The mitochondrial genome of the mountain wooly tapir, *Tapirus pinchaque* and a formal test of the effect of altitude on the adaptive evolution of mitochondrial protein coding genes in odd-toed ungulates

Edgar G. Gutiérrez[1], Jorge Ortega[1], Avery Savoie[2] and J. Antonio Baeza[2,3,4]*

Abstract

Background The harsh conditions of high-altitude environments are known to drive the evolution of physiological and morphological traits in endothermic animals. These conditions are expected to result in the adaptive evolution of protein coding genes encoded in mitochondrial genomes that are vital for the oxidative phosphorylation pathway. In this study, we formally tested for signatures of adaptive evolution on mitochondrial protein coding genes in *Tapirus pinchaque* and other odd-toed ungulates inhabiting high-elevation environments.

Results The AT-rich mitochondrial genome of *T. pinchaque* is 16,750 bp long. A phylomitogenomic analysis supports the monophyly of the genus *Tapirus* and families in the Perissodactyla. The ratio of non-synonymous to synonymous substitutions demonstrated that all mitochondrial genes undergo purifying selection in *T. pinchaque* and other odd ungulates living at high elevations. Over this negative background selection, Branch Models suggested that *cox3* and *nad6* might be undergoing stronger purifying selection than other mitochondrial protein coding genes. Furthermore, Site Models suggested that one and four sites in *nad2* and *nad5*, respectively, could be experiencing positive selection. However, these results were supported by Likelihood Ratio Tests but not Bayesian Empirical Bayes posterior probabilities. Additional analyses (in DataMonkey) indicated a relaxation of selection strength in *nad6*, evidence of episodic diversifying selection in *cob*, and revealed episodic positive/diversifying selection signatures for two sites in *nad1*, and one site each in *nad2* and *nad4*.

Conclusion The mitochondrial genome of *T. pinchaque* is an important genomic resource for conservation of this species and this study contributes to the understanding of adaptive evolution of mitochondrial protein coding genes in odd-toed ungulates inhabiting high-altitude environments.

Keywords Adaptive evolution, Mitochondrial genes, Odd-toed ungulates, Selection, Selective pressure, *Tapirus pinchaque*

*Correspondence:
J. Antonio Baeza
baeza.antonio@gmail.com
Full list of author information is available at the end of the article

Gutiérrez *et al. BMC Genomics*   (2023) 24:527

Page 2 of 21

Background

During their evolutionary radiation, mammals have colonized a wide variety of environments, including high-altitude habitats characterized by extreme conditions, such as low atmospheric pressure, reduced oxygen (including hypoxia), high solar radiation, and low temperatures [1, 2]. Consequently, high-altitude mammals exhibit a set of specialized morphological and physiological traits that have been considered adaptations to the harsh conditions characteristic of these environments [3]. Typical convergent putative adaptations in high-altitude organisms include, among others, small body sizes (relative to their closely related species) that decrease energy demand, and dense long coats to minimize heat dissipation (e. g., in some sheep breeds) [3–5]. Physiological strategies to cope with hypoxic conditions at high altitudes include hypoxia-tolerance coupled with low metabolic rates to reduce oxygen demand [6, 7] and high hemoglobin concentrations plus increased hemoglobin binding-affinity for oxygen, as reported in highland mammals such as the yak [8], the alpaca [9], and the deer mouse [10], among others [11]. Most recently, specialized endothelial cells in the lungs of yaks have been discovered that might help individuals to thrive in the cold, low-oxygen conditions characteristic of high-altitude environments [12].

The genomic underpinnings of adaptation to high altitude environments are not well understood but have been extensively studied during the last decade (in the yak; *Bos grunniens*, the Tibetan gray wolf; *Canis lupus chanco*, and sheep; *Ovis aries*) [12–15]. For instance, in the nuclear genome, candidate genes that have contributed to adaptation to high altitudes include EPAS1 (endothelial PAS domain protein 1; also known as HIF-2A) and EGLN1 (egl-9 family hypoxia inducible factor 1; also known as HIF-prolylhydroxylase 2). These two genes are involved in the response to hypoxic stress [14, 15] and have been shown to experience strong signatures of positive selection [7, 11, 14, 16]. The EPAS1 gene directly regulates key genes such as erythropoietin and vascular endothelial growth factor, while the EGLN1 gene negatively regulates the activity of hypoxia-inducible factor-1 alpha (HIF-1A), a transcriptional complex that plays a central role in oxygen homeostasis in mammals [14, 17]. Adaptive evolution in mitochondrial genome (mtDNA) of mammals inhabiting high-altitude environments has been explored in a few species (Tibetan horses [18], alpacas [19], and camelids [20], among a few others).

Mitochondria are responsible for supplying cellular energy to the cells in the form of adenosine triphosphate (ATP) through the oxidative phosphorylation system (OXPHOS) carried out by the electron transport chain and ATP synthase [21]. Furthermore, ATP is used for generating heat and maintaining body temperature in endothermic animals, including mammals [1]. Importantly, hypoxic conditions, characteristic of high-altitude environments, are known to affect physiological processes involving metabolic performance and increases in Reactive Oxygen Species (ROS) [22]. ROS produces superoxide $(O_2.^-)$ and hydrogen peroxide (H_2O_2) that triggers oxidative damage to lipids, proteins, and deoxyribonucleic acid (DNA) [23, 24]. Similarly, hypoxia inhibits the transcription of mitochondrial DNA and damages the structure and function of mitochondria. Consequently, mitochondrial protein-coding genes are expected to experience selective pressures in species that are continuously exposed to harsh environmental conditions [2, 25, 26]

The mammalian mtDNA is a circular, double-stranded molecule that hosts its own genetic material [23, 27]. Invariably mammalian mtDNA contains 22 transfer RNA genes (tRNA), two ribosomal genes (rRNA), and 13 protein coding genes (PCG´s). The latter genes encode polypeptides that have a key role in the OXPHOS [2, 19]. Specifically, these OXPHOS-related polypeptides (7 subunits of the NADH dehydrogenase complex, 3 subunits of the cytochrome c oxidase, 2 subunits of ATP synthase, and the cytochrome b subunit of the cytochrome bc1 complex) are involved in electron transfer and aerobic respiration [22]. Changes in mitochondrial function driven by mutations in genes involved in oxidative phosphorylation, are likely essential for adaptation to adverse conditions (including hypoxia) typical of high-altitude environments [23].

The order Perissodactyla comprises the odd-toed ungulates and contains three families: Equidae, Rhinocerotidae, and Tapiridae [28]. The family Tapiridae comprises a single extant genus, *Tapirus*, represented by four extant species [29]; two exclusively distributed in South America (*Tapirus terrestris* and *Tapirus pinchaque*); *T. bairdii* which is distributed in Southern Mexico, Central America, and the Pacific coast of Colombia, and *Tapirus indicus*, that inhabits Southern Thailand and through peninsular Malaysia to Indonesia [30–32]. A fifth species of *Tapirus* was proposed nearly a decade ago [33]; however, analysis using mitochondrial markers (i. e., *cox1, cox2* and *cob* genes) demonstrated that haplotypes from individuals identified as *T. kabomani* clustered together with haplotypes from individuals identified as *T. terrestris*, and haplotypes did not segregate according to putative species [34]. Thus, *T. kabomani* is currently considered a synonym of *T. terrestris*.

Among tapirs, the mountain or wooly tapir, *Tapirus pinchaque* inhabits the tropical montane Andean forests and 'paramos' between 2,000 and 4,800 m above sea level (masl) in Colombia, Ecuador, and northern Peru [35, 36] (Fig. 1). By contrast to *T. pinchaque*, all other congeneric

Gutiérrez *et al. BMC Genomics*   (2023) 24:527

Page 3 of 21

	T. terrestris	*T. indicus*	*T. bairdii*	*T. pinchaque*
Body length (m)	1.80 - 2.21	1.85 - 2.40	1.80 - 2.50	1.8 max
Body weight (kg)	180 -320	250 - 400	150 - 300	136 - 250
Coat variation	Short and smooth	Short and smooth	Short and smooth	Long, thick, and woolly
Altitude variation (masl)	Up to 1800	Up to 2000	0 to 3500	2000 to 4800
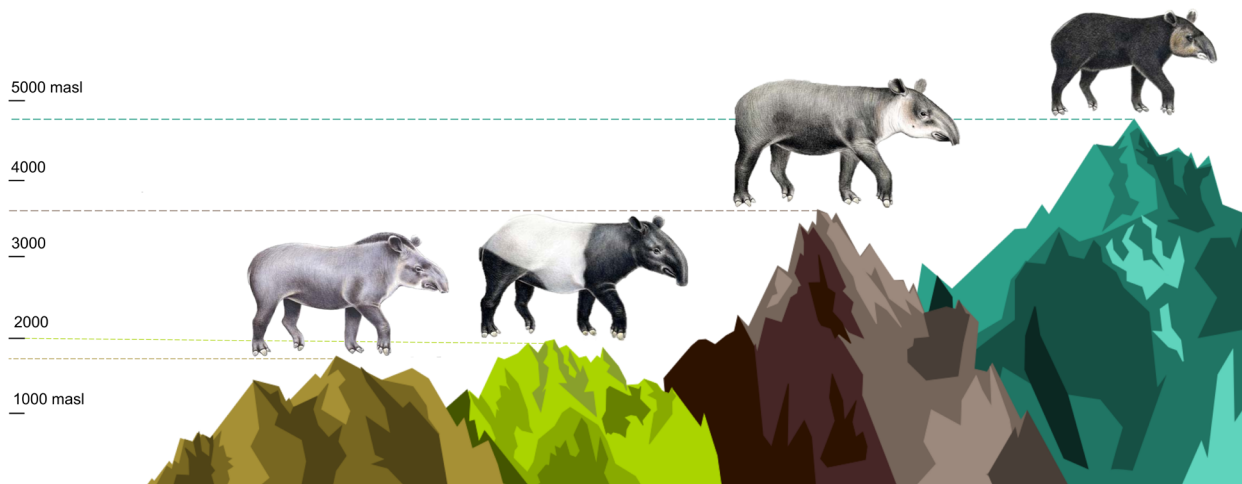
Fig. 1 Ecological and phenotypic characteristics of *Tapirus pinchaque* and congeneric species. Tapir species ilustration credit: Stephen Nash (published with permission)

species primarily prefer relatively lower altitude habitats, e. g., *Tapirus terrestris* inhabits lowland forests up to 1,800 masl [37, 38] while *T. indicus* can be found up to 2,000 masl, with a greater presence in lower altitude forests (125 to 1000 m) [39]. Lastly, the distribution of *T. bairdii* is limited to the states of Campeche, Chiapas, Oaxaca, Quintana Roo, and Veracruz in Mexico, from sea level to altitudes not higher than 2000 masl [40]. However, much of the literature describes the distribution in Central America from sea level up to 3600 m [41–43]. Other than a white line around its lips, *T. pinchaque* is distinguished from other relatives by a thick wooly coat (composed of 2–4 cm long hair) and the presence of a thick inner layer of fat [35, 44] that might well be interpreted as adaptations to high altitude. T*apirus pinchaque* is also the smallest species (1.8 m maximum length) [45] in the genus (*T. terrestris*, *T. bairdii*, and *T. indicus* range in body length between 1.8 to 2.5 m in adult individuals) [37, 40, 46] (Fig. 1). The smaller size of *T. pinchaque* is probably related to lower energy expenditure, which has been suggested to be another adaptive phenotypic trait for living at high altitudes [4]. Furthermore, mountain tapirs often swim and wallow in water bodies, a behavior likely used to cool their bodies due to high radiation

during the day at high altitude [47]. The genomic basis of such adaptations in *T. pinchaque*, however, remains unknown. We focus on the mtDNA and we are particularly interested in studying the effect of altitude on the adaptive evolution of protein coding genes in the mountain tapir and odd-toed ungulates.

This study aims to explore adaptive evolution in the mtDNA of *T. pinchaque* and other species belonging to the order Perissodactyla that inhabit high-altitude environments. To achieve the aforementioned goal, first, we assembled and characterized in detail the complete mitochondrial genome of the mountain tapir *Tapirus pinchaque* profiting from low-pass (=low coverage) next generation short-read sequencing. We describe the mitochondrial genome of this species in detail following suggestions in Baeza [48]. Specifically, the nucleotide composition of the entire mtDNA, the codon usage of PCG′s, the secondary structure of each tRNA, the secondary structure and tandem repeats within the putative control region (CR) were calculated, and then compared against other congeneric extant species. After an analysis of mitogenomic features in the genus *Tapirus*, we explored the phylogenetic position of *T. pinchaque* among extant congeneric species and generated

Gutiérrez *et al. BMC Genomics*   (2023) 24:527

Page 4 of 21

a phylogenetic hypothesis for the Perissodactyla based on mitochondrial protein-coding genes. Lastly, we used a set of bioinformatic tools to estimate the ratio ω of non-synonymous (d_N) to synonymous (d_S) substitutions rate ($ω = d_N/d_S$) in order to identify signatures of adaptive evolution (i. e., positive selection) on PCG´s in members of the Perissodactyla adapted to live in high altitude environments. High-altitude species, including *T. pinchaque* were compared against species living below 2,000 masl, using Branch Models (BM), Site Models (SM), and Branch-Site Models (BSM), along sequences or across branches of a phylogeny [49, 50]. We predict that the protein coding genes in the mitocondrial genome of *T. pinchaque* and other species belonging to the order Perissodactyla that inhabit high-altitude environments exhibit signatures of positive selection compared to closely related species inhabiting low-altitude environments. The genomic resource generated in this study and analyses provide new insights into the molecular mechanisms responsible for adaptation to high-altitude environments in *T. pinchaque* and other representatives of Perissodactyla.

Importantly, other than its peculiar high-altitude lifestyle, *T. pinchaque* is considered an important member of its community: more than 50 species of seeds have been found in feces from individuals inhabiting the Ecuadorian Andes highlighting the putative role of this large mammal in seed dispersal [34, 51]. Despite the potentially crucial ecological role of this species, there is a very small number of genetic studies for *T. pinchaque* [34]. Thus, the assembly and detailed description of its mitochondrial genome coupled with the formal test of adaptive evolution in mitochondrial protein coding genes in this and other odd-toed ungulates represents a step forward to improve genomic resources that can aid in the conservation of this species currently experiencing major environmental issues [52, 53].

Results and discussion

The pipeline GetOrganelle [54] assembled (with an average coverage of 169.5×per nucleotide) and circularized the complete mitochondrial chromosome (GenBank accession no. OQ420428) of *T. pinchaque* (Fig. 2). The complete sequence of *T. pinchaque* mtDNA is 16,750 bp long, and exhibits a length similar to that found in *T. terrestris* (16,772 bp) [55], *T. bairdii* (16,697 bp) [56], *T. indicus* (16,717 bp) [57], as well as in other members of the order Perissodactyla (e. g., Rhinocerotidae: *Ceratotherium simum* – 16,832 bp and *Rhinoceros unicornis* – 16,829 bp; Equidae: *Equus asinus* – 16,813 bp and *Equus caballus* – 16,504 bp) [58–61].

The mitochondrial genome of *T. pinchaque* encodes 13 PCG´s, two rRNA genes (rrnS [12S ribosomal RNA]

and rrnL [16S ribosomal RNA]), 22 tRNAs, and exhibits a relatively long (1,302 bp) non-coding putative CR or D-Loop (Table 1). In the mtDNA of *T. pinchaque*, all PCG´s are encoded in the positive or heavy strand (H), except *nad6* which is located in the negative or light strand (L) (Fig. 2). Equally, the two rRNA's and 14 tRNA genes are encoded in the H strand. The remaining eight tRNA genes are located in the L strand (Fig. 2). The gene order herein reported for *T. pinchaque* is identical to that previously observed in the congeneric *T. terrestris* [55], *T. bairdii* [56], and *T. indicus* [57]. Likewise, mitochondrial synteny in the genus *Tapirus* is identical to that reported for members of the closely related families Equidae (e. g., Tibetan wild ass; *Equus kiang* and Qingyang donkey; *Equus asinus*) [62, 63] and Rhinocerotidae (e. g., *Rhinoceros sondaicus* and *Diceros bicornis*) [64] with complete mitochondrial genomes available in GenBank.

Nucleotide usage in the complete mitochondrial sequence of *T. pinchaque* was as follows: T = 27.98%, C = 25.25%, A = 34.36%, G = 12.41%. The G + C content was 37.66%, while the A + T content was 62.3%. The aforementioned A + T content value is similar to those previously reported in the genus *Tapirus* (i. e., *T. terrestris*; 62.22%, and *T. bairdii*; 62.00%; Table 2). The A + T content in *T. indicus* (NC023838) has not been reported [57]. Herein, we estimated nucleotide usage for *T. indicus*, that exhibits the lowest AT-content (61.9%) compared to the rest of the congeneric species. A somewhat lower AT-content has been reported in representatives of the sister clade Rhinocetoridae: *Rhinoceros unicornis* (59.8%) [59], *Rhinoceros sondaicus* (59.8%) [64], and *Dicerorhinus sumatrensis* (58.4%) [65]. The AT-rich nucleotide usage in *Tapirus* and allies is often assumed to be driven by the asymmetric nature of the replication process in mtDNA [66]. Furthermore, AT-content might be influenced by differential mutation pressure at synonymous and non-synonymous sites in PCG´s [67].

In *Tapirus pinchaque*, 10 PCG´s used ATG as the start codon and the remaining three PCG´s (*nad2*, *nad3*, and *nad5*) used ATA as start codons. Eight of the PCG´s used the conventional TAA stop codon and *cob* gene used AGA as stop codon. The remaining PCG´s (*nad2*, *cox3*, *nad3*, and *nad4*) ended with an incomplete stop codon (T). The relative abundance of start and stop codons in PCG´s of *T. pinchaque* agrees with that previously reported for other species of *Tapirus* (i. e., *T. terrestris*, *T. bairdii*, and *T. indicus*) [55–57] as well as for other representatives of the order Perissodactyla [60, 61]. Similarly, incomplete stop codons (T), like those found in four PCG´s (*nad2*, *cox3*, *nad3*, and *nad4*), have been previously observed in *T. terrestris* [55], *T. bairdii* [56], *T. indicus* [57], as well as in *Ceratotherium simum* [58], in

Gutiérrez *et al. BMC Genomics*   (2023) 24:527

Page 5 of 21

Fig. 2 Circular representation of the complete mitochondrial genome in *Tapirus pinchaque*. The 13 PCG´s, 22 tRNA's, 2 rRNA's, and the control region are annotated. Photo credit: David Sifry (published with permission)

Rhinoceros unicornis [59], and in other odd-toed ungulates (e. g., *Equus africanus somalies*) [68].

In the mitochondrial PCG´s of *T. pinchaque*, the eight codons with the highest Relative Synonymous Codon Usage (RSCU) values (=overrepresented codons) were CTA(Leu) (2.77), TCA(Ser) (2.66), CGA(Arg) (2.48), GTA(Val) (2.23), ACA(Thr) (2.10), TGA(Trp) (1.94), CAA(Gln) (1.93), AAA(Lys) (1.92) (Fig. 3, Supplementary Table S1). In turn, underrepresented codons included CGG(Arg) (0.06), TGG(Trp) (0.06), GCG(Ala) (0.07), CAG(Gln) (0.07), and AAG(Lys) (0.08) (Fig. 3, Supplementary Table S1). This is the second time that a

detailed analysis of codon usage has been conducted in the family Tapiridae; the first analysis was conducted in *Tapirus bairdii*, and our results are consistent with those reported there [56]. No studies focusing on codon usage of mitochondrial PCG´s were found in other Perissodactyla species, however, the results described above are consistent with studies in other mammals [69, 70]. It should be noted that the RSCU results showed that most of the overrepresented codons (RSCU > 1.9) are A-ending, while underrepresented codons (RSCU < 0.06) are G-ending. The result of the RSCU analysis supports the notion that nucleotide usage is affected by patterns of mutational

Table 1 Mitochondrial genome of *Tapirus pinchaque*. Arrangement and annotation. * = incomplete stop codon

Name	Type	Start	Stop	Strand	Length (bp)	Start	Stop	Continuity
trnF (ttc)	tRNA	1	68	+				0
rrnS	rRNA	69	1037	+				-1
trnV(gta)	tRNA	1037	1103	+				-2
rrnL	rRNA	1102	2682	+				0
trnL2(tta)	tRNA	2683	2757	+				2
nad1	PCG	2760	3716	+	957	ATG	TAA	0
trnI(atc)	tRNA	3716	3785	+				-3
trnQ(caa)	tRNA	3783	3855	-				2
trnM(atg)	tRNA	3858	3926	+				0
nad2	PCG	3927	4968	+	1042	ATA	Taa*	0
trnW(tga)	tRNA	4969	5037	+				5
trnA(gca)	tRNA	5043	5111	-				1
trnN(aac)	tRNA	5113	5185	-				32
trnC(tgc)	tRNA	5219	5284	-				0
trnY(tac)	tRNA	5285	5351	-				1
cox1	PCG	5353	6897	+	1545	ATG	TAA	-3
trnS2(tca)	tRNA	6895	6963	-				7
trnD(gac)	tRNA	6971	7037	+				0
cox2	PCG	7038	7721	+	684	ATG	TAA	3
trnK(aaa)	tRNA	7725	7791	+				1
atp8	PCG	7793	7996	+	204	ATG	TAA	-43
atp6	PCG	7954	8634	+	681	ATG	TAA	-1
cox3	PCG	8634	9417	+	784	ATG	Taa*	0
trnG(gga)	tRNA	9418	9486	+				0
nad3	PCG	9487	9832	+	346	ATA	Taa*	1
trnR(cga)	tRNA	9834	9901	+				0
nad4l	PCG	9902	10,198	+	297	ATG	TAA	-7
nad4	PCG	10,192	11,569	+	1378	ATG	Taa*	0
trnH(cac)	tRNA	11,570	11,638	+				0
trnS1(agc)	tRNA	11,639	11,697	+				0
trnL1(cta)	tRNA	11,698	11,767	+				6
nad5	coding	11,774	13,588	+	1815	ATA	TAA	-17
nad6	coding	13,572	14,096	-	525	ATG	TAA	3
trnE(gaa)	tRNA	14,100	14,168	-				5
cob	coding	14,174	15,313	+	1140	ATG	AGA	0
trnT(aca)	tRNA	15,314	15,382	+				0
trnP(cca)	tRNA	15,383	15,448	-				
CR		15,449	16,750	+				

bias or natural selection, being a driving force of variation in codon usage [71].

The mtDNA of *T. pinchaque* encoded 22 tRNA genes. The tRNA's ranged in length from 59 bp (tRNA-Ser1) to 75 bp (tRNA-Leu2). The secondary structures of the tRNA's detected here (Fig. 4), were very similar to those reported in *T. bairdii* [56]. All tRNA in the *T. pinchaque* mtDNA, except for tRNA-Ser1, exhibited a standard 'cloverleaf' secondary structure as indicated by the web server Forna [72]. In *T. pinchaque*, the tRNA-Ser1 has a missing D-arm, and subsequently, it is a relatively short gene (59 bp in length) compared to the other tRNA lengths. Importantly, the loss of stem-loop structures in the D-arm of the tRNA-Ser1 gene appears to be a common feature reported in eumetazoans, including mammals [73]. This is the second study in which the secondary structures of tRNA's are described in the family Tapiridae.

Table 2 Nucleotide usage values in different Perissodactyla species, including the *Tapirus* species analyzed in this study

Species	T(U)%	C%	A%	G%	Length (bp)
Tapirus pinchaque	27.9	25.3	34.4	12.4	16,750
Tapirus indicus	27.2	25.8	34.6	12.3	16,717
Tapirus bairdii	29.8	25.6	32.2	12.3	16,697
Tapirus terrestris	27.8	25.4	34.4	12.4	16,766
Rhinoceros unicornis	26.2	27.5	33.7	12.7	16,829
Dicerorhinus sumatrensis	25.3	28.3	33.1	13.3	16,466
Diceros bicornis	25.9	27.9	33.5	12.6	16,594
Rhinoceros sondaicus	26.3	27.4	33.5	12.8	16,315

The *rrnS* (12S) and *rrnL* (16S) genes in the mitochondrial genome of *T. pinchaque* are 969 and 1,579 bp in length, respectively. The 12S gene is located between the *trnF* and *trnV* genes and exhibits a nucleotide content equal to A = 38.6%, T = 23.1%, C = 22.1%, G = 16.2% while the 16S gene is located between *trnV* and *trnL2* and has a nucleotide content equal to A = 38.3%, T = 24.6%, C = 21.5%, G = 15.6%. The two rRNA genes are AT-rich (*rrnS* = 61.7% and *rrnL* = 62.9%). These AT-rich values are

consistent with those previously reported for *T. terrestris* (12S = 61.9% and 16S*rrnL* = 62.9%) [55] and *T. bairdii* (12S = 61.7% and 16S = 62.6%) [56]. It should be noted that the nucleotide composition for the 12S gene in *T. indicus* is A = 38.6%, C = 22.6%, T = 22.1%, and G = 16.7% [57]. Therefore, this gene is AC-rich (61.2%), a genomic feature that differs from that reported in other species of *Tapirus*, including *T. pinchaque*. The 16S gene of *T. indicus* exhibits an AT-content (63.5%) [57] similar to that reported for other species of *Tapirus*. Interestingly, several Old World odd-toed ungulates exhibit a very similar nucleotide use in both rRNA genes, such as in *Rhinoceros unicornis* (12S = 61.4% A + C, and 16S = 63.1% A + T) [59], *Rhinoceros sondaicus* (12S = 60.9% A + C, and 16S = 62.8% A + T) [64], and in *Dicerorhinus sumatrensis* (12S = 62.0% A + C, and 16S = 61.0% A + T) [65].

The length of the CR in the mitochondrial sequence of *T. pinchaque* is 1,301 bp long and exhibits a nucleotide composition equal to A = 33.8%, T = 28.3%, C = 25.1%, and G = 12.7%. We noted that the length of the control region in *T. pinchaque* is slightly longer than that of *T. bairdii* (1,247 bp) [56] and *T. indicus* (1,268 bp) [57]. The length of this non-coding region varies moderately in the order Perissodactyla and ranges from a minimum length
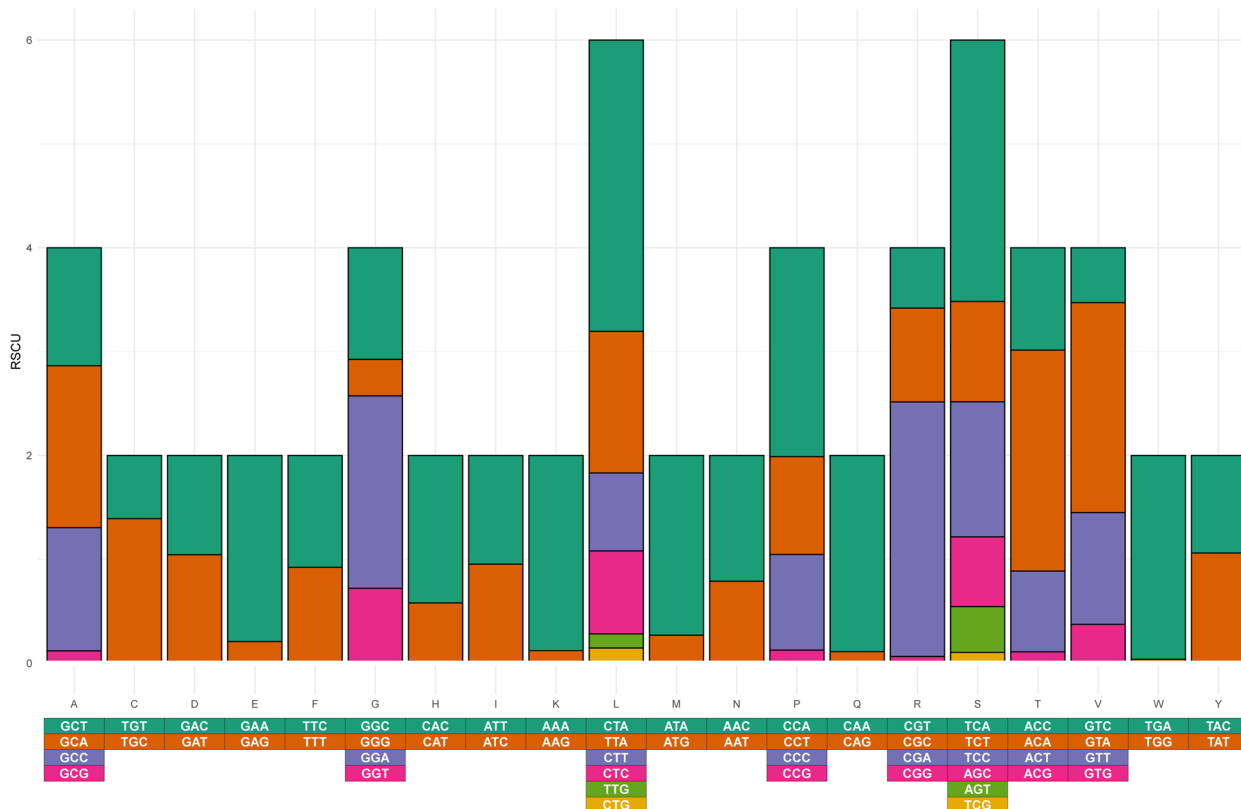
Fig. 3 Relative synonymous codon usage (RSCU) analysis in all 13 mitochondrial protein-coding genes of *Tapirus pinchaque*

Gutiérrez *et al. BMC Genomics*   (2023) 24:527

Page 8 of 21
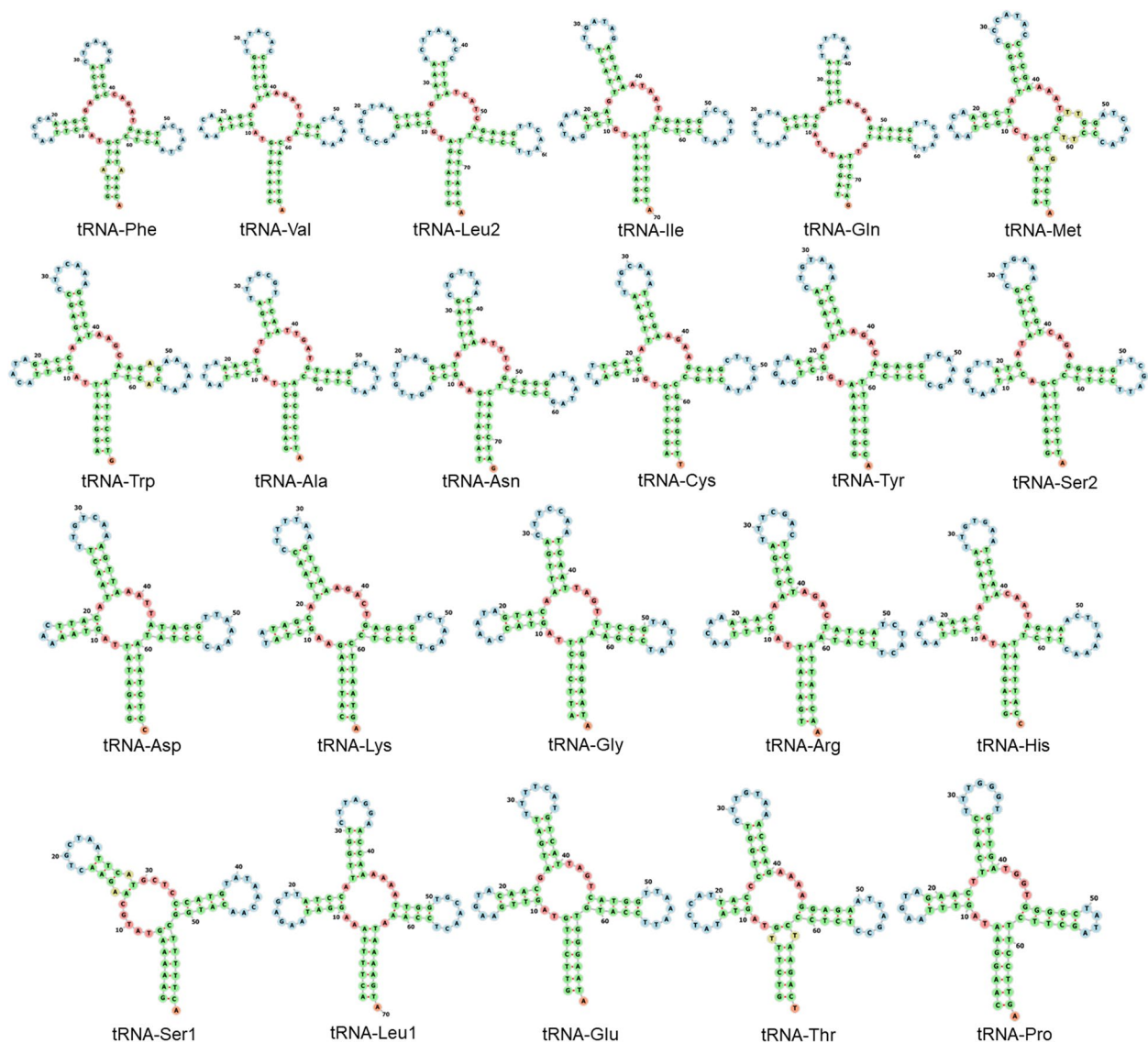
Fig. 4 Secondary structure of the 22 tRNA genes present in the mitochondrial genome of *Tapirus pinchaque*

of 1,086 bp in the Turkish Anatolian donkey; *Equus asinus* [74] to a maximum of 1,376 bp in the Indian rhinoceros; *R. unicornis* [59]. The length of the CR is known to vary extensively among mammals [75], and the differences in length are likely driven by the fast mutation rate reported for this region, especially when compared to that of mitochondrial PCG´s [76].

Within the CR of *T. pinchaque*, eight microsatellites were detected, most of them dinucleotide motifs usually repeated three times (Supplementary Table S2). The existence of microsatellite repeats in the CR is characteristic of the mitochondrial genome in mammals [75, 77]. A tandem repeat sequence in the CR [5'-(ACA TAC GTA TAC)$_{26}$–3' motif] was found to begin at position 16,143

and end at 16,456 of the complete mtDNA sequence (Fig. 5, Supplementary Table S3). Tandem repeats are also reported in the congeneric species *T. indicus* and *T. bairdii* [56, 57] as well as in other representatives of the Perissodactyla (e. g., *Equus caballus* and *Rhinoceros unicornis*) [59, 78]. We note that all these tandem repeats exhibit purine/pyrimidine alternation, a feature conserved in mammalian mtDNA's [59, 74, 77]. A total of 20 possible secondary structures, all of them containing variable number and sizes of stem-loops (Supplementary Fig. S1), were revealed for the CR of *T. pinchaque*. The Gibbs free energy [ΔG] of these 20 structure predictions ranged in values between [ΔG] = -288.9 kcal/mol to [ΔG] = -287.6 kcal/mol (Supplementary Fig. S1).

The three functional domains within the CR, commonly reported in mammals [79], were recognized in *T. pinchaque* (Fig. 5). These are the central conserved domain and two flanking variable domains, namely the Extended Termination Associated Sequence (ETAS) domain and the Conserved Sequence Block (CSB) domain. The CSB, Central, and ETAS domains are 654, 334, and 315 bp in length, respectively. These domain lengths are within the range of those found in other tapir species, such as *T. indicus* and *T. bairdii* [56, 57].

The ETAS domain, located at the 5'-end of the CR, is A + T rich (63.2%) and contain extended ending associated sequence 1 (ETAS-1) and a second sequence homologous to the termination-associated sequence 2 (ETAS-2; Fig. 4). Within the ETAS-1 sequence, a functional conserved motif 5'-GCCCCAT-3' was identified (Fig. 5), which is known as a D-loop stop point [80]. Previous studies have also identified this motif in another mammalian CR [81, 82]. The Central domain sequence is also A + T rich (57.2%). A high degree of conservation was observed in the Central domain of the mtDNA of *T. pinchaque* and other species of *Tapirus* as well as other members of the Perissodactyla whose CR structure has been analyzed in detail [56, 79]. The CSB domain, positioned at the 5'-end of the CR, is A + T rich (64.2%) and contains three conserved sequence blocks (CSB-1, CSB-2, and CSB -3) that have been suggested to be functionally important for mtDNA replication and transcription [83]. We note that the relatively long tandem repeat [with motif: 5'-(ACATACGTATAC)$_{26}$–3'] is located between the CSB-1 and CSB-2 regions (Fig. 4), as reported before in other odd-toed ungulates [56, 57, 78]. In general, detailed analyses of the CR in the order Perissodactyla are rare. Additional studies focusing on the CR organization are needed to further improve the understanding of the function of this region in mammals and beyond.

Phylomitogenomics of odd-toed ungulates

Our phylogenetic analysis using a Maximum Likelihood approach confirmed the monophyly of the order Perissodactyla; all representatives of the family Tapiridae, including *T. bairdii*, *T. indicus*, *T. terrestris*, and *T. pinchaque*, all members of the family Equidae, and all species belonging to the family Rhinocerotidae used in our analysis clustered together into a single fully supported (bootstrap value [bv] = 100) clade (Fig. 6). In the Perissodactyla, the family Equidae occupied an early branching position; it was sister to a second clade comprising members of the families Rhinocerotidae and Tapiridae (bv = 91). Within the family Tapiridae, *T. indicus* occupied an early branching position; it was sister to a second fully supported monophyletic clade (bv = 100) comprising all three remaining species of *Tapirus*. In the latter clade, *T. bairdii* was sister to a well-supported clade (bv = 98) containing *T. terrestris* and *T. pinchaque*. The phylogenetic relationships among the different species of Equidae and Rhinocerotidae are identical to those found by previous studies [56, 64, 84].

The effect of altitude on mitochondrial adaptive evolution

In our first selective pressure analysis, all the Ka/Ks ratios for all the 13 PCG´s of *T. pinchaque* exhibited values < 1, suggesting stronger purifying selection and evolutionary constraints in these mitochondrial genes (Fig. 7). Our results of the Ka/Ks comparisons between *T. pinchaque* and the other tapir species showed that *atp8* (Ka/Ks < 0.4; $p > 0.05$) and *nad6* (Ka/Ks < 0.35; $p < 0.05$) were the genes with the highest Ka/Ks values (Fig. 7). These results could be explained by the action of less intense purifying selection in *atp8* and *nad6* compared to the remaining mitochondrial PCG´s [85]. Regarding the lowest values (Ka/Ks ≤ 0; $p ≤ 0.05$), they were observed in the *nad4l* and in the cytochrome c oxidase subunit genes (*cox1*, *cox2*, and *cox3*; Fig. 7). Therefore, we suggest that the last four mentioned genes experience the strongest purifying selection. The results of the selective pressure analyses reported here are consistent with the results in all the PCG´s reported for *Tapirus bairdii* [56]. Until now, there have been few analyses of selective pressure in mitochondrial PCG´s within the family Tapiridae. Similarly, mitochondrial PCG´s selective pressure analyses are rare in the order Perissodactyla. A single previous report found a similar pattern of purifying selection of mitochondrial PCG´s in members of the genus *Equus* [86]. In general, previous research in mammals is also in agreement with the general pattern of mtDNA genes experiencing purifying selection, e. g., the representatives of the Caprini tribe [1], and several species of the family Mustelidae [87]. Our results agree with the hypothesis that purifying selection constraints, in general, the evolution of PCG´s, probably to maintain functionality of the OXPHOS pathway [19, 88].

In the second analysis performed to examine for signatures of positive selection in all 13 mitochondrial

(See figure on next page.)

Fig. 5 Organization and structure of the control region in the mitochondrial genome of *Tapirus pinchaque*. **a** Visual representation of the CR and the three functional domains (ETAS, central, and CSB domains). **b** Nucleotide sequence of the complete control region; locations of the ETAS-1 and ETAS-2 (underlined in green), the functional conserved motif 5'-GCCCCAT-3' (in bold); the large highly conserved regions within the central domain (underlined in purple), as well as the CSB1, CSB2, and CSB3 blocks (underlined in orange) are shown. The long repetitive motif is indicated by the dotted green line and **c**) A possible secondary structure for this repetitive region is represented
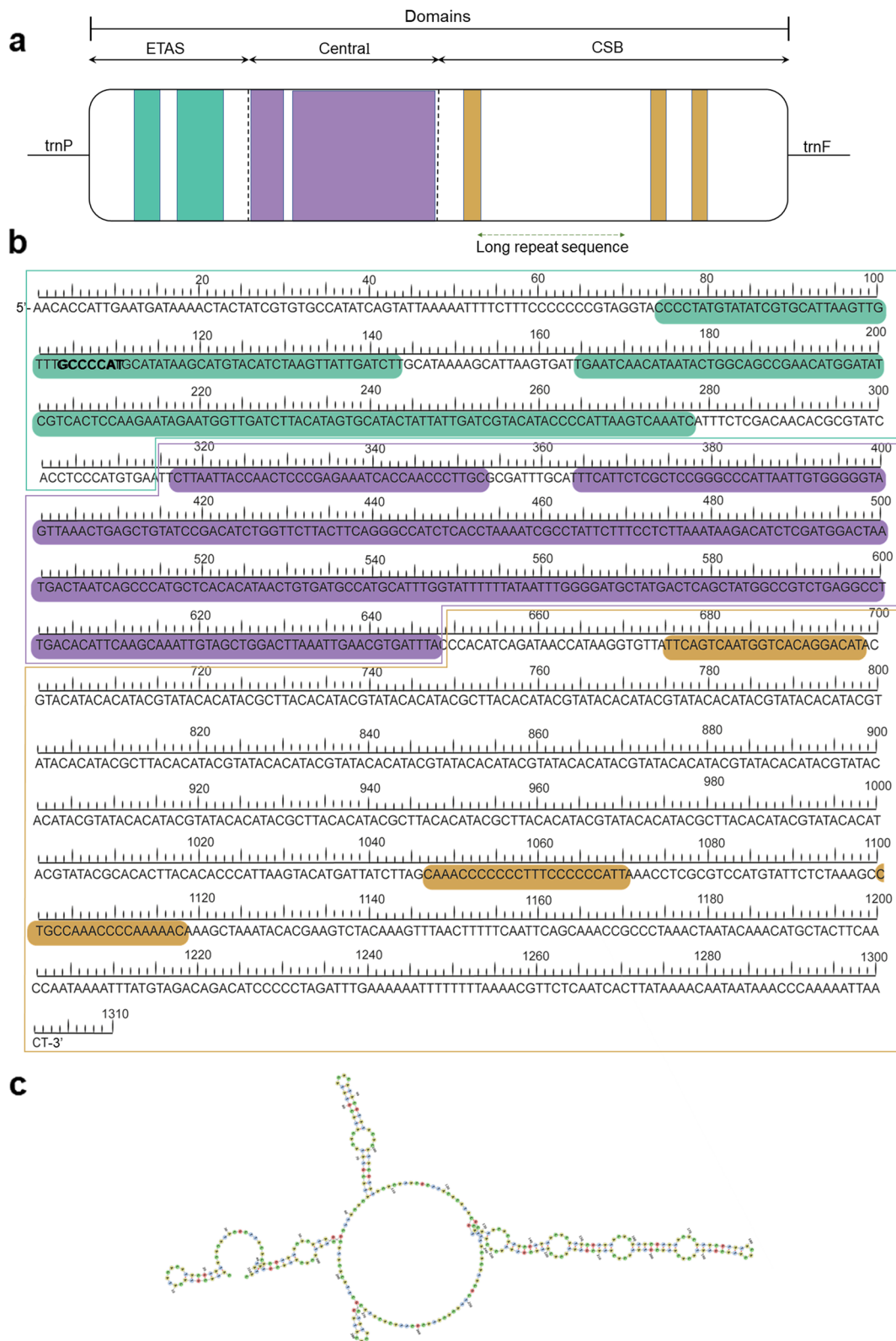
Fig. 5 (See legend on previous page.)

Gutiérrez *et al. BMC Genomics*   (2023) 24:527

Page 11 of 21
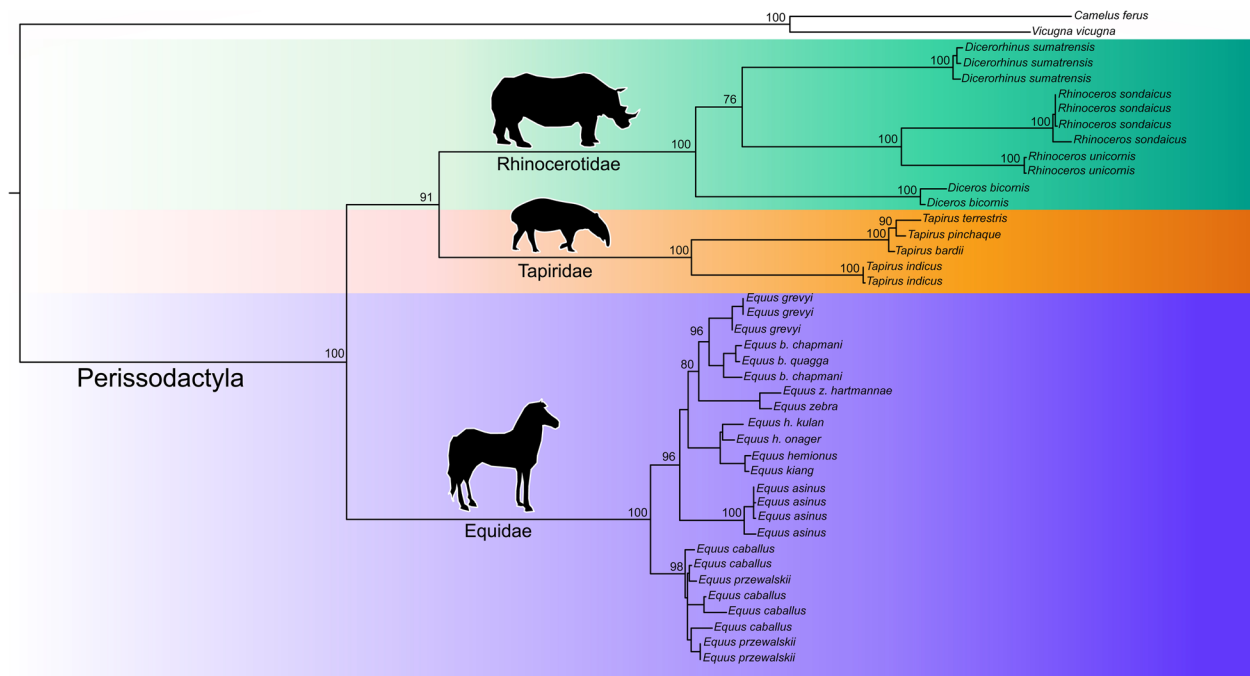
Fig. 6 Phylogenetic position of *Tapirus pinchaque* and phylogenetic relationships in the order Perissodactyla. Phylogenetic analysis generated by Maximum Likelihood based on a concatenated alignment of amino acids belonging to the 13 protein-coding genes present in the mitochondrial genome of 40 representatives of the order Perissodactyla plus two outgroup species. Numbers near internal nodes above the branches represent bootstrap values
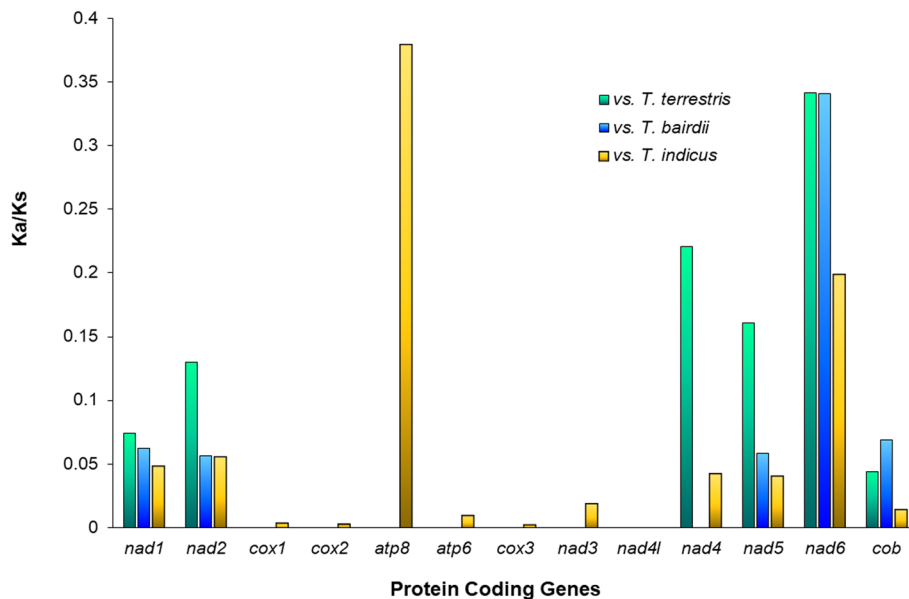
Fig. 7 Analysis of selective pressure in the PCG´s of *Tapirus pinchaque vs.* other species of *Tapirus.* Value of the K_A/K_S ratio for each of the 13 protein-coding genes present in the mitochondrial genome are provided

genes using BM's in EasyCodeML, we observed that all the 13 mitochondrial PCG´s have values of $\omega < 1$ in the foreground branches (Table 3). Based on likelihood ratio tests, nonetheless, the two-ratio model provided a statistically significant better fit than the one-ratio model in two genes (*cox3* [$\omega = 0.1136$; $p = 0.0489$], and *nad6*

Gutiérrez *et al. BMC Genomics*   (2023) 24:527

Page 12 of 21

Table 3 Branch-Model results with EasyCodeML. Selective pressure analysis of the mitochondrial protein-coding genes (four high-altitude species as the foreground branches)

Gene	Model	Ln L	Parameter estimates		LRT *p*-value	ω for foreground branches
atp6	M2	-3635.876	ω0 = 0.033	ω1 = 0.010	0.482	-
	M0	-3636.123	ω = 0.033			
atp8	M2	-863.670	ω0 = 0.289	ω1 = 2.267	0.337	-
	M0	-863.208	ω = 0.288			
cox1	M2	-7087.656	ω0 = 0.005	ω1 = 0.010	0.6183	-
	M0	-7087.780	ω = 0.005			
cox2	M2	-3217.650	ω0 = 0.012	ω1 = 0.010	0.561	-
	M0	-3217.819	ω = 0.012			
cox3	M2	-3345.092	ω0 = 0.005	ω1 = 0.114	0.048	ω1 = 0.11356
	M0	-3347.030	ω = 0.006			
cytb	M2	-6018.723	ω0 = 0.015	ω1 = 0.063	0.116	-
	M0	-6019.97	ω = 0.015			
nad1	M2	-4453.074	ω0 = 0.011	ω1 = 0.114	0.123	-
	M0	-4454.265	ω = 0.011			
nad2	M2	-5078.249	ω0 = 0.036	ω1 = 0.042	0.896	-
	M0	-5078.257	ω = 0.036			
nad3	M2	-1915.324	ω0 = 0.022	ω1 = 119.627	0.983	-
	M0	-1915.324	ω = 0.021			
nad4	M2	-6990.917	ω0 = 0.021	ω1 = 0.010	0.309	-
	M0	-6991.435	ω = 0.021			
nad4l	M2	-1616.032	ω0 = 0.026	ω1 = 0.010	0.641	-
	M0	-1616.141	ω = 0.026			
nad5	M2	-9747.676	ω0 = 0.034	ω1 = 0.080	0.335	-
	M0	-9748.141	ω = 0.034			
nad6	M2	-2826.731	ω0 = 0.049	ω1 = 0.315	0.015	ω1 = 0.31501
	M0	-2829.661	ω = 0.052			

[ω = 0.3150; *p* = 0.0155] Table 3). Our results suggest that *cox3* and *nad6* could be experiencing stronger purifiying selection than other mitochondrial PCG's, and that putative amino acid changes occurring in these two genes might be fine-tuning the efficiency of proton translocation, modulation of the redox potential of Complex I, and the production of ROS in odd-toed ungulates inhabiting high altitude environments [89]. Our results contrast to what has been reported in other studies; subunits of the *COX* gene complex (e. g., *cox1* and *cox2* genes) and NADH gene complex (e. g., *nad2, nad3, nad5,* and *nad6* genes) may be under positive selection, at least to some extent, in the high-altitude Tibetan antelope (*Pantholops hodgsonii*), Tibetan horse (*E. caballus*), and plateau pika (*Ochotona curzoniae*) [2, 18, 90].

We also performed a BSM analysis to detect signatures of positive selection in the four high-altitude species (i. e., foreground branches) when compared to low-altitude species (background branches). For all PCG's, model A (which allows positive selection; ω > 1) did not represent a better fit to our dataset compared to model A_{null} (which allows negative and neutral selection; $0 < ω < 1$), as indicated by a Likehood Ratio Test (LRT) in all PCG's (*p* > 0.05, Supplementary Table S4). Unexpectedly, and in contrast to our results, Wang et al. [91] detected signatures of positive selection, using the same BSM analysis, in a single mitochondrial gene (*nad5*) but using a disparate set of species that included mammal, aves, reptiles, amphibian, and ray-finned fishes.

Finally, SMs were used to identify sites of the sequences under positive selection in each PCG of the four high-altitude species of Perissodactyla. The M7 *vs.* M8 comparison showed high statistical significance (LRT test *p* < 0.01) in five different amino acid residues belonging to two PCG's (Supplementary Table S5): the *nad2* gene with one amino-acid residue (Valine in position 157, Supplementary Table S5) and the *nad5* gene with four amino-acid residues (Serine in position 27; Asparagine in position 149; Threonine in position 420, and Glutamic Acid in position 532; Supplementary Table S5).

Gutiérrez *et al. BMC Genomics*   (2023) 24:527

Page 13 of 21

Therefore, the existence of positive selection in these two genes is inferred; however, none of the sites exhibited significant posterior probability values greater than 0.95, as indicated by the Bayes empirical Bayes (BEB) approach. A previous study by Ning et al. [2] on different horse *Equus caballus* populations inhabiting different altitudes detected signatures of positive selection in two *nad6* residues (118 K and 162 L) from populations living above 2,000 masl.

In the web server DataMonkey [92], the RELAX analysis [93] indicated that *nad6* experienced a relaxation in selection strength ($k = 0.34$; LRT = 4.57; $p = 0.032$; Table 4) in high-altitude odd-toed ungulates. In turn, no indication that the strength of natural selection relaxed or intensified along branches leading to high-altitude odd-toed ungulates was found in the other 12 mitochondrial PCG´s (all LRT´s with $p > 0.05$; Table 4). The adaptive Branch-Site Random Effects Likelihood (aBSREL) [94] analysis found no evidence of episodic diversifying selection in any of the 13 mitochondrial PCG´s in high-altitude odd-toed ungulates. On the other hand, the Branch-Site Unrestricted Statistical Test for Episodic Diversification (BUSTED) analysis [95] (that included site-to-site synonymous rate variation) provided evidence of episodic diversifying selection in *cob* (LRT, $p = 0.038$) in high-altitude odd-toed ungulates (Table 4). The Mixed Effects Model of Evolution (MEME) analysis [96] revealed statistically significant episodic positive / diversifying selection signature in two sites (245; $p = 0.048$ and 257; $p = 0.040$) in *nad1*, one site (152; $p = 0.046$) in *nad2*, and one site

(100; $p = 0.040$) in *nad4* gene (Table 4). The Fast Unconstrained Bayesian AppRoximation (FUBAR) analysis [97] found no evidence of sites experiencing positive selection in any of the mitochondrial PCG´s. Instead, a large number of sites under negative selection was detected in most of the 13 mitochondrial genes, with *cox1* exhibiting the largest number of sites ($n = 176$) (Table 4) followed by *cob* and *nad4* (164 and 104 sites, respectively) under purifying selection (Table 4). Lastly, a Single-Likelihood Ancestor Counting (SLAC) analysis [98] found evidence of codons undergoing purifying selection ($\omega < 1$) at sites 158 ($p = 0.037$) and 184 ($p = 0.049$) of the *cob* and *nad5* genes, respectively (Table 4).

We have explored if adaptive evolution occurs in mitochondrial PCG´s of *T. pinchaque* and other odd-toed ungulates inhabiting high-altitude environments and found a signature of positive selective pressure at various residues in different PCG´s. Our results are in line with previous studies examining adaptive evolution in PCG´s of odd-toed ungulates (e. g., [13, 18, 99]) and other vertebrates and invertebrates colonizing harsh environments (e. g., [100–102]). Furthermore, previous studies have found evidence of adaptive evolution in mitochondrial PCG´s in bats [103], birds [104], snakes [105], and penguins [106], among other vertebrates [19]. Overall, our studies support the notion that changes in physiological/ metabolic demand and other environmental conditions (e. g., temperature, oxygen availability) in vertebrates may result in selective pressures acting upon mitochondrial-encoded genes to meet the demands of the animal´s

Table 4 Estimation of selection signatures using different models in the webserver DataMonkey. K denotes selection intensity parameter; $\omega < 1$ indicates purifying selection; $\omega > 1$ indicates diversifying selection; > 0.95 indicates significant posterior probability

gene	BM	BSM		SM		
	RELAX	aBSREL	BUSTED	MEME	FUBAR	SLAC
	K value (*p*-value;LR)	# sites (*p*-value)		sites $\omega > 1$ (*p*-value)	# sites $\omega < 1$ (> 0.95)	site $\omega < 1$ (*p*-value)
atp6	33.54 (0.337; 0.92)	–	–	–	27	–
atp8	0.23 (0.229; 1.45)	–	–	–	1	–
cox1	0.82 (0.88; 0.02)	–	–	–	176	–
cox2	11.77 (0.723; 0.13)	–	–	–	62	–
cox3	5.82 (0.080; 3.07)	–	–	–	18	–
cob	1.99 (0.156; 2.01)	–	**1 (0.038)**	–	164	**158 (0.037)**
nad1	1.68 (0.261; 1.26)	–	–	**245 (0.048), 257 (0.040)**	37	–
nad2	5.43 (0.899; 0.02)	–	–	**152 (0.046)**	33	–
nad3	0.93 (1.0; 0.00)	–	–	–	10	–
nad4	1.09 (0.924; 0.01)	–	–	**100 (0.040)**	104	–
nad4l	43.17 (0.316; 1.0)	–	–	–	5	–
nad5	0.41 (0.205; 1.61)	–	–	–	82	**184 (0.049)**
nad6	**0.34 (0.032; 4.57)**	–	–	–	9	–

Gutiérrez *et al. BMC Genomics*　　(2023) 24:527

Page 14 of 21

lifestyles and habitats in which they live [88]. The potential for adaptive evolution of mitochondrial PCG´s needs to be considered when exploring for evolutionary significant units in odd-toed ungulates and other species of conservation concern.

Conclusions

In summary, we have assembled, annotated, and characterized in detail the complete mtDNA of the smallest *Tapirus* species, the mountain wooly tapir, *T. pinchaque*. We inferred that, overall, all 13 mitochondrial PCG´s are exposed to negative / purifying selection. Nonetheless, further analyses of adaptive molecular evolution indicated that some residues experience adaptive evolution in species of odd-toed ungulates that have colonized high-altitude environments. Specifically, Branch Models indicated that *cox3* and *nad6* might be undergoing strong purifying selection in ungulates colonizing high-altitude habitats. Furthermore, Site Models suggested that one and four sites in *nad2* and *nad5*, respectively, could be experiencing positive selection. Lastly, additional analyses (RELAX, aBSREL, BUSTED, MEME, FUBAR, and SLAC) indicated a relaxation of selection strength in *nad6*, evidence of episodic diversifying selection in *cob*, and revealed episodic positive/diversifying selection signatures for two sites in *nad1*, one site each in *nad2* and *nad4*. The complete mitochondrial genome of *T. pinchaque* will support conservation initiatives in this remarkable mammal (i. e., non-intrusive bioprospecting and biomonitoring, see [107]) and our results indicate that the colonization of high-altitude habitats have resulted, at least to some extent, in the adaptive evolution of mitochondrial PCG´s in *T. pinchaque* and other odd-toed ungulates.

Methods

Field collection and sequencing

A hair sample preserved in 70% ethanol belonging to one specimen of *T. pinchaque* was donated to one of us (JO) by the Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais, Minas Gerais, Brazil. Total genomic DNA was extracted from the hair sample with the Qiagen Blood and Tissue Kit (Qiagen, Hilden, Germany) using the manufacturer's instructions. Extracted DNA was then shipped to the Savannah River Ecology Laboratory, University of Georgia, Aiken, for next-generation sequencing. Illumina paired-end (PE) shotgun libraries were prepared using the standard protocol of the Nextera™ DNA Sample Prep Kit (Epicentre®) and sequenced in an Illumina HiSeq sequencer (Illumina, San Diego, CA, USA) using 2×200 cycles. A total of 5,507,122 pairs of reads were generated for *T. pinchaque* and made available in FASTQ format by the sequencing facility. The totality of these reads was used for assembling the mitochondrial genome.

Mitochondrial genome assembly

The mitochondrial genome of *T. pinchaque* was de novo-assembled using the pipeline GetOrganelle v1.6.4 [54]. The mitochondrial genome of the congeneric *T. terrestris*, available in GenBank (KJ417810), was used as a seed. The run used k-mer sizes of 21, 55, 85, and 115.

Mitochondrial genome annotation and characterization

The newly assembled mitochondrial genome was first annotated in silico using the web servers MITOS and MITOS2　(http://mitos2.bioinf.uni-leipzig.de/index.py) [108] with the vertebrate genetic code. Annotation and manual curation of all the features found in the mitochondrial genome, including the PCG´s start and stop codons corrections, were carried out using the ExPASy v. 3.0 translate tool (https://web.expasy.org/translate/) [109] and MEGA v. X [110]. Subsequent visualization of each manually curated mitochondrial genome was performed with the Chloroplot webserver (https://www.frontiersin.org/articles/10.3389/fgene.2020.576124/full) [111].

The nucleotide composition of the mitochondrial genome was estimated in the software MEGA v. X. A codon usage analysis of the PCG´s from the studied species was conducted using the Sequence Manipulation Suite: Codon Usage online tool v. 2 (https://www.bioinformatics.org/sms2/codon_usage.html) [112]. Next, EZcodon, as implemented in the web server EZmito v. 2021.11 (http://ezmito.unisi.it/ezcodon) [113], was used to calculate Relative Synonymous Codon Usage (RSCU). RSCU is defined as the ratio of the observed frequency of codons to the expected frequency given that all the synonymous codons for the same amino acid are used equally [69, 71].

The secondary structure of each tRNA was predicted using the software MiTFi [114] as implemented in MITOS2 (http://mitos2.bioinf.uni-leipzig.de/ index.py) [115] and the web server Forna (http://rna.tbi.univie.ac.at/forna/) [72] was used for visualization of each tRNA.

The putative CR of the mitochondrial genomes was examined in detail, first using the BioPHP-microsatellite repeat finder (http://insilico.ehu.es/mini_tools/microsatellites/) [116] web server to detect microsatellites. The CR was then analyzed using the web server Tandem Repeat Finder v. 4.09.1 (https://tandem.bu.edu/trf/trf.html) [117] to detect repetitive tandem sequences. Lastly, secondary structure predictions of the CR in the mitochondrial genome were modeled using the RNA structure Web Server tool v. 6.4 (http://rna.urmc.rochester.edu/RNAstructureWeb) [118]. All default options were used for exploring the secondary structure of the

CR, except for the temperature that was set to 285.15°K for *T. pinchaque*, representing the 'thermal' environment inhabited by this species.

Phylomitogenomics of odd-toed ungulates

We examined the phylogenetic position of *T. pinchaque* among other representatives of the genus *Tapirus* and Perissodactyla based on translated PCG´s, following the protocol detailed in Ennis et al. [56]. The newly assembled mitochondrial genome of *T. pinchaque* together with the mitochondrial genomes of 39 other specimens belonging to extant odd-toed ungulates available in the NCBI's GenBank database were used for phylogenetic reconstruction using the software MitoPhAST v. 3 [119]. Two different species of artiodactyls (order Artiodactyla), namely, *Vicugna vicugna*, and *Camelus ferus*, belonging to the family Camelidae, were used as outgroups during the phylogenetic analysis. The software MitoPhAST first extracted all 13 PCG´s nucleotide sequences from the species available in GenBank as well as those from *T. pinchaque* and aligned each set of PCG nucleotide sequences after translating them into amino acids with the software Clustal Omega v. 1.2.4 [120]. Next, poorly aligned regions were removed with the software trimAl v. 1.2 [121] to then partition the dataset and estimate the best fitting models of sequence evolution using the program ProtTest v. 3.4.2 [122]. Finally, the concatenated and partitioned PCG amino acid alignments were used to conduct a maximum likelihood (ML) phylogenetic tree analysis in the software IQ-TREE v. 1.6.12 [123]. The robustness of the ML tree topology was ascertained by 1,000 bootstrap pseudoreplicates of the tree search.

Selective pressure analyses

First, we measured selective pressures acting on each of the 13 PCG´s in the studied species. For this purpose, the number of non-synonymous substitutions per non-synonymous site (Ka), synonymous substitutions per synonymous site (Ks), and the Ka / Ks ratio were estimated using the software KaKs_calculator v. 2.0 [124]. Neutral selection is given by Ka /Ks = 1; while Ka / Ks < 1 denotes purifying (negative) selection, and Ka / Ks > 1 implies diversifying (positive) selection in a particular PCG [124]. For this analysis, Ka and Ks values were based on pairwise comparisons between each PCG sequence from the species herein studied, and all other species of *Tapirus* with mitochondrial genomes available in GenBank (*T. terrestris*, NC_053962.1; *T. indicus*, KJ417810.1 and *T. bairdii*, NC_063943). The γ-MYN was used during calculations to account for variable mutation rates along the length of each gene sequence [124].

Next, we examined for signatures of positive selection in all 13 mitochondrial PCG´s that could be driven by colonization of high-altitude environments in *T. pinchaque* and other representatives of the Perissodactyla using the program EasyCodeML v.1.0 [50, 125]. EasyCodeML is a codon-based method for detecting signatures of positive selection that estimates ω (d_N/d_S) across a phylogenetic tree. We used as input files, the phylogeny previously obtained with MitoPhAST and the multiple sequence alignments of each mitochondrial PCG´s with the software Muscle v. 5 as implemented in MEGA. The possible effect of a particular ecological condition (e. g., high-altitude in this study) on selective pressures proceeds after the user assigns 'foreground' and 'background' branches in the phylogenetic tree, with EasyCodeML [126]. Foreground branches correspond to those leading to terminal nodes (species) that live in high-altitude habitats: represented by *T. pinchaque*, *T. bairdii*, *Equus burchellii quagga*, and *Equus kiang* in our study. In turn, background branches are those leading to 36 other terminals from representatives of the Perissodactyla including in the phylogeny and inhabiting regions below 2000 masl.

In EasyCodeML, we first used Branch Models (BM) [126] to test for statistically significant differences in ω among branches of the phylogenetic tree. We fitted three different models to the sequence dataset (aligned amino acids strings for each PCG separately), namely, the one-ratio model, the two-ratio model, and the free-ratio model. The one-ratio model (M0) assumes that ω is constant throughout the phylogenetic tree, the free-ratio model allows an independent ω for each branch in the tree, and the two-ratio model assumes that specific (e. g., foreground) branches have an ω that is different from that throughout the rest of the (e. g., background) branches in the phylogenetic tree [125, 127]. We conducted pairwise comparisons among the three different models using independent LRT's [128]. The degrees of freedom were the difference in the number of free parameters between models. Evidence for positive selection and the effect of a particular environmental condition on PCG´s adaptive evolution is provided when the two-ratio model provides a better fit than the M0 model to the sequence data [126]. Also, heterogeneity in the dN/dS ratio between lineages can be evidenced when comparing the free-ratio model against the M0. This heterogeneity between lineages may be due to positive selection or relaxed selection constraints [127].

Considering that positive selection can often act on only a few sites and during short periods of evolutionary time, we also used the BSM in EasyCodeML to test for statistically significant differences in ω among sites along pre-specified lineages in our phylogeny [49]. The BSM assume heterogeneous ω across sites and across branches [125]. Two models were fitted to our datasets (aligned PCG´s): the model A and the model A_{null} [49, 125].

Gutiérrez *et al. BMC Genomics*   (2023) 24:527

Page 16 of 21

Model A allows positive selection (with ω > 1) to occur along pre-specified branches (i. e., foreground branches) while for the other branches (i. e., background branches) only negative selection and neutral evolution (0 < ω < 1) is allowed. In turn, Model A$_{null}$ allows negative selection (i. e., 0 < ω < 1) and neutral evolution (i. e., ω = 1) to occur along all branches [49, 125]. We conducted pairwise comparisons among the two different models using independent LRT's [128]. We predicted Model A to be a better fit to our dataset than the null model if high-altitude environments drive adaptive evolution of codons along branches in odd-toed ungulates.

Lastly, we used the SM in EasyCodeML, to investigate selection pressure among sites of each mitochondrial gene sequence [126]. We fitted six different codon substitution models to our dataset: M0 (one ω ratio for all sites), M1a (nearly neutral, ω = 1), M2a (positive selection, ω > 1), M7 (ω varies according to β distribution, 0 < ω < 1), M8 (β and ω > 1), and M8a (β and ω = 1) in which ω-ratio is allowed to vary among sites [49, 126]. Two pairs of site model contrasts (M1a *vs.* M2a and M7 *vs.* M8) are particularly effective for the detection of positive selection [128]. Support for positive selection can be identified if M2a provides a better fit than M1a. Similarly, positive selection can be identified if the M8 model provides a better fit than M7 or M8a [125]. The M8 vs. M7 offers a very stringent test of positive selection [126; 129]. We conducted pairwise comparisons among the different models using independent LTR [128, 129]. Finally, we apply the BEB method to identify sites under positive selection or relaxed from purifying selection in foreground branches with significant LRTs in the SM test [129].

In addition to our analyses in EasyCodeML, we use codon models in the HyPhy v. 2.0 package [130] as implemented on the web server DataMonkey v. 1.10.0 (http://www.datamonkey.org/) [92] to detect signatures of adaptive evolution in mitochondrial PCG's from odd-toed ungulates inhabiting high-altitude environments. One BM (i. e., RELAX) [93] and two BSMs (aBSREL [94] and BUSTED [95]) were executed with the same input dataset previously used in the EasycodeML analyses that contained aligned sequences from the four species of interest plus the other 38 odd-toed ungulates (including the two species as outgroup) from low-altitude habitats.

RELAX determines if the strength of selection has relaxed or intensified along pre-specified branches (i. e., leading to high-altitude odd-toed ungulates) in a phylogenetic tree [93]. RELAX fits two models (null, with three ω classes applied to the entire phylogeny vs. alternative, introducing the selection intensity parameter *k* and modified and fixed ω^k classes) to the dataset and uses the LRT to compare the null and alternative models. The parameter *k* estimated in RELAX represents

a measure of selection strength with values greater or lower than 1 indicating that selection strength has intensified or relaxed, respectively [93]. A better fit of the alternative model to the dataset is useful for identifying shifts in the stringency of natural selection on a given gene [92]. Neighbor-Joining trees used for the calculation were automatically generated in the web server DataMonkey using the General Time-Reversible (GTR) substitution model.

aBSREL tests whether a proportion of sites have evolved under positive selection in pre-specified branches (i. e., leading to high-altitude odd-toed ungulates) in a phylogeny [92]. aBSREL models' heterogeneity in ω at both site and branch levels and uses the corrected Akaike's Information Criterion (AICc) to estimate the 'optimal' number of ω rate classes for each branch in the phylogeny. After fitting a full adaptive model with the 'optimal' number of ω rate classes to the dataset, an LRT is performed to compare the fit of the full model against a null model that does not allow branches to experience positive selection in the phylogeny [93]. In turn, BUSTED identifies whether a gene has undergone positive selection in at least one site and at least one branch of interest in a phylogeny. As in previous methods, BUSTED compares the fit of two models to the dataset; an alternative model (also called unrestricted model) against a null model (or constrained model) that does not allow positive selection on the branches of interest. In addition, BUSTED calculates "Evidence Ratios" (ER) for each site. The ER is a likelihood ratio that the unrestricted (alternative) model represents a better fit to the dataset compared to the null (constrained) model [95].

Lastly, three different Site-Model analyzes were executed using only the four high-altitude odd-toed ungulates as an input file: MEME [96], FUBAR (Fast, Unconstrained Bayesian AppRoximation [97], and SLAC (Single-Likelihood Ancestor Counting [98]. MEME tests whether sites in a set of branches evolve under positive or diversifying episodic selection and employs a mixed-effects maximum likelihood approach to infer two ω-rate classes and the probability that each site will evolve under each respective ω-rate class [96]. We selected a *p*-value < 0.05 to decrease false positive rates in this test [96]. FUBAR uses a Bayesian approach to infer dN and dS per site in each alignment and reports posterior probabilities (ranging from 0 to 1) to determine the existence of positive selection [97]. We changed the default posterior probability value from 0.9 to 0.95 to minimize false positive results in our analyses [97]. SLAC identifies positive and negative selection at specific sites in a codon alignment through a combination of maximum likelihood and a counting (Suzuki-Gojobori modified version) approach to infer dN and dS per site [98].

Abbreviations

aBSREL	Adaptive Branch-Site Random Effects Likelihood
AICc	Akaike's Information Criterion
Ala	Alanine amino acid
Arg	Arginine amino acid
ATP	Adenosine triphosphate
atp	Subunits of ATP synthase
BEB	Bayesian Empirical Bayes
BM	Branch Model
BSM	Branch-Site Model
BUSTED	Branch-Site Unrestricted Statistical Test for Episodic Diversification
bv	Bootstrap value
cm	Centimeters
cob	Cytochrome b subunit of the cytochrome bc1 complex
cox	Subunits of the cytochrome c oxidase
CR	Control Region
CSB	Conserved Sequence Block
d_N	Non-synonymous substituions
d_S	Synonymous substitutions
EGLN1	Egl-9 family hypoxia inducible factor 1
EPAS1	Endothelial PAS domain protein 1
ER	Evidence Ratio
ETAS	Extended Termination Associated Sequence
FUBAR	Fast, Unconstrained Bayesian AppRoximation
Gln	Glutamine aminoacid
GTR	General Time-Reversible
H	Heavy strand
H_2O_2	Hydrogen peroxide
HIF-1A	Hypoxia-inducible factor-1 alpha
Ka	Non-synonymous substitutions per non-synonymous site
Ks	Synonymous substitutions per synonymous site
L	Negative or light strand
Leu	Leucine aminoacid
LRT	Likehood Ratio Test
Lys	Lysine amino acid
m	Meters
M0	One-ratio model
masl	Meters above sea level
MEME	Mixed Effects Model of Evolution
ML	Maximun Likelihood
mtDNA	Mitochondrial genome
nad	Subunits of the NADH dehydrogenase complex
NCBI	National Center for Biotechnology Information
O_2^-	Superoxide
OXPHOS	Oxidative phosphorylation system
PCG´s	Protein coding genes
PE	Paired-end
ROS	Reactive oxygen species
rRNA	Ribosomal genes
rrnL	16S ribosomal RNA
rrnS	12S ribosomal RNA
RSCU	Relative Synonymous Codon Usage
Ser	Serine amino acid
SLAC	Single-Likelihood Ancestor Counting
SM	Site Model
T	Incomplete stop codon
Thr	Threonine amino acid
tRNA	Transfer RNA genes
Trp	Tryptophan amino acid
Val	Valine aminoacid
$\omega = d_N/d_S = Ka/Ks$	The ratio of non-synonymous to synonymous substitutions

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-023-09596-8.

> **Additional file 1: Table S1.** Codon usage analysis of the protein coding genes in the mitochondrial DNA of *Tapirus pinchaque*. **Table S2.** Characteristics of the microsatellite repeat sequences detected in the control region of the mtDNA of *T. Pinchaque*. **Table S3.** Tandem repeat sequence detected in the Control Region of the mtDNA of *Tapirus pinchaque*.
>
> **Additional file 2: Supplementary Figure S1.** Secondary structure predictions of the Control Region.
>
> **Additional file 3: Table S4.** Branch-Site Model results with EasyCodeML. Selective pressure analysis of the mitochondrial protein-coding genes (four high-altitude species as the foreground branches). **Table S5.** Site-Model results with EasyCodeML. Selective pressure analysis of the mitochondrial protein-coding genes (four high-altitude species as the foreground branches).

Authors' contributions

E.G.G., J.O., and J.A.B. contributed to study development and design. Analysis was performed by E.G.G., A.S., and J.A.B. The first draft of the manuscript was written by E.G.G. and J.A.B. with the support of J.O. E.G.G., J.O., and J.A.B. read and approved the final manuscript.

Availability of data and materials

Complete mitochondrial genome sequence data have been deposited in the NCBI GenBank with accession no. OQ420428.

Declarations

Ethics approval and consent to participate

All methods were performed in accordance with the relevant guidelines and regulations.

Consent for publication

Not applicable.

Competing interests

The authors have not disclosed any conflict of interest.

Author details

[1]Laboratorio de Bioconservación y Manejo, Posgrado en Ciencias Químicobiológicas, Departamento de Zoología, Escuela Nacional de Ciencias Biológicas, Instituto Politécnico Nacional, Prolongación Carpio y Plan de Ayala S/N, Col. Santo Tomás, Ciudad de México C.P. 11340, Mexico. [2]Department of Biological Sciences, 132 Long Hall, Clemson University, Clemson, SC 29634, USA. [3]Smithsonian Marine Station at Fort Pierce, 701 Seaway Drive, Fort Pierce, FL 34949, USA. [4]Departamento de Biología Marina, Facultad de Ciencias del Mar, Universidad Católica del Norte, Larrondo 1281, Coquimbo, Chile.

Gutiérrez *et al. BMC Genomics*   (2023) 24:527

Page 18 of 21

References

1. Hassanin A, Ropiquet A, Couloux A, Cruaud C. Evolution of the mitochondrial genome in mammals living at high altitude: new insights from a study of the tribe Caprini (Bovidae, Antilopinae). J Mol Evol. 2009;68:293–310. https://doi.org/10.1007/s00239-009-9208-7.
2. Ning T, Xiao H, Li J, Hua S, Zhang YP. Adaptive evolution of the mitochondrial NADH6 gene in the domestic horse. Genet Mol Res. 2010;9:144–50. https://doi.org/10.4238/vol9-1gmr705.
3. Friedrich J, Wiener P. Selection signatures for high-altitude adaptation in ruminants. Anim Genet. 2020;51:157–65. https://doi.org/10.1111/age.12900.
4. Yang J, Li W-R, Lv F-H, He S-G, Tian S-L, et al. Whole-genome sequencing of native sheep provides insights into rapid adaptations to extreme environments. Mol Biol Evol. 2016;33:2576–92. https://doi.org/10.1093/molbev/msw129.
5. Edea Z, Dadi H, Dessie T, Kim K-S. Genomic signatures of high-altitude adaptation in Ethiopian sheep populations. Genes Genomics. 2019;41:973–81. https://doi.org/10.1007/s13258-019-00820-y.
6. Hochachka PW, Buck LT, Doll CJ, Land SC. Unifying theory of hypoxia tolerance: molecular/metabolic defense and rescue mechanisms for surviving oxygen lack. PNAS USA. 1996;93(18):9493–8. https://doi.org/10.1073/pnas.93.18.9493.
7. Li M, Pan D, Sun H, Zhang L, Cheng H, et al. The hypoxia adaptation of small mammals to plateau and underground burrow conditions. AMEM. 2021;4:319–28. https://doi.org/10.1002/ame2.12183.
8. Weber RE, Lalthantluanga R, Braunitzer G. Functional characterization of fetal and adult yak hemoglobins: an oxygen binding cascade and its molecular basis. Arch Biochem Biophys. 1988;263:199–203. https://doi.org/10.1016/0003-9861(88)90628-5.
9. Piccinini M, Kleinschmidt T, Jurgens KD, Braunitzer G. Primary structure and oxygen-binding properties of the hemoglobin from guanaco (*Lama guanacoe*, Tylopoda). Biol Chem Hoppe Seyler. 1990;71:641–8. https://doi.org/10.1515/bchm3.1990.371.2.641.
10. Storz JF, Sabatino SJ, Hoffmann FG, Gering EJ, Moriyama H, Ferrand N, et al. The molecular basis of high-altitude adaptation in deer mice. PLoS Genet. 2007;3:e45. https://doi.org/10.1371/journal.pgen.0030045.
11. Gou X, Wang Z, Li N, Qiu F, Xu Z, Yan D, et al. Whole-genome sequencing of six dog breeds from continuous altitudes reveals adaptation to high-altitude hypoxia. Genome Res. 2013;24:1308–15. https://doi.org/10.1101/gr.171876.113.
12. Gao X, Wang S, Wang Y-F, Li S, Wu S-X, Yan R-G, et al. Long read genome assemblies complemented by single cell RNA-sequencing reveal genetic and cellular mechanisms underlying the adaptive evolution of yak. Nat Commun. 2022;13:4887. https://doi.org/10.1038/s41467-022-32164-9.
13. Qiu Q, Zhang G, Ma T, Qian W, Wang J, et al. The yak genome and adaptation to life at high altitude. Nat Genet. 2012;44:946–9. https://doi.org/10.1038/ng.2343.
14. Zhang W, Fan Z, Han E, Hou R, Zhang L, et al. Hypoxia Adaptations in the Grey Wolf (*Canis lupus chanco*) from Qinghai-Tibet Plateau. PLoS Genet. 2014;10(7):e1004466. https://doi.org/10.1371/journal.pgen.1004466.
15. Gorkhali N, Dong K, Yang M, Song S, Kader A, Shrestha BS, et al. Genomic analysis identified a potential novel molecular mechanism for high-altitude adaptation in sheep at the Himalayas. Sci Rep. 2020;6:29963. https://doi.org/10.1038/srep29963.
16. Song S, Yao N, Yang M, Liu X, Dong K, Zhao Q, et al. Exome sequencing reveals genetic differentiation due to high-altitude adaptation in the Tibetan cashmere goat (*Capra hircus*). BMC Genomics. 2016;17:122. https://doi.org/10.1186/s12864-016-2449-0.
17. Buroker NE, Ning X-H, Zhou Z-N, Li K, Cen W-J, Wu X-F, et al. EPAS1 and EGLN1 associations with high altitude sickness in Han and Tibetan Chinese at the Qinghai-Tibet Plateau. Blood Cells Mol Dis. 2012;49(2):67–73. https://doi.org/10.1016/j.bcmd.2012.04.004.
18. Xu S, Luosang J, Hua S, He J, Ciren A, Wang W, Tong X, et al. High altitude adaptation and phylogenetic analysis of Tibetan horses based on the mitochondrial genome. J Genet Genomics. 2007;34:720–9. https://doi.org/10.1016/S1673-8527(07)60081-2.
19. da Fonseca RR, Johnson WE, O'Brien SJ, Ramos MJ, Antunes A. The adaptive evolution of the mammalian mitochondrial genome. BMC Genomics. 2008;9:119. https://doi.org/10.1186/1471-2164-9-119.
20. Di Rocco F, Zambelli AD, Vidal Rioja LB. Identification of camelid specific residues in mitochondrial ATP synthase subunits. J Bioenerg Biomembr. 2009;41:223–8. https://doi.org/10.1007/s10863-009-9221-y.
21. Tang JX, Thompson K, Taylor RW, Oláhová M. Mitochondrial OXPHOS biogenesis: co-regulation of protein synthesis, import, and assembly pathways. Int J Mol Sci. 2020;21:3820. https://doi.org/10.3390/ijms21113820.
22. Wallace DC. A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: a dawn for evolutionary medicine. Annu Rev Genet. 2005;39:359–407. https://doi.org/10.1146/annurev.genet.39.110304.095751.
23. Luo Y, Yang X, Gao Y. Mitochondrial DNA response to high altitude: a new perspective on high-altitude adaptation. Mitochondrial DNA. 2013;24(4):313–9. https://doi.org/10.3109/19401736.2012.760558.
24. Lee J, Song C-H. Effect of reactive oxygen species on the endoplasmic reticulum and mitochondria during intracellular pathogen infection of mammalian Cells. Antioxidants. 2021;10:872. https://doi.org/10.3390/antiox10060872.
25. Gao WX, Liu JZ, Wu LP, Cai MC. Studies of hypoxic rat brain mitochondrial transcription activity in vitro. Chin J Physiol. 2001;17:323–6.
26. Sharma S, Singh Y, Sandhir R, Singh S, Ganju L, Kumar B, Varshney R. Mitochondrial DNA mutations contribute to high altitude pulmonary edema via increased oxidative stress and metabolic reprogramming during hypobaric hypoxia. Biochim Biophys Acta Bioenerg. 1989;1862(8):148431. https://doi.org/10.1016/j.bbabio.2021.148431.
27. Xu SQ, Yang YZ, Zhou J, Jing GE, Chen YT, et al. A mitochondrial genome sequence of the Tibetan antelope (*Pantholops hodgsonii*). Genomics Proteomics Bioinformatics. 2005;3:5–1. https://doi.org/10.1016/S1672-0229(05)03003-2.
28. Ashley MV, Norman JE, Stross L. Phylogenetic Analysis of the perissodactylan family Tapiridae using mitochondrial cytochrome C oxidase (COII) sequences. J Mamm Evol. 1996;3:315–26. https://doi.org/10.1007/BF02077448.
29. Schoch RM. A review of the Tapiroids. In: Prothero DR, Schoch RM, editors. The Evolution of Perissodactyls. New York: Oxford University Press; 1989. p. 298–320.
30. Scherler L, Becker D, Berger J-P. Tapiridae (Perissodactyla, Mammalia) of the swiss molasse basin during the Oligocene-Miocene transition. J Vertebr Paleontol. 2011;31(2):479–96. https://doi.org/10.1080/02724634.2011.550360.
31. Rayan DM, Mohamad SW, Dorward L, Aziz SA, Clements GR, Christopher WCT, et al. Estimating the population density of the Asian tapir (*Tapirus indicus*) in a selectively logged forest in Peninsular Malaysia. Integr Zool. 2012;7:373–80. https://doi.org/10.1111/j.1749-4877.2012.00321.x.
32. Simpson BK, Shukor MN, Magintan D. Food Selection of the Malayan Tapir (*Tapirus indicus*) Under Semi-Wild Conditions. AIP Conf Proc. 2014;1571:317–24. https://doi.org/10.1063/1.4858676.
33. Couzzol MA, Clozato CL, Holanda EC, Rodrigues FHG, Nienow S, Thoisy B, et al. A new species of tapir from the Amazon. J Mamm. 2013;94:1331–45. https://doi.org/10.1644/12-MAMM-A-169.1.
34. Ruiz-García M, Vásquez C, Sandoval S, Kaston F, Luengas-Villamil K, Shostell JM. Phylogeography and spatial structure of the lowland tapir (*Tapirus terrestris*, Perissodactyla: Tapiridae) in South America. Mitochondrial DNA A. 2016;27(4):2334–42. https://doi.org/10.3109/19401736.2015.1022766.
35. Downer CC. The mountain tapir, endangered 'flagship' species of the high Andes. Oryx. 1996;30:45–58. https://doi.org/10.1017/S0030605300021384.
36. Lizcano DJ, Pizarro V, Cavelier J, Carmona J. Geographic distribution and population size of the mountain tapir (*Tapirus pinchaque*) in Colombia. J Biogeogr. 2002;29:7–15.
37. Padilla M, Dowler RC. Tapirus terrestris. Mamm Species. 1994;481(2):1–8.
38. Flesher KM, Medici EP. The distribution and conservation status of *Tapirus terrestris* in the South American Atlantic Forest. Neotrop Biol. 2022;17(1):1–19. https://doi.org/10.3897/neotropical.17.e71867.

Gutiérrez *et al. BMC Genomics*   (2023) 24:527

Page 19 of 21

39. Holden J, Yanuar A, Martyr D. The Asian Tapir in Kerinci Seblat National Park, Sumatra: evidence collected through photo-trapping. Oryx. 2003;37(1):34–40. https://doi.org/10.1017/S0030605303000097.

40. March IJ, Naranjo EJ. Tapir (*Tapirus bairdii*). In: Ceballos G, Oliva G, editors. Los mamíferos silvestres de México, Comisión Nacional para el Conocimiento y Uso de la Biodiversidad y Fondo de Cultura Económica, CDMX, Mexico; 2005. 496-497.

41. Reid F. A field guide to the mammals of Central America and southeast Mexico. Oxford University Press; 1997.

42. Naranjo EJ, Vaughan C. Ampliación del ámbito altitudinal del tapir Centroamericano (*Tapirus bairdii*). Rev Biol Trop. 2000;48:724.

43. González-Maya JF, Schipper J, Benítez A. Elevational distribution and abundance of Baird's tapir (*Tapirus bairdii*) at different protection areas in the Talamanca region of Costa Rica. Tapir Cons. 2009;18:29–35.

44. Eisenberg JF, Groves CP, MacKinnon K. Tapirs. In: Parker SP, editor. Grzimek's encyclopedia of mammals. München: McGraw-Hill Inc; 1990. p. 598–608.

45. Downer CC. Status and action plan of the mountain tapir (*Tapirus pinchaque*). In: Brooks DM, Bodmer RE, Matola S, editors. Tapirs status survey and conservation action plan. Gland, Cambridge: IUCN/SSC Tapir Specialist Group, IUCN; 1997.

46. Traeholt C, Novarino W, bin Saaban S, Shwe NM, Lynam A, Zainuddin Z, Simpson B, bin Mohd S. *Tapirus indicus*. The IUCN Red List of Threatened Species; 2016. https://www.iucnredlist.org/search?query=tapirus%20indicus&searchType=species Accessed 21 Sept 2022.

47. Nowak RM. Walker's mammals of the world. 6th ed. Baltimore: Johns Hopkins University Press; 1999.

48. Baeza JA. An introduction to the Special Section on Crustacean Mitochondrial Genomics: Improving the assembly, annotation, and characterization of mitochondrial genomes using user-friendly and open-access bioinformatics tools, with decapod crustaceans as an example. J Crustac Biol. 2022;42:1–4. https://doi.org/10.1093/jcbiol/ruac012.

49. Zhang J, Nielsen R, Yang Z. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol Biol Evol. 2005;22:2472–9. https://doi.org/10.1093/molbev/msi237.

50. Yang Z. PAML 4: Phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24:1586–91. https://doi.org/10.1093/molbev/msm088.

51. Downer CC. Observations on the diet and habitat of the mountain tapir (*Tapirus pinchaque*). J Zool. 2001;254:279–91. https://doi.org/10.1017/S0952836901000796.

52. Ortega-Andrade HM, Prieto-Torres DA, Gómez-Lora I, Lizcano DJ. Ecological and geographical analysis of the distribution of the mountain tapir (*Tapirus pinchaque*) in Ecuador: importance of protected areas in future scenarios of global warming. PLoS One. 2015;10(3):e0121137. https://doi.org/10.1371/journal.pone.0121137.

53. More A, Devenish C, Carrillo-Tavara K, Piana RP, Lopez-Malaga C, Vega-Guarderas Z, Nuñez-Cortez E. Distribution and conservation status of the mountain tapir (*Tapirus pinchaque*) in Peru. J Nat Conserv. 2022;66:126–30. https://doi.org/10.1016/j.jnc.2022.126130.

54. Jin J-J, Yu W-B, Yang J-B, Song Y, dePamphilis CW, et al. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. 2020;21:241. https://doi.org/10.1186/s13059-020-02154-5.

55. Arnason U, Adegoke JA, Gullberg A, Harley EH, Janke A, Kullberg M. Mitogenomic relationships of placental mammals and molecular estimates of their divergences. Gene. 2008;421:37–51. https://doi.org/10.1016/j.gene.2008.05.024.

56. Ennis CC, Ortega J, Baeza JA. First genomic resource for an endangered neotropical mega-herbivore: the complete mitochondrial genome of the forest-dweller (Baird's) tapir (*Tapirus bairdii*). PeerJ. 2022;10:e13440. https://doi.org/10.7717/peerj.13440.

57. Muangkram Y, Wajjwalku W, Kaolim N, Buddhakosai W, Kamolnorranath S, Siriaroonrat. The complete mitochondrial genome of the Asian tapirs (*Tapirus indicus*): the only extant Tapiridae species in the old world. Mitochondrial DNA A. 2016;27(1):413–5. https://doi.org/10.3109/19401736.2014.898283.

58. Xu X, Arnason U. The Complete Mitochondrial DNA Sequence of the White Rhinoceros, *Ceratotherium simum*, and Comparison with the mtDNA Sequence of the Indian Rhinoceros. Rhinoceros unicornis Mol

59. Phylogenet Evol. 1997;7(2):189–94. https://doi.org/10.1006/mpev.1996.0385.

59. Xu X, Janke A, Arnason U. The Complete Mitochondrial DNA Sequence of the Greater Indian Rhinoceros, *Rhinoceros unicornis*, and the Phylogenetic Relationship Among Carnivora, Perissodactyla, and Artiodactyla (+ Cetacea). Mol Biol Evol. 1996;13(9):1167–73. https://doi.org/10.1093/oxfordjournals.molbev.a025681.

60. Sun Y, Jiang Q, Yang C, Wang X, Tian F, Wang Y, et al. Characterization of the complete mitochondrial genome of Dezhou donkey (*Equus asinus*) and evolutionary analysis. Curr Genet. 2016;62:383–90. https://doi.org/10.1007/s00294-015-0531-9.

61. Pei J, Chu M, Bao P, Sha Z, Ding X, Yan P, Guo X. The complete mitochondrial genome of the Sanhe horse (*Equus caballus*). Conserv Genet Res. 2019;11:11–4. https://doi.org/10.1007/s12686-017-0951-6.

62. Luo Y, Chen Y, Liu F, Jiang C, Gao Y. Mitochondrial genome sequence of the Tibetan wild ass (*Equus kiang*). Mitochondrial DNA. 2011;22(1–2):6–8. https://doi.org/10.3109/19401736.2011.588221.

63. Guo X, Bao P, Pei J, Ding X, Liang C, Yan P, Lu D. Complete mitochondrial genome of Qingyang donkey (*Equus asinus*). Conserv Genet Resour. 2017;9:269–71. https://doi.org/10.1007/s12686-016-0670-4.

64. Margaryan A, Sinding MHS, Liu S, Vieira FG, Chan YL, et al. Recent mitochondrial lineage extinction in the critically endangered Javan rhinoceros. Zool J Linn Soc. 2020;190:372–83. https://doi.org/10.1093/zoolinnean/zlaa004.

65. Steiner CC, Houck ML, Ryder OA. Genetic variation of complete mitochondrial genome sequences of the Sumatran rhinoceros (*Dicerorhinus sumatrensis*). Conserv Genet. 2018;19:397–408. https://doi.org/10.1007/s10592-017-1011-1.

66. Gissi C, Iannelli F, Pesole G. Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. Heredity. 2008;101:301–20. https://doi.org/10.1038/hdy.2008.62.

67. Jia W, Higgs PG. Codon usage in mitochondrial genomes: distinguishing context-dependent mutation from translational selection. Mol Biol Evol. 2008;25(2):339–51. https://doi.org/10.1093/molbev/msm259.

68. Hou H-Y, Chang R-Y, Cheng Y-N, Jang-Liaw N-H. Complete mitochondrial genome sequence for the Somali wild ass *Equus africanus somaliensis*. Conserv Genet Resour. 2019;11:413–7. https://doi.org/10.1007/s12686-018-1035-y.

69. Lamelas L, Aleix-Mata G, Rovatsos M, Marchal JA, Palomeque T, Lorite P, Sánchez A. Complete Mitochondrial Genome of Three Species of the Genus *Microtus* (Arvicolinae, Rodentia). Animals. 2020;10:2130. https://doi.org/10.3390/ani10112130.

70. Uddin A, Mazumder TH, Barbhuiya PA, Chakraborty S. Similarities and dissimilarities of codon usage in mitochondrial ATP genes among fishes, aves, and mammals. IUBMB Life. 2020;72:899–914. https://doi.org/10.1002/iub.2231.

71. De Mandal S, Mazumder TH, Panda AK, Kumar SN, Jin F. Analysis of synonymous codon usage patterns of HPRT1 gene across twelve mammalian species. Genomics. 2020;112:304–31. https://doi.org/10.3390/biology10111078.

72. Kerpedjiev P, Hammer S, Hofacker IL. Forna (force-directed RNA): Simple and effective online RNA secondary structure diagrams. Bioinformatics. 2015;31(20):3377–9. https://doi.org/10.1093/bioinformatics/btv372.

73. Jühling F, Mörl M, Hartmann RK, Sprinzl M, Stadler PF, Pütz J. tRNAdb 2009: Compilation of tRNA sequences and tRNA genes. Nucleic Acids Res. 2009;37:D159–62. https://doi.org/10.1093/nar/gkn772.

74. İbiş O. Mitogenome Characterization of Turkish Anatolian Donkey (*Equus asinus*) and Its Phylogenetic Relationships. Türkiye Tarımsal Araştırmalar Dergisi - Turkish J Agric Res. 2019;6(3):257–67. https://doi.org/10.19159/tutad.569435.

75. Wilkinson GS, Chapman AM. Length and sequence variation in evening bat D-loop mtDNA. Genetics. 1991;128:607–17. https://doi.org/10.1093/genetics/128.3.607.

76. Saccone C, Lanave C, Pesole G, Sbisa E. Peculiar features and evolution of mitochondrial genome in mammals. In: DiMauro S, Wallace DC, editors. Mitochondrial DNA in human pathology. New York: Raven Press; 1993. p. 27–37.

77. Ghivizzani SC, Mackay SLD, Madsen CS, Laipis PJ, Hauswirth WW. Transcribed heteroplasmic repeated sequences in the porcine mitochondrial DNA D-loop region. J Mol Evol. 1993;37:36–47. https://doi.org/10.1007/BF00170460.

Gutiérrez *et al. BMC Genomics*   (2023) 24:527

Page 20 of 21

78. Xu X, Arnason U. The complete mitochondrial DNA sequence of the horse, *Equus caballus*: extensive heteroplasmy of the control region. Gene. 1994;148:357–62. https://doi.org/10.1016/0378-1119(94)90713-7.

79. Sbisá E, Tanzariello F, Reyes A, Pesole G, Saccone C. Mammalian mitochondrial D-loop region structural analysis: identification of new conserved sequences and their functional and evolutionary implications. Gene. 1997;205(1–2):125–40. https://doi.org/10.1016/S0378-1119(97)00404-6.

80. Doda JN, Wright CT, Clayton DA. Elongation of displacement-loop strands in human and mouse mitochondrial DNA is arrested near specific template sequences. PNAS USA. 1981;78:6116–20. https://doi.org/10.1073/pnas.78.10.6116.

81. Douzery E, Randi E. The mitochondrial control region of Cervidae: Evolutionary patterns and phylogenetic content. Mol Biol Evol. 1997;14:1154–66. https://doi.org/10.1093/oxfordjournals.molbev.a025725.

82. Skorupski J. Characterisation of the Complete Mitochondrial Genome of Critically Endangered *Mustela lutreola* (Carnivora: Mustelidae) and Its Phylogenetic and Conservation Implications. Genes. 2022;13:125. https://doi.org/10.3390/genes13010125.

83. Shadel GS, Clayton DA. Mitochondrial DNA maintenance in vertebrates. Annu Rev Biochem. 1997;66:409–35. https://doi.org/10.1146/annurev.biochem.66.1.409.

84. Willerslev E, Thomas M, Gilbert P, Binladen J, Ho SYW, et al. Analysis of complete mitochondrial genomes from extinct and extant rhinoceroses reveals lack of phylogenetic resolution. BMC Evol Biol. 2009;9:95.

85. Stewart JB, Freyer C, Elson JL, Wredenberg A, Cansu Z, Trifunovic A, Larsson N-G. Strong Purifying Selection in Transmission of Mammalian Mitochondrial DNA. PLoS Biol. 2008;6(1):e10. https://doi.org/10.1371/journal.pbio.0060010.

86. Achilli A, Olivieri A, Soares P, Lancioni H, Kashani BH, Perego UA, et al. Mitochondrial genomes from modern horses reveal the major haplogroups that underwent domestication. PNAS USA. 2012;109(7):2449–54. https://doi.org/10.1073/pnas.1111637109.

87. Wei Q, Zhang H, Wu X, Sha W. The selective constraints of ecological specialization in Mustelidae on mitochondrial genomes. Mamm Res. 2020;65(1):85–92. https://doi.org/10.1007/s13364-019-00461-2.

88. James JE, Piganeau G, Eyre -walker A. The rate of adaptive evolution in animal mitochondria. Mol Ecol. 2016;25:67–78. https://doi.org/10.1111/mec.13475.

89. Zhuang X, Cheng CHC. ND6 gene lost and found: evolution of mitochondrial gene rearrangement in Antarctic notothenioids. Mol Biol Evol. 2010;27(6):1391–403. https://doi.org/10.1093/molbev/msq026.

90. Luo Y, Gao W, Gao Y, Tang S, Huang Q, et al. Mitochondrial genome analysis of *Ochotona curzoniae* and implication of cytochrome c oxidase in hypoxic adaptation. Mitochondrion. 2008;8(5–6):352–7. https://doi.org/10.1016/j.mito.2008.07.005.

91. Wang X, Zhou S, Wu X, Wei Q, Shang Y, Sun G, et al. High-altitude adaptation in vertebrates as revealed by mitochondrial genome analyses. Ecol Evol. 2021;11:15077–84. https://doi.org/10.1002/ece3.8189.

92. Weaver S, Shank SD, Spielman SJ, Li M, Muse SV, Pond SLK. Datamonkey 2.0: A modern web application for characterizing selective and other evolutionary processes. Mol Biol Evol. 2018;35:773–7. https://doi.org/10.1093/molbev/msx335.

93. Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Pond SLK. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. Mol Biol Evol. 2015;32:1342–53.

94. Murrell B, Weaver S, Smith MD, Wertheim JO, Murrell S, Aylward A, et al. Gene-wide identification of episodic selection. Mol Biol Evol. 2015;32(5):1365–71. https://doi.org/10.1093/molbev/msv035.

95. Wertheim JO, Murrell B, Smith MD, Pond SLK, Scheffler K. RELAX: detecting relaxed selection in a phylogenetic framework. Mol Biol Evol. 2015;32(3):820–32. https://doi.org/10.1093/molbev/msu400.

96. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Pond SLK. Detecting individual sites subject to episodic diversifying selection. PLoS Genet. 2012;8(7):e1002764. https://doi.org/10.1371/journal.pgen.1002764.

97. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, et al. FUBAR: a fast, unconstrained Bayesian approximation for inferring selection. Mol Biol Evol. 2013;30(5):1196–205. https://doi.org/10.1093/molbev/mst030.

98. Pond SLK, Frost SDW. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol Biol Evol. 2005;22(5):1208–22. https://doi.org/10.1093/molbev/msi105.

99. Ge R-L, Cai Q, Shen Y-Y, San A, Ma L, Zhang Y, et al. Draft genome sequence of the Tibetan antelope. Nat Commun. 2013;4:1858. https://doi.org/10.1038/ncomms2860.

100. Li X-D, Jiang G-F, Yan L-Y, Li R, Mu Y, Deng W-A. Positive Selection drove the adaptation of mitochondrial genes to the demands of flight and high-altitude environments in grasshoppers. Front Genet. 2018;9:605. https://doi.org/10.3389/fgene.2018.00605.

101. Panis DD, Lambertucci SA, Wiemeyer G, Dopazo H, Almeida FC, Mazzoni CJ, et al. Mitogenomic analysis of extant condor species provides insight into the molecular evolution of vultures. Sci Rep. 2021;11:17109. https://doi.org/10.1038/s41598-021-96080-6.

102. Niu Y, Zhang X, Xu T, Li X, Zhang H, Wu A, et al. Physiological and biochemical adaptations to high altitude in tibetan frogs Nanorana parkeri. Front Physiol. 2022;3:942037. https://doi.org/10.3389/fphys.2022.942037.

103. Shen Y-Y, Liang L, Zhu Z-H, Zhou W-P, Irwin DM, Zhang Y-P. Adaptive evolution of energy metabolism genes and the origin of flight in bats. PNAS USA. 2010;107:8666–71. https://doi.org/10.1073/pnas.0912613107.

104. Shen Y-Y, Shi P, Sun Y-B, Zhang Y-P. Relaxation of selective constraints on avian mitochondrial DNA following the degeneration of flight ability. Genome Res. 2009;19:1760–5. https://doi.org/10.1101/gr.093138.109.

105. Castoe TA, Jiang ZJ, Gu W, Wang ZO, Pollock DD. Adaptive evolution and functional redesign of core metabolic proteins in snakes. PLoS One. 2008;3:e2201. https://doi.org/10.1371/journal.pone.0002201.

106. Ramos B, González-Acuña D, Loyola DE, Johnson WE, Parker PG, Massaro M, et al. Landscape genomics: natural selection drives the evolution of mitogenome in penguins. BMC Genomics. 2018;19:53. https://doi.org/10.1186/s12864-017-4424-9.

107. Baeza JA. Mitochondrial genomes assembled from non-invasive eDNA metagenomic scat samples in the endangered Amur tiger *Panthera tigris altaica*. PeerJ. 2022;10:e14428. https://doi.org/10.7717/peerj.14428.

108. Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, et al. MITOS: Improved de novo metazoan mitochondrial genome annotation. Mol Phylogenet Evol. 2013;69(2):313–9. https://doi.org/10.1016/j.ympev.2012.08.023.

109. Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, De Castro E, et al. ExPASy: SIB bioinformatics resource portal. Nucleic Acids Res. 2012;40(W1):W597–603.

110. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018;35(6):1547–9. https://doi.org/10.1093/molbev/msy096.

111. Zheng S, Poczai P, Hyvönen J, Tang J, Amiryousefi A. Chloroplot: an online program for the versatile plotting of organelle genomes. Front Genet. 2020;11:576124. https://doi.org/10.3389/fgene.2020.576124.

112. Stothard P. The Sequence Manipulation Suite: JavaScript programs for analyzing and formatting protein and DNA sequences. Biotechniques. 2000;28:1102–4. https://doi.org/10.2144/00286ir01.

113. Cucini C, Leo C, Iannotti N, Boschi S, Brunetti C, Pons J, et al. EZmito: a simple and fast tool for multiple mitogenome analyses. Mitochondrial DNA B. 2021;6(3):1101–9. https://doi.org/10.1080/23802359.2021.1899865.

114. Jühling F, Pütz J, Bernt M, Donath A, Middendorf M, Florentz C, Stadler PF. Improved systematic tRNA gene annotation allows new insights into the evolution of mitochondrial tRNA structures and into the mechanisms of mitochondrial genome rearrangements. Nucleic Acids Res. 2012;40(7):2833–45. https://doi.org/10.1093/nar/gkr1131.

115. Donath A, Jühling F, Al-Arab M, Bernhart SH, Reinhardt F, Stadler PF, et al. Improved annotation of protein-coding genes boundaries in metazoan mitochondrial genomes. Nucleic Acids Res. 2019;47(20):10543–52. https://doi.org/10.1093/nar/gkz833.

116. Bikandi J, San Millán R, Rementeria A, Garaizar J. *In silico* analysis of complete bacterial genomes: PCR, AFLP-PCR, and endonuclease restriction. Bioinformatics. 2004;20:798–9. https://doi.org/10.1093/bioinformatics/btg491.

117. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27(2):573–80. https://doi.org/10.1093/nar/27.2.573.

118. Bellaousov S, Reuter JS, Seetin MG, Mathews DH. RNAstructure: web servers for RNA secondary structure prediction and analysis. Nucleic Acids Res. 2013;41:W471–4. https://doi.org/10.1093/nar/gkt290.

Gutiérrez *et al. BMC Genomics*   (2023) 24:527

Page 21 of 21

119. Tan MH, Gan HM, Schultz MB, Austin CM. MitoPhAST, a new automated mitogenomic phylogeny tool in the post-genomic era with a case study of 89 decapod mitogenomes including eight new freshwater crayfish mitogenomes. Mol Phylogenet Evol. 2015;85:180–8. https://doi.org/10.1016/j.ympev.2015.02.009.

120. Sievers F, Higgins DG. Clustal Omega for making accurate alignments of many protein sequences. Protein Sci. 2017;27(1):135–45. https://doi.org/10.1002/pro.3290.

121. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25:1972–3. https://doi.org/10.1093/bioinformatics/btp348.

122. Abascal F, Zardoya R, Posada D. ProtTest: selection of best-fit models of protein evolution. Bioinformatics. 2005;21:2104–5. https://doi.org/10.1093/bioinformatics/bti263.

123. Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32:268–74. https://doi.org/10.1093/molbev.

124. Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. Genomics Proteomics Bioinformatics. 2010;8(1):77–80. https://doi.org/10.1016/S1672-0229(10)60008-3.

125. Gao F, Chen C, Arab DA, Du Z, He Y, Ho SYW. EasyCodeML: A visual tool for analysis of selection using CodeML. Ecol Evol. 2019;9:3891–8. https://doi.org/10.1002/ece3.5015.

126. Yang Z, Nielsen R. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol Biol Evol. 2002;19(6):908–17. https://doi.org/10.1093/oxfordjournals.molbev.a004148.

127. Yang Z. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol Biol Evol. 1998;15(5):568–73. https://doi.org/10.1093/oxfordjournals.molbev.a025957.

128. Anisimova M, Bielawski JP, Yang Z. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. Mol Biol Evol. 2001;18:1585–92.

129. Yang Z, Wong WS, Nielsen R. Bayes empirical Bayes inference of amino acid sites under positive selection. Mol Biol Evol. 2005;22:1107–18. https://doi.org/10.1093/molbev/msi097.

130. Pond SLK, Frost SDW, Muse SV. HyPhy: hypothesis testing using phylogenies. Bioinformatics. 2005;21(5):676–9. https://doi.org/10.1093/bioinformatics/bti079.

Publisher's Note