Published in final edited form as:

Nat Ecol Evol. 2023 August ; 7(8): 1267–1286. doi:10.1038/s41559-023-02094-w.

North African fox genomes show signatures of repeated introgression and adaptation to life in deserts

Joana L. Rocha[1,2,3,4,✉], **Pedro Silva**[1,2,3], **Nuno Santos**[1,3], **Mónia Nakamura**[1,2,3], **Sandra Afonso**[1,3], **Abdeljebbar Qninba**[5], **Zbyszek Boratynski**[1,3], **Peter H. Sudmant**[4,6], **José C. Brito**[1,2,3], **Rasmus Nielsen**[4,6,7,9,10,✉], **Raquel Godinho**[1,2,3,8,9,10,✉]

[1]CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Campus de Vairão, Universidade do Porto, Vairão, Portugal.

[2]Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Porto, Portugal.

[3]BIOPOLIS Program in Genomics, Biodiversity and Land Planning, CIBIO, Campus de Vairão, Vairão, Portugal.

[4]Department of Integrative Biology and Department of Statistics, University of California Berkeley, Berkeley, CA, USA.

[5]Laboratory of Geophysics and Natural Hazards, Geophysics, Natural Patrimony and Green Chemistry Research Center (GEOPAC), Institut Scientifique, Mohammed V University of Rabat, Rabat, Morocco.

[6]Center for Computational Biology, University of California, Berkeley, CA, USA.

[7]Globe Institute, University of Copenhagen, Copenhagen, Denmark.

[8]Department of Zoology, University of Johannesburg, Auckland Park, South Africa.

[✉]**Correspondence and requests for materials** should be addressed to Joana L. Rocha, Rasmus Nielsen or Raquel Godinho. joana_rocha@berkeley.edu; ras_nielsen@berkeley.edu; rgodinho@cibio.up.pt.

Author contributions

R.G. and J.L.R. conceived and designed the study with R.N. R.G. and J.L.R. initiated the project. J.L.R., N.S., M.N., R.G., Z.B., J.C.B. and A.Q. collected the samples. J.C.B. and A.Q. provided logistical support in North Africa. J.L.R. performed DNA laboratory work and processed the sequencing data. S.A. performed RNA laboratory work. J.L.R. and N.S. processed samples for physiological analysis. R.N. oversaw all statistical and computational analyses, which were performed by J.L.R. (phenotypic and whole-genome data analyses) and P.S. (transcriptomics). P.H.S., R.G. and R.N. helped with editing and reviewing the paper for submission. J.L.R., R.G. and R.N. wrote the paper with input from all authors.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Code availability

All code written for this project is available on GitHub (https://doi.org/10.5281/zenodo.7826260)[132] archived from repository https://github.com/joanocha/HOTFOXES.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41559-023-02094-w.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41559-023-02094-w.

Peer review information *Nature Ecology & Evolution* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

[9]These authors jointly supervised this work: Rasmus Nielsen, Raquel Godinho.

[10]These authors contributed equally: Rasmus Nielsen, Raquel Godinho.

Abstract

Elucidating the evolutionary process of animal adaptation to deserts is key to understanding adaptive responses to climate change. Here we generated 82 individual whole genomes of four fox species (genus *Vulpes*) inhabiting the Sahara Desert at different evolutionary times. We show that adaptation of new colonizing species to a hot arid environment has probably been facilitated by introgression and trans-species polymorphisms shared with older desert resident species, including a putatively adaptive 25 Mb genomic region. Scans for signatures of selection implicated genes affecting temperature perception, non-renal water loss and heat production in the recent adaptation of North African red foxes (*Vulpes vulpes*), after divergence from Eurasian populations approximately 78 thousand years ago. In the extreme desert specialists, Rueppell's fox (*V. rueppellii*) and fennec (*V. zerda*), we identified repeated signatures of selection in genes affecting renal water homeostasis supported by gene expression and physiological differences. Our study provides insights into the mechanisms and genetic underpinnings of a natural experiment of repeated adaptation to extreme conditions.

As arid and hyperarid regions worldwide are increasingly becoming drier and warmer, it has never been more pressing to study adaptive evolution in deserts[1–3]. Foxes of the genus *Vulpes* are a group of canid species with geographic distributions that overlap with some of Earth's harshest deserts[4]. The Rueppell's fox (*V. rueppellii*) is a desert-adapted species widely distributed in hot arid and hyperarid regions of North Africa and the Middle East (Fig. 1a,b). In North Africa, Rueppell's foxes share their range with a more distantly related congener, the fennec (*V. zerda*), and both species are remarkably well-adapted to the extreme temperatures and water scarcity in the Sahara, the largest hot desert on the planet[5]. Fennec and Rueppell's foxes have reduced mass-adjusted basal metabolic rate and non-renal water loss in comparison to canid species living in habitats with water availability, suggesting an increased ability to reduce cutaneous and respiratory (non-renal) water loss[6–8]. Fennec foxes also have increased ability to produce hyperosmotic urine, an adaptation allowing increased water retention in the kidney, though it is unknown if this ability is shared with the Rueppell's fox[9–11].

The sister species to the Rueppell's fox is the red fox (*V. vulpes*), a widely distributed and ecologically plastic species found in a variety of habitats across the Northern Hemisphere (Fig. 1a,b)[12]. Rueppell's and red foxes are both probably of Eurasian origin but differ fundamentally in ecology, morphology, behaviour and physiology, suggesting divergent evolution and adaptations to different ecological niches. However, the two sister species overlap in range at the outer edges of south Asian deserts and northern fringes of the Sahara, where red fox desert morphs have been described[4,5]. The Sahara and its northern outer edges thus represent a remarkable natural laboratory for addressing questions regarding red fox dispersal, interactions with ancestral desert-dwelling species and natural selection processes.

Despite decades of studies on canid adaptive evolution[13–19], our understanding of the spatiotemporal emergence and genetic basis of repeated and successful colonization of

extreme hot desert environments remains limited. Here, we generated 82 whole genomes comprising four *Vulpes* species to gain insight on the origins, evolution and genetic underpinnings of unique and shared fox adaptations to life in the Sahara (Fig. 1). We also gathered gene expression and physiological data to inform the functional effects of genetic variation targeted by selection in desert-dwelling foxes (Supplementary Data and Supplementary Tables 1–3 in Supplementary Information).

Results

Divergence, differentiation and dispersal into North Africa

To uncover the origins and evolutionary history of desert adaptation in foxes, we sequenced the complete genomes (~4.5× coverage) of 30 North African and 18 Eurasian red foxes and 24 North African Rueppell's foxes (Fig. 1a,b; Methods; Supplementary Table 1). North African red foxes were sampled along a climatic gradient ranging from dry subhumid to hyperarid regions, thus providing a broad representation of red fox diversity across its geographic range in North Africa. In addition, we sequenced ten genomes belonging to two divergent North African species: desert specialist fennec fox ($n = 5$) and the pale fox ($n = 5$; *V. pallida*), a species endemic to the semi-arid Sahel (Fig. 1a,b). We analysed our data together with data from other canid species (Supplementary Table 2), using a suite of statistical methods that take genotype uncertainty into account by working directly on genotype likelihoods (Methods)[20]. The genome assemblies of the arctic fox[21] and domestic dog (*Canis lupus familiaris*)[22] were both used as reference in parallel analyses to address issues of reference biases (Methods; Supplementary Fig. 1).

Average whole-genome (Fig. 1c,d and Supplementary Fig. 2) and mitochondrial DNA-based (Extended Data Fig. 1 and Supplementary Figs. 3 and 4) phylogenetic analyses and ancestral range reconstruction for several *Vulpes* species supported that Rueppell's and red foxes are sister species and confirmed the phylogenetic placement of the pale fox as outgroup to the other groups of *Vulpes*[12] (Supplementary Tables 4 and 5). These analyses further showed that Rueppell's foxes and red foxes have a Eurasian origin with independent North African colonizations and share a common ancestor with Asian endemic foxes and the Eurasian and North American arctic fox clade, after they split from the fennec (Fig. 1c and Extended Data Fig. 1).

Though past studies have placed the Middle East as the cradle of modern red fox genetic diversity[23], little is known of how the genetic diversity and past population dynamics seen in Middle Eastern and European red foxes[24] relates to those currently inhabiting the African continent. Principal component and admixture analyses first separated Eurasian and North African red foxes and then Eurasian populations into Europe and Asia (Fig. 2a and Supplementary Fig. 5a). We found higher levels of admixture between Middle Eastern and European red foxes than between Middle Eastern and North African red foxes and a single red fox from the Sinai Peninsula exhibiting a small proportion of North African ancestry (Figs. 2b and Supplementary Fig. 6). Consistent with this, the levels of red fox genome-wide differentiation were greater between North Africa and the Middle East (global $F_{st} = 0.13$) and Europe (global $F_{st} = 0.19$), than between the Middle East and Europe (global $F_{st} = 0.09$). This suggests more limited gene flow between North Africa and Eurasian red foxes

than among Eurasian populations, consistent with red fox isolation in North Africa (Fig. 2a). Higher rates of linkage disequilibrium (LD) decay in North African compared to Eurasian red foxes seem to suggest a bottleneck in population size affecting Eurasian red foxes and/or more admixture LD caused by population structure (Fig. 2c). We found evidence for genetic continuity throughout the Sahara within North African Rueppell's fox, with eastern (Egypt) and western (Mauritania and Western Sahara) individuals clustering together (Supplementary Figs. 5b and 6).

To explore when Rueppell's fox and red fox diverged and infer their joint demographic history in North Africa, we used GADMA[25] with dadi[26]. The best-fitting model estimated a divergence time of ~576 thousand years ago (ka), with subsequent non-zero, two-way asymmetric migration (Fig. 2d). Despite the inherent uncertainty in the genome-wide mutation rate estimate and variance in generation time we found divergence time estimates between the two species to be in accordance with the earliest fossil records of red fox[27] and within the coalescent time interval of 476–788 kyr estimated with mitogenomes from several *Vulpes* species (Methods; Supplementary Table 5). The absence of admixture between North African Rueppell's fox and red fox in admixture plots (Fig. 2b; $K = 2$) suggests that the interspecific gene flow is not a recent phenomenon.

We next set out to resolve the demographic history of red fox dispersal into North Africa. The model with highest composite likely-hood found using GADMA and dadi dates the divergence of Eurasian and North African populations to ~78 ka (Supplementary Table 7), centred around a Saharan humid period or wet phase that ended 72 ka (refs. 28–30). In North Africa, several different 'green Sahara' periods spanning marine isotope stages 5e, 4 and 3 are thought to have enabled human expansion inside and outside the continent and across an otherwise inhospitable desert[28–32]. Such a humid period could also have enabled an ancestral red fox population to colonize North Africa and expand to regions that were previously uninhabitable for the species, thus setting the stage for divergence from Eurasian red foxes and introgression with Rueppell's fox as subsequent climate change caused population isolation (Fig. 4a). Consistent with this, the best-fitting model supports a population bottleneck in the red fox North African population after divergence from Eurasian populations, followed by a recent expansion. The model also supports the maintenance of the ancestral population size in Eurasia immediately after divergence from North Africa followed by asymmetrical two-way migration (Supplementary Table 7), suggesting population movements between Eurasia and North Africa during subsequent African humid periods. The absence of substantially admixed individuals in admixture plots (Fig. 2b) suggests that this is not a recent or ongoing phenomenon and, thus, red foxes have probably remained isolated ever since the Sahara attained its current desert profile.

Ancient introgressive-hybridization events

Past introgression among ecophysiologically similar and geographically overlapping desert-dwelling foxes (Fig. 1a,b) could have facilitated evolutionary responses to climate change in arid regions. Motivated by this, we used *D* statistics[33] and TreeMix admixture graphs[34] to test for introgression among North African foxes (Methods). We found a significant excess of allele sharing between North African red foxes and Rueppell's fox relative to

Eurasian red foxes (Fig. 3a, top, Extended Data Fig. 2, Supplementary Figs. 7 and 8 and Supplementary Table 6) providing evidence of past interspecific gene flow consistent with demographic inference (Fig. 2d and Supplementary Table 7). Remarkably, the same analyses further suggested introgression between Rueppell's fox and the highly divergent fennec fox (Fig. 3a, bottom). A sliding window analysis of the introgression statistic f_d (ref. 35) revealed several genomic regions shared between North African Rueppell's and red foxes (Fig. 3b, top) and a large ~25 megabase (Mb) region shared between Rueppell's fox and fennec (Fig. 3b, bottom). To further examine the evolutionary history of these regions, we compared the estimated average genome-wide tree with tree topologies for genes with the most extreme values of the empirical distribution for f_d (above the 99th percentile; Fig. 3c,d, Supplementary Tables 8 and 9, Extended Data Figs. 3 and 4 and Supplementary Fig. 9) and for the 25 Mb region (Fig. 3e,f, Supplementary Tables 10 and 11 and Extended Data Fig. 5).

In contrast to the topology observed for the genome-wide tree, nearly all North African red fox individuals appeared more closely related to Rueppell's fox than to other red foxes at *SLC6A16* (Fig. 3c), encoding a member of a family of neurotransmitter transporters involved in cellular response to oxidative stress[36,37], which shows the second strongest signal of introgression (Fig. 3b, top, Extended Data Fig. 4 and Supplementary Fig. 9). Although this discordance is consistent with introgression, it could also be explained by incomplete lineage sorting (ILS)[38]. To disentangle these two scenarios, we compared levels of genome-wide differentiation (F_{ST}) and genetic divergence (d_{XY}) between Rueppell's fox and North African red fox with those estimated at shared regions. Consistent with an introgression scenario, d_{xy} levels across *SLC6A16* were lower between North African red fox and Rueppell's fox (<0.5% of the genome-wide empirical distribution) (Fig. 3d). Additionally, F_{ST} levels across *SLC6A16* were lower between Rueppell's fox and North African red fox and higher between the North African red fox population and Eurasian red fox (Supplementary Figs. 10–12). We also found similar gene-species tree discordances for other genes associated with glycogen storage and starch digestion (*MGAM*) and pigmentation (*HPS5*), although with less convincing patterns of reduced F_{ST} and d_{XY} between North African Rueppell's fox and red fox that would otherwise allow us to favour an introgression scenario over ILS (Supplementary Figs. 10–12).

Unexpectedly, we found a very unusual topology at the *CRADD* gene tree (top outlier in Fig. 3b, top), in which some Eurasian red foxes and the arctic fox formed a highly divergent outgroup to other *Vulpes* (Extended Data Fig. 3). While highlighted as a putative locus of introgression between Rueppell's fox and North African red fox, the high f_d score at the *CRADD* region instead represents deeply divergent alleles more closely related between arctic and red foxes and predating all extant members of *Vulpes* (see Extended Data Fig. 3a for analysis with additional members of Canidae). The observed gene-species tree discordance in Eurasian red fox could either result from ancient introgression with a highly divergent as-yet-unidentified canid lineage or, alternatively, reflect an ancestral polymorphism maintained by balancing selection, for example spatially varying selection. This hypothesis would, however, require some level of recombination suppression in the region[39].

Next, we explored the genomic region shared between fennec fox and Rueppell's fox (Fig. 3b, bottom). Enrichment analysis on the entire 25 Mb outlier region identified several gene ontology (GO) terms that were over-represented (Supplementary Table 11; false discovery rate (FDR) $q < 0.05$). The functions of individual genes include thyroid hormone metabolism (*SULT1B1*), stress response and metabolic fasting (*SULT1D1*), temperature homeostasis, regulation of heat generation and feeding behaviour (*NMU*) and pigmentation (*KIT*), all related to traits of possible importance to life in arid environments. The gene tree for this region unambiguously grouped Rueppell's fox as sister species to the fennec (Fig. 3e and Extended Data Fig. 5). However, the pairwise divergence between the two desert-dwelling species is not much different from the genome-wide average (Extended Data Fig. 6 and Supplementary Fig. 13) and genome-wide introgression between fennec and Rueppell's fox is no longer significant when the 25 Mb region is removed (Supplementary Table 6), suggesting that introgression did not happen directly from fennec to Rueppell's fox but from another fennec-related donor species (see Extended Data Fig. 1 for mtDNA species relationships). Sliding window analyses at the affected 25 Mb region further revealed that Rueppell's fox and red fox revert to being sister taxa at the *KIT* gene (Extended Data Fig. 7). *KIT* is a well-known case of an autosomal gene with paralogues located on red fox supernumerary chromosomes (Bs), which may vary in number across individuals[40]. Consistent with this, we found that red fox and Rueppell's fox individuals carry different numbers of copies of *KIT* that were strongly correlated with the copy number variation pattern of some other autosomal genes, which probably are also located on the same supernumerary chromosome (Supplementary Fig. 14). Under a scenario of ancient interspecific gene flow, we would expect *KIT* to be highly polymorphic in Rueppell's fox, as the result of divergence between ancestral fennec-like autosomal alleles and the unaffected supernumerary alleles, which both map to the same region in the arctic fox and dog reference genomes. Consistent with this, Rueppell's fox showed a peak in genetic diversity in *KIT* not found in other foxes (Extended Data Fig. 7). Past correspondence analysis using chromosome painting map the affected 25 Mb region from chromosome 13q of the domestic dog to chromosomes 2p in red fox, 12p in arctic fox and part of chromosome 9 in fennec (Fig. 3f)[41]. We suggest that a major structural rearrangement, such as the translocation of a chromosomal arm, could explain how loci at this divergent region were maintained in strong LD. We also note that the data do not allow us to distinguish between introgression and the hypothesis of an ancient trans-species polymorphism, with a fennec-related divergent recombination-suppressed haplotype seemingly becoming fixed in Rueppell's fox and repeatedly lost in arctic and red foxes (Extended Data Fig. 5).

Red fox dispersal into Africa followed by climate adaptation

Red foxes living in increasingly aridenvironmentsat theSaharan northern edge for numerous generations may have faced unique selective pressures, offering a unique opportunity to understand how climatic oscillations over the last ~78 kyr promoted and shaped adaptive evolution (Fig. 4a). To investigate signatures of positive selection specific to North African red fox, we looked for highly differentiated genomic regions (population branch statistics, PBS) and/or regions with reduced diversity (α) in this population relative to Eurasian red fox and Rueppell's fox. A sliding window analysis revealed several genomic regions with high PBS and $-\log_{10}(\alpha)$, hinting at selection in North African red fox (Fig. 4b,c). We

then focused on the regions with the most extreme values of the empirical distribution (above the 99.95% percentile) for both methods and identified overlapping genes (Fig. 4d and Supplementary Tables 12–14). We note that similar approaches to these have been successful in identifying genes targeted by natural selection underlying unique adaptations to extreme selective pressures in other mammalian populations with similar sample sizes[13,42,43]. Then, we examined expression patterns in blood samples (n = 3 Rueppell's fox, n = 2 red fox in North Africa, n = 5 in Iberia) for the major outlier genes (Supplementary Tables 3 and 15 and Supplementary Figs. 15 and 16).

Remarkably, the second largest increase in $-\log_{10}(\alpha)$ was *SLC12A2* (Fig. 4c), a gene encoding a sodium chloride transporter that plays a vital role in the regulation of ion balance and cell volume and has been associated with a 70% decrease in saliva secretion and sweat in knockout mice, as well as dry mouth disease and impaired salivary and sweat glands in humans[44,45]. We additionally found that *SLC12A2* is significantly less expressed in North African red fox and Rueppell's fox compared with Eurasian red fox ($\log_2$ fold-change = −1.13, adjusted $P_{(\text{North Africa versus Iberia})}$ = 0.00113; Fig. 4e and Extended Data Fig. 8). The strong selection signal in *SLC12A2* in North African fox could thus potentially be related to changes in expression at this gene, possibly resulting in reduced amount of evaporative water loss through panting, the main mechanism for cooling in canids or sweating. Additionally, enrichment analysis on top regions with reduced diversity identified several biological processes in North African red fox related to water homeostasis and regulation of water loss via skin, not found in Eurasian populations, though these were non-significant after correcting for multiple testing (FDR $q > 0.05$; Supplementary Table 14 versus Supplementary Tables 16–18).

While some of the selection signals unique to North African red foxes may be related to other environmental factors, such as local pathogens (for example, *CD163*, ranked first in $-\log_{10}(\alpha)$); Fig. 4c and Extended Data Fig. 8), many top hits harboured candidate genes associated with traits of potential relevance to life in arid regions. For instance, a highly differentiated and low diversity gene (ranked third in $-\log_{10}(\alpha)$; Fig. 4) is *TRPA1*, encoding for a transient receptor potential (thermoTRP) channel protein that is gated by noxious cold or heat and acts as sensor for pain and environmental irritants[46–48]. Several thermoTRP sensors associated with reduced temperature sensitivity and increased cold tolerance have been reported to be under selection in arctic sled dogs[13], mammoths[49] and humans in high-latitude environments[50]. North African foxes are nocturnal, with denning behaviour during the day[4] and, though arid regions in North Africa are classified as 'hot', they may experience extreme daily temperature oscillations[51]. We thus suggest that reduced thermosensitivity to both extreme cold and heat can be adaptive in arid regions with high isothermality and that the genetic variation of North African red fox probably has been targeted by selection on phenotypes linked to bidirectional thermoception. Another candidate gene under selection in North African red fox is *THADA* (ranked 14th in PBS; Supplementary Tables 12 and 13), a gene involved in heat production that has been widely associated with evolution of reduced metabolic rates in warmer climates to prevent overheating and climate adaptation in humans[52]. We additionally found a region harbouring *ZNF516*, a gene involved in cold-induced thermogenesis[53] to be under selection in Eurasian red fox (Extended Data Fig. 9 and Supplementary Table 16).

Rueppell's fox and fennec physiology and genetic adaptation

Our results suggest that North African red foxes have unique adaptations related to tolerance of extreme temperatures and evaporative water retention that may have increased survival chances in the increasingly arid regions at the Saharan fringes over the past ~78 kyr, providing a prime example of recent adaptation to rapidly changing climates (Fig. 4). This raises the question of what signatures of selection are to be found in other fox species adapted to life at the core of deserts for increasingly longer time periods such as Rueppell's fox (476–788 ka) and the fennec (4.8–6.2 million years ago (Ma)). To address this, we performed a genome-wide selection scan using a PBS-like method implemented in Ohana[54] that can be applied to several ancestry components. We ran Ohana in sliding windows to detect regions that strongly deviate in Rueppell's fox and fennec from the genome-wide covariance structure using a likelihood ratio test and we then focused on the top differentiated genes (above the 99.95th percentile; Fig. 5a,b). We found signals of selection in genes related to renal water homeostasis in both fennec (*KIRREL1*, *CACNA1H*) and Rueppell's fox (*WFS1*), genes related to pigmentation (*IRF4*) in the fennec and related to thyroid hormone metabolism (*SLCO4A1*), temperature homeostasis (*NTSR1*), thermotaxis (*OPN4*) and cellular response to ultraviolet radiation (*PTPRK*) in Rueppell's fox (Supplementary Tables 19–21).

Deficiency of *WFS1* in knockout mice leads to impairment of vasopressin (AVP) secretion and decreased ability to concentrate urine with low volume under water deprivation[55]. Remarkably, *WFS1* is a top candidate gene for selection in Rueppell's fox (ranked second) and is also overexpressed in Rueppell's fox compared to red fox ($\log_2$ fold-change = 1.83, adjusted $P_{(\text{Rueppell's versus red foxes})} = 0.0013$) (Fig. 5c). AVP plays critical roles in water reabsorption at the kidneys, resulting in increased urine-concentrating ability and reduced urine volume. Lineage-specific differences in this hormone have been associated with increased urine-concentrating abilities and other potential adaptations related to water retention in desert-dwelling mammals, with higher levels in some desert-dwellers compared to non-desert counterparts[11,56–58]. It is, therefore, possible that selection may have targeted AVP or AVP-related pathways in Rueppell's fox to a greater extent than red fox. To investigate further, we tested if Rueppell's fox had higher water retention abilities compared to red fox from North Africa and Eurasia by examining physiological parameters using blood samples comprising 11 red foxes, 13 Rueppell's foxes and 10 fennecs (Supplementary Tables 3, 22 and 23, Extended Data Fig. 10 and Supplementary Figs. 17 and 18). We used a linear model that included sex, weight, body mass index and an indicator of whether the individuals were caught free-ranging or captive as covariates (Supplementary Table 23). We found that levels of copeptin, a biomarker for AVP, were significantly higher in Rueppell's fox compared to red fox (Fig. 5d) and that physiological differences are more likely attributable to differences between species, rather than the result of sex, body mass or different plastic responses in the status of captivity versus free-ranging animals (Supplementary Table 23 and Extended Data Fig. 10). Notably, the maximum urine osmolality observed for Rueppell's fox was 5,126 mOsm kg^{-1}, whereas maximum urine osmolality observed for red fox was 3,027 mOsm kg^{-1} (Supplementary Fig. 18). We also note that the highest urine-concentrating ability for one Rueppell's fox in our dataset is comparable to that observed in kangaroo rats[59] and slightly above the maximum recorded

value for the fennec fox (4,022 mOsm kg^{-1}), which are remarkably adapted to subsist without water[9]. We thus suggest that selection acting on mutations changing expression of *WFS1* may explain increased vasopressin levels and urine-concentrating abilities in Rueppell's fox. Notably, the *WFS1* gene is within the highly divergent 25 Mb region found in Rueppell's fox genome, which suggests that the species may have inherited increased capacity to retain water under dehydration from a desert-dwelling ancestor (Fig. 3e,f).

The top candidate gene for selection in fennec, *KIRREL1* (Fig. 5b), encodes a protein produced in kidney podocytes that plays an important role in the normal development and function of the glomerular permeability (Fig. 5e and Supplementary Tables 20 and 21). In humans, mutant KIRREL1 proteins fail to localize the podocyte cell membrane, resulting in nephrotic syndrome, while knockout mice for this gene exhibit loss of glomerular filtration function[60]. Notably, fennecs also have higher AVP levels and seem to have higher urine-concentrating ability than red foxes (Fig. 5d, Supplementary Fig. 18 and Supplementary Tables 22 and 23). *KIRREL1* is therefore a likely candidate gene that may be related to fennecs' increased kidney function. Rueppell's fox and fennecs also had significantly lower levels of thyroid hormone thyroxine (T4) than did red fox (Extended Data Fig. 10 and Supplementary Table 23), which potentially underlie previous evidence of reduced metabolic rate and non-renal water losses in these species[7,10,61]. The selective advantages of lower metabolism in deserts would include reduced energy demand, low energy expenditures, lower production of metabolic heat and subsequent reduced need for evaporative cooling[7,62,63]. While we did not find thyroid-related genes under selection in fennec, we found a solute carrier for T4, *SLCO4A1*, to be under selection in Rueppell's fox (ranked first). We also noted that the 25 Mb region probably introgressed into Rueppell's fox was enriched for sulfotransferase genes (*SULT*s, FDR $P = 0.017$), one of which involved is in thyroid hormone metabolism (Supplementary Table 11). These, and many of the candidate genes for desert adaptation identified in this study, remain promising targets for future studies (Supplementary Text).

Discussion

North African foxes comprise an entirely new genomic system in which to examine adaptation to extreme conditions in distinct lineages of the same genus. In this study, we combined genomic resources with ecophysiological and expression data to elucidate how climate adaptation has shaped the genomes of these iconic species. By modelling the demographic histories of Rueppell's fox and red fox we were able to demonstrate that Rueppell's fox differentiated from the red fox ~576 ka and evolved to inhabit some of the most extreme hot deserts on Earth, including the Sahara, while red fox radiated from Eurasia into different regions of the globe, having later dispersed into North Africa around 78 ka. We further show that the spatiotemporal emergence of North African red fox following divergence from Eurasian populations probably relates to the oscillating palaeoclimatic history of the Sahara[30]. North African fox species thus represent evolutionary replicates of a natural experiment of desert adaptation at different timescales.

Research of the last decade has demonstrated that introgression is a widespread phenomenon in nature and often of great adaptive significance due to its potential to rapidly introduce

combinations of beneficial alleles[38,64–70]. There is also a growing body of empirical evidence showing that introgression is highly pervasive in canids[71,72], with remarkable case studies of species such as coyotes and grey wolves[73] and Tibetan and Himalayan wolves and dogs[14] carrying genetic material introgressed from highly divergent and unidentified lineages. In this work, we identify introgressive-hybridization events among geographically overlapping fox species facing similar environmental challenges related to life in arid and hyperarid regions, suggesting that introgression plays an important role in desert adaptation, similarly to what has previously been reported in seasonal and high-altitude environments[14,64,74,75]. The recent dispersal of red fox into the increasingly arid regions of North Africa was probably facilitated by introgression from Rueppell's fox of at least one region harbouring *SLC6A16*. We additionally show that Rueppell's foxes carry a massive ~25 Mb long block of genes harbouring highly divergent alleles in strong LD, which probably represents an inversion or a translocation, introgressed from an as-yet-undefined fox lineage closely related to the fennec. We note, however, that a trans-species polymorphism resulting from ILS of a structural rearrangement could also explain the observed pattern, although this would require the independent loss of the fennec-related recombination-suppressed haplotype in arctic and red foxes and in different species from the South African (Cape fox, *V. chama*), North American (kit fox *V. macrotis* and swift fox *V. velox*) and Asian (Bengal's fox *V. bengalensis*, Tibetan fox *V. ferrilata* and corsac fox *V. corsac*) fox clades not featured in this study. Future sequencing efforts on reference-quality haplotype-resolved genomes for these *Vulpes* species will help to resolve the structure of this region in Rueppell's foxes and test, for example, if it is phylogenetically closer to the fennec or to the Blanford's fox (*V. cana*), the only extant long-diverged sister species to the fennec which overlaps in range with Rueppell's fox in the Middle East[76]. We would also expect the three Asian species to be phylogenetically closer to red foxes at this region, then followed by the arctic fox clade and only then by Rueppell's fox and the fennec. Genome-wide scans for signatures of adaptation further revealed that the gene with second strongest signal of selection in Rueppell's fox, *WFS1*, is within this 25 Mb region and shows a significant increase in expression in Rueppell's fox. Knockout mice for *WFS1* suffer from an impairment of vasopressin (AVP) that prevents them from concentrating urine under water deprivation[55] and this gene may, therefore, underlie Rueppell's fox increased ability to concentrate urine via increased AVP production. Other candidate genes identified herein remain promising targets for future functional studies identifying the exact molecular and physiological mechanisms of action.

We find another striking signal of deep introgression at *CRADD*, a gene associated with body size differences and presumably under selection in arctic sled dogs[13] and humans[77]. The occurrence of deeply divergent alleles shared between Eurasian red fox and arctic fox at this locus could reflect two introgression events involving an unknown out-group to all extant members of *Vulpes* and both arctic fox and Eurasian fox: either as independent introgressions from the unknown lineage into each of the two species or only into one species which then served as an intermediate donor to the other. We note that similar introgression scenarios have been proposed to explain deeply divergent *EPAS1* alleles in Tibetan and Himalayan wolves and dogs[14]. Alternatively, this result could reflect an ancestral polymorphism predating the origin of all extant *Vulpes* and maintained across

multiple speciation events for millions of years by evolutionary forces such as balancing selection, for example in the form of spatially varying selection, acting on a structural variant keeping alleles tightly linked via recombination suppression[70,78,79]. Future long-read sequencing of additional *Vulpes* species and red fox and arctic fox populations from diverse localities across their geographic range will enable further investigations on the geographical extent of gene-species tree discordances and to test for a model of spatially varying selection where different environmental settings select for genotypes from a common gene pool matching local conditions[79]. This will not only help elucidate the unsorted history of this genomic region but further allow for a rigorous assessment on the extent and long-term impacts of stochastic lineage sorting and interspecific gene flow throughout the history of diversification of the genus.

Overall, our selection scans suggest that some aspects of evolutionary adaptation in North African foxes depend on temporal scale, with selection targeting renal water retention and kidney function in older core desert-dwelling species such as Rueppell's fox and fennec, to a higher degree than what we observe in desert-edge dwelling North African red foxes which, presumably, have more access to water[5,12]. We find that the adaptation of North African red fox to increased aridity via reduced non-renal water loss through mechanisms such as panting and sweating[44,45], probably was facilitated by selection on genes such as *SLC12A2*. Signatures of selection on genes related to temperature perception (*TRPA1*) and thermoregulation (*THADA*) mirrors previously reported climate adaptations of arctic sled dogs (*TRPC4* and *TRPV2*), mammoths (*TRPA1* and other thermoTRPs) and human populations (*THADA* and *TRPM8*) in colder climates[13,49,50,80].

Desertification of the Sahara and expansion of arid regions worldwide are expected to dramatically increase, which may further intensify pre-existing selective pressures on desert-dwelling systems[3,81]. As terrestrial biodiversity continues to decline it has never been more pressing to move beyond the standard model systems to study the mechanisms driving dispersal movements and adaptation over the last Quaternary climatic oscillations in natural populations, to elucidate the ways species may adapt to ongoing and future climatic changes[82,83]. Here, we find that introgressive-hybridization events combined with selection for tolerance to extreme temperatures and increased capacity to preserve water may have enabled the survival of North African foxes in desert regions at different evolutionary scales. We propose that genetic variation shared among core desert species is an important component of recent or ongoing adaptation to rapidly changing climates in foxes living at the edge of deserts. Our findings may thus help address questions concerning the persistence and adaptive capacity of biodiversity currently challenged by desertification.

Methods

Blood and tissue collection

We obtained blood and/or tissue samples from 97 individual foxes of the genus *Vulpes* from specimens captured in the field or held in captivity and from road-killed animals collected in North Africa, the Middle East and Europe (Supplementary Tables 1 and 3 and Supplementary Data). North African fox captures took place in two different areas of Morocco where three *Vulpes* species—red fox (*Vulpes vulpes*), Rueppell's fox (*Vulpes*

rueppellii) and fennec fox (*Vulpes zerda*) are known to occur: the Draa Valley at the region of Guelmim-Es Semara and the Western Sahara at the region of Dakhla-Oued Ed-Dahab. Before trapping, a preliminary field inspection was performed in each region surveying the terrain in search for carnivore signs. With the help of local Saharawi people, we were able to gather information on areas potentially visited by foxes to set camera-traps, namely through personal observations of live and/or road-killed animals, compatible dens, tracks and faeces.

Following the survey, North African foxes were captured using a combination of double-door Tomahawk cage traps ($n = 10$), Collarum snares ($n = 1$) and padded foothold traps ($n = 3$) set across different line transects ranging between 8 and 10 km. The three types of trap were combined to maximize capture success and selected on the basis of efficiency and selectivity in capturing foxes while ensuring the safety of captured animals. All traps were set using odour-free tools and gloves to prevent human scent contamination and Tomahawks were camouflaged with local vegetation to increase the likelihood of individual interest in cage traps. In total, 361 trap × nights (14 with Collarum, 132 with padded foothold traps and 215 with cage traps) were performed in Dakhla-Oued Ed-Dahab and 340 trap × nights (100 with Collarum and 240 with cage traps) were performed in Guelmim-Es Semara. Twelve cage traps were prebaited for one night before being activated. Both Collarum and padded foothold traps were set coupled to remote satellite trap-alarms. Traps were monitored twice every day, in the early morning and late afternoon and baited at the second monitoring with canned sardines (cage traps), Collarum bait (Collarum traps) and a combination of canine scents (leg-hold). Additional free-ranging red foxes ($n = 5$) captured in the region of Alto Minho (Portugal) and Asturias (Spain), following similar logistics to our North African sampling, were incorporated in our data collection. To maximize sampling size, 17 captive foxes available at public and private zoos and biological parks were also sampled (Supplementary Tables 1 and 3).

A total of 36 animals (free-ranging or captive-held; Supplementary Table 3) were anaesthetized by intramuscular injection of a mixture of ketamine (catalogue no. 1170069; Imalgene) and medetomidine (catalogue no. 7418335; Domitor) by a certified wildlife veterinarian. The animals were weighed, sexed and measured for age (juvenile <1 yr versus adult >1 yr), body, hindfoot, tail and ear length. The health status of the individual was checked every 10 min during the whole procedure. Blood was recovered by venipuncture of the arm veins 30 min after administration of the anaesthetic and preserved in both clotting and EDTA tubes. Serum/plasma was separated from blood cells through centrifugation at 1,500*g* for 10 min for biochemistry measurements (plasma/serum) and genomic DNA extraction (cells). When available, urine was retrieved using a sterilized catheter. Blood and urine samples were kept refrigerated ~8–72 h after captures. If the body mass of the animal was sufficiently high, additional whole-blood was collected and stored in PAXgene Blood RNA tubes (catalogue no. 762165; QIAGEN) for RNA extraction (Iberian red fox, $n = 5$; North African red fox, $n = 2$; Rueppell's fox, $n = 4$). These tubes were stored at room temperature (RT) for no more than 2 d, followed by refrigerated storage at −20 and then −80 °C). Free-ranging (wild) individuals were additionally ear punched using an 8 mm livestock biopsy ear punch. The fresh tissues were stored at RT in 2 ml Eppendorf tubes with 96% ethanol for DNA extraction. After sampling, free-ranging animals were marked with a microchip (Freevision Technologies) to control for potential recapture and

individuals were safely released into their environment. Chemical immobilization was reversed by the intramuscular injection of atipamezole (Revertor) after administration of an anti-inflammatory drug.

To increase our sampling set for genomic analysis, we made use of an additional set of 61 tissue samples collected from road-killed animals during past field expeditions conducted in North Africa, Europe and the Middle East throughout November 2003 to January 2018. This set of samples consisted of 16 Rueppell's foxes, 40 red foxes and 5 pale foxes (*V. pallida*), all preserved in 96% ethanol (Supplementary Table 1). The geographical location of all samples included in this study were recorded with a global positioning system on the WGS84 datum.

All foxes captured in this study were sampled under a scientific permit signed by the Moroccan entity Le Haut Comissaire aux Eaux et Forets et a la Lute Contre la Desertification and under a collaboration protocol between the Research Centre in Biodiversity and Genetic Resources (CIBIO-InBIO, University of Porto, Portugal) and the Institute Scientifique de Rabat (Morocco) and protocols between CIBIO-InBIO and the respective Zoos and Biological Parks from which samples were obtained. Animals were allowed to recover under optimized safety conditions and released unharmed, not showing any signs of trauma, as attested by the recapture of two individuals. All sample exportation/importation was performed under declaration of the Institute for the Conservation of Nature and Forests and CITES listed specimens were accompanied with an appropriate CITES permit (18PTLX00948R, 17US697233/9 and DE-0072/2018).

Phenotyping

In total, 34 individual *Vulpes* samples were phenotyped in this study (n = 13, Rueppell's fox; n = 11, red fox; n = 10, fennec; Supplementary Table 3). This comprised 21 males (61.7%) and 13 females (38.2 %) evenly distributed within species, most of which were adults (Supplementary Table 3). Blood samples were analysed for a total of 12 physiological parameters at a commercial laboratory (Inno, Braga, Portugal), including hormones total thyroxine (T4) and aldosterone, blood metabolites (urea, uric acid and creatinine), electrolytes (sodium, potassium and chloride), protein (albumin) and fat (cholesterol), plasma and urine osmolality (mOsm kg^{-1}). Additionally, a Sandwich Canine Copeptin ELISA Kit (catalogue no. MBS007635; MyBioSource) was performed to measure copeptin levels, a biomarker for vasopressin (AVP), following the manufacturer's instructions. AVP is the main hormone responsible for regulating urine-concentrating ability. The upregulation of this hormone stimulates the insertion of water channels (aquaporins) into the collecting tubule of the kidney, increasing water reabsorption and reducing urine volume[85–87]. Copeptin is a 39-amino acid glycopeptide that comprises the C-terminal part of the AVP precursor (CT-proAVP). The two are released in equimolar amounts into the bloodstream by prevasopressin and, while AVP decays rapidly ex vivo to nearly undetectable levels, leading to pre-analytical errors, copeptin remains stable in plasma after collection, therefore constituting an accurate surrogate marker for AVP[88,89]. As such, values for copeptin ex vivo correlate well with the values of vasopressin in vivo.

DNA and RNA extraction

From our collection of 97 individuals, 89 individual fox specimens were selected for whole-genome sequencing (Supplementary Table 1). These comprised sister species Rueppell's fox (n = 29) and red fox (n = 50; 32 from North Africa + 18 from Eurasia), as well as two outgroup species, the fennec (n = 5) and pale fox (n = 5) (Supplementary Table 1). Genomic DNA was extracted using the Qiagen DNeasy Blood & Tissue Kit (catalogue no. 69504, QIAGEN), following the manufacturer's protocol. Before the incubation step of the extraction protocol, 4 µl of 100 mg ml^{-1} RNase A was added and incubated for 10 min at RT to ensure complete removal of RNA. Road-kill samples were left at RT overnight in 1 ml of PBS buffer to clean the tissues and deplete them of potential bacterial contaminants before DNA extraction. Because elution buffers containing high concentrations of EDTA might inhibit the end-repair step of library preparation, DNA samples were eluted in EB buffer (10 mM Tris-Cl, pH 8.5) that comes with MinElute PCR Purification Kit (catalogue no. 28004, QIAGEN), as opposed to the AE buffer (10 mM Tris-Cl, 0.5 mM EDTA; pH 9.0) that comes with the Qiagen DNeasy Blood & Tissue Kit. The quality and concentration of DNA extraction was assessed for each sample with Nanodrop (A260/280 and A260/230 ratios; ThermoFisher Scientific) and Qubit dsDNA Kit (catalogue no. Q32850, ThermoFisher Scientific) quantification methods and DNA length integrity was checked via gel electrophoresis. RNA was extracted using PAXgene Blood RNA Kit (catalogue no. 762174, QIAGEN) for 11 specimens of Rueppell's fox (n = 4) and North African and Iberian red fox (n = 5 Iberia; n = 2 North Africa; Supplementary Table 3). All samples had sufficiently high concentration and RNA integrity number (RIN) scores (>8.2), except for one individual (one Rueppell's fox sample, VR03; RIN = 6.8; Supplementary Table 3),

Library preparation and sequencing

Aliquots of 100 µl containing 0.5–2.5 µg of DNA were sheared to fragment sizes ranging from 200 to 600 base pairs (bp) using a Bioruptor sonication device (Diagenode). Sonication protocols varied according to DNA quality. Fresh tissue and blood samples were sonicated with seven cycles of 15 s of shearing (on) followed by 90 s of pause (off). Road-kill samples received the same timing of on and off but varying cycles from one to six based on the degree of fragmentation. After shearing, samples were selected for an average fragment size of 350 bp using Agencourt AMPure XP beads (catalogue no. 082A63881, Beckman Coulter) as in the double size selection step of Illumina's Truseq DNA PCR-free sample preparation protocol. Highly degraded road-kill samples for which the large fragment removal step was not necessary were cleaned for fragments <200 bp using low fragment single size selection with 1.6× of Agencourt AMPure XP bead suspension. Sequencing libraries were built in batches of 20 samples at a time using the KAPA Hyper Prep Kit PCR-free protocol (catalogue no. KK8505, KAPA Biosystems) following the manufacturer's instructions, adjusting overall volumes of each reaction step to half. Each library preparation session included a negative control. Libraries were run on a 2% agarose gel to identify potential adaptor dimers, which were removed through 1× bead-clean up, quantified under Kapa Library Quantification Kit qPCR protocol and pooled equimolar on the basis of quantitative PCR results into a total of two pools. Pool volumes were then bead-cleaned up with 1× ratio and quantified using the Qubit dsDNA high-sensitivity assay. The quality,

size distribution and concentration of pooled libraries was checked using a DNA 1,000 chip on a Bioanalyzer 2100 instrument (Agilent Technologies). Due to the TruSeq style Y-shaped adaptors, part of unamplified single-stranded libraries could not be accurately run on an agarose gel or Bioanalyzer. To determine sizing more accurately, we first amplified a small portion of the final library. This was done for quality-control purposes only and the amplified product was not sequenced. For each sequencing pool, a test run was performed on a single MiSeq run using the v.2 Nano kit (catalogue no. MS-103–1001; Illumina) to ensure even sample representation and then each pool was sequenced 150 bp paired-end in a total of two S4 lanes on the Illumina NovaSeq6000 Sequencing System (University of San Francisco). Samples were sequenced to an expected average raw coverage of 5×, with 20% of the reads expected to be lost to sequencing duplicates. RNA sequencing libraries were built following each step of a cost-effective customized protocol by ref. 90, a method that is strand-specific (that is, the sequencing reads maintain the strand orientation of the transcripts) and uses oligo-dT enrichment[90]. Sequencing was performed on one lane of an Illumina HiSeq X Ten platform (Macrogen).

Whole-genome sequencing data quality control

Read processing, site quality filtering and SNP calling.—Adaptor sequences were trimmed from pair-end Illumina raw reads with cutadapt v.1.18 using a quality threshold of 20 to trim low-quality ends (5′ and 3′) from the reads before adaptor removal. FastQC v.0.11.8 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used to evaluate the quality of paired-end raw reads and potential adaptor contamination before and after adaptor trimming, resulting in the detection of poly-G sequences. To circumvent this problem, we used the --nextseq-trim=20 option in cutadapt v.1.18 (ref. 91), which trims reads below a quality threshold of 20 before adaptor removal with the specification that high-quality G bases at the 3′ end are ignored. Trimmomatic v.0.38 (ref. 92) was then used to perform a sliding window trimming, cutting bases once the average Phred quality within a 4 bp sliding window falls below 20 (SLIDINGWINDOW = 4:20) and to remove reads that are shorter than 30 bp (MINLEN = 30). Retained read pairs were aligned to the arctic fox (*V. lagopus*) reference genome[21] using the Burrows–Wheeler Aligner (BWA) v.0.7.17 (ref. 93) with the mem option and processed with samtools v.1.9 (ref. 94) to remove unmapped reads, reads with an unmapped mate, reads failing quality checks, reads whose alignment was not primary and reads with mapping quality below 15. Read pairs were also independently aligned to the domestic dog genome (*Canis lupus familiaris*) publicly available at the National Center for Biotechnology Information (NCBI) (CanFam3.1, GCF_000002285.3) using the same methods. The genome assemblies of the arctic fox[21] and domestic dog (*C. lupus familiaris*)[22] were both independently used as reference, as opposed to the currently available red fox genome assembly, to address issues of reference biases. Sequencing duplicates of the resulting bam files were marked for removal with Picard v.2.20.0 (http://broadinstitute.github.io/picard/). Read-group IDs were added to each bam file and libraries were merged to sample level using samtools addreplacerg and merge options, respectively. Lastly, local realignments of reads around indels were performed with GATK v.3.5.0 (https://www.broadinstitute.org/GATK) using RealignerTargetCreator to identify suspicious intervals requiring realignment which were then realigned using IndelRealigner. Finally,

read depth and coverage were determined for realigned bams using bedtools (Supplementary Table 1).

FastQC results identified severe bacterial and adaptor contaminations in four Rueppell's fox road-killed samples and as a result, the final coverage was below 1× in these individuals after all quality-control steps. These individuals were excluded from downstream analysis (Supplementary Table 1). Given that our remaining sample set included samples from road-killed individuals for which postmortem DNA damage is a possibility, we followed the conventions of ancient DNA research and used MapDamage v.2.09 (ref. 95) which allows both the quantification of damage patterns and the downscaling of the quality score of potential damaged bases in such samples. Postmortem DNA damage usually results in cytosine deamination, which increases the transition/transversion ratio. Misincorporation plots that display damage patterns at the 5′ and 3′ ends of reads showed no difference between samples taken from live individuals and road-killed animals. Because none of our samples had the profile characteristic of samples with postmortem DNA damage, we opted not to downscale quality scores.

Snakefiles used for read processing and evaluation metrics throughout different quality-control filtering steps can be found in ngsQCrun.py and ngsQCevaluate.py at https://github.com/joanocha/NGS-quality-control, respectively. High-coverage whole-genome sequencing data for other canid species were downloaded from GenBank at NCBI and received similar read processing (Supplementary Table 2). The arctic fox and the dog assemblies were aligned with nucmer in Mummer v.3 (ref. 96) with -c 800 (setting the minimum length of a cluster of matches to 800) and setting the minimum alignment identity to 90 (--delta-filter −1 -q -i 90). This allowed us to match the scaffolds of the arctic fox reference genome to the dog reference genome, which is assembled into chromosomes and has the best annotation among canid genomes (Supplementary Fig. 1).

After read processing, reliable genomic positions for population genetic analyses were obtained using snpCleaner v.2.4.1, a Perl script that works with bcftools vcf file format (https://github.com/tplinderoth/ngsQC/tree/master/snpCleaner). This was done by filtering our 85 bam files with the following criteria: (1) at least 80% of the individuals had to be covered by at least one sequencing read to retain the site, (2) sites with read depth above 50× in each individual (that is, 50 times the number of individuals in the dataset) were excluded, (3) only reads with mapping quality of 30 and base quality of 20 were included, (4) only read pairs mapped with proper-paired orientation were included, (5) sites with an excess of heterozygotes (minimum $P = 1 \times 10^{-6}$) were considered potential mapping errors and excluded, and (6) if one of the two alternative alleles was biased with respect to the read base quality (minimum $P = 1 \times 10^{-10}$), mapping quality (minimum $P = 1 \times 10^{-4}$) or distance from the end of the read (minimum $P = 1 \times 10^{-4}$), the site was excluded. The subset of retained sites (positions) from all samples will, hereafter, be designated as $\text{all}^{\text{sites}}$. Similar analyses were performed for a subset of the data comprising only ingroup foxes (red fox and Rueppell's fox; $\text{ingroup}^{\text{sites}}$) or using our dataset with additional canid genomes to be used as outgroups in phylogenetic analysis (Supplementary Table 2).

For single-nucleotide polymorphism (SNP) calling and down-stream population genetic analyses, we used the probabilistic framework based on genotype likelihoods implemented in ANGSD v.0.933 (ref. 20) as an alternative to calling genotypes. This framework accommodates uncertainty regarding genotypes associated with low-coverage datasets such as ours. We estimated allele frequencies and genotype likelihoods for the set of quality-controlled sites defined above (all^{sites}) using the flags: -minMapQ 30 -minQ 20 -remove_bads 1 -uniqueOnly 1 -only_proper_pairs 1 -GL 1 -doMaf 1 -doMajorMinor 1 -doGlf 2 -SNP_pval 1×10^{-5}. Additionally, a position file of all retained SNPs (only variable sites) was created from the minor allele frequencies output to be used in downstream analysis (hereafter, all^{snps}). ANGSD was run with -sites all^{snps} to estimate the site frequency spectrum (here-after, SFS) of each group of populations and/or species determined by unsupervised clustering analysis. Briefly, ANGSD computes posterior probabilities of sample allele frequency for each site, which is then used to estimate the SFS. The shape of the SFS was used as a proxy for the adequacy of data quality-control measurements and subsequent estimates of the SFS served as input for downstream analyses (namely F_{ST}, Pi-based selection scans and demography). We examined the SFS using all^{snps} estimated with -SNP_pval 1×10^{-6} and noticed it led to an apparent deficiency of the singleton category. While this category usually does not affect downstream analyses because it is removed by filtering on minor allele frequency filters, it can have a strong effect on demographic and population-size analyses. To avoid under-calling of rare alleles we, therefore, set a less stringent cutoff of SNPpval 1×10^{-5} for most of our analyses, unless otherwise noted.

For mitochondrial DNA reads, consensus sequences were called for positions with a minimum coverage of ten and at least 80% support in covered reads. Our 85 complete mitochondrial genome alignment was also combined with available complete mitochondrial genomes from other fox species downloaded from Genbank (Supplementary Table 2). The combined datasets were aligned using MAFFT v.7.407 (ref. 97) and the hypervariable region (D-loop) was excluded from downstream analyses using mitogenomes due to alignment ambiguities. With the exclusion of this region the mitogenome sequence alignment was 15,356 bp. We also trimmed our alignment to include ~1,130 bp of the cytochrome *b* (cytb) region so that we could perform downstream analysis with other available fox sequences (Supplementary Table 2). The mtDNA alignments were eye-checked using SeaView v.4.7 (ref. 98).

Relatedness estimation.—We ran NGSrelate[99] in ANGSD to investigate the extent of relatedness in our dataset, as most population genetic analyses require that individuals are not closely related to each other and since this study is based on a low-coverage genomic dataset. NGSrelate allows estimation of relatedness coefficients from geno-type likelihoods instead of called genotypes and thereby accounts for the uncertainty inherent in low-depth sequencing data. NGSrelate was run on three datasets (or bam lists): 32 North African, 18 Eurasian red fox individuals and 25 Rueppell's fox individuals. For each bam list we used ANGSD to estimate allele frequencies and genotype likelihoods with the following options: -minMapQ 30 -minQ 20 -remove_bads 1 -uniqueOnly 1 -only_proper_pairs 1 -gl 1 -doMajorMinor 1 -doMaf 1 -doGlf 3 for a selected set of sites including only autosomal SNPs (-sites all^{snps} without X chromosome). Then, we looked into the k0, k1

and k2 coefficients for each pair of individuals within each dataset. Briefly, k0, k1 and k2 correspond to the proportions of the genome where the pair of individuals analysed share 0, 1 and 2 alleles identical by descent, respectively, which means that the less related two individuals are the higher proportion of alleles with 0 identity by descent (the higher the k0). Pairs of individuals with a k0 <0.75 were considered to be closely related and for these cases we removed the individual with the highest number of close relatives until no pair of close relatives was left. NGSrelate was run using standard expectation maximization (EM) for optimization, allowing up to 500,000 EM iterations for each pair and a stopping criterion of 1×10^{-12} difference in likelihood between two consecutive EM iterations. This was performed ten times with different random number seeds to ensure the convergence of the EM algorithm. Ultimately one single pair of k0 = 0.4 was identified for North African red fox individuals and two pairs for Rueppell's fox, leading to the removal of one individual (with lowest average genome-wide coverage) for each pair of related foxes (Supplementary Table 1). We did not run NGSrelate on outgroup species due to our limited sample size. Downstream whole-genome and mtDNA analyses proceeded with a final dataset comprising 82 foxes (Supplementary Table 1). As such, we repeated the steps performed in site quality filtering and SNP calling but for a bam list of 82 individuals (all^{sites} and all^{snps} without related samples).

RNA-seq quality control and differential expression analyses

Reads were mapped to three different reference genomes and overlaps with annotated features were counted with the splice-aware aligner STAR[100]. The reference genomes used were the domestic dog (Can-Fam3.1, GCF_000002285.3), the red fox (vulVul2.2, GCF_003160815.1), downloaded from NCBI, and the arctic fox[21]. In the case of the domestic dog and the red fox genomes, the respective RefSeq gene/transcript annotations in GTF format were used for read overlap counting. For the arctic fox, the dog annotation was mapped to the arctic fox assembly using Liftoff 1.5.1 (ref. 101) and the resulting GTF annotation was then used for overlap counting in STAR. Mapping results were inspected with Qualimap[102] and MultiQC[103] (Supplementary Fig. 15). To confirm the genetic relationships between the RNA-seq samples, genotypes were called using the reads mapped to the dog genome with BCFtools mpileup/call[104]. The resulting genotypes were filtered with VCFtools[105] to keep only bi-allelic loci with a minimum coverage of four reads, MAF of 0.05 and missing at most in three samples. The dataset was then pruned for loci in high LD (variants with $r^2 > 0.2$ in 50 kilobase (kb) windows and step size of ten variants) and genotypes were used in a principal components analysis (PCA) in PLINK 1.90 (ref. 106) (Supplementary Fig. 16a). Fragment counts (counts of read pairs overlapping annotated genes) from STAR were imported and processed using R 4.0.3 3 (2020). Only genes with at least ten overlapping fragments (between all samples) were considered for the downstream analyses. Normalization, expression estimation and differential expression were performed with the DESeq2 package[107]. For exploratory analyses, namely PCAs based on gene expression, gene counts were transformed using the rlog transformation (Supplementary Fig. 16b–d). In the differential expression analyses, the significance threshold was set at 5%. Differential expression was tested between the following pairs of sample groups: Rueppell's fox versus red fox; all North African samples versus all Iberian samples; North African red

fox versus Rueppell's fox; all Rueppell's fox versus Iberian red fox; North African versus Iberian red fox (Supplementary Table 15).

Mitochondrial DNA phylogenetic analysis

Maximum-likelihood trees for mitogenomes and cytb alignments including representatives from all fox species with available DNA information (Supplementary Table 2) were generated with MEGA X 10.1.5 using the general time reversible model with 1,000 bootstrap samples[108]. The consensus phylogenies were used as inputs for ancestral range reconstruction under a maximum-likelihood framework with the R package BioGeoBEARS[109,110]. Species were assigned to North Africa, Eurasia and North America, according to their current geographic distribution range (accessed from the International Union for Conservation of Nature (IUCN https://www.iucnredlist.org/, 2020). We did not assign red fox to North America, as it has been established that red fox colonized North America from Eurasia and we were primarily interested in ancestral ranges[23,111]. Furthermore, our red fox samples are of Eurasian and North African origin. We ran BioGeoBEARS under three biogeographical models: DEC (dispersal-extinction-cladogenesis), DIVA (dispersal-vicariance analysis) and BayArea (Bayesian inference of historical biogeography for discrete areas), all of which were tested with and without founder-event speciation (*j* parameter)[109]. We then chose the best-fitting model (DEC, DEC + J, DIVALIKE, DIVALIKE + J, BAYAREALIKE and BAYAREALIKE + J) on the basis of likelihood ratio test, the Akaike Information criterion and corrected AIC (Supplementary Table 4).

From the whole-mitochondrial genome alignment dataset, Bayesian phylogenetic trees and estimates of the time to the most recent common ancestor (t_{MRCA}) were performed using BEAST v.1.10.4 (ref. 112). We used the general time reversible model with a site heterogeneity modelofgamma andinvariantsites(GTR + G + I), assuggestedby JModel Test[113]. We ran the settings with an uncorrelated relaxed log-normal clock model to allow rate variation among lineages. The Yule tree prior assumes an (unknown) constant lineage birth rate for each branch in the tree and it is usually suitable for trees describing the relationships between individuals from different species. However, this speciation process does not allow for extinction. The birth–death process, on the other hand, assumes that at any point in time every lineage can undergo speciation or go extinct and, given the observed cases of extinction in Canidae and *Vulpes* genus fossil record[27], we assumed this tree to be the better suited to our dataset. Nodes of the tree were calibrated using the t_{MRCA} using both fossil and secondary calibrations estimated from previous studies[114–116]. Fossil calibrations were set using a gamma prior. The t_{MRCA} for *Vulpes* species and raccoon dog was set to 8.2 Ma (refs. 27,114). The oldest fossil assigned to Old World *Vulpes* is *V. riffautae* around 7 Ma (ref. 27). We assumed that this fossil is the basal group to all other *Vulpes* in our dataset and used it to calibrate the t_{MRCA} for pale fox and the remaining fox species. Using secondary calibrations estimated from previous studies as priors, the divergence time between fennec fox and other species was set to 4.72 ± 0.5 Ma. The t_{MRCA} for arctic fox and reminder species was set to 3.17 ± 0.09 Ma (ref. 115). The t_{MRCA} for corsac/Tibetan fox (*V. corsac/V. ferrilata*) and reminder *Vulpes* was set to 1.8 ± 0.5 Ma and t_{MRCA} for corsac and Tibetan fox to 1.02 ± 0.3 Ma (ref. 114). The t_{MRCA} for Bengal's fox (*V. bengalensis*),

red fox and Rueppell's fox were left unknown. We performed a single run with 150×10^6 iterations and sampled states every 1,000 iterations. We used Tracer v.1.7.1 (ref. 112) and effective sample size values to investigate convergence (Supplementary Table 5). We next constructed a median-joining network for whole mitogenomic and cytb haplogroups using PopART v.1 (ref. 117) to assess the relationships between the different haplotypes at a population level (Supplementary Table 2).

Population genetic analyses

To understand how different ingroup and outgroup species of *Vulpes* are related at the whole-genome level, how genetically divergent and differentiated they are from each other and what the population history of North African and Eurasian red foxes is, we explored our final low-coverage dataset comprising 82 whole genomes using different methods (see Supplementary Table 1 for information on average coverage per individual and geographic distribution). Principal component analysis (PCA) was executed in PCangsd using separate Beagle-format files generated in ANGSD with -sites all^{snps} for all 82 foxes (Supplementary Fig. 2) and for 48 red foxes (Supplementary Fig. 5a) and 24 Rueppell's foxes (Supplementary Fig. 5b), filtering for a minor allele frequency (-minMaf) of 0.05. Eigen decomposition was performed on the estimated covariance matrix using the eigen function in R to obtain a PCA plot. Global F_{ST} estimates based on the joint SFS (jSFS) between major clusters defined by the PCA, were computed in ANGSD. Global F_{ST} estimates were performed between species to observe the extent to which different *Vulpes* were differentiated from each other, as past evidence from microsatellite data had suggested high genetic similarity for Rueppell's fox and North African red fox[12]. The extent of genetic differentiation seen between North African and Eurasian red fox and between Rueppell's fox and red fox was used to assess the most appropriate methods to detect signatures of selection. As input for the different SFSs, we used a set of SNPs called with all individuals (all^{snps}).

To infer global admixture proportions and the covariance structure of allele frequencies in Rueppell's fox and red fox (ingroups), we used the structure model implemented in the Ohana tool suite by assuming two to six ancestral populations (*K*, Supplementary Fig. 6). For each value of *K* in Ohana[54] we used 32 independent executions with different random seeds and report the ones that reached the best likelihood. This was done using as input in a Beagle-format file generated in ANGSD with Rueppell's fox and red fox individuals altogether, using a set of quality-controlled SNPs called from all individuals (-sites ingroup^{snps}). To measure patterns of LD we filtered Beagle-format files for red fox and Rueppell's fox to include SNPs with at least 1× coverage in all individuals. These were then used as inputs for ngsLD. As a last step, we ran both the Rscript provided by ngsLD and a customized python script to draw the distribution of R^2 against the physical distance between SNPs. LD patterns were used to define window size for down-stream analysis on a sliding window. We also estimated whole-genome pairwise genetic distances between individuals using ngsDist, a programme that takes the uncertainty of a genotype's assignation into account (v.1.0.9; ref. 118). We ran ngsDist with 100 bootstraps on different datasets generated with SNPs called for all 82 whole genomes generated in this study (all^{snps}) and with additional whole genomes from other species, which were mapped and

processed together with our dataset in a parallel analysis. We ran FastME[119] to infer a tree for each of the resulting matrixes and then RAxML[120] to place support on the main tree and further confirmed that species relationships were resolved (bootstrap support >95%).

Detection of ancient admixture events

To elucidate possible patterns of gene flow between Rueppell's fox and red fox we performed an ABBA-BABA test (*D* statistics) as implemented in -doAbbababa2 of ANGSD[121]. This is based on the di-allelic patterns of alleles between four groups of individuals—H1, H2, H3, H4—and allows testing for the correctness of a hypothetical genetic relationship between the four groups, where H4 is the outgroup and the correctness of the tree corresponds to absence of gene flow between H3 and either H2 or H1. This method supports low-depth sequencing data and is calculated using all the reads of the genomes (no -sites or -rf filter provided). We first performed *D* statistics using the arctic fox as H4 (outgroup), Rueppell's fox as H3 and Eurasian and North African red foxes as H1 and H2 and tested for the correctness of this relationship between these four lineages. We additionally tested for the correctness of the relationships between the four taxa used in this study: pale fox (H4), fennec (H3) and Rueppell's fox (H2) and all red foxes (H1). These analyses were repeated using bams generated from mapping to the domestic dog genome and using the dog as outgroup (H4). Genome-wide analyses were run on blocks of 5 Mb followed by the Rscript DSTAT as implemented in ANGSD to obtain genome-wide level *D* statistics, *Z* scores and *P* values for significance (Supplementary Table 6).

For comparisons in which significant genome-wide introgression was found, the ABBA-BABA test was performed on 200, 100 and 50 kb windows to find potentially introgressed outlier regions or genes. *D* statistic and *Z* scores were computed on windows along the length of the genome. Since spurious outliers may also arise in regions of reduced diversity, we adapted ABBA-BABA on windows to calculate f statistics[35]. Briefly we ran doAbbababa2 on windows to get numerators for H1,H2,H3,H4 and H1,H2,H2,H4 and H1,H3,H3,H4. We then merged the .abbababa2 outputs to estimate the fraction of introgression, $f$hom=numerator(H1,H2,H3,H4)/numerator(H1,H3,H3,H4) and fraction admixture fd = numerator(H1,H2,H3,H4)/numerator(H1,HD,HD,H4) , where HD = H2 or HD = H3, depending on which maximized the frequency of the derived allele (https://github.com/joanocha/ngsIntrogression/). Alternatively, we ran ANGSD with -doHaplo 1 -doCounts 1 -minMinor 1 to call pseudohaploids on the basis of sampling of a single base directly from bam files and the resulting file was used to run fourPopWindows.py in genomics general (https://github.com/simonhmartin/genomics_general) and calculate f statistics on sliding windows (Fig. 3b). Genomic windows above the 99th percentile of the empirical distribution of f_d were considered outlier candidates for introgression between H2 (or P2) and H3 (or P3), presented as ranked lists from highest to lowest value (Supplementary Tables 8 and 10). Both approaches, which differ fundamentally on user choice to either work with genotype likelihoods or pseudohaploids of a single base directly from bams, resulted in the same outliers. The GTF annotation resulting from mapping the arctic fox to the dog assembly with Liftoff v.1.5.1 (ref. 101) was then used for overlap with genomic windows.

TreeMix v.1.13 (ref. 34) was also used to investigate the phylogenetic placement of Rueppell's fox and red fox relative to the arctic fox, fennec and pale fox and to further investigate the directionality of possible introgressive-hybridization events. We first ran ANGSD to generate allele frequency files for the following groups: North African red fox (n = 30), Eurasian red fox (n = 18), Rueppell's fox (n = 24), arctic fox (n = 1), fennec (n = 5) and pale fox (n = 5). Another version was run with allele frequency files (hereafter, maf files) for Iberian (n = 4) and Middle Eastern (n = 13) red foxes. Each of the maf files was generated using a set of SNPs called with all individuals and forcing the major allele frequency to be that of the ancestral (-minMapQ 30 -minQ 20 -remove_ bads 1 -uniqueOnly 1 -only_proper_pairs 1 -GL 1 -doMaf 1 -doMajorMinor 5, -SNP_pval 1e-5). These were performed using reads mapped to the dog genome, which was used as -anc in ANGSD. The different maf files were merged on the basis of chromosome, position and major allele and adapted to be used as input for TreeMix. We estimated minor allele counts by multiplying the EM frequency estimate of the maf file by twice the number of individuals (2Nind) and major allele counts by subtracting the minor allele count to 2Nind (input_for_treemix.py in ngsIntrogression/). We then ran TreeMix with options -k 500, -global, -noss and -se allowing zero to three migration events, with the pale fox as outgroup (Supplementary Figs. 7 and 8).

Tree topologies for potentially introgressed regions and genetic divergence

We investigated the phylogenetic pattern for major outlier regions from the window-based ABBA-BABA analyses (windows with f_d above the 99th percentile of the empirical distribution). First, we constructed a genome-wide tree for *Vulpes* using one individual representative of each species (highest coverage individual per species and within Eurasian and North African red foxes; Supplementary Table 1). Selected bams were executed in ANGSD with -doHaplo 1 -doCounts 1 -minMinor 1 to call pseudohaploids on the basis of sampling of a single base. The resulting file was transposed into a fasta format alignment which was used as input in IQ-TREE v.1.6.12 and run with -m TEST (JModelTest model finder) and -bb 1000 (number of bootstraps). The resulting maximum-likelihood consensus tree was then plotted in R (Fig. 3c,e, left). We compared this tree topology (>95% bootstrap support) to that estimated using ngsDist to ensure that species relationships were resolved (Fig. 1c). This workflow was also used to generate maximum-likelihood trees for potentially introgressed genes or genomic regions, with the difference that ANGSD was run for all the individuals from our dataset (Fig. 3c,e, Extended Data Fig. 3a, Supplementary Fig. 9 and Extended Data Figs. 5a,b and 7a,b) and with the -r option to provide the coordinates of the reference genome for the candidate gene. ANGSD was also run with -doHaplo 1 and -r with additional available genomes (Supplementary Table 2). To augment the phylogenetic analyses, we performed PCA using a single read sampling approach in ANGSD using SNPs that were called for potentially introgressed regions based on genotype likelihoods (Extended Data Figs. 3b, 4 and 5c).

Finally, we used allele frequency files generated in ANGSD to perform scans of genetic divergence (d_{XY}) between species and red fox populations in 50 kb windows with a 10 kb slide and 25 kb windows with a 5 kb slide using PopGenomicsTools (https://github.com/tplinderoth/). These estimates were used to investigate whether potentially introgressed regions had lower genetic divergence between donor and recipient species than genome-

wide levels. Similarly, we used the estimated jSFS between different fox lineages to compute F_{ST} and nucleotide diversity (π) on sliding windows with ANGSD and then investigated whether potentially introgressed regions had lower genetic differentiation between donor and recipient species than genome-wide levels (Supplementary Figs. 10–13 and Extended Data Figs. 6 and 7c). As input for the different SFSs, we used a set of SNPs called with all individuals (all^{snps}).

Demographic analyses

The jSFS (or 2D-SFS) of North African Rueppell's fox and red fox was obtained using realSFS using the same set of quality-controlled SNPs (ingroups^{snps}) excluding potentially introgressed regions and the X chromosome. This served as input for estimating a demographic history of the two species in North Africa with dadi[26]. To jointly estimate demographic parameters such as ancestral and current effective population sizes, population divergence times and migration rates and a best-fitting model we ran dadi as implemented in GADMA (genetic algorithm for demographic analyses)[25]. This software performs an automatic and unsupervised global search for a best-fitting demographic model using a genetic algorithm based on dadi parameter estimates and composite likelihood scores (Supplementary Table 7). For all demographic analyses we assumed a generation time of 2 and a mutation rate similar to wolves (~4.5×10^{-9}) (ref. 122). Thus far, the conventional generation time used to model the demographic history of red fox has been one[23,123]. However, this is unlikely to match the generation time in natural populations. For instance, the generation time is presumed to be three for Rueppell's fox (IUCN, 2021; https://www.iucnredlist.org/fr/species/23053/46197483) and 3.4 for kit fox[124]. GADMA was run independently for 100 model replicates and allowing for changes in the effective population size of the ancestral of the two populations. We then checked whether the resulting best models and demographic parameters were compatible with the BEAST estimates of divergence time for mtDNA (Supplementary Table 5) and evidence from whole-genome analyses, including ABBA-BABA tests (Supplementary Table 6) and admixture graphs (Supplementary Fig. 7). The jSFS of North African and Middle Eastern red foxes served as input to explore the evolutionary origin of red foxes using the same approach to that of the evolutionary history of North African Rueppell's fox and red fox (Supplementary Table 7). European red fox samples were not analysed due the degree of divergence from Middle Eastern red fox.

Genome-wide signatures of selection in North African red fox

To detect signatures of positive selection in North African red fox we used the PBS[125], which is a summary of a three-way comparison of allele frequencies between a focal group, a closely related ingroup and an outgroup. PBS specifically tests for loci where allele frequencies in the focal group are especially differentiated from those in both of the other populations, which could be signatures of positive selection. A population's PBS value thus represents the amount of allele frequency change at a given locus since the focal population diverged from other two. To test the hypothesis of selection on North African red fox we used Eurasian red fox as the closest ingroup and Rueppell's fox as out-group. To test the hypothesis of selection on Eurasian red fox, North African foxes were used as the other ingroup and Rueppell's fox as outgroup. All PBS analyses were performed on the variance

component files printed with ANGSD for the three-way comparison using FstPBS.py in the ngsSelection tool suite (https://github.com/joanocha/ngsSelection/). This script uses the printed files that contain the variance between population and total variance per SNP to calculate three-way weighted estimates of F_{ST} and PBS on sliding windows of base pairs or SNPs (upon user specification). On the basis of considerations of the rate of decay of LD in the populations (Fig. 2c), we chose to work on windows of 50 kb with a 10 kb slide (genome-wide) and 25 kb with a 5 kb slide (for an enlarged view of candidate genes).

We then used ANGSD to compute genome-wide estimates of average number of pairwise differences (π) on windows of 50 kb with a 10 kb slide and 25 kb with a 5 kb slide, for each of the three groups. To further investigate selection, we looked for windows of reduced diversity in the focal group relative to other regions of the genome and the other populations, which could be signatures of selective sweeps. To account for variation in π common to all foxes, due to factors such a variation in the mutation rate or background selection and to focus on reduced diversity unique to the focal group, we calculated $\alpha = \pi^{pop1} / (\pi^{pop1} + \pi^{pop2} + \pi^{pop3})$, where pop1 is the focal group, pop2 is the closely related ingroup and pop3 is the outgroup. Windows with higher $-\log_{10}(\alpha)$ relative to the rest of the genome might harbour genes putatively under selection (https://github.com/joanocha/ngsSelection/).

These analyses were performed simultaneously for reads mapped to the arctic fox and domestic dog reference genomes. We focused on candidate genes with extreme outlier regions under both methods (PBS and $-\log_{10}(\alpha)$) (ranked in Supplementary Tables 12, 13, 16 and 17) and tested for expression differences between groups (section on RNA-seq quality control and differential expression analyses) (Supplementary Figs. 15 and 16, Extended Data Fig. 8 and Supplementary Table 15). We then ran PBS on individual SNPs to detect the top variants under selection for each of the top candidate genes for reads mapped to the domestic dog reference genome and inspected their location using the UCSC Genome Browser[126].

Genome-wide signatures of selection in core desert-dwelling species

PBS and α are powerful methods to find signatures of selection in red fox because global estimates of F_{ST} between North Africa and Eurasian red foxes were sufficiently low (global $F_{ST} \approx 0.13$) and we had data from their immediate outgroup, Rueppell's fox. While PBS has proved effective for identifying population-specific selection in humans and canid populations adapted to extreme environmental conditions[13,125,127], F_{ST}-based methods are not useful when genetic differentiation between the focal group and the closely related ingroup is high ($F_{ST} > 0.2$; ref. 128). In sharp contrast to past F_{ST} estimates using microsatellites (global $F_{ST} \approx 0.14$ between Rueppell's fox and North African red fox and global $F_{ST} \approx 0.13$ between Rueppell's fox and Eurasian red fox; ref. 12), we found Rueppell's fox and red fox to be highly differentiated using whole-genome data (global $F_{ST} \approx 0.52$ between Rueppell's fox and North African red fox and global $F_{ST} \approx 0.51$ between Rueppell's fox and Eurasian red fox). The same rational would apply to attempts of finding signatures of selection in the fennec. As such, we used two alternative approaches to detect species-specific signatures of selection in either Rueppell's fox or fennec.

Multispecies Ohana

We ran Ohana[54] to obtain admixture plots and trees for our multispecies dataset (n = 24 Rueppell's fox, n = 48 red fox, n = 5 fennec, n = 5 pale fox), using as input a beagle file with SNPs called from all individuals (all^{snps}). Using the results from K = 4, we ran the selscan option in Ohana, which uses a maximum-likelihood method to detect SNPs that strongly deviated from the genome-wide covariance structure. This method is distinct to PBS in the sense that it focuses on genetic components, instead of predefined populations and can incorporate more than three branches in a tree, thus having greater power of analysis when more than three groups are present. It is also distinct from PBS in the sense that it uses the genome-wide covariance structure to form a null hypothesis that can be tested against using a likelihood ratio test for each SNP. The resulting likelihood ratio test can thus be used to identify regions in a focal group that experienced a larger than expected change in allele frequency as compared to the prediction of the genome-wide pattern. In this study, we took advantage of this approach to test for selection in the branches leading to fennec fox and Rueppell's fox.

For each SNP we introduced a scalar variable (h; https://github.com/jade-cheng/ohana/wiki/) that is multiplied onto the variance associated with the focal population/species and therefore allows for two models: one in which the focal population-specific allele frequency changes can be predicted from the genome-wide covariance pattern and another that allows for larger changes in focal populations (higher variance components) than expected from the genome-wide pattern. In regions that have experienced selection, we would expect increased LD and multiple SNPs showing a strong signal. However, even if there has been selection affecting a window there may also be many SNPs that do not show strong allele frequency differences because of low LD with the causal SNP. For that reason, we calculated the mean log-likelihood ratio for the top 500 SNPs in windows of 50 kb, sliding every 10 kb (https://github.com/joanocha/ngsSelection/Ohana/). Windows with mean top likelihood ratios above the 99.95th percentile of the empirical distribution were considered to be outliers and genes in the windows were identified and classified as potential targets of selection (Fig. 5 and Supplementary Tables 19 and 20). We denote the log-likelihood ratio calculated for fennec as LLR_F, Rueppell's fox as LLR_{Ru} and the one calculated for the red fox lineage (Eurasian + North African red fox) as LLR_{Red}.

We noted that many of the outlier regions putatively under selection were common to both Rueppell's fox (LLR_{Ru}) and red fox (LLR_{Red}), nominally suggesting that selection has been strong on both species and targeted inherited ancestral standing genetic variation. This is probably an artefact caused by the fact that the nearest outgroup to Rueppell's fox and red fox is quite divergent and, for that reason, it is difficult to distinguish between selection on the red fox branch and Rueppell's fox branch. For genes truly under selection in, say, red fox, if not allowing selection in the red fox branch, an increase in length of the Rueppell's fox branch will still lead to an improved likelihood compared to a model of no selection in which neither branch is elongated. To address this problem, we used a two-step inferential approach. We first determined if there was evidence for selection in either lineage using $LLR = \max\{LLR_{Red}, LLR_{Ru}\}$. For windows with evidence of selection, we then further investigated which lineage was the most likely target of selection using the statistic: d_{LR}

= $LLR_{Ru} - LLR_{Red}$, where a positive value indicates more evidence that the selection is in Rueppell's fox and a negative value indicates that the selection probably was in the red fox. We note that when calculated for a single SNP and not in windows, $Ratio_{LR}$ can be interpreted as a Bayes factor. We then focused on the 99.95th percentile of the empirical distribution for d_{LR} in Rueppell's fox (Supplementary Table 19). Finally, we looked to see if candidate genes were differentially expressed between Rueppell's fox and red fox (section on RNA-seq quality control and differential expression analyses) (Fig. 5).

Intraspecific polymorphism to interspecific divergence.—We calculated the test statistic for a chi-square test of homogeneity akin to the Hudson–Aguade–Kreitman (HKA) test[129,130], as an alternative method for identifying signatures of selection specifically on Rueppell's fox and on fennec, following a similar approach to that of ref. 131. We note that a main difference between this approach and the previously discussed approach for identifying selection is that HKA-related tests can identify selection using divergent species with evidence of repeated selective fixations or recent selective sweeps, as indicated by a reduced ratio of polymorphism to divergence compared to the rest of the genome (https://github.com/joanocha/ngsSelection/). Using both tests provides an opportunity to detect selection acting on different timescales. To perform these analyses on the low-coverage data we used the ANGSD allele frequency estimates (obtained by running ANGSD with the flag -doMajorMinor 5 for all[snps]) to calculate counts of A, B, C and D defined as: A, number of polymorphic sites in focal species; B, number of polymorphic sites in red fox; C, number of fixed differences between focal species and both red fox and dog; and D, number of fixed differences between red fox and both focal species and dog.

This analysis was repeated with both the fennec and Rueppell's fox as focal species. Because we are working on genotype likelihoods, we defined as fixed SNPs those having an estimated minor allele frequency above 0.97 in one species and below 0.03 in the other. We further explored other threshold combinations (0.99 and 0.01; 0.95 and 0.05). We considered as polymorphic the SNPs whose minor allele frequency ranged between the minimum and maximum frequency thresholds. Then we tested for the null hypothesis that the A/C on windows of 50 kb sliding every 10 kb are equal to A/C genome-wide with a Pearson's chi-square test. To find signatures of positive selection specific to the focal species we performed a homogeneity test for reduction of polymorphic sites (low polymorphism-to-divergence ratio in fox species of interest relative to red fox). In this test, the null hypothesis is that $A/C = B/D$ (or $A/B = C/D$), which is then tested with a Pearson's chi-square test on the 2 × 2 contingency table. The nominal P values calculated from these tests are not calibrated correctly because they do not take the variance of the coalescence process within species into account when there is LD among SNPs. For that reason, we provide the $-\log_{10}(P)$ as a raw score that should not be interpreted probabilistically, similarly to ref. 131. Then windows were ranked based on the homogeneity test score, filtering for windows in which:

- C was higher than A
- the ratio of C/A was always greater than the ratio of D/B
- the ratio of C/A was always higher than $C_{\text{genome-wide}}/A_{\text{genome-wide}}$

- the ratio of D/B was always lower than $D_{\text{genome-wide}}/B_{\text{genome-wide}}$
- HKA chi^2 value higher in the focal group (Rueppell's fox or fennec) than in red fox
- HKA score in red fox <1.3

The rationale behind these conditions, also used for polar bears[131], was to ensure that we were targeting outlier windows specific to the Rueppell's fox or the fennec lineage not found under selection in red fox. We reported the top-20 ranked genes for homogeneity test scores for fennec. No windows were left for Rueppell's fox after applying the same filtering (Supplementary Table 21).

Functional enrichment analysis

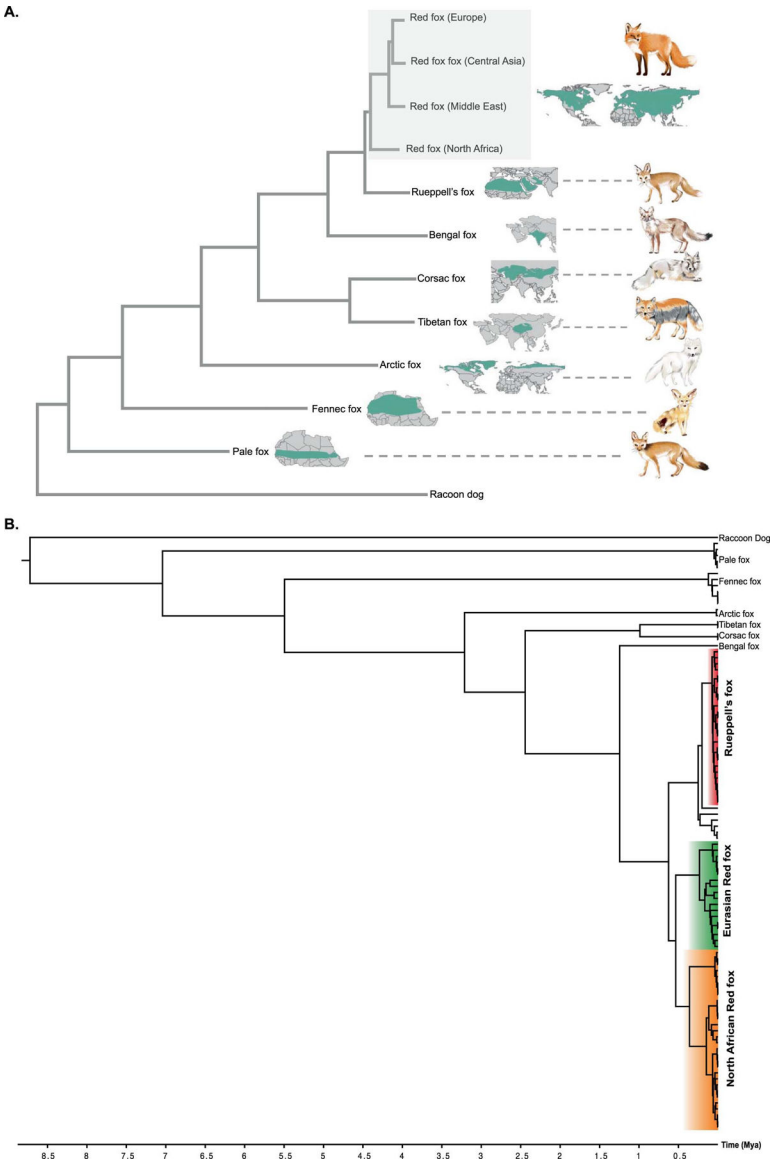
We computed enrichment of GO terms (http://geneontology.org/) for putatively introgressed regions between North African fox (that is, above the 99.9th percentile for ABBA-BABA on windows) using biological process or molecular function. We also computed enrichment of GO terms in the regions of the genome that displayed high $-\log_{10}(a)$ values when using North African fox or Eurasian red fox as focal populations and with high d_{LR} and $\text{LLR}_{\text{fennec}}$ (above the 99.95th percentile). The GO terms were considered to be significantly enriched if FDR $P < 0.05$. Both P values and corrected (FDR) P values were reported (Supplementary Tables 9, 11, 14 and 18). We repeated these analyses with the GOrilla toolkit (http://cbl-gorilla.cs.technion.ac.il/) using two unranked lists of genes: the target list for potentially introgressed genes and the top genes under selection and a background list of the annotated genes.

Differences in physiological parameters between core desert fox and red fox

To further explore evidence for desert adaptation in desert-dwelling fox we tested for species-specific significant differences in physiological parameters between Rueppell's fox, fennec and red fox (Methods; Supplementary Table 22). Descriptive statistics of the data were performed in R, including summary statistics for each parameter and boxplots (Supplementary Table 22 and Extended Data Fig. 10). As an exploratory approach to the data, pairwise t-tests comparing the three different fox species were performed on each parameter, as well as Welch's two-sample t-test comparing males versus females, North African versus European red fox and captive versus free-ranging individuals. Juveniles were excluded from the dataset. We then tested for significant species-specific differences in these parameters, while accounting for additional factors (covariates) that may affect the results of exploratory t-tests: status (captured free-ranging in the wild or held captive), sex (female or male), weight (continuous variable measured in kilograms) and body mass index (continuous variable measured in kilogram per square metre). Body mass index (weight divided by square of body length) was used as an alternative covariate to body length to account for the high correlation found betweenweightandotherbodymeasurements(hindfootandtaillength) (Supplementary Fig. 17). The linear model was implemented in R using the lm function (Supplementary Table 23). A Shapiro–Wilk test of normality was performed and QQ-plots inspected to determine if the residuals were normally distributed. To test for homogeneity of variances, both Levene's

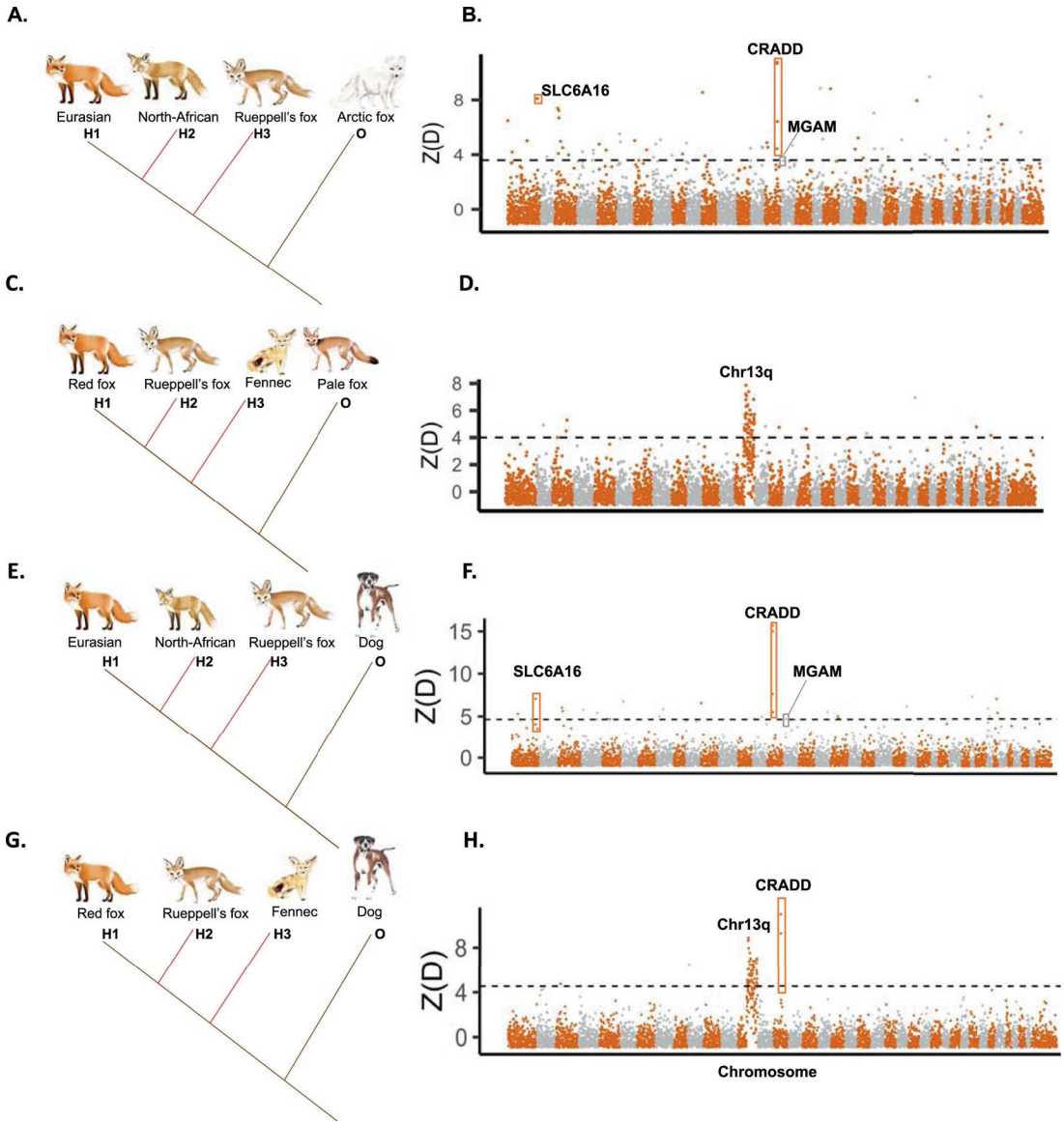
and Barttlet's tests were performed in R. For the parameters in which the Shapiro–Wilk test did not lead to the rejection of the null hypothesis of the residuals being normally distributed, we based our conclusions on the *P* values provided by lm. For the cases in which assumptions of normality were violated, as well as the assumptions of homogeneity of variances, the response variable was log-transformed (Supplementary Table 23). We also used certain physiological parameters as covariates to account for putative correlated responses (Supplementary Table 23). Finally, we were not able to collect enough urine samples for inferential statistics ($n = 3$ Rueppell's fox and $n = 3$ red fox). As such, we compared the maximum values for urine-concentrating ability obtained in this study with 121 mammalian species published in a recent review[11] by plotting these values along an aridity gradient. While acknowledging the inherit limitation that different individuals in this review were captured under different circumstances (some dehydrated, others water-fed and others unknown) and urine osmolality was measured with different methods, this allowed us to gain a general perspective on where our values lie in the general distribution of mammalian maximum urine osmolalities (Supplementary Fig. 18).

Extended Data



Extended Data Fig. 1 |. Mitogenomics of Vulpes using the raccoon dog as an outgroup (related to Fig. 1).

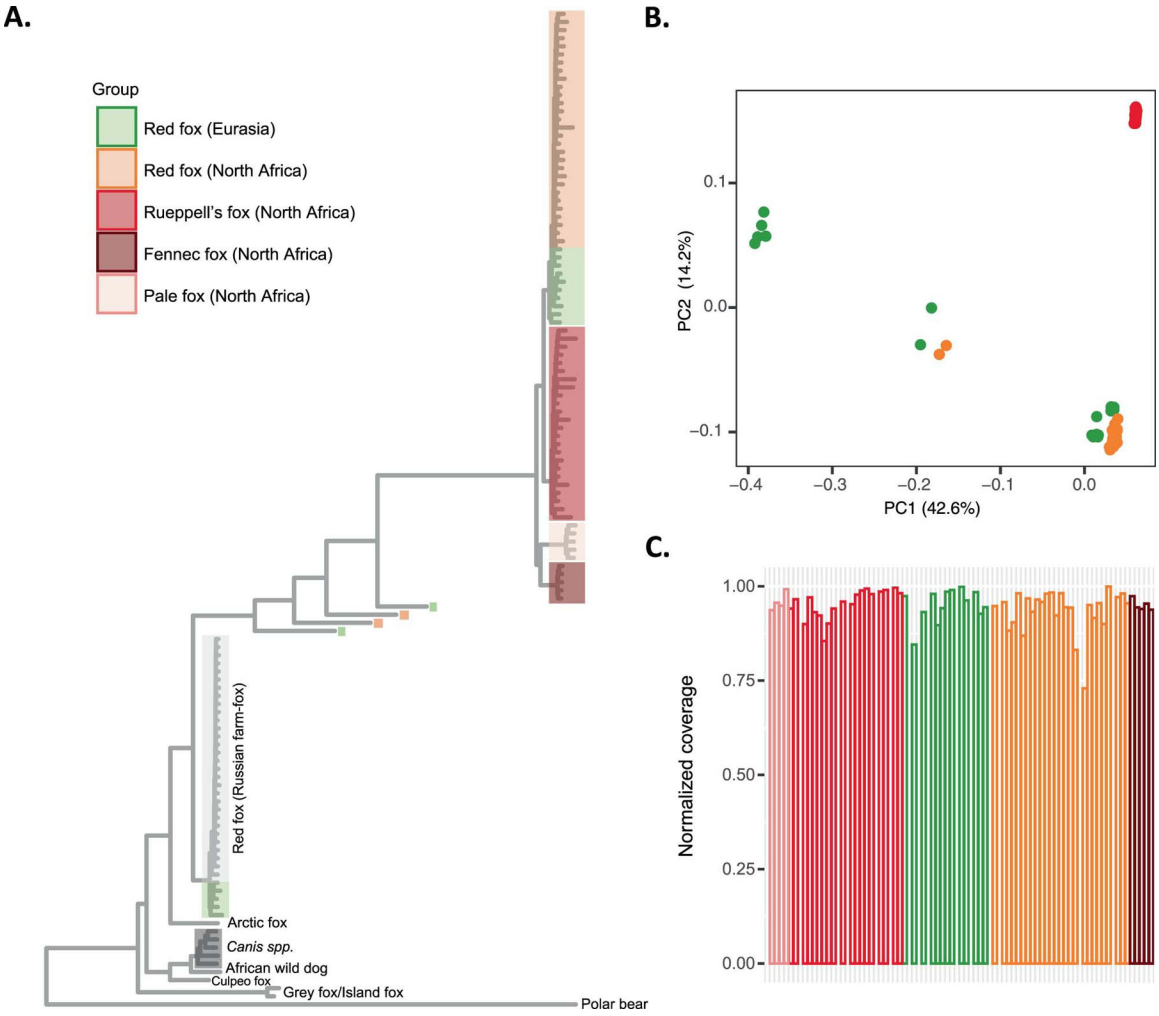
(a) Maximum-likelihood tree followed by species distribution range downloaded from IUCN (2021). Fox illustrations by Margarida Laranjeira Rocha. **(b)** Bayesian phylogenetic tree and time to the most recent common ancestor estimates; timescale in Ma represented bellow (see Supplementary Table 5 for the median time estimates and 95% HPD for each node). Accession numbers of published sequences are available in Supplementary Table 2. North African red foxes are coloured in orange, Eurasian red foxes green and Rueppell's fox in red. Rueppell's foxes are outgrouped by a non-coloured clade of 6 red foxes (n = 1 North Africa, n = 6 Middle East), revealing a paraphyletic pattern in red foxes.



Extended Data Fig. 2 |. ABBA-BABA on windows using the arctic fox (A-D) and the domestic dog (E-H) as reference genomes (related to Fig. 3).

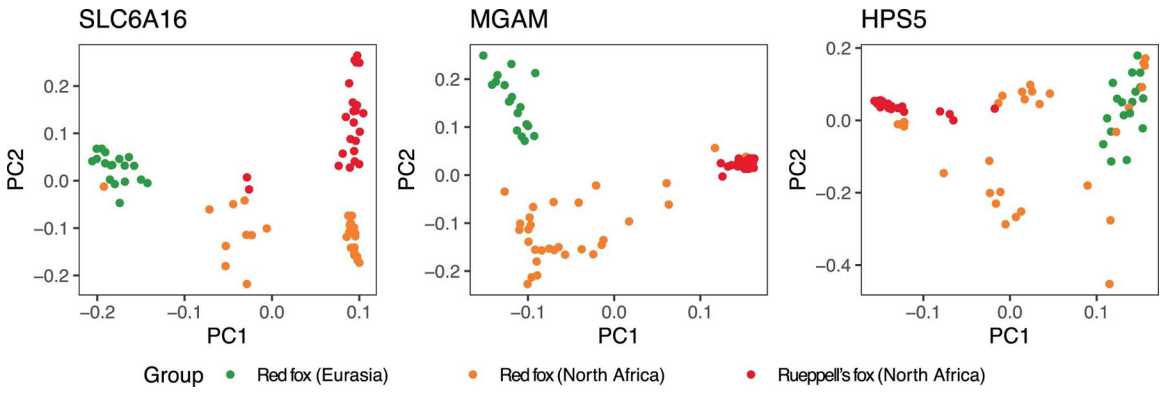
D statistics between Rueppell's fox (*Vulpes rueppellii*) and North African red fox (*V. vulpes*) (**a–b, e–f**) and between Rueppell's fox and fennec fox (*V. zerda*) (**c–d, g–h**). **(a, c, e, g)** The pattern of ABBA/BABA in **A** and **E** is as follows: red fox Eurasia (H1) and North African red foxes (H2) have a recent common ancestor that diverged from the ancestral of the Rueppell's fox (H3), using the arctic fox as outgroup (O) in **A** and the dog as outgroup (O) in **E**. The pattern of ABBA/BABA in **C** and **G**: all red foxes (H1) Rueppell's foxes (H2) have a recent common ancestor that diverged from the ancestral of the fennec fox (H3), using the pale fox as outgroup (O) in **C** and the dog as outgroup (O) in **E**. Illustrations by Margarida Laranjeira Rocha. (**b, d, f, h**) Manhattan plot of the Z-transformed D statistic for non-overlapping windows of 200 kb across all chromosomes, according to pattern described in **A, C, E, G** respectively. Each dot represents a window and windows from the same

chromosome have the same colour. Line threshold represents regions with Z score values above the 99th percentiles of the empirical distribution.

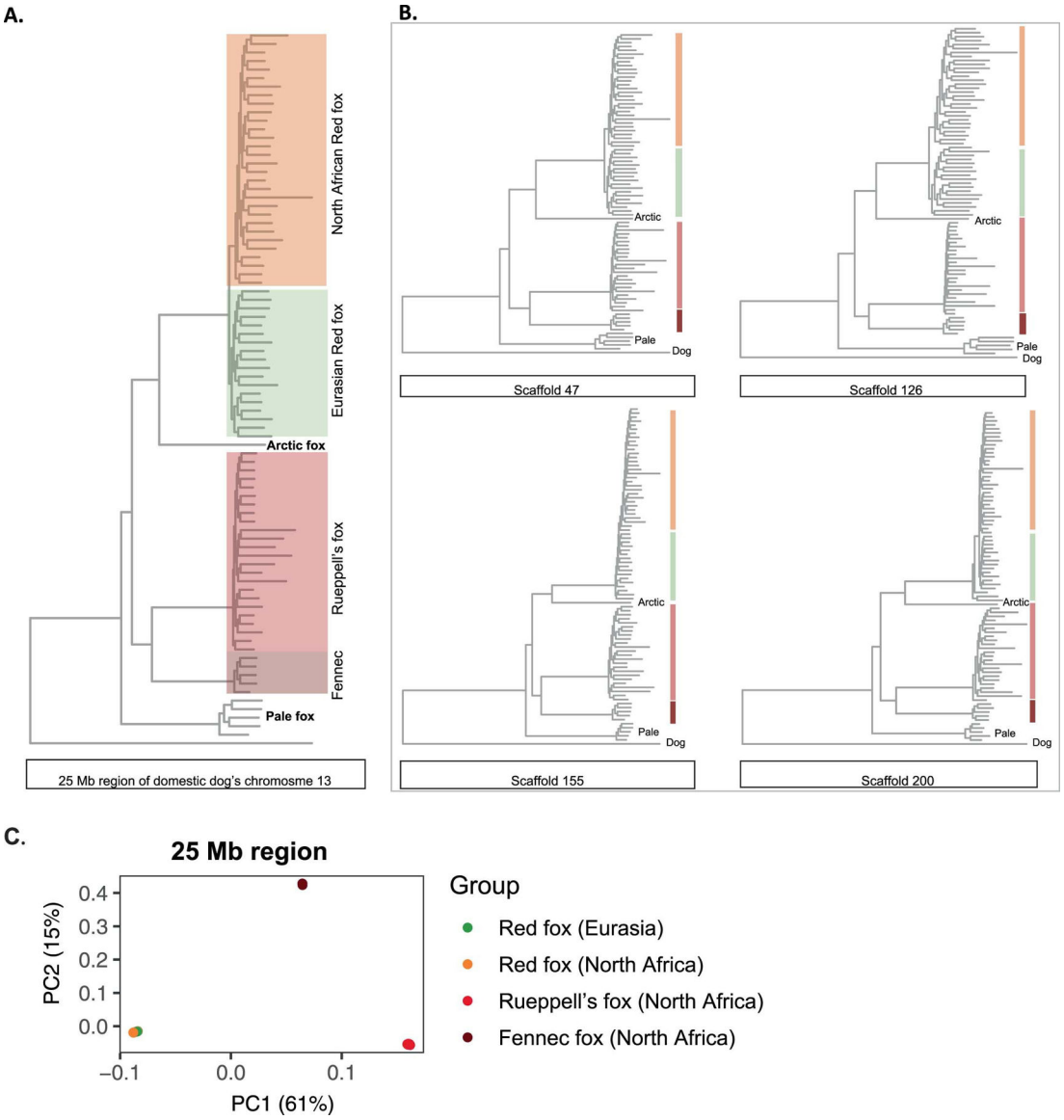


Extended Data Fig. 3 |. CRADD gene – top outlier above the 99th percentile of the empirical distribution for f_d and Z(D) in ABBA-BABA on windows (related to Fig. 3).

a) Maximum-likelihood tree estimated with IQtree for a sequence alignment of pseudohaploids comprising the CRADD gene (~ 180 kb) based on sampling of a random base (samples from our dataset are highlighted in the legend of the figure, see Supplementary Table 2 for additional genomes with legend in the tree). **(b)** PCA with 48 red foxes and 24 Rueppell's using 23,016 SNPs called from a 1.3 Mbp region containing CRADD; **(c)** Read depth-based estimate for CRADD per individual (depth for CRADD divided by the genome-wide coverage per individual). There is no evidence for CRADD copy number variation.

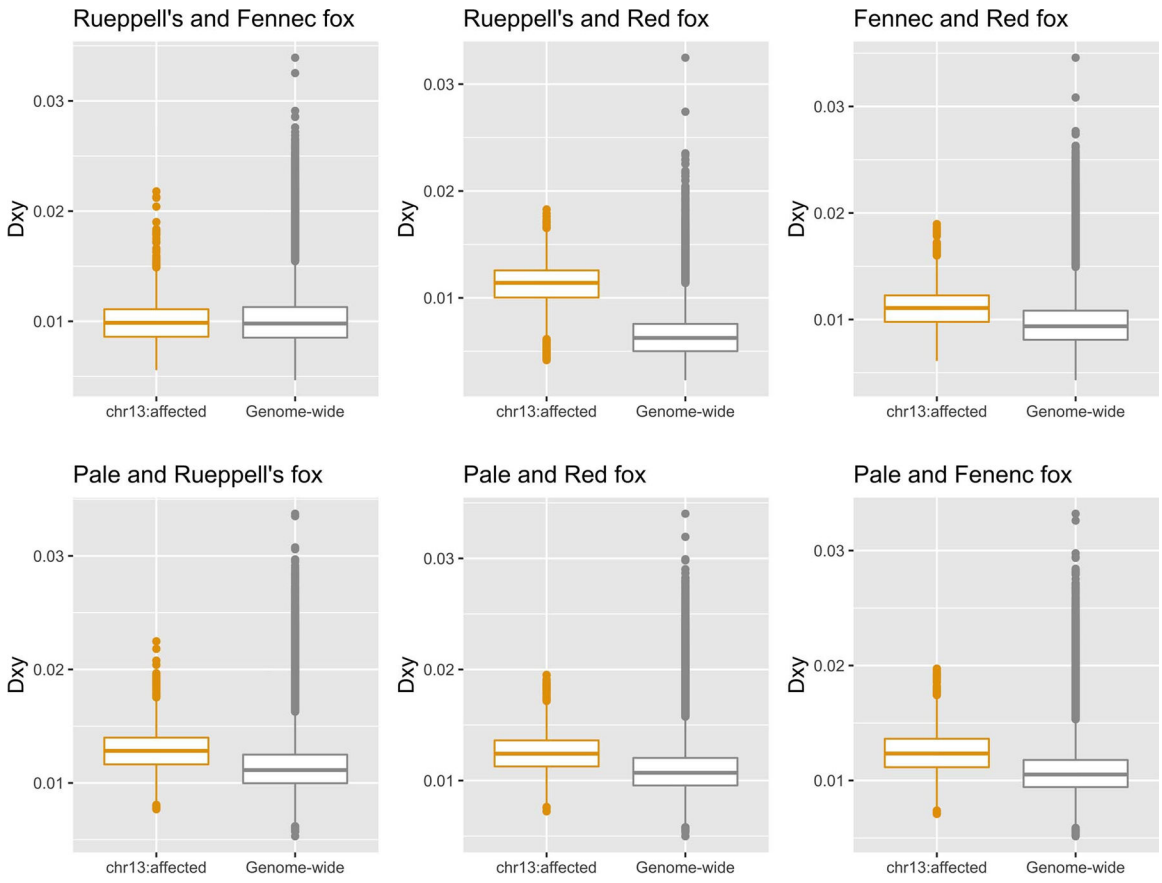


Extended Data Fig. 4 |. Principal component analysis for SLC6A16 (ranked 2nd) and two other genes (MGAM, HPS5) above the 99th percentile of f_d empirical distribution (related to Fig. 3). PCA based on sampling of a single read at each site. North African red foxes individuals (n = 30) unambiguously cluster closer to Rueppell's (n = 24) than other red foxes (n = 18). Arrow highlights hidden North African red fox individuals clustering within Rueppell's fox (MGAM). See also Figure S9 for gene tree topologies.



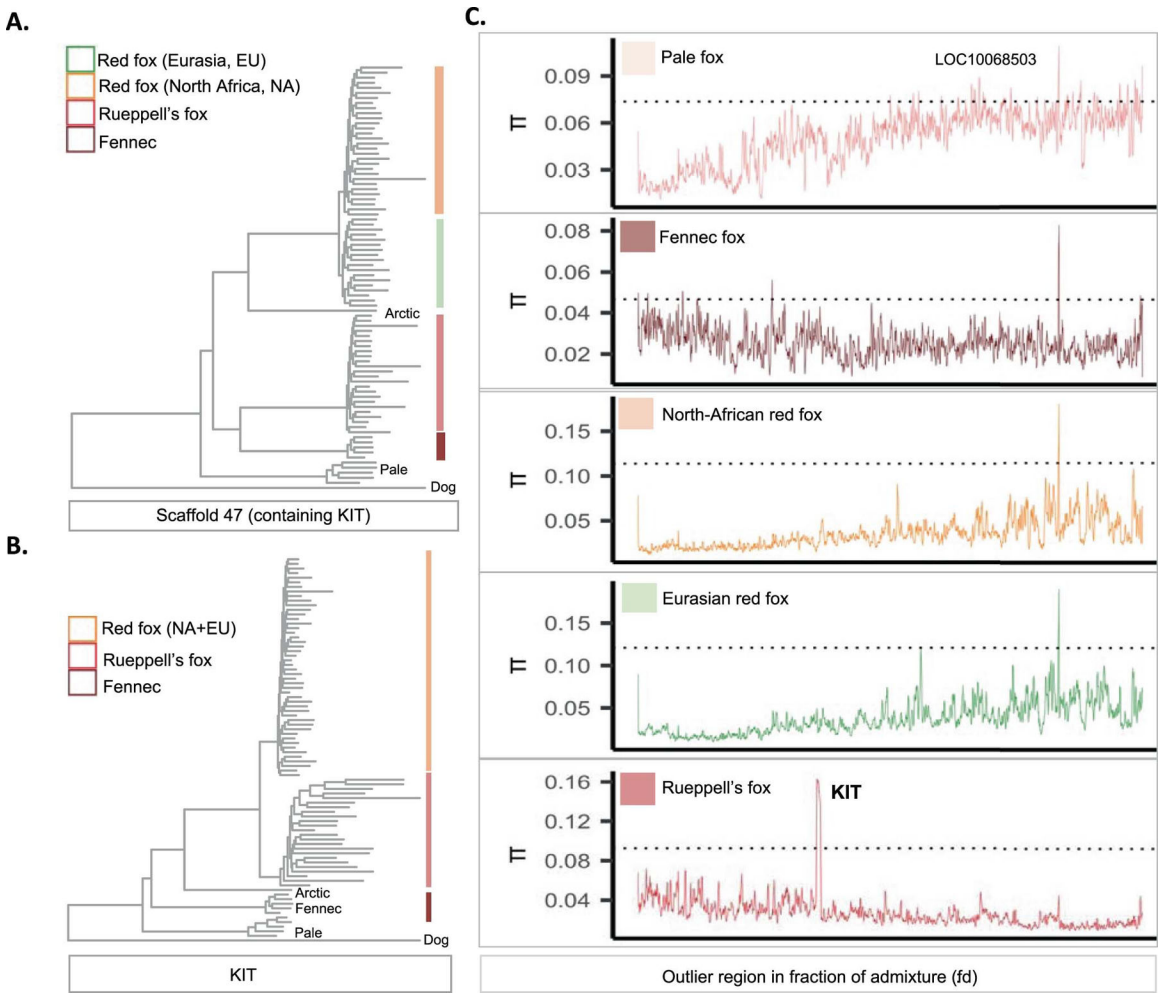
Extended Data Fig. 5 |. Tree topologies and Principal Component Analysis on the ~ 25 Mb outlier region above the 99th percentile of the empirical distribution for f_d in ABBA-BABA on windows (related to Fig. 3).

Maximum-likelihood tree estimated with IQtree for the affected chromosomal region called with ANGSD (-doHaplo 1) using both **(a)** the dog genome (NC_006595.3, 38236300–63241923), and **(b)** the arctic fox as references (scaffolds 47, 126, 155, 200); **(c)** PCA of the 25 Mb region in 48 red foxes, 24 Rueppell's foxes and 5 fennecs, based on sampling of a single read at each site. PC1 first separates Rueppell's and fennec and PC2 separates Rueppell's and red foxes.



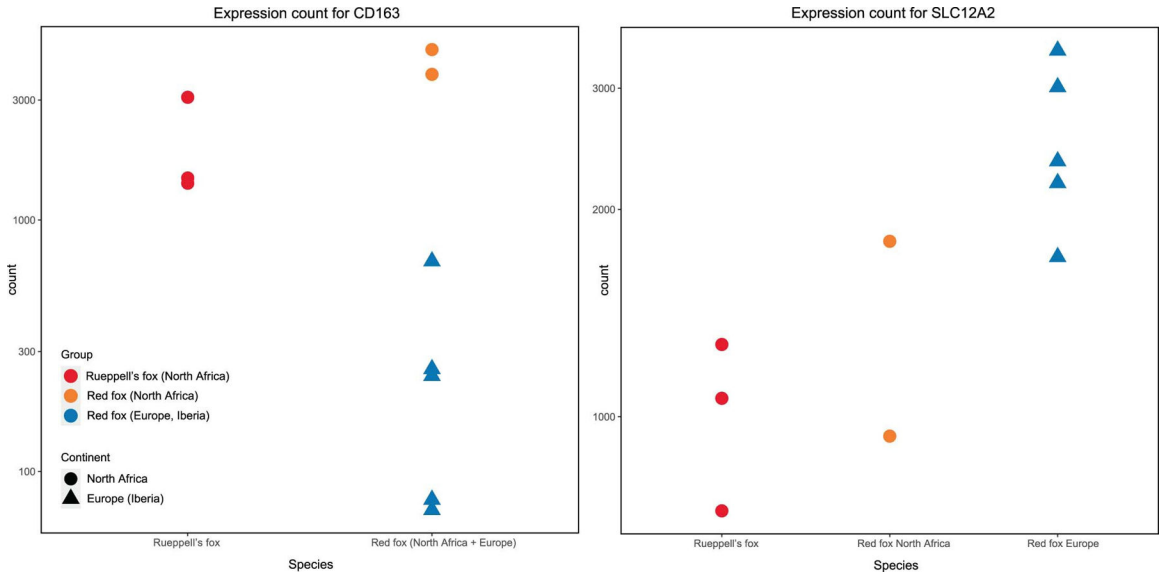
Extended Data Fig. 6 |. Genetic Divergence for the 25 Mb region above the 99th percentile of the empirical distribution for f_d on windows (related to Fig. 3).

Estimated genetic divergences (d_{XY}) between different pairs of *Vulpes* species on sliding windows of 50 kb overlapped every 10 kb for the affected chromosomal region and for all the autosomes (excluding the CRADD gene). Boxes indicate upper and lower quartiles; centre line represents median; whiskers extend to minimum and maximum values within 1.5x interquartile range; points show outliers beyond whiskers. Sample sizes for pairs of species are as follows: n = 24 Rueppell's fox, n = 48 Red fox.



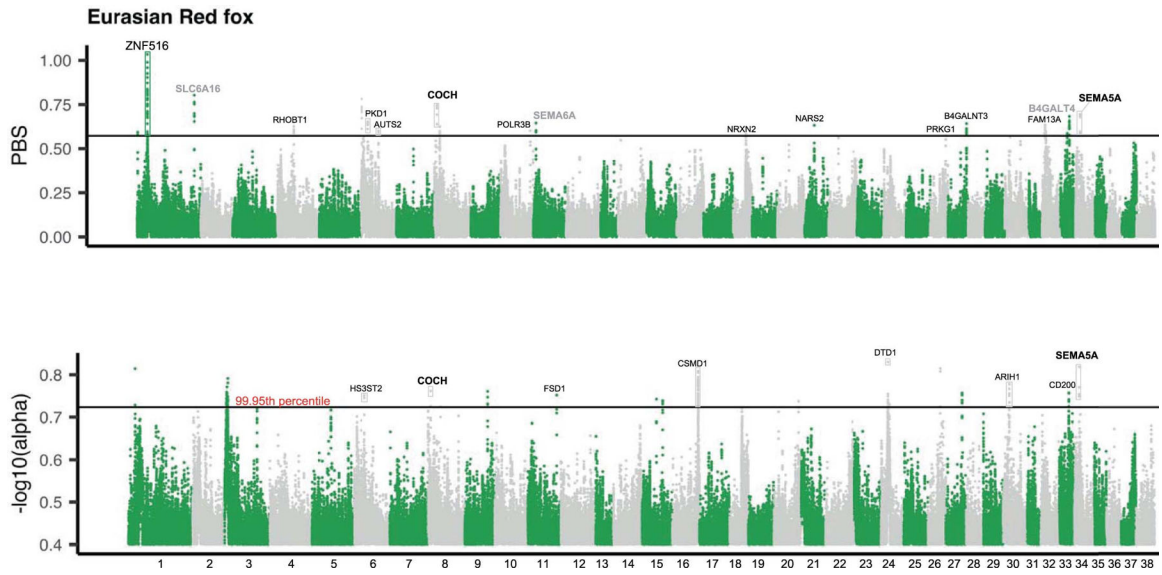
Extended Data Fig. 7 |. Evidence for potential introgression of the 25 Mb region supported by KIT (related to Fig. 3).

(a) Tree topology for one of the four scaffolds comprising the 25 Mb region above the 99th percentile of the empirical distribution for f_d in ABBA-BABA on windows; fennec and Rueppell's are sister species. **(b)** Tree topology for KIT; Rueppell's (n = 24) and red foxes (n = 48) become sister species, as in the average genome-wide tree. **(c)** Nucleotide diversity for the 25 Mb outlier region in 4 *Vulpes* species; KIT stands as being highly polymorphic in Rueppell's fox, suggesting the simultaneous mapping of divergent alleles likely introduced via introgression and alleles located in the B chromosomes which are closer to red foxes. Dotted lines show the 99.95th percentile of the empirical distribution of nucleotide diversity for the region. The fact that KIT is not highly polymorphic in red foxes, which also have autosomal and B alleles mapped to the same region further confirms our hypothesis of shared structural variation between Rueppell's fox and the fennec.



Extended Data Fig. 8 |. Gene expression differences between North African and Eurasian red foxes and Rueppell's foxes (related to Fig. 4).

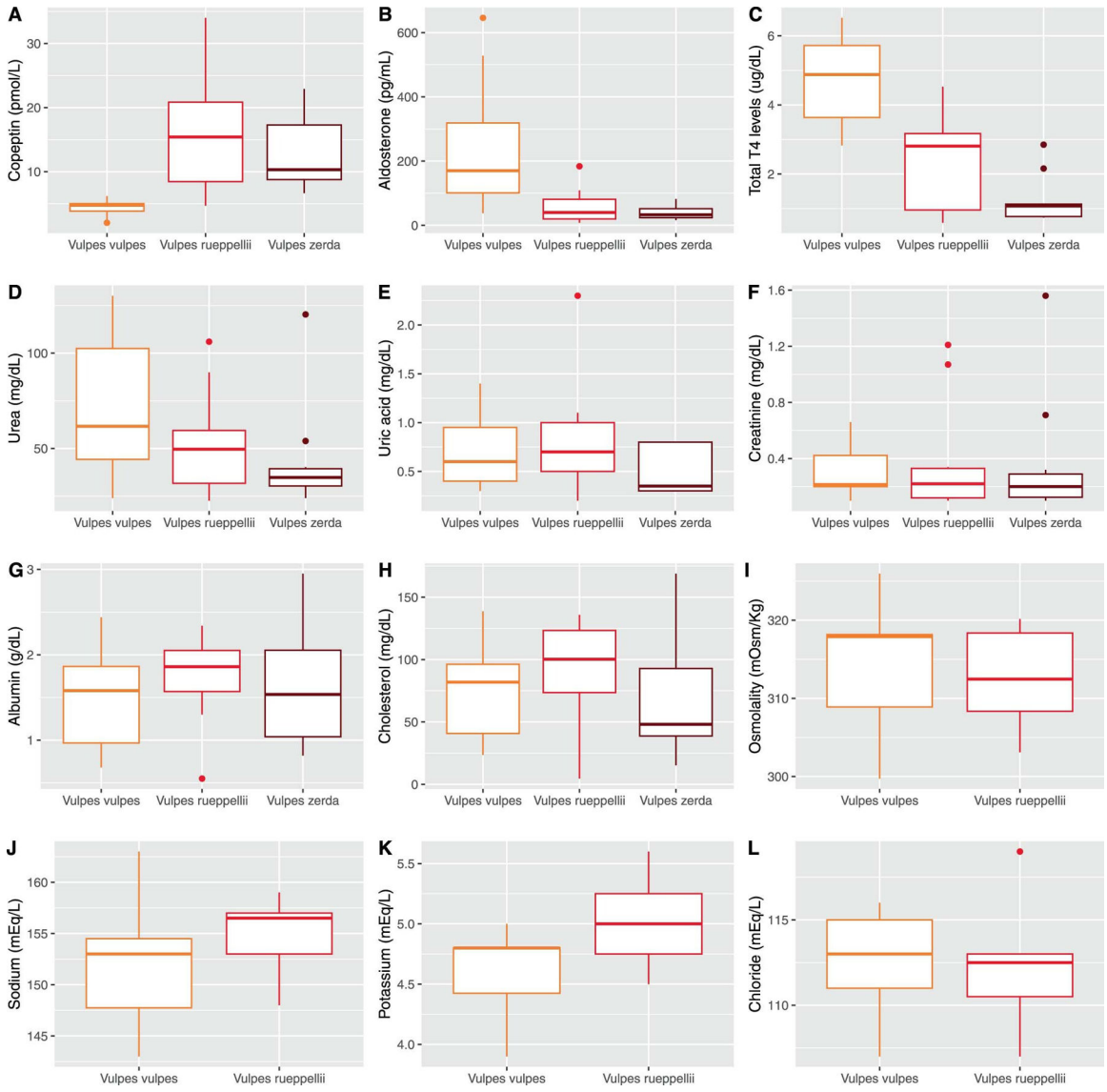
The plots represent top candidate genes putatively under selection in North African red foxes (CD163, SLC12A2) for which differential expression was significant between North African and Eurasian foxes (p-adjusted CD163 = 9.3e-08, p-adjusted SLC12A2 = 0.001), and/or North African red fox and Eurasian red foxes (p-adjusted CD163 = 2.5e-06; p-adjusted SLC12A2 = 0.079). The arctic fox transcriptome was used as reference.



Extended Data Fig. 9 |. Selection specific to Eurasian red foxes (related to Fig. 4).

Genome-wide selection scans on windows of 50 kb with a 10 kb slide across autosomal chromosomes, displayed as in the dog genome (x axis): population branch statistics (PBS) and alpha (negative logarithm of alpha) were estimated using Eurasian red foxes as focal group, North African red foxes as closest ingroup and Rueppell's fox as outgroup; genes in grey are genes that stand out above the 99th percentile of the empirical distribution of

f statistics in ABBA-BABA on windows for introgression between Rueppell's and North African red foxes. The CRADD region and the 25 Mb region identified in Fig. 3 in main text were removed from the analysis. Lines show the 99.95th percentile of the empirical distribution. Names of genes within the highest peaks are shown (non-label peaks represent no overlap with genes). The arctic fox (*V. lagopus*) was used as reference genome. We note that other genes not displayed in the figure can overlap the outlier regions; a full list can be found in Supplementary Tables S16–S17.



Extended Data Fig. 10 |. Physiological differences between three species of *Vulpes* (related to Supplementary Tables 22–23 and Fig. 5).

Boxplots of 12 physiological parameters measured from the serum/plasma of red fox (*V. vulpes*; n = 10), desert-dwelling Rueppell's fox (*V. rueppellii*; n = 13), and fennec fox (*V. zerda*; n = 10): (**a**) Copeptin (used as biomarker for vasopressin; pmol/L), (**b**) Aldosterone (pg/mL), (**c**) Total T4 (ug/dL), (**d**) Urea (mg/dL), (**e**) Uric acid (mg/dL), (**f**) Creatinine

(mg/dL); (**g**) Albumin (g/dL), (**h**) Cholesterol (mg/dL), (**i**) Osmolality (mOSm/Kg), (**j**) Sodium (mEq/L), (**k**) Potassium (mEq/L), (**l**) Chloride (mEq/L); Juveniles were excluded. Boxes indicate upper and lower quartiles; centre line represents median; whiskers extend to minimum and maximum values within 1.5× interquartile range; points show outliers beyond whiskers. Additional pairwise t-tests comparing the three different fox species were performed with Bonferroni correction for multiple comparisons on each parameter, as well as Welch's two-sample t-test comparing males versus females, North African versus European red foxes, and captive versus free-ranging individuals. All tests are two-sided. No significant differences between species were found except for A-C ($p_{Copeptin}$ = 0.00037; $p_{Aldosterone}$ = 0.0034; p_{T4} = 8. 8e-05) and between red fox and fennec ($p_{Copeptin}$=0.02273; $p_{Aldosterone}$ = 0.0027; p_{T4} = 1.0e-06). No significant differences were found between Rueppell's and fennec for copeptin (AVP) and Aldosterone ($P_{Copeptin}$= 0.55044; $P_{Aldosterone}$ = 1.00; P_{T4} = 0.16), nor between captive and wild-caught individuals ($P_{Copeptin}$=0.9811; $P_{Aldosterone}$ = 0.3522; P_{T4} = 0.3867). After correcting for the potentially confounding factors caused by sex, weight, body mass index and whether the individuals were caught free-ranging or captive as covariates, we still found the same pattern of significance (Supplementary Table 23). This seems to suggest that, at least for these three hormones (A-C), there is no evidence of significant intraspecific differences in red foxes (n = 4 North Africa and n = 6 Iberia), and that the differences observed are not due to individuals being free-ranging or held in captivity, though this should be taken with caution given the small sample size (Supplementary Table 23).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank M. Samlali, B. Habib and M. Milanese for their assistance in sampling collection in North Africa. We thank J. V. López Bao, H. R. Maior and F. Alvares for support and assistance in sampling collection in Iberia and J. F. Layna (Spain), V. Soeiro (Parque Biológico de Gaia, Portugal), Temara Zoo and Dream Village (Morocco) for access to foxes in captivity. We thank L. Smith (UC Berkeley), P. Ribeiro, S. Mourão and S. Lopes (CIBIO, UP) for support during laboratory work and generation of data. We thank A. Múrias, C. Pacheco, D. Lobo, M. S. Ferreira, D. Brandt, D. Aguillar-Gomez, L. Pipes, H. Wang and T. Linderoth for helpful discussions during data analysis and interpretation of results. We thank L. Dalén and A. Gaur for kindly sharing the arctic fox reference genome and Bengal fox mitogenome, respectively. We acknowledge M. Laranjeira Rocha for the illustrations featured in this work. This project was supported by the Portuguese Foundation for Science and Technology, FCT (PTDC/BIA-EVF/31902/2017 granted to R.G.) and the National Institutes of Health (R01GM138634 granted to R.N.). J.L.R., N.S., M.N., J.C.B. and R.G. worked under scholarships or research contracts from FCT (SFRH/ BD/ 116397/2016, SFRH/BPD/116596/2016, SFRH/BD/144087/2019, CEECINST/00014/2018/CP1512/CT0001 and 2021/00647/CEECIND, respectively).

Data availability

The sequencing data (genomic and transcriptomic raw reads) generated in this study are available through NCBI repositories linked to BioProject accession number PRJNA951250. Physiological and morphological data per individual are available from GitHub (https://doi.org/10.5281/zenodo.7826260)[132].

References

1. Williams M Climate Change in Deserts: Past, Present and Future (Cambridge Univ. Press, 2014).
2. Riddell EA et al. Exposure to climate change drives stability or collapse of desert mammal and bird communities. Science 371, 633–638 (2021). [PubMed: 33542137]
3. Liu Y & Xue Y Expansion of the Sahara desert and shrinking of frozen land of the Arctic. Sci. Rep 10, 4109 (2020). [PubMed: 32139761]
4. Sillero-Zubiri C et al. Canids: Foxes, Wolves, Jackals and Dog s(IUCN, 2004).
5. Brito JC et al. Biogeography and conservation of taxa from remote regions: an application of ecological-niche based models and GIS to North-African canids. Biol. Conserv 142, 3020–3029 (2009).
6. Williams JB et al. Energy expenditure and water flux of Rüppell's foxes in Saudi Arabia. Physiol. Biochem. Zool 75, 479–488 (2002). [PubMed: 12529849]
7. Williams JB et al. A phylogenetic analysis of basal metabolism, total evaporative water loss, and life-history among foxes from desert and mesic regions. J. Comp. Physiol. B 174, 29–39 (2004). [PubMed: 14564467]
8. Careau V et al. Basal metabolic rate of Canidae from hot deserts to cold arctic climates. J. Mammal 88, 394–400 (2007).
9. Noll-Banholzer U Water balance and kidney structure in the fennec. Comp. Biochem Physiol. A 62, 593–597 (1979).
10. Maloiy GMO et al. Thermoregulation and metabolism in a small desert carnivore: the fennec fox (Fennecus zerda) (Mammalia). J. Zool 198, 279–291 (1982).
11. Rocha JL et al. Convergent evolution of increased urine-concentrating ability in desert mammals. Mamm. Rev 51, 482–491 (2021).
12. Leite JV et al. Differentiation of North African foxes and population genetic dynamics in the desert —insights into the evolutionary history of two sister taxa, Vulpes rueppellii and Vulpes vulpes. Org. Divers Evol 15, 731–745 (2015).
13. Sinding MHS et al. Arctic-adapted dogs emerged at the Pleistocene–Holocene transition. Science 368, 1495–1499 (2020). [PubMed: 32587022]
14. Wang MS et al. Ancient hybridization with an unknown population facilitated high-altitude adaptation of canids. Mol. Biol. Evol 37, 2616–2629 (2020). [PubMed: 32384152]
15. von Holdt B et al. EPAS1 variants in high altitude Tibetan wolves were selectively introgressed into highland dogs. PeerJ 2017, e3522 (2017).
16. Li Y et al. Population variation revealed high-altitude adaptation of Tibetan mastiffs. Mol. Biol. Evol 31, 1200–1205 (2014). [PubMed: 24520091]
17. Wang G-D et al. Genetic convergence in the adaptation of dogs and humans to the high-altitude environment of the Tibetan plateau. Genome Biol. Evol 6, 2122–2128 (2014). [PubMed: 25091388]
18. Miao B et al. Genomic analysis reveals hypoxia adaptation in the Tibetan mastiff by introgression of the gray wolf from the Tibetan plateau. Mol. Biol. Evol 34, 734–743 (2017). [PubMed: 27927792]
19. Signore AV et al. Adaptive changes in hemoglobin function in high-altitude Tibetan canids were derived via gene conversion and introgression. Mol. Biol. Evol 36, 2227–2237 (2019). [PubMed: 31362306]
20. Korneliussen TS et al. ANGSD: analysis of next generation sequencing data. BMC Bioinformatics 15, 356 (2014). [PubMed: 25420514]
21. Hasselgren M et al. Genomic and fitness consequences of inbreeding in an endangered carnivore. Mol. Ecol 30, 2790–2799 (2021). [PubMed: 33955096]
22. Lindblad-Toh K et al. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature 438, 803–819 (2005). [PubMed: 16341006]
23. Statham MJ et al. Range-wide multilocus phylogeography of the red fox reveals ancient continental divergence, minimal genomic exchange and distinct demographic histories. Mol. Ecol 23, 4813–4830 (2014). [PubMed: 25212210]

24. McDevitt AD et al. Next-generation phylogeography resolves post-glacial colonization patterns in a widespread carnivore, the red fox (Vulpes vulpes), in Europe. Mol. Ecol 31, 993–1006 (2022). [PubMed: 34775636]
25. Noskova E et al. GADMA: genetic algorithm for inferring demographic history of multiple populations from allele frequency spectrum data. Gigascience 9, giaa005 (2020). [PubMed: 32112099]
26. Gutenkunst RN et al. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet 5, 1000695 (2009).
27. Bartolini Lucenti S & Madurell-Malapeira J Unraveling the fossil record of foxes: an updated review on the Plio-Pleistocene Vulpes spp. from Europe. Quat. Sci. Rev 236, 106296 (2020).
28. Hoffmann DL et al. Timing and causes of North African wet phases during the last glacial period and implications for modern human migration. Sci. Rep 6, 36367 (2016). [PubMed: 27808237]
29. Drake N & Breeze P in Africa from MIS 6–2: Population Dynamics and Paleoenvironments (eds Jones SC& Stewart BA) 103–122 (Springer, 2016).
30. Pausata FSR et al. The greening of the Sahara: past changes and future implications. One Earth 2, 235–250 (2020).
31. Osborne AH et al. A humid corridor across the Sahara for the migration of early modern humans out of Africa 120,000 years ago. Proc. Natl Acad. Sci. USA 105, 16444–16447 (2008). [PubMed: 18936490]
32. Castañeda IS et al. Wet phases in the Sahara/Sahel region and human migration patterns in North Africa. Proc. Natl Acad. Sci. USA 106, 20159–20163 (2009). [PubMed: 19910531]
33. Green RE et al. A draft sequence of the neandertal genome. Science 328, 710–722 (2010). [PubMed: 20448178]
34. Pickrell JK & Pritchard JK Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet 8, 1002967 (2012).
35. Martin SH et al. Evaluating the use of ABBA–BABA statistics to locate introgressed loci. Mol. Biol. Evol 32, 244–257 (2015). [PubMed: 25246699]
36. Ainsworth HC et al. Genetic epidemiology in kidney disease. Nephrol. Dial. Transplant 32, ii159–ii169 (2017). [PubMed: 28201750]
37. Pramod AB et al. SLC6 transporters: structure, function, regulation, disease association and therapeutics. Mol. Asp. Med 34, 197–219 (2013).
38. Hedrick PW Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. Mol. Ecol 22, 4606–4618 (2013). [PubMed: 23906376]
39. Charlesworth D Balancing selection and its effects on sequences in nearby genome regions. PLoS Genet 2, e64 (2006). [PubMed: 16683038]
40. Makunin AI et al. Genes on B chromosomes of vertebrates. Mol. Cytogenet 7, 99 (2014). [PubMed: 25538793]
41. Graphodatsky AS et al. Phylogenomics of the dog and fox family (Canidae, Carnivora) revealed by chromosome painting. Chromosome Res 16, 129–143 (2008). [PubMed: 18293108]
42. Ilardo MA et al. Physiological and genetic adaptations to diving in sea nomads. Cell 173, 569–580 (2018). [PubMed: 29677510]
43. Campbell-Staton SC et al. Ivory poaching and the rapid evolution of tusklessness in African elephants. Science 374, 483–487 (2021). [PubMed: 34672738]
44. Stödberg T et al. SLC12A2 mutations cause NKCC1 deficiency with encephalopathy and impaired secretory epithelia. Neurol. Genet 6, e478 (2020). [PubMed: 32754646]
45. Evans RL et al. Severe impairment of salivation in Na+/K+/2Cl− cotransporter (NKCC1)-deficient mice. J. Biol. Chem 275, 26720–26726 (2000). [PubMed: 10831596]
46. Sinica V et al. Human and mouse TRPA1 are heat and cold sensors differentially tuned by voltage. Cells 9, 57 (2019). [PubMed: 31878344]
47. Xu M et al. TRPV1 and TRPA1 in lung inflammation and airway hyperresponsiveness induced by fine particulate matter (PM2.5). Oxid. Med. Cell Longev 2019, 7450151 (2019). [PubMed: 31281589]

48. Sinica V & Vlachová V Transient receptor potential ankyrin 1 channel: an evolutionarily tuned thermosensor. Physiol. Res 70, 363 (2021). [PubMed: 33982589]
49. Lynch VJ et al. Elephantid genomes reveal the molecular bases of woolly mammoth adaptations to the Arctic. Cell Rep 12, 217–228 (2015). [PubMed: 26146078]
50. Key FM et al. Human local adaptation of the TRPM8 cold receptor along a latitudinal cline. PLoS Genet 14, e1007298 (2018). [PubMed: 29723195]
51. Ward D The Biology of Deserts 2nd edn (Oxford Univ. Press, 2016).
52. Moraru A et al. THADA regulates the organismal balance between energy storage and heat production. Dev. Cell 41, 72–81 (2017). [PubMed: 28399403]
53. Dempersmier J et al. Cold-inducible Zfp516 activates UCP1 transcription to promote browning of white fat and development of brown fat. Mol. Cell 57, 235–246 (2015). [PubMed: 25578880]
54. Cheng JY et al. Detecting selection in multiple populations by modeling ancestral admixture components. Mol. Biol. Evol 39, msab294 (2022).
55. Kurimoto J et al. Deficiency of WFS1 leads to the impairment of AVP secretion under dehydration in male mice. Pituitary 1, 3 (2021).
56. Bridges TE & James NV The hypothalamo-neurohypophysial system of native Australian desert rodents. The vasopressin and oxytocin contents of Notomys alexis and Pseudomys australis compared with those of the laboratory rat and mouse in different states of water balance. Aust. J. Exp. Biol. Med. Sci 60, 265–283 (1982).
57. Donald J & Pannabecker TL in Sodium and Water Homeostasis (eds Hyndman K & Pannabecker T) 191–211 (Springer, 2015).
58. Schwimmer H & Haim A Physiological adaptations of small mammals to desert ecosystems. Integr. Zool 4, 357–366 (2009). [PubMed: 21392308]
59. Beuchat CA Body size, medullary thickness, and urine concentrating ability in mammals. Am. J. Physiol. Regul. Integr. Comp. Physiol 258, R298–R308 (1990).
60. Solanki AK et al. Mutations in KIRREL1, a slit diaphragm component, cause steroid-resistant nephrotic syndrome. Kidney Int 96, 883–889 (2019). [PubMed: 31472902]
61. Noll-Banholzer U Body temperature, oxygen consumption, evaporative water loss and heart rate in the fennec. Comp. Biochem Physiol. A 62, 585–592 (1979).
62. Ishikawa A & Kitano J in Thyroid Hormone (ed. Agrawal NK) Ch. 2 (IntechOpen, 2012).
63. Geffen E & Girard I in The Swift Fox: Ecology and Conservation of Swift Foxes in a Changing World (eds Sovada MA & Carbyn L) 223–229 (Univ. Regina Press, 2003).
64. Ferreira MS et al. The legacy of recurrent introgression during the radiation of hares. Syst. Biol 70, 593–607 (2021). [PubMed: 33263746]
65. Jones MR et al. Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. Science 360, 1355–1358 (2018). [PubMed: 29930138]
66. Edelman NB et al. Genomic architecture and introgression shape a butterfly radiation. Science 366, 594–599 (2019). [PubMed: 31672890]
67. Oziolor EM et al. Adaptive introgression enables evolutionary rescue from extreme environmental pollution. Science 364, 455–457 (2019). [PubMed: 31048485]
68. Witt KE & Huerta-Sánchez E Convergent evolution in human and domesticate adaptation to high-altitude environments. Phil. Trans. R. Soc. B 374, 20180235 (2019). [PubMed: 31154977]
69. Cao YH et al. Historical introgression from wild relatives enhanced climatic adaptation and resistance to pneumonia in sheep. Mol. Biol. Evol 38, 838–855 (2021). [PubMed: 32941615]
70. Marques DA et al. A combinatorial view on speciation and adaptive radiation. Trends Ecol. Evol 34, 531–544 (2019). [PubMed: 30885412]
71. vonHoldt BM et al. Admixture mapping identifies introgressed genomic regions in North American canids. Mol. Ecol 25, 2443–2453 (2016). [PubMed: 27106273]
72. Koepfli KP et al. Genome-wide evidence reveals that African and Eurasian golden jackals are distinct species. Curr. Biol 25, 2158–2165 (2015). [PubMed: 26234211]
73. Gopalakrishnan S et al. Interspecific gene flow shaped the evolution of the genus Canis. Curr. Biol 28, 3441–3449 (2018). [PubMed: 30344120]

74. Racimo F et al. Archaic adaptive introgression in TBX15/WARS2. Mol. Biol. Evol 34, msw283 (2016).
75. Huerta-Sánchez E et al. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature 512, 194–197 (2014). [PubMed: 25043035]
76. Geffen E in The Biology and Conservation of Wild Canids (eds Macdonald DW& Sillero-Zubiri C) 199–206 (Oxford Univ. Press, 2004).
77. Kimura R et al. Gene flow and natural selection in oceanic human populations inferred from genome-wide SNP typing. Mol. Biol. Evol 25, 1750–1761 (2008). [PubMed: 18524786]
78. Jamie GA & Meier JI The persistence of polymorphisms across species radiations. Trends Ecol. Evol 35, 795–808 (2020). [PubMed: 32408996]
79. Thomas L et al. Spatially varying selection between habitats drives physiological shifts and local adaptation in a broadcast spawning coral on a remote atoll in Western Australia. Sci. Adv 8, 9185 (2022).
80. Cardona A et al. Genome-wide analysis of cold adaptation in indigenous Siberian populations. PLoS ONE 9, e98076 (2014). [PubMed: 24847810]
81. Vale CG & Brito JC Desert-adapted species are vulnerable to climate change: insights from the warmest region on Earth. Glob. Ecol. Conserv 4, 369–379 (2015).
82. Leclère D et al. Bending the curve of terrestrial biodiversity needs an integrated strategy. Nature 585, 551–556 (2020). [PubMed: 32908312]
83. Rocha JL et al. Life in deserts: the genetic basis of mammalian desert adaptation. Trends Ecol. Evol 36, 637–650 (2021). [PubMed: 33863602]
84. Global Aridity and PET Database (CGIAR-CSI, 2009);https://cgiarcsi.community/data/global-aridity-and-pet-database/
85. Knepper MA et al. Molecular physiology of water balance. N. Engl. J. Med 372, 1349–1358 (2015). [PubMed: 25830425]
86. Sands JM & Layton HE Advances in understanding the urine-concentrating mechanism. Annu. Rev. Physiol 76, 387–409 (2014). [PubMed: 24245944]
87. Xu MM & Wang DH Water deprivation up-regulates urine osmolality and renal aquaporin 2 in Mongolian gerbils (Meriones unguiculatus). Comp. Biochem Physiol. A 194, 37–44 (2016).
88. Morgenthaler NG Copeptin: a biomarker of cardiovascular and renal function. Congest. Heart Fail 16, S37–S44 (2010). [PubMed: 20653710]
89. Morgenthaler NG et al. Copeptin: clinical use of a new biomarker. Trends Endocrinol. Metab 19, 43–49 (2008). [PubMed: 18291667]
90. Hou Z et al. A cost-effective RNA sequencing protocol for large-scale gene expression studies. Sci. Rep 5, 9570 (2015). [PubMed: 25831155]
91. Martin M Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17, 10 (2011).
92. Bolger AM et al. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120 (2014). [PubMed: 24695404]
93. Li H Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM Preprint at https://arxiv.org/abs/1303.3997 (2013).
94. Li H et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079 (2009). [PubMed: 19505943]
95. Jónsson H et al. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics 29, 1682–1684 (2013). [PubMed: 23613487]
96. Kurtz S et al. Versatile and open software for comparing large genomes. Genome Biol 5, R12 (2004). [PubMed: 14759262]
97. Katoh K & Standley DM MAFFT Multiple Sequence Alignment Software Version 7: improvements in performance and usability. Mol. Biol. Evol 30, 772–780 (2013). [PubMed: 23329690]
98. Gouy M et al. SeaView Version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol. Biol. Evol 27, 221–224 (2010). [PubMed: 19854763]

99. Korneliussen TS & Moltke I NgsRelate: a software tool for estimating pairwise relatedness from next-generation sequencing data. Bioinformatics 31, 4009–4011 (2015). [PubMed: 26323718]
100. Dobin A et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21 (2013). [PubMed: 23104886]
101. Shumate A & Salzberg SL Liftoff: accurate mapping of gene annotations. Bioinformatics 37, 1639–1643 (2021). [PubMed: 33320174]
102. Okonechnikov K et al. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics 32, 292–294 (2016). [PubMed: 26428292]
103. Ewels P et al. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32, 3047–3048 (2016). [PubMed: 27312411]
104. Li H A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27, 2987–2993 (2011). [PubMed: 21903627]
105. Danecek P et al. The variant call format and VCFtools. Bioinformatics 27, 2156–2158 (2011). [PubMed: 21653522]
106. Chang CC et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4, 7 (2015). [PubMed: 25722852]
107. Love MI et al. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 550 (2014). [PubMed: 25516281]
108. Kumar S et al. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol 35, 1547–1549 (2018). [PubMed: 29722887]
109. Matzke NJ Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. Syst. Biol 63, 951–970 (2014). [PubMed: 25123369]
110. Matzke NJ Probabilistic Historical Biogeography: New Models for Founder-Event Speciation, Imperfect Detection, and Fossils Allow Improved Accuracy and Model-Testing PhD thesis, Univ. California (2013).
111. Sacks BN et al. Nuclear genetic analysis of the red fox across its Trans-Pacific range. J. Hered 109, 573–584 (2018). [PubMed: 29889225]
112. Drummond AJ & Rambaut A BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol 7, 214 (2007). [PubMed: 17996036]
113. Posada D jModelTest: phylogenetic model averaging. Mol. Biol. Evol 25, 1253–1256 (2008). [PubMed: 18397919]
114. Perini FA et al. The evolution of South American endemic canids: a history of rapid diversification and morphological parallelism. J. Evol. Biol 23, 311–322 (2010). [PubMed: 20002250]
115. Kumar V et al. Genetic signatures of adaptation revealed from transcriptome sequencing of Arctic and red foxes. BMC Genom 16, 585 (2015).
116. Zhao C et al. The complete mitochondrial genome of the Tibetan fox (Vulpes ferrilata) and implications for the phylogeny of Canidae. C. R. Biol 339, 68–77 (2016). [PubMed: 26868757]
117. Leigh JW & Bryant D popart: full-feature software for haplotype network construction. Methods Ecol. Evol 6, 1110–1116 (2015).
118. Vieira FG et al. Improving the estimation of genetic distances from next-generation sequencing data. Biol. J. Linn. Soc 117, 139–149 (2016).
119. Lefort V et al. FastME 2.0: a comprehensive, accurate, and fast distance-based phylogeny inference program. Mol. Biol. Evol 32, 2798–2800 (2015). [PubMed: 26130081]
120. Stamatakis A RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313 (2014). [PubMed: 24451623]
121. Soraggi S et al. Powerful inference with the D-statistic on low-coverage whole-genome data. G3 Genes Genomes Genet 8, 551–566 (2018).
122. Koch EM et al. De novo mutation rate estimation in wolves of known pedigree. Mol. Biol. Evol 36, 2536–2547 (2019). [PubMed: 31297530]
123. Kutschera VE et al. A range-wide synthesis and timeline for phylogeographic events in the red fox (Vulpes vulpes). BMC Evol. Biol 13, 114 (2013). [PubMed: 23738594]

124. Lonsinger RC et al. Evaluating effective population size and genetic diversity of a declining kit fox population using contemporary and historical specimens. Ecol. Evol 8, 12011–12021 (2018). [PubMed: 30598795]

125. Yi X et al. Sequencing of fifty human exomes reveals adaptations to high altitude. Science 329, 75–78 (2010). [PubMed: 20595611]

126. Kent WJ et al. The Human Genome Browser at UCSC. Genome Res 12, 996–1006 (2002). [PubMed: 12045153]

127. Malaspinas AS et al. A genomic history of Aboriginal Australia. Nature 538, 207–214 (2016). [PubMed: 27654914]

128. Crawford JE & Nielsen R Detecting adaptive trait loci in nonmodel systems: divergence or admixture mapping? Mol. Ecol 22, 6131–6148 (2013). [PubMed: 24128338]

129. Hudson RR et al. A test of neutral molecular evolution based on nucleotide data. Genetics 116, 153–159 (1987). [PubMed: 3110004]

130. Zhai W et al. An investigation of the statistical power of neutrality tests based on comparative and population genetic data. Mol. Biol. Evol 26, 273–283 (2009). [PubMed: 18922762]

131. Liu S et al. Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. Cell 157, 785–794 (2014). [PubMed: 24813606]

132. Rocha JL et al. Data and code from: North-African fox genomes show signatures of repeated introgression and adaptation to life in deserts. GitHub 10.5281/zenodo.7826260 (2023).

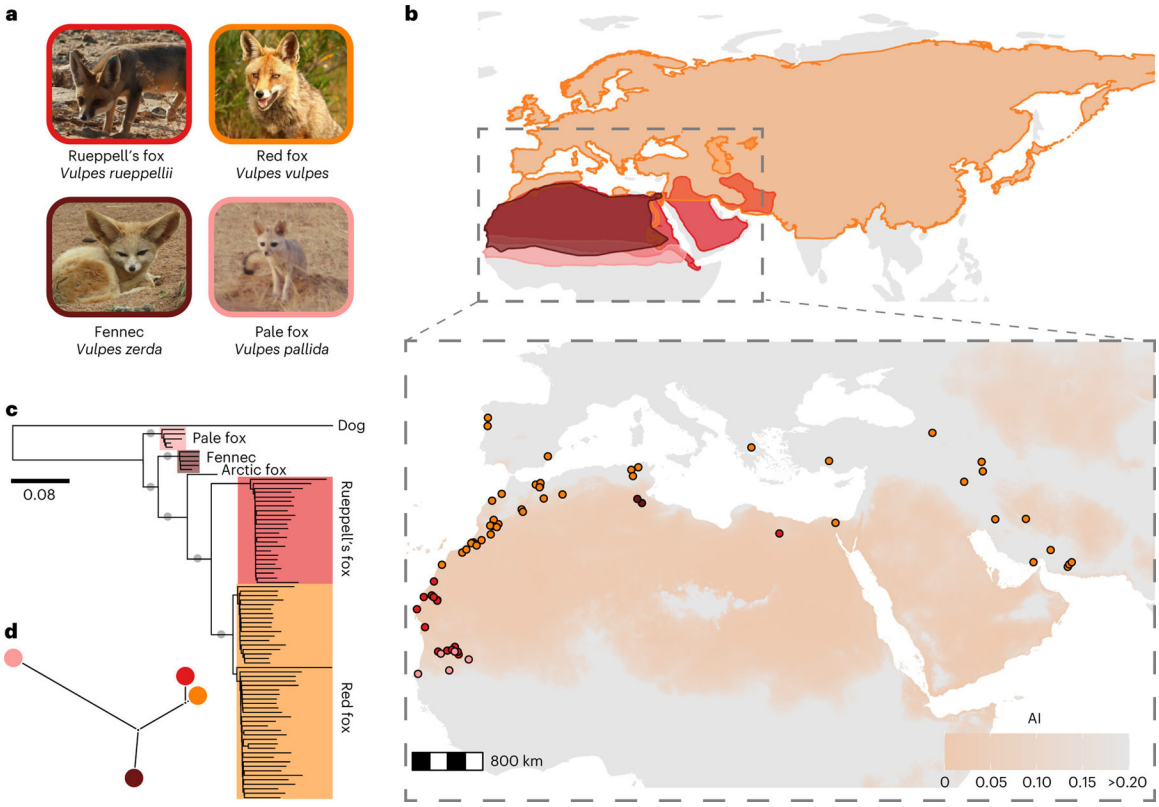


Fig. 1 |. Species relationships and geographic distribution of North African foxes.

a, Fox species (genus *Vulpes*) featured in this study. Colours outlining images denote the species in subsequent figures (photos by M.N. (Rueppell's fox), R. Rocha (red fox), A.Q. (fennec fox) and J.C.B. (pale fox)). **b**, North African fox geographic distribution. Colour shading on map corresponds to the North African and Eurasian geographic distribution of species featured in **a**. Shapefiles were downloaded from IUCN, 2021 (https://www.iucnredlist.org/); enlargement of the outlined region shows samples (n = 24 Rueppell's foxes, n = 48 red foxes, n = 5 fennec, n = 5 pale foxes) coloured according to species depicted over a projection of the aridity index (AI)[84]. Colour for AI ranges from 0 to 0.2, to define arid and hyperarid regions (AI < 0.2) and demark them from non-arid areas (AI > 0.2). **c**, RAxML average whole-genome tree constructed with genetic distances among 83 foxes on the basis of genotype likelihoods using the dog as outgroup. Genetic distances were calculated with ngsDist with genome-wide polymorphic sites. This was performed through concatenation of genomic data, which typically leads to long external branches and short internal branches within species. Grey circles indicate branches with >95% bootstrap support. **d**, Phylogenetic tree inferred with Ohana structure analysis for K = 4 using the four *Vulpes* species sequenced in this study (see Supplementary Fig. 2 for PCA and Extended Data Fig. 1 and Supplementary Figs. 3 and 4 for mtDNA analyses).

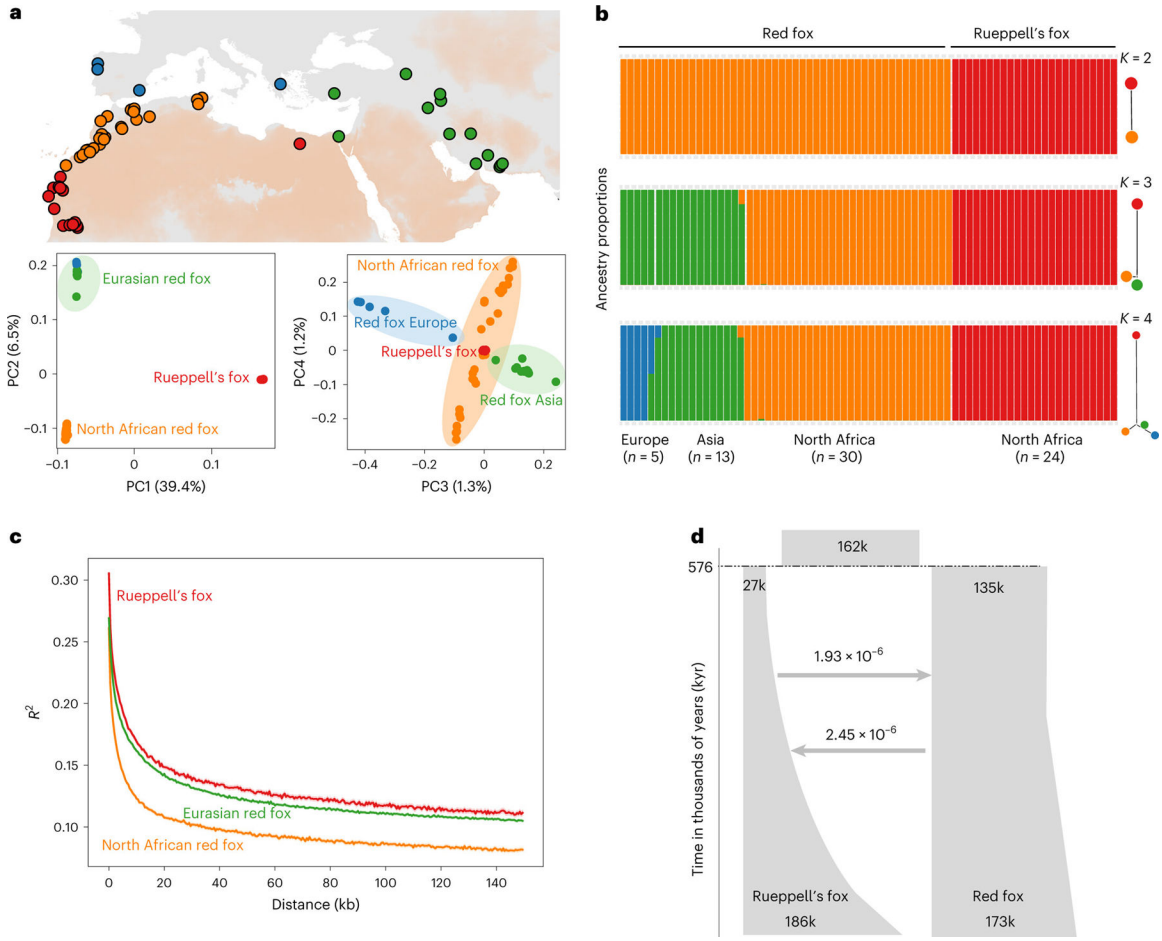


Fig. 2 |. Genetic structure, diversity and joint evolutionary history of North African red fox and Rueppell's fox.

a, PCA. Ellipses highlight geographical partition of genetic diversity according to continental region (Supplementary Fig. 6). **b**, Admixture proportions inferred with Ohana structure analysis and trees for $K = 2$–4. Each bar shows the inferred ancestry fraction for an individual. Labels above and below bars identify the species and geographical location (Supplementary Fig. 6). **c**, LD distribution of Rueppell's fox (red), North African red fox (orange) and Eurasian red fox (green). **d**, The joint demographic history of Rueppell's fox and red fox in North Africa inferred using diffusion approximation to the SFS as implemented in dadi using GADMA. Divergence times are represented in thousands of years. The estimated effective population sizes and the migration rates are also indicated (Supplementary Table 7).

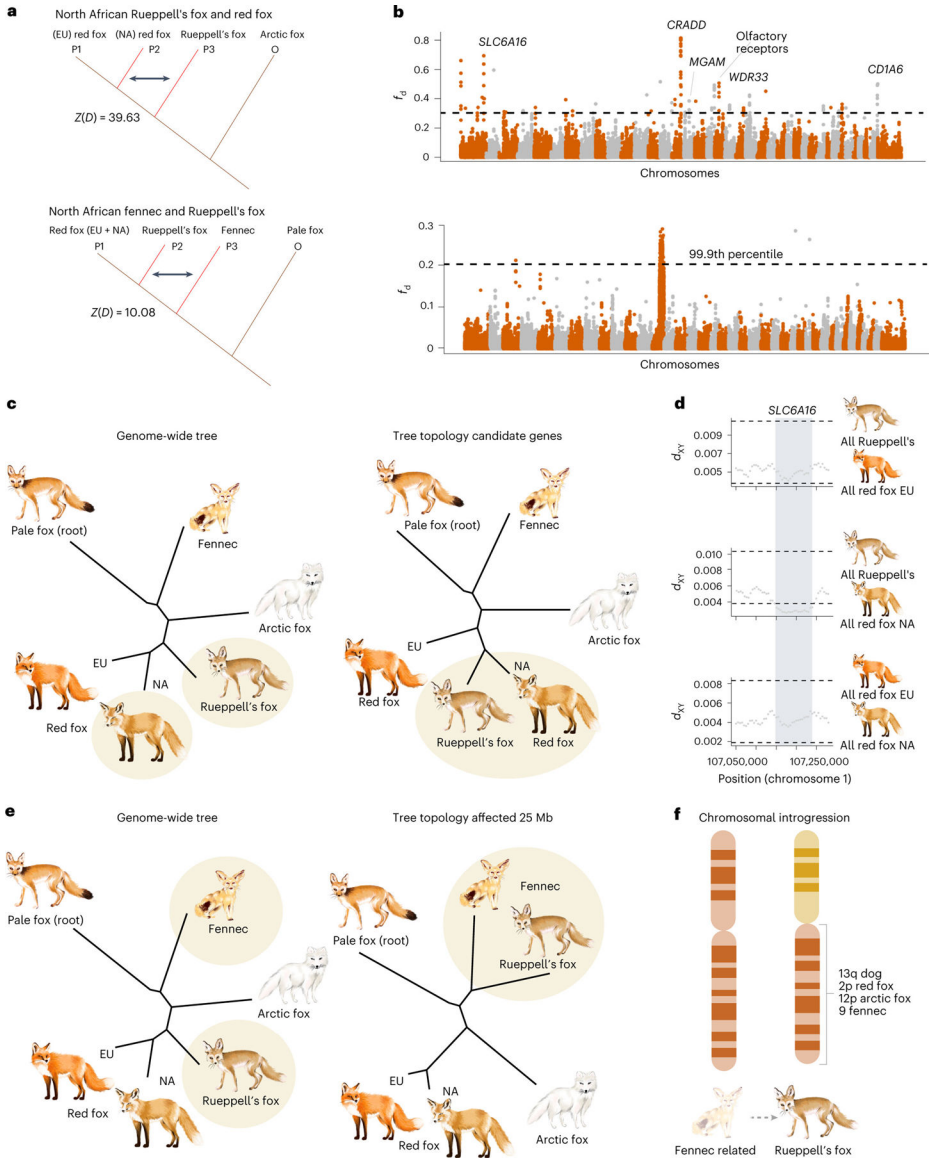


Fig. 3 |. Introgression in North African foxes.

a, Genome-wide $Z(D)$ for introgression between Rueppell's fox and North African red fox (top) and between Rueppell's fox and fennec (bottom; see also Supplementary Fig. 7). **b**, Manhattan plot of fraction of admixture (f_d) on 100 kb windows with 20 kb overlap across all chromosomes. Eurasian (EU) red foxes were used as P1, North African (NA) red foxes as P2 and Rueppell's fox as P3, with the arctic fox as outgroup (top), and all red foxes were used as P1, Rueppell's fox as P2 and fennec as P3, with the pale fox as outgroup (bottom). Windows from the same chromosome have the same colour. Dark line threshold separates regions above the 99.9th percentile (Supplementary Tables 8–11). The arctic fox was used as reference genome and scaffolds were ordered with respect to the dog genome (Supplementary Fig. 1). See also Extended Data Fig. 2 for $Z(D)$ on windows with different outgroups and reads mapped to the dog genome. **c**, Estimated tree topologies for *Vulpes* across whole-genome (left) and a 200 kb region harbouring SLC6A16 (right) ranked second

above the 99.9th percentile of the empirical distribution for f_d between Rueppell's fox and North African red fox (see Extended Data Figs. 3 and 4 and Supplementary Fig. 9 for deeper exploration of outliers). **d**, Genetic divergence (d_{XY}) between different pairs of fox lineages on sliding windows of 25 kb overlapped every 5 kb for chromosome harbouring gene *SLC6A16*. Threshold lines mark the 95% confidence interval of the genome-wide empirical distribution for d_{XY} (Supplementary Figs. 10–12). **e**, Estimated tree topologies for *Vulpes* across whole-genome (left) and scaffolds in the arctic fox genome comprising ~ 25 Mb above the 99th percentile of the empirical distribution of f_d between fennec and Rueppell's fox (Extended Data Fig. 5). **f**, Schematic representation for the hypothetical origin of a large chromosomal region harbouring alleles in strong LD in Rueppell's fox: a 25 Mb block shared between Rueppell's fox and a fox closely related to the fennec highlights a putatively adaptive chromosomal translocation (see Extended Data Figs. 5–7 and Supplementary Figs. 13 and 14 for more indepth analyses of the region). Fox illustrations by M. Laranjeira Rocha.

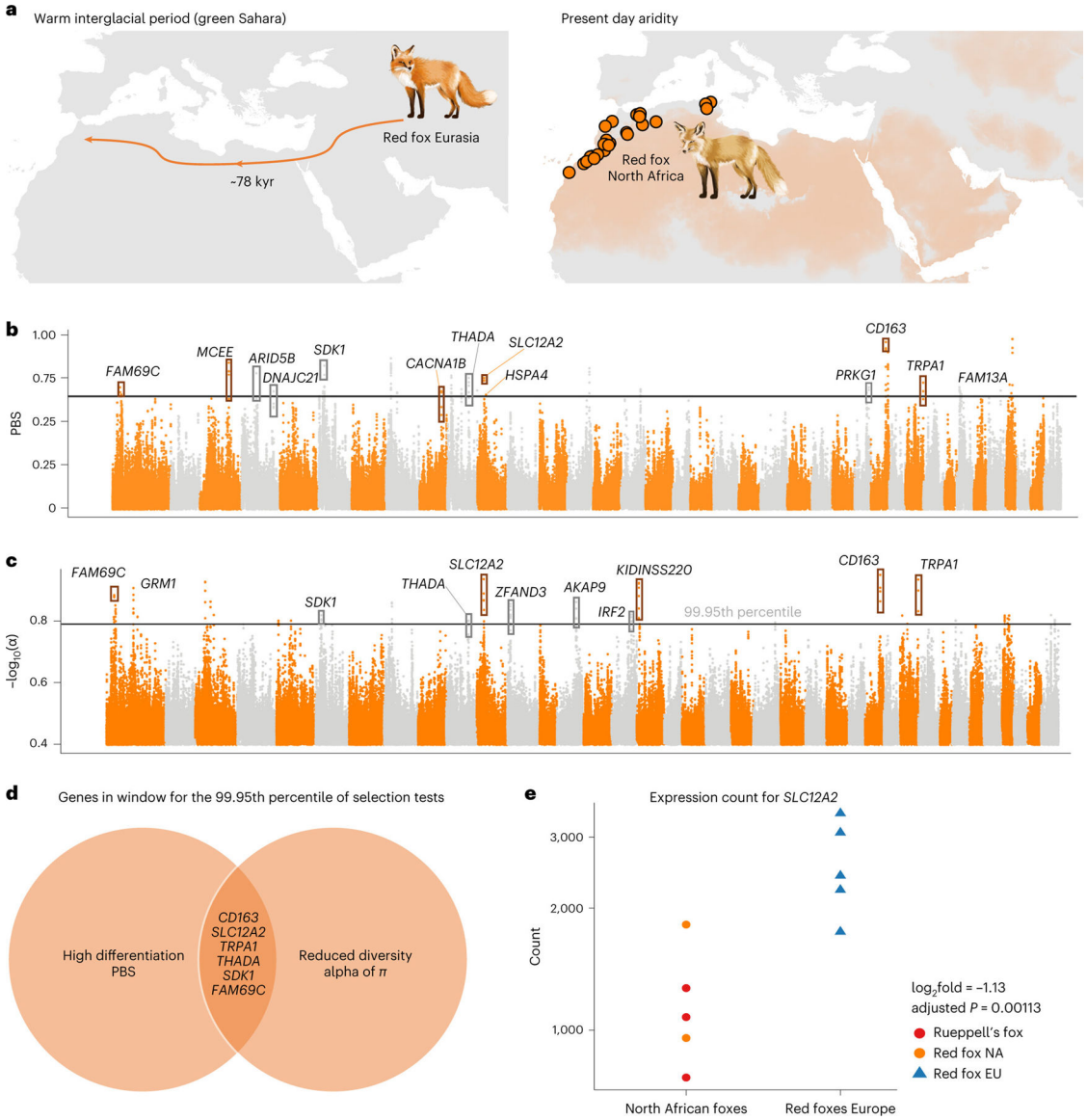


Fig. 4 |. Demographic and selection history of North African red fox.

a, Illustration of the demographic history of red fox North African colonization and divergence from Eurasia inferred by diffusion approximation to the SFS as implemented in dadi using GADMA. Map shows the time and presumable colonization route by which the red fox entered North Africa at the end of a warm interglacial period (left) and the sampling sites (orange dots) of red fox in North Africa. Fox illustrations by M. Laranjeira Rocha. **b**,**c**, Manhattan plots for genome-wide selection scans on windows of 50 kb with a 10 kb slide across autosomal chromosomes (*x* axis) using North African red foxes as focal group, Eurasian red foxes as closest ingroup and Rueppell's fox as outgroup (see Extended Data Fig. 9 for Eurasian red fox as focal group). Threshold lines show the 99.95th percentile of the empirical distribution PBS and negative logarithm of α. Windows from the same chromosome have the same colour. Names of genes within the highest peaks are shown (outliers in both methods are highlighted in bold). The arctic fox was used as reference

genome and scaffolds were ordered with respect to the dog genome (Supplementary Fig. 1). We note that other genes not displayed in the figure can overlap the outlier regions; a full list can be found in Supplementary Tables 12 and 13. **d**, Intersection of regions in the 0.5% tails of the genome-wide empirical distribution for selection tests based in high allele-frequency differentiation (PBS) and reduced genetic diversity (α). **e**, *SLC12A2* messenger RNA expression levels in blood of North African foxes ($n = 3$ Rueppell's foxes and $n = 2$ red foxes North Africa) and Eurasian foxes (red fox Iberia, $n = 5$); this gene is significantly less expressed in North African foxes ($\log_2$ fold-change =−1.13, adjusted $P =$ 0.00113; Extended Data Fig. 8).

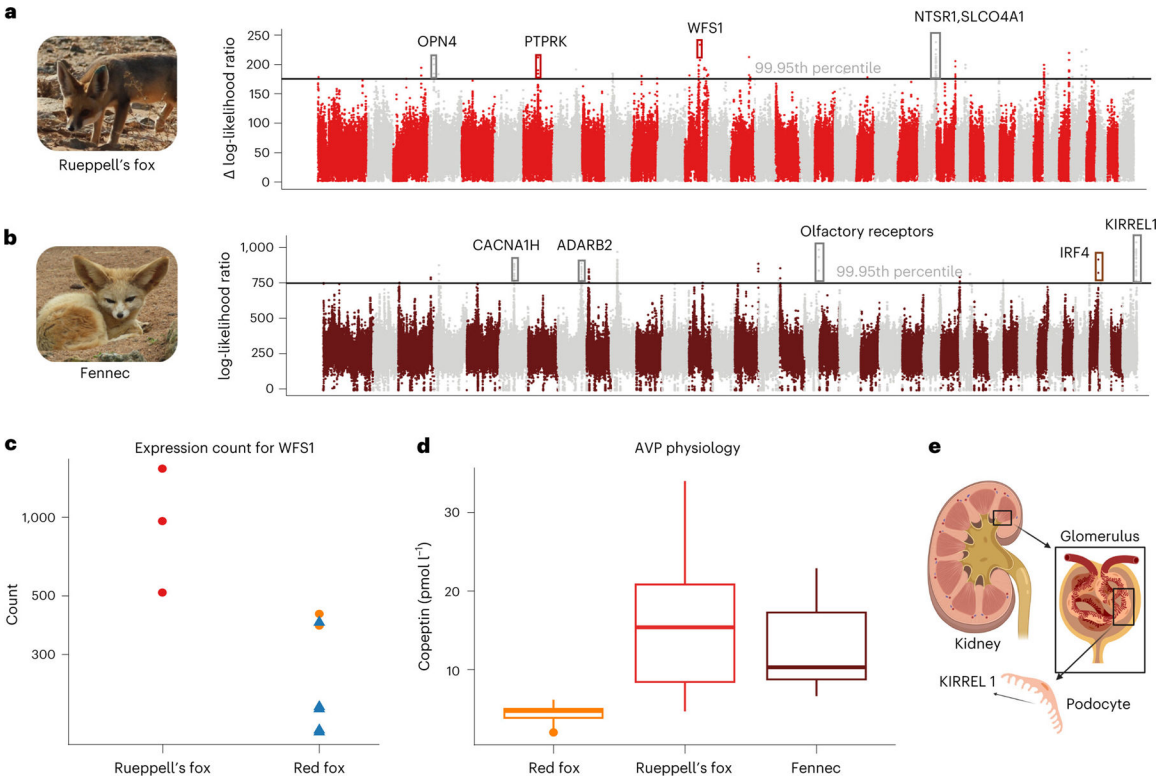


Fig. 5 |. Genome-wide selection scans for Rueppell's fox and fennec using Ohana based on the structure and tree results from *K* = 4 highlight unique adaptations supported by expression and physiological differences.

a,**b**, Manhattan plots of the log-likelihood ratio values for the top 500 SNPs on windows of 50 kb with a 10 kb slide across chromosomes (*x* axis) for Rueppell's fox (**a**) and fennec (**b**). This test compares a model in which the target/focal component experiences faster allele frequency changes in a specific genomic region or locus (selection hypothesis) than expected from the genome-wide distribution of allele frequency changes (global estimate). To account for selection specific to Rueppell's fox and not in their common ancestral with red foxes we subtracted the log-likelihood ratio of Rueppell's fox and red fox for the same windows (delta log-likelihood ratio). Threshold lines show the 99.95th percentile of the empirical distribution. Names of genes within the highest peaks are shown (full list can be found in Supplementary Tables 19 and 20). Scaffolds are ordered with respect to the dog genome with each colour representing a different chromosome. Photos by M.N (**a**) and A.Q. (**b**). **c**, WFS1 mRNA expression levels in blood of Rueppell's fox ($n = 3$) and red foxes ($n = 2$ North Africa, $n = 5$ Iberia). **d**, Blood levels of copeptin (biomarker for vasopressin, AVP) in red foxes ($n = 11$), Rueppell's fox ($n = 13$) and fennec ($n = 12$). Boxes indicate upper and lower quartiles; centre line represents median; whiskers extend to minimum and maximum values within 1.5× interquartile range; points show outliers beyond whiskers (see Supplementary Text, Supplementary Tables 22 and 23, Extended Data Fig. 10 and Supplementary Figs. 17 and 18 for other physiological parameters). **e**, Illustration of the location of podocytes—cells involved in ensuring size- and charge-selective ultrafiltration

at the kidney glomerulus— where protein KIRREL1 is produced and required for proper function of the glomerular filtration barrier (created with BioRender.com).