% median confidence interval to the highest (**c**).

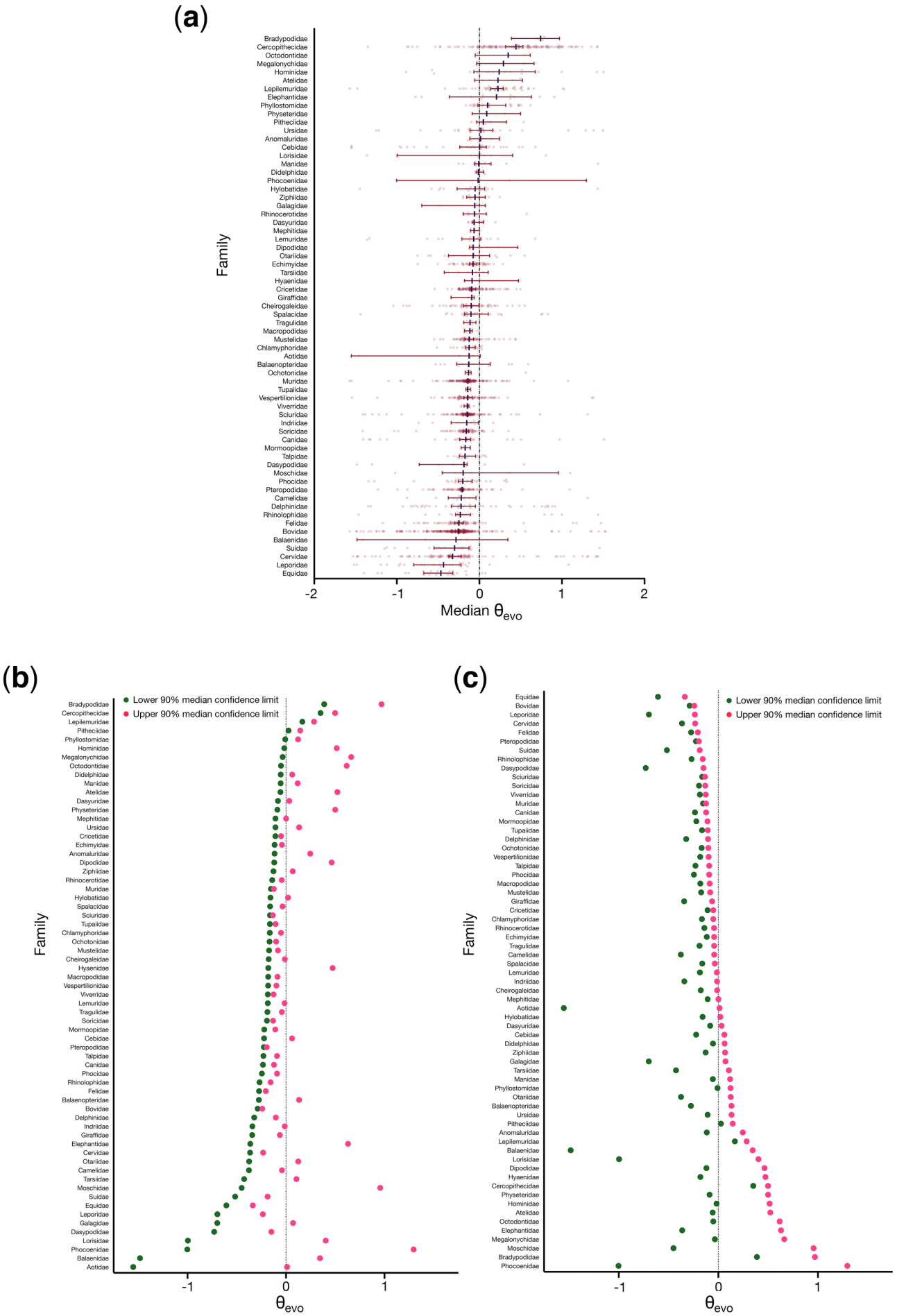


Figure 3: Mammalian families differ in their likelihood of potentially efficacious mitochondrial protein substitutions. The analyses in (**a**), (**b**), and (**c**) were performed as in Figure 2, except mammalian families assigned five or more edges are analyzed.

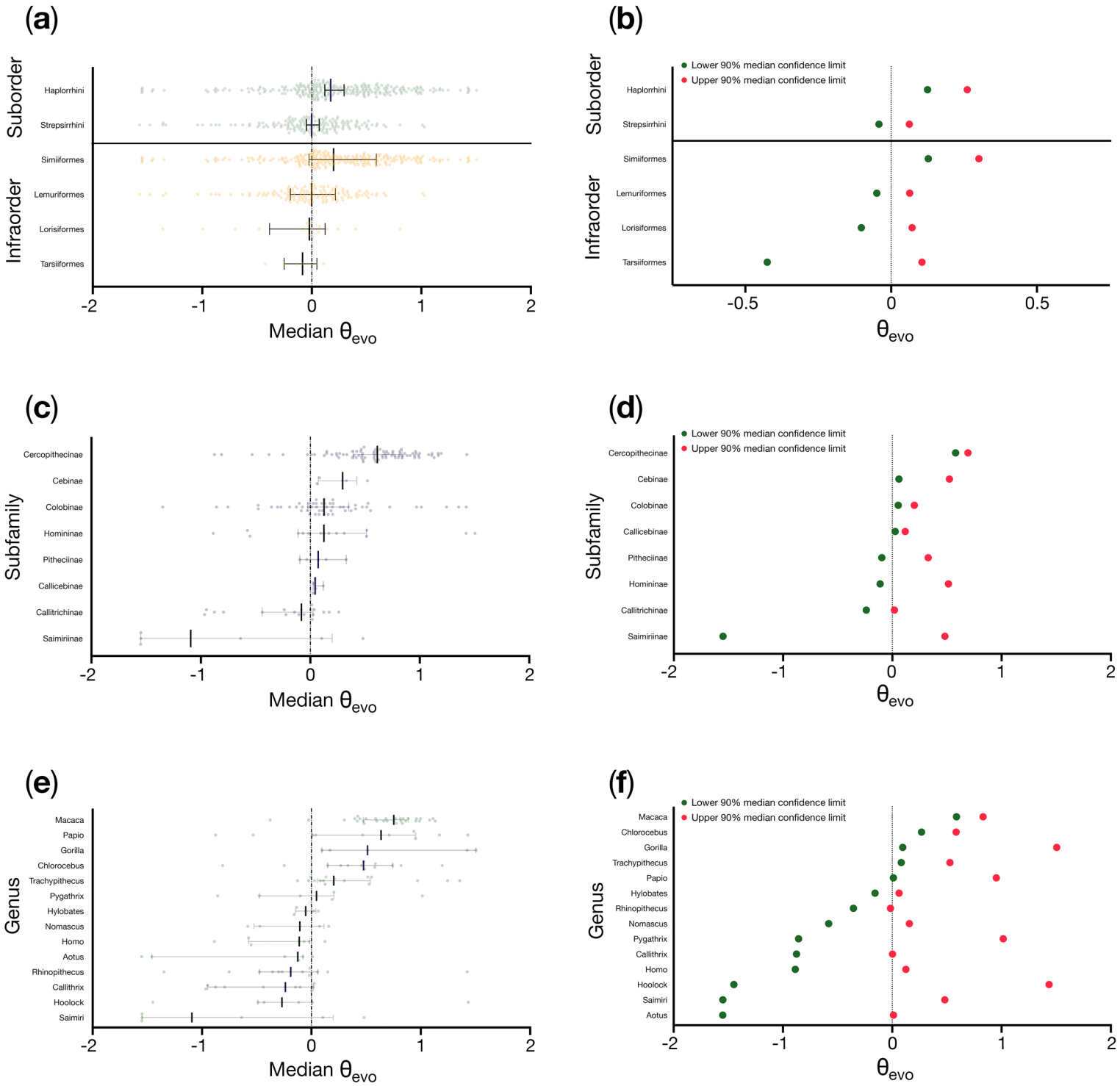


Figure 4: Substantial mitochondrial protein evolution is prevalent across several primate clades. Analyses were carried out as in Figures 2 and 3, except primate suborders and infraorders (**a** and **b**), subfamilies (**c** and **d**), or genera (**e** and **f**) assigned five or more edges are included.

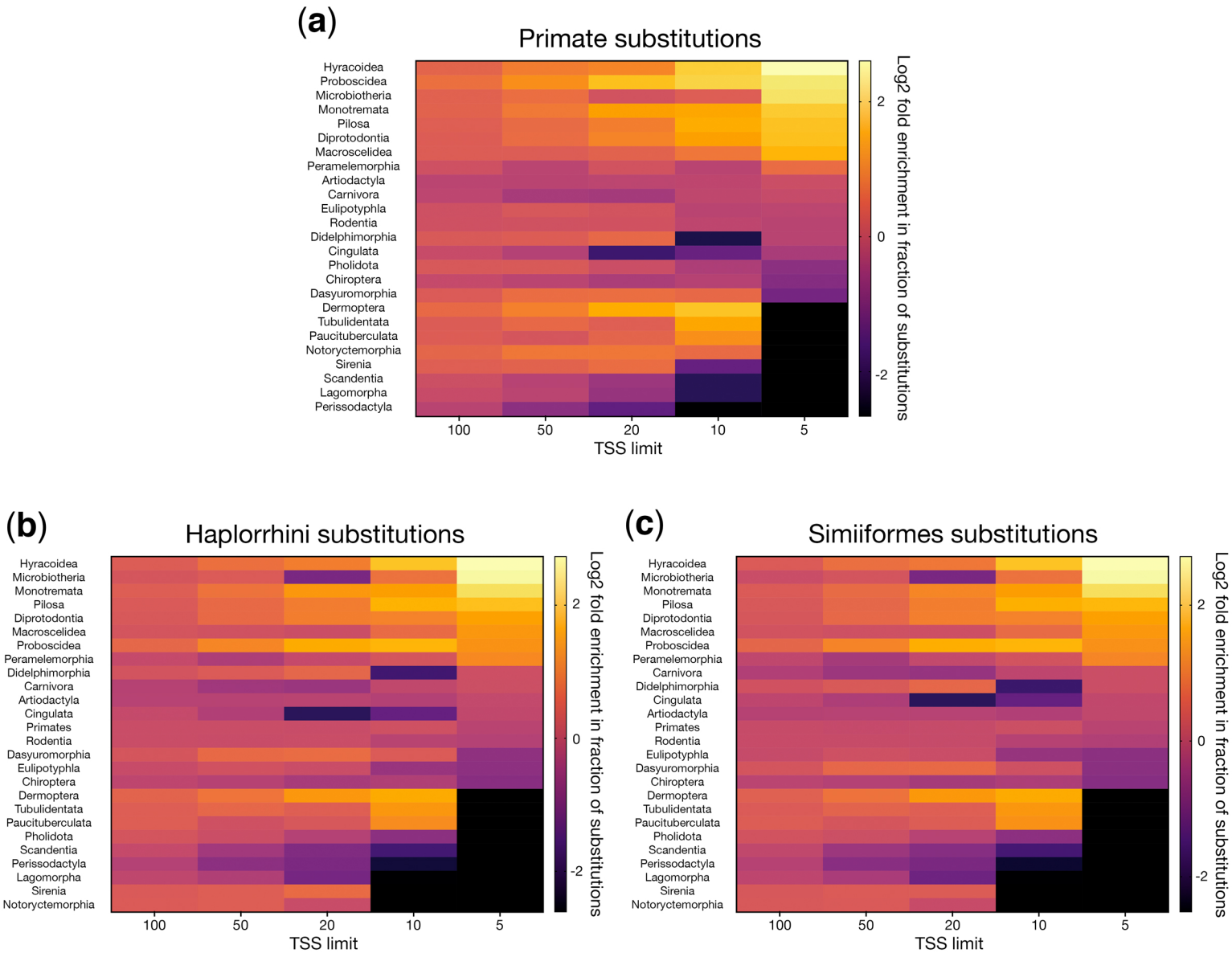


Figure 5: A search for specific primate substitutions at highly conserved mitochondrial protein positions outside of selected clades. Specific substitutions found along edges assigned to (**a**) primates, (**b**) haplorhines, and (**c**) simians were located in edges outside of the selected clade, and those substitutions were assigned to mammalian orders. The $\log_2$ fold change was used to compare the proportion of these substitutions found in each order when a TSS cut-off was applied to the proportion of those found in each order when the dataset was unfiltered by conservation level.

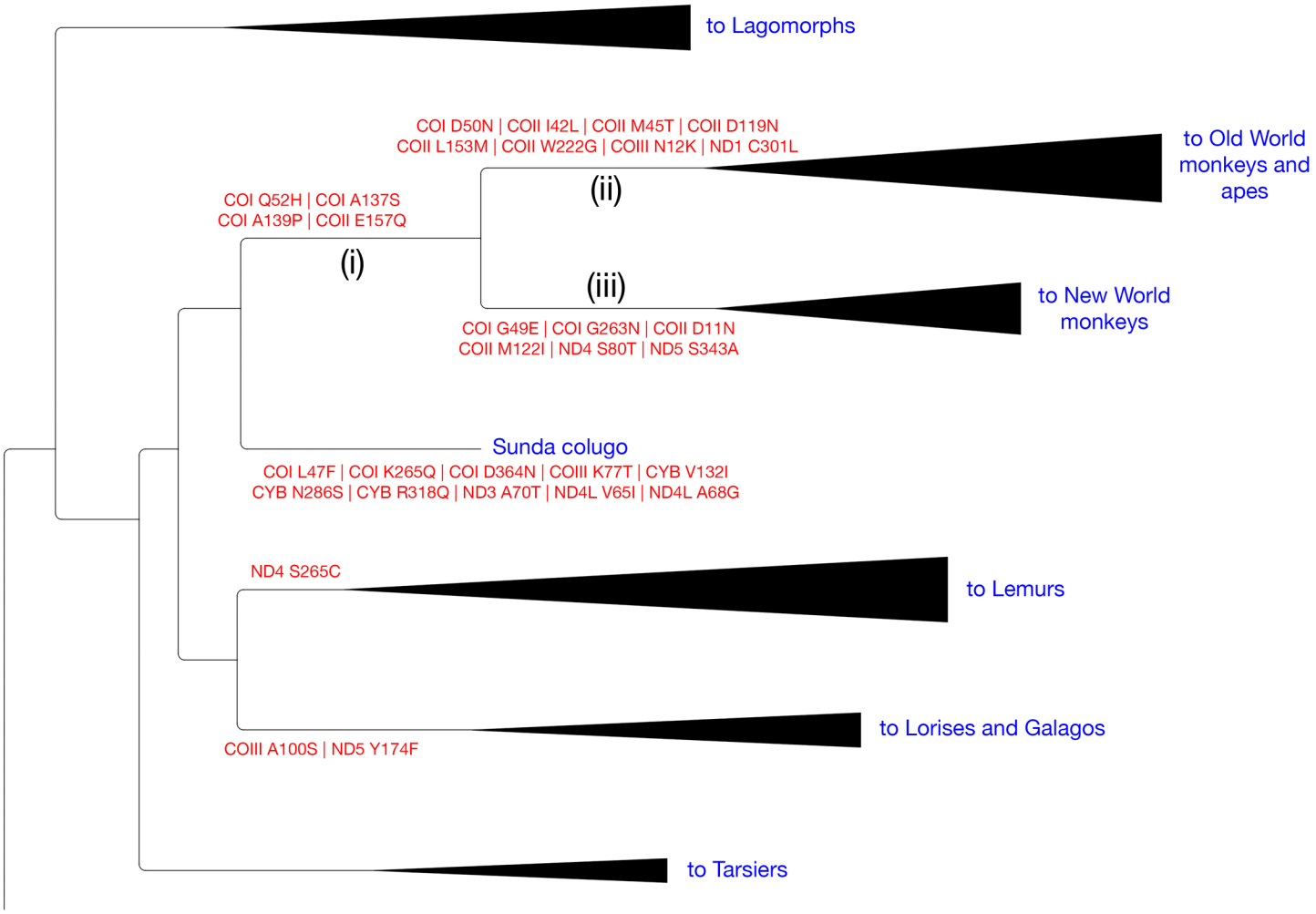
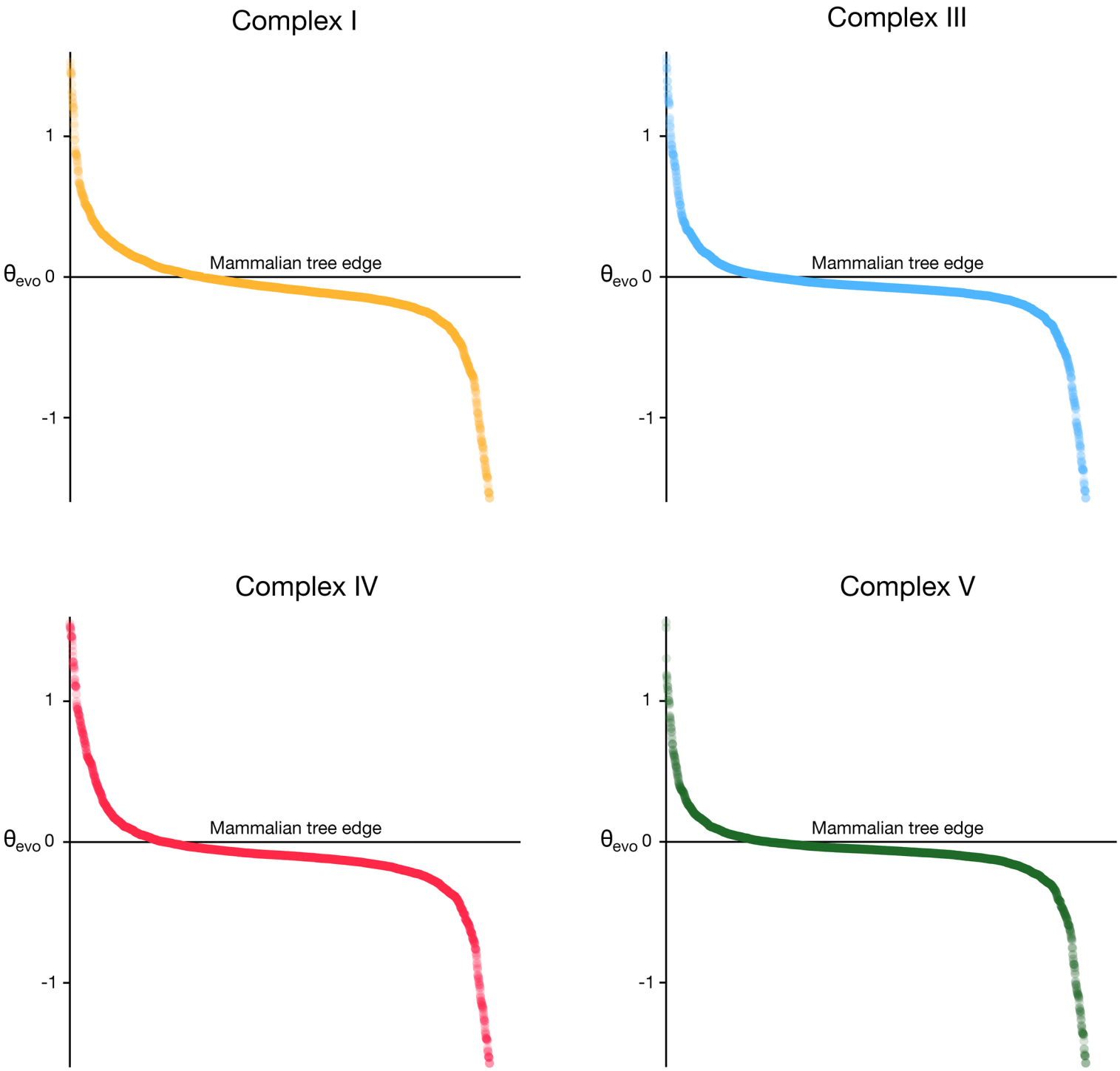
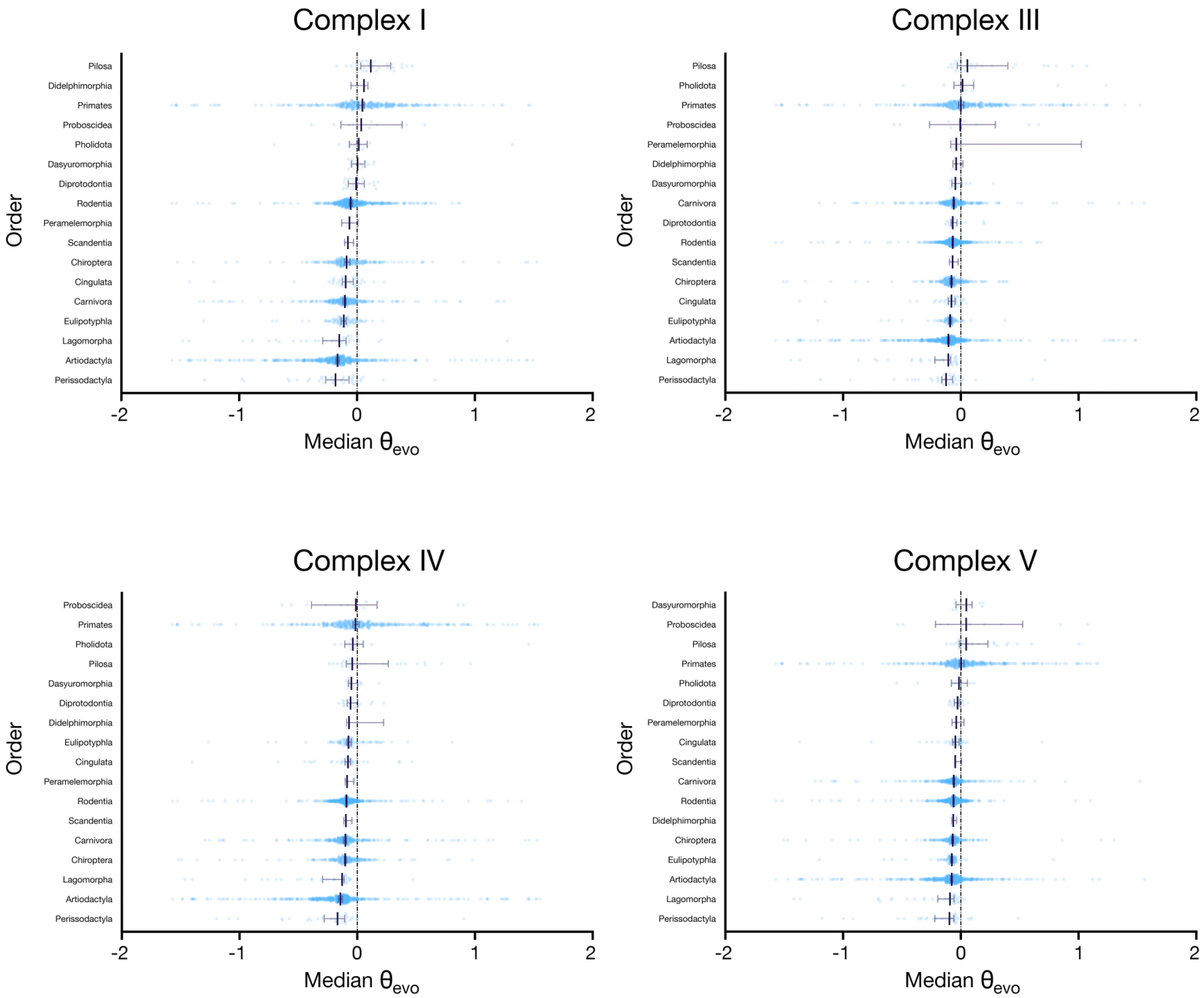


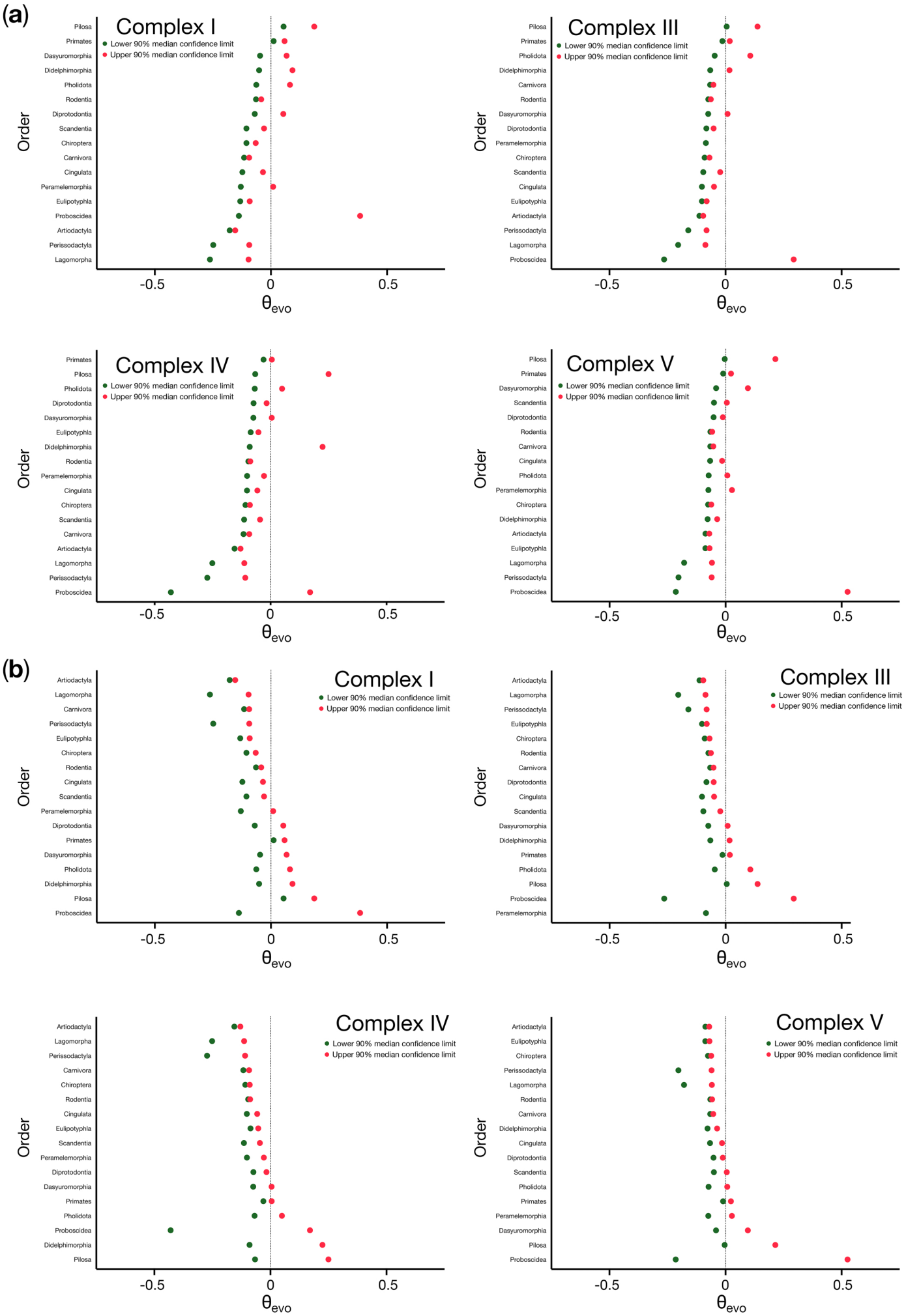
Figure 6: Substitutions at highly conserved mitochondrial protein positions are predicted at edges representing early simian evolution. A portion of the inferred mammalian mtDNA tree is provided, and all substitutions occurring at positions with a TSS of five or lower are shown. Edge (**i**) leads to the node parental to Old World monkeys, apes, and New World monkeys, edge (**ii**) leads to Old World monkeys and apes, and edge (**iii**) leads to New World monkeys. All alignment positions are indexed to *Bos taurus* reference sequences.



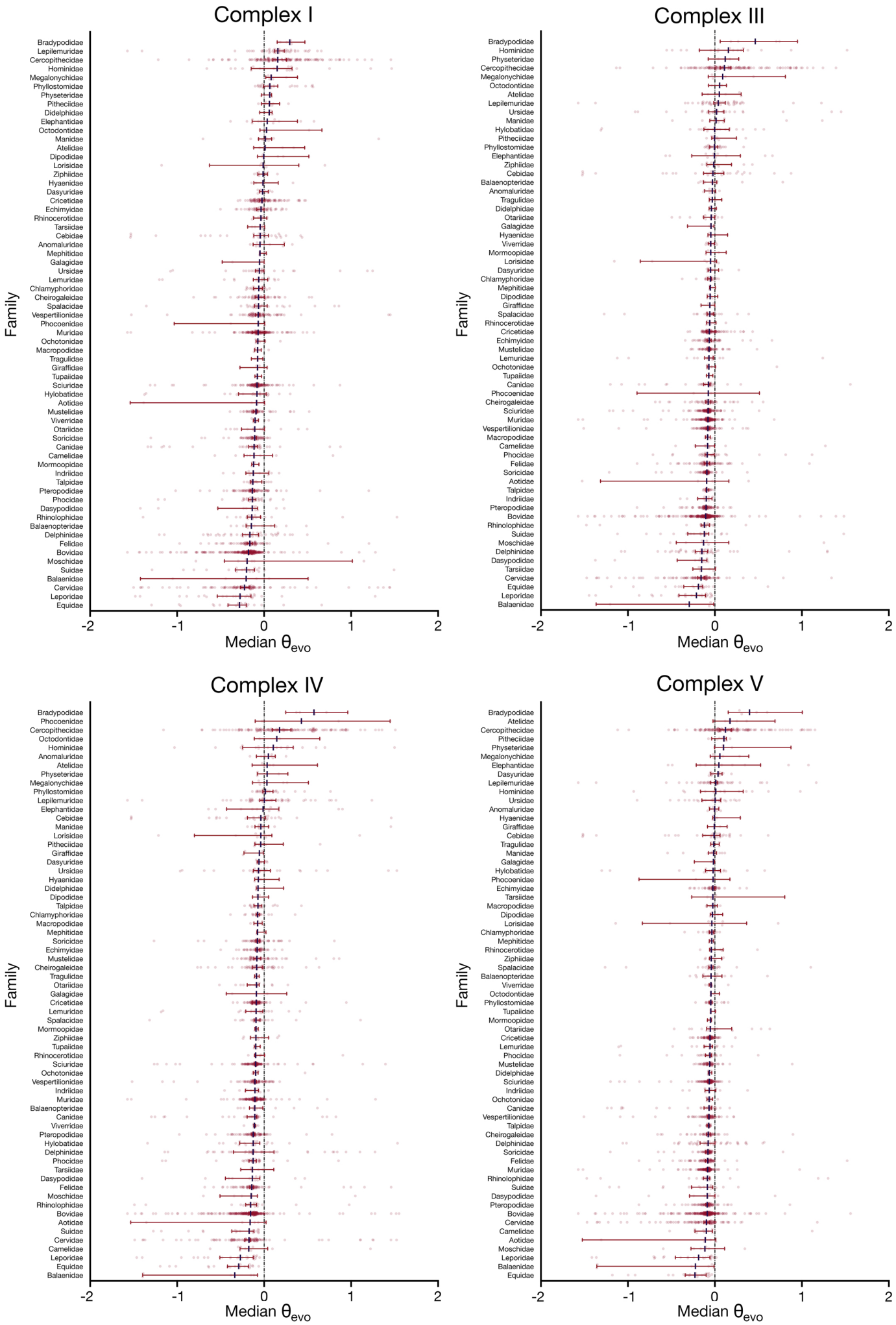
Supplemental Figure S1: θ_{evo} calculated for each analyzed edge for specific OXPHOS complexes. Analysis was performed as in Figure 1f, except that SPCSs calculated from protein positions in Complex I, Complex III, Complex IV, or Complex V were used to generate θ_{evo} values.



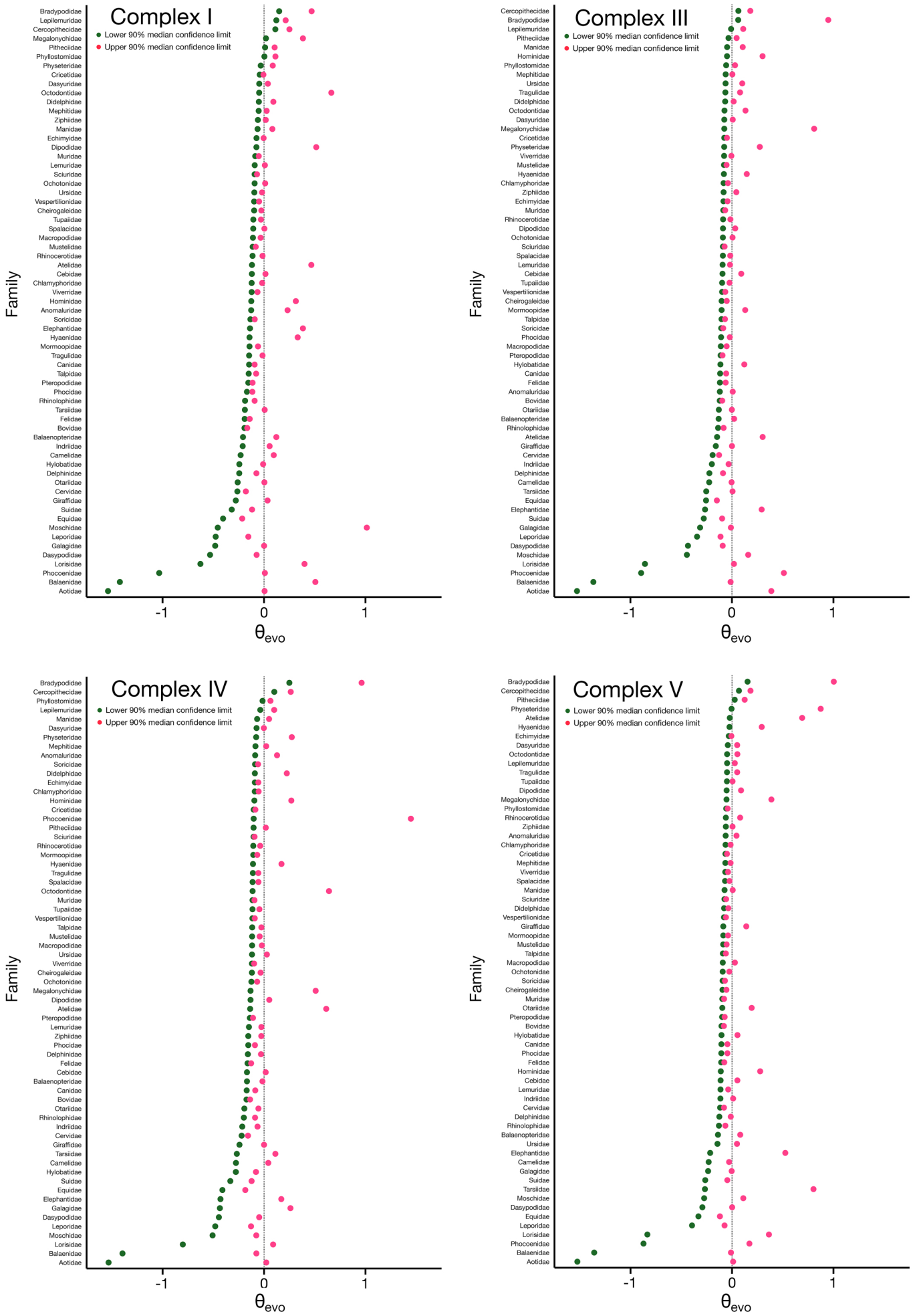
Supplemental Figure S2: Mammalian orders differ in their propensity for potentially efficacious mitochondrial protein substitutions at specific OXPHOS complexes (median calculations). Analysis was performed as in Figure 2a, except that θ_{evo} values were obtained by analysis of Complex I, Complex III, Complex IV, or Complex V polypeptides.



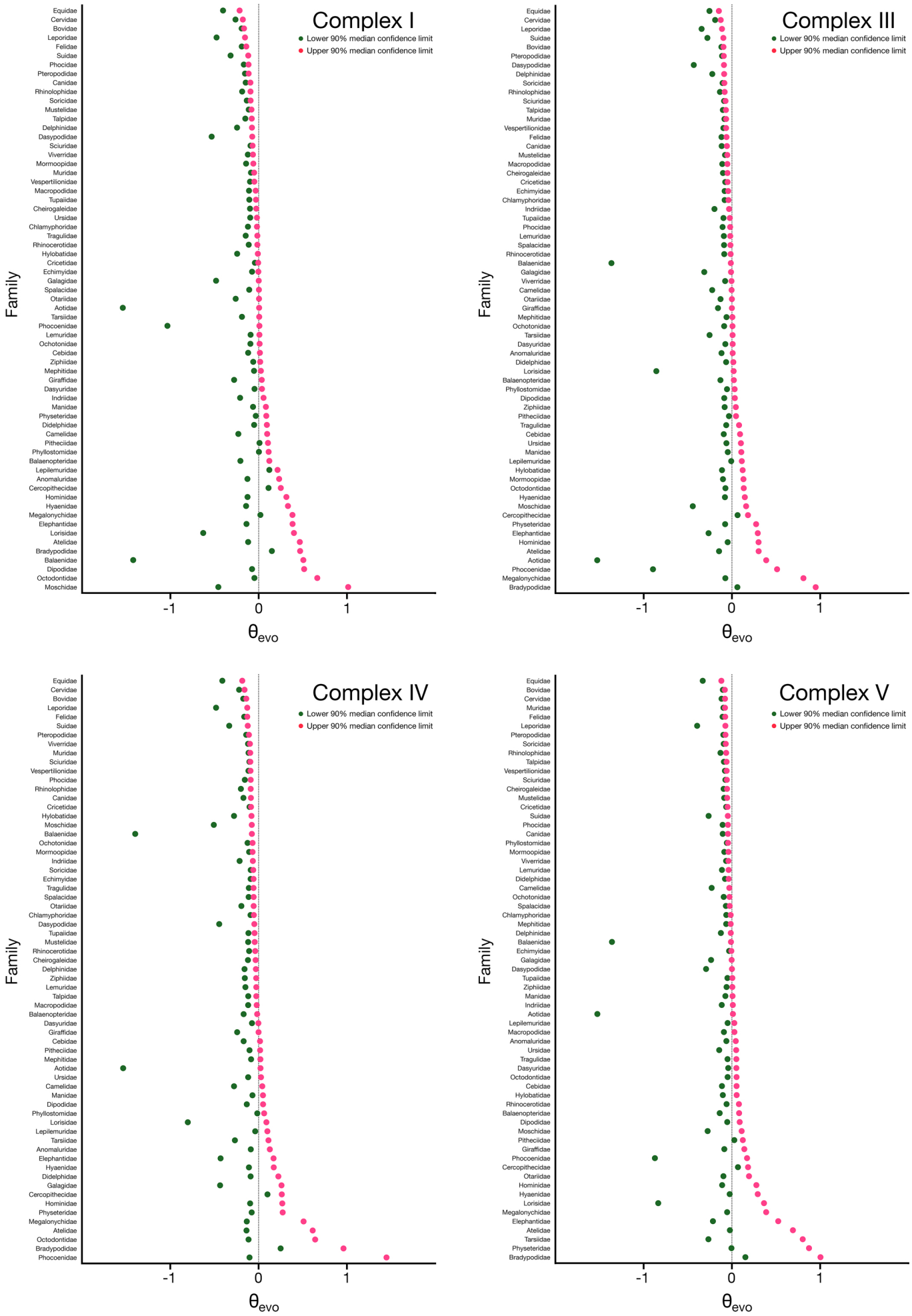
Supplemental Figure S3: Mammalian orders differ in their propensity for potentially efficacious mitochondrial protein substitutions at specific OXPHOS complexes (median confidence intervals). Analysis was performed as in (**a**) Figure 2b or (**b**) Figure 2c, except that θ_{evo} values were obtained by analysis of Complex I, Complex III, Complex IV, or Complex V proteins.



Supplemental Figure S4: Mammalian families differ in their propensity for potentially efficacious mitochondrial protein substitutions at specific OXPHOS complexes (median calculations). Analysis was performed as in Figure 3a, except that θ_{evo} values were obtained by analysis of Complex I, Complex III, Complex IV, or Complex V subunits.



Supplemental Figure S5: Mammalian families differ in their propensity for potentially efficacious mitochondrial protein substitutions at specific OXPHOS complexes (median confidence intervals ordered by lower 90% median confidence limit). Analysis was performed as in Figure 3b, except that θ_{evo} values were obtained by analysis of Complex I, Complex III, Complex IV, or Complex V proteins.



Supplemental Figure S6: Mammalian families differ in their propensity for potentially efficacious mitochondrial protein substitutions at specific OXPHOS complexes (median confidence intervals ordered by upper 90% median confidence limit). Analysis was performed as in Figure 3c, except that θ_{evo} values were obtained by analysis of Complex I, Complex III, Complex IV, or Complex V polypeptides.

Supplemental File 1: All predicted protein substitutions along all edges at positions containing less than 2% gaps across input and ancestral sequences are listed, along with associated inferred taxonomy information, TSS, and branch length. All alignment positions refer to the *Bos taurus* reference sequences.

Supplemental File 2: The TSS value calculated for each mitochondrial protein alignment position using the mammalian reference sequence dataset is provided. All alignment positions refer to the *Bos taurus* reference sequences.

Supplemental File 3: SPCS and θ_{evo} outputs are provided for analyses across all mitochondria-encoded positions, as well as for focused analyses of specific OXPHOS complexes and individual proteins.

Supplemental File 4: A GenBank flat file containing RefSeq entries for mammalian mtDNAs and the entry for the reptile *Anolis punctatus*.

Supplemental File 5: A maximum likelihood inferred tree generated by RAxML-NG analysis of concatenated and aligned protein coding sequences from mammalian and *Anolis punctatus* mtDNAs.

Supplemental File 6: Bootstrap replicates were generated from the alignment of concatenated protein coding sequences. Felsenstein's Bootstrap Proportions (Felsenstein 1985) were calculated and used to label the maximum likelihood inferred tree of mammalian mtDNAs.

Supplemental File 7: Bootstrap replicates were generated using concatenated mammalian mtDNA coding sequences. Transfer Bootstrap Expectations (Lemoine et al. 2018) were calculated and used to label the maximum likelihood inferred tree of mammalian mtDNAs.

Supplemental File 8: PAGAN tree output using aligned amino acid sequences and the rooted maximum likelihood inferred tree as input.