Ancient and modern DNA reveal dynamics of domestication and cross-continental dispersal of the dromedary

Faisal Almathen[a,b,1], Pauline Charruau[c,d,e,1], Elmira Mohandesan[c,d,1], Joram M. Mwacharo[b], Pablo Orozco-terWengel[f], Daniel Pitt[f], Abdussamad M. Abdussamad[g], Margarethe Uerpmann[h], Hans-Peter Uerpmann[h], Bea De Cupere[i], Peter Magee[j], Majed A. Alnaqeeb[k], Bashir Salim[l], Abdul Raziq[m], Tadelle Dessie[n], Omer M. Abdelhadi[o], Mohammad H. Banabazi[p], Marzook Al-Eknah[q], Chris Walzer[d], Bernard Faye[r], Michael Hofreiter[s], Joris Peters[t,u], Olivier Hanotte[b,2], and Pamela A. Burger[c,2]

[a]Department of Veterinary Public Health and Animal Husbandry, College of Veterinary Medicine and Animal Resources, King Faisal University, 400 Al-Hasa, Saudi Arabia; [b]Genetics, Ecology and Evolution, School of Life Sciences, University of Nottingham, NG7 2RD Nottingham, United Kingdom; [c]Research Institute of Wildlife Ecology, Vetmeduni Vienna, 1160 Vienna, Austria; [d]Institute of Population Genetics, Vetmeduni Vienna, 1210 Vienna, Austria; [e]Department of Ecology and Evolution, University of Lausanne, 1015 Lausanne, Switzerland; [f]School of Biosciences, Cardiff University, Cardiff CF10 3AX, Wales, United Kingdom; [g]Department of Animal Science, Faculty of Agriculture, Bayero University, PMB 3011, Kano State, Nigeria; [h]Institut für Naturwissenschaftliche Archäologie, Abteilung Archäozoologie, Universität Tübingen, 7207 Tuebingen, Germany; [i]Department of Paleontology, Royal Belgian Institute of Natural Sciences, 1000 Brussels, Belgium; [j]Department of Classical and Near Eastern Archaeology, Bryn Mawr College, Bryn Mawr, PA 19010; [k]Department of Biological Sciences, Faculty of Science, Kuwait University, Safat 13060, Kuwait; [l]Department of Parasitology, Faculty of Veterinary Medicine, University of Khartoum, Khartoum-North 13314, Sudan; [m]Faculty of Veterinary and Animal Sciences, Lasbela University of Agriculture, Water and Marine Sciences, Uthal 90150, Pakistan; [n]Animal Biosciences, International Livestock Research Institute, Addis Ababa 1000, Ethiopia; [o]Department of Animal Production, Faculty of Natural Resources and Environmental Studies, University of Kordofan, Khartoum 11111, Sudan; [p]Department of Biotechnology, Animal Science Research Institute of Iran, 3146618361 Karaj, Iran; [q]Department of Clinical Studies, College of Veterinary Medicine and Animal Resources, King Faisal University, 1757 Al-Hasa, Saudi Arabia; [r]Centre de Coopération Internationale en Recherche Agronomique pour le Développement-Environnements et Sociétés, UMR 112, Campus International de Baillarguet, TAC/112A, 34398 Montpellier, France; [s]Department of Mathematics and Natural Sciences, Evolutionary and Adaptive Genomics, Institute for Biochemistry and Biology, University of Potsdam, 14476 Potsdam, Germany; [t]Department of Veterinary Sciences, Institute of Palaeoanatomy, Domestication Research and the History of Veterinary Medicine, Ludwig-Maximilians-Universität München (LMU Munich), 80539 Munich, Germany; and [u]Staatliche Naturwissenschaftliche Sammlungen Bayerns, Bavarian State Collection of Anthropology and Palaeoanatomy, 80333 Munich, Germany

Edited by Melinda A. Zeder, National Museum of Natural History, Santa Fe, NM, and approved April 4, 2016 (received for review October 6, 2015)

Dromedaries have been fundamental to the development of human societies in arid landscapes and for long-distance trade across hostile hot terrains for 3,000 y. Today they continue to be an important livestock resource in marginal agro-ecological zones. However, the history of dromedary domestication and the influence of ancient trading networks on their genetic structure have remained elusive. We combined ancient DNA sequences of wild and early-domesticated dromedary samples from arid regions with nuclear microsatellite and mitochondrial genotype information from 1,083 extant animals collected across the species' range. We observe little phylogeographic signal in the modern population, indicative of extensive gene flow and virtually affecting all regions except East Africa, where dromedary populations have remained relatively isolated. In agreement with archaeological findings, we identify wild dromedaries from the southeast Arabian Peninsula among the founders of the domestic dromedary gene pool. Approximate Bayesian computations further support the "restocking from the wild" hypothesis, with an initial domestication followed by introgression from individuals from wild, now-extinct populations. Compared with other livestock, which show a long history of gene flow with their wild ancestors, we find a high initial diversity relative to the native distribution of the wild ancestor on the Arabian Peninsula and to the brief coexistence of early-domesticated and wild individuals. This study also demonstrates the potential to retrieve ancient DNA sequences from osseous remains excavated in hot and dry desert environments.

anthropogenic admixture | *Camelus dromedarius* | demographic history | paleogenetics | wild dromedary

The dromedary (*Camelus dromedarius*) is one of the largest domestic ungulates and one of the most recent additions to livestock. Known as the "ship of the desert" (1), it enabled the transportation of people and valuable goods (e.g., salt, incense, spices) over long distances connecting Arabia, the Near East, and North Africa. This multipurpose animal has outperformed all other domestic mammals, including the donkey, in arid environments and continues to provide basic commodities to millions of people inhabiting marginal agro-ecological zones. In the current context of advancing desertification and global climate change, there is renewed interest in the biology and production traits of the species (2), with the first annotated genome drafts having been recently released (3, 4).

Significance

The dromedary is one of the largest domesticates, sustainably used in arid and hostile environments. It provides food and transport to millions of people in marginal agricultural areas. We show how important long-distance and back-and-forth movements in ancient caravan routes shaped the species' genetic diversity. Using a global sample set and ancient mitochondrial DNA analyses, we describe the population structure in modern dromedaries and their wild extinct ancestors. Phylogenetic analyses of ancient and modern dromedaries suggest a history of restocking from wild animals from the southeast coast of the Arabian Peninsula. Dromedaries now extend the list of species for which classic models of domestication from a single center and from wild conspecific individuals in isolation are rejected.

Author contributions: O.H. and P.A.B. designed research; F.A., P.C., E.M., and A.M.A. performed research; F.A., P.O.-T.W., A.M.A., M.U., H.-P.U., B.D.C., P.M., M.A.A., B.S., A.R., T.D., O.M.A., M.H.B., M.A.-E., C.W., B.F., M.H., J.P., O.H., and P.A.B. contributed new reagents/analytic tools; F.A., P.C., E.M., J.M.M., P.O.-t.W., D.P., and P.A.B. analyzed data; and F.A., P.C., E.M., P.O.-t.W., M.H., J.P., O.H., and P.A.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The sequences reported in this paper have been deposited in the Genbank database (accession nos. JX946206–JX946273, KF719283–KF719290, and KT334316–KT334323).

See Commentary on page 6588.

[1]F.A., P.C., and E.M. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: pamela.burger@vetmeduni.ac.at or olivier.hanotte@nottingham.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1519508113/-/DCSupplemental.

In contrast to other livestock species, the evolutionary history and domestication of Old World camelids (Camelini) have remained largely unexplored because of the scarcity of camel bone assemblages from well-dated archaeological contexts (5). Following the Pleistocene, the wild dromedary retreated to ecologically favored areas (i.e., mangrove habitats) on the Arabian Peninsula (6), a rather small geographic region compared with the native distributions of the wild ancestors of other domesticates (*SI Appendix*). The domestication of the dromedary likely happened in the late second millennium BCE as deduced from: (*i*) diachronic osteometric analysis illustrating a significant decrease in bone size in remains dating to the very end of the second or beginning of the first millennium BCE (ca. 1,100–800 BCE) (7–12); (*ii*) changes in the cultural context, i.e., increased representation of dromedary bones in settlement refuse vs. large concentrations in sites without architecture, e.g., site of Al-Sufouh, United Arab Emirates (UAE); and (*iii*) figurines and representations of indubitably domesticated dromedaries (13). Based on the available zooarchaeological records, it is assumed that the wild one-humped camel did not survive the start of the CE (8, 9, 12, 14), in contrast to the wild ancestors of most other livestock species (15, 16). Small numbers of domesticated dromedaries likely arrived in Mesopotamia by the second quarter of the first millennium BCE, but there, as well as in northeast Africa, larger herds appeared only during Late Antiquity and/or early medieval times (fourth to seventh centuries CE) (1, 11, 17). If its use as "camelry" in warfare was minor compared with the horse (1), the dromedary was readily adopted as beast of burden and continued fulfilling this role well into the 20th century CE in caravans sometimes encompassing thousands of animals (18, 19).

In the present study, we address the questions of domestication and demographic history of the dromedary across its geographic range, combining information from ancient DNA Sanger and next-generation sequencing data of wild and early-domestic dromedary osseous remains with modern nuclear (microsatellites) and mitochondrial genetic diversity. Our results show that the domestication process and the current diversity of the species were shaped by early introgression from the wild as well as by human-mediated factors.

Results and Discussion

Little Population Structure in Modern Dromedaries, a Consequence of Cross-Continental Back-and-Forth Movements. By examining modern genetic diversity and its global distribution, it is possible to gain insight into the domestication process, because, in the absence of recurrent introgression, populations close to the putative domestication centers are assumed to retain higher levels of ancestral polymorphism (20). Such distribution of genetic diversity has been suggested to explain the frequently observed negative correlation between genetic diversity and the geographic distance from the place of origin in numerous livestock species (20–25). In the case of the dromedary, before the introduction of the domestic form, there had been no representatives of *Camelus* on the African continent since the Late Pleistocene, and the Holocene native distribution of wild dromedaries seems to have been restricted to the Arabian Peninsula (6, 7). Modern dromedary populations from the Arabian Peninsula therefore were expected to display the highest level of genetic diversity and variation. To test this expectation, we combined two comprehensive datasets encompassing 759 mitochondrial (867 bp; end of cytochrome B, tRNAs threonine and proline, beginning of control region; MT-CR) and 970 multiloci (17 autosomal microsatellites) genotypes, covering five defined geographical regions (26): Eastern Africa (EAF, $n = 170$), Western and Northern Africa (WNAF, $n = 233$), North Arabian Peninsula (NAP, $n = 349$), South Arabian Peninsula (SAP, $n = 181$), and Southern Asia including Australia (SAS, $n = 150$) (Dataset S1).

Shared genetic diversity and population structure in modern dromedaries. In contrast to the hypothesis that the greatest ancestral variation is retained close to the area of domestication (20), we observed similar amounts of heterozygosity (H_E: 0.58–0.63) and allelic richness (*Ar*: 4.88–6.47) among the different populations (Bonferroni corrected Wilcoxon rank-sum test; $P > 0.05$) (*SI Appendix*, Table S1). This finding precluded any conclusion about the existence of an ancestral population or a geographic center of dispersion (for comparisons with other camelids, see *SI Appendix*). Shared diversity also was revealed by the analysis of molecular variance with 95.7% (nuclear) and 95.3% (mtDNA) of the variation distributed within populations. Hence, we investigated genetic population structure in modern dromedaries disregarding their geographic origins. Mitochondrial median-joining network (MJN) analysis (27) split the 76 haplotypes into two haplogroups, H_A and H_B, containing six major haplotypes (H_A: *A1* and *A2*; H_B: *B1–4*) (Fig. 1*B*). This partition was supported by Bayesian phylogenetic analysis [posterior probability (PP) = 0.98] (*SI Appendix*, Fig. S1). No phylogeographic pattern was detectable, because the six major haplotypes were observed across the global range of the species (Fig. 1*A*). In contrast, with the nuclear structure analysis we retrieved an optimal number of two ancestral populations (*SI Appendix*, Fig. S2*A*), clearly separating EAF dromedaries from all other populations (Fig. 2). This separation also is reflected in the 3D factorial correspondence analysis (*SI Appendix*, Fig. S3) and in the limited population differentiation (nuclear $F_{ST} = 0.013–0.070$) (*SI Appendix*, Table S2), a plausible consequence of the intense back-and-forth movements that characterized the use of dromedaries in cross-continental trading.

Genetic distinctiveness of East African dromedaries. Modern EAF dromedaries exhibit the lowest nuclear ($H_E = 0.58$, $Ar = 4.48$) but the highest mtDNA ($H_d = 0.79$, $\theta_\pi = 3.62$) diversity of all populations (*SI Appendix*, Table S1). These elevated values could, in principle, be explained by a large proportion of ancestral diversity in the mtDNA or by a cryptic population structure not accounted for in the analysis (28). Although 85% of the investigated haplotypes belonged to H_B, dromedaries in Eastern Africa exhibited a more balanced ratio between H_A (38%) and H_B (62%) (Fig. 1*A*). These results may be interpreted as the consequence of a random founder effect followed by successive gene flow with a restricted number of sires. Globalization of genetic diversity might not have affected the EAF as much as other populations, likely because of its isolation from the northern part of the continent by eco-geographical obstacles (e.g., the Ethiopian Plateau and the swamps of the Sudd), physiological constraints (humidity, food plants, lack of salt, disease) and, perhaps most importantly, cultural barriers (*SI Appendix*, Fig. S4).

Subtle population structure within the SAP. To investigate subtle population structure that might have been masked by the high distinctiveness of EAF, we excluded the latter from structure analysis and observed nine independent clusters (Fig. 2 and *SI Appendix*, Fig. S2*B*). Despite substantial admixture, two dromedary populations (Awarik and Awadi; Dataset S1) from an isolated mountainous region in southwestern Saudi Arabia segregated. Dromedaries from Oman and UAE separated from the cluster containing Southern Asian individuals, whereas WNAF and NAP populations shared common ancestry and genetic diversity. Within the latter only the Hadana breed (Dataset S1) appeared to have a contrasting genetic makeup (Fig. 2).

Introduction of Arabian dromedaries into Africa. The absence of genetic structure between WNAF and NAP ($\phi_{ST} = 0.006$; $P < 0.001$; $F_{ST} = -0.002$; $P > 0.05$) points to an extensive exchange of dromedaries introduced into northeastern Africa from the Arabian Peninsula via the Sinai (*SI Appendix*, Fig. S4), possibly starting in the early first millennium BCE and intensifying in the Ptolemaic period (1, 17). From here, dromedaries spread across northern Africa, but their adoption into local economies may have been slow, considering that the first unequivocal evidence for their presence in northwestern Africa comes from archaeological layers dating to the fourth to the seventh century CE (Late Antiquity/Early Middle Ages) (*SI Appendix*). Although WNAF-NAP showed close cross-continental affinities with Southern Arabian and Asian dromedaries, the two African populations were genetically the most distant (EAF/WNAF-NAP $\phi_{ST} = 0.164$; $F_{ST} = 0.040$; $P < 0.001$), in contrast with their geographical proximity. The lowest pairwise genetic distances for Eastern African dromedaries were actually measured with the SAP populations (*SI Appendix*, Table S2), suggesting a few possible routes for domestic dromedaries to be introduced to Eastern Africa. These involve the

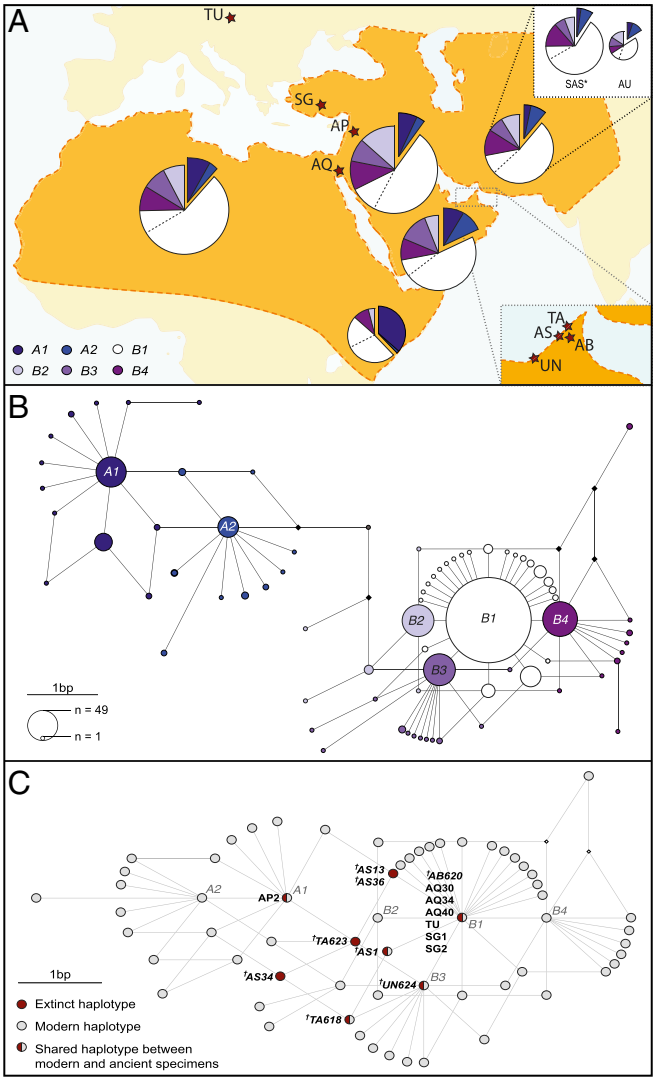


Fig. 1. Representation of the mitochondrial haplotypes retrieved from 759 modern dromedaries and 15 archaeological specimens. (*A*) Geographical distribution of the modern haplogroups across the species range (delimited by orange dashed line). Pie charts are proportional to sample sizes of the five distinctive regions (Dataset S1). Haplogroups were defined according to Bayesian analysis of population structure (BAPS) clustering (*SI Appendix*). The proportion of singletons diverging from B1 by one or two mutations (seventh cluster) is depicted by the dotted line within B1 (white). The chart in the upper right corner represents haplogroups retrieved from Southern Asian (SAS*; $n = 87$) and Australian (AU; $n = 38$) dromedaries. Stars depict locations of the archaeological sites: SG, Sagalassos, Turkey (Early Byzantine, 450–700 CE); TU, Tulln, Austria (second Ottoman–Habsburg war, ca. 1683 CE); AP, Apamea, Syria (Early Byzantine, 400–600 CE); AQ, Aqaba, Jordan (Mamluk and Ottoman periods, 1260–1870 CE). The *Inset* in the lower right corner shows sites in the UAE: AB, Al-Buhais (5000–4000 BCE); AS, Al-Sufouh (ca. 2400–1400 BCE); TA, Tell Abraq (Late Bronze–Iron Age, 1260–500 BCE); UN, Umm-an-Nar (Early Bronze Age, 3000–2000 BCE). (*B*) MJN displaying 76 haplotypes grouped into two maternal lineages, H_A (*A1* and *A2*) and H_B (*B1–4*). Haplotypes diverging from *A1* and *A2* and from *B1–4* are colored according to BAPS clustering (*SI Appendix*). Circles are proportional to the sample size. Small diamonds represent median vectors corresponding to missing haplotypes or homoplasies. (*C*) Parsimonious representation of the occurrence and sharing of mitochondrial haplotypes (531 bp) between modern (light gray) and ancient (dark red) samples. Wild dromedary samples are marked with a dagger (†). Taxonomic determinations of ancient specimens are detailed in *SI Appendix*. Umm-an-Nar's sample (UN624) was represented assuming the most frequent nucleotide (nt15486: G). In the case of the alternative allele (nt15486: A), UN624 shared its haplotype with the specimen from Tell Abraq (TA623) (*SI Appendix*). For both networks, consensus network of all shortest trees is shown; branch lengths are proportional to number of mutations.

transfer from the Arabian Peninsula by boat either directly across the Gulf of Aden or further north across the Red Sea to Egypt and then traveling south along the western coast of the Red Sea to northwestern Sudan, Eritrea, and Ethiopia (*SI Appendix*, Fig. S4). A seaborne introduction appears likely, because there is increasing evidence that the southern Arabian Peninsula played an important role in domestication [e.g., African wild ass (29)] and in the transfer of crops and livestock [e.g., zebu cattle, fat-tailed sheep (30, 31)] between South Asia and the African continent. Additional evidence for a separate introduction might come from socio-ethological observations; today's Eastern African dromedaries are used largely for milk production rather than for riding and transportation, and this use could be rooted in practices associated with the early stage of dromedary husbandry in the southern Arabian Peninsula (1, 7).

Representation of the global genetic diversity in Australian dromedaries. An interesting observation concerns the genetic makeup of the Australian population. Although animals were imported from a single geographic area (northwest of the Indian subcontinent) between the 1860s and 1920s (2, 32), domestic and feral Australian dromedaries possess all mtDNA haplogroups observed in the global population (Fig. 1*A*) and nuclear diversity similar to that of the global population (Fig. 2 and *SI Appendix*, Table S3). This diversity mirrors the extensive admixture in the dromedary population of the Old World through historical cross-continental exchanges that was already attained by the middle of the 19th century.

Domestication of Dromedaries and Restocking from the Wild in the Southeast Coast of the Arabian Peninsula.

Ancient mitochondrial haplotypes in early-domestic and wild (extinct) dromedaries. In absence of phylogeographic signals supporting the hypothesis of ancestral populations, we investigated the historic genetic repartition before the intensive gene flow induced by large-scale back-and-forth movements. Because poor DNA preservation in arid regions poses significant technical challenges (33), there are only a few findings from hot areas, where ancient DNA (aDNA) contributed significantly to the understanding of prehistoric events (34–37). In this study, we retrieved aDNA from up to 7,000-y-old wild dromedary specimens originating from archaeological contexts in the Arabian desert (*SI Appendix*, Table S4). We successfully amplified 531-bp mtDNA using 10 overlapping primer pairs (*SI Appendix*, Table S5) from eight wild dromedary bones from the sites Al-Sufouh (AS), Tell Abraq (TA), Umm-an-Nar (UN), and Al-Buhais (AB) in the UAE and from seven early-domesticated dromedary specimens excavated in Apamea (AP; Syria), Aqaba (AQ; Jordan), Sagalassos (SG; Turkey), and Tulln (TU; Austria) (Fig. 1*A*). No novel mitochondrial haplotypes were retrieved in the early-domesticated individuals, because six of them (AQ30, AQ34, AQ40, SG1, SG2, and TU) exhibited MT-CR sequences identical to those of the modern dromedaries belonging to the frequent haplotype *B1* (Fig. 1*C*). Only the Syrian specimen was characteristic of the rare haplotype *A1* (AP2) (Fig. 1*C*). This finding implies that both haplogroups (H_A and H_B) were already present in the Levantine herds of the fourth to seventh century CE. Different estimates of the time to the most recent common ancestor (TMRCA) of H_A and H_B [>5,700 y ago (ya)] (*SI Appendix*, Table S6) predate the assumed period of domestication during the end of the second or beginning of the first millennium BCE (7, 8, 12, 14), suggesting that at least two, but more likely a minimum of six wild maternal lineages were captured during the process of domestication. The eight ancient wild dromedary samples from four different locations in the UAE presented at least six different mitochondrial haplotypes (Fig. 1*C*) with a diversity of $\theta_\pi = 1.643$ and $H_d = 0.929$ (*SI Appendix*, Table S1). At least three of these remains (AS1, AB620, and TA618) shared their respective haplotypes with modern dromedaries belonging to haplogroup H_B. The last three retrieved haplotypes were unique to wild camels (AS13 with AS36, AS34, TA623) and occupied an intermediate position between the modern haplogroups H_A and H_B (Fig. 1*C*; see *SI Appendix* for UN624).

Wild dromedaries from the southeast coast of the Arabian Peninsula contribute to the domestic gene pool. The sharing of MT-CR sequences characteristic of H_B haplotypes between wild and modern

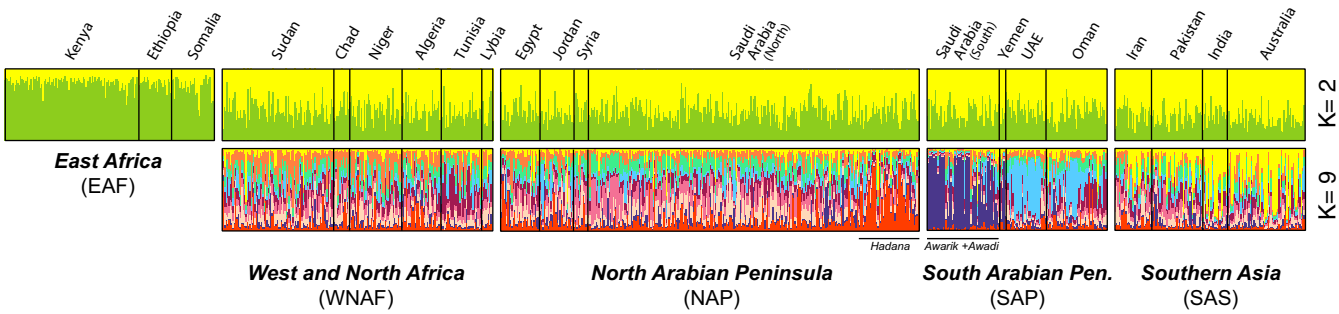


Fig. 2. Individual assignment (structure) plots of 970 (global dataset) and 810 dromedaries (excluding EAF) for a theoretical number of ancestral genetic populations (K) set at 2 and 9, respectively. Optimal clustering solution determined with DeltaK is reported in *SI Appendix*, Fig. S2. Sample sizes of the distinctive regions and countries are presented in *SI Appendix*, Table S1 and Dataset S1.

dromedaries from the same geographical region (today's UAE) illustrates the contribution of ancient relatives of these wild dromedary populations to the modern domestic gene pool. Although the wild specimens in our sample set come from a limited geographical distribution, large prehistoric faunal assemblages from sites dating from 5000–500 BCE in other parts of the Arabian Peninsula, such as coastal Yemen (38), have not yielded wild dromedary remains so far, indicating that at the time people started domesticating dromedaries, the native distribution of the wild ancestor of the one-humped camel already may have been limited to the Arabian southeast coast. This finding, together with the low frequency of H_A in modern dromedaries, suggests that the *A*-haplotypes were already present in lower frequency in the ancestral wild dromedary population, or, alternatively, were restricted to regions where there has been less intense archaeological research and/or poor faunal preservation.

Dynamics of Dromedary Domestication.

Population expansion in the context of domestication. In the context of domestication, molecular signals of sudden expansion are often interpreted as population growth or diffusion of domesticates across a wider geographic range (39). From the mtDNA, we obtained negative values of Tajima's D (-1.735; $P = 0.021$) and Fu's F_S (-87.48; $P < 10^{-5}$), which, in the absence of selection, indicate past demographic expansion. In the MJN analysis, we distinguished two haplogroups harboring six haplotypes at high frequencies, from which singletons radiate differing by one or two mutations (Fig. 1*B*). We could not reject the hypothesis that the pairwise differences between sequences of *A1* and *A2* and *B1–4* and their respective "derived" haplotypes were distributed according to a Poisson distribution, which indicates sudden expansion (40) and provides support for multiple contributions of ancestral female lineages to the current gene pool of modern dromedaries (*SI Appendix*). The Bayesian Skyline Plot (BSP) obtained from modern and early-domesticated maternal sequences (448 bp) shows a rise of the domestic N_e, around 600 ya [95% highest posterior density (HPD): 300–1,000 ya] (Fig. 3). This finding coincides with the Arab expansion in general and with the rise of the Ottoman Empire, the conquest of Constantinople (1453 CE), and of Southern Asia, including the Red Sea coasts, in the following century (41). Once Medina and Mecca had become part of the Empire (in the early 16th century CE), dromedaries were widely used for long-distance trade along the ancient Incense Route and for pilgrim transport (42) (*SI Appendix*, Fig. S4). There is tentative evidence that trade between southwest and southeast Arabia began as early as the first centuries of the first millennium BCE. This exchange was almost certainly camel-borne (13).

Approximate Bayesian computation inferences of domestication scenarios. Four scenarios can potentially explain the patterns of genetic diversity recorded in modern dromedaries: at the time of domestication, the initial gene pool was captured from: (*i*) one unique and diverse wild dromedary population; (*ii*) a primary small population of domesticates, with subsequent introgression of wild lineages into the early-domesticated gene pool; (*iii*) two independent source populations, each represented by one of the two observed ancestral lineages; or (*iv*) two source populations at successive time periods. Using approximate Bayesian computation (ABC) algorithms (43) on a combined mitochondrial and microsatellite dataset ($n = 642$), we simulated these four different scenarios (*SI Appendix*, Fig. S6). We obtained realistic PPs for up to 11 historic and demographic parameters (*SI Appendix*, Fig. S7), with the exception of the first scenario, for which the N_e of "Pop 2" was larger than 10^8 individuals and could not be reduced to a biologically meaningful value, and the time of divergence between populations was around 50 ya (generation time of 5 y). Thus, the remaining scenarios were compared to assess the one that best fit the data. The highest PP and Bayes Factor (BF) (*SI Appendix*, Tables S7 and S8) were obtained for the second scenario involving one domestication mode with introgression from a wild unsampled source population. In all pairwise comparisons the second scenario had a higher probability, with the BF ranging from ~63 to ~10^{23}. The remaining comparisons had substantially smaller BF values, mostly lower than 1 (*SI Appendix*, Table S8). This endorsement of the second scenario mirrors recent studies in pigs and other livestock in which a model incorporating continuous gene flow between a wild and a domestic species was better supported than traditional

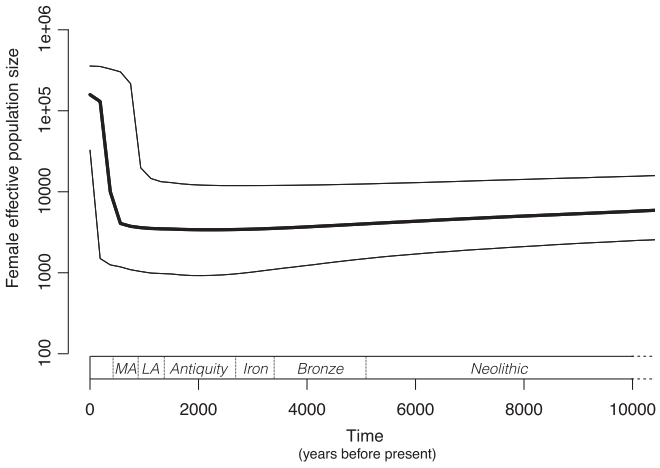


Fig. 3. BSP derived from the alignment of 759 modern with seven early-domesticated dromedary MT-CR sequences. The thick solid line depicts the median estimate of N_e, with black thin lines delimiting the 95% HPD. We used the archaeological dating of the wild and early-domesticated dromedary samples (*SI Appendix*, Table S4) to estimate the substitution rate $\mu = 1.232 \times 10^{-06}$ substitution·site^{-1}·y^{-1} (95% HPD: 4.435×10^{-07}, 2.213×10^{-06}). LA, Late Antiquity; MA, Middle Ages.

hypotheses assuming reproductive isolation (15, 16). Because wild and early-domesticated dromedaries coexisted in the Arabian Peninsula for only a short time [probably less than 2,000 y (8)], the period of potential gene flow was rather short compared those for cattle (16), pigs (15), or horses (25, 44, 45). This short period for potential gene flow, together with the possible existence of genomic islands of domestication, as recently proposed in pigs (15), likely explains the maintenance of the domestic phenotype in dromedaries. However, in the absence of complete genomes from wild dromedaries, this question requires further investigation.

Regarding the later introgression of an unsampled wild gene pool, the poor knowledge of the Holocene distribution of wild one-humped camels on the Arabian Peninsula is a limiting factor. Concentrations of bones indicative of larger camel herds have been found only in Neolithic to Bronze Age contexts on the eastern coast of the Arabian Peninsula (8, 14, 46, 47). The presence of pre-Iron Age camel remains in the Southern Levant has been controversial, because these specimens were considered to be intrusive to the archaeological context or unreliably ^{14}C-dated (9, 48).

Population bottlenecks predating domestication. Using coalescent simulations based on microsatellite diversity (MSVAR 1.3) in modern dromedaries, we captured several signals of severe bottlenecks (N_e reductions up to 65-fold) predating domestication (~8,600 ya in EAF; ~5,100 ya in the other populations) (*SI Appendix*, Fig. S8 and Table S9). The genetic distinctiveness of the EAF population, which could be a consequence of a random founder effect, might explain the precocity of its N_e decline. The drastic population reduction observed across all populations possibly relates to abrupt worldwide climate events, which triggered a general cooling and drying of the northern hemisphere, causing region-wide crop failures and the collapse of several civilizations (49–55). By the time cultural control over the wild dromedary was initiated, its native population and distribution may already have become diminished (*SI Appendix*, Fig. S5) and increasingly disjointed before the global extinction of the wild populations less than two millennia after the appearance of the domestic form (8, 14).

Given the environmental context in which the wild dromedary would have evolved, it can be assumed that its native distribution and population size were generally quite restricted compared with the ancestral ranges of other livestock species before domestication. As suggested by the environmental context of the archaeological findings, the wild ancestors of *C. dromedarius* spent part of their lives foraging in coastal habitats including mangroves (6). Salt is crucial to the health of camels (47, 56), and feeding in coastal habitats might have offered possibilities to enhance salt intake because of sea spray and the presence of halophyte vegetation. Because in prehistoric times mangroves may have occurred on the coastal southern Arabian Peninsula, the possibility that this region also sustained a wild dromedary population cannot be excluded. However, elevated sea levels and the lack of (zoo)archaeological investigations in the southern Arabian Peninsula may explain why genetic screening of the ancestral diversity remains incomplete.

Conclusion

The dromedary's fundamental role in the tradition of cross-continental caravan networks gave rise to an intense sharing of genetic variation, blurring genetic signals about ancestral diversity and possible center of domestication. Nevertheless, using a large modern DNA dataset in combination with a number of ancient sequences, we were able to support a scenario with an initial domestication followed by consecutive introgression from wild populations echoing findings from other species (57), such as horses (25, 44, 45), cattle (16), and pigs (15). Interestingly, in dromedaries, this restocking occurred from an unsourced wild "ghost" population, a pattern thus far observed in only few other domestic species (e.g., pigs and dogs). A remarkable feature in the history of dromedary domestication is the substantial genetic diversity of the domestic population, given the temporally and geographically restricted coexistence of early-domestic animals and their wild ancestors, which already were heading to extinction when the domestic form emerged. Modern dromedary populations largely maintained and consolidated this ancestral diversity, often lost in other livestock, underlining their potential to adapt sustainably to future challenges of desertification and climate change.

Materials and Methods

Modern and Ancestral Genetic Diversity. Hair, blood, and saliva samples were collected commensally during routine veterinary treatments, and all owners agreed to the analysis; no further specific permissions were required from the Ethics Committee of the Vetmeduni Vienna for this study. To infer the genetic diversity, population structure, and differentiation of the modern and ancient dromedary populations, we performed genetic analyses on a total of 1,083 modern dromedaries originating from 21 countries, seven early-domesticated (400–1870 CE) specimens, and eight wild dromedary specimens (5000–1000 BCE) (Fig. 1*A*). Wild dromedaries were classified based on the archaeological context (*SI Appendix*) and morphological differentiation (12). Detailed information about samples is given in Dataset S1; collection, wet-laboratory, and in silico procedures are given in *SI Appendix*, Table S4.

Population Genetics and Demographic Analysis. Genetic diversity estimators, genetic distances on the nuclear and mitochondrial data, and neutrality tests (mtDNA) are detailed in *SI Appendix*. Test of the goodness of fit for the Poisson distribution to the pairwise differences between the haplotypes and minimal mitochondrial diversity in the initial pool of domesticated camels (*SI Appendix*) followed Luikart et al. (58). Historical population demographic dynamics were assessed using the 448-bp MT-CR alignment from modern, early-domesticated, and wild samples. The birth–death skyline plot serial model (59) was implemented in BEAST 2.2.0 (60), accounting for serial samples taken at different time points (*SI Appendix*, Table S4). The resulting substitution rate was used to compute BSPs for domestic and wild dromedaries separately (*SI Appendix*). Coalescent simulations based on microsatellite diversity were implemented in MSVAR 1.3 (61, 62). The model assumes a single stable ancestral population N_1 at some time t_1 ago that experienced a demographic change (bottleneck or expansion) starting at time t and changed exponentially in size to the current population N_0. We simulated two different demographic scenarios by choosing (*i*) larger prior lognormal distribution values for N_0 than for N_1 (expansion) and (*ii*) vice versa (a bottleneck). In the absence of a species-specific microsatellite mutation rate in camels, we chose an average mammalian microsatellite mutation rate (63) of 10^{-4} (rate variation: 10^{-3}–10^{-5}) (*SI Appendix*).

ABC Inferences of Four Alternative Domestication Scenarios. To test the hypotheses of one independent or multiple domestication scenarios vs. restocking from the wild, we used ABCtoolbox (43) on the combined (n = 642) mitochondrial and microsatellite dataset. For each of the four scenarios (*SI Appendix*, Fig. S6) we simulated a large number of datasets (1,000,000) using Fastsimcoal2 (64) under the coalescent model drawing parameter values from prior distribution ranges (*SI Appendix*, Table S10). We tested a maximum of 11 historical parameters and generated 15 summary statistics for each simulation in Arlequin3.5 (65) (*SI Appendix*, Table S11). Summary statistics with highest pairwise correlations (R correlation test with Spearman's rho statistics; *SI Appendix*, Fig. S9) were removed, resulting in 12 summary statistics for further analysis. With the 5,000 simulations closest to the observed dataset, we evaluated model differentiation with the R package abc (66) (*SI Appendix*, Fig. S10) and assessed model fit with the ABC-GLM postsampling adjustment step built into ABCtoolbox (43, 67) to calculate marginal densities and probability of each scenario. Marginal distributions of each scenario were used to calculate PPs and BF for each pairwise comparison between scenarios; the alternative hypothesis can be rejected if the BF between two scenarios is greater than three (43, 68).

ACKNOWLEDGMENTS. We specifically thank the camel owners; the research group at the Management of Scientific Centres and Presidential Camels; His Highness the President Sheikh Khalifia Bin Zayed Al Nahayan; R. Saleh, D. Tabbaa, I. Awaddi, A. Perret, J. Perret, P. Spencer, S. Shahkarami, and M. Ismail, who collected samples; the International Livestock Research Institute; S. A. Jasim, Director of Archaeology in the Emirate of Sharjah, UAE, who facilitated research at al-Buhais and Tell Abraq; C. Vogl, C. Schlötterer, and M. Müller for scientific support; and the editor and two reviewers for their thoughtful comments on an early version of the manuscript. We acknowledge the use of the High Performance Computing Cluster, Organisms and Environment Division, School of Biosciences, Cardiff University. Funding was provided by Grant KFU, F289 from King Faisal University, Saudi Arabia (to F.A.), Grant NBAF419 from the Natural Environment Research Council, Sheffield University, United Kingdom (to O.H.); European Research Council Consolidator Grant 310763 GeneFlow (to M.H.); and Austrian Science Fund Grants P21084-B17 and P24706-B25 (to P.A.B.). P.A.B. is the recipient of Austrian Programme for Advanced Research and Technology Fellowship 1106/12 from the Austrian Academy of Sciences.

1. Bulliet R (1975) *The Camel and the Wheel* (Columbia Univ Press, New York), p 327.
2. Faye B, Grech S, Korchani T (2004) Le dromadaire, entre féralisation et intensification. *Anthropozoologica* 39(2):391–398.
3. Fitak RR, Mohandesan E, Corander J, Burger PA (2016) The de novo genome assembly and annotation of a female domestic dromedary of North African origin. *Mol Ecol Resour* 16(1):314–324.
4. Wu H, et al. (2014) Camelid genomes reveal evolution and adaptation to desert environments. *Nat Commun* 5:5188.
5. Peters J, von den Driesch A (1997) The two-humped camel (*Camelus bactrianus*): New light on its distribution, management and medical treatment in the past. *J Zool (Lond)* 242:651–679.
6. Peters J (1998) *Camelus thomasi* Pomel, 1893, a possible ancestor of the one-humped camel? *Mamm Biol* 63:372–376.
7. Grigson C (2012) Camels, copper and donkeys in the Early Iron Age of the Southern Levant: Timna revisited. *Levant* 44(1):82–100.
8. Uerpmann H, Uerpmann M (2002) The appearance of the domestic camel in southeast Arabia. *J Oman Stud* 12:235–260.
9. Grigson C (2014) The history of the camel bone dating project. *Anthropozoologica* 49(2):225–235.
10. Iamoni M (2009) The Iron Age ceramic tradition in the Gulf: A re-evaluation from the Omani perspective. *Proc Semin Arab Study* 39:223–236.
11. Rowley-Conwy P (1988) The camel in the Nile Valley: New radiocarbon accelerator (AMS) dates from Qasr Ibrim. *J Egypt Archaeol* 1:245–248.
12. Uerpmann M, Uerpman HP (2012) Archeozoology of camels in South-Eastern Arabia. *Camels in Asia and North Africa. Interdisciplinary Perspectives on Their Significance in Past and Present*, eds Knoll E, Burger P (Academy of Sciences Press, Vienna), pp 109–122.
13. Magee P (2015) When was the dromedary domesticated in the ancient Near East? *Zeitschrift für Orient-Archaeologie* 8:253–278.
14. von den Driesch A, Obermaier H (2007) The hunt for wild dromedaries during the 3rd and 2nd millennia BC on the United Arab Emirates coast. Camel bone finds from the excavations at Al Sufouh 2 Dubai, UAE. *Skeletal Series and Their Socio-Economic Context, Documenta Archaeolbiologiae*, eds Grupe G, Peters J (Marie Leidorf, Rahden, Germany), pp 133–167.
15. Frantz LAF, et al. (2015) Evidence of long-term gene flow and selection during domestication from analyses of Eurasian wild and domestic pig genomes. *Nat Genet* 47(10):1141–1148.
16. Park SD, et al. (2015) Genome sequencing of the extinct Eurasian wild aurochs, *Bos primigenius*, illuminates the phylogeography and evolution of cattle. *Genome Biol* 16:234.
17. Midant-Reynes B, Braunstein-Silvestre F (1977) Le chameau en Egypte. *Orientalia* 46: 337–362.
18. Heiss J (2012) Caravans from South Arabia: Roads and organization. *Camels in Asia and North Africa - Interdisciplinary Perspectives on Their Past and Present Significance*, eds Knoll E, Burger P (Academy of Science Press, Vienna), pp 131–139.
19. Pesce A, Grabato Pesce E (1984) *Marvel of the Desert: The Camel in Saudi Arabia* (Immel, London).
20. Troy CS, et al. (2001) Genetic evidence for Near-Eastern origins of European cattle. *Nature* 410(6832):1088–1091.
21. Hanotte O, et al. (2002) African pastoralism: Genetic imprints of origins and migrations. *Science* 296(5566):336–339.
22. Peter C, et al.; ECONOGENE Consortium (2007) Genetic diversity and subdivision of 57 European and Middle-Eastern sheep breeds. *Anim Genet* 38(1):37–44.
23. Kijas JW, et al.; International Sheep Genomics Consortium (2009) A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *PLoS One* 4(3):e4668.
24. Cañón J, et al.; ECONOGENE Consortium (2006) Geographical partitioning of goat diversity in Europe and the Middle East. *Anim Genet* 37(4):327–334.
25. Warmuth V, et al. (2012) Reconstructing the origin and spread of horse domestication in the Eurasian steppe. *Proc Natl Acad Sci USA* 109(21):8202–8206.
26. Food and Agriculture Organization of the United Nations (2011) *Molecular Genetic Characterization of Animal Genetic Resources — FAO Animal Production and Health Guidelines* (FAO, Rome).
27. Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16(1):37–48.
28. Hudson RR (1991) Gene genealogies and the coalescent process. *Oxford Surveys in Evolutionary Biology*, eds Futuyma D, Antonovics J (Oxford Univ Press, Oxford), Vol 7, pp 1–44.
29. Rosenbom S, et al. (2015) Genetic diversity of donkey populations from the putative centers of domestication. *Anim Genet* 46(1):30–36.
30. Boivin N, Fuller D (2009) Shell middens, ships and seeds: Exploring coastal subsistence, maritime trade and the dispersal of domesticates in and around the ancient Arabian Peninsula. *J World Prehist* 22:113–180.
31. Gifford-Gonzalez D, Hanotte O (2011) Domesticating animals in Africa: Implications of genetic and archaeological findings. *J World Prehist* 24:1–23.
32. Rangan H, Kull C (2010) The Indian Ocean and the making of Outback Australia. *Indian Ocean Studies — Cultural, Social and Political Perspectives*, eds Moorthy S, Jamal Y (Routledge, London), pp 45–72.
33. Hofreiter M, et al. (2015) The future of ancient DNA: Technical advances and conceptual shifts. *BioEssays* 37(3):284–293.
34. Bollongino R, et al. (2013) 2000 years of parallel societies in Stone Age Central Europe. *Science* 342(6157):479–481.
35. Fernández E, et al. (2014) Ancient DNA analysis of 8000 B.C. near eastern farmers supports an early neolithic pioneer maritime colonization of Mainland Europe through Cyprus and the Aegean Islands. *PLoS Genet* 10(6):e1004401.
36. Meiri M, et al. (2013) Ancient DNA and population turnover in southern levantine pigs—signature of the sea peoples migration? *Sci Rep* 3:3035.
37. Orlando L, et al. (2006) Geographic distribution of an extinct equid (*Equus hydruntinus*: Mammalia, Equidae) revealed by morphological and genetical analyses of fossils. *Mol Ecol* 15(8):2083–2093.
38. Vagedes K (2013) Arabia Felix — Das Land der Königin von Saba' im Spiegel von Tierknochen aus archäologischen Ausgrabungen im Jemen. *Beiträge zur Anthropologie und Paläoanatomie, Documenta Archaeobiologiae*, eds Grupe G, et al. (Marie Leidorf, Rahden, Germany), pp 163–262.
39. Bruford MW, Bradley DG, Luikart G (2003) DNA markers reveal the complexity of livestock domestication. *Nat Rev Genet* 4(11):900–910.
40. Slatkin M, Hudson RR (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* 129(2):555–562.
41. Kennedy H (2007) *The Great Arab Conquests: How the Spread of Islam Changed the World We Live in* (Da Capo, Philadelphia).
42. Gauthier-Pilters H, Dagg AI (1983) *The Camel: Its Evolution, Ecology, Behavior, and Relationship to Man* (Univ of Chicago Press, Chicago).
43. Wegmann D, Leuenberger C, Neuenschwander S, Excoffier L (2010) ABCtoolbox: A versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics* 11:116.
44. Der Sarkissian C, et al. (2015) Evolutionary genomics and conservation of the endangered Przewalski's horse. *Curr Biol* 25(19):2577–2583.
45. Schubert M, et al. (2014) Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proc Natl Acad Sci USA* 111(52):E5661–E5669.
46. Köhler-Rollefson IU (1991) *Camelus dromedarius*. *Mamm Species* 375:1–8.
47. Peck EF (1939) Salt intake in relation to contagious necrosis and arthritis of one-humped camels (*Camelus dromedarius*) in British Somaliland. *Vet Rec* 46(51): 1355–1360.
48. Hedges R, Housley R, Law I, Perry C, Gowlett J (1987) Radiocarbon dates from the Oxford AMS System: Datelist 6. *Archaeometry* 29:289–291.
49. Alley RB, Ágústsdóttir AM (2005) The 8k event: Cause and consequences of a major Holocene abrupt climate change. *Quat Sci Rev* 24(10-11):1123–1149.
50. Cheng H, et al. (2009) Timing and structure of the 8.2 kyr B.P. event inferred from δ18O records of stalagmites from China, Oman, and Brazil. *Geology* 37(11): 1007–1010.
51. Kuper R, Kröpelin S (2006) Climate-controlled Holocene occupation in the Sahara: Motor of Africa's evolution. *Science* 313(5788):803–807.
52. Gibbons A (1993) How the Akkadian Empire was hung out to dry. *Science* 261(5124):985.
53. Kaniewski D, et al. (2008) Middle East coastal ecosystem response to middle-to-late Holocene abrupt climate changes. *Proc Natl Acad Sci USA* 105(37):13941–13946.
54. Kaniewski D, et al. (2010) Late second–early first millennium BC abrupt climate changes in coastal Syria and their possible significance for the history of the Eastern Mediterranean. *Quat Res* 74(2):207–215.
55. Weiss H, et al. (1993) The genesis and collapse of third millennium north mesopotamian civilization. *Science* 261(5124):995–1004.
56. Schwartz HJ (1992) The camel (*Camelus dromedarius*) in Eastern Africa. *The One-Humped Camel (C. dromedarius) in Eastern Africa: A Pictorial Guide to Diseases, Health Care, and Management*, eds Schwartz HJ, Dioli M (Joseph Markgraf Scientific Books, Weikersheim, Germany), pp 1–7.
57. Larson G, Fuller DQ (2014) The evolution of animal domestication. *Annu Rev Ecol Evol Syst* 45:115–136.
58. Luikart G, et al. (2001) Multiple maternal origins and weak phylogeographic structure in domestic goats. *Proc Natl Acad Sci USA* 98(10):5927–5932.
59. Stadler T, Kühnert D, Bonhoeffer S, Drummond AJ (2013) Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proc Natl Acad Sci USA* 110(1):228–233.
60. Bouckaert R, et al. (2014) BEAST 2: A software platform for Bayesian evolutionary analysis. *PLOS Comput Biol* 10(4):e1003537.
61. Beaumont MA (1999) Detecting population expansion and decline using microsatellites. *Genetics* 153(4):2013–2029.
62. Storz JF, Beaumont MA (2002) Testing for genetic evidence of population expansion and contraction: An empirical analysis of microsatellite DNA variation using a hierarchical Bayesian model. *Evolution* 56(1):154–166.
63. Rooney AP, Honeycutt RL, Davis SK, Derr JN (1999) Evaluating a putative bottleneck in a population of bowhead whales from patterns of microsatellite diversity and genetic disequilibria. *J Mol Evol* 49(5):682–690.
64. Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M (2013) Robust demographic inference from genomic and SNP data. *PLoS Genet* 9(10):e1003905.
65. Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 10(3):564–567.
66. Csilléry K, François O, Blum MGB (2012) ABC: An R package for approximate Bayesian computation (ABC). *Methods Ecol Evol* 3(3):475–479.
67. Leuenberger C, Wegmann D (2010) Bayesian computation and model selection without likelihoods. *Genetics* 184(1):243–252.
68. Kass RE, Raftery AE (1995) Bayes factors. *J Am Stat Assoc* 90(430):773–795.

SUPPLEMENTARY INFORMATION

Sample collection and geographical distribution

In this study, we sampled 1,083 modern dromedaries from 21 countries across the species range (*Dataset1*). Sampling covered five defined geographical regions (1), namely: Eastern Africa (EAF; n = 170), Western and Northern Africa (WNAF; n = 233), North Arabian Peninsula (NAP; n = 349), South Arabian Peninsula (SAP; n = 181), and Southern Asia (SAS; n=150). From the 1860s until the 1920s about 20,000 camels were imported to Australia from northwest regions of the Indian subcontinent (2, 3). For this reason, and as confirmed by our population genetic structure and phylogenetic analyses (Figs. 2 and S3), we included Australian dromedaries into the Southern Asian population. In parallel, we carried out DNA analyses on seven early-domesticated dromedary specimens excavated in Apamea, Syria (Early Byzantine: 400-600 CE (Common Era)); Sagalassos, Turkey (Early Byzantine: 450-700 CE); Aqaba, Jordan (Mamluk and Ottoman Periods: 1260-1870 CE); and Tulln, Austria (2nd Turkish war *circa* 1683 CE) (Figs. 1 and S4). In addition, eight wild dromedary specimens originating from the United Arabian Emirates (UAE) were genetically investigated. The wild specimens were excavated from archaeological sites of Al-Buhais (5000-4000 BCE (Before Common Era); Umm an-Nar (Early Bronze Age: 3000-2000 BCE); Al-Sufouh (*ca.* 2400-1400 BCE); and Tell Abraq (Late Bronze – Iron Age: 1260-500 BCE) (Figs. 1 and S4). Detailed information on the modern and archaeological sampling is available in *Dataset1* and Table S4, respectively.

Classification of the wild dromedaries. Here we provide a detailed description of the archaeological context of the specimens from Al-Sufouh and Tell Abraq, UAE, in support of the classification of these findings as "wild dromedaries". Apart from their large size, there are several indications that the *Camelus* bone specimens collected at Al-Sufouh pertain to wild animals. The site's environmental setting appears very particular, as the faunal assemblage (> 80,000 specimens) is heavily dominated by marine fish and molluscs (> 120 taxa) (4). Together with other species that died naturally, they represent the natural taphocoenosis typical of littoral settings. Amongst the marine gastropods, the most frequent and edible large mud creeper (*Terebralia palustris*) is very conspicuous. It is a typical inhabitant of mangrove forests and khors (*i.e.*, tidal creek systems). The latter ecotopes are characterized by broad intertidal flats dotted with supratidal islets typically vegetated with haplophyte species. Khor environments must have been attractive to camels, considering the necessity of salt intake to their well-being (5). The location of Al-Sufouh and its faunal composition clearly illustrates that Bronze Age communities living on the coast deliberately

exploited such situations: 99.4% of the terrestrial mammalian assemblage totalling 17,911 specimens can be assigned to the one-humped camel (NISP = 17,812; minimum number of individuals: 123), whilst sheep, goat, cattle, dog, gazelle, Arabian oryx and striped hyena taken together account for only 0.6% of the assemblage. From a taxonomic viewpoint, the mammalian fauna from Al-Sufouh clearly contrasts with settlement refuse from contemporaneous Bronze Age contexts, which usually are characterized by a heavy dominance of small livestock and cattle. With less than 100 potsherds, some flints objects including an arrowhead as well as an axe and arrowhead made of copper, material culture at the site is scanty and therefore particular as well. A survey near Al-Sufouh moreover revealed the lack of contemporaneous habitation nearby, excluding the possibility that the camel remains represent carcass refuse from large livestock butchering at the settlement's periphery. The latter is also contradicted by the high frequency of long bones in the assemblage, body parts mostly removed to be processed more intensely (*e.g.*, cooking) to obtain the marrow. Interestingly, demographic profiling based on dental remains showed that only a single animal out of 29 evaluable individuals proved younger than two years. Animals older than six years accounted for a quarter of the assemblage, whilst more than two thirds represented animals aged between two and six years, the majority of these being older than four years when killed. Based on the pelvic remains (n = 70), it can be concluded that amongst sexually mature individuals, stallions numbered twice as many as mares. Sex-related demographic profiling thus suggests that meat provisioning targeted mainly young adult males, which clearly contradicts human management aiming at camels predestined for labour and/or trade. The presence of cut and chop marks on the bones shows that dismembering took place on the spot, whilst the skeletal part distribution implies that the skins with the foot bones still attached as well as particularly meat parts (*e.g.,* shoulder region) were likely removed to be processed elsewhere (4). In sum, the site's ecological setting, archaeology and faunal composition as well as the morphology, relative frequency, age and body part distribution, ratio male to female and comparably large size of the remains excavated allow concluding that during the 3rd and 2nd millennium BCE, the khor site of Al-Sufouh was a suitable place to hunt wild dromedaries, and bachelor males in particular.

Tell Abraq is a major mounded Tell site in the north of the UAE. The excavations have focused on habitation levels that stretch from *c.* 1500 BC to 500 BC on the southern side of the mound. Specimen TA618 comes from the vertical defined unit (*e.g.*, Locus) 5111, which was a part of a filling event that occurred as a large surround wall was constructed around the site. The material from within the filling event is dated by three ^{14}C-dates that were taken from the lowest, middle, and upper deposits. The three dates are statistically the same and

when combined suggest a chronology of between 1300 and 1100 BC, a chronological range which is in complete agreement with the artifacts from these levels which date to the transitional Late Bronze Age/Iron Age I period. TA623 comes from Locus 5163. This deposit is part of a collapse level of the large surround wall noted above. The deposit contains Iron Age II ceramics (1000-600 BC) with a very small quantity of Iron Age III ceramics (600-300BC). A ^{14}C sample from an ash layer below this collapse deposit has an upper limit in the eighth century BC suggesting that this collapse layer dates to the second phase of the Iron Age II period to the early Iron Age III period.

Holocene distribution of wild dromedaries in the Arabian Peninsula. While the wild specimens in our sample set come from a limited geographical distribution, large prehistoric faunal assemblages from 5000-500 BCE, sites in other parts of the Arabian Peninsula, such as coastal Yemen (6), did not yield wild dromedary remains so far, indicating that at the time people started its domestication, the native distribution of the wild ancestor of the one-humped camel may already have been limited to the Arabian Southeast coast. The poor knowledge of the Holocene distribution of wild one-humped camels on the Arabian Peninsula is a further limiting factor. To date, concentrations of bones indicative of larger camel herds have only been found in Neolithic to Bronze Age contexts on the eastern coast of the Arabian Peninsula (4, 5, 7, 8). Zooarchaeological evidence for pre-Iron Age camel population in the Southern Levant is not equivocal, since some of these specimens turned out to be intrusive based on ^{14}C-dates, whereas in other cases the stratigraphical position was considered insecure because of superposing later occupations (9, 10). Conceivably, wild dromedaries may have found suitable habitat in the interior of the Arabian Peninsula as well. However, the current state of archaeo(zoo)logical research in this vast region does not allow verifying this assumption.

DNA extraction

Modern samples. Hair, blood and saliva were collected during routine veterinary treatments. Hair samples were digested with a modified lysis buffer (11) and DNA was extracted using the DNeasy® Blood and Tissue Kit (Qiagen, Hilden, DE). Blood and saliva were blotted on FTA® Cards (Whatman Inc, New Jersey, US). DNA was extracted from blood and saliva using FTA® Purification Reagent following the manufacturer's protocol. DNA from Australian camels was provided by Peter Spencer (Murdoch University, Perth, AU).

Ancient samples. Early-domesticated and wild dromedary specimens were prepared in a dedicated and highly contained ancient DNA laboratory (Paleogenetic Core Facility, ArchaeoBioCenter, LMU Munich, Germany). DNA was extracted from bone material

following a range of standard contamination precautions. Authentication criteria for aDNA studies, such as multiple DNA extractions, independent PCR amplification and parallel extraction/ PCR controls were performed following protocols previously described (12, 13). Extractions were conducted in batches of seven samples in the presence of blank controls.

DNA genotyping and sequencing

Modern samples. From 1,083 dromedaries sampled in this study, we successfully genotyped 970 individuals with 17 microsatellite loci (Table S12) as well as 20 Bactrian camels (*C. bactrianus*) to confirm the absence of introgression between the two domestic forms. This set of markers was selected according to recommendations from the joint Food and Agricultural organization of the United Nations (FAO) and International Society for Animal Genetics (ISAG) panel on livestock genetic diversity. We selected 759 individuals for sequencing a continuous 867 bp mitochondrial fragment (nt15112 - nt15978; numbering according to GenBank Accession number NC_009849.1) spanning the end of cytochrome B (184 bp), tRNA threonine and proline (134 bp) and the beginning of the control region (MT-CR; 549 bp) until the short tandem repeat. Sequencing was performed in both directions using an in-house MegaBACE 500 sequencer (GE Healthcare) or outsourced. Mitochondrial sequences were aligned with CODONCODE ALIGNER 3.7.1 (Codon Code Corporation); unique and novel mitochondrial haplotypes were deposited in GenBank (Accession numbers JX946206-JX946273, KF719283-KF719290). The final overlapping data set of mitochondrial and nuclear markers consisted of 646 individuals (*Dataset1*).

Ancient samples. For the 15 ancient specimens, we amplified a 531 bp fragment of MT-CR and preceding tRNAs (nt15347 - nt15877) using ten overlapping primer pairs (Table S5) or genomics technology on the Illumina platform (see methods below; Table S4). Similarly to the modern haplotypes, ancient mitochondrial sequences were edited and aligned with CODONCODE ALIGNER. Ancient mitochondrial haplotypes were deposited in GenBank (Accession numbers KT334309-KT334323). In the dromedary sample from Umm-an-Nar (UN624; Table S4), the determination of the nucleotide nt15486 (G/A; numbering according to GenBank Accession number NC_009849.1) remained ambiguous despite six repetitions from two independent extractions. Assuming the most frequent nucleotide (nt15486: G), UN624 sample represented a MT-CR fragment identical to the modern haplotype *B3* (Fig. 1c). In the case of the alternative allele (nt15486: A), UN624 shared its haplotype with the specimen from Tel Abraq (TA623; Fig. 1c; Table S4).

Preparation of Illumina sequencing libraries for early-domesticated specimens. Prior to next-generation library construction, the 80 bp fragments (including the primers) of MT-CR were amplified to verify the success of DNA extraction. The library preparation, indexing and capture enrichment for six early-domesticated samples (Tulln specimen was not included; Table S4) were performed in a dedicated aDNA laboratory at the University of York (UK), following the standard contamination precautions (14). The double-stranded libraries (DSL) were produced directly from the aDNA extracts as well as the extraction blank and water control, following Meyer *et al.* (15) with minor modifications as described in Zhang *et al.* (16). Indexing PCRs were performed to create indexed libraries for the individual samples.

In-solution hybridization capture. MtDNA from the six early-domesticated dromedaries was enriched in the barcoded Illumina libraries by in-solution hybridization capture, using Mycroarray's Mybait kit according to manufacturer's instructions. We performed the capture enrichment for the entire dromedary mtDNA using 827 unique 80 bp custom designed baits that were tiled every 20 bp (4 x tiling). Following the capture procedure the entire enriched libraries were amplified in 40 µl reaction volume containing 1x AmpliTaq Gold buffer, 2mM $MgCl_2$, 0.1 mg/ml BSA, 0.25 mM dNTPs, 0.75 µM of each primary library amplification primer (IS5 and IS6) from Meyer and Kircher (17), 0.05 U AmpliTaq Gold and 20µl library template. The post-capture PCR programme consisted of initial denaturation at 94°C for 10 min followed by 10-20 cycles of 94°C for 30 sec, 60°C for 45 sec, 72°C for 45 sec and a final extension of 72°C for 5 min. Following the post capture amplification, the indexed libraries were pooled in equimolar ratio and single-end (SE) sequenced on one lane of the HiSeq2000 Illumina platform (National High-throughput DNA Sequencing Centre, University of Copenhagen, Denmark). Although we attempted to capture the entire mtDNA, here we used the 531-bp fragment of MT-CR and preceding tRNAs (nt15347 - nt15877) in accordance with the modern dromedary data set.

High-throughput data pre-processing. A total of 22,851,585 SE reads from six early-domesticated samples were trimmed for adapter and index sequences using the software CUTADAPT (18). Initially the reads shorter than 25 nucleotides were discarded to reduce the chance of spurious hits against the reference genome. The individual read collections were then assembled against the dromedary mitochondrial reference sequence (Genbank Accession number NC_009849.1) by using the BURROWS-WHEELER ALIGNMENT TOOL v.0.7.3a (19) with the parameters *-l* 1024 *-i* 0 *-o* 2 *-n* 0.03 *-t* 6, as optimized in Schubert *et al.* (20). The PCR duplicates were removed using MARKDUPLICATES as implemented in PICARD tools (www.picard.sourceforge.net). The reads were filtered for mapping quality lower than 20 and

the consensus and the SNPs were called using SAMTOOLS package v.0.1.19 (21). The filtered SNPs output of BCFTOOLS (part of SAMTOOLS) was transformed into a file for haplogroup calling. The assembly was checked by eye at each informative SNP position to identify sequencing reads conflicting with the reference sequence. Only SNPs that were covered by three unique read with different start and end positions within the region nt15347-nt15878 were accepted for downstream analysis. In general, aDNA sequences show damage patterns including fragmentation and C-T misincorporation at the 5' and G-A at the 3' end (22-24). To confirm the authenticity of the aDNA sequence data, we used the python script *mapdamage2* (http://geogenetics.ku.dk/publications/mapdamage/) (25) to identify these aDNA damage patterns in all sequences mapping to the dromedary mitochondrial genome. The damage pattern for one early-domesticated sample representing each archaeological site is illustrated in Fig S11.

Inferences of population structure and genetic distance within the modern dromedary stock

Nuclear data. Without using any prior information about the location of the samples (loc-prior), we investigated the potential number of genetic clusters (K) and whether these clustering solutions reflect geographically defined populations. We used the mixed ancestry admixture model implemented in STRUCTURE (26), which assumes that each individual derived its ancestry from 1 to K populations. Ten independent simulations for each K ($2 \leq K \leq 11$) were performed to estimate the true number of populations using 50,000 iterations after a burn-in of 10,000 Markov Chain Monte Carlo (MCMC). We determined the best clustering solution by calculating DeltaK in STRUCTURE HARVESTER (27). Results from the multiple runs were concatenated using CLUMPP (28) and displayed in R. To investigate subtle population structure that might have been masked by the high genetic distinctiveness of EAF, we excluded EAF in the analysis and re-ran STRUCTURE (Figs. 2 and S2). Despite an important amount of admixture, genetic grouping reflected populations slightly different from the one defined by the FAO (Fig. 2; *Dataset1*) (1). Despite significant positive F_{IS} values, all these clustering solutions did not seem to result from strong inbreeding (Table S1). Additionally, we estimated the degree of population structure applying a non-model based approach like the multidimensional factorial correspondence analysis (FCA) implemented in GENETIX 4.05.2 (29). No introgression with Bactrian camels was detected in our dromedary sample-set using the 17 nuclear markers.

Mitochondrial data. We applied a Bayesian Analysis of Population Structure implemented in BAPS 5.3 using the 'clustering of linked loci' model (30). We performed five independent

runs for each of the specified prior upper bound values for the numbers of clusters (*i.e.*, 2-10). BAPS revealed seven clusters (PP = 1), corresponding to the six most frequent haplotypes (*A1-2, B1-4*; Fig. 1b) and an additional cluster grouping the singletons diverging from *B1* by one mutation.

Genetic distance. Analysis of molecular variance (AMOVA), nuclear (F_{ST}) and mitochondrial (ϕ_{ST}) pairwise values were calculated with ARLEQUIN3.5 (31). Both pairwise genetic distances (F_{ST} and ϕ_{ST}) displayed EAF as the most distant population (Table S2).

Genetic diversity based on nuclear and mitochondrial DNA

Genetic diversity. For each modern dromedary population, observed (H_O) and expected (H_E) heterozygosities, total (TNA) and mean number of alleles (MNA) were calculated in MICROSATELLITE TOOLKIT (32). To compare allelic diversity between populations, we calculated allelic richness (*Ar*) for each population based on the rarefaction approach implemented in FSTAT 2.9.3.2 (33). Inbreeding coefficients (F_{IS}) were calculated using GENETIX 4.05.2. Analysis of deviations of allele frequencies from the Hardy–Weinberg equilibrium (HWE) and the null alleles (null a.) were estimated with CERVUS V. 3.0.7. With the selected set of 17 microsatellite loci we detected a total of 158 alleles among the 970 genotyped animals (Table S12). The majority of the loci exhibited substantial polymorphisms. Multiple testing for all markers considering the global dromedary population showed that only six out of the 17 loci were in Hardy–Weinberg equilibrium (HWE). After removing all EAF individuals from the analysis, which is the most distinct genetic population, we found 11 loci in HWE. When HWE was tested in just the EAF population, only one locus (YWLL59) showed significant deviation (Table S12). Mitochondrial haplotype (H_d), average number of pairwise differences (θ_π; (34)) and Watterson θ_w (based on the number of segregating sites; (35)) were computed in ARLEQUIN 3.5. The HKY+G substitution model with gamma correction (α= 0.0221) was selected as the best-fit model to the 76 unique dromedary sequences based on the Akaike Information Criterion with correction for small sample size (AICc) in the program jModelTest v.0.1.1 (36), using Maximum Likelihood (ML) tree as base tree for the likelihood calculations. As the HKY model is not implemented in the program ARLEQUIN3.5, the more inclusive Tamura-Nei (TN) model with the same parameters for ti/tv rate and a gamma correction of α = 0.02 was used. For the eight ancient wild dromedary samples the Kimura-2-Parameter (K2P) model with gamma correction of α = 0.05 was selected as the best-fit evolutionary model based on AICc. In the wild dromedary sample from Umm-an-Nar, UAE (UN624; Table S4), despite multiple repetitions, the determination of the

nucleotide nt15486 remained ambiguous (G/A). Therefore we estimated the haplotype diversity parameters for the wild dromedaries using both alternative alignments (Table S1).

Comparison of genetic diversity levels among the camelids. Overall nuclear heterozygosity (H_E = 0.630 ± 0.183) corresponded to the variation previously reported in dromedaries (37-40) and was comparable to the estimates obtained for the populations separately (Table S1). While the amount of heterozygosity and allelic richness did not differ significantly among WNAF-NAP, SWAP (Southwestern Arabian Peninsula population grouping the *Awadi* and *Awarik* Arabian camels), SEAP (Southeastern Arabian Peninsula) and SAS populations (Bonferroni corrected Wilcoxon-Rank-Sum test; *P*-value > 0.05), modern EAF camels exhibit the lowest nuclear diversity in terms of H_E (0.579 ± 0.175) and *Ar* (4.48; Table S1). Mitochondrial haplotype diversities (H_d) in dromedaries were slightly lower (0.71-0.79; overall: 0.74; Table S1) than the estimates obtained for other camelids, such as domestic Bactrian camels (0.60-0.93; overall: 0.73; (41)), vicuñas (0.72-0.90; overall: 0.76; (42)), and guanacos (0.6-0.81; overall: 0.75), with the exception of a population in Tierra del Fuego that presents an H_d of 0.36 (43). The highest H_d and θ_π (0.793 and 3.617, respectively) were measured in EAF, slightly exceeding the estimates for the populations confined to the Arabian Peninsula (Table S1). These elevated values of H_d and θ_π could, in principle, be explained by an unaccounted cryptic population structure in EAF (44), or by a large proportion of ancestral diversity in the mtDNA. While 85% (n = 646) of the investigated haplotypes pertained to H_B, and both haplogroups could not be assigned to specific geographical areas, camels in EAF exhibited a more balanced ratio between H_A (38%) and H_B (62%) (Fig. 1a). In contrast to the hypothesis of retained ancestral variation, EAF presented one of the lowest nuclear heterozygosity (H_O and H_E) among the populations tested (Table S1). These results can be interpreted as the consequence of a random founder effect followed by successive gene flow with a restricted number of sires.

Cultural context of EAF's genetic set-up. The uniqueness of the genetic setting of the EAF may result from geographical as well as cultural barriers. Indeed, prior to the introduction of one-humped camels, communities with economies based on cattle and/or small livestock pastoralism were already distributed in Eastern Africa, and in this cultural landscape, camel keeping may well have prospered especially in landscapes submarginal to the raising of cattle, sheep and goat. Moreover, many areas located around the Horn of Africa have remained infested with trypanosomes, which probably constrained the expansion of camel husbandry once established in arid East Africa. Finally, it is worth noting that the dromedary dung from the site of Qasr Ibrim in South Egypt dated around 740 BCE provides us with a *terminus ante*

quem for the species' appearance in the African continent (10). Its presence in South Egypt opens up the possibility of a third potential route of camel imports into Africa, namely by vessel across the Red Sea to coastal south-eastern Egypt or north-western Sudan (Fig. S4). However, the presence of camel remains in 1st millennium BCE sites along the western Red Sea coast of present-day Egypt and Sudan is necessary to confirm this hypothesis.

Phylogenetic relationship and divergence time estimates

The relationships between the different mitochondrial haplotypes were investigated by constructing a Median Joining Network (MJN) using the program NETWORK 4.6.1.0 (45). The MJN including the 15 ancient specimens was constructed based on 531 bp of MT-CR (nt15347-nt15877; numbering according to GenBank NC_009849.1). The dromedary's mtDNA haplotype phylogeny (Fig. S1) was inferred using the Bayesian approach implemented in the program MRBAYES V.3.2.1 (46) using two independent Markov Chain Monte Carlo (MCMC) runs of 2 million generations each. Trees were sampled every 1000 generations; the first 25% being discarded as burn-in.

Divergence time estimates. In general, estimation of divergence time requires calibration points that approximate the divergence between an outgroup and the clade of interest; usually these time inferences require the assumption of a constant mutation rate (μ) over time and across taxa. Therefore, we tested the null hypothesis of a constant evolutionary rate with the molecular clock test implemented in the program TREEPUZZLE V.5.2 (47). We used a subset of data composed of one sequence per dromedary haplotype (n = 76) to reduce the number of parameters and provided a ML tree rooted with the Bactrian camel (GenBank accession number NC_009628.1) as the starting tree (PhyML V.2.4.4; (48). We noted that the assumption of a constant rate of change among the camelids was rejected previously (49). For this reason, we performed an additional molecular clock test in PAML V.4.6 (50) using a mid-point rooted starting tree obtained from the 76 dromedary haplotypes without an out-group. We used the program FIGTREE V.1.3 (51) to define the mid-point root from a consensus unrooted ML tree built in the program PhyML. In the rooted-tree approach, we rejected the molecular clock hypothesis at a significance level of 5% (rooted tree: df = 75; $\Delta = 161.74 > \chi^2_{\alpha = 0.05} = 96.22$). However, using the unrooted tree, we failed to reject the molecular clock hypothesis (unrooted tree: df = 74; $\Delta = 62.51 < \chi^2_{\alpha = 0.05} = 95.08$). Failure to reject the molecular clock hypothesis allowed us to estimate the divergence time based on the relationship between time and genetic distance between clades ($D = 2\mu T$, where D is the sequence divergence between clades, μ is the mutation rate in units of substitutions per site

per year or generation and T is the divergence time between two clades). Computation of *D* was performed in ARLEQUIN3.5 using the net number of nucleotides between populations (D_A) and the coalescent method (*tau*), which accounts for the effect of unequal sizes of the derived populations (52). Time to the most recent common ancestor (TMRCA) between H_A and H_B was estimated as TMRCA = (distance/length of sequence)/(2μ).

We used different mutations rates (μ) to estimate the TMRCA of H_A and H_B: (*i*) the mutation rate estimate inferred from cattle MT-CR sequences and aDNA calibration (μ = $6.94x10^{-07}$ [$4.52x10^{-07}$, $9.35x10^{-07}$] sub/site/y; (53); and (*ii*) the two mutation rates deduced from the estimated *Camelus* species split ($4.4x10^{6}$ [$1.9x10^{6}$ - $7.2x10^{6}$] ya; (54) and the distances D_A (268.25 substitutions for the 867-bp fragment) and *tau* (258.08 sub/867-bp) measured between the Bactrian and the dromedary sequences (μ_{DA} = $3.516x10^{-08}$ [$2.149x10^{-08}$ - $8.142x10^{-08}$] sub/site/y; μ_{tau} = $3.383x10^{-08}$ [$2.067x10^{-08}$ - $7.833x10^{-08}$] sub/site/y). The estimated D_A between H_A and H_B (10.90973 sub/867-bp) translates into a TMRCA comprised between 6,700 and 304,000 ya, while the distance *tau* results into TMRCA estimates ranging from 5,700 and 260,000 ya (Table S6). While the TMRCA inferred from the divergence between Bactrian and dromedary camels violates the molecular clock assumption, the TMRCA deduced from the mutation rate estimate based on the bovine MT-CR ($TMRCA_{DA}$ = 9,066 [6,729 – 13,920] ya; $TMRCA_{tau}$ = 7,761 [5,761 – 11,917] ya; Table S6) are likely to be much shorter than real TMRCA due to the fact that they focused on the non-coding part of the mitochondrial genome, only.

Bayesian inferences of divergence time. In addition, we estimated the TMCRA of H_A and H_B using the MT-CR (448 bp) fragment from modern and ancient dromedary samples with BEAST 2.2.0 (55). We dated the tips with the archaeological dates of the ancient samples (Table S4). We applied a relaxed lognormal clock, coalescent Bayesian skyline and TN93 nucleotide substitution model and ran MCMC for 100,000,000 generations with initial 1,000,000 steps discarded as burn-in. Examination of the autocorrelation times of the MCMC plots indicated that runs were optimal, as was revealed by the convergence of the posterior distribution with adequate ESS (> 100) for all parameters. We estimated the parameter root height, which represents the total height of the tree about 9 kya (TMRCA = 8,932 y [95% HPD 7,000 - 14,191]). We applied a second approach in BEAST 2.2.0 using the 859 bp fragment from 759 modern sequences and the substitution rate μ = $1.232x10^{-06}$ sub/site/y [95% HPD: $4.4353x10^{-07}$, $2.2132x10^{-06}$], which was estimated previously from the tip dates using the 448 bp CR fragment including all ancient samples. This second approach delivered a TMRCA of 6,094 ya [95% HPD: 5,829 - 6,374]. Overall and independently of the approach,

with a minimum TMCRA dated to 5,700 ya (Table S6) it seems unlikely that the actual divergence between H_A and H_B happened subsequently or concordantly to domestication.

Demographic history analysis

For each pre-defined population, evaluation of possible population expansions were assessed using the neutrality tests of Tajima's *D* (56) and Fu's F_S, which have been shown to be especially sensitive to population expansion (57) as implemented in the program ARLEQUIN3.5.

Estimating population expansion. Assuming a constant mutation rate, pairwise differences of a population that underwent sudden expansion in the context of domestication are distributed according a Poisson distribution (58). To test the goodness of fit of a Poisson distribution to the observed pairwise differences between the modern haplotypes, we compared the empirical log-likelihood values with the ones obtained for 1,000 simulated Poisson distributions (with parameter of the simulated distribution equals to $\lambda_{empirical}$) using a chi-square test. In the cases where the data fit a Poisson distribution, the single parameter of the empirical Poisson distribution, lambda (λ), is an estimate of the rate of mutation (μ) occurring in a period of time (*t* in generation); consequently, we inferred μ with the formula $\lambda = 2\mu T$. Minimal mitochondrial diversity required in the initial pool of domesticated camels was inferred by estimating μ for different evolutionary scenarios. From the modern MJN (Fig. 1b) and the phylogram (Fig. S1) we distinguished six haplogroups corresponding to the BAPS clustering solution of K = 6, distributed into two haplogroups H_A and H_B. As the divergence between H_A and H_B likely predated domestication (see TMRCA calculations above), at least one haplotype representing of each of the two haplogroups should have been present in the initial domestic pool. Within the ancestral lineage H_A, two haplogroups were centered on the haplotypes *A1* and *A2*; while in H_B four grouped around the haplotypes *B1*, *B2*, *B3* and *B4* (Fig. 1b). The topology of these six haplotypes at high frequency, from which singletons dispersed in a star-like shape of one- or two-step mutations, and the uniformity of the external branch lengths (Fig. S1) suggest population expansion. We first postulated that the entire diversity within the lineages H_A and H_B was generated since the time of domestication (strong bottleneck). We thus tested whether the number of substitutions on the external branches followed a Poisson distribution (58). Under the assumption that the initial pool of domesticated camels consisted of individuals representative of the most frequent mitochondrial haplotypes, *A1* for H_A and *B1* for H_B, the distribution of the substitutions was not significantly different from a Poisson distribution ($P > 0.05$). Similarly the distribution of the substitutions of the second scenario assuming one unique wild source population consisting of all the six haplotypes, did not differ

significantly from a Poisson distribution. Using the relationship between λ (unique parameter of the Poisson distribution), mutation rate and time ($\lambda = 2\mu t$) and assuming that domestication commenced t = 600 generations ago (*ca.* 3000 y with a generation time of 5 y), we estimated the different μ. Our first scenario, involving the contribution of the six most frequent haplotypes to the initial domestic pool, yielded estimation of $\mu_{(A1+A2+B1+B2+B3+B4)} = 0.08$ sub/site/Myr. Yet, the second scenario, where the minimum initial domestication pool was assumed to be formed of the two most frequent haplotypes within each haplogroup (*A1* and *B1*), resulted in an estimate of $\mu_{(A1+B1)} = 0.22$ sub/site/Myr. Whereas both estimates were in the same order of magnitude as μ calculated from coding parts of mammalian mitochondrial genomes (59), they remained slower than that of the one calculated for the MT-CR in cattle (*e.g.*, 0.694 sub/site/Myr; (53). Nonetheless, the first scenario yielded similar estimate, $\mu_{(A1+A2+B1+B2+B3+B4)}$, that the one calculated for the 867 bp fragment using the TMRCA between the dromedary and Bactrian camels ($\mu_{DA} = 0.035$ [0.021 – 0.081] sub/site/Myr; $\mu_{tau} =$ 0.034 [0.021 - 0.078] sub/site/Myr; see above method in '*Phylogenetic relationship and divergence time estimates*'). As we could not argue against the contribution of at least six ancestral female lineages to the current gene pool and that comparable amount of maternal diversity was observed in goats, cattle and donkeys (60-64), we presumed the first scenario as the most plausible. Hence in this perspective, the initial diversity was remarkably high relative to the distribution of the wild one-humped camel on the coastal Arabian Peninsula and to the brief co-existence (*e.g.*, less than two millennia) of wild and early-domesticated individuals postulated on the basis of the current archaeofaunal record of that region (4, 8). High mitochondrial DNA diversity in ungulate species such as goat, horse and cattle has been interpreted as a sign of recurrent introgression during the early stage of domestication (60, 61, 65).

Coalescent simulations to infer demographic changes

Coalescent simulations with mitochondrial data. To assess historical population demographic dynamics, we used the MT-CR (448 bp) on the combined modern, early-domesticated and wild dromedary data set consisting of 774 individual (not collapsed) haplotypes. We applied the birth-death skyline plot serial model implemented in BEAST 2.2.0, which accounts for serial samples taken at different time points (55, 66). We used the archaeological dating of the extinct wild and early-domesticated samples (Table S4) to date the tips and estimated the substitution rate with a relaxed lognormal clock model from the combined ancient and modern samples. We used the resulting substitution rate $\mu = 1.232 \times 10^{-06}$ sub/site/y [95% HPD: 4.435×10^{-07}, 2.213×10^{-06}] to compute Bayesian Skyline plots (BSP) for the domestic

and the wild dromedaries separately under the stepwise constant function. To infer ancestral gene trees, we used the TN93 substitution model. Each MCMC sample was based on a run of 200 million generations, sampled every 1,000 generations with the initial 20 million generations discarded as burn-in. Runs were repeated twice using different random number seeds to confirm consistency of the generated skyline plot and refine skyline parameters for acceptance of effective sample sizes (ESS). Convergence of the chains to the likelihood stationary distribution was systematically confirmed by visual inspection of the plotted posterior estimates following analysis and visualization with the program TRACER V.1.5.1 (http://beast.bio.ed.ac.uk/Tracer). Examination of the autocorrelation times of the MCMC plots indicated that runs were optimal, as was revealed by the convergence of the posterior distribution with adequate ESS (> 100) for all parameters. The "tree" and "log" files from the two independent runs were combined using LOGCOMBINER V.1.6.1 and the combined files were used to generate the BSP for each dataset. The final BSPs were displayed in R using the output values imported from TRACER.

Within the constraints of the number of ancient wild camels (n = 8) examined, we see a potential signal of a sudden population decline around 6,000 - 8,000 ya, followed by a slow expansion (Fig. S5) until they disappeared *ca.* 2,000 ya (67). However, the low sample sizes can lead to unreliable BSPs (68) and we observed large Bayesian credible intervals (CIs; Fig. S5), especially towards recent times, which is compatible with many possible demographic trajectories, including a simple flat line (no demographic change). Although we tried to exclude a possible false signal in the ancient wild camels with the re-analysis of 100 randomly down-sampled (n = 8) modern datasets, of which none resulted in a bottleneck, we acknowledge the limitation of this analysis with so few samples.

Coalescent simulations with nuclear data. For the inference of more recent demographic history, we used coalescent simulations implemented in MSVAR 1.3 (69, 70) with the microsatellite dataset. The model assumes a single stable ancestral population N_1 at some time t_1 ago that experienced a demographic change (bottleneck or expansion) starting at time t and subsequently changed exponentially in size to the current population N_0. We simulated two different demographic scenarios by choosing (*i*) larger prior distribution values for the current population size N_0 than the ancestral N_1 (expansion) and (*ii*) vice versa, larger priors for N_1 than N_0 (decline or bottleneck). To assess the independency of the posterior estimates for the parameters N_0, N_1 and t, for each scenario we tested various prior distributions (Fig. S8). In absence of a species-specific microsatellite mutation rate in camels we choose an average mammalian mutation rate (71, 72) of $1x10^{-4}$ sub/site/y allowing a rate variation between 10^{-3} and 10^{-5}. As the method convergences slower if the sample size is large (more than 200

chromosomes per locus) we sub-sampled 100 individuals from the five populations, respectively. We run three coalescent simulations for each population with $2.5x10^9$ iterations of the MCMC algorithm discarding the first 20% as burn-in. Convergence of the chains from each population simulated with four different priors, respectively, were assessed with the Gelman and Rubin's diagnostic (73) implemented in the R package *boa* (74). Gelman-Rubin's convergence tests of the MCMC algorithm for the independent runs and each variable resulted in values below the threshold of 1.1 (75). However, convergence could not be reached after $2.5x10^9$ MCMC iterations for the parameters N_0 and t in the SEAP and SAS groups. Although the current N_e and the timing of the bottleneck could not be estimated in the SEAP and SAS populations (Gelman-Rubin diagnostic values >1.1 (76)), we can assume a comparable demographic history based on the similarity of genetic makeup of these groups.

The oldest bottleneck, detected from the maternal sequences of the wild dromedaries (Fig. S5) and the nuclear polymorphisms of the EAF (Fig. S8; Table S9), possibly relates to the abrupt worldwide climate event that occurred *ca.* 8,200 ya when the glacial lake Agassiz drained into the northern Atlantic ocean causing a general cooling and drying of the northern hemisphere that lasted between two and four centuries (77, 78). It could also correspond to climatic change observed in the Eastern Sahara around 7,300 ya, when desiccation resulted in southward shifting of the desert margin and the sub-Saharan spreading of pastoralism (79). Despite rather large CIs (Table S9), the time point estimates of the bottleneck detected from the nuclear polymorphisms of the other populations (*e.g.*, combined WNAF-NAP, SAP and SAS) coincided with the archeological-deduced onset of the domestication process, but as well concurred with two additional abrupt climatic events occurring between 4,200 and 2,500 ya. The first of these two events (*ca*. 4,200 to 3,900 ya) was marked by an increase in wind circulation and aridification of the Middle East and was a potential cause for the synchronous collapse of the Akkadian empire and populations in neighbouring regions (80, 81). The second event (*ca*. 3,500 to 2,500 ya) had a stronger effect on the region and lasted much longer, causing region-wide crop failures marking the collapse of the Ugarit kingdom (82, 83) and ending the late Bronze Age. By the time cultural control over the wild one-humped dromedary was initiated, its native distribution may already have become increasingly disjointed due to anthropogenic activities.

Approximate Bayesian Computation (ABC) inferences of four alternative domestication scenarios

To test the hypotheses of one independent or multiple domestication scenarios we studied the genealogical history of the dromedary populations using ABCTOOLBOX (84) on a combined

mitochondrial and microsatellite dataset (n = 642). We followed the population structure observed in STRUCTURE (best-fitting K = 2; Fig. S2) and restricted the analysis to two populations (EAF *vs.* WNAF_NAR_SAR_SAS_combined) to avoid overparametrization of the models tested (85). We acknowledge that we might capture only a simplified version of the real demographic history while we reduced parameters to fit the four categories of events: population divergence, discrete change of effective population size, admixture and sampling. We tested four scenarios, in which we hypothesized (*i*) one domestication, (*ii*) one domestication with consecutive admixture from a wild unsampled source population, (*iii*) two independent domestications at variable time points, and (*iv*) two domestications at serial time points (Fig. S6). For each scenario we simulated a large number ($1x10^6$) of datasets under the coalescent model drawing their parameter values from a prior distribution range (Table S10). We estimated the following historic and demographic parameters: effective population size of the sampled modern populations (N1 and N2), the wild unsampled source populations (NW1, NW2), and the ancestral population (NA); and the time of domestication (t_dom), admixture (t_adm) between a wild and a domestic population, and divergence (t_div) between the two unsampled wild populations (Fig. S6). The genetic variation within and between populations was summarized in 15 summary statistics (Table S11): as population specific summary statistics we used the mean number of alleles (Obs0_K) and mean genetic diversity (Obs0_H) across loci, the mean of pairwise differences (Obs1_Pi), the segregating sites (Obs1_S) and the private segregating sites (Obs1_PrS); for population pairwise comparisons we used the mean total genetic diversity (Obs0_tot_H), the pairwise F_{ST} for microsatellite (Obs0_FST) (86), mean number of alleles (Obs0_tot_K) (87) across loci for two populations, number of haplotypes (Obs1_tot_K), and the mean of pairwise differences (Obs1_PI_2_1). The mutational model for the microsatellites was the Strict Stepwise Mutation model allowing variation in mutation rate across loci following a gamma distribution. For the mtDNA we used the default finite site model. We assessed if the observed summary statistics occurred within the 95% quantiles of the simulated summary statistics by generating a density distribution for each statistic and calculating the 2.5 and 97.5 percentile of the distribution (Fig. S12). Correlations between the summary statistics and their respective significances were estimated using Spearman's rho statistics and the function *cor.test* in R. Graphical representation of these results was obtained using a modified script of the *plotcorr* function from the *ellipse* R package (Fig. S9). Summary statistics with a high correlation (Obs1_S_1/_2, Obs0_tot_H; Table S11) were removed from the analysis of the final dataset.

We run a cross-validation for model selection using the function *cv4postpr* in the R package *abc* (88) to evaluate if the twelve summary statistics provide enough statistical power to

discriminate the four scenarios. We used the summary statistics of the 5000 simulations closest to the observed data and randomly selected 1000 as sample (nval) for cross-validation with a single tolerance rate (tols) of 0.05 and the method 'mnlogistic' based on multinomial logistic regression. The model misclassification plots showing the clear separation between the four scenarios are displayed in Fig. S10.

The posterior distribution of each parameter was performed using the GLM approach implemented in ABCTOOLBOX (89). To identify the best-fitting scenario, the marginal distributions of each scenario were used to calculate the scenario's probability (Table S7), which corresponds to the proportion of the retained simulations that presented a lower or equal likelihood under the inferred GLM as compared to the observed data (89). We also used the marginal densities to calculate Bayes factors (BF) for each pairwise comparison between scenarios (84). The highest support was estimated for scenario *(ii)* using both the scenario probabilities and the Bayes Factors (Table S8). The resulting posterior distributions of the parameter values of the four scenarios are presented in Figure S7 and Table S13.

SUPPLEMENTARY TABLES

Table S1. Genetic diversity of the modern domestic and wild dromedary populations

Sample source	No. of samples	Microsatellite (17 loci) Genetic diversity H_E (*SD*)	H_O (*SD*)	Allelic diversity TNA	MNA (*SD*)	*Ar*	F_{IS}
EAF	160	0.579 (0.175)	0.532 (0.179)	97	5.71 (3.48)	4.882	0.082***
WNAF-NAP	524	0.634 (0.175)	0.602 (0.169)	146	8.59 (5.36)	6.469	0.051***
SWAP	64	0.562 (0.197)	0.513 (0.181)	90	5.29 (3.46)	5.259	0.089***
SEAP	77	0.620 (0.214)	0.578 (0.208)	103	6.06 (3.42)	5.899	0.067***
SAS	145	0.617 (0.200)	0.559 (0.187)	121	7.12 (4.87)	6.375	0.092***
Total	**970**	**0.630 (0.183)**	**0.577 (0.170)**	**158**	**9.29 (5.45)**	**6.468**	**0.085****

Sample source	No. of samples	mtDNA (867 bp) Variable sites	No. of haplotypes	H_d (*SD*)	θ_π (*SD*)	θ_w (*SD*)	Tajima's *D*	Fu's F_S
EAF	74	16	15	0.793 (0.028)	3.617 (2.056)	3.282 (1.150)	0.297 ns	-1.77 ns
WNAF-NAP	410	33	42	0.712 (0.023)	2.000 (1.249)	5.006 (1.282)	-1.617 *	-26.67***
SAP (SWAP-SEAP)	150	23	32	0.764 (0.034)	2.595 (1.547)	4.119 (1.244)	-1.042 ns	-20.81***
SAS	125	19	22	0.711 (0.042)	2.012 (1.263)	3.518 (1.129)	-1.197 ns	-11.76**
Total	**759**	**47**	**76**	**0.743 (0.016)**	**2.353 (1.421)**	**6.520 (1.477)**	**-1.712** **	**-87.48****

Sample source	No. of samples	mtDNA (531 bp) Variable sites	No. of haplotypes	H_d (*SD*)	θ_π (*SD*)	θ_w (*SD*)	Tajima's *D*	Fu's F_S
EAF	74	11	12	0.767 (0.029)	2.120 (1.324)	2.257 (0.872)	-0.165 ns	-2.45 ns
WNAF-NAP	410	23	35	0.700 (0.023)	1.430 (0.874)	3.489 (0.976)	-1.510 *	-28.38 ***
SAP (SWAP-SEAP)	150	18	28	0.754 (0.035)	1.806 (1.160)	3.223 (1.033)	-1.193 *	-21.63 ***
SAS	125	13	17	0.690 (0.043)	1.485 (1.000)	2.407 (0.854)	-1.003 ns	-8.73 *
Total	**759**	**33**	**59**	**0.731 (0.016)**	**1.621 (1.062)**	**4.578 (1.122)**	**-1.661** **	**-27.01** ***

Sample source	No. of samples	aDNA (531 bp) Variable sites	No. of haplotypes	H_d (*SD*)	θ_π (*SD*)	θ_w (*SD*)	Tajima's *D*	Fu's F_S
Wild^G	8	3	7	0.964 (0.077)	1.679 (1.096)	1.157 (0.781)	1.855 ns	-5.16 **
Wild^A	8	3	6	0.929 (0.084)	1.643 (1.078)	1.157 (0.781)	1.728 ns	-3.18 **

According to the clustering solutions (Fig. 2 and *SI Appendix,* Fig. S3), dromedaries from Western and Northern Africa (WNAF) were grouped with the ones originating from North Arabian Peninsula (NAP) for calculation of the nuclear and mitochondrial diversity estimators. Due to the sub-structure existing at the nuclear level (Fig. 2) in South Arabian Peninsula (SAP), individuals from this region have been divided accordingly into Southwestern (SWAP) and Southeastern (SEAP) populations. Australian individuals were included in the Southern Asian (SAS) group due to their shared ancestry (Fig. 1 and 2; Table S4). EAF: Eastern Africa population; H_E: Expected heterozygote frequency; H_O: Observed heterozygote frequency; *TNA*: total number of alleles; *MNA*: mean number of alleles per locus; *Ar*: allelic richness per locus calculated for a population based on minimum sample size of 60 diploid individuals; F_{IS}: Inbreeding coefficient; H_d: haplotype diversity; θ_π: theta estimator based on the mean number of nucleotide differences; θ_w: theta estimator based on the segregating sites; *SD*: standard deviation values. Significance: **P* <0.1; ***P* <0.01; ****P* <0.001; ns: not significant. G/A Genetic diversity estimators calculated with the two potential haplotypes of UN624 at nt15486 (either allele G or A; *SI Appendix*).

Table S2. Population pairwise distances based on the 867 bp mtDNA sequences (ϕ_{ST}; above diagonal) and 17 microsatellite loci (F_{ST}; below diagonal).

ϕ_{ST} / F_{ST}	EAF	WNAF- NAP	SWAP	SEAP	SAS
EAF	-	0.164^{***}	0.077^{***}	0.081^{***}	0.155^{***}
WNAF-NAP	0.040^{***}	-	0.004^{ns}	0.014^{***}	0.000^{ns}
SWAP	0.070^{***}	0.035^{***}	-	-0.006^{ns}	0.006^{ns}
SEAP	0.051^{***}	0.013^{***}	0.033^{***}	-	0.012^{ns}
SAS	0.060^{***}	0.013^{***}	0.037^{***}	0.018^{***}	-

*: *P*-value < 0.1; ***: *P*-value < 0.001; ns: not significant

Table S3. Genetic diversity of the Southern Asian dromedaries inferred from mitochondrial and 17 microsatellites data.

	mtDNA (867 bp)						
Populations	No. drom.	Haplotypes	Var. sites	H_d	θ_π	θ_w	
Australia (AU)	38	11	13	0.814 (0.052)	3.153 (1.670)	3.094 (1.216)	
Iran (IR)	30	12	15	0.717 (0.090)	3.396 (1.788)	3.786 (1.484)	
Pakistan (PK)	38	7	8	0.588 (0.088)	1.793 (1.059)	1.904 (0.848)	
India (BD)	19	3	2	0.632 (0.073)	0.804 (0.606)	0.572 (0.427)	
Southern Asia	125	22	19	0.711 (0.042)	2.502 (1.358)	3.518 (1.129)	
	Microsatellite (17 loci)						
Populations	No. drom.	TNA	MNA	*Ar*	H_E	H_O	F_{IS}
Australia (AU)	59	99	5.82 (3.54)	1.60	0.604 (0.206)	0.544 (0.183)	0.100
Iran (IR)	28	98	5.76 (3.60)	1.61	0.616 (0.191)	0.574 (0.194)	0.070
Pakistan (PK)	39	99	5.82 (3.78)	1.62	0.617 (0.193)	0.561 (0.209)	0.092
India (BD)	19	81	4.76 (2.44)	1.57	0.574 (0.257)	0.588 (0.273)	-0.025
Southern Asia	145	121	7.12 (4.87)	1.62	0.617 (0.200)	0.559 (0.187)	0.092

No. drom.: sample size; H_d: haplotype diversity; θ_π: theta estimator based on the mean number of nucleotide differences; θ_w: theta estimator based on the segregating sites; TNA: total number of alleles; MNA: mean number of alleles per locus; *Ar*: allelic richness per locus calculated for a population based on minimum sample size of one diploid individual; H_E: expected heterozygote frequency; H_O: observed heterozygote frequency; F_{IS}: Inbreeding coefficient. Standard deviation values are indicated between brackets.

Table S4. Geographical locations and archaeological information of the early-domestic and wild dromedary specimens successfully amplified for the 531-bp MT-CR fragment.

Sample ID	Site	Sector-level	Date (archeological period)	Sequencing Technology	GenBank Accession Number
Wild					
AB620	Al-Buhais (BHS 18), UAE	-	5000-4000 BCE	Sanger sequencing	KT334320
UN624	Umm an-Nar, UAE	-	Early Bronze Age (3000-2000 BCE)	Sanger sequencing	KT334323
AS1	Al Sufouh 2, UAE	A-C7-lvl7	2400-1400 BCE	Sanger sequencing	KT334316
AS13	Al Sufouh 2, UAE	A-C7-lvl3	2400-1400 BCE	Sanger sequencing	KT334317
AS34	Al Sufouh 2, UAE	A-1N-lvl5	2400-1400 BCE	Sanger sequencing	KT334318
AS36	Al Sufouh 2, UAE	A-C7-lvl4	2400-1400 BCE	Sanger sequencing	KT334319
TA618	Tell Abraq, UAE	Locus 5111	Transition Late Bronze - Early Iron Age (*ca.* 1260 - 1130 BCE)	Sanger sequencing	KT334321
TA623	Tell Abraq, UAE	Locus 5163	Iron Age II (*ca.* 800 - 500 BCE)	Sanger sequencing	KT334322
Early-domestic					
AP2	Apamea, Syria	G357	Early Byzantine (400-600 CE)	Ds-DNA library In-solution capture NGS sequencing	KT334309
SG1	Sagalassos, Turkey	92N7	Early Byzantine (450-550 CE)	Ds-DNA library In-solution capture NGS sequencing	KT334313
SG2	Sagalassos, Turkey	98PQ35	Early Byzantine (450-700 CE)	Ds-DNA library In-solution capture NGS sequencing	KT334314
AQ30	Aqaba, Jordan	D3-14	Mamluk (1260-1456 CE)	Ds-DNA library In-solution capture NGS sequencing	KT334310
AQ34	Aqaba, Jordan	D3-30	Mamluk (1260-1456 CE)	Ds-DNA library In-solution capture NGS sequencing	KT334311
AQ40	Aqaba, Jordan	D5A-1	Ottoman (1456-1870 CE)	Ds-DNA library In-solution capture NGS sequencing	KT334312
TU	Tulln an der Donau, Austria	SE 6684	2nd Ottoman-Habsburg war (*ca.* 1683)	Sanger sequencing	KT334315

Table S5. Primer pairs used to amplify the 531-bp MT-CR fragment (nt15347-15877) in ancient dromedary samples.

Primer Name	Sequence (5' to 3')	Tm ° C	Product size (bp)
Ancient_mtDNA_F1	RCCACACCCTCCCTAAGACT	60.51	92
Ancient_mtDNA_R1	CGGAGGTCAGGGGGTAGT	59.91	
Ancient_mtDNA_F2	CACCCAAAGCTGGAATTCTT	59.17	100
Ancient_mtDNA_R2	GGCATGAYATGTGGTTTTTAG	58.01	
Ancient_mtDNA_F3	ACGGCAATAGCCCTTGAGTA	59.73	97
Ancient_mtDNA_R3	CAACGCGTGCTGTGACAT	60.50	
Ancient_mtDNA_F4	GCGTRCATGAAACCTCAATA	59.69	90
Ancient_mtDNA_R4	TATATGCATGGGGCAAACAA	59.78	
Ancient_mtDNA_F5	TGTTTGCCCCATGCATATAA	59.78	85
Ancient_mtDNA_R5	TGCGTATTGACTGGAAATGA	57.70	
Ancient_mtDNA_F6	CRCATTATGTCAAATCATTTCC	59.33	99
Ancient_mtDNA_R6	CTGCYRAGCGGGTTGATGAT	60.24	
Ancient_mtDNA_F7	CCGCGTGAAATCATCAACC	62.41	94
Ancient_mtDNA_R7	TGCCTGGTAAAGTTCCGGTAT	60.73	
Ancient_mtDNA_F8	CATCCATTGTGGGGGTTTCT	61.90	86
Ancient_mtDNA_R8	AGTGTGGGCGATTTTAGGTG	59.99	
Ancient_mtDNA_F9	GGACCATCTCACCTAAAATCG	58.52	80
Ancient_mtDNA_R9	GGCATGGGCTGATTAGTCATT	61.22	
Ancient_mtDNA_F10	GGCATCTGGTTCTTACTTCAGG	60.13	100
Ancient_mtDNA_R10	GGCATGGGCTGATTAGTCATT	61.22	

Primer sequences were designed using *Camelus dromedarius* mitogenome as reference (accession number: NC_009849.1)

Table S6. Estimates of the time to the most recent common ancestor (TMRCA) of the two haplogroups H_A and H_B inferred with different approaches.

***(i)* TMRCA calculation based on the mutation rate inferred from cattle MT-CR (μ = 6.94x10-07 [4.52x10-07, 9.35x10-07] sub/site/y; (53))**

tau (H_A-H_B) = 9.34 sub /867bp

μ (MT-CR cattle)	TMRCA
4.52E-07	11,917
6.94E-07	**7,761**
9.35E-07	5,761

D_A (H_A-H_B) = 10.91 sub /867bp

μ (MT-CR cattle)	TMRCA
4.52E-07	13,920
6.94E-07	**9,066**
9.35E-07	6,729

***(ii)* TMRCA calculation based on the genetic distance between dromedary - Bactrian camel for the 867bp fragment**

Assuming a divergence time between dromedary - Bactrian camel to $4.4x10^6$ [$1.9x10^6$ - $7.2x10^6$] ya (54)

(iia) Drom-Bac._D_A= 268.25 sub /867bp

tau (H_A-H_B) = 9.34 sub /867bp

μ (Drom-Bac._D_A)	TMRCA
2.15E-08	250,692
3.52E-08	**153,200**
8.14E-08	66,155

D_A (H_A-H_B) = 10.91 sub /867bp

μ (Drom-Bac._D_A)	TMRCA
2.15E-08	292,824
3.52E-08	**178,948**
8.14E-08	77,273

(iib) Drom-Bac._*tau*= 258.08 sub /867bp

tau (H_A-H_B) = 9.34 sub /867bp

μ (Drom-Bac._*tau*)	TMRCA
2.07E-08	260,571
3.38E-08	**159,237**
7.83E-08	68,762

D_A (H_A-H_B) = 10.91 sub /867bp

μ (Drom-Bac._*tau*)	TMRCA
2.07E-08	304,364
3.38E-08	**185,999**
7.83E-08	80,318

***(iii)* Bayesian inference of TMRCA by incorporating tip date information from ancient samples on the 448 bp MT-CR fragment in BEAST 2.2.0**

TMRCA	95% HPD
8,933	7,000 – 14,191

***(iv)* Bayesian inference of TMRCA using the substitution rate μ = 1.232x10-6 sub/site/y [95% HPD: 4.4353x10 - 7, 2.2132x10-06], which was estimated previously from the tip dates using the 448 bp CR fragment including all ancient samples**

TMRCA	95% HPD
6,094	5,829 – 6,374

Table S7. Marginal densities and posterior probabilities (PP) for scenarios *(i)* - *(iv)*.

Scenario	$\log_{10}$ (marginal density)	PP
i	-4.119	0.013
ii	5.993	0.212
iii	-17.000	0.002
iv	4.193	0.035

Table S8. Log_{10} Bayes factors (BF) calculated for all pairwise comparisons between scenarios *(i)* - *(iv)*.

Scenario	*i*	*ii*	*iii*	*iv*
i		10.1122	-12.8808	8.3123
ii	-10.1122		-22.9930	-1.7999
iii	12.8808	22.9930		21.1931
iv	-8.3123	1.7999	-21.1931	

The scenario given on the header row is the hypothesis tested with the scenario in the first column being the alternative hypothesis. Shaded cells indicate BF values >3 ($\log_{10}(3) = \sim 0.477$).

Table S9. Current (N_0), ancestral (N_1) effective population size and time since bottleneck (t) and their 95% highest probability density (HPD) intervals inferred in MSVAR and *boa* R Package.

Population	Run	N_0	HPD	N_1	HPD	t	HPD
EAF	1	407	93-1,778	19,055	3,236-114,815	8,710	1,072-58,884
	2	347	81-1,514	16,982	3,162-91,201	7,762	1,096-48,978
	3	437	102-1,778	16,982	2,951-100,000	8,710	1,380-58,884
	4	398	95-1,585	18,621	3,388-97,724	9,333	1,585-57,544
All-excl.EAF	1	476	78-2,407	14,594	3,043-66,755	4,270	450-32,032
	2	427	67-2,347	11,646	2,459-54,746	3,475	335-30,576
	3	654	123-2,954	17,197	3,491-83,685	7,177	871-56,087
	4	584	123-2,444	13,759	2,874-62,912	5,549	834-34,310
WNAF-NAP	1	389	48-2,344	7,413	1,318-38,019	3,311	309-30,903
	2	295	21-2,570	6,607	1,202-37,154	2,344	63-52,481
	3	562	93-2,691	8,913	1,585-54,954	5,129	427-60,256
	4	427	76-2,399	7,413	1,259-4,1687	3,802	407-36,308
SEAP	1	na	na	8,128	1,479-39,811	na	na
	2	na	na	5,370	933-28,184	na	na
	3	na	na	6,918	1,259-32,359	na	na
	4	na	na	6,918	1,288-38,019	na	na
SWAP	1	166	18-1,230	10,000	1,584-58,884	3,715	324-37,154
	2	162	24-977	8,128	1,318-50,119	3,631	457-30,200
	3	112	2-1,175	10,233	1,698-61,660	2,570	45-33,884
	4	91	5-912	6,607	977-38,019	1,660	62-24,547
SAS	1	na	na	7,244	1,445-39,811	na	na
	2	na	na	7,079	1,259-37,154	na	na
	3	na	na	8,710	1,585-46,774	na	na
	4	na	na	7,244	1,259-3,5481	na	na

N_0 = current effective population size (N_e); N_1 = ancestral N_e; t = time in years since demographic event started; HPD = 95% highest probability density interval; na = not applicable. While Gelman-Rubin's diagnostic test indicated reasonable convergence for WNAF-NAP and SWAP populations (values <1.1; (75)), convergence could not be reached after $2.5*10^9$ MCMC iterations for the parameters N_0 and t in the SEAP and SAS groups. EAF = East Africa; All-excl.EAF = all populations combined without EAF; WNAF-NAP = Western North Africa and Northern Arabian Peninsula; SEAP = Southeast Arabian Peninsula, SWAP = Southwest Arabian Peninsula; SAS = Southeast Asia including Australia.

Table S10. Prior boundaries of the parameters used to generate the four scenarios.

	***Scenario* (*i*)**		***Scenario* (*ii*)**		***Scenario* (*iii*)**		***Scenario* (*iv*)**	
Parameter	*Min*	*Max*	*Min*	*Max*	*Min*	*Max*	*Min*	*Max*
DNA_MUTATION	1.00E-08	1.00E-06	1.00E-08	1.00E-06	1.00E-08	1.00E-06	1.00E-08	1.00E-06
MSAT_MUTATION	1.00E-05	1.00E-02	1.00E-05	1.00E-02	1.00E-05	1.00E-02	1.00E-05	1.00E-02
GAMMA	8	15	8	15	8	15	8	15
LOG_N1	1	6	1	6.5	1	8	1	6
LOG_N2	1	8	1	8	1	8.5	1	8.5
LOG_NA	1	8	1	8	1	8	1	9
LOG_NW1			1	8	1	5	1	7
LOG_NW2					1	5		
Migrants (m)			0	1				
tadm			10	5,000				
tdiv					10	15,000		
tdom	10	8,000	10	25,000			10	8,000
tdom1					10	7,000		
tdom2					10	7,000	10	3,000

DNA_MUTATION: rate per site per generation, with a generation time assumed to be 5y; MSAT_MUTATION: rate per locus per generation; GAMMA: gamma distribution of the msat mutation rate; LOG N1/N2/NA/NW1/NW2: Log of the estimated effective population size of population 1/ 2/ ancestral/ wild 1/ wild 2; Migrants: proportion of population 1 made of migrants from population 2; tadm: time of admixture between populations, in generations; tdom/1 /2: time of domestication, in generations.

Table S11. Fifteen summary statistics of the observed dataset (n = 642) generated by ARLEQUIN.

Statistic	*Obs. Values*
OBS0_K_1	5
OBS0_K_2	8.8235
OBS0_TOT_K	9
OBS0_H_1	0.5638
OBS0_H_2	0.6282
OBS0_TOT_H	**0.6268**
OBS0_FST_2_1	0.042
OBS1_S_1	**16**
OBS1_S_2	**42**
OBS1_PRS_1	3
OBS1_PRS_2	29
OBS1_TOT_S	45
OBS1_PI_1	3.61132
OBS1_PI_2	2.07592
OBS1_PI_2_1	3.1901

Obs0 refers to statistics calculated from the microsatellite data. Obs1 refers to statistics calculated from the mtDNA data. Statistics in bold were removed before the final analysis due to high correlation to other statistics estimated with the Spearman rho correlation analysis (Fig. S9).

Population specific summary statistics: the mean number of alleles (Obs0_K) and mean genetic diversity (Obs0_H) across loci, the mean of pairwise differences (Obs1_Pi), the segregating sites (Obs1_S) and the private segregating sites (Obs1_PrS).

Population pairwise comparisons: the mean total genetic diversity (Obs0_tot_H), the pairwise F_{ST} for microsatellite (Obs0_FST), mean number of alleles (Obs0_tot_K) across loci for two populations, number of haplotypes (Obs1_tot_K), and the mean of pairwise differences (Obs1_PI_2_1).

Table S12. Information about the seventeen microsatellite loci.

Marker	Reference	nA	H_O	H_E	PIC	null a.	HWE global population	HWE without EAF	HWE only EAF
CMS09	(90)	10	0.685	0.725	0.683	0.029	NS	NS	NS
CMS13	(90)	9	0.661	0.716	0.671	0.042	7.03E-04	NS	NS
CMS15	(90)	12	0.705	0.773	0.739	0.046	9.13E-06	NS	NS
CMS18	(90)	5	0.369	0.402	0.359	0.041	NS	1.21E-05	NS
CMS25	(90)	8	0.570	0.637	0.567	0.056	9.41E-06	3.07E-05	NS
CMS50	(90)	15	0.786	0.865	0.849	0.048	4.10E-07	NS	NS
CMS121	(90)	14	0.709	0.761	0.727	0.036	1.61E-03	NS	NS
CVRL01R†	(91)	24	0.809	0.868	0.858	0.035	2.49E-04	2.18E-03	NS
CVRL04R†	(91)	7	0.611	0.643	0.570	0.024	NS	NS	NS
CVRL05R†	(91)	12	0.617	0.667	0.616	0.041	1.30E-04	NS	NS
CVRL06R†	(91)	5	0.316	0.328	0.294	0.017	NS	NS	NS
CVRL08	(91)	3	0.296	0.338	0.282	0.065	2.78E-04	2.32E-03	NS
LCA66	(92)	7	0.677	0.735	0.688	0.041	NS	NS	NS
VOLP10	(93)	12	0.723	0.795	0.763	0.048	6.00E-08	NS	NS
VOLP32	(93)	3	0.342	0.348	0.288	0.009	NS	NS	NS
YWLL44	(94)	10	0.527	0.626	0.574	0.081	6.47E-11	4.01E-10	NS
YWLL59	(94)	2	0.401	0.483	0.366	0.092	3.10E-07	2.30E-04	3.00E-04
Average over all loci	(95)	9.29	0.577	0.582	0.630				

	5'-3' Forward sequence	5'-3' Reverse sequence
CVRL1R†	GGGCAAGCTTGACTTGACTT	TGCTTATCATGCACGAGGTC
CVRL4R†	CTTTCTGAACTTCTGTTGTCTGC	AAACCTGCAAGTTCTCAGTTTAAG
CVRL5R†	TCTTCCTGGTCCATATCTTGTAGAC	CACTGGTCCCTGTCATTATGC
CVRL6R†	AATTCTGACCAGGAGTCTGCTT	AGTCCATGAGCAAGTGAATGAA

The set of markers was selected according to recommendations from the joint Food and Agricultural organization of the United Nations (FAO) and International Society for Animal Genetics (ISAG) panel on livestock genetic diversity. Genotypes from locus CMS17 were excluded from the analysis, as this marker, which was developed for *C. bactrianus*, was found monomorphic in *C. dromedarius*. Parameter estimations were performed with CERVUS 3.0.7.

†modified from Mariasegaram *et al.* (91); n_A: number of alleles; PIC: Polymorphism Information Content – value of a marker for detecting polymorphism within the dromedary population; HWE: non-deviation from Hardy-Weinberg Equilibrium expectation (NS: non significant *P*-value > 0.01); null a.: estimated frequency of null allele.

Table S13. Posterior density distributions of historical and demographic parameters in the best-fitting scenario (*ii*) estimated with ABCTOOLBOX

Parameter	DNA MUTATION	GAMMA	LOG N1	LOG N2	LOG NA	LOG NW1	MSAT MUTATION	MIGRANTS	tadm	tdom
Mode	1.095E-07	13.241	4.040	5.538	2.688	2.970	7.63E-04	0.84925	85.23	2783.39
Median	2.312E-07	11.773	4.051	5.552	3.369	3.973	1.51E-03	0.71201	680.54	6333.29
Quantile 50 Lower bound	1.185E-07	10.086	3.582	5.086	2.305	2.645	7.46E-04	0.52397	319.87	3081.82
Quantile 50 Upper bound	3.923E-07	13.279	4.529	6.016	4.674	5.550	2.53E-03	0.85322	1185.80	11661.10
Quantile 90 Lower bound	3.159E-08	8.540	2.922	4.424	1.326	1.437	1.47E-04	0.22328	61.54	654.12
Quantile 90 Upper bound	6.773E-07	14.543	5.227	6.672	6.620	7.277	4.66E-03	0.96650	2145.68	19923.50
Quantile 95 Lower bound	1.976E-08	8.285	2.712	4.212	1.165	1.229	6.68E-05	0.14043	29.54	315.33
Quantile 95 Upper bound	7.586E-07	14.749	5.452	6.880	7.100	7.586	5.51E-03	0.98266	2490.23	21684.50
Quantile 99 Lower bound	9.994E-09	8.048	2.306	3.803	1.021	1.036	1.39E-06	0.03751	3.88	31.80
Quantile 99 Upper bound	8.685E-07	14.935	5.877	7.278	7.674	7.884	6.87E-03	0.99629	3025.11	23639.30
HPD 50 Lower bound	2.481E-05	11.166	3.570	5.116	1.704	1.915	3.91E-05	0.69349	10.00	22.00
HPD 50 Upper bound	-2.457E-05	14.191	4.482	5.995	3.920	4.623	1.44E-03	0.97485	661.96	6297.88
HPD 90 Lower bound	2.725E-05	8.774	2.907	4.447	1.000	1.141	3.53E-05	0.34173	10.00	22.00
HPD 90 Upper bound	-2.668E-05	14.719	5.201	6.628	5.889	6.874	3.70E-03	0.99998	1740.20	17092.40
HPD 95 Lower bound	2.774E-05	8.493	2.714	4.236	1.000	1.000	3.30E-05	0.23621	10.00	22.00
HPD 95 Upper bound	-2.706E-05	14.894	5.450	6.874	6.593	7.226	4.66E-03	0.99998	2141.41	19979.30
HPD 99 Lower bound	2.704E-05	8.1407	2.299	3.814	1.000	1.000	3.43E-05	0.09048	10.00	22.00
HPD 99 Upper bound	-2.621E-05	15	5.864	7.261	7.508	7.789	6.41E-03	0.99998	2843.52	22991.70

DNA_MUTATION: rate per mitochondrial site per generation, with a generation assumed to be 5y; MSAT_MUTATION: rate per autosomal locus per generation; GAMMA: gamma distribution of the microsatellite mutation rate; LOG N1/N2/NA/NW1/NW2: Log of the estimated effective population size of population 1/ 2/ ancestral/ wild 1/ wild 2; MIGRANTS: proportion of population 1 made of migrants from population 2; tadm: time of admixture between populations, in generations; tdom/1 /2: time of domestication, in generations.

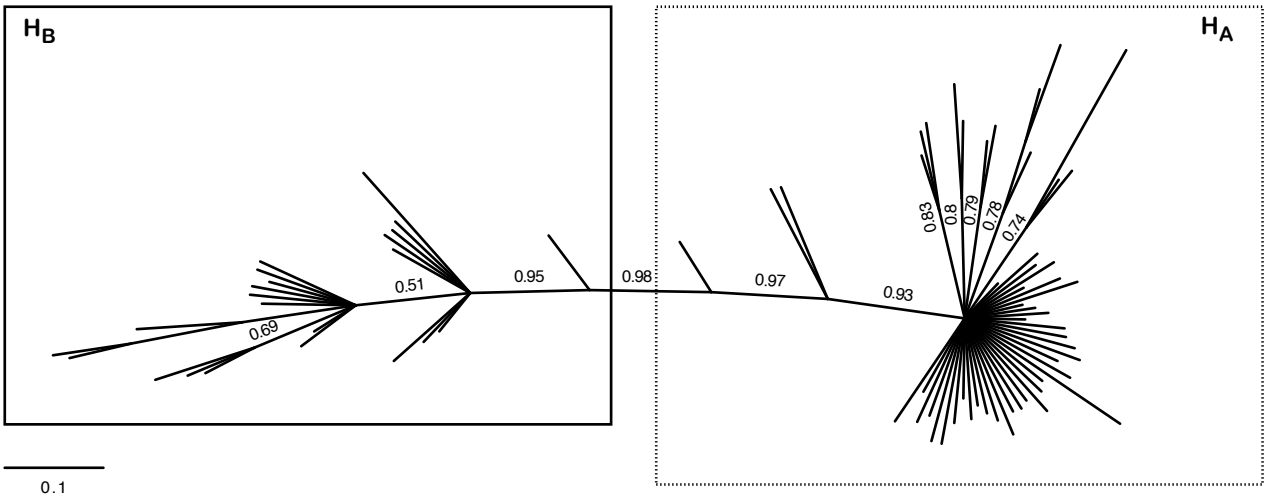


Figure S1. Consensus Bayesian phylogeny of the 76 modern dromedary haplotypes, resolving into two haplogroups (H_A and H_B).

Values above the branches indicated the posterior probabilities (PP).

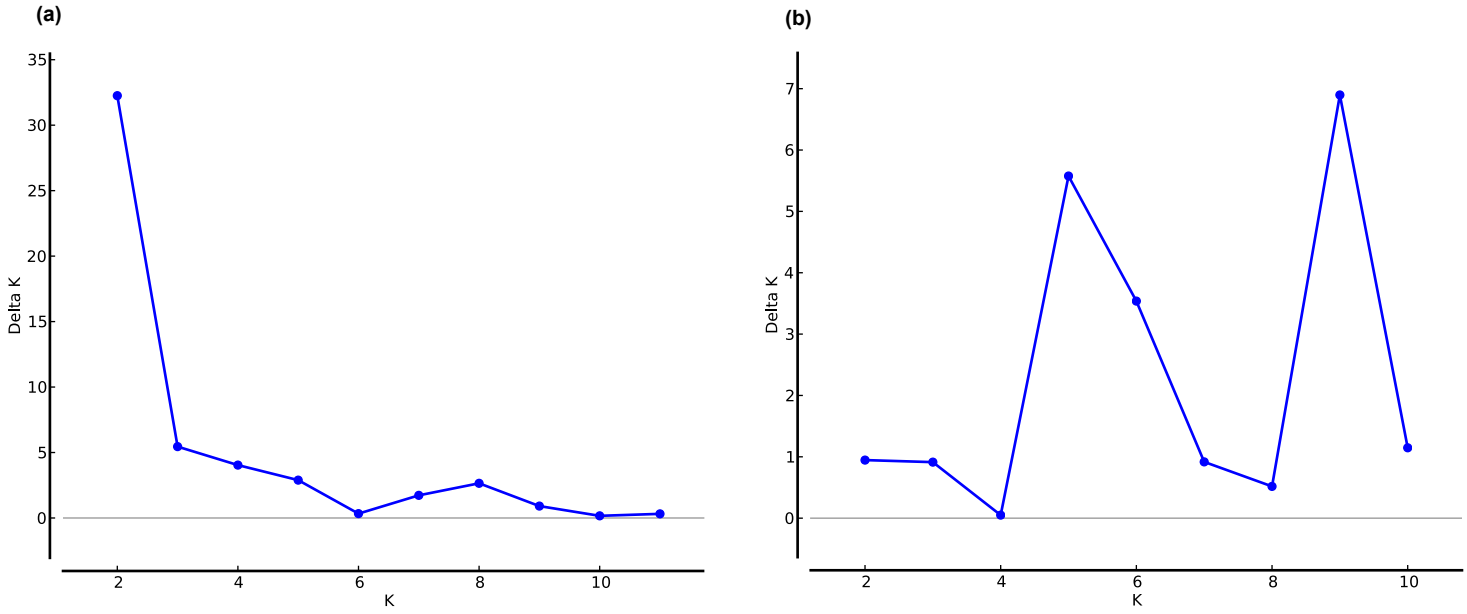


Figure S2. Delta K analysis for a different number of clusters (K).

(a) For the global modern sample set consisting of 970 dromedaries, Delta K showed a peak at K = 2, suggesting two clusters as the optimal solution. **(b)** Excluding the EAF individuals from the dataset (n=810 dromedaries), Delta K showed a peak at K = 9, suggesting nine clusters as the optimal solution.

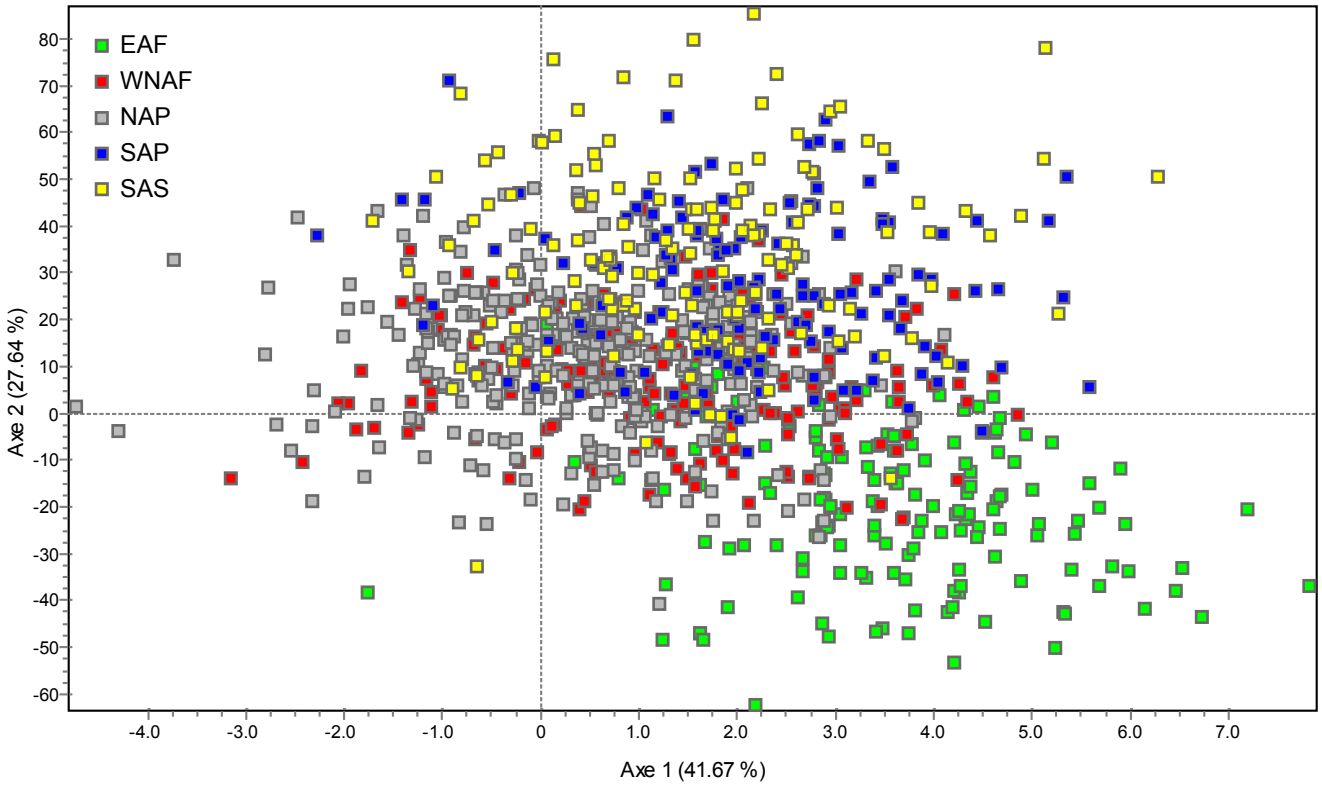


Figure S3. Factorial correspondence analysis (FCA) of 970 modern dromedaries based on 17 microsatellite loci.

The individuals are colored according to their geographical origin. The axes 1–2 explain 69.31% of the variation among the populations and separate most of the EAF individuals from the rest of the population. EAF: East African population (n = 160); WNAF: Western and Northern African populations (n = 207); NAP: Northern Arabian Peninsula (n = 317); SAP: Southern Arabian Peninsula (n = 141); SAS: Southern Asian population (n = 145).

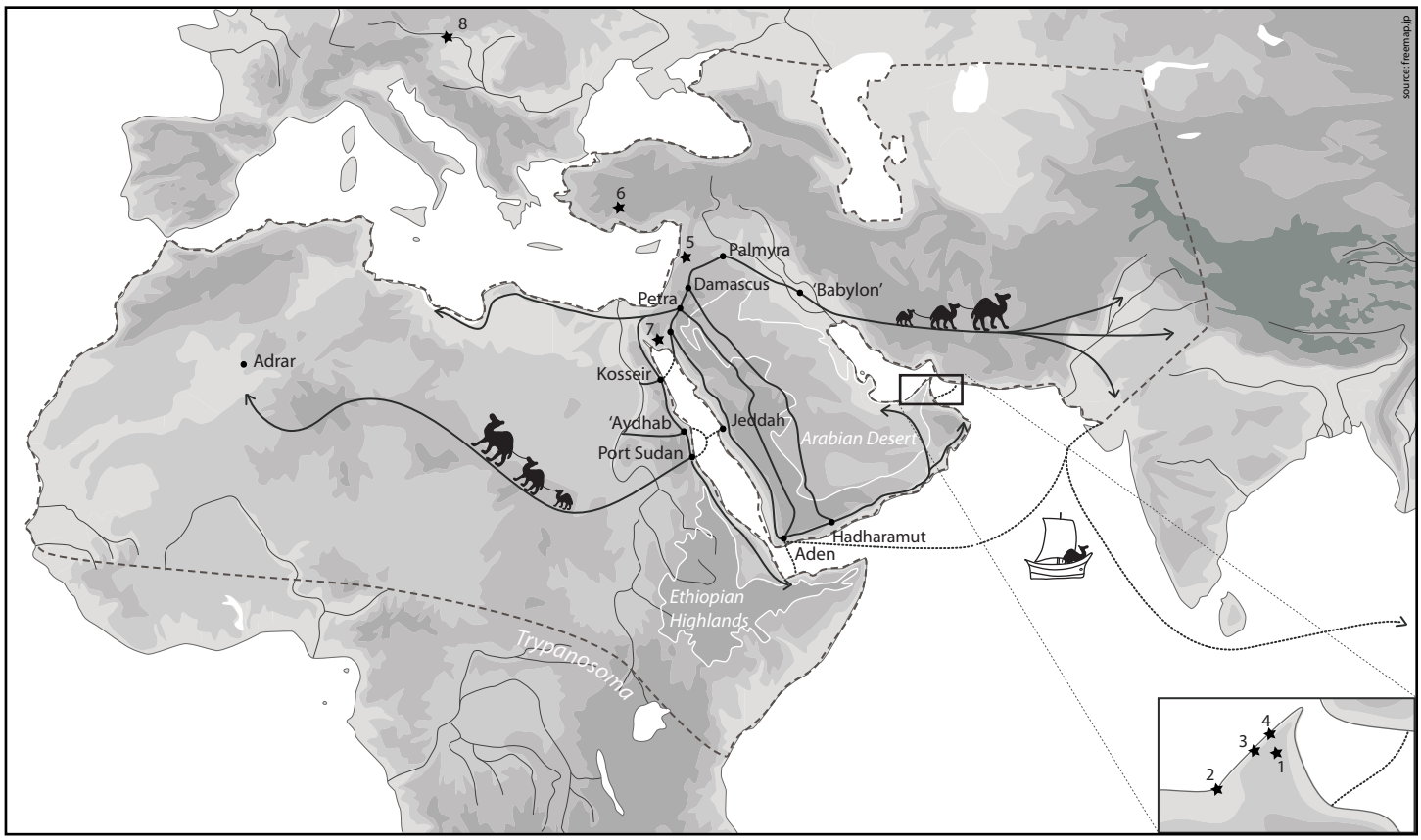


Figure S4. Schematic representation of the historical network of caravan routes (*i.e.*, Incense and Silk routes) according to descriptions from Bulliet (1) and Heiss (16).

Archaeological sites from which the ancient specimens originated are pictured with a black star (right-bottom-corner zoom on the UAE peninsula – 1: Al-Buhais (5000-4000 BCE); 2: Umm-an-Nar (Early Bronze Age: 3000-2000 BCE); 3: Al-Sufouh (ca. 2400-1400 BCE); 4: Tell Abraq (Late Bronze – Iron Age: 1260-500 BCE); main map – 5: Apamea, Syria (Early Byzantine: 400-600 CE); 6: Sagalassos, Turkey (Early Byzantine: 450-700 CE); 7: Aqaba, Jordan (Mamluk and Ottoman Periods: 1260-1870 CE); 8: Tulln, Austria (2^{nd} Turkish war *ca.* 1683 CE)).
The historical repartition of domestic dromedaries (depicted with dashed lines) is bordered on the south by areas infested with Trypanosoma and included some geographical barriers as Ethiopian Highlands and Arabian Desert (surrounded with white lines).
The land route (depicted by solid lines) from Aden to North Arabian Peninsula was part of the Incense Road and consisted of three main itineraries, namely *i)* al tariq Tihama (or Tihama road) along the coastal plains, *ii)* al tariq al jibal (or the highland road), and *iii)* al tariq al sufla (or the lower road) via eastern Arabian Desert. On the western coast of the Red Sea existed a trading route connecting the Horn of Africa to Petra and Damascus via Port Sudan, 'Aydhab and Myos Hormos (near today's Kosseir). The trans-Saharan route passed through the major centers of southern Saharan rock art such as Darfur (western Sudan), Ennedi and Tibesti (Chad), Tassili and Ahaggar (Algeria) up to Adrar (Algeria) and linked these regions with the upper Nile valley. A second route, bordering the Mediterranean coast, connected Northwestern Africa to the North of the Arabian Peninsula from where caravans were leaving to Southern Asia along the Silk Road. Two major sea routes (pictured with dotted lines) connected the Arabian Peninsula to the African continent: *i)* the southern one, from Hadharamut and Aden to the Horn of Africa ('Land of Punt'), and *ii)* from Jiddah to 'Aydhab and Port Sudan. Sea routes were also used between the Gulf of Oman and Iran and between South Arabian Peninsula (Aden) and Indian subcontinent, known as the Spice Route. The most contemporary migration route started in the 1860s and linked Pakistan to Australia where several thousands of camels were imported.

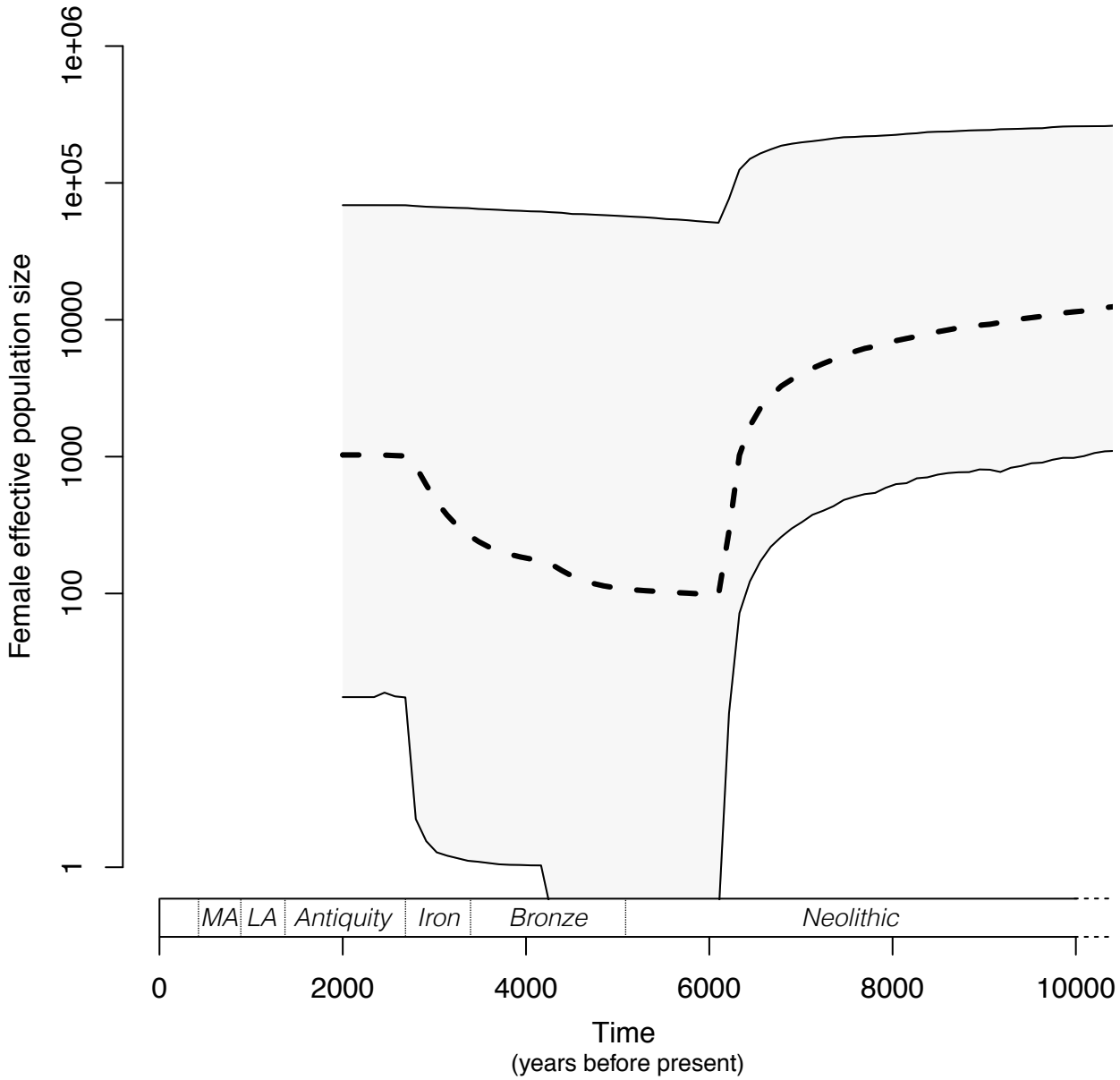


Figure S5. Bayesian skyline plot derived from the alignment of eight MT-CR sequences from wild dromedaries showing the female effective population size (N_e) fluctuations for the past ten millennia.

The thick dashed line depicts the median estimate of N_e with black lines delimiting the 95% HPD. Substitution rate $\mu = 1.232 \times 10^{-06}$ sub/site/y [95% HPD: 4.4353×10^{-07}, 2.2132×10^{-06}] was used to compute the BSP under the stepwise constant function.
For comparison, approximate archaeological time periods have been added to the time scale (in years before present). LA: Late Antiquity; MA: Middle Age.

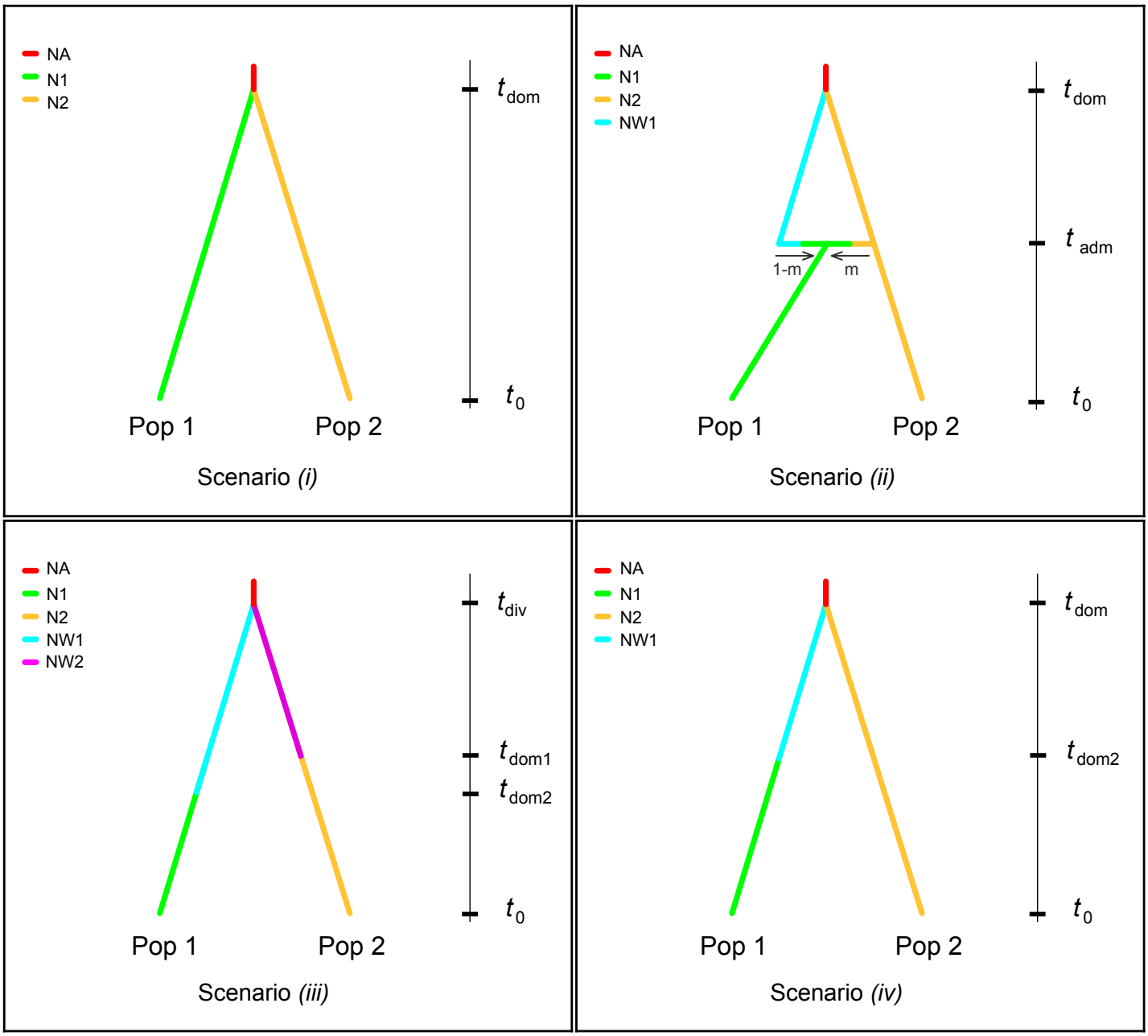
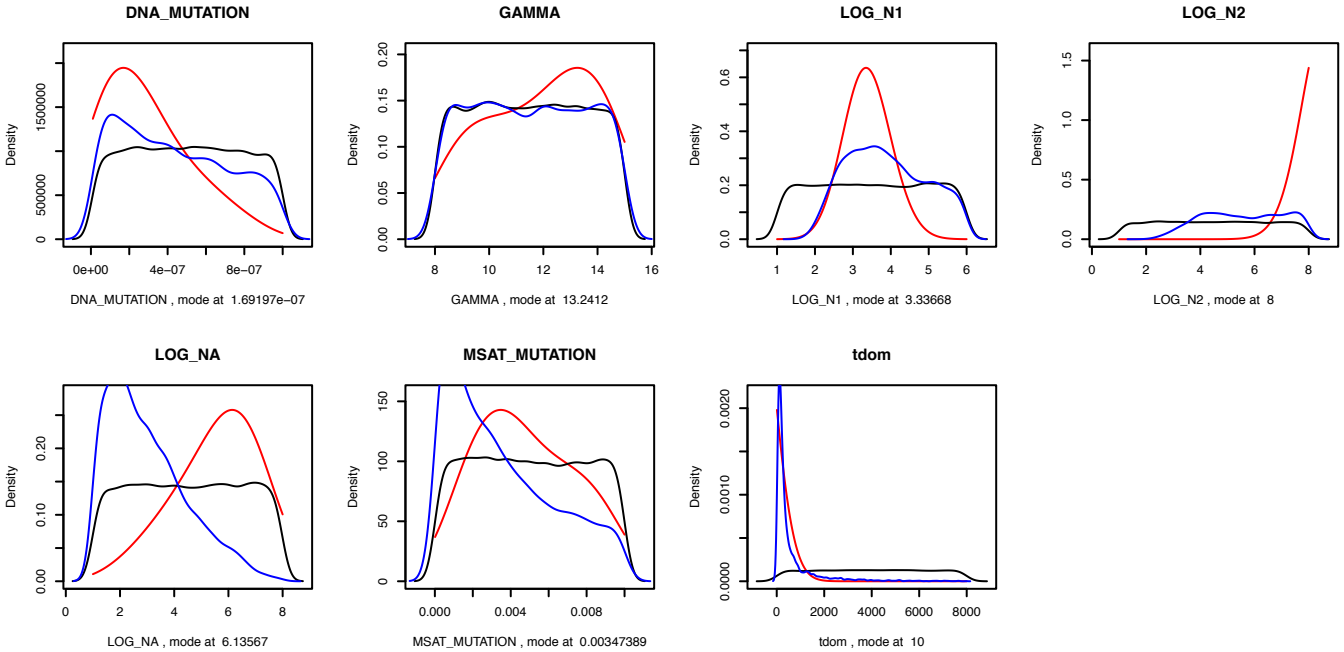


Figure S6. Four different scenarios of domestication simulated with ABCToolbox on the combined mitochondrial and nuclear dataset (n = 642).

Scenarios: *(i)* one domestication, *(ii)* one domestication with consecutive admixture from a wild unsampled source population, *(iii)* two independent domestications at the variable time points, and *(iv)* two domestications at serial time points. In scenario *(iii)*, t_{dom1} and t_{dom2} do not constrain each other.
Pop1 = EAF (n = 62); Pop2 = WNAF_NAP_SAP_SAS_combined (n = 580);
NA = ancestral N_e; N1 = N_e of Pop1; N2 = N_e of Pop2; NW1 = N_e of a wild unsampled source population 1; NW2 = N_e of a wild unsampled source population 2; t_{dom} = time of domestication; $t_{dom1/2}$ = time of domestication at variable time points; t_{div} = time of divergence of the two unsampled wild ancestral populations; t_{adm} = time of admixture (introgression) from a wild unsampled population; t_0 = present. Time is not to scale.

(a) Scenario *(i)*



(b) Scenario *(ii)*

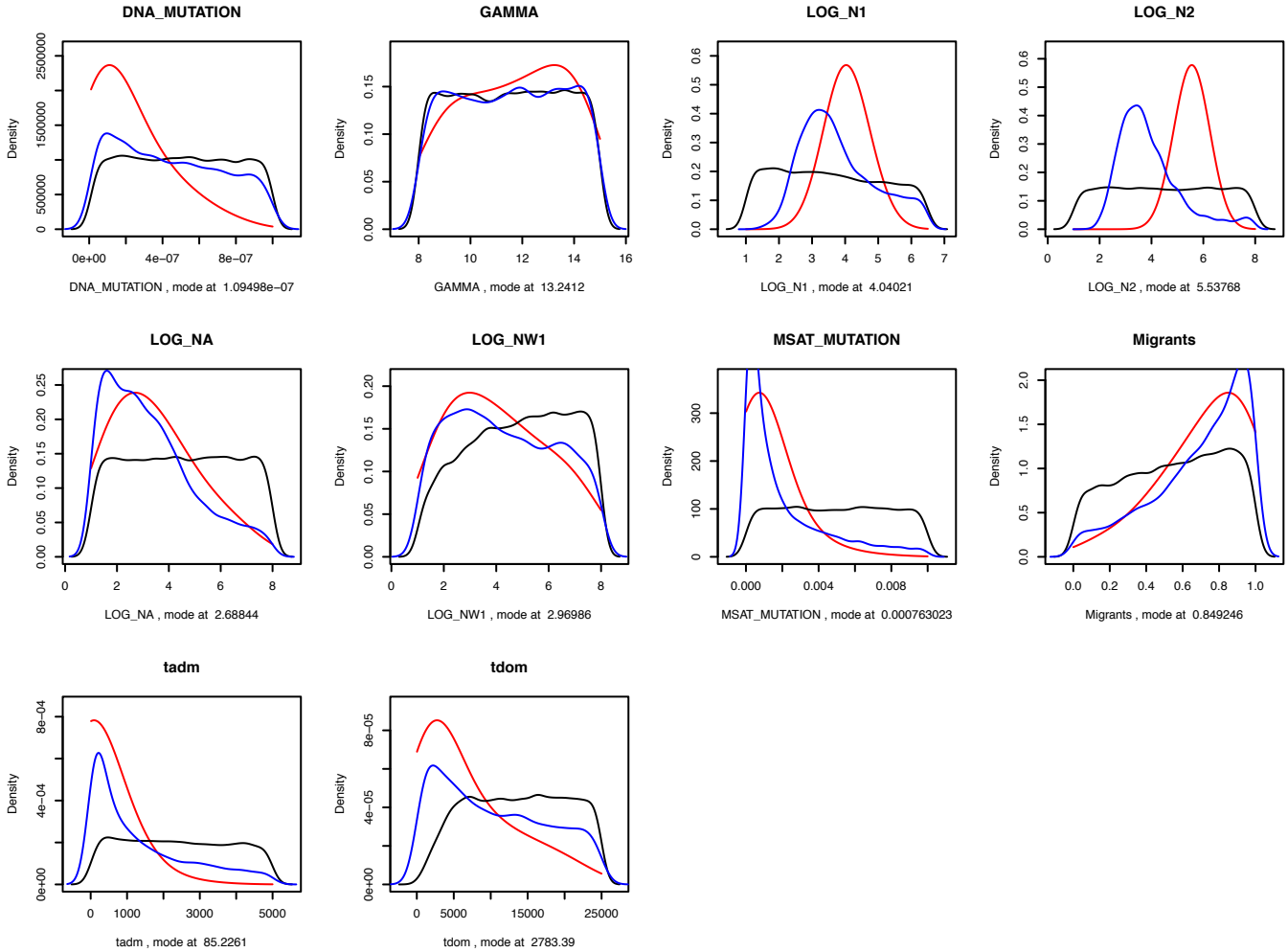
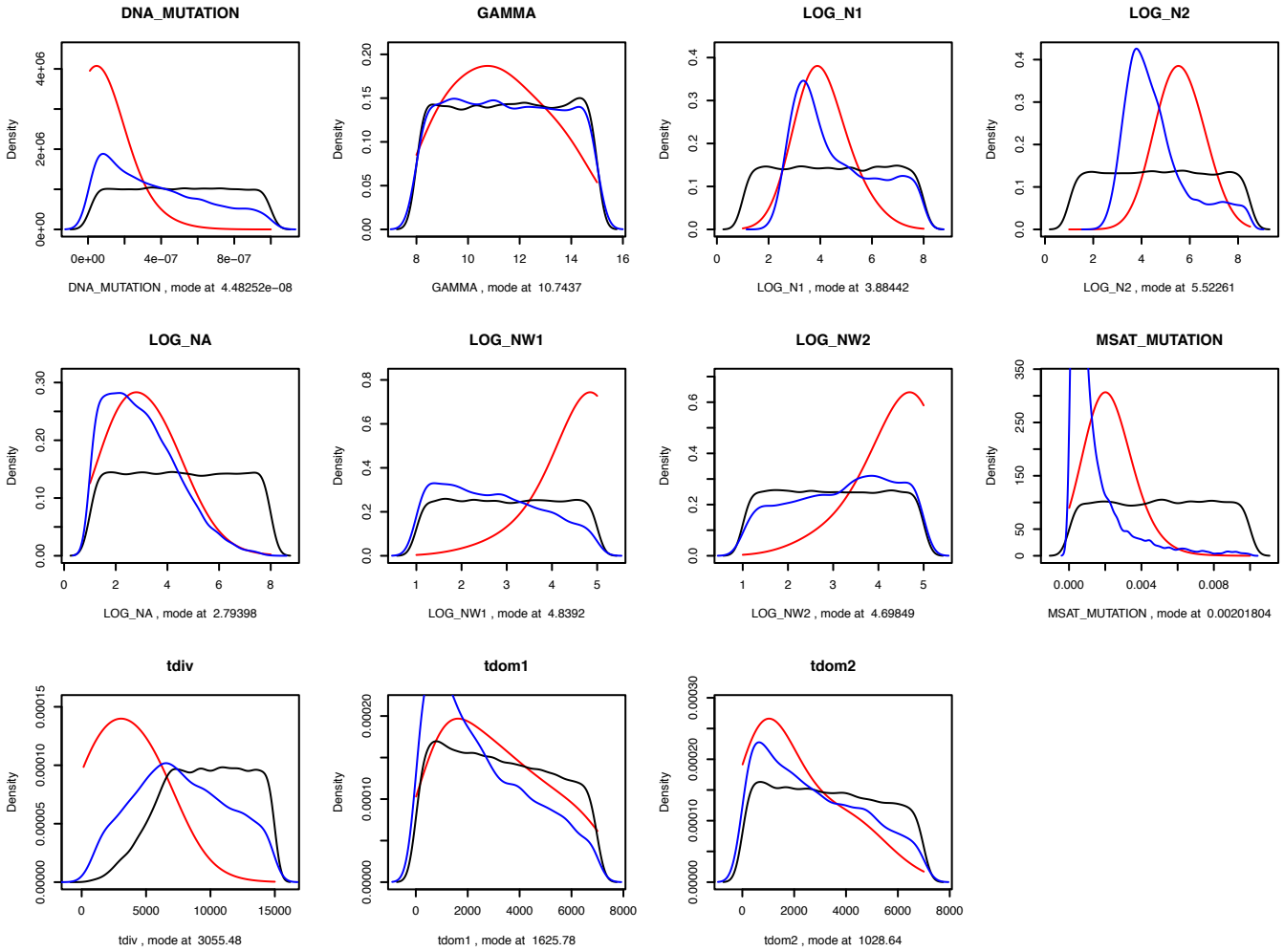


Figure S7. Posterior (in red), marginal (in blue), and prior (in black) distributions of the demographic parameters for the four domestication scenarios (**a-d**), plotted using the 5,000 closest simulations to the observed dataset.

DNA_MUTATION: rate per site per generation, with a generation time assumed to be 5y; MSAT_MUTATION: rate per locus per generation; GAMMA: gamma distribution of the msat mutation rate; LOG N1/N2/NA/NW1/NW2: Log of the estimated effective population size of population 1/ 2/ ancestral/ wild 1/ wild 2; MIGRANTS: proportion of population 1 made of migrants from population 2;t_{dom}: time of domestication, in generation; $t_{dom1/2}$: time of domestication at variable time points; t_{adm}: time of admixture (introgression) between populations.

(c) Scenario *(iii)*



(d) Scenario *(iv)*

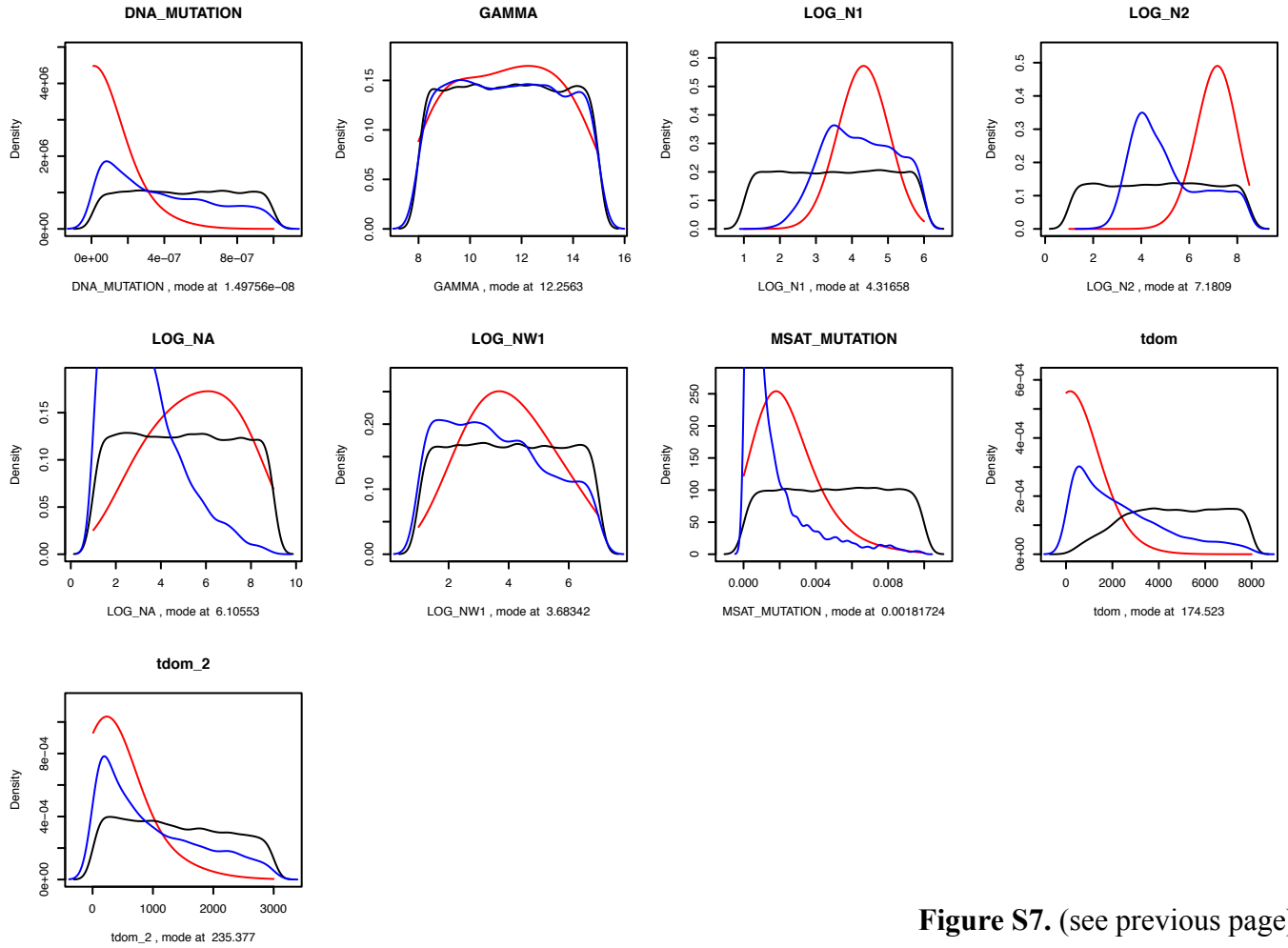


Figure S7. (see previous page)

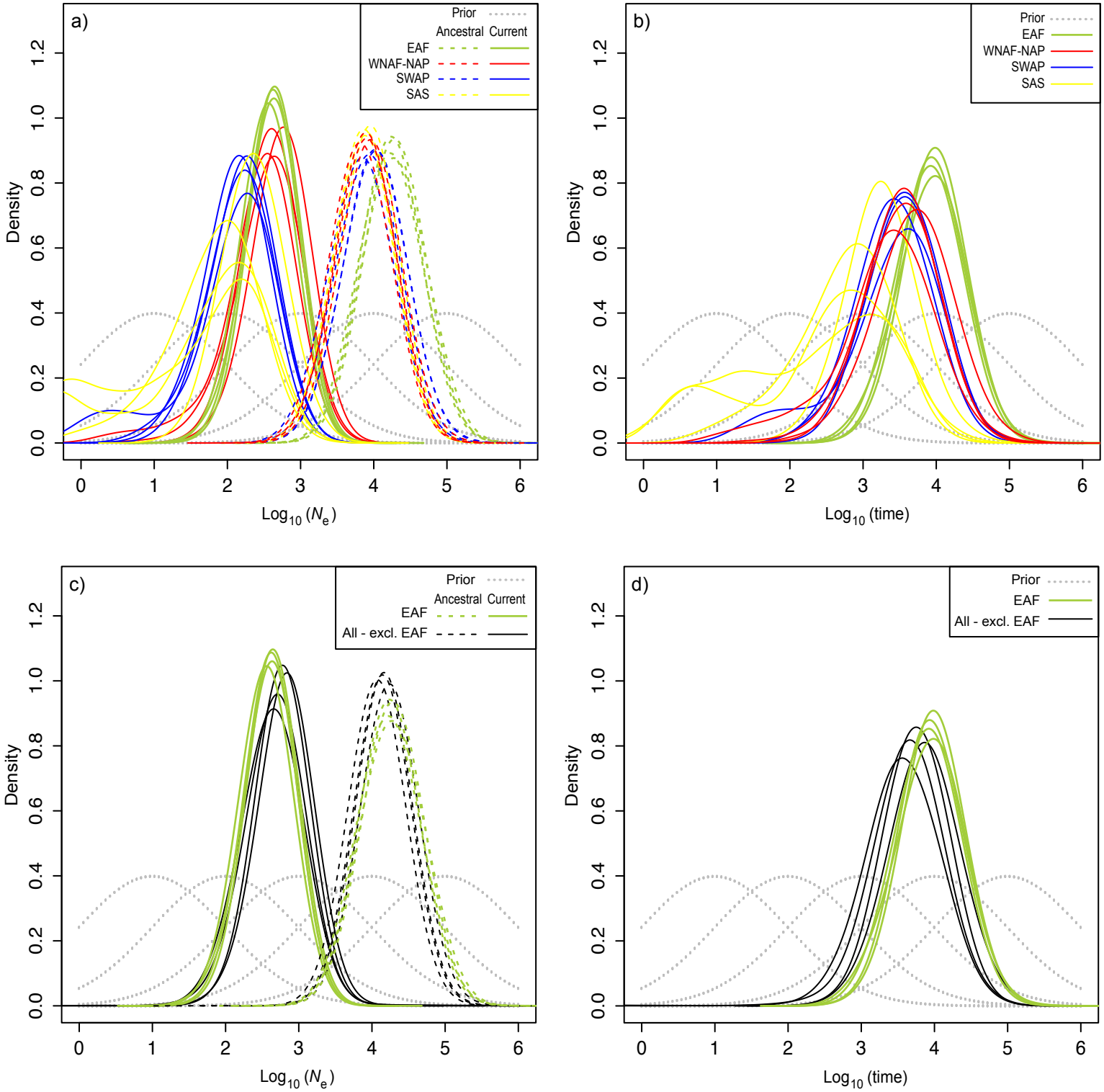


Figure S8. Density plots of MSVAR results showing ancestral (dashed lines) and current (solid lines) effective population size (N_e) estimations of **(a)** the different dromedary populations and **(c)** all populations combined excluding EAF, as well as **(b, d)** the time since their respective declines.

Coalescent simulations were run using the 17-nuclear-loci dataset. Different priors are shown in grey dotted lines. Details are available in Table S9. EAF = East Africa, WNAF-NAP = Western-Northern Africa and Northern Arabian Peninsula, SWAP = Southwest Arabian Peninsula, SAS = Southern Asia, including Australia.

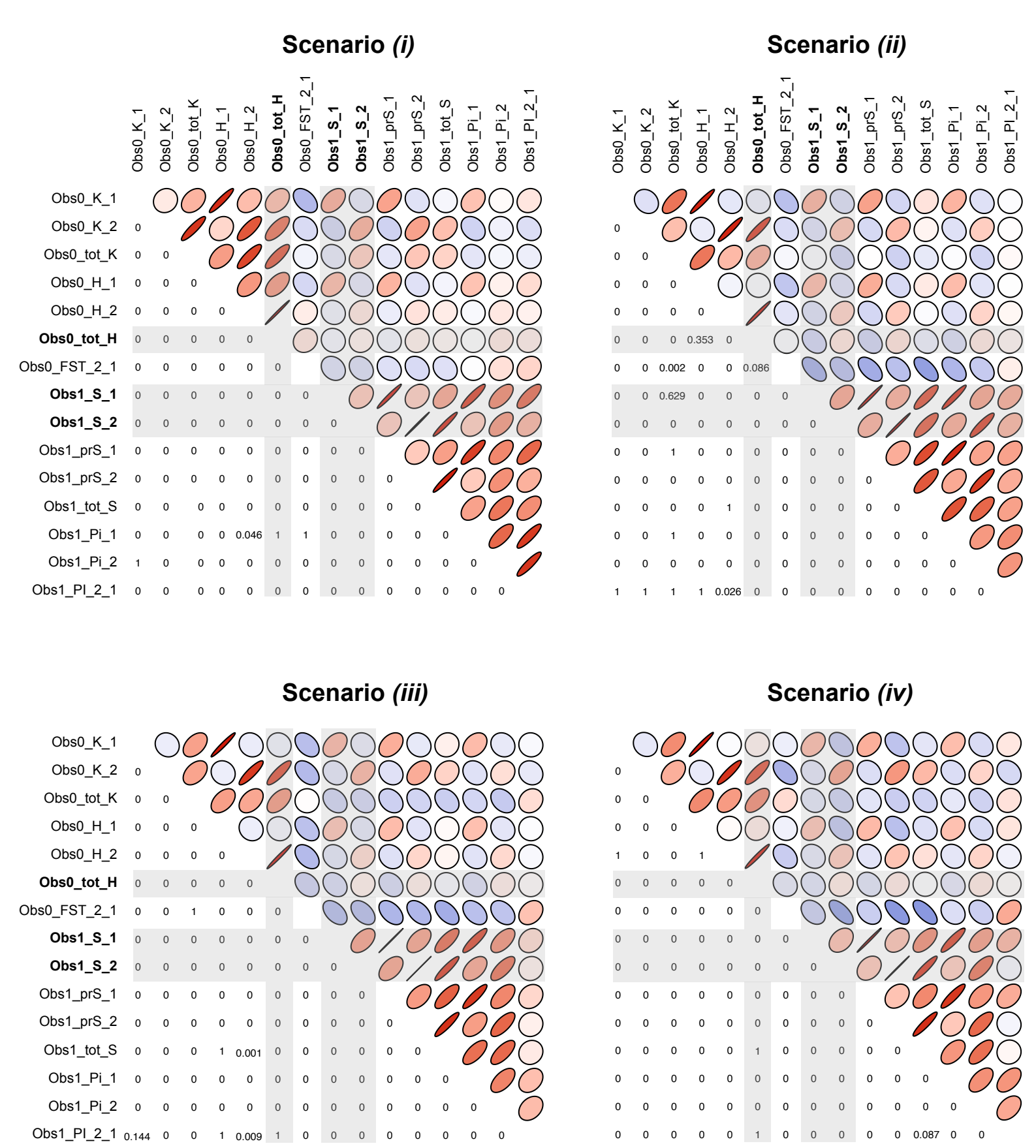


Figure S9. Spearman's rho correlation analysis of 15 simulated summary statistics used in ABCTOOLBOX.

Ellipse glyphs in the upper part of the matrix are shaped to match the corresponding Spearman's coefficient. Colors (blue to red scale) emphasize the sign of the slope. Bonferroni corrected *P*-values are presented on the lower part of the matrix. Population specific summary statistics: the mean number of alleles (Obs0_K) and mean genetic diversity (Obs0_H) across loci, the mean of pairwise differences (Obs1_Pi), the segregating sites (Obs1_S) and the private segregating sites (Obs1_PrS). Population pairwise comparisons: mean total genetic diversity (Obs0_tot_H), pairwise F_{ST} for microsatellite (Obs0_FST), mean number of alleles (Obs0_tot_K) across loci for two populations, number of haplotypes (Obs1_tot_K), and the mean of pairwise differences (Obs1_PI_2_1).

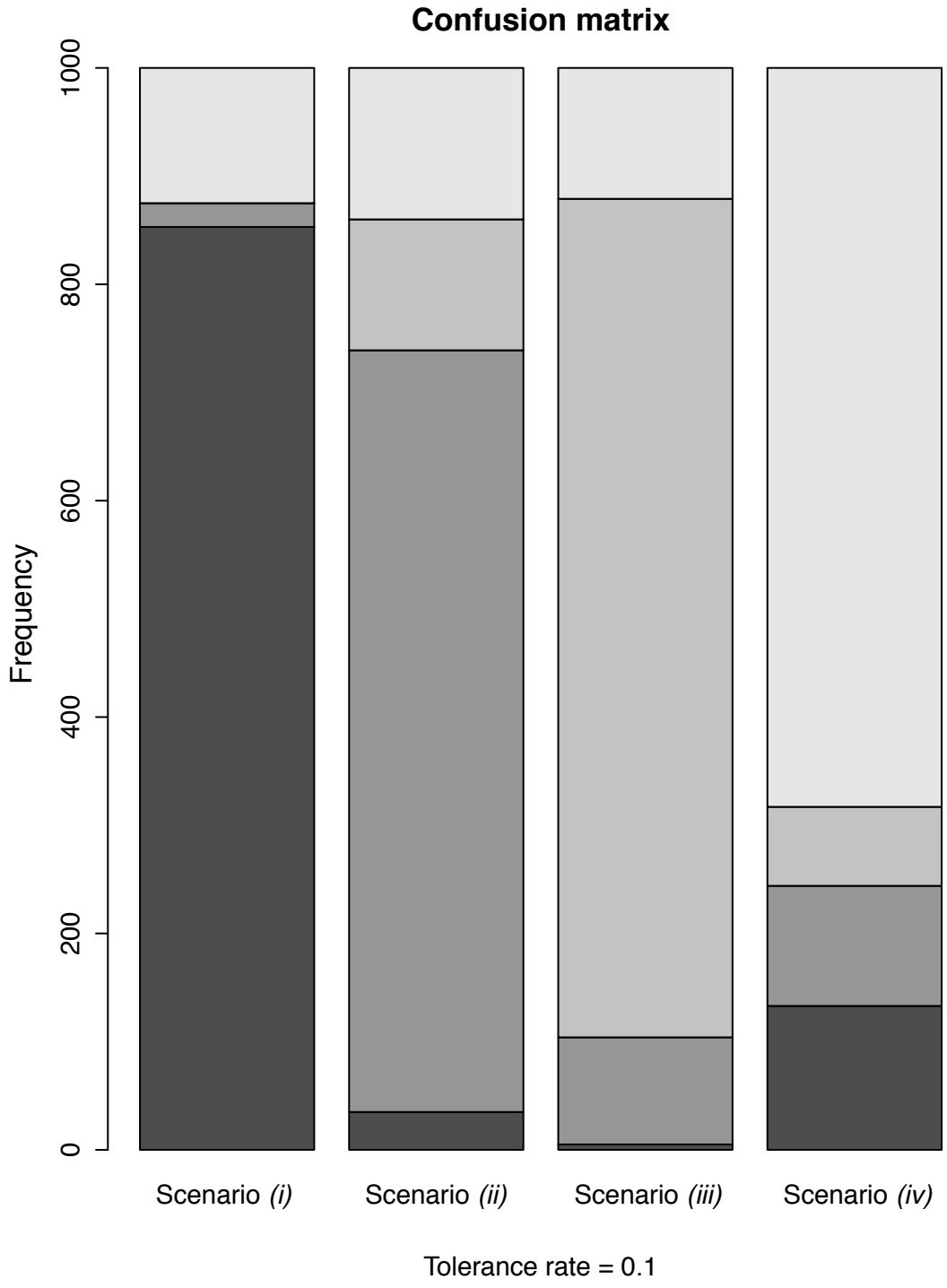


Figure S10. Model missclassification of the four different domestication scenarios using the R package *abc*.

The confusion matrix is based on 1,000 samples of each scenario. Each grey shade from dark to light corresponds to the scenarios (*i*) to (*iv*), respectively. Scenario (*ii*) clearly differentiated based on the twelve summary statistics (Fig. S9; Table S11).

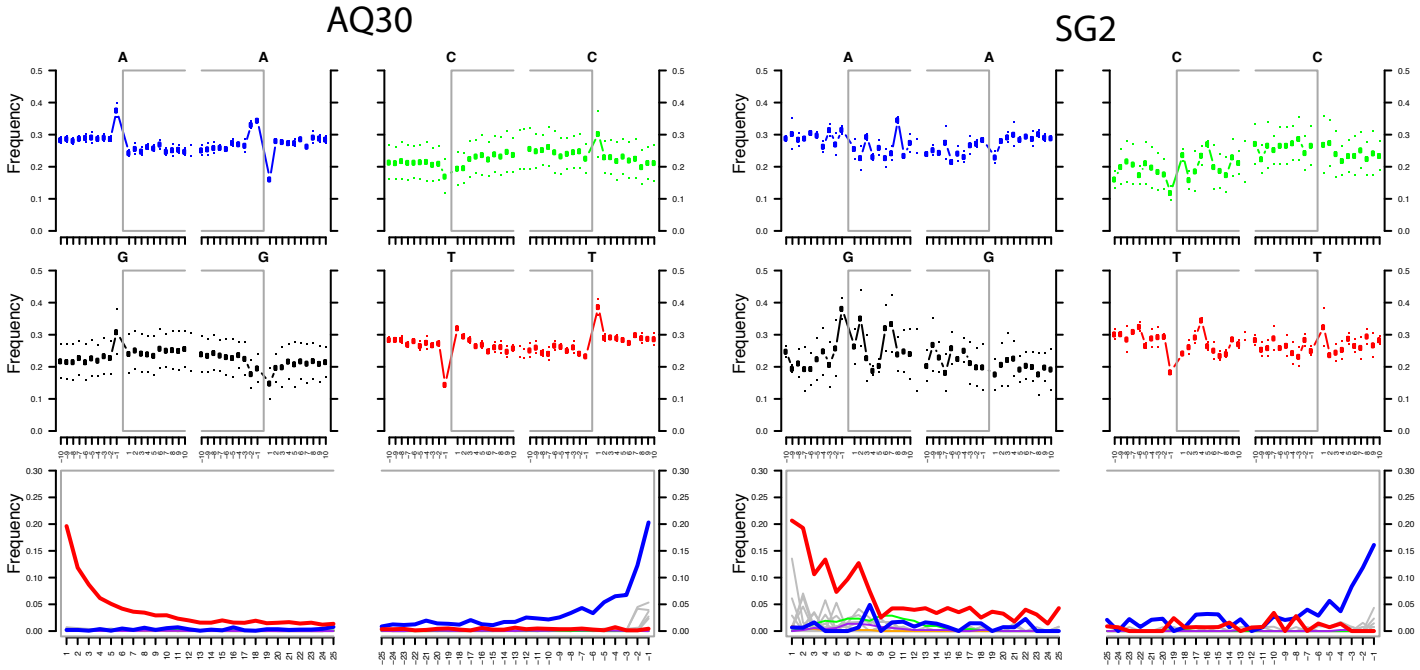


Figure S11. Examples of damage patterns in early-domestic dromedary sequences obtained from the samples collected in Aqaba (AQ30) and Sagalassos (SG2) archaeological sites.

The base frequencies at the 5' and 3' ends of the strand breaks are depicted (top and middle). Frequencies are shown for A, G, C and T for the 10 bases at the 5' and 3' ends of the breaking sites. Note the excess of purines (A and G) at the first nucleotide position preceding the strand break. The gray square brackets show the start and end of the molecules (strand break). The C to T nucleotide misincorporation at the first and the last 25 bases is shown (bottom). There is an increase in frequency of T at the 5' and A at the 3' end, which is a typical pattern for aDNA damage.

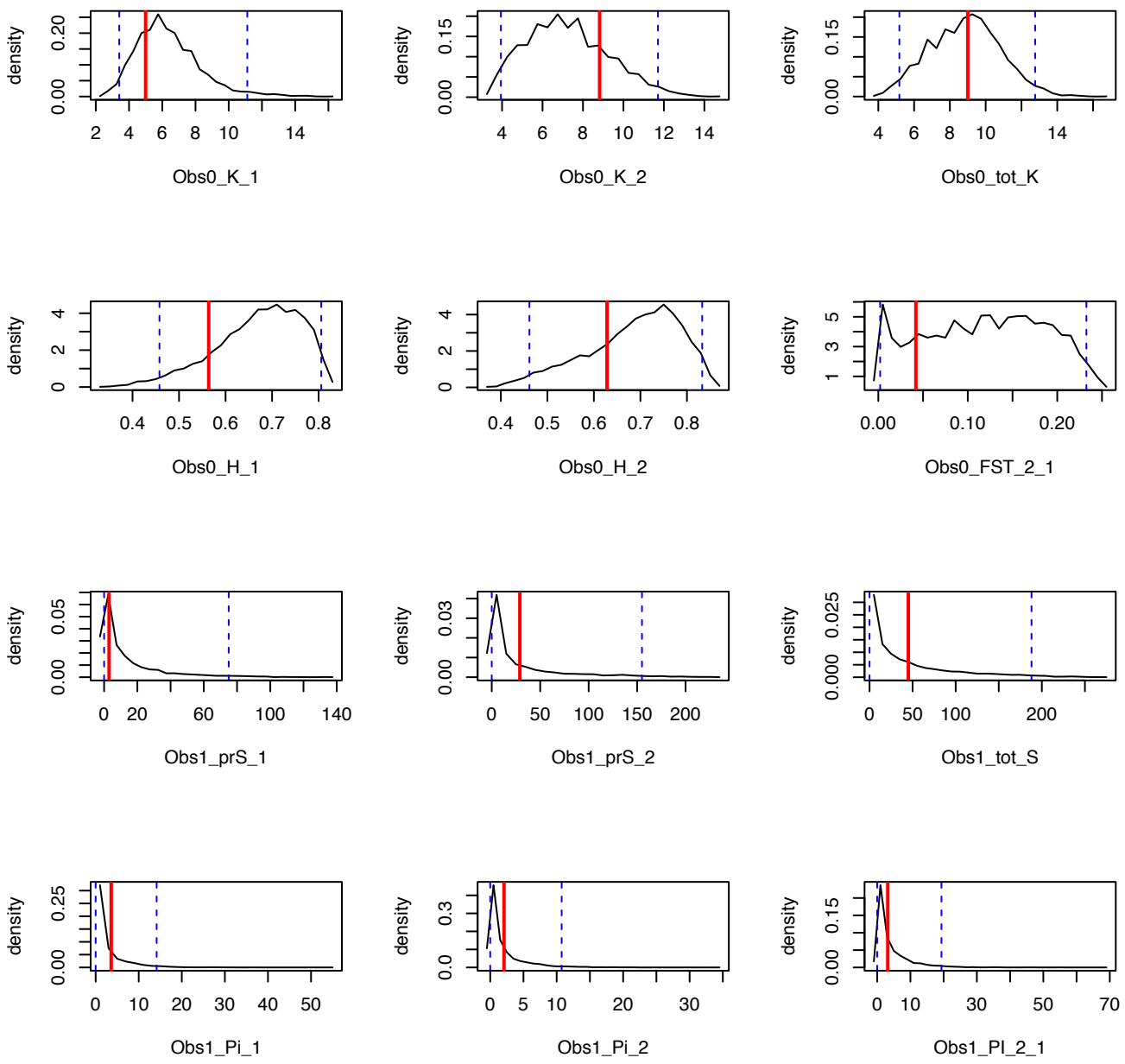


Figure S12. Density distributions of the twelve summary statistics (black) with 2.5 and 97.5 quantiles (blue).

Distributions were generated from the 5,000 simulations closest to the observed dataset of scenario *(ii)*. Corresponding observed summary statistic of each plot is shown in red.

REFERENCES

1. FAO (2012) FAOSTAT - Live animals, Camels. http://faostat.fao.org
2. Rangan H, Kull C (2010) The Indian Ocean and the making of Outback Australia. *Indian Ocean studies – Cultural, social and political perspectives*, eds Moorthy S & Jamal Y (Routledge, London), pp 45–72.
3. Faye B, Abdallah H, Almathen F, Harzallah B, Al-Mutairi S (2011) *Camel biodiversity – Camel phenotypes in the Kingdom of Saudi Arabia* (Camel Breeding, Protection and Improvement Center, Riyadh) p 69.
4. von den Driesch A, Obermaier H (2007) The hunt for wild dromedaries during the 3rd and 2nd millennia BC on the United Arab Emirates coast. Camel bone finds from the excavations at Al Sufouh 2 Dubai, UAE. *Skeletal series and their socio-economic context,* Documenta Archaeobiologiae, eds Grupe G & Peters J (Verlag Marie Leidorf GmbH, Rahden, Germany), pp 133–167.
5. Peck EF (1939) Salt intake in relation to contagious necrosis and arthritis of one-humped camels (*Camelus dromedarius*) in British Somaliland. *Veterinary Record* 51:1355-1360.
6. Vagedes K (2013) Arabia Felix - das Land der Königin von Saba' im Spiegel von Tierknochen aus archäologischen Ausgrabungen im Jemen. *Beiträge zur Anthropologie und Paläoanatomie,* Documenta Archaeobiologiae, eds Grupe G*, et al.* (Verlag Marie Leidorf GmbH, Rahden, Germany), pp 163-262.
7. Köhler-Rollefson IU (1991) *Camelus dromedarius*. *Mammalian Species* 375:1-8.
8. Uerpmann H, Uerpmann M (2002) The appearance of the domestic camel in south-east Arabia. *Journal of Oman Studies* 12:235–260.
9. Hedges R, Housley R, Law I, Perry C, Gowlett J (1987) Radiocarbon dates from the Oxford AMS System: Datelist 6. *Archaeometry* 29:289-291.
10. Grigson C (2014) The history of the camel bone dating project. *Anthropozoologica* 49(2):225-235.
11. Pfeiffer I, Volkel I, Taubert H, Brenig B (2004) Forensic DNA- typing of dog hair: DNA-extraction and PCR amplification. *Forensic Science International* 141:149–151.
12. Rohland N, Hofreiter M (2007) Ancient DNA extraction from bones and teeth. *Nature Protocols* 2(7):1756-1762.
13. Rohland N*, et al.* (2010) Genomic DNA sequences from mastodon and woolly mammoth reveal deep speciation of forest and savanna elephants. *PLoS Biology* 8(12):e1000564.
14. Knapp M, Clarke AC, Horsburgh KA, Matisoo-Smith EA (2012) Setting the stage – Building and working in an ancient DNA laboratory. *Annals of Anatomy - Anatomischer Anzeiger* 194(1):3-6.
15. Meyer M*, et al.* (2012) A high-coverage genome sequence from an archaic Denisovan individual. *Science* 338(6104):222-226.
16. Zhang H*, et al.* (2013) Morphological and genetic evidence for early Holocene cattle management in northeastern China. *Nature Communications* 4.
17. Meyer M, Kircher M (2010) Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harbor Protocols* 2010(6):pdb.prot5448.
18. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal* 17(1):10-12.
19. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25(14):1754-1760.
20. Schubert M*, et al.* (2012) Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics* 13(1):178.
21. Li H*, et al.* (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16):2078-2079.
22. Briggs AW*, et al.* (2007) Patterns of damage in genomic DNA sequences from a Neandertal. *Proceedings of the National Academy of Sciences* 104(37):14616-14621.
23. Krause J*, et al.* (2010) A complete mtDNA genome of an early modern human from Kostenki, Russia. *Current Biology* 20(3):231-236.
24. Sawyer S, Krause J, Guschanski K, Savolainen V, Pääbo S (2012) Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS ONE* 7(3):e34131.
25. Jónsson H, Ginolhac A, Schubert M, Johnson PLF, Orlando L (2013) mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* 29(13):1682-1684.
26. Pritchard J, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945-959.
27. Earl D, vonHoldt B (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4:359–361.
28. Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23(14):1801-1806.

29. Belkir K, Borsa P, Goudet J, Chikhi L, Bonhomme F (1999) Genetix, logiciel sous Windows TM pour la génétique des populations. (Laboratoire Génome et Populations, CNRS UPR 9060, Université de Montpellier II, Montpellier, France).
30. Corander J, Tang J (2007) Bayesian analysis of population structure based on linked molecular information. *Mathematical Biosciences* 205:19–31.
31. Excoffier L, Lischer H (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10:564–567.
32. Park S (2001) Trypanotolerance in West African Cattle and the Population Genetic Effects of Selection. Ph.D. thesis (University of Dublin, Dublin, Ireland).
33. Goudet J (1995) FSTAT (Version 1.2): A Computer Program to Calculate F-Statistics. *Journal of Heredity* 86:485–486.
34. Nei M (1987) *Molecular Evolutionary Genetics* (Columbia University Press, New York, USA) p 512
35. Watterson G (1975) On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* 7:256–276.
36. Posada D (2008) jModelTest: Phylogenetic Model Averaging. *Molecular Biology and Evolution* 25:1253–1256.
37. Mburu D*, et al.* (2003) Genetic diversity and relationships of indigenous Kenyan camel (*Camelus dromedarius*) populations: implications for their classification. *Animal Genetics* 34:26–32.
38. Nolte M, Kotze A, van der Bank FH, Grobler JP (2005) Microsatellite markers reveal low genetic differentiation among southern African *Camelus dromedarius* populations. *South African Journal of Animal Science* 35:152–161.
39. Schulz U*, et al.* (2005) The Majorero camel (*Camelus dromedarius*) breed. *Animal Genetic Resources Information* 36:61–71.
40. Spencer P, Woolnough AP (2010) Assessment and genetic characterisation of Australian camels using microsatellite polymorphisms. *Livestock Science* 129:241–245.
41. Chuluunbat B, Charruau P, Silbermayr K, Khorloojav T, Burger PA (2014) Genetic diversity and population structure of Mongolian domestic Bactrian camels (Camelus bactrianus). *Animal Genetics* 45(4):550-558.
42. Marín J*, et al.* (2008) Mitochondrial DNA variation and systematics of the guanaco (*Lama guanicoe*, Artiodactyla: *Camelidae*). *Journal of Mammalogy* 89:269–281.
43. González BA*, et al.* (2014) Maintenance of Genetic Diversity in an Introduced Island Population of Guanacos after Seven Decades and Two Severe Demographic Bottlenecks: Implications for Camelid Conservation. *PLoS ONE* 9(3):e91714.
44. Hudson RR (1991) Gene genealogies and the coalescent process. *Oxford Surveys in Evolutionary Biology*, eds Futuyma D & Antonovics J (Oxford University Press, Oxford), Vol 7, pp 1-44.
45. Bandelt H, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* 16:37–48.
46. Huelsenbeck J, Ronquist F (2001) MrBayes: Bayesian inference of phylogeny. *Bioinformatics* 17:754–755.
47. Schmidt H, Petzold E, Vingrom M, von Haeseler A (2003) Molecular phylogenetics: parallelized parameter estimation and quartet puzzling. *Journal of Parallel and Distributed Computing* 63:719–727.
48. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* 52:696–704.
49. Cui P*, et al.* (2007) A complete mitochondrial genome sequence of the wild two-humped camel (*Camelus bactrianus ferus*): an evolutionary history of camelidae. *BMC Genomics* 8:241.
50. Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* 24:1586–1591.
51. Rambaut A (2009) FigTree v1.3.1 (University of Edinburgh, Edinburgh, UK).
52. Gaggiotti O, Excoffier L (2000) A simple method of removing the effect of a bottleneck and unequal population sizes on pairwise genetic distances. *Proceedings of the Royal Society Biological Sciences Series B* 267:81–87.
53. Ho SYW*, et al.* (2011) Bayesian Estimation of Substitution Rates from Ancient DNA Sequences with Low Information Content. *Systematic Biology* 60:1-10.
54. Wu H*, et al.* (2014) Camelid genomes reveal evolution and adaptation to desert environments. *Nature Communications* 5:1-9.
55. Bouckaert R*, et al.* (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol* 10(4):e1003537.
56. Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:597–601.
57. Fu Y (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147:915–925.
58. Slatkin M, Hudson R (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* 129:555–562.

59. Nabholz B, Glémin S, Galtier N (2008) Strong variations of mitochondrial mutation rate across mammals—the longevity hypothesis. *Molecular Biology and Evolution* 25:120–130.
60. Luikart G*, et al.* (2001) Multiple maternal origins and weak phylogeographic structure in domestic goats. *Proceedings of the National Academy of Sciences of the United States of America* 98:5927-5932.
61. Beja-Pereira A*, et al.* (2006) The origin of European cattle: evidence from modern and ancient DNA. *Proceedings of the National Academy of Sciences of the United States of America* 103:8113–8118.
62. Troy C*, et al.* (2001) Genetic evidence for Near-Eastern origins of European cattle. *Nature* 410:1088–1091.
63. Naderi S*, et al.* (2008) The goat domestication process inferred from large-scale mitochondrial DNA analysis of wild and domestic individuals. *Proceedings of the National Academy of Sciences of the United States of America* 105:17659–17664.
64. Kimura B*, et al.* (2011) Ancient DNA from Nubian and Somali wild ass provides insights into donkey ancestry and domestication. *Proceedings of the Royal Society Biological Sciences Series B* 278:50-57.
65. Achilli A*, et al.* (2012) Mitochondrial genomes from modern horses reveal the major haplogroups that underwent domestication. *Proceedings of the National Academy of Sciences of the United States of America* 109:8202–8206.
66. Stadler T, Kühnert D, Bonhoeffer S, Drummond AJ (2013) Birth–death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proceedings of the National Academy of Sciences of the United States of America* 110(1):228-233.
67. Uerpmann M, Uerpman HP (2012) Archeozoology of camels in South-Eastern Arabia. *Camels in Asia and North Africa. Interdisciplinary perspectives on their significance in past and present*, eds Knoll E & Burger P (Academy of Science Press, Vienna, Austria), pp 109–122.
68. Heller R, Chikhi L, Siegismund HR (2013) The confounding effect of population structure on Bayesian skyline plot inferences of demographic history. *PLoS One* 8(5):e62992.
69. Beaumont MA (1999) Detecting population expansion and decline using microsatellites. *Genetics* 153(4):2013-2029.
70. Storz JF, Beaumont MA (2002) Testing for genetic evidence of population expansion and contraction: an empirical analysis of microsatellite DNA variation using a hierarchical Bayesian model. *Evolution* 56(1):154-166.
71. Brinkmann B, Klintschar M, Neuhuber F, Hühne J, Rolf B (1998) Mutation rate in human microsatellites: Iinfluence of the structure and length of the tandem repeat. *The American Journal of Human Genetics* 62(6):1408-1415.
72. Rooney AP, Honeycutt RL, Davis SK, Derr JN (1999) Evaluating a putative bottleneck in a population of bowhead whales from patterns of microsatellite diversity and genetic disequilibria. *Journal of Molecular Evolution* 49(5):682-690.
73. Brooks SP, Gelman A (1998) General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* 7(4):434-455.
74. Smith BJ (2007) boa: an R package for MCMC output convergence assessment and posterior inference. *Journal of Statistical Software* 21(11):1-37.
75. Gelman A*, et al.* (2004) *Bayesian Data Analysis* (Boca Raton, USA) 3rd Ed p 675.
76. Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. *Statistical Science* 7(4):457-472.
77. Alley RB, Ágústsdóttir AM (2005) The 8k event: cause and consequences of a major Holocene abrupt climate change. *Quaternary Science Reviews* 24(10–11):1123-1149.
78. Cheng H*, et al.* (2009) Timing and structure of the 8.2 kyr B.P. event inferred from δ18O records of stalagmites from China, Oman, and Brazil. *Geology* 37(11):1007-1010.
79. Kuper R, Kröpelin S (2006) Climate-Controlled Holocene Occupation in the Sahara: Motor of Africa's Evolution. *Science* 313(5788):803-807.
80. Gibbons A (1993) How the Akkadian Empire was hung out to dry. *Science* 261(5124):985.
81. Weiss H*, et al.* (1993) The genesis and collapse of third millennium north Mesopotamian civilization. *Science* 261(5124):995-1004.
82. Kaniewski D*, et al.* (2008) Middle East coastal ecosystem response to middle-to-late Holocene abrupt climate changes. *Proceedings of the National Academy of Sciences of the United States of America* 105(37):13941-13946.
83. Kaniewski D*, et al.* (2010) Late second–early first millennium BC abrupt climate changes in coastal Syria and their possible significance for the history of the Eastern Mediterranean. *Quaternary Research* 74(2):207-215.
84. Wegmann D, Leuenberger C, Neuenschwander S, Excoffier L (2010) ABCtoolbox: a versatile toolkit for approximate Bayesian computations. *BMC Bionformatics* 11:1-7.
85. Csilléry K, Blum MGB, Gaggiotti OE, François O (2010) Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution* 25(7):410-418.
86. Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* 38(6):1358-1370.

87. Rannala B, Mountain JL (1997) Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences of the United States of America* 94(17):9197-9201.
88. Csilléry K, François O, Blum MGB (2012) abc: an R package for approximate Bayesian computation (ABC). *Methods in Ecology and Evolution* 3(3):475-479.
89. Leuenberger C, Wegmann D (2010) Bayesian computation and model selection without likelihoods. *Genetics* 184(1):243-252.
90. Evdotchenko D, Han Y, Bartenschlager H, Preuss S, Geldermann H (2003) New polymorphic microsatellite loci for different camel species. *Molecular Ecology Notes* 3:431-434.
91. Mariasegaram M*, et al.* (2002) Isolation and characterization of eight microsatellite markers in *Camelus dromedarius* and cross-amplification in *C. bactrianus* and *Lama pacos*. *Animal Genetics* 33:385-387.
92. Penedo M, Caetano AR, Cordova K (1999) Eight microsatellite markers for South American camelids. *Animal Genetics* 30:166-167.
93. Obreque V*, et al.* (1998) Characterization of 10 polymorphic alpaca dinucleotide microsatellites. *Animal Genetics* 29:461-462.
94. Lang K, Wang Y, Plante Y (1996) Fifteen polymorphic dinucleotide microsatellites in llamas and alpacas. *Animal Genetics* 27:293.
95. FAO (2011) *Molecular genetic characterization of animal genetic resources - FAO Animal Production and Health Guidelines* (FAO, Rome, Italy).
96. Bulliet R (1975) *The camel and the wheel.* (Columbia University Press, New York, USA) p 327.
97. Heiss J (2012) Caravans from South Arabia: roads and organization. *Camels in Asia and North Africa - Interdisciplinary perspectives on their past and present significance*, eds Knoll E & Burger P (Academy of Science Press, Vienna, Austria), pp 131-139.