


SCIENTIFIC REPORTS



OPEN

Mitogenome Sequencing in the Genus *Camelus* Reveals Evidence for Purifying Selection and Long-term Divergence between Wild and Domestic Bactrian Camels

Elmira Mohandesan^{1,2,3}, Robert R. Fitak⁴, Jukka Corander^{5,6}, Adiya Yadamsuren⁷, Battsetseg Chuluunbat⁸, Omer Abdelhadi⁹, Abdul Raziq¹⁰, Peter Nagy¹¹, Gabrielle Stalder¹, Chris Walzer^{1,12}, Bernard Faye¹³ & Pamela A. Burger¹ 

The genus *Camelus* is an interesting model to study adaptive evolution in the mitochondrial genome, as the three extant Old World camel species inhabit hot and low-altitude as well as cold and high-altitude deserts. We sequenced 24 camel mitogenomes and combined them with three previously published sequences to study the role of natural selection under different environmental pressure, and to advance our understanding of the evolutionary history of the genus *Camelus*. We confirmed the heterogeneity of divergence across different components of the electron transport system. Lineage-specific analysis of mitochondrial protein evolution revealed a significant effect of purifying selection in the concatenated protein-coding genes in domestic Bactrian camels. The estimated $dN/dS < 1$ in the concatenated protein-coding genes suggested purifying selection as driving force for shaping mitogenome diversity in camels. Additional analyses of the functional divergence in amino acid changes between species-specific lineages indicated fixed substitutions in various genes, with radical effects on the physicochemical properties of the protein products. The evolutionary time estimates revealed a divergence between domestic and wild Bactrian camels around 1.1 [0.58–1.8] million years ago (mya). This has major implications for the conservation and management of the critically endangered wild species, *Camelus ferus*.

The genus *Camelus* is an interesting model to study the role of natural selection in the mitogenome as a driving force for adaptive divergence. The three extant *Camelus* species exist in very different habitats, each encompassing environmental pressures potentially resulting in the evolution of adaptations specific to each species. Dromedaries (*Camelus dromedarius*) inhabit semi-arid and hot desert regions of North and East Africa, Arabian Peninsula and southwest Asia^{1,2}, whereas domestic Bactrian camels (*Camelus bactrianus*) are distributed over

¹Research Institute of Wildlife Ecology, Vetmeduni Vienna, Savoyenstrasse 1, 1160, Vienna, Austria. ²Institute of Population Genetics, Vetmeduni Vienna, Veterinärplatz 1, 1210, Vienna, Austria. ³Institute for Molecular Evolution and Development, University of Vienna, Althanstrasse 14, 1090, Vienna, Austria. ⁴Department of Biology, Duke University, Durham, NC, 27708, USA. ⁵Department of Biostatistics, University of Oslo, N-0317, Oslo, Norway. ⁶Department of Mathematics and Statistics, University of Helsinki, FIN-00014, Helsinki, Finland. ⁷Mammalian Ecology Laboratory, Institute of Biology, Mongolian Academy of Sciences, Peace avenue-54b, Bayanzurh district, Ulaanbaatar, 210351, Mongolia. ⁸Laboratory of Genetics, Institute of Biology, Mongolian Academy of Sciences, Peace avenue-54b, Bayanzurh district, Ulaanbaatar, 210351, Mongolia. ⁹University of Khartoum, Department for Meat Sciences, Khartoum, Sudan. ¹⁰Lasbela University of Agriculture, Water and Marine Sciences, Regional Cooperation for Development (RCD) Highway, Uthal, Pakistan. ¹¹Farm and Veterinary Department, Emirates Industry for Camel Milk and Products, PO Box 294239, Dubai, Umm Nahad, United Arab Emirates. ¹²International Takhi Group - Mongolia, Baigal Ordon, Ulaanbaatar, Mongolia. ¹³CIRAD-ES, UMR 112, Campus International de Baillarguet, TA C/112A, 34398, Montpellier, France. Correspondence and requests for materials should be addressed to E.M. (email: elmira.mohandesan@univie.ac.at) or P.A.B. (email: pamela.burger@vetmeduni.ac.at)

central (Kazakhstan, Iran) and eastern (Russia, Mongolia, China) Asia, ranging from rocky mountains to flat and cold (semi-) deserts^{3–5}. The wild two-humped camels (*Camelus ferus*) once were distributed over eastern and central Asia, adapted to arid and cold plains and hills^{4,6}. Today, this critically endangered wild species (IUCN 2014) can only be found in the Chinese Taklimakan and Lop Noor deserts and in the Mongolian Great Gobi Strictly Protected Area “A”, with a population census indicating as few as 2,000 individuals remain^{7,8}.

The ancestors of *Camelus* emerged on the North American continent 35–40 mya. After splitting into New World (Lamini) and Old World (Camelini) camels around 16.3 mya^{9,10}, the latter migrated via the Bering land bridge to the eastern hemisphere (the Old World), while the ancestors of llamas and alpacas spread to South America. Although the differentiation of the two domestic Old World camel species, dromedary and Bactrian camel, has long been established and estimated at around 4.4–8 mya^{9,10}, the wild two-humped camel (*C. ferus*) has been recognized only lately as a separate species^{11–14}. However, a reliable estimate for the divergence time between domestic and wild Bactrian camels has been missing. Contrary to the last remaining wild relatives of the domestic Bactrian camel, the wild ancestors of the dromedary became extinct less than one millennium after the domestic form appeared (3,000–4,000 years ago), contributing to its domestication with multiple introgressions¹⁵.

Inferring the evolutionary and demographic history of wild and domestic species using mitochondrial DNA (mtDNA) has been established for a long time¹⁶. The relatively small mitogenome (16–17 kb) encodes a suite of 13 proteins, which interact with the nuclear DNA to build the oxidative phosphorylation (OXPHOS) pathway. Although these proteins are involved in key processes in the cells, the evolution of mtDNA has long been considered nearly neutral, and many studies have neglected the direct impact of positive and/or negative selection in shaping mitogenome diversity. A growing number of studies, however, favor the view that variation in the mitogenome is not selectively neutral (e.g., refs 17–21). Purifying selection has been implicated as a predominant force in shaping mitogenome evolution by limiting the accumulation of deleterious mutations at evolutionary constrained sites (e.g., cytochrome oxidase subunits (COX), NADH dehydrogenase subunits (ND), and cytochrome b (CYTB))^{17,22–25}.

Positive selection has been associated with shaping the mitogenome diversity in various species, such as snakes²⁶, salmonids²⁷, killer whales²⁸, and across the mammalian phylogeny^{29,30}. Positive selection in the mitochondrial genes (e.g., COX, ATPase complexes, CYTB) has been shown to be involved in adaptation to different environmental pressures such as high altitude^{31–33} or temperature^{18,34–36}.

The most common approach in detecting selection in protein-coding genes involves estimating the rates of nonsynonymous (dN) and synonymous (dS) nucleotide substitutions and calculating the rates ratio as $\omega = dN/dS$. In the McDonald-Kreitman (MK) test, the neutrality index (NI), or the excess of polymorphisms within species compared with the fixed substitutions between species, is used as an indicator to investigate the role of selection in shaping mitochondrial divergence between species³⁷.

In the case of mtDNA, its high mutation rate can create intensive background noise hiding potential positively selected sites of adaptive value. Moreover, the commonly used methods such as dN/dS ratio and McDonald-Kreitman test are biased towards moderately conserved protein-coding genes³⁸. The accurate diagnosis of individual amino acid changes with adaptive or non-adaptive effects is mathematically difficult because multiple nonsynonymous changes are required to satisfy the condition of $dN > dS$ in a statistically significant manner³⁹. One way to circumvent this problem is to analyze specific amino acid changes with adaptive values. Insights about selection acting on specific mitochondrial genes can be gained from branch-site tests on single codons⁴⁰ and also by estimating lineage-specific deviations of dN/dS ratios in single genes⁴¹. Additional methods target the detection of positive selection at the mitochondrial level by comparing patterns of physicochemical changes of amino acids in the context of their effects on protein structure^{42–44}. The technique of targeting changes in amino acid properties should be more sensitive than the dN/dS approach^{45,46}, and results in a high degree of resolution for the diagnosis of individual amino acid changes with adaptive values.

In this study, we screened for signatures of natural selection and divergence in 27 mitogenomes of the three different Old World camelid species. Specifically, we used a hierarchical approach and investigated (i) mitogenome evolution in the genus *Camelus* under a neutral model versus a model of natural selection; (ii) the evolution of mitochondrial protein-coding genes focusing on lineage-specific evolutionary changes; (iii) selection pressure on single sites in protein-coding genes within the *Camelus* lineage using a codon-based analysis; and (iv) changes in the physicochemical properties of functional proteins caused by amino acid replacements.

Finally, to gain knowledge about the evolutionary history of Old World camelids we estimated divergence times between domestic Bactrian camels and their wild, critically endangered relatives as well as between one- and two-humped camels using a Bayesian approach. Our study contributes to the understanding of evolutionary processes in Old World camels, and we provide support for the conservation and management of the critically endangered wild species *Camelus ferus*.

Results

Mitochondrial gene diversity and differentiation among Camelus. We analyzed 27 complete mitogenomes of Old World camels (Table S1), including three sequences from GenBank (*C. dromedarius*: NC_009849.1, *C. bactrianus*: NC_009628.2, *C. ferus*: NC_009629.2) to investigate signals of selection in *Camelus* possibly due to adaptation to different environmental regimes. With a mean read coverage of 20X [10–26X] after correcting for a maximal read coverage of 30, the alignment of all mitogenomes displayed a total length of 16,390 bp excluding repetitive regions and contained 1,347 variable sites (1,262 parsimony-informative, 85 singletons). We calculated the mitogenome diversity within each of the three Old World camel species and found higher haplotype (H_d) and nucleotide diversity (π) in dromedaries ($H_d = 1.0$, $\pi = 0.00169$) compared to domestic ($H_d = 0.952$, $\pi = 0.00132$) and wild ($H_d = 0.822$, $\pi = 0.00131$) Bactrian camels, respectively (Table 1). Tajima's D^{47} (-2.003 , $P < 0.01$) as well as Fu and Li's F^{48} (-2.580 , $P < 0.02$) were significantly negative in dromedaries, whereas in wild camels they were positive (2.331, $P < 0.05$ and 1.772; $P < 0.02$, respectively). For the domestic Bactrian

| Species | n | Length ^a (bp) | S | h | H _d (SD) | k | θ _π (SD) | θ _S (SD) | Tajima's D | Fu and Li's F test | Ramos-Onsins and Rozas (R2) statistic | Fu's Fs statistic |
|-----------------------|----|--------------------------|-----|----|---------------------|--------|---------------------|---------------------|-------------|--------------------|---------------------------------------|-------------------|
| <i>C. dromedarius</i> | 10 | 16375 | 131 | 10 | 1.0 (0.045) | 27.644 | 0.00169 (0.00097) | 0.00283 (0.00116) | -2.00336** | -2.57981** | 0.2330 ns | -1.206** |
| <i>C. bactrianus</i> | 7 | 16385 | 53 | 6 | 0.952 (0.096) | 21.619 | 0.00132 (0.00028) | 0.00132 (0.00061) | -0.00364 ns | 0.29791 ns | 0.1555 ns | 1.466 ns |
| <i>C. ferus</i> | 10 | 16383 | 41 | 5 | 0.822 (0.097) | 21.444 | 0.00131 (0.00017) | 0.00088 (0.00038) | 2.33046* | 1.77232** | 0.2547 ns | 6.667 ns |

Table 1. Mitogenome diversity in the three Old World camel species. ^aAlignment excluding repetitive parts of the CR, gaps and missing data; n = number of samples; S = segregating sites; h = number of haplotypes; H_d = haplotype diversity; k = average number of nucleotide differences; θ_π = theta estimator based on the mean number of nucleotide differences; θ_S = Watson's theta estimator based on the number of segregating sites; SD = standard deviation values. Statistical significance: *P < 0.05; **P < 0.02, ns = not significant.

camels we could not reject the hypothesis of selective neutrality for Tajima's D (-0.004, P > 0.10) and Fu and Li's F (0.298, P > 0.10) (Table 1). The tests for past population growth based on the number of pairwise nucleotide differences considering a small sample size, *i.e.*, Ramos-Onsins and Rozas R2 statistic⁴⁹, as well as Fu's Fs (-1.206, P < 0.002) resulted in a significant expansion signal only for dromedaries (Table 1).

The average number of nucleotide differences between species (n) and the net nucleotide substitutions per site (D_a) between the pairwise species comparisons were calculated as n = 520.9/588.9 and D_a = 0.069/0.070 (dromedary *vs.* domestic/wild Bactrian camel) and n = 160.7 and D_a = 0.018 (domestic *vs.* wild Bactrian camel) (Tables S2–4). We observed substantial differentiation in protein-coding genes between the pairwise comparisons of Old World camels (dromedary *vs.* Bactrian camel, dromedary *vs.* wild camel and Bactrian *vs.* wild camel). This resulted in 862, 874 and 223 fixed synonymous and nonsynonymous substitutions (D_s + D_n), leading to a permanent replacement (D_n) of 84, 90 and 25 amino acids, respectively (Tables S2–4). In all pairwise comparisons, net divergence (D_a) varied substantially between genes; however, it was lower in the two rRNA subunits (12S, 16S) and in the concatenated tRNAs in comparison with protein-coding genes (Tables S2–4).

To investigate the heterogeneity of divergence across gene classes with different functions we examined the divergence level in protein-coding genes classified into different mitochondrial components of the electron transport system (ETS) complexes. In each pairwise comparison among the three species, we observed the highest differentiation in ETS III (D_a(db) = 0.092, D_a(df) = 0.099, D_a(bf) = 0.023) and the lowest in ETS IV (D_a(db) = 0.068, D_a(df) = 0.069, D_a(bf) = 0.017). The D_a(bf) between domestic and wild Bactrian camels showed low differentiation among all ETS complexes (D_aETS I = 0.022, D_aETS III = 0.023, D_aETS IV = 0.017, D_aETS V = 0.018) (Table S5).

Signals of purifying selection in the genus *Camelus*. *Mitogenome protein evolution under a neutral model versus a model of natural selection.* The global MK test³⁷ performed pairwise on the concatenated protein-coding genes (11,379 bp) in the three Old World camel species showed an excess of amino acid polymorphism in dromedary and domestic Bactrian camels (P_n/P_s = 0.21) as well as a deficiency of interspecific fixed amino acid replacements (D_n/D_s = 0.10). This resulted in a NI of 2.0 (P = 0.006) for dromedary and domestic Bactrian camels, indicative of purifying selection. A similar NI of 2.17 (P = 0.035) was detected in the pairwise comparison between domestic and wild Bactrian camels. In contrast, between dromedary and wild Bactrian camels we could not identify any signal of positive or negative selection (NI = 1.41, P = 0.244). Moreover, we compared the one- and two-humped camels, resulting in NI = 1.46 (P = 0.041) (Table S6).

Next, we investigated lineage-specific mtDNA protein evolution by performing the "extended MK test"⁵⁰ for both concatenated protein-coding genes as well as each gene separately. In dromedaries, the NI indicated an excess of nucleotide polymorphism in ND6 in respect to its divergence from the domestic (NI = 13.33, P = 0.009) and wild Bactrian camels (NI = 14.66, P = 0.007), suggesting purifying selection against mildly deleterious mutations in this gene (Tables S7 and 8). In domestic Bactrian camels, we detected significant effects of purifying selection in concatenated protein-coding genes with respect to dromedaries (NI = 3.47, P = 0.001) and wild Bactrian camels (NI = 2.97, P = 0.0083) (Tables S9 and 10). The excess of replacement polymorphism was not significant in the wild Bactrian camel lineage (Tables S11 and 12). The significant results of the MK test (global and extended) indicating signals of purifying selection in the mitogenomes of the genus *Camelus* are summarized in Table 2.

Evolution of mitochondrial protein-coding genes focusing on the lineage-specific evolutionary changes. To test for selective pressure acting on a particular *Camelus* lineage, we used a maximum likelihood (ML) approach and examined variation in the ratio of dN/dS (ω) on the codon-based alignments of ten different mammalian species including Old World camels (*C. dromedarius*, *C. bactrianus*, *C. ferus*), New World camels (*Lama glama*, *L. guanicoe*, *Vicugna pacos*, *V. vicugna*), *Bos taurus*, *Pantholops hodgsonii* and *Homo sapiens* (see Table S1 for accession numbers). In the background model, the uniform dN/dS (ω) ratio across the reconstructed phylogeny (HKY + G + I) of concatenated protein-coding genes was estimated at ω = 0.035. In the species-specific model, which was a significant improvement over the background model (likelihood ratio test [LRT], χ² = 179.7, df = 64, P < 0.001), the highest dN/dS ratio (ω = 0.063) was estimated in wild camels followed by dromedaries (ω = 0.049) and domestic Bactrian camels (ω = 0.041) (Fig. 2). The two-humped camel lineage displayed an ω of 0.025 preceding its separation into the domestic and wild Bactrian camels.

*Selection pressure on single sites in protein-coding genes within the *Camelus* lineage using a codon-based analysis.* Using the branch-site model A (foreground lineage *vs.* background lineages) implemented in CODEML

| | Selection test | mtDNA gene | NI (<i>P</i>) |
|---------------------------------------|----------------|------------|-----------------|
| dromedary vs. domestic Bactrian camel | g-MKT | PC genes | 2 (0.006**) |
| | e-MKT | ND6 | 13.33 (0.009**) |
| dromedary vs. wild Bactrian camel | e-MKT | ND6 | 14.66 (0.007**) |
| domestic Bactrian camel vs. dromedary | e-MKT | PC genes | 3.47 (0.001**) |
| domestic vs. wild Bactrian camel | e-MKT | PC genes | 2.17 (0.035*) |
| domestic vs. wild Bactrian camel | e-MKT | PC genes | 2.97 (0.008**) |
| one- vs. two-humped camels | g-MKT | PC genes | 1.48 (0.041*) |

Table 2. Summary of the significant results of the global and extended MK tests indicating purifying selection in *Camelus* mitogenomes. The two-tailed *P*-values of Fisher's exact tests are shown. *P*-values with asterisk are significant with the following significant levels **P* < 0.05; ***P* < 0.01. PC = concatenated protein coding genes, g-MKT = global MKT, e-MKT = extended MKT, NI = neutrality index.

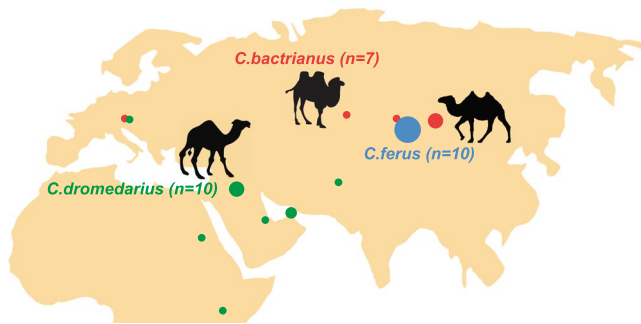


Figure 1. Geographical locations of the samples used in this study, including three published mitogenomes obtained from GenBank. The size of each circle is proportional to the sample size for that location. Colors correspond to the three Old World camel species, and the number of individuals per species are shown within parentheses. Detailed sample information is given in Table S1. The map and figures are adopted from <https://openclipart.org> and are licensed under Creative Commons Zero 1.0 License.

(PAML)⁵¹, the Bayes empirical Bayes (BEB) analysis indicated a total of 18 sites under positive selection in the entire *Camelus* lineage (Table S13). We detected 4 sites (in ND2, ATP6), 10 sites (in ND1, ND2, COX3, ND3, ND4, ND5) and 12 sites (in ND2, ATP8, ND4, ND5, CYTB) under positive selection in dromedaries, Bactrian camels and wild camels, respectively (Table S13). However, the significance of the detected sites could not be confirmed in the subsequent LRTs (Table S13).

Changes in the physicochemical properties of functional proteins by amino acid replacements. The analysis of functional divergence in amino acid changes was performed using TreeSAAP software⁵², which takes into account the magnitude of the impact of the amino acid replacement on local physicochemical properties of the protein. Changes with a magnitude ≥ 6 and *P* < 0.001 were considered as indicating positive directional selection for a given physicochemical property⁵². From a total of 84 (dromedary vs. Bactrian camel), 90 (dromedary vs. wild camel) and 25 (domestic vs. wild Bactrian camel) fixed nonsynonymous substitutions between *Camelus* species, we found evidence of positively selected sites in ATP8 (1 site), ATP6 (2 sites), ND3 (2 sites), ND5 (15 sites) and ND6 genes (4 sites) (Table S14). In these protein-coding genes, we observed 12, 14 and 12 ancestral amino acids in dromedary, domestic and wild Bactrian camels, respectively (Table S14). These amino acid changes affect physicochemical properties such as alpha-helical tendencies (P_{α}), equilibrium constant (ionization of COOH) (pK'), polar requirement (P_r), surrounding hydrophobicity (H_p), power to be at the middle of alpha-helix (α_m), isoelectric point (pHi), and solvent accessible reduction ratio (R_a).

Divergence time estimations among Camelidae. We used a Bayesian approach to estimate the 95% credibility intervals (CIs) of divergence times between the camelid species (Fig. 2 and Table 3). We performed five independent runs of MULTIDIVTIME⁵³, all of which indicated good convergence (all corrected scale reduction factors = 1.0). Furthermore, the multivariate potential scale reduction factor (MPSRF)^{54,55} was 1.001 and 1.01 for analyses including all sites and only first and second positions, respectively, supporting that the results across all runs had reached a stationary distribution. The mean and 95% CI for the divergence between Camelidae and Bovidae was estimated at 41.4 [24.7–60.6] mya. Within the Camelidae, the mean divergence time between the Camelini and Lamini tribes was 15.8 [CI 9.2–23.2] mya. The mean divergence between one-humped and two-humped camels occurred around 5.3 [CI 2.9–7.7] mya, and domestic and wild Bactrian camels split at about 1.1 [CI 0.58–1.8] mya. The estimates for divergence times when restricted to only the first and second codon positions were similar, albeit slightly higher than estimates from all sites (Table 3). For example, this analysis suggested the divergence between domestic and wild Bactrian camels had occurred 2.3 [CI 0.71–4.1] mya.

| Phylogenetic node | All sites | First and second protein-coding positions only |
|--|---------------------|--|
| Camelidae & Bovidae | 41.4 [24.7–60.6] | 40.1 [24.0–58.9] |
| Camelini & Lamini | 15.8 [9.2–23.2] | 19 [11.5–28.5] |
| <i>C. dromedarius</i> & [<i>C. bactrianus</i> + <i>C. ferus</i>] | 5.3 [2.9–7.7] | 7.3 [3.7–11.6] |
| <i>C. bactrianus</i> & <i>C. ferus</i> | 1.1 [0.58–1.8] | 2.3 [0.71–4.1] |

Table 3. Divergence time estimates among Camelidae. Mean and 95% posterior credibility intervals (within brackets) of divergence time estimates for the four nodes of interest produced using MULTIDIVTIME. All values are given in millions of years ago (mya).

The mtDNA saturation analysis showed that the %Ti did not approach 50% until a sequence divergence of around 0.23 and 0.12 for all sites and first and second positions, respectively (Figure S1). The mean sequence divergence across all genes for Bovidae *vs.* Camelidae pairwise comparisons was 0.23 and 0.10 for all sites and first and second positions, respectively. This suggests that the Bovidae-Camelidae divergence time may be underestimated due to saturation, whereas shallower nodes are less likely to be affected.

Discussion

We investigated the evolutionary history of Old World camelids and studied the role of natural selection in shaping mitochondrial diversity. The three species are well-adapted to different environmental regimes as their habitats are spread over different altitudes and temperatures. Particularly, we were interested in estimating the divergence time between the domestic and wild Bactrian camel because a reliable estimate had not been obtained previously.

Diversity and differentiation in the mitochondrial genes among Old World camels. We observed higher haplotype and nucleotide diversity in the ten investigated dromedary mitogenomes compared to the seven domestic and ten wild Bactrian camels (Table 1), respectively. The geographically larger distribution of the dromedary samples over three continents and seven countries *versus* the domestic Bactrian camel specimen collected from Mongolia, Kazakhstan and Austria (with potential Mongolian origin) in principle could account for these differences. Similarly, the wild two-humped camel samples originated from one national park, the Strictly Protected Area “A” in the Mongolian Gobi desert (Table S1), a consequence of their small census size and limited geographic habitat in Mongolia and China.

During the process of domestication, population growth and dispersion of animals across a wider geographic range can be captured from molecular signals of sudden expansion⁵⁶. In this study we obtained significant negative values of Tajima’s *D*, and Fu and Li’s *F* test in dromedaries (Table 1), which can indicate demographic expansion and/or positive selection. Evidence for recent population expansion in the context of dromedary domestication was confirmed by partial mitochondrial ND5, CYTB and CR analysis from over 600 individuals throughout their range¹⁵. On the contrary, the positive values of Tajima’s *D* and Fu and Li’s *F* test in the wild two-humped camels are indicative of a decrease in population size and/or balancing selection. These results are in agreement with a reduced genetic variation and small effective population size observed in wild Bactrian camels^{7,57}.

Overall, we calculated a mitogenome differentiation of 6.9% and 7% between dromedaries and domestic and wild Bactrian camels (Tables S2 and 3), respectively; whereas the sequence differences among two-humped camels showed 1.8% (Table S4). This corresponds to previous estimates (1.9%–2%) on partial mitochondrial genome sequences including ND5, CYTB and CR^{13,14}.

Investigating the heterogeneity of divergence across the different gene classes the rRNAs (12S and 16S) and tRNAs, which are required for mitochondrial protein synthesis, were highly conserved. This is in agreement with general slow divergence rates of mitochondrial RNA genes observed in other mammals^{58–62}. The observed rates of molecular evolution in different protein-coding genes of the ETS complexes reflected the differences in their functional constraints. In the ETS complexes the rates of molecular evolution followed the order CYTB > NDs > ATPs > COXs, with the majority of fixed differences observed in CYTB and the fewest in the COX genes. Functional modifications in the CYTB protein may be involved in physiological adaptation to different thermal environments^{25,30,36}. The high variation observed in CYTB in the *Camelus* lineage could thus be explained with more specialized metabolic requirements in camels, like the adaptation to low-energy diet or living at higher altitudes in the case of Bactrian camels. A pattern of high variation in CYTB was also observed in other species with specialized metabolic requirements, such as the elephant, dugong, sloth and pangolin²⁹. Higher rates of evolution in ND genes compared to the ATP and COX were reported previously in different species of insects, fish and mammals^{63,64}. As reported in other studies, nonsynonymous mutations in the ND genes are generally less deleterious than mutations in the COX genes, which oftentimes greatly alter functional properties of the protein and fitness^{24,31,65,66}. When we took into account the different sequence lengths of the ETS complexes, we observed the highest number of fixed nonsynonymous substitutions in the ATP genes in pairwise comparison between dromedary *versus* domestic and wild Bactrian. In all comparisons, the COX genes contained the lowest number of nonsynonymous substitutions, indicating their high evolutionary constraint in the *Camelus* lineage.

Signatures of purifying selection in the mitogenomes of *Camelus*. *Mitogenome protein evolution under a model of natural selection.* The pattern of intraspecific polymorphism in dromedary and interspecific divergence to domestic/wild Bactrian camel indicated the presence of purifying selection in the ND6 gene. This result showed that selection prevents the accumulation of slightly deleterious mutations in ND6, which is suggested to be one of the active subunits of the enzyme involved in OXPHOS pathways⁶⁷. No evidence of selection was found in other protein-coding genes in dromedaries, probably a result of insensitivity of the extended MK test for the restricted number of polymorphisms and divergence in species with recent population expansion³⁴. In domestic Bactrian camels, we did not detect signals of purifying selection in any particular gene; however, we observed a significant level of purifying selection over the concatenated protein-coding regions.

Evolution of mitochondrial protein-coding genes focusing on the lineage-specific evolutionary changes. The low dN/dS ratio estimated for the *Camelus* lineage ($dN/dS = 0.033$) indicated the effect of purifying selection on the mitochondrial protein-coding genes in this lineage. Purifying selection has been shown to be the main driver in shaping mitogenome diversity^{17,22–25}. However, it has been observed that domesticated species have a higher dN/dS ratio than their wild counterparts; a result possibly attributed to relaxed purifying selection during domestication (e.g., yak⁶⁸, goat⁶⁹). We observed this pattern in South American camelids, but contrary to this hypothesis, in wild two-humped camels we estimated slightly higher dN/dS ratio ($dN/dS = 0.063$) in comparison to the two domestic species (dromedaries $dN/dS = 0.049$; Bactrian camels $dN/dS = 0.041$). Wild camels are a sister species to the domestic Bactrian camels and have never been domesticated. Their original population size has been reduced over 80% during the last century and probably numbers less than 2,000 individuals^{7,8}. The current extremely small population size might be a reason for the relaxed pressure of purifying selection acting on the mitochondrial protein-coding genes, which could be explained by the stronger effect of genetic drift than natural selection in populations with smaller size^{70–73}. Furthermore, compared with other domesticated ungulates, domesticated camels (dromedaries and Bactrian camels) have maintained unusually high levels of genetic variation and lack the extensive secondary bottlenecks experienced during breed formation¹⁵. The combination of the small population size in wild camels and the chance of more effective purifying selection in the domesticated species could be responsible for this result. Nonetheless, we conclude that purifying selection acts to maintain the stability and efficiency of the OXPHOS system, which are the requirements for aerobic cell respiration and energy production.

*Selection pressure on single sites in protein-coding genes within *Camelus*.* Despite the low dN/dS ratio values across the examined *Camelus* phylogeny indicating purifying selection, the BEB analysis of the branch-site model A recovered 18 codons potentially under positive selection (BEB > 50%, including one codon BEB > 95%) in seven genes (Table S13). However, their significance could not be further confirmed with the LRT. The BEB approach is not a direct estimator of positive selection and is used to allow sampling error in parameter estimates. Considering the small sample size and possibly large parameter estimate errors, there is very little power to detect individual sites under selection. Similarly, Matosiuk *et al.*⁵⁸ identified 12 sites in roe deer lineage that may have been positively selected, which their significance could not be confirmed with subsequent LRT.

Changes in the physicochemical amino acid properties of functional proteins. We detected candidate positions for positive selection at fixed nonsynonymous substitutions in ATP8 (1 site), ATP6 (2 sites), ND3 (2 sites), ND5 (15 sites) and ND6 genes (4 sites) when comparing species-specific lineages (Table S14). The observed amino acid changes can affect physicochemical properties such as alpha-helical tendencies, equilibrium constant (ionization of COOH), polar requirement, surrounding hydrophobicity, power to be at the middle of alpha-helix, isoelectric point, and solvent accessible reduction ratio. The ancestral amino acid states are equally represented in dromedary, domestic and wild Bactrian camels (Table S14).

The detection of positive selection by comparing physicochemical changes of amino acids is very sensitive, as in highly conserved genes even single amino acid changes can be adaptive if they cause biochemically advantageous characterization. However, a single base change seems too weak to be identified as a positively selected site with the dN/dS ratio approach^{45,46,74,75}. As a reaction to this potential weakness in dN/dS , we applied a physicochemical property model in TreeSAAP and characterized positive destabilizing selection that correlated with seven biochemical functional shifts in local regions of the proteins. McClellan *et al.* (2005) argued that when this radical amino acid changes are favored by selection, they potentially result in local directional shifts in biochemical function, structure, or both, representing the unambiguous signature of molecular adaptation⁴⁶. The observed changes in the amino acid properties can affect protein functions in different ways. First, we identified an increase or decrease in the alpha-helical tendency (P_{α}) in the ND5 gene caused by five or three amino acid changes, respectively. Increasing the alpha-helical tendencies results in a longer, more rigid alpha helix, which makes interactions with other amino acid motifs more difficult. On the contrary, decreasing this property allows for a more flexible and open alpha helix, which increases the accessibility at protein interfaces⁷⁶. Second, we detected an increase (eight amino acid changes) or decrease (three changes) in the equilibrium constant for the ionization of COOH (pK') in ND6 and ND5, respectively, which make the region more or less water-soluble and hydrophilic. An increase in hydrophilicity is very important regarding a reduced reactive oxygen species (ROS) production and increased longevity during calorie restriction⁷⁷. An increase in surrounding hydrophobicity (H_p) property in ND3 gene, indicated that the regions surrounding these amino acid sites become less hydrophilic and more hydrophobic by introducing an isoleucine instead of a threonine.

Long-term divergence between domestic and wild Bactrian camels. Until recently, the critically endangered wild two-humped camels have been either considered to be feral (domestic runaways; reviewed in Peters & van den Driesch⁷⁸), or classified as a subspecies of the domestic Bactrian camel, *Camelus bactrianus*

*ferus*¹². However, earlier studies on partial mitochondrial genes indicated a high amount of differentiation (1.8–2%; refs 13, 14 and 79) between domestic and wild two-humped camels, and the International Commission of Nomenclature⁸⁰ designated the species status “*Camelus ferus*” to the wild camel⁸¹. Genome-wide comparison between the two species confirmed their differentiation¹², however, no divergence time was investigated. Here, we estimated the split between domestic Bactrian camels and their wild relatives to 1.1 [0.58–1.8] mya using all sites, and to 2.3 [0.71–4.1] mya using first and second protein-coding positions only (Table 3); long before domestication started around 4000–6000 ya⁸². With no mitochondrial haplotype sharing and a high number of fixed nucleotide polymorphisms ($n = 233$) between them, we conclude that wild camels are not only a separate species but also not the direct ancestors of contemporary domestic Bactrian camels. Our modern livestock consequently was domesticated from different ancestral populations, which had diverged from their wild sister species around 1.1–2.3 mya, and became extinct after founding the modern domestic Bactrian camel gene pool. A similar evolutionary history can be observed in the Przewalski and the domestic horse^{83,84}. Finally, our divergence time estimates between Camelidae and Bovidae 41.4 [24.7–60.6] mya as well as between Camelini and Lamini 15.8 [9.2–23.2] mya were in line with the latest genome-wide comparisons between the different species (42.7 [33.2–51.0] and 16.3 [9.4–25.3], respectively)⁹. Considering that most wild relatives of our domesticates (including llama and alpaca) are still alive, our results emphasize the evolutionary importance of *Camelus ferus* as the last wild representatives of the Old World camelids.

Methods

Sampling and sequencing of complete mtDNA. Genomic DNA was extracted from whole blood of 24 specimens of Old World camelids (Table S1 and Fig. 1), using Master Pure™ DNA purification kit for blood (Epicentre version III). This sampling was designed to span the known phylogenetic tree of Old World camelids, and to include mtDNA of species that are well adapted to hot, arid (nine dromedaries) and cold, dry (nine wild and six domestic Bactrian camels) environments. Blood samples were retrieved commensally either during routine veterinary treatment or during a radio-collaring project for Mongolian wild camels. All data sets were collected within the frames of the legal requirements of Austria and Mongolia. Capture and collaring of wild camels was conducted within a cooperation agreement between the International Takhi Group and the Mongolian Ministry of Nature, Environment and Tourism (signed on 15.02.2001 and renewed on 27.01.2011). The 500 bp paired-end Illumina library preparation and sequencing of each sample on a single lane of Illumina HiSeq 2000 platform was performed within the framework of a whole-genome re-sequencing study of Old World camelids (Fitak *et al.* in preparation). Three reference mitogenomes were retrieved from GenBank (Accession numbers NC_009849.1, NC_009628.2, NC_009629.2).

Read processing and alignments. The raw sequence reads for each individual were trimmed at the 3' end to a minimum phred-scaled base quality of 20 and minimum length of 50 bp (base pairs) using POPOOLATION v1.2.2⁸⁵. For each species, the trimmed reads were mapped against the corresponding mitochondrial reference genome (*C. dromedarius* Genbank Accession no: NC_009849.1, *C. bactrianus*: NC_009628.2, and *C. ferus*: NC_009629.2), using BWA v0.7⁸⁶ with the following parameters (-l 200 -n 0.01 -o 1 -e 12 -d 12 -t 6). Duplicate reads were removed using Picard-tools ‘MarkDuplicates’ v.1.89 (<http://www.picard.sourceforge.net>) and only the filtered reads, which were properly paired and unambiguously mapped with mapping quality score >20 were used for further analyses. The consensus sequence for each sample was generated using CLC Genomics Workbench 6.5.1 (<http://www.clcbio.com>). The consensus mitogenomes were aligned with CodonCode Aligner v.3.7.1 (Codon Code Corporation, USA). The mitochondrial sequences of all 13 protein-coding genes, two rRNA subunits (12S, 16S) and control region of the three camel species were identified using the corresponding reference sequences of dromedary, domestic and wild Bactrian camel. The overlapping fragments of all sequences were examined to ensure complete sequence coverage. To avoid nuclear insertions of mitochondrial sequences (Numts), all amino acid sequences of protein-coding genes were screened for frameshift mutations and/or early stop codons. As additional quality control we visually screened the sequence reads for misalignments and/or large gaps using Integrative Genomics Viewer v.2.3⁸⁷.

Mitochondrial gene diversity and differentiation among Camelus. Summary statistics of the 27 camel mitogenomes, within and between species were calculated with DnaSP v5.10.1⁸⁸. Within species diversity parameters included the number of polymorphic (segregating) sites (S), number of haplotypes (h), haplotype (H_d) and nucleotide (π) diversity, average number of nucleotide differences (k) and Watterson’s theta (θ_s) based on the number of segregating sites (Table 1). The software was also used to carry out neutrality tests including Tajima’s D ⁴⁷, Fu and Li’s F ⁴⁸, Fu’s F_s ⁸⁹, and Ramos-Onsins and Rozas R_2 statistic⁴⁹. Both D and F can detect historical population expansions and contractions, whereas F_s and R_2 explicitly test past population expansion and are more powerful⁴⁹. When testing for past expansions, R_2 is preferred for small sample sizes ($n \leq 20$) and F_s for larger sample sizes⁴⁹. Significance and 95% confidence intervals of the F_s and R_2 statistics were calculated with 1000 coalescent simulations. Pairwise differentiation comparisons between species comprised the average number of nucleotide differences between species (n), number of fixed synonymous (D_s) and nonsynonymous substitutions (D_n) between species, and a sliding window analysis (window size 100 bp, step size 25 bp) of the number of net nucleotide substitutions per site (D_w) between species-specific lineages. The summary statistics of differentiation in pairwise comparisons between species were calculated for the protein-coding genes, rRNAs, tRNAs and the control region separately, with the overall differentiation as a sum of individual gene divergence.

Mitochondrial genome evolution and screening for natural selection. To address our main question, whether mitogenomes in the genus *Camelus* evolve under a model of neutrality or natural selection, we employed four different approaches (*i–iv*): first, we performed a global MK test³⁷ on the concatenated

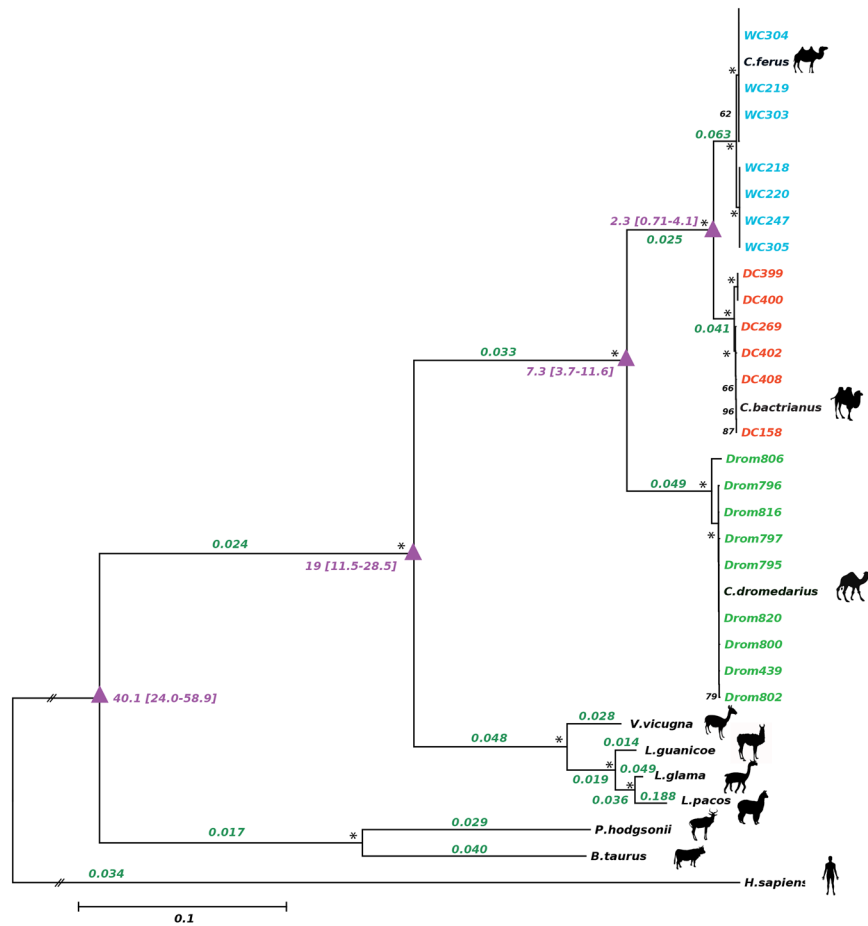


Figure 2. The Maximum likelihood phylogenetic tree inferred from 11,379 bp concatenated protein-coding genes from 10 different species. The HKY substitution model with a gamma distribution and proportion of invariant sites (+GI) was used to model evolutionary rate differences. Lineage-specific dN/dS values for concatenated protein-coding genes are indicated in green over each branch. The estimated divergence time (mya; using first and second protein-coding positions) of each evolutionary lineage is shown in purple color, with confidence intervals in brackets. The statistical support for each node is based on 100 bootstrap replicates, with values (in black) on nodes indicating bootstrap supports above 50%; an asterisk indicates a value of 100%. Branch lengths are scaled as the number of base pair changes per site, with the exception of the human outgroup sequence which is rescaled for aesthetic purposes. The figures depicted on the phylogenetic tree are adopted from <https://openclipart.org>, and are licensed under Creative Commons Zero 1.0 License.

protein-coding genes (11,379 bp), using DnaSP v5.10.1⁸⁸. We calculated the neutrality index ($NI = (P_n/P_s)/(D_n/D_s)$; polymorphic sites/ fixed sites among species) at non-synonymous (dN) and synonymous (dS) sites in pairwise comparison between the three species. The significance of the NI values was assessed with the two-tailed Fisher's exact test. Next, to explicitly examine which *Camelus* lineage had either a deficiency or excess of amino acid polymorphisms and as such might be affected by natural selection, we performed an extended MK test⁵⁰ on the concatenated protein-coding genes, and each of 13 protein-coding genes separately. In the extended MK test we calculated polymorphisms within one species and the divergence (fixed SNPs) at nonsynonymous and synonymous sites in pairwise comparison.

In our second approach, we examined whether mitochondrial protein-coding genes in the *Camelus* lineage show any specific patterns of evolutionary rates. We assessed the variation in the ratio of dN to dS among the branches of a phylogenetic maximum likelihood (ML) tree using MEGA6⁹⁰, with the best-fitting model HKY + G + I according to the Bayesian Information Criterion (BIC). Using the ML method implemented in the CODEML program of PAML v.4.8⁵¹, we applied the background and the branch-specific model on the 11,379 bp concatenated alignment (including gaps) of the mitochondrial protein-coding genes from ten different species (34 individuals). We included Old World (*C. dromedarius*, *C. bactrianus*, *C. ferus*) and New World camels (*Lama glama*, *L. guanicoe*, *Vicugna pacos*, *V. vicugna*), *Bos taurus*, *Pantholops hodgsonii* and *Homo sapiens*; all accession numbers are given in Table S1. While in the background model a single value of the dN/dS ratio (ω) is considered among all sites and branches, the branch-specific model allows a varying ω among the branches in the phylogeny⁹¹. We used a LRT to determine whether or not the branch-specific model was a significant improvement over the background model.

Third, we tested for selection in single sites of the protein-coding genes within the genus *Camelus* using the branch-site model A⁴⁰ in CODEML. In this model, the ω value is allowed to vary among different branches as well as in the nucleotide sites of a particular branch (foreground lineage). We tested a null model (H0), in which the foreground branch (*Camelus sp.*) may have different proportions of sites under neutral selection than the background (i.e., relaxed purifying selection), and an alternative model (H1), in which the foreground branch may have a proportion of sites under positive selection. If in the alternative model a site was identified as potentially fixed due to positive selection ($dN/dS > 1$), its posterior probability was estimated with the Bayes empirical Bayes (BEB) approach⁵¹. For each model we retrieved the log likelihood values from which we computed the likelihood ratio tests (LRT) for the significance (*P*-value) of the sites detected under positive selection.

Finally, we investigated the effect of amino acid changes on the protein function, using TreeSAAP⁵². This program uses statistical models of molecular evolution and phylogenetic trees to predict positive selection based on amino acid physicochemical properties. TreeSAAP takes into account the magnitude of the impact of the amino acid replacements on local physicochemical properties and determines whether the observed degree of amino acid changes deviates from the neutral expectation. Radical magnitudes of changes ≥ 6 , with $P \leq 0.001$, are considered as an evidence for positive directional selection for a given physicochemical property⁹². To strengthen the accuracy and avoid false positive results we only used amino acid properties with the accuracy of detecting selection higher than 85% (i.e., 11 out of 31 sites were excluded) as recommended in⁹². In case of detecting fixed amino acid changes with effects on the physicochemical properties of a protein, we determined the ancestral state of the amino acid preceding the split of one- and two-humped camels, and domestic and wild Bactrian camels with the ML analysis in MEGA6⁹⁰. We constructed the phylogenetic tree for each protein coding gene in eight different species (Camelini, Lamini, and *B.taurus*) in MEGA as described above and used these trees as input into TreeSAAP. We restricted the sites analyzed to only the fixed nonsynonymous substitutions in pairwise comparisons in the three camel species. This analysis allowed us to investigate whether the domestic species have more amino acid changes with a predicted negative or positive impact on the respective protein due to artificial selection during domestication.

Divergence time estimation. We estimated the divergence times among *Camelidae* with MULTIDIVTIME⁵³, which calculates posterior credibility intervals (CIs) of divergence times using a Bayesian approach and assuming a relaxed molecular clock with autocorrelation among rates. This method has been shown to produce posterior CIs that contain the true divergence time of simulated datasets $\geq 95\%$ of the time as long as the assumption of autocorrelation among rates was not violated⁹³. We followed the protocol outlined by Rutschmann⁹⁴ for the 13 protein-coding DNA sequences. The analysis can be summarized as follows: for each gene's alignment we determined the best-fit nucleotide substitution model and generated a ML tree topology in MEGA6 according to the procedure above. We calculated model parameters, including nucleotide frequencies, the transition/transversion rate ratio, and rate heterogeneity among sites for each gene using 'baseml' in PAML v4.8⁵¹. Next, branch lengths and a variance-covariance matrix were estimated using the 'estbranches' function in MULTIDIVTIME against the ML tree topology for the 10-species dataset described above. We then ran five independent chains of MULTIDIVTIME to estimate divergence times, with each chain containing 10^5 burn-in steps followed by 10^6 steps sampled every 100th step (10^4 samples in total). We parameterized the MULTIDIVTIME runs using $r_{tm} = 6.5$, $r_{rate} = 0.07$, and $brownmean = 0.3$ and with equivalent standard deviations, respectively. The time between the ingroup root and median of the tips (r_{tm}) corresponded to 1 unit ≈ 10 million years, and was retrieved from the TimeTree database⁹⁵ as the mean time of the most recent common ancestor between the families Bovidae and Camelidae (65 mya). The rate (r_{rate}) was calculated as r_{tm} divided by the median branch length between the ingroup root and tips. We set $brownmean$ to approximately $2/r_{tm}$. We included a prior to constrain the node connecting the Camelini and Lamini tribes to 8–30 mya based upon both fossil^{96, 97}, and molecular^{98, 99} evidence.

Prior to estimating posterior probabilities, we assessed convergence of the independent MULTIDIVTIME chains using the R package 'boa' v1.1.8¹⁰⁰ to ensure the runs reached a stationary distribution. We examined each chain and its parameters (divergence time and rate for each node) independently for convergence using the corrected scale reduction factors (CSRF) from^{54, 55}. A chain is suggested to have converged when the CSRF is 1, or more specifically, when the 97.5% upper confidence limit of the estimate is ≤ 1.2 ⁵⁴. Furthermore, we also examined for convergence across all parameters and chains simultaneously using the multivariate potential scale reduction factor (MPSRF)^{54, 55}. The MPSRF converges to 1 upon reaching stationarity when the number of steps is reasonably large. After convergence was verified, all chains were combined and the posterior density mean and 95% CIs were estimated for the parameters. We repeated the above analysis in MULTIDIVTIME using only the first and second codon positions with a modified $r_{rate} = 0.012$.

Because mtDNA evolves rapidly, it is often subject to the effects of saturation (multiple mutations at a single site) – especially at deep divergence times^{101, 102}. As a result, divergence estimates need to take into account the possible effects of saturation. To determine the potential for saturation effects in our dataset, we used the DNA saturation analysis suggested in ref. 101. Briefly, at low levels of divergence the ratio of transitions to transversions (Ti/Tv) in mtDNA is relatively high. However, as divergence increases, the chance for multiple transversions at a site increases, thus the Ti/Tv ratio approaches 1:2 when base frequencies are equal and phylogenetic information is eventually lost. Therefore, the simultaneous comparison of divergence and Ti/Tv ratio (as a proxy for saturation) serve as an indicator for phylogenetic information¹⁰¹. We calculated the proportion of transitions (%Ti)¹⁰¹ to avoid instances where no transversions occurred and compared it with pairwise sequence divergence to assess the potential for saturation. Regions of low %Ti ($\leq 50\%$)¹⁰¹ were defined as likely affected by saturation. We performed the DNA saturation analysis separately for all sites in the protein-coding sequences and also for the first and second codon positions.

Availability of data. The complete camel mitochondrial genomes obtained in this study are deposited in Genbank with accession numbers listed below: Drom439 (Qatar): KU605072, Drom795 (Saudi Arabia): KU605073, Drom796 (Saudi Arabia): KU605074, Drom797 (Saudi Arabia): KU605075, Drom801A (Austria): KU605076, Drom802 (UAE, Dubai): KU605077, Drom806 (Kenya): KU605078, Drom816 (Sudan): KU605079, Drom820 (Pakistan): KU605080, DC158 (Austria): KU666460, DC269 (Kazakhstan): KU666461, DC399 (Mongolia): KU666462, DC400 (Mongolia): KU666463, DC402 (Mongolia): KU666464, DC408 (Mongolia): KU666465, WC214 (Mongolia): KU666451, WC216 (Mongolia): KU666452, WC218 (Mongolia): KU666453, WC219 (Mongolia): KU666454, WC220 (Mongolia): KU666455, WC247 (Mongolia): KU666456, WC303 (Mongolia): KU666457, WC304 (Mongolia): KU666458, WC305 (Mongolia): KU666459.

References

- Bourzat, D. & Wilson, R. T. Research on the dromedary in Africa. *Scientific and Technical Review*. **6**, 383–389 (1987).
- Wardeh, M. F. Classification of the Dromedary Camels. *Journal Camel Science*. **1**, 1–7 (2004).
- Faye, B. & Konuspaveva, G. The Encounter between Bactrian and Dromedary Camels in Central Asia. In: Knoll, E.-M. and Burger, P. editors. *Camels in Asia and North Africa*. Vienna: Interdisciplinary perspectives on their significance in past and present. Austrian Academy of Sciences Press, Wien (Austria). p. 27–33 (photos p. 248–250) (2012).
- Grubb, P. Order Artiodactyla. In: Wilson, D. E., Reeder, D. M. editors. *Mammal species of the world: a taxonomic and geographic reference*. Baltimore: Johns Hopkins University Press. p. 637–722 (2005).
- Potts, D. Camel hybridization and the role of *Camelus bactrianus* in the Ancient Near East. *Journal of the Economic and Social History of the Orient*. **47**, 143–165 (2004).
- Reading, R. P., Mix, H., Lhagvasuren, B. & Blumer, E. S. Status of wild Bactrian camels and other large ungulates in south-western Mongolia. *Oryx*. **33**, 247–255 (1999).
- Yadamsuren, A., Dulamtseren, E. & Reading, R. P. The Conservation Status and Management of Wild Camels in Mongolia. In: Knoll, E. M., Burger, P. A. editors. *Camels in Asia and North Africa*. Vienna: Interdisciplinary perspectives on their significance in past and present. Austrian Academy of Sciences Press, Wien (Austria). p. 45–54 (2012).
- Lei, Y., Hare, J., Guoying, Y. & Yun, C. The Status of the Wild Camel in China. In: E. M., Knoll and P. A., Burger (Eds) *Camels in Asia and North Africa*. Interdisciplinary perspectives on their significance in past and present. Austrian Academy of Sciences, Vienna, p.55–60 (2012).
- Wu, H. *et al.* Camelid genomes reveal evolution and adaptation to desert environments. *Nat Commun*. **5**, 5188, doi:10.1038/ncomms6188 (2014).
- Kadwell, M. *et al.* Genetic analysis reveals the wild ancestors of the llama and the alpaca. *Proc Biol Sci*. **268**, 2575–84 (2001).
- Burger, P. A. The history of Old World camelids in the light of molecular genetics. *Trop Anim Health Prod*. **48**, 905–13, doi:10.1007/s11250-016-1032-7 (2016).
- Jirimutu. *et al.* Genome sequences of wild and domestic bactrian camels. *Nat. Commun*. **3**, 1202, doi:10.1038/ncomms2192 (2012).
- Silbermayr, K. *et al.* Isolation and characterization of nine new microsatellite loci in the domestic Bactrian camel (*Camelus bactrianus*) and amplification in the wild Bactrian camel (*C. ferus*). *Mol Ecol Res*. **10**, 1106–1108, doi:10.1111/j.1755-0998.2010.02916.x (2010).
- Ji, R. *et al.* Monophyletic origin of domestic bactrian camel (*Camelus bactrianus*) and its evolutionary relationship with the extant wild camel (*Camelus bactrianus ferus*). *Anim. Gen.* **40**, 377–82, doi:10.1111/j.1365-2052.2008.01848.x (2009).
- Almathen, F. *et al.* Ancient and modern DNA reveal dynamics of domestication and cross-continental dispersal of the dromedary. *Proc Natl Acad Sci USA* **113**, 6707–6712, doi:10.1073/pnas.1519508113 (2016).
- Avise, J. C. *Molecular Markers, Natural History, and Evolution*, 2nd edn. Sinauer Associates, Sunderland, MA (2004).
- Meiklejohn, C. D., Montooth, K. L. & Rand, D. M. Positive and negative selection on the mitochondrial genome. *Trends Genet.* **23**, 259–263 (2007).
- Ballard, J. W. O. & Rand, D. M. The population biology of mitochondrial DNA and its phylogenetic implications. *Annu Rev Ecol Evol Syst* **36**, 621–642 (2005).
- Ballard, J. W. O. & Whitlock, M. C. The incomplete natural history of mitochondria. *Mol. Ecol.* **13**, 729–744 (2004).
- Blier, P. U., Dufresne, F. & Burton, R. S. Natural selection and the evolution of mtDNA-encoded peptides: evidence for intergenomic co-adaptation. *Trends Genet.* **17**, 400–406 (2001).
- Ballard, J. W. O. & Kreitman, M. Is mitochondrial-DNA a strictly neutral marker? *Trends Ecol Evol*. **10**, 485–488 (1995).
- Kerr, K. C. Searching for evidence of selection in avian DNA barcodes. *Mol Ecol Resour.* **11**, 1045–1055, doi:10.1111/j.1755-0998.2011.03049.x (2011).
- Rutledge, L. Y., Patterson, B. R. & White, B. N. Analysis of *Canis* mitochondrial DNA demonstrates high concordance between the control region and ATPase genes. *BMC Evol Biol.* **10**, 215 (2010).
- Bazin, E., Glémin, S. & Galtier, N. Population size does not influence mitochondrial genetic diversity in animals. *Science*. **312**, 570–572 (2006).
- Ruiz-Pesini, E., Mishmar, D., Brandon, M., Procaccio, V. & Wallace, D. C. Effects of purifying and adaptive selection on regional variation in human mtDNA. *Science*. **303**, 223–226 (2004).
- Castoe, T. A., Jiang, Z. J., Gu, W., Wang, Z. O. & Pollock, D. D. Adaptive evolution and functional redesign of core metabolic proteins in snakes. *PLoS ONE*. **3**, e2201, doi:10.1371/journal.pone.0002201 (2008).
- Garvin, M. R., Bielawski, J. P. & Gharrett, A. J. Positive Darwinian selection in the piston that powers proton pumps in complex I of the mitochondria of Pacific salmon. *PLoS One*. **6**, e24127, doi:10.1371/annotation/8a7e2019-c039-4a54-b797-e9adb2ce4efc (2011).
- Foote, A. D. *et al.* Positive selection on the killer whale mitogenome. *Biol Lett.* **7**, 116–118, doi:10.1098/rsbl.2010.0638 (2011).
- da Fonseca, R. R., Johnson, W. E., O'Brien, S. J., Ramos, M. J. & Antunes, A. The adaptive evolution of the mammalian mitochondrial genome. *BMC Genomics*. **9**, 119, doi:10.1186/1471-2164-9-119 (2008).
- Mishmar, D. *et al.* Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci USA* **100**, 171–176 (2003).
- Scott, G. R. *et al.* Molecular evolution of cytochrome C oxidase underlies high-altitude adaptation in the bar-headed goose. *Mol Biol Evol.* **28**, 351–63, doi:10.1093/molbev/msq205 (2011).
- 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.* **68**, 293–310 (2009).
- Luo, Y. *et al.* Mitochondrial genome analysis of *Ochotona curzoniae* and implication of cytochrome c oxidase in hypoxic adaptation. *Mitochondrion*. **8**, 352–357, doi:10.1016/j.mito.2008.07.005 (2008).
- Melo-Ferreira, J. *et al.* The elusive nature of adaptive mitochondrial DNA evolution of an arctic lineage prone to frequent introgression. *Genome Biol Evol.* **6**, 886–96, doi:10.1093/gbe/evu059 (2014).
- Silva, G., Lima, F. P., Martel, P. & Castilho, R. Thermal adaptation and clinal mitochondrial DNA variation of European anchovy. *Proc Biol Sci.* **281**, 1792, doi:10.1098/rspb.2014.1093 (2014).
- Fontanillas, P., Depraz, A., Giorgi, M. S. & Perrin, N. Nonshivering thermogenesis capacity associated to mitochondrial DNA haplotypes and gender in the greater white-toothed shrew, *Crocidura russula*. *Mol Ecol.* **14**, 661–670 (2005).
- McDonald, J. H. & Kreitman, M. Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature*. **351**, 652–654 (1991).

38. Biswas, S. & Akey, J. M. Genomic insights into positive selection. *Trends Genet.* **22**, 437–446 (2006).
39. Crandall, K. A., Kelsey, C. R., Imamichi, H., Lane, H. C. & Salzman, N. P. Parallel evolution of drug resistance in HIV: failure of nonsynonymous/synonymous substitution rate ratio to detect selection. *Mol Biol Evol.* **16**, 372–382 (1999).
40. 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.* **22**, 2472–2479 (2005).
41. Toll-Riera, M., Laurie, S. & Alba, M. M. Lineage-specific variation in intensity of natural selection in mammals. *Mol Biol Evol.* **28**, 383–398 (2011).
42. McClellan, D. A. & McCracken, K. G. Estimating the influence of selection on the variable amino acid sites of the cytochrome b protein functional domains. *Mol Biol Evol.* **18**, 917–925 (2001).
43. Xia, X. & Li, W. H. What amino acid properties affect protein evolution? *J Mol Evol.* **47**, 557–564 (1998).
44. Hughes, A. L., Ota, T. & Nei, M. Positive darwinian selection promotes charge profile diversity in the antigen-binding cleft of class I major-histocompatibility-complex molecules. *Mol Biol Evol.* **7**, 515–524 (1990).
45. Porter, M. L., Cronin, T. W., McClellan, D. A. & Crandall, K. A. Molecular characterization of crustacean visual pigments and the evolution of pancrasteacean opsins. *Mol Biol Evol.* **24**, 253–268 (2007).
46. McClellan, D. A. *et al.* Physicochemical evolution and molecular adaptation of the cetacean and artiodactyl cytochrome b proteins. *Mol Biol Evol.* **22**, 437–455 (2005).
47. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics.* **123**, 585–595 (1989).
48. Fu, Y. X. & Li, W. H. Statistical tests of neutrality of mutations. *Genetics.* **133**, 693–709 (1993).
49. Ramos-Onsins, S. E. & Rozas, J. Statistical properties of new neutrality tests against population growth. *Mol Biol Evol.* **19**, 2092–2100 (2002).
50. Messer, P. W. & Petrov, D. A. Frequent adaptation and the McDonald-Kreitman test. *Proc Natl Acad Sci USA* **110**, 8615–8620, doi:10.1073/pnas.1220835110 (2013).
51. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
52. Woolley, S., Johnson, J., Smith, M. J., Crandall, K. A. & McClellan, D. A. TreeSAAP: selection on amino acid properties using phylogenetic trees. *Bioinformatics.* **19**, 671–672 (2003).
53. Thorne, J. L. & Kishino, H. Divergence time and evolutionary rate estimation with multilocus data. *Syst Biol.* **51**, 689–702 (2002).
54. Brooks, S. & Gelman, A. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics.* **7**, 434–455 (1998).
55. Gelman, A. & Rubin, D. B. Inference from iterative simulation using multiple sequences. *Statistical Science.* **7**, 457–511 (1992).
56. Bruford, M. W., Bradley, D. G. & Luikart, G. DNA markers reveal the complexity of livestock domestication. *Nat Rev Genet.* **4**, 900–910 (2003).
57. Charruau, P. Insights from evolutionary history and population genetics for domestic and wildlife conservation – cases of the Old World camelids and cheetahs. Thesis, University of Veterinary Medicine, Vienna (Austria) (2012).
58. Matosiuk, M., Sheremetyeva, I. N., Sheremetyev, I. S., Saveljev, A. P. & Borkowska, A. Evolutionary neutrality of mtDNA introgression: evidence from complete mitogenome analysis in roe deer. *J Evol Biol.* **27**, 2483–94 (2014).
59. Yang, L. *et al.* Species identification through mitochondrial rRNA genetic analysis. *Sci Rep.* **4**, 4089, doi:10.1038/srep04089 (2014).
60. Abhyankar, A., Park, H. B., Tonolo, G. & Luthman, H. Comparative sequence analysis of the non-protein-coding mitochondrial DNA of inbred rat strains. *PLoS One.* **4**, e8148, doi:10.1371/journal.pone.0008148 (2009).
61. Hellberg, M. E. No variation and low synonymous substitution rates in coral mtDNA despite high nuclear variation. *BMC Evol Biol.* **6**, 24 (2006).
62. Vilmi, T., Moilanen, J. S., Finnilä, S. & Majamaa, K. Sequence variation in the tRNA genes of human mitochondrial DNA. *J Mol Evol.* **60**, 587–597 (2005).
63. Ballard, J. W. Comparative genomics of mitochondrial DNA in members of the *Drosophila melanogaster* subgroup. *J Mol Evol.* **51**, 48–63 (2000).
64. Zhang, F. & Broughton, R. E. Mitochondrial-nuclear interactions: compensatory evolution or variable functional constraint among vertebrate oxidative phosphorylation genes? *Genome Biol Evol.* **5**, 1781–91, doi:10.1093/gbe/evt129 (2013).
65. Blier, P. U., Breton, S., Desrosiers, V. & Lemieux, H. Functional conservatism in mitochondrial evolution: insight from hybridization of arctic and brook charrs. *J Exp Zool B Mol Dev Evol.* **306**, 425–432 (2006).
66. Montooth, K. L., Abt, D. N., Hofmann, J. W. & Rand, D. M. Comparative genomics of *Drosophila* mtDNA: Novel features of conservation and change across functional domains and lineages. *J Mol Evol.* **69**, 94–114, doi:10.1007/s00239-009-9255-0 (2009).
67. Brandt, U. Energy converting NADH: quinone oxidoreductase (complex I). *Annu Rev Biochem.* **75**, 69–92 (2006).
68. Colli *et al.* Whole mitochondrial genomes unveil the impact of domestication on goat matrilineal variability. *BMC Genomics.* **16**, 1115, doi:10.1186/s12864-015-2342-2 (2015).
69. Wang *et al.* Domestication relaxed selective constraints on the yak mitochondrial genome. *Mol Biol Evol.* **28**, 15553–15556, doi:10.1093/molbev/msq336 (2010).
70. Furlan, E. *et al.* Small population size and extremely low levels of genetic diversity in island populations of the platypus, *Ornithorhynchus anatinus*. *Ecol Evol.* **2**, 844–57, doi:10.1002/ece3.195 (2012).
71. Frankham, R., Ballou, J. D. & Briscoe, D. A. Introduction to conservation genetics. 2nd ed. Cambridge, U. K: Press Syndicate for the Univ. of Cambridge (2010).
72. Willi, Y., Van, Buskirk, J. & Hoffmann, A. A. Limits to the adaptive potential of small populations. *Ann Rev Ecol Evol Syst.* **37**, 433–458 (2006).
73. Wright, S. Classification of the factors of evolution. Cold Spring Harbor Symposia on Quantitative Biology. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press. **20**, 16–24 (1955).
74. McClellan, D. A. Directional Darwinian Selection in proteins. *BMC Bioinformatics.* **14**, 1–8 (2013).
75. Hughes, A. L. Looking for Darwin in all the wrong places: the misguided quest for positive selection at the nucleotide sequence level. *Heredity.* **99**, 364–73 (2007).
76. Burkin, D. J., Kim, J. E., Gu, M. & Kaufman, S. J. Laminin and alpha7beta1 integrin regulate agrin-induced clustering of acetylcholine receptors. *J Cell Sci.* **113**, 2877–86 (2000).
77. Beckstead, W. A., Ebbert, M. T., Rowe, M. J. & McClellan, D. A. Evolutionary pressure on mitochondrial cytochrome b is consistent with a role of Cytb17T affecting longevity during caloric restriction. *PLoS One.* **4**, e5836, doi:10.1371/journal.pone.0005836 (2009).
78. Peters, J. von den Driesch. The two-humped camel (*Camelus bactrianus*): New light on its distribution, management and medical treatment in the past. *J Zool.* **242**, 651–679 (1997).
79. Jianlin, H., Jiexia, Q., Zhenming, M., Yaping, Z. & Wen, W. Rapid communication: three unique restriction fragment length polymorphisms of EcoRI, PvuII, and ScaI digested mitochondrial DNA of Bactrian camels (*Camelus bactrianus fesus*) in China. *J. Anim. Sci.* **77**, 2315–6 (1999).
80. International Commission on Zoological Nomenclature, 2003. Opinion 2027 (Case 3010): Usage of 17 specific names based on wild species which are pre-dated by or contemporary with those based on domestic animals (Lepidoptera, Osteichthyes, Mammalia): conserved. <http://iczn.org/content/biodiversity-studies>. Accessed on 8 Feb 2016.
81. Gentry, A., Clutton-Brock, J. & Groves, C. P. The naming of wild animal species and their domestic derivatives. *J Archeol Sci.* **31**, 645–651 (2004).

82. Han, J., Quan, J. X., Han, J. L. & Men, Z. M. Genetic differentiation between *Camelus bactrianus ferus* and *C. bactrianus* inferred from mitochondrial DNA RFLPS. In: Reading RP, Enkhbileg, D., Galbaatar, T. editors. Ecology and Conservation of Wild Bactrian Camels (*Camelus bactrianus ferus*) Ulaanbaatar, Mongolia: Mongolian Conservation Coalition & Admon Printing. p. 65–70 (2002).
83. Orlando, L. *et al.* Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature*. **499**, 74–48, doi:10.1038/nature12323 (2013).
84. Goto, H., Ryder, O. A. & Fisher, A. R. A massively parallel sequencing approach uncovers ancient origins and high genetic variability of endangered Przewalski's horses. *Genome Biol Evol*. **3**, 1096–106, doi:10.1093/gbe/evr067 (2011).
85. Kofler, R. *et al.* PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS One*. **6**, e15925, doi:10.1371/journal.pone.0015925 (2011).
86. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. **25**, 1754–60 (2009).
87. Robinson, J. T. *et al.* "Integrative Genomics Viewer". *Nature Biotechnology*. **29**, 24–26 (2011).
88. Librado, P. & Rozas, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*. **25**, 1451–1452 (2009).
89. Fu, Y. X. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*. **147**, 915–25 (1997).
90. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729 (2013).
91. Yang, Z. & Nielsen, R. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *J Mol Evol*. **46**, 409–418 (1998).
92. McClellan, D. A. & Ellison, D. D. Assessing and improving the accuracy of detecting protein adaptation with the TreeSAAP analytical software. *Int J Bioinform Res Appl*. **6**, 120–133 (2010).
93. Battistuzzi, F. U., Filipski, A., Hedges, B. A. & Kumar, S. Performance of relaxed-clock methods in estimating evolutionary divergence times and their credibility intervals. *Mol Biol Evol*. **27**, 1289–1300 (2010).
94. Rutschmann, F. Bayesian molecular dating using PAML/multidivtime. A step-by-step manual. University of Zurich, Switzerland. Available at www.plant.ch/software.html (2004).
95. Hedges, S. B., Dudley, J. & Kumar, S. TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics*. **22**, 2971–2972 (2006).
96. Harrison, J. A. Revision of the Camelinae (Arteriodactyla, Tylopoda) and description of the new genus *Alforjas*. *Paleont. Cont.* **95**, 1–20 (1979).
97. Webb, S. D. Pleistocene llamas of Florida, with a brief review of the Lamini. In Pleistocene mammals of Florida (ed. Webb, S. D.). University Press of Florida, Gainesville. p. 170–214 (1974).
98. Cui, P. *et al.* A complete mitochondrial genome sequence of the wild two-humped camel (*Camelus bactrianus ferus*): an evolutionary history of camelidae. *BMC Genomics*. **8**, 241 (2007).
99. Stanley, H. F., Kadwell, M., Wheeler & Jane, C. Molecular evolution of the family Camelidae: a mitochondrial DNA study. *Proc R Soc Lond B*. **256**, 1–6 (1994).
100. Smith, B. J. boa: an R package for MCMC output convergence assessment and posterior inference. *Journal of Statistical Software*. **21**, 1–37 (2007).
101. Roe, A. D. & Sperling, F. A. H. Patterns of evolution of mitochondrial cytochrome c oxidase I and II DNA and implications for DNA barcoding. *Molecular Phylogenetics and Evolution*. **44**, 325–345, doi:10.1016/j.ympev.2006.12.005 (2007).
102. Zheng, U., Peng, R., Kuro-o, M. & Zeng, X. Exploring Patterns and Extent of Bias in Estimating Divergence Time from Mitochondrial DNA Sequence Data in a Particular Lineage: A Case Study of Salamanders (Order Caudata). *Molecular Biology and Evolution*. **28**, 2521–2535, doi:10.1093/molbev/msr072 (2003).

Acknowledgements

We are very grateful to the Wild Camel Protection Foundation for the support in sample collection, to W. Asaad and R. Saleh for facilitating sample collection in Syria and Jordan and to G. Gassner, First Austrian Camel Riding School, and H. Burgstaller, Graefliche Kamelheimat, for reference samples from Austria. We acknowledge general support from C. Schlotterer, Institute of Population Genetics, Vetmeduni Vienna, Austria. EM and RRF were supported by the Austrian Science Foundation (FWF) grant P24706-B25 to PAB, recipient of an APART fellowship (11506) of the Austrian Academy of Sciences. JC was partially funded by the COIN Centre of Excellence, Academy of Finland. The authors wish to acknowledge CSC – IT Center for Science, Finland, for computational resources.

Author Contributions

EM wrote the paper and performed laboratory work and bioinformatic analyses. RRF and PAB wrote the paper and performed data analysis. JC provided computational resources and revised the manuscript. AD, BC, OA, AR, PN, GS, CW, BF provided the samples and revised the manuscript.

Additional Information

Supplementary information accompanies this paper at doi:10.1038/s41598-017-08995-8

Competing Interests: The authors declare that they have no competing interests.

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

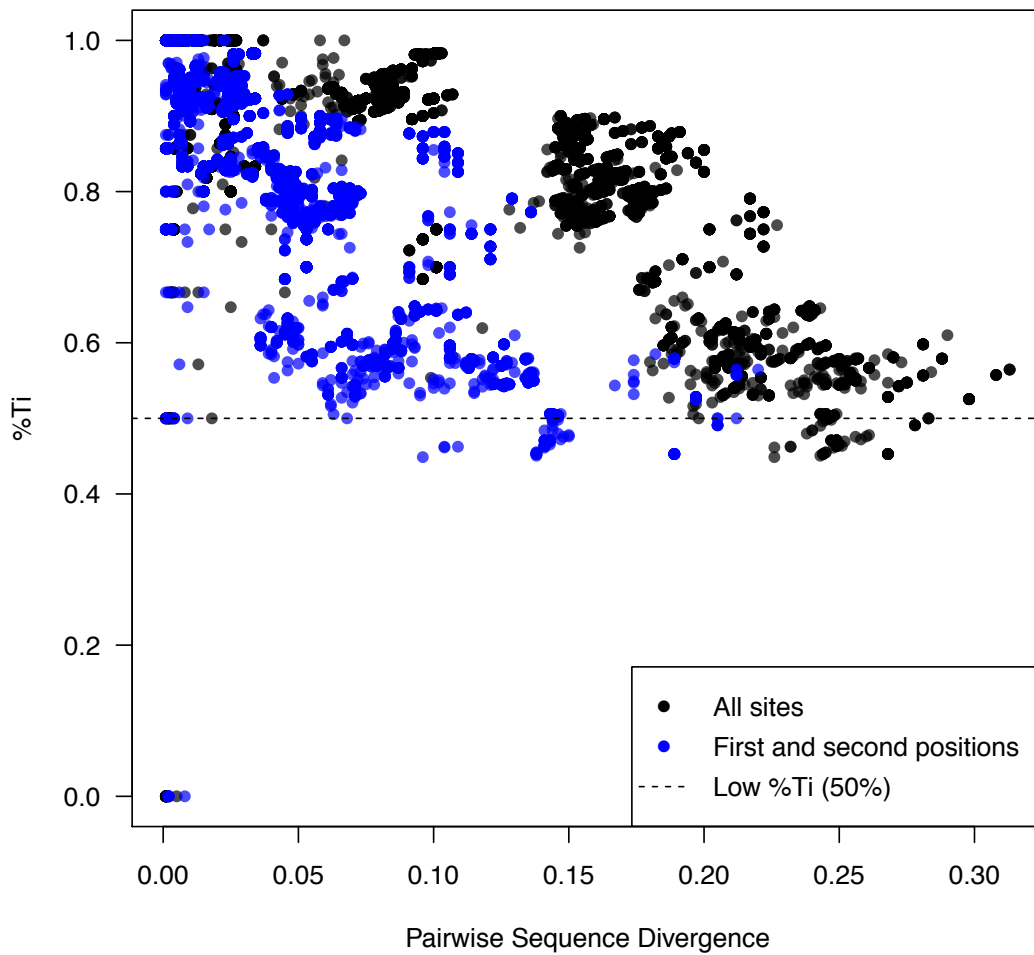


Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative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

© The Author(s) 2017

1 **Supplementary Information**
2
3 **Mitogenome Sequencing in the Genus *Camelus* Reveals Evidence for Purifying**
4 **Selection and Long-term Divergence between Wild and Domestic Bactrian**
5 **Camels**
6
7 **Elmira Mohandesan, Robert R. Fitak, Jukka Corander, Adiya Yadamsuren,**
8 **Battsetseg Chuluunbat, Omer Abdelhadi, Abdul Raziq, Peter Nagy, Gabrielle**
9 **Stalder, Chris Walzer, Bernard Faye, Pamela A. Burger**

10 Figure S1: The mtDNA saturation analysis performed on all sites (black) in the
11 protein-coding sequences, and first and second positions (blue). Regions of low %Ti
12 ($\leq 50\%$) corresponding to sequence divergence of around 0.23 and 0.12 were defined
13 as likely affected by saturation.



14
15
16

17 Table S1: List of the 24 mitochondrial genomes of Old world camelids (*C.*
 18 *dromedarius*, *C.bactrianus*, *C.ferus*) obtained in this study. The geographic location
 19 of samples, and the length of the recovered mtDNA for each sample are mentioned.
 20 The Genbank accession numbers of the mtDNA of 10 different species used in this
 21 study are also listed.
 22

| Sample ID | Species | Geographical location | GenBank ID | Length (bp) |
|-----------------------|-----------------------|-----------------------------|-------------|-------------|
| Dromedary reference | <i>C. dromedarius</i> | United Arab Emirates: Dubai | NC_009849.1 | 16,643 |
| Drom439 ¹ | <i>C. dromedarius</i> | Qatar | KU605072 | 16,379 |
| Drom795 | <i>C. dromedarius</i> | Saudi Arabia | KU605073 | 16,379 |
| Drom796 | <i>C. dromedarius</i> | Saudi Arabia | KU605074 | 16,379 |
| Drom797 | <i>C. dromedarius</i> | Saudi Arabia | KU605075 | 16,379 |
| Drom801A ² | <i>C. dromedarius</i> | Austria | KU605076 | 16,379 |
| Drom802 | <i>C. dromedarius</i> | UAE, Dubai | KU605077 | 16,379 |
| Drom806 | <i>C. dromedarius</i> | Kenya | KU605078 | 16,375 |
| Drom816 | <i>C. dromedarius</i> | Sudan | KU605079 | 16,379 |
| Drom820 | <i>C. dromedarius</i> | Pakistan | KU605080 | 16,379 |
| Bactrian reference | <i>C. bactrianus</i> | Mongolia, China | NC_009628.2 | 16,659 |
| DC158 ³ | <i>C. bactrianus</i> | Austria | KU666460 | 16,385 |
| DC269 | <i>C. bactrianus</i> | Kazakhstan | KU666461 | 16,385 |
| DC399 | <i>C. bactrianus</i> | Mongolia | KU666462 | 16,385 |
| DC400 | <i>C. bactrianus</i> | Mongolia | KU666463 | 16,385 |
| DC402 | <i>C. bactrianus</i> | Mongolia | KU666464 | 16,385 |
| DC408 | <i>C. bactrianus</i> | Mongolia | KU666465 | 16,385 |
| Wild camel reference | <i>C. ferus</i> | Mongolia | NC_009629.2 | 16,680 |
| WC214 | <i>C. ferus</i> | Mongolia | KU666451 | 16,383 |
| WC216 | <i>C. ferus</i> | Mongolia | KU666452 | 16,383 |
| WC218 | <i>C. ferus</i> | Mongolia | KU666453 | 16,384 |
| WC219 | <i>C. ferus</i> | Mongolia | KU666454 | 16,383 |
| WC220 | <i>C. ferus</i> | Mongolia | KU666455 | 16,384 |
| WC247 | <i>C. ferus</i> | Mongolia | KU666456 | 16,384 |
| WC303 | <i>C. ferus</i> | Mongolia | KU666457 | 16,383 |
| WC304 | <i>C. ferus</i> | Mongolia | KU666458 | 16,383 |
| WC305 | <i>C. ferus</i> | Mongolia | KU666459 | 16,384 |
| Llama | <i>L. glama</i> | | AP003426.1 | 16,597 |
| Guanaco | <i>L. guanicoe</i> | | NC_011822.1 | 16,649 |
| Alpaca | <i>L. pacos</i> | | Y19184.1 | 16,652 |
| Vicuna | <i>V. vicugna</i> | | NC_013558.1 | 16,084 |
| Cattle | <i>B. taurus</i> | | AF492351.1 | 16,338 |
| Tibetan Antelope | <i>P. hodgsonii</i> | | NC_007441.1 | 16,498 |
| Tibetan human | <i>H.sapiens</i> | | KF056318.1 | 16,569 |

23

24 ¹ The individual was sampled in Jordan close to the border to Saudi Arabia, however,
 25 its origin from Qatar was confirmed.

26 ² The individual's origin can be traced back to North Africa and the Canary Islands.

27 ³ The individual was sampled in Austria, however, genetically (based on 20
 28 microsatellites) it clustered with Mongolian Bactrian camels.

29 Table S2: Summary statistics of mitochondrial gene differentiation between
 30 dromedaries and domestic Bactrian camels.
 31

| Gene | Length (bp) | n | D _a | D _s | D _n |
|----------|-------------|-------|----------------|----------------|----------------|
| ND1 | 954 | 41.1 | 0.080 | 68 | 3 |
| ND2 | 1041 | 44.3 | 0.078 | 62 | 15 |
| COX1 | 1542 | 44.8 | 0.063 | 96 | 0 |
| COX2 | 681 | 27.5 | 0.077 | 50 | 1 |
| COX3 | 783 | 29.1 | 0.069 | 48 | 2 |
| AP6 | 678 | 28.3 | 0.078 | 44 | 7 |
| AP8 | 201 | 10.1 | 0.093 | 12 | 6 |
| CYTB | 1137 | 56.8 | 0.092 | 93 | 9 |
| ND3 | 348 | 13.4 | 0.073 | 20 | 4 |
| ND4 | 1377 | 61.5 | 0.081 | 101 | 7 |
| ND4L | 294 | 13.6 | 0.087 | 24 | 0 |
| ND5 | 1818 | 81.8 | 0.083 | 120 | 24 |
| ND6 | 525 | 25.8 | 0.090 | 40 | 6 |
| 12S rRNA | 967 | 18.1 | 0.035 | NA | NA |
| 16S rRNA | 1563 | 20.1 | 0.024 | NA | NA |
| tRNAs | 1508 | 22.5 | 0.027 | NA | NA |
| CR | 953 | 22.3 | 0.042 | NA | NA |
| Total | 16370 | 520.9 | 0.069 | 778 | 84 |

32
 33 n = average number of nucleotide differences between species; D_a: the number of net
 34 nucleotide substitutions per site between species; D_s: number of fixed synonymous
 35 substitution between species; D_n: number of fixed non-synonymous substitution
 36 between species.
 37

38 Table S3: Summary statistics of mitochondrial gene differentiation between
 39 dromedaries and wild Bactrian camels.
 40

| Gene | Length (bp) | n | D _a | D _s | D _n |
|----------|-------------|-------|----------------|----------------|----------------|
| ND1 | 954 | 41.9 | 0.079 | 68 | 4 |
| ND2 | 1041 | 44.8 | 0.077 | 60 | 15 |
| COX1 | 1542 | 50.9 | 0.060 | 90 | 0 |
| COX2 | 681 | 30.3 | 0.081 | 52 | 1 |
| COX3 | 783 | 34.0 | 0.077 | 55 | 1 |
| AP6 | 678 | 28.2 | 0.077 | 44 | 7 |
| AP8 | 201 | 10.5 | 0.098 | 12 | 7 |
| CYTB | 1137 | 61.4 | 0.099 | 95 | 15 |
| ND3 | 348 | 13.2 | 0.070 | 20 | 3 |
| ND4 | 1377 | 59.3 | 0.078 | 97 | 8 |
| ND4L | 294 | 13.5 | 0.084 | 23 | 0 |
| ND5 | 1818 | 85.4 | 0.084 | 124 | 23 |
| ND6 | 525 | 27.7 | 0.097 | 44 | 6 |
| 12S rRNA | 967 | 20.4 | 0.038 | NA | NA |
| 16S rRNA | 1563 | 20.4 | 0.023 | NA | NA |
| tRNAs | 1508 | 22.9 | 0.027 | NA | NA |
| CR | 953 | 23.3 | 0.043 | NA | NA |
| Total | 16370 | 588.9 | 0.070 | 784 | 90 |

41
 42 n = average number of nucleotide differences between species; D_a: the number of net
 43 nucleotide substitutions per site between species; D_s: number of fixed synonymous
 44 substitution between species; D_n: number of fixed non-synonymous substitution
 45 between species.
 46

47
48
49
50

Table S4: Summary statistics of mitochondrial gene differentiation between domestic and wild Bactrian camels.

| Gene | Length (bp) | n | D _a | D _s | D _n |
|----------|-------------|-------|----------------|----------------|----------------|
| ND1 | 954 | 13.6 | 0.023 | 20 | 1 |
| ND2 | 1041 | 12.9 | 0.020 | 17 | 2 |
| COX1 | 1542 | 12.1 | 0.014 | 21 | 0 |
| COX2 | 681 | 7.5 | 0.018 | 12 | 0 |
| COX3 | 783 | 10.7 | 0.022 | 16 | 1 |
| AP6 | 678 | 6.1 | 0.016 | 11 | 0 |
| AP8 | 201 | 2.9 | 0.025 | 4 | 1 |
| CYTB | 1137 | 16.0 | 0.023 | 20 | 6 |
| ND3 | 348 | 2.6 | 0.014 | 4 | 1 |
| ND4 | 1377 | 18.4 | 0.022 | 27 | 3 |
| ND4L | 294 | 7.3 | 0.045 | 13 | 0 |
| ND5 | 1818 | 19.3 | 0.016 | 21 | 7 |
| ND6 | 525 | 9.4 | 0.030 | 12 | 3 |
| 12S rRNA | 967 | 3.5 | 0.006 | NA | NA |
| 16S rRNA | 1563 | 4.8 | 0.004 | NA | NA |
| tRNAs | 1508 | 6.0 | -0.00005 | NA | NA |
| CR | 953 | 6.8 | 0.012 | NA | NA |
| Total | 16370 | 160.7 | 0.018 | 198 | 25 |

51
52
53
54
55
56

n = average number of nucleotide differences between species; D_a: the number of net nucleotide substitutions per site between species; D_s: number of fixed synonymous substitution between species; D_n: number of fixed non-synonymous substitution between species.

57 Table S5: Pairwise divergence estimates of the electron transport system (ETS) complexes between the three Old world camel species.

58

| ETS complex | mitochondrial genes | Concatenated sequence (bp) | D _a (db) | D _a (df) | D _a (bf) | D _n (db) | D _n (df) | D _n (bf) | D _n (per 1000 bp) | | |
|-------------|-------------------------------------|----------------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|------------------------------|------|------|
| ETS I | ND1, ND2, ND3, ND4L, ND4, ND5, ND6, | 6,357 | 0.082 | 0.081 | 0.022 | 59 | 59 | 17 | 9.28 | 9.2 | 2.6 |
| ETS III | CYTB | 1,137 | 0.092 | 0.099 | 0.023 | 9 | 15 | 6 | 7.91 | 13.1 | 5.2 |
| ETS IV | COX1 COX2 COX3 | 3,006 | 0.068 | 0.069 | 0.017 | 3 | 2 | 1 | 0.99 | 0.66 | 0.33 |
| ETS V | ATP6 ATP8 | 879 | 0.081 | 0.082 | 0.018 | 13 | 14 | 1 | 14.7 | 15.9 | 1.13 |

59

60 D_a: number of net nucleotide substitutions per site between species; D_n: number of fixed nonsynonymous substitution between species; d:

61 dromedary; b: Bactrian camel and f: wild two-humped camel.

62

63

64

65 Table S6: Global MK test on 11,379 bp concatenated protein-coding genes conducted on pairwise comparisons between dromedaries,
 66 domestic and wild Bactrian camels, respectively. The two-tailed *P*-values of Fisher's exact tests are shown. *P*-values with asterisk are
 67 significant with the following significant levels * $P < 0.05$; ** $P < 0.01$.
 68

| | Synonymous Substitutions | | Non-synonymous Substitutions | | Neutrality index (NI) | <i>P</i> -value |
|-----------------------------------|--|----------------------------------|--|----------------------------------|-----------------------|-----------------|
| | Fixed differences between species (Ds) | Polymorphic sites among all (Ps) | Fixed differences between species (Dn) | Polymorphic sites among all (Pn) | | |
| dromedary vs. domestic Bactrian | 778 | 125 | 84 | 27 | 2 | 0.006** |
| dromedary vs. wild Bactrian camel | 784 | 123 | 90 | 20 | 1.41 | 0.244 |
| domestic vs. wild Bactrian camel | 198 | 62 | 25 | 17 | 2.17 | 0.035* |
| One- vs two-humped camels | 678 | 348 | 75 | 57 | 1.48 | 0.041* |

69
70

71 Table S7: Polymorphisms within *C. dromedarius* species and divergence to *C. bactrianus* at synonymous and nonsynonymous positions
72 in the mitochondrial protein coding genes. The two-tailed *P*-values of Fisher's exact tests are shown. *P*-values with asterisk are
73 significant with significance level of ** $P < 0.01$.
74

| | Fixed synonymous differences (Ds) | Synonymous polymorphic sites (Ps) | Fixed nonsynonymous differences (Dn) | Nonsynonymous polymorphic sites (Pn) | Neutrality index | <i>P</i> -value |
|--------------------|-----------------------------------|-----------------------------------|--------------------------------------|--------------------------------------|------------------|-----------------|
| ND1 | 68 | 8 | 3 | 1 | 2.83 | 0.385 |
| ND2 | 62 | 9 | 15 | 2 | 0.91 | 1 |
| COX1 | 96 | 11 | 0 | 0 | - | 1 |
| COX2 | 50 | 4 | 1 | 0 | 0 | 1 |
| ATP8 | 12 | 0 | 6 | 1 | - | 0.368 |
| ATP6 | 44 | 5 | 7 | 1 | 1.25 | 1 |
| COX3 | 48 | 8 | 2 | 0 | 0 | 1 |
| ND3 | 20 | 3 | 4 | 0 | 0 | 1 |
| ND4L | 24 | 2 | 0 | 0 | - | 1 |
| ND4 | 101 | 15 | 7 | 2 | 1.92 | 0.352 |
| ND5 | 120 | 17 | 24 | 4 | 1.17 | 0.759 |
| ND6 | 40 | 2 | 6 | 4 | 13.33 | 0.009** |
| CYTB | 93 | 9 | 9 | 0 | 0 | 1 |
| Concatenated genes | 778 | 93 | 84 | 15 | 1.49 | 0.178 |

75
76

77 Table S8: Polymorphism within *C. dromedarius* and divergence to *C. ferus* at synonymous and nonsynonymous positions in protein
 78 coding genes in mtDNA. The two-tailed *P*-values of Fisher's exact tests are shown. The *P*-values with asterisk are significant with
 79 significance level of ** $P < 0.01$.
 80

| | Fixed synonymous differences (Ds) | Synonymous polymorphic sites (Ps) | Fixed nonsynonymous differences (Dn) | Nonsynonymous polymorphic sites (Pn) | Neutrality index | <i>P</i> -value |
|--------------------|-----------------------------------|-----------------------------------|--------------------------------------|--------------------------------------|------------------|-----------------|
| ND1 | 68 | 8 | 4 | 1 | 2.12 | 0.453 |
| ND2 | 60 | 9 | 15 | 2 | 0.88 | 1 |
| COX1 | 90 | 11 | 0 | 0 | - | 1 |
| COX2 | 52 | 4 | 1 | 0 | 0 | 1 |
| ATP8 | 12 | 0 | 7 | 1 | - | 0.4 |
| ATP6 | 44 | 5 | 7 | 1 | 1.25 | 1 |
| COX3 | 55 | 8 | 1 | 0 | 0 | 1 |
| ND3 | 20 | 3 | 3 | 0 | 0 | 1 |
| ND4L | 23 | 2 | 0 | 0 | - | 1 |
| ND4 | 97 | 15 | 8 | 2 | 1.61 | 0.629 |
| ND5 | 124 | 17 | 23 | 4 | 1.26 | 0.750 |
| ND6 | 44 | 2 | 6 | 4 | 14.66 | 0.007** |
| CYTB | 95 | 9 | 15 | 0 | 0 | 0.600 |
| Concatenated genes | 784 | 93 | 90 | 15 | 1.40 | 0.249 |

81
82

83 Table S9: Polymorphism within *C. bactrianus* and divergence to *C. dromedarius* at synonymous and nonsynonymous positions in protein
 84 coding genes in mtDNA. The two-tailed *P*-values of Fisher's exact tests are shown. The *P*-values with asterisk are significant with
 85 significance level of ** $P < 0.01$.
 86

| | Fixed synonymous differences (Ds) | Synonymous polymorphic sites (Ps) | Fixed nonsynonymous differences (Dn) | Nonsynonymous polymorphic sites (Pn) | Neutrality index | <i>P</i> -value |
|--------------------|-----------------------------------|-----------------------------------|--------------------------------------|--------------------------------------|------------------|-----------------|
| ND1 | 68 | 3 | 3 | 1 | 7.55 | 0.200 |
| ND2 | 62 | 4 | 15 | 1 | 1.03 | 1 |
| COX1 | 96 | 0 | 0 | 0 | - | 1 |
| COX2 | 50 | 0 | 1 | 0 | - | 1 |
| ATP8 | 12 | 1 | 6 | 1 | 2 | 1 |
| ATP6 | 44 | 2 | 7 | 0 | 0 | 1 |
| COX3 | 48 | 0 | 2 | 2 | - | 0.004 |
| ND3 | 20 | 0 | 4 | 0 | - | 1 |
| ND4L | 24 | 1 | 0 | 0 | - | 1 |
| ND4 | 101 | 8 | 7 | 1 | 1.80 | 0.483 |
| ND5 | 120 | 3 | 24 | 3 | 5 | 0.071 |
| ND6 | 40 | 4 | 6 | 0 | 0 | 1 |
| CYTB | 93 | 6 | 9 | 3 | 5.16 | 0.056 |
| Concatenated genes | 778 | 32 | 84 | 12 | 3.47 | 0.001** |

87
88

89 Table S10: Polymorphism within *C. bactrianus* and divergence to *C. ferus* at synonymous and nonsynonymous positions in protein
 90 coding genes in mtDNA. The two-tailed *P*-values of Fisher's exact tests are shown. The *P*-values with asterisk are significant with
 91 significance level of ** $P < 0.01$.
 92

| | Fixed synonymous differences (Ds) | Synonymous polymorphic sites (Ps) | Fixed nonsynonymous differences (Dn) | Nonsynonymous polymorphic sites (Pn) | Neutrality index | <i>P</i> -value |
|--------------------|-----------------------------------|-----------------------------------|--------------------------------------|--------------------------------------|------------------|-----------------|
| ND1 | 20 | 3 | 1 | 1 | 6.66 | 0.300 |
| ND2 | 17 | 4 | 2 | 1 | 2.12 | 0.521 |
| COX1 | 21 | 0 | 0 | 0 | - | 1 |
| COX2 | 12 | 0 | 0 | 0 | - | 1 |
| ATP8 | 4 | 1 | 1 | 1 | 4 | 1 |
| ATP6 | 11 | 2 | 0 | 0 | - | 1 |
| COX3 | 16 | 0 | 1 | 2 | - | 0.017 |
| ND3 | 4 | 0 | 1 | 0 | - | 1 |
| ND4L | 13 | 1 | 0 | 0 | - | 1 |
| ND4 | 27 | 8 | 3 | 1 | 1.12 | 1 |
| ND5 | 21 | 3 | 7 | 3 | 3 | 0.328 |
| ND6 | 12 | 4 | 3 | 0 | 0 | 1 |
| CYTB | 20 | 6 | 6 | 3 | 1.66 | 0.664 |
| Concatenated genes | 198 | 32 | 25 | 12 | 2.97 | 0.008** |

93
 94

95 Table S11: Polymorphism within *C. ferus* and divergence to *C. dromedarius* at synonymous and nonsynonymous positions in protein
 96 coding genes in mtDNA. The two-tailed *P*-values of Fisher's exact tests are shown.
 97

| | Fixed synonymous differences (Ds) | Synonymous polymorphic sites (Ps) | Fixed nonsynonymous differences (Dn) | Nonsynonymous polymorphic sites (Pn) | Neutrality index | <i>P</i> -value |
|--------------------|-----------------------------------|-----------------------------------|--------------------------------------|--------------------------------------|------------------|-----------------|
| ND1 | 68 | 4 | 4 | 0 | 0 | 1 |
| ND2 | 60 | 4 | 15 | 1 | 1 | 1 |
| COX1 | 90 | 3 | 0 | 0 | - | 1 |
| COX2 | 52 | 3 | 1 | 0 | 0 | 1 |
| ATP8 | 12 | 0 | 7 | 0 | - | 1 |
| ATP6 | 44 | 0 | 7 | 0 | - | 1 |
| COX3 | 55 | 4 | 1 | 0 | 0 | 1 |
| ND3 | 20 | 0 | 3 | 1 | - | 0.166 |
| ND4L | 23 | 1 | 0 | 0 | - | 1 |
| ND4 | 97 | 2 | 8 | 0 | 0 | 1 |
| ND5 | 124 | 7 | 23 | 1 | 0.77 | 1 |
| ND6 | 44 | 0 | 6 | 1 | - | 0.137 |
| CYTB | 95 | 2 | 15 | 1 | 3.16 | 0.370 |
| Concatenated genes | 784 | 30 | 90 | 5 | 1.45 | 0.401 |

98
99

100 Table S12: Polymorphism within *C. ferus* and divergence to *C. bactrianus* at synonymous and nonsynonymous positions in protein
 101 coding genes in mtDNA. The two-tailed *P*-values of Fisher's exact tests are shown.
 102

| | Fixed synonymous differences (Ds) | Synonymous polymorphic sites (Ps) | Fixed nonsynonymous differences (Dn) | Nonsynonymous polymorphic sites (Pn) | Neutrality index | <i>P</i> -value |
|--------------------|-----------------------------------|-----------------------------------|--------------------------------------|--------------------------------------|------------------|-----------------|
| ND1 | 20 | 4 | 1 | 0 | 0 | 1 |
| ND2 | 17 | 4 | 2 | 1 | 2.12 | 0.521 |
| COX1 | 21 | 3 | 0 | 0 | - | 1 |
| COX2 | 12 | 3 | 0 | 0 | - | 1 |
| ATP8 | 4 | 0 | 1 | 0 | - | 1 |
| ATP6 | 11 | 0 | 0 | 0 | - | 1 |
| COX3 | 16 | 4 | 1 | 0 | 0 | 1 |
| ND3 | 4 | 0 | 1 | 1 | - | 0.333 |
| ND4L | 13 | 1 | 0 | 0 | - | 1 |
| ND4 | 27 | 2 | 3 | 0 | 0 | 1 |
| ND5 | 21 | 7 | 7 | 1 | 0.42 | 0.651 |
| ND6 | 12 | 0 | 3 | 1 | - | 0.25 |
| CYTB | 20 | 2 | 6 | 1 | 1.66 | 1 |
| Concatenated genes | 198 | 30 | 25 | 5 | 1.32 | 0.574 |

103

104 Table S13: Potential positively selected sites detected by branch-site model A
 105 implemented in CODEML (PAML) with the correspondent Bayes empirical Bayes
 106 (BEB) values. The significance of the detection of sites under positive selection on
 107 each of the selected branch was assessed by Likelihood Ratio Test (LRT). n.s: non
 108 significant values of LRT. *Posterior probability in BEB > 95%, Posterior probability
 109 50% - 95% without asterisk.
 110

| Foreground branch | Gene | Codon | Amino acid | BEB | Δ LRT | d.f | P-value |
|------------------------|-----------------|-------|------------|--------|--------------|-----|---------|
| <i>C. dromedarius</i> | ND2 | 98 | V | 0.517 | 0 | 1 | n.s |
| | | 219 | L | 0.504 | 0 | 1 | n.s |
| | ATP6 | 1 | V | 0.671 | 0 | 1 | n.s |
| | | 38 | V | 0.592 | 0 | 1 | n.s |
| <i>C. bactrianus</i> | ND1 | 15 | V | 0.515 | -0.46 | 1 | n.s |
| | ND2 | 243 | L | 0.627 | 0.07 | 1 | n.s |
| | COX3 | 41 | I | 0.543 | 0 | 1 | n.s |
| | | 165 | I | 0.723 | 0 | 1 | n.s |
| | ND3 | 86 | T | 0.844 | 0 | 1 | n.s |
| | ND4 | 60 | S | 0.572 | 0 | 1 | n.s |
| | ND5 | 161 | A | 0.701 | 0 | 1 | n.s |
| | | 543 | T | 0.634 | 0 | 1 | n.s |
| | | 592 | I | 0.697 | 0 | 1 | n.s |
| | | 595 | I | 0.56 | 0 | 1 | n.s |
| | <i>C. ferus</i> | ND2 | 62 | T | 0.576 | 0 | 1 |
| ATP8 | | 44 | T | 0.758 | 0 | 1 | n.s |
| ND4 | | 48 | N | 0.704 | 0 | 1 | n.s |
| | | 435 | A | 0.693 | 0 | 1 | n.s |
| ND5 | | 470 | N | 0.677 | 0 | 1 | n.s |
| | | 495 | L | 0.678 | 0 | 1 | n.s |
| | | 536 | V | 0.675 | 0 | 1 | n.s |
| CYTB | | 4 | I | 0.638 | 0 | 1 | n.s |
| | | 14 | M | 0.529 | 0 | 1 | n.s |
| | | 39 | V | 0.598 | 0 | 1 | n.s |
| | | 349 | M | 0.618 | 0 | 1 | n.s |
| | 353 | V | 0.654 | 0 | 1 | n.s | |
| <i>Camelus</i> lineage | ND2 | 227 | L | 0.649 | 0 | 1 | n.s |
| | | 347 | Y | 0.625 | 0 | 1 | n.s |
| | COX2 | 94 | V | 0.589 | -0.55 | 1 | n.s |
| | ATP8 | 22 | V | 0.584 | 0.94 | 1 | n.s |
| | | 34 | P | 0.546 | 0.94 | | n.s |
| | | 38 | S | 0.648 | 0.94 | | n.s |
| | COX3 | 40 | S | 0.944 | 2.80 | 1 | n.s |
| | ND3 | 19 | I | 0.537 | 0.95 | 1 | n.s |
| | ND4 | 4 | Y | 0.514 | 3.84 | 1 | n.s |
| | | 50 | L | 0.848 | 3.84 | | n.s |
| | | 82 | H | 0.953* | 3.84 | | n.s |
| | | 183 | V | 0.918 | 3.84 | | n.s |
| | | 187 | H | 0.644 | 3.84 | | n.s |
| | | 248 | L | 0.673 | 3.84 | | n.s |
| | | 438 | S | 0.508 | 3.84 | | n.s |
| | 458 | L | 0.636 | 3.84 | | n.s | |
| ND5 | 183 | A | 0.766 | 0.11 | 1 | n.s | |
| | 432 | N | 0.838 | 0.11 | | n.s | |

111 Table S14: Fixed nonsynonymous substitutions between species-specific lineages of camels, their peptide chain locations, inferred
 112 ancestral states, and impact on local physicochemical properties are shown. Pairwise comparisons between three camel species
 113 (dromedary, domestic and wild Bactrian) are shown with abbreviations db, df and bf. The ancestral states of the amino acid changes were
 114 inferred using MEGA6. Magnitude of the amino acid changes (categories ≥ 6) and the affected physicochemical properties are inferred
 115 with TreeSAAP (v.3.2): Increase (\uparrow) and decrease (\downarrow) in amino acid properties in a given branch is indicated.
 116

| Gene | Amino Acid site | Location | Ancestral state | <i>C. dromedarius</i> | <i>C. bactrianus</i> | <i>C. ferus</i> | Magnitude of change | | |
|------|-----------------|------------------|-----------------|-----------------------|----------------------|-----------------|--|--|---|
| | | | | | | | <i>C. dromedarius</i> --> <i>C. bactrianus</i> | <i>C. dromedarius</i> --> <i>C. ferus</i> | <i>C. bactrianus</i> --> <i>C. ferus</i> |
| ATP8 | 36 | Nontransmembrane | Ala | Asp | Ala | Ala | 7 *** \downarrow , P_r | 7 *** \downarrow , P_r | |
| ATP6 | 47 | Nontransmembrane | His | His | Tyr | Tyr | 7 *** \downarrow , α_m | 7 *** \downarrow , α_m | |
| | 51 | Nontransmembrane | Arg | Gln | Arg | Arg | 7 *** \uparrow , pHi | 7 *** \uparrow , pHi | |
| ND3 | 30 | Nontransmembrane | Ile | Thr | Ile | Ile | 8 *** \uparrow , R_a 8 *** \uparrow , H_p | | |
| | 87 | Transmembrane | Thr | Thr | Ile | Thr | 8 *** \uparrow , R_a 8 *** \uparrow , H_p | | 8 *** \downarrow , R_a |
| ND5 | 69 | Nontransmembrane | Ile | Ile | Met | Met | | 8 *** \uparrow , pK' | |
| | 130 | Transmembrane | Ile | Val | Ile | Ile | | 8 *** \downarrow , pK' | |
| | 162 | Nontransmembrane | Ala | Ala | Thr | Ala | 6 *** \downarrow , P_a | | |
| | 195 | Nontransmembrane | Thr | Thr | Ala | Ala | 6 *** \uparrow , P_a | | |
| | 214 | Transmembrane | Met | Ile | Met | Met | | 8 *** \uparrow , pK' | |
| | 272 | Transmembrane | Thr | Met | Thr | Thr | 6 *** \downarrow , P_a | | |
| | 375 | Transmembrane | Ile | Ile | Val | Val | | 8 *** \uparrow , pK' | |
| | 394 | Nontransmembrane | Met | Thr | Met | Met | 6 *** \uparrow , P_a | | |
| | 398 | Nontransmembrane | Ala | Ala | Thr | Thr | 6 *** \downarrow , P_a | | |
| | 414 | Transmembrane | Ile | Val | Ile | Ile | | 8 *** \downarrow , nK' | |

| | | | | | | | | | |
|-----|-----|---------------|-----|-----|-----|-----|--|-------------------------|--|
| | 599 | Transmembrane | Ile | Ile | Ala | Ala | | 8 ^{***} ↑, pK' | |
| ND6 | 12 | Transmembrane | Ile | Ile | Ile | Val | | 8 ^{***} ↑, pK' | |
| | 18 | Transmembrane | Ile | Ile | Val | Val | | 8 ^{***} ↑, pK' | |
| | 102 | Transmembrane | Ser | Ile | Ser | Ala | | 8 ^{***} ↑, pK' | |
| | 105 | Transmembrane | Val | Ile | Val | Val | | 8 ^{***} ↑, pK' | |

117
118 P_{α} : Alpha-helical tendencies, pK': Equilibrium constant (ionization of COOH), P_r : Polar requirement, H_p : Surrounding hydrophobicity,
119 α_m : Power to be at the middle of alpha-helix, pHi : Isoelectric point, R_a : Solvent accessible reduction ratio. The location of the peptides
120 containing the amino acid changes were predicted by the program Phobius. The P -values with asterisk are significant with following
121 significant levels *** $P < 0.001$.