

Life in Deserts: a genomic history of adaptation and introgression in North-African foxes

Joana Maria Laranjeira Rocha

Doctoral Programme in Biodiversity, Genetics and Evolutiom

Departamento de Biologia, Faculdade de Ciências da Universidade do Porto

2021

Orientador

Dr. Maria Raquel Barata Godinho, Investigadora Auxiliar, Centro de Investigação em Biodiversidade e Recursos Geneticos (CIBIO-InBIO)

Coorientador

Professor Dr. Rasmus Nielsen, Full Professor, University of California, Berkeley

To my sister, Margarida,

To Gonzalo,

I dedicate this thesis.

“The words nostalgia for sand, nostalgia for solitude, nostalgia for space are only figures of speech, and explain nothing. But for the first time, on board a ship seething with people crowded upon one another, I seemed to understand the desert. The Sahara may be more lively than a capital, and the most crowded city is deserted if the essential poles of life lose their magnetism.”

— *Antoine de Saint-Exupéry in Letters to a hostage*



Rock art, Western Sahara, January 2018. Photo: Joana L. Rocha

Foreword

In compliance with the no. 2 of article 4 of the General Regulation of Third Cycles of the University of Porto and with the article 31 of the Decree-Law no. 74/2006, of 24 March, with the alteration introduced by the Decree-Law no. 230/2009, of 14 September, the results of already published works were totally used and included in some of the chapters of this dissertation. As these works were performed in collaboration with other authors, the candidate clarifies that, in all these works, participated in obtaining, interpreting, analyzing and discussing the results, as well as in the writing of the published forms.

The Faculty of Sciences of the University of Porto was the home institution of the candidate, and the work was supervised by Dr. Raquel Godinho (CIBIO-InBIO - Research Center in Biodiversity and Genetic Resources and Faculty of Sciences of the University of Porto) and co-supervised by Professor Rasmus Nielsen (Center for Theoretical Genetics, Department of Integrative Biology, University of California, Berkeley). The work was developed at CIBIO-InBIO and at UC-Berkeley in the context of the Doctoral Programme in Biodiversity, Genetics and Evolution (BIODIV). The candidate was financially supported by Fundação para a Ciência e Tecnologia (FCT, Portugal) through the attribution of the PhD grant SFRH/BD/116397/2016. The work was partially funded through FCT project grant “HOTFOXES” (PTDC/BIA-EVL/31902/2017) and by a National Institutes of Health (NIH) grant (R01GM138634).



Acknowledgments

Words won't be enough to express how grateful I am for the people and institutions who have helped, taught, and supported me for the past four and half years. To all of them, thank you!

First, I would like to thank my two supervisors Dr. Raquel Godinho and Professor Dr. Rasmus Nielsen for all the time dedicated to this project, for sharing their knowledge, and for helping me grow as a scientist during all these years. Both your ideas and inputs were instrumental to this work, and your supportive words crucial to my academic progress and personal growth.

Raquel, thank you for changing my life and giving me the opportunity of following my dreams. If it wasn't for your guidance and your outstanding efforts to make this project possible, I would not be where I am now. Thank you for holding me high every time I felt discouraged by logistical challenges or unexpected results. Above all, I will always treasure our friendship.

Rasmus, thank you for accepting to be my advisor and receiving me so well in your lab. Thank you for being such a great educator, for making me think critically, inspiring me to write shorter, and for being so patient and incredibly generous with your time. Above all, thank you for supporting me in so many ways during these challenging times and for being so kind. I learned so much from you.

To the ECOGEN group my heartful thank you for being good colleagues and for accommodating our meetings 8 hours away. Thank you for all your valuable feedback and suggestions during our lab meetings, particularly Pedro Silva, Carolina Pacheco, and Diana Lobo.

To past and current members of the Nielsen lab I am eternally thankful for the warm welcoming. You are all as incredibly talented as you are kindhearted. I am particularly grateful to Tyler, Diana and Debora for all the help and feedback. I could not ask for better colleagues, nor better friends.

My gratitude also goes to all collaborators and team members that contributed to this work with different levels of expertise. To Professor Abdeljebbar Qninba, I thank the possibility of sampling in Morocco. To Jose C. Brito and Pedro Silva I thank all the help with analyses, discussions, feedback and insights at the different states of this work. To Monia Nakamura and Nuno Santos I am grateful for our expeditions and for teaching me so much during field work.

I would also like to express my gratitude to all the Saharawi who helped us in the field, including Mr. Abd-Lakrim in the province of Assa-Guelmin, and Mr. Mohammed Samlali and Mr. Habib in Dakhla.

In particular, I would like to thank Mr. Habib for not letting us drive into dangerous territory, for allowing us to store samples at his freezer, and for leading us to the foxes. Thank you for keeping the door of the desert open for us.

To Professor Nuno Ferrand I thank my interest for evolutionary biology, and for watching me grow.

To Prof. Paulo Celio Alves my heartful appreciation for his vital role in the Doctoral Programme, and for all his generosity and willingness to listen and helping in every way.

To Prof. Peter Sudmant I thank all the incredible support, discussions and harness of enthusiasm. Peter, the way you think about science and your passionate personality inspire me every day!

I also thank the institutions or entities that contributed to the work developed in this thesis. Those include Centro de Testagem Molecular at CIBIO-InBIO, Temara Zoo, Dream Village, Parque Biologico de Gaia and Projecto Lobo, and the Evolutionary Genomics Lab (EGL) at UC-Berkeley. I am particularly grateful to Lydia Smith, Liliana Farelo, Sofia Mourao, Diana Costa, Nuno Santos and Sandra Afonso for all the help and assistance during experimental laboratorial analysis.

On the personal side, I am grateful for the network of friends I've build along the way, a group of inspiring kind people scattered around the world whom I deeply miss. To the friends from Porto, I thank all the amazing moments. To the friends from BioDiv, I thank our comradery. To Carolina and Mafalda I thank the love and support and for being amazing role models. To Ben, Kelsey, Mallory, Gonzalo, Diana, Debora, Mauri, Ian, Moi, Daniel, Rohit, Manju, Remi, Tyler and Begonia I thank you for making Berkeley feel like home and giving me the happiest times of my life. Diana, Ben, thank you for being the best friends in the world and for thinking I am better than I am.

To my family, I thank you for your unconditional love. I thank my mother for her constant support and stoicism. I thank my father for my passionate views and faith in people. I thank both my parents for giving me an amazing sister, who I thank for the wonderful fox illustrations that are displayed throughout this thesis. I thank my grandmothers for teaching me how to celebrate life. I miss you.

Finally, Gonzalo Benegas, I thank you for your love, and for the green waves of life.

Abstract

Deserts are the driest environments on the planet. For the last fifty years, deserts captivated many evolutionary ecologists, who turned to these natural laboratories to study how biological design is challenged by different aspects of the environment, including extreme thermal amplitudes, intense UV radiation and scarce resources. However, little is known about the genetic mechanisms underlying desert adaptation. The main aim of this thesis was to advance our understanding of the evolutionary processes driving adaptative evolution in deserts by reviewing the state-of-the-art on the genetic basis of mammalian desert adaptation, and by studying the demographic and adaptive history of North-African foxes (*Vulpes sp.)* living in the outer edges and core desert regions of the Sahara.

The first manuscript of this thesis offers a synthesis on the recent but growing body of work on the genomics of mammalian desert adaptation. By linking emerging genomic research with classical studies on desert eco-physiology, we were able to identify shared patterns of adaptation at the genetic, pathway and phenotypic level associated with fat and water metabolism and insulin response in biological systems with different evolutionary ages, occupying distinct ecological niches and at distinct stages of adaptive evolution in different deserts of the world. These patterns were consistent with the complexity and variety of environmental stressors associated with desert life. Surprisingly, we could not draw major patterns on the nature of desert adaptive genetic variation, nor on the sources by which adaptive genetic variants may arise. Moreover, studies of desert adaptation are challenged by a lack of accurate genotype–phenotype–environment maps, owing to logistical issues and expense of research in desert regions. Motivated by this, we encourage development of systems that facilitate functional analyses but also acknowledge the need for more studies on a wider variety of desert mammals. Lastly, we explore the limits and future potential of using genomics to address desert adaptation and provide explicit recommendations for future research.

In the second manuscript, we investigated the role of climatic variables as predictors of phenotypic variation. In particular, we focused on urine-concentrating ability, one of the most celebrated textbook examples of desert adaptive phenotypes in mammals, reviewed from the literature in 121 mammalian species. By integrating urine osmolality data with bioclimatic variables and phylogenetic comparative methods not only were we able to explicitly test the general hypothesis that desert-dwelling mammals have repeatedly evolved greater ability to concentrate urine, but also demonstrated that the mean annual aridity index of a species' geographic range largely predicts its urine-concentrating ability. This work contributes to the establishment of the general hypothesis that aridity is one of the main selective pressures driving phenotypic evolution in different desert mammalian lineages. Future studies investigating association of aridity and other bioclimatic variables

with desert adaptive phenotypes and genotypes are needed, as they hold the potential to elucidate future adaptative responses to increasingly arid conditions.

In the third manuscript, we studied the genetic basis and evolutionary history of desert adaptation in North-African foxes. Taking advantage of a large sample collection, we were able to able to generate whole genomes for 82 foxes belonging to 4 distinct species, including habitat generalist red fox (*V. vulpes*), sister species and desert adapted Rueppell's fox (*V. rueppellii*), distantly related desert fennec fox (*V. zerda*) and the basal Sahelian endemic pale fox (*V. pallida*). By combining state-of-the-art genomics with eco-physiological and expression data, we were able to study the evolutionary mechanisms at play during desert adaptation throughout different evolutionary scales and provide insights into the genetic basis underlying the ability of different fox species to thrive in desert environments. We show that past gene flow between North-African red foxes and desert-specialist Rueppell's fox likely facilitated red fox recent adaptation to an expanding Sahara, after they split from Eurasian populations ca. 78 Kya. Further comparative analysis identified a highly unusual and divergent 25 Mb genomic region in Rueppell's fox that can be explained as the result of an ancient introgression from another fox related to the fennec (*V. zerda*). Scans for signatures of adaptation revealed signals of selection at genes associated with water retention through reduced panting and sweating, temperature perception and adaptive thermogenesis in North-African red fox, and genes affecting renal water homeostasis in Rueppell's and fennec fox. These discoveries suggest that both standing adaptive genetic variation and introgression with older resident desert species have been key drivers of desert adaptation in North-African foxes and are likely to be critical components of future evolutionary responses to climate change.

Overall, this thesis contributes to a better understanding of the evolutionary mechanisms driving adaptation in desert environments and provides a glimpse into the origins and evolution of some of the most charismatic and elusive desert-dwelling species. With the first two manuscripts, we bridge evolutionary genomics and classical eco-physiology. Our work in North-African foxes provides an empirical example of desert adaptation and underscores the important roles of introgressive-hybridization events and ancestral standing genetic variation as sources for adaptive genetic variation at different spatiotemporal scales. This work also provides important insights on the genes and gene changes potentially underlying adaptation in an entirely new system. As desertification of the Sahara and expansion of arid regions worldwide are expected to dramatically increase, preserving adaptive variation in these umbrella species may help conservation efforts focused on rescuing biodiversity currently challenged by desertification.

Keywords

Deserts, aridity, adaptation, genomics, physiology, convergent evolution, introgressive-hybridization, ancestral introgression, standing genetic variation, North-African foxes, Sahara.

Resumo

Os desertos são os ambientes mais áridos do planeta, e durante os últimos cinquenta anos cativaram muitos ecologistas e biólogos evolutivos que os têm utilizado como laboratórios naturais para estudar de que forma os sistemas biológicos podem ser moldados por diferentes fatores abióticos, incluindo amplitudes térmicas extremas, radiação ultravioleta intensa e recursos escassos. No entanto, pouco se sabe sobre os mecanismos genéticos subjacentes à adaptação ao deserto. Esta tese teve como principal objetivo contribuir para a compreensão geral dos processos evolutivos responsáveis pela adaptação de mamíferos aos desertos, através da revisão do estado da arte sobre a base genética da adaptação deste grupo de vertebrados de deserto, e através do estudo da história demográfica e adaptativa de raposas do Norte de Africa (*Vulpes sp.*) que vivem nas fronteiras e nas regiões centrais do deserto do Saara.

O primeiro manuscrito desta tese contribui com uma síntese dos estudos mais recentes sobre a genómica da adaptação ao deserto em mamíferos. Através do enquadramento de estudos genómicos e estudos clássicos de eco-fisiologia, foi possível identificar padrões de adaptação ao nível genético, funcional e fenotípico relacionados com o metabolismo de gordura e água, e resposta à insulina. Estes padrões são compartilhados por sistemas biológicos com diferentes idades evolutivas, com nichos ecológicos distintos e em estágios distintos de adaptação em diferentes desertos do mundo. Os padrões de adaptação encontrados são ainda consistentes com a complexidade e variedade de pressões ambientais associadas à vida no deserto. Contrariamente ao esperado, não se encontraram padrões gerais de variação genética adaptativa ao deserto, ou relacionados com as diferentes vias pelas quais as variantes genéticas adaptativas podem surgir. Além disso, os estudos de adaptação ao deserto carecem de mapas de genótipo-fenótipo-ambiente, sobretudo devido a dificuldades logísticas e despesas associadas a investigação em regiões desérticas. Neste contexto, este trabalho de revisão encoraja ao desenvolvimento de sistemas que facilitem análises funcionais, mas também reconhece a necessidade de estudos mais abrangentes que incluam diferentes organismos de deserto. Por último, explora-se os limites e o potencial de usar análises genómicas para estudar adaptação ao deserto, e sugerem-se possíveis vias para estudos futuros.

No segundo manuscrito, investiga-se o papel preditivo de variáveis climáticas na variação fenotípica. Este trabalho foca-se na capacidade de concentração de urina, um dos exemplos mais emblemáticos de fenótipos adaptativos em mamíferos de deserto, que foi aqui revista para 121 espécies de mamíferos. Ao integrar dados de osmolalidade de urina com variáveis bioclimáticas e métodos de filogenética comparada, foi possível testar a hipótese geral de que os mamíferos de deserto evoluíram repetidamente uma maior capacidade de concentrar urina, mas também demonstramos pela primeira vez que o índice de aridez média anual da distribuição geográfica de

uma espécie pode ser usado para prever a sua capacidade de concentrar urina. Este trabalho contribui para o estabelecimento da hipótese geral de que a aridez é uma das principais pressões seletivas que impulsionam a evolução fenotípica em diferentes linhagens de mamíferos de deserto. Futuras associações entre o índice de aridez/variáveis bioclimáticas, e fenótipos adaptativos e/ou genótipos permitirão elucidar respostas adaptativas em ambientes cada vez mais áridos.

No terceiro manuscrito, estudamos as bases evolutivas da variação genética adaptativa das raposas do Norte de África. Aproveitando uma grande coleção de amostras obtidas no âmbito deste trabalho ou de trabalhos anteriores, geraram-se genomas completos para 82 raposas pertencentes a 4 espécies distintas, incluindo a generalista raposa vermelha (*V. vulpes*), a sua espécie irmã de deserto raposa de Rueppell (*V. rueppellii*), a espécie divergente de deserto feneco (*V. zerda*) e a raposa pálida do Sahel (*V. pallida*). Ao combinar genómica com dados fisiológicos e transcriptómica, foi possível estudar os mecanismos evolutivos responsáveis pela adaptação ao deserto em diferentes escalas evolutivas e entender melhor a base genética subjacente à capacidade de prosperar em ambientes desérticos em espécies diferentes. Nomeadamente, este estudo revela que a existência de fluxo génico entre a raposa vermelha do Norte de África e a raposa de Rueppell terá facilitado a adaptação mais recente da raposa vermelha a um Saara em expansão, depois de esta se ter separado das populações da Eurásia ca. 78 Kya. Uma análise comparativa adicional identificou uma região genómica de 25 Mb na raposa de Rueppell altamente divergente que pode ser explicada como o resultado de uma introgressão antiga por via de uma espécie de raposa relacionada com o feneco. Testes focados na deteção de assinaturas de adaptação revelaram sinais de seleção em genes relacionados com uma maior retenção de água através da redução da respiração ofegante e suor, perceção de temperatura e termogénese adaptativa na raposa vermelha do Norte de África, e genes que afetam a retenção de água nos rins na raposa de Rueppell e feneco. Estes resultados sugerem que variação genética pré-existente e variação genética herdada por introgressão a partir de espécies que residem há mais tempo no deserto, terão permitido a adaptação das raposas do Norte da África e desempenharão um papel critico em futuras respostas evolutivas às mudanças climáticas.

De uma maneira geral, os estudos realizados nesta tese contribuem para um melhor entendimento dos mecanismos evolutivos que conduzem a adaptação em ambientes desérticos, e oferecem uma perspetiva sobre a origem e história evolutiva de algumas das espécies mais carismáticas e evasivas do deserto. Nos dois primeiros manuscritos, fazemos uma ponte entre a genómica evolutiva e a fisiologia comparada clássica. O nosso trabalho sobre raposas norte-africanas contribui com um exemplo empírico de adaptação ao deserto, e exalta a importância da hibridização introgressiva e da variação genética pré-existente como fontes de variação genética adaptativa em diferentes escalas espácio-temporais. Este trabalho também fornece informações relevantes sobre os genes e as mudanças genéticas potencialmente subjacentes à adaptação a um sistema inteiramente novo. Uma vez que a desertificação do Saara e a expansão das regiões áridas

em todo o mundo deverão aumentar dramaticamente, preservar a variação adaptativa em espécies guarda-chuva pode ajudar a conservar biodiversidade atualmente ameaçada pela desertificação.

Palavras chave

Desertos, aridez, adaptação, genómica, fisiologia, evolução convergente, hibridação introgressiva, introgressão ancestral, variação genética pré-existente, raposas norte-africanas, Saara.

Table of contents

Foreword……..……6
Acknowledgments……..7
Abstract…….....9
Keywords……..11
Resumo……12
Palavras Chave……14
Table of contents……15
List of figures and Tables……17
Abbreviations ……...24

Chapter I. General Introduction……..……...26

1.1. Repeated evolutionary adaptation in terrestrial extremes ……...26
1.2. The nature of adaptive variation ……..30
1.3. Foxes as a model to study the genetic basis of adaptation ……32
1.3.1. On the origins of foxes ……..37
1.3.2. The intriguing evolutionary history of Red and Rueppell's foxes……39
1.4. Objectives ……..…….....43
1.5. Work outline ……..…….....43
1.6. References……..…….....45

Chapter II. On genes and phenotypes underlying desert adaptation in mammals…….57

PAPER 1. Life in Deserts: The genetic basis of mammalian desert adaptation ……...…..58

PAPER 2. Convergent evolution of increased urine concentrating ability in desert mammals……..95

Chapter III. Genomic insights into the adaptive history of North-African foxes……128

PAPER 3. North-African fox genomes reveal signatures of ancient introgression and adaptation to life in deserts……..…129

Chapter IV. General discussion……231

4.1. On the origins and evolution of North-African *Vulpes*……...232
4.2. Introgression in canids and implications to desert life……..……...233
4.3. Insights into complex evolutionary histories……...……...237
4.3. A gradual understanding of the mechanisms underlying desert adaptation……...238
4.5. Challenges in dissecting the genetic basis of traits in non-model organisms……...239
4.6. Concluding Remarks and Perspectives……..……...240

List of Figures and Tables

Chapter I

Figure 1.1. Phylogeny of the family Canidae based on 14kb DNA sequence from 12 exons and four introns (adapted from from Lindblad-Toh et al. 2005). The numbers above and below each branching point indicate two different measures of branching point support (bootstrapping and Bayesian posterior probability values). The three main clades of canids are indicated by red (red fox-like canids), green (South American fox-like canids) and blue (wolf-like canids) colors. Example species (underlined) are as follows: Arctic fox (*Vulpes lagopus*), red fox (*Vulpes vulpes*), raccoon dog (*Nyctereutes procyonoides*) and bat-eared fox (*Otocyon megalotis*), crab-eating fox (*Cerdocyon thous*), maned wolf (*Chrysocyon brachyurus*), bush dog (*Speothos venaticus*), domestic dog (*Canis familaris*), gray wolf (*Canis lupus*), golden jackal (*Canis aureus*), coyote (*Canis latrans*), African jackal (*Canis mesomelas*) and gray fox (*Urocyon cinereoargenteus*).

Figure 1.2. Distribution range for the 12 extant species of the genus *Vulpes*: (A) Bengal fox, *V. bengalensis (By Mikael Bauer on iNaturalist.org); (B)* Rueppell's fox, *V. rueppellii (By Mickaël Villemagne, iNaturalist.org);* (C) Red fox, *V. vulpes (By Yellowstone National Park on Flickr. Public domain);* (D) Corsac fox, *V. corsac (By Wilm Stan, Wikimedia commons. Public Domain);* (E) Tibetan fox, *V. ferrilata (By pfaucher, iNaturalist.org);* (F) Arctic fox, *V. lagopus (By Emma, Flickr);* (G) Kit fox, *V. macrotis (By Seth Topham. Public domain);* (H) Swift fox, *V. velox (By leptim, iNaturalist.org);* (I) Fennec fox, *V. zerda (By pfaucher, iNaturalist.org);* (J) Blanford's fox *V. cana (Eyal Bartov, Wikimedia);* (K) *V. chama (By Yathin S Krishnappa, Wikimedia);* (L) Pale fox, *V. pallida (By pfaucher, iNaturalist.*org). Distribution ranges were downloaded from IUCN.

Figure 1.3. Illustration of *Vulpes* cytochrome b gene phylogeny and respective distribution range maps. Phylogenetic tree was inferred by using the Maximum Likelihood method and General Time Reversible model in MEGA. Sequences were downloaded from GenBank for each species. Distribution ranges were downloaded from IUCN (2020) and plotted in R and are followed by scientific illustration of each species (illustrations by *Margarida Tangerina*). Analyses performed under the scope of this thesis.

Figure 1.4. Representation of mean Annual aridity index (A) and Annual mean temperature (B) in a cytochrome b phylogenetic tree with 11 *Vulpes* species represented in Figure 1.3. Aridity index was obtained from the CGIAR-CSI Global Aridity database (Trabucco and Zomer 2009), and annual mean temperature (BIO1) from the WorldClim database (Fick and Hijmans 2017). Distribution ranges were downloaded from IUCN (2020) and Geographical Information System was used to calculate mean values for each environmental variable within the geographic range polygon of each species. The mean aridity index for the distribution range of different *Vulpes sp* can be classified into discrete categories as follows: hyper-arid (Rueppell's fox *V.rueppellii*; Fennec, *V. zerda*; dark red), arid (Kit fox, *V. macrotis*; Blanford's fox, *V. cana*; Pale fox *V. pallida*; red), semi-arid (Corsac fox, *V. corsac*; Swift fox, *V. velox*; orange-red), dry sub-humid (Tibetan fox, *V. ferrilata*; Bengal fox, *V. bengalensis*; yellow), humid (Red fox, *V. vulpes*; Arctic fox, *V. lagopus*; blue). Ancestral states for mean AI and BIO1, estimated using maximum-likelihood methods (implemented with the R package 'phytools'), were projected onto the internal edges and nodes of the tree using a color gradient. These analyses were performed under the scope of this thesis and are presented to serve as an illustrative representation of species ecological contexts. Since trees species distribution ranges and environmental variables both rely on present-day information, assumptions on ancestral environments are inherently imprecise and should thus be interpreted with caution.

Figure 1.5. Regions with overlapping distribution between sister species Rueppell's (*V. rueppellii*; *photo by Jesper Särnesjö, Wikimedia*), Red fox (*V. vulpes*; *photo by Paul Cools on iNaturalist.org*) and basal group Fennec fox (*V. zerda; photo by pfaucher, iNaturalist.org*). Rueppell's fox is found in the deserts of North-Africa and the Middle East (yellow). In North Africa, Rueppell's shares most of the range of Fennec fox (orange) and overlaps with sister species Red fox (photo by *Aissa Djamel Filali, iNaturalist.org*) at the Saharan borders and/or in regions with water availability. Distribution ranges were downloaded from IUCN and plotted in R.

Figure 1.6. Illustration of 10 out of 44 morphological subspecies of Red fox (A) and 5 subspecies of Rueppell's fox (B), though many are doubtful and do not translate genetic variability.

Chapter II: PAPER 1

Figure 1. Genomic approaches used to identify genes underlying desert adaptation: A-population-based, B-comparative-based methods, and C-differential expression analyses in

natural (I) and laboratory (II) conditions; A-C: different silhouettes represent distinct species, silhouettes in orange represent desert populations/species, and grey silhouettes represent non-desert populations/species under comparison and outgroups; double-headed arrows represent the comparison being made, one-headed arrows represent the group or lineage targeted by the method; X symbol represents a condition of experimental dehydration or water-deprivation, as opposed to animals being given water. D-schematic representation synthesizing and relating some of the recently discovered genes and functional gene classes to previously described phenotypes. Functional connections result from gene ontology enrichment analyses on sets of candidate genes and/or gene-function annotation (Gene Cards, OMIM, UniProt). Some of the highlighted genes identified with population-based and comparative-based methods on whole-genomes and DNA sequences of candidate genes found differentially expressed are represented with superscripts as follows: En – Endurers (1-Bactrian Camel, *Camelus bactrianus*, 2-Dromedary camel, *C. dromedarius*, 3-Taklamakan and 4- Barki sheep, *Ovis aries* 5-Barki goat, *Capra hircus*, 6-Arid-dwelling Asian *sheep*); 6-Ev – Evaders (1- Cactus mouse, *Peromyscus eremicus*, 2- Canyon mouse *P. citrinus*, 3- House mouse, *Mus musculus*, 4-Patagonian olive mouse, *Abrothrix olivacea*, 5-Banner-tailed kangaroo rat, *Dipodomys spectabilis*, 6-Bailey's pocket mouse, *Chaetodipus baileyi*); H – Human*s* (Aboriginal Australians).

Figure 2. Shared evidence of selection in genes (A) and GO terms and pathways (B) obtained from comparative and population genomics in evader (Ev), endurer (En) and human (H) case studies. Examples reviewed here include: Cactus mouse (*Peromyscus eremicus*) [27], Canyon mouse (*P. citrinus*) [30], Tarim red deer (*Cervus elaphus yarkandensis*) [25], arid-dwelling Asian and Taklamakan desert sheep (*Ovis aries*) [22], Bactrian camel (*Camelus bactrianus*) [51,53], Dromedary camel (*C. dromedarius*) [51], North-African Barki sheep (*O. aries*) and goat (*Capra hircus*) [26], Nubian Ibex (*Capra nubiana*) [58], southern African Khoe-San [49] and Aboriginal Australians [21] (*Homo sapiens*). Significance of gene and GO overlap was calculated using Fisher's exact test on a contingency table. Significant overlap is represented by colors standing out against the lightest background (marginal p-values > 0.05).

Figure I. The different steps for future research attempting to integrate available methodologies to address the genomic basis of desert adaptation. In part A we suggest simultaneous collection of genomic, transcriptomic (across different tissues and also blood), phenotypic (see Box1) and climatic data (e.g. WorldClim database). Data collection under common garden and field reciprocal transplant experiments are suggested for model organisms (e.g. *Mus* and *Peromyscus* sp). For common garden experiments we suggest different combinations of environmental stressors. Different silhouettes represent distinct

species, silhouettes in orange represent desert populations/species, and grey silhouettes represent non-desert populations/species; double-headed arrows represent the comparison being made; X symbol represents a condition of experimental dehydration or water-deprivation, as opposed to animals being given water. In part B we illustrate the diverse ways in which different datasets can be combined for methods in population and comparative genomics, depending on data availability and the evolutionary context of the study-system of choice. The strength of selection studies is that they are powerful for identifying candidates for further testing. In part C we suggest experimental validation of candidate genes under selection (CRISPR-Cas 9 and cell culture).

Table I. Classic physiological systems allowing mammalian survival in deserts. The phenotypes listed may themselves be adaptive or may be part of a correlated plastic response.

Chapter II: PAPER 2

Figure I. Urine concentrating ability in 121 mammal species, measured as maximum urine osmolality (mosmol/Kg), along species distribution range mean annual aridity index.

Figure 1. Quantile-quantile plots for the normal distribution of variables (A, C, E) and log10 of variables (B, D, F) used in this study of urine-concentrating ability of desert mammals.

Figure 2. Ancestral state reconstruction for mean annual aridity index (AI, left) and maximum urine osmolalities (mOsm/kg; right) in a phylogenetic tree with 121 mammalian species.

Figure 3. Maximum urine osmolality (mOsm/kg) in mammals and mean aridity index (AI). (A) Schematic illustration of the endurer-evader-evaporator concept proposed by Willmer et al. (2005) and classification of mammals into arid, semi-arid, dry sub-humid and humid habitat occupiers according to Trabucco and Zomer (2009). (B) Boxplot for maximum mOsm in 121 endurers, evaders, and evaporators with ranges with different AIs. (C) Scatterplot for maximum mOsm in 121 mammal species.

Figure 4. Spearman correlations for: (A) log10 maximum urine osmolality (measured in mOsm/kg) and log10 body mass (kg); (B) log10 maximum urine osmolality (measured in mOsm/kg) and mean annual aridity index; and (C) log10 body mass (kg) and mean annual aridity index. *P*-values are not presented, as the plots do not take phylogenetic affinity into account.

Table 1. Phylogenetic signal (λ) for variables with significant *P*-values. Variables for which λ is close to 1, and with *P*-values < 0.05, have significant phylogenetic signals.

Table 2. PGLS models for predicting mammalian log10(maximum urine osmolality). Model notation refers to the variables used as predictors: mean AI = annual aridity index, BM = log10 body mass (kg); C = condition (fed-F, given salt load-S, protein load-P, unknown condition-X, and combinations of these, all relative to dehydrated-D); M = method (addition of solutes-A, freezing point-F, vapour pressure-V, and combinations of these, all relative to unknown method-?); Di = diet (frugivorous-Fr, granivorous-G, herbivorous-H, insectivorous-I, omnivorous-O, and CI-both carnivorous and insectivorous, all relative to strictly carnivorous-C). Significant *P*-values are highlighted in bold; AIC = Akaike information criterion, s.e. = standard error.

Table 3. PGLS models for predicting mammalian log10 maximum urine osmolality (mOsm/kg) in the dataset including only dehydrated individuals. Model notation refers to the combination of variables used as predictors while taking phylogeny into account: mean AI = annual aridity index; BM = log10 body mass (kg); C = condition (fed-F, given salt load-S, protein load-P, unknown condition-X, and combinations of these, all relative to dehydrated-D); M = method (addition of solutes-A, freezing point-F, vapour pressure-V, and combinations of these, all relative to unknown method-?); Di = diet (frugivorous-Fr, granivorous-G, herbivorous-H, insectivorous-I, omnivorous-O, and CI-both carnivorous and insectivorous, all relative to strictly carnivorous-C). Significant *P*-values are highlighted in bold; AIC = Akaike information criterion, s.e. = standard error.

Table 4. PGLS models for predicting log10 mass-adjusted basal metabolic rate (ml of O_2 per g per day) ranked by AIC (Akaike information criterion). Model notation refers to the combination of variables used as predictors while taking phylogeny into account. Variables are BM = log10 body mass (g); mean AI = annual aridity index (continuous), cAI = categorical AI (arid, semi-arid, dry sub-humid, humid), Di = diet (carnivorous-C, frugivorous-Fr, granivorous-G, herbivorous-H, insectivorous-I, omnivorous-O, carnivorous and insectivorous-CI); and mean temperature values for BIO1 = annual mean temperature, BIO5= maximum temperature of warmest month, BIO6 = minimum temperature of coldest month, BIO7 = temperature annual range. Significant variables are highlighted in bold (*P*-value < 0.05). When λ is zero, the covariance between species is zero and the linear model approaches that of a non-phylogenetic regression, while when λ is 1, the evolution of the residual error follows a Brownian model.

Chapter III: PAPER 3

Figure 1. Sampling localities and overall species relationships. (A) Geographic location of the samples (points) and distribution ranges (lines) for Rueppell's and Red fox used for population genetic analyses (see Table S1 for sampling details and Figure S2) [Photo credit: Ricardo Rocha, Monia Nakamura]. North-African red fox individuals were sampled across an aridity gradient ranging from dry-sub humid to hyper-arid regions. We used a color range for Aridity Index (AI) from 0 to 0.2, to define arid (0.05 < AI <0.2) and hyper-arid regions (AI<0.05) and demark them from non-desert areas (AI > 0.2). Dark grey lines delimit the distribution range of Red foxes, and beige line border the distribution range of Rueppell's fox (B) RaxML whole-genome tree constructed with genetic distances among 83 foxes based on genotype likelihoods using the Dog as outgroup. Genetic distances were calculated with ngsDist with genome-wide polymorphic sites, filtering for a minor allele frequency of 0.05. This was performed through concatenation of genomic data, which typically leads to long external and short internal branches within species. (C) Admixture proportions and trees inferred with Ohana structure analysis for K=3 (top) and K=4 (bottom). Each bar shows the inferred ancestry fraction for an individual. Labels above and below bars identify the species and geographical location.

Figure 2. Introgression in North African foxes. Manhattan plot of fraction of admixture ($f\hat{}_d$) on non-overlapping windows of 200kb across all chromosomes (A) Eurasian red foxes were used as H1, North African red foxes as H2 and Rueppell's fox as H3, with the arctic fox as outgroup; (B) All red foxes were used as H1, Rueppell's foxes as H2, and fennec fox as H3, with the pale fox as outgroup; this is followed by a zoom-in view on 50kb windows (some examples highlighted). Windows from the same chromosome have the same color. Dark line threshold separates regions above the 99th percentile. We note that other genes not displayed in the figure can overlap the outlier regions. The arctic fox was used as reference genome and scaffolds were ordered with respect to the dog. (**C**) Estimated tree topologies for *Vulpes*: whole-genome tree (left), 200kb region harboring gene MGAM (center), and 4 scaffolds in the arctic fox genome (scaffold 47,155,126, 200) above the 99th percentile of the empirical distribution (right). Regions marked with a rectangle were selected for further investigations of gene-species tree discordance.

Figure 3. Selection specific to North-African Red foxes. Genome-wide selection scans on windows of 50 kb with a 10kb 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. Dashed lines show the 99.95th percentile of the PBS empirical distribution and negative logarithm of alpha Pi. Names of genes within the highest peaks are shown. The arctic fox 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 Tables S12-S13.

Figure 4. Genome-wide selection scans for Rueppell's and fennec fox using Ohana based on the structure and tree results from K = 4. (A) Manhattan plot of the log-likelihood ratio values for the top 500 SNPs on windows of 50 kb with a 10kb slide across chromosomes (x-axis) for Rueppell's fox and fennec fox. To account for selection specific to Rueppell's and not in their common ancestral with Red foxes we subtracted the log-likelihood ratio of Rueppell's and Red fox for the same windows (delta log-likelihood ratio). Dashed 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 Tables S19-20). (B) Representation of the selection study performed in (A): structure plot corresponding to four *Vulpes*, followed by the global tree and representation of the likelihood ratio test implemented in *Ohana*. 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). (C) WFS1 mRNA expression levels in blood of Rueppell's (*n*=3) and Red foxes (*n*=2 North Africa, *n*=5 Iberia).

Chapter IV

Figure 4.1. Phylogenetic relationships of *Vulpes* spp. and geographically overlapping species with Rueppell's fox other than the red fox (related to Figure 2B in Main Text of Chapter III).

Abbreviations

AIC – Akaike Information Criterion
AICc – corrected AIC
ANOVA – analysis of variance
bp – base pairs
D – *D* statistic
DEC – Dispersal-Extinction-Cladogenesis
DIVA – Dispersal-Vicariant-Analysis
DNA – Deoxyribonucleic acid
d_{XY} – genetic distance
EM – Expectation Maximization
f^d – fraction of admixture
FDR – false discovery rate
f^hom – admixture proportion or fraction of introgression
F_{ST} – fixation index
GO – Gene Ontology
GPS – global positioning system
GWAS – genome-wide association study
HPD – high posterior density
kb – kilobase pairs
kya – thousands of years ago
LD – linkage disequilibrium
LRT – likelihood ratio test
LLR – log-likelihood ratio
MAF – Minor allele frequency
Mb – mega base pairs
MIS – Marine Isotope state
ML – maximum likelihood
mtDNA – mitochondrial DNA
Mya – millions of years ago
PBS – Population Branch Statistics
PCA – principal component analysis
PCR – polymerase chain reaction
PGLS – phylogenetic generalized least squares
RNA-seq - Ribonucleic acid (RNA)-sequencing
SD – standard deviation
SE – standard error
SFS – site frequency spectrum
SNP – single nucleotide polymorphism
t - population divergence time
TMRCA – time to the most recent common ancestor
μ – mutation rate
π - nucleotide diversity

Chapter I.

General Introduction

How species adapt in order to fit the ecological circumstances of new environments is a fundamental question in evolutionary biology. Within the last ten years, the advent of cost-effective Next Generation sequencing technologies and increasing availability of large genomic datasets for many populations and species has allowed researchers to directly address this question (Stapley et al. 2010). Understanding the mechanisms by which species in the past successfully adapted to the physical and climatic rigors of their environment is important for evaluating the possibility of evolutionary rescue of biodiversity currently challenged by habitat loss and climate change (Hedrick and Fredrickson 2010; Corlett 2017; Breed et al. 2019; Exposito-Alonso et al. 2019; Hoffmann et al. 2020). However, providing evidence that a given trait or phenotype is the outcome of natural selection can be challenging, as differences in morphology, physiology and behavior may be due to plastic responses, or may have evolved neutrally due to random genetic drift. Quantifying selection in nature requires measurement of traits affecting fitness - i.e. survivorship and/or lifetime reproductive success (Linnen and Hoekstra 2009). Detecting such traits in the wild and then being able to explicitly correlate them with individual genotypes and the environment is still a key challenge to modern evolutionary biologists working with non-model organisms. On the other hand, moving beyond the standard model systems and broadening the study of the mechanisms driving adaptation in natural populations or species has never been more pressing, as terrestrial biodiversity is decreasing, despite ambitious aims for genetic rescue and conservation management (Leclère et al. 2020).

1.1. Repeated evolutionary adaptation in terrestrial extremes

Extreme terrestrial habitats comprise many distinct selective regimes broadly characterized by low resource availability and extreme climates, ranging from hot deserts at low latitudes, to extreme cold settings at the poles, and extreme cold hypoxic environments at high altitudes (Wilmer et al. 2005; Bell 2012). As bellwethers for global change, extreme environments are ideal settings to address questions concerning the persistence and adaptive capacity of species (Vale and Brito 2015; Colella et al. 2020; Riddell et al. 2021). These are home to populations and/or species that have been pushed to their

physiological limits by temperature, water, oxygen, and diet for multiple generations (Bell 2012), and are therefore likely to exhibit features outside the scope of phenotypic plasticity, and random genetic drift, that were selected for the ability to survive and reproduce successfully in extreme conditions. Environmental extremes are also associated with the repeated evolution of similar adaptive traits in species that share the same harsh conditions and yet are distantly related (Wilmer et al. 2005; Bell 2012). In this sense, these environments are prime natural laboratories to investigate common molecular mechanisms underlying adaptation, and the populations or species that their inhabit make some of the best-known examples of Darwinian natural selection.

Modern human populations inhabiting extreme conditions have long been at the center of investigations of the genetic basis of adaptation, and a classic textbook example is the adaptation to the extreme hypoxic environment of the Tibetan plateau, which has an average elevation over 4500 meters (Yi et al. 2010; Ilardo and Nielsen 2018). Tibetan highlanders have lower hemoglobin concentration levels than Han Chinese and Europeans, and it has been suggested throughout decades of physiological research that the selective advantage of this trait under hypoxic conditions includes not having the cardiovascular risks associated to the exacerbated high hemoglobin production response of acclimatized lowlanders (Beall et al. 2010; Lorenzo et al. 2014; Simonson et al. 2015). Population genomics comparing native Tibetan highlanders with Han Chinese and Europeans then demonstrated that Tibetans are able to live in extreme high altitudes as the result of positive selection at a hypoxia pathway gene, EPAS1, involved in the regulation of hemoglobin concentration (Yi et al. 2010). Other human populations that inhabit high-altitudes at distinct regions of the globe, such as the Ethiopian highlanders (Alkorta-Aranburu et al. 2012; Huerta-Sánchez et al. 2013) and the Aymaras from Bolivia (Crawford et al. 2017), adopted different evolutionary mechanisms to cope with the same demanding high-altitude extreme conditions, as genome-wide scans for signatures of selection in these populations found that different genes were involved.

Humans also successfully adapted to life in the extreme cold high-arctic, in particular to cope with the high levels of cholesterol and omega-3 fatty acids present in their lipid-rich diets, which can lead to cardiovascular diseases. Genome-wide scans for signatures of genetic adaptation in Greenland Inuit populations identified several positively selected genes encoding the rate-limiting enzymes in the synthesis of long chain polyunsaturated fatty acids (PUFAs), among which omega-3 fatty acids are included (Fumagalli et al. 2015). It has been suggested that adaptive mutations in the Inuit could potentially decrease the rate of endogenous synthesis of PUFAs, resulting instead in short chain PUFAs, which could help prevent some of the health risks associated to high-fat diets (Ilardo and Nielsen 2018). Adding to adaptation to extreme diets, adaptive tolerance to cold temperatures is another adaptive feature associated with human life in extreme cold environments, as suggested by signatures of positive

selection in a gene, TRPM8, encoding a receptor involved in sensing and reacting to cold through physiological thermoregulation (Key et al. 2018). It has been suggested that a genomic variant in this gene (which exists at high frequency in high latitudes) could potentially diminish perception of cold (Ilardo and Nielsen 2018).

A fundamental aspect of studying extreme-environment adaptation in humans has been the possibility of taking advantage of numerous genomic and phenotypic resources to perform genotype-phenotype-fitness associations, thus explicitly addressing the genetic mutations (variants) underlying adaptive traits. Achieving similar levels of understanding in wild animals is constrained by lack of both genomic and phenotypic resources, and/or lack of segregating genetic variation that can be linked to specific traits. As such, research on the genetic basis of adaptive traits in extreme environments (or other local adaptation scenarios), has been biased towards advantageous systems, which only represent a limited range of adaptive genetic variation (Smith et al. 2020). Despite challenges to genotype-phenotype-fitness associations, the existing genomic studies in non-model organisms living in extreme environments, particularly mammals, have made significant contributions to our understanding of the mechanisms driving adaptation (Cho et al. 2013; Li et al. 2014; Liu et al. 2014; Welch et al. 2014; Lynch et al. 2015; Figueiró et al. 2017; Witt and Huerta-Sánchez 2019). In particular, these studies have shown that, even in evolutionary biology, life finds a way of repeating itself. For example, population genomics approaches comparing highlander dog breed Tibetan mastiff (*Canis lupus familiars*) with closely related lowland dogs found that the EPAS1 gene found in human Tibetans is also under selection in Tibetan dogs (Gou et al. 2014; Li et al. 2014; Wang et al. 2014a). EPAS1 was also found to be under selection in other domestic animals (Witt and Huerta-Sánchez 2019) and wolves (*Canis lupus chanco*) (Zhang et al. 2014a) sharing the same extreme hypoxic conditions. Though it is not clear if the same adaptive variants at EPAS1 were targeted by selection, the Tibetan plateau represents a classic case of convergent evolution at the gene level.

Sister species polar *(Ursus maritimus*) and brown bears (*Ursus actos*) provide another example of the relevance of studies of adaptative evolution in non-model organisms. Unlike brown bears, polar bears are uniquely adapted to life in the extreme cold High-Arctic. While brown bears are omnivores, polar bears have a high-fat hypercarnivorous diet consisting almost exclusively of marine mammals (Miller et al. 2012; Liu et al. 2014; Welch et al. 2014; Rinker et al. 2019). By taking advantage of the fact that the two sister species are recently diverged (less than one million of years), yet ecologically distinct, different genomic studies were able to identify genes putatively under selection involved in fur pigmentation, as well genes involved in fatty-acid metabolism and adipose tissue development (Miller et al. 2012; Liu et al. 2014; Welch et al. 2014; Rinker et al. 2019). Among the fat-metabolism genes is APOB, an apolipoprotein involved in low-density lipoprotein particle clearance, regulation of cholesterol storage

and metabolism. Beneficial mutations at fat-metabolism genes thus may potentially underlie polar bears' ability to cope with the high-cholesterol diets that provide them with the metabolic heat necessary to endure the cold High-Arctic (Liu et al. 2014). Notably, APOB was found to be under selection in cold-dwelling North American grey wolves (Schweizer et al. 2016). Notably, different genes involved in fatty-acid metabolism, cardiovascular functions, and fur color have also been the target of natural selection in Arctic foxes (*Vulpes lagopus*) (Kumar et al. 2015).

Other case studies for which there is also shared evidence of selection in the same genes, genes from the same family, and/or in genes with similar functions in cold environments include the extinct woolly mammoths (*Mammuthus primigenius*) and arctic-dwelling Greenland sled dogs. Comparative analysis using ancient genomes from woolly mammoths found unique amino-acid changes relative to other extant elephantid species in genes enriched for functions related to circadian biology, skin and hair development and physiology, lipid metabolism, adipose development and physiology, and temperature sensation (Lynch et al. 2015). The later encode temperature-sensitive transient receptor potential (thermoTRP) channels that sense noxious cold TRPM8 – also identified in humans (Key et al. 2018) – innocuous warmth (TRPV3 and TRPV4), noxious cold and/or heat (TRPA1) (Lynch et al. 2015). Similarly, a recent study aiming a elucidating the evolutionary origins and genetic basis of adaptation in Greenland sled dogs, found two transient receptor potential channels, TRPC4 and TRPV2, putatively under selection (Sinding et al. 2020). Furthermore, sled dogs also showed genomic signatures of selection in genes involved in coping with high intake of fatty acids (Sinding et al. 2020), including APOO, which encodes for an apolipoprotein family member that helps regulate fatty acid metabolism, just as APOB in polar bears (Liu et al. 2014; Sinding et al. 2020). Together, these results suggest that adaptive tolerance to extreme cold and extreme dietary regimes are the two main features characterizing mammalian adaptive evolution in the cold arctic regions.

Overall, the last decade of genomic research on adaptation to terrestrial extremes has significantly contributed to a better understanding of diverse ways in which different species/populations sharing similar selective pressures meet their functional requirements, providing evidence for distinct case studies in which there are multiple solutions to the same adaptive problem, and cases in which selection yields repeatable results. In particular, these studies have allowed researchers to dissect different levels of convergent evolution, which may concern the same genes, or similar pathways and functional categories, and/or the same phenotypes. As such, despite outstanding challenges in establishing genotype-phenotype-fitness correlations, future research at a wider taxonomic range adapted to life in terrestrial extremes has the potential to lead to many exciting discoveries on the origins of adaptive genetic variation, and how different species respond to shifting selective pressures at distinct evolutionary time scales.

1.2. Different sources of adaptive genetic variation

The first step towards the understanding of the mechanisms by which natural populations and/or species adapt to new environments is to elucidate the origins of genetic variants or molecular changes that translate into phenotypic and fitness differences targeted by selection. There are two possible sources of adaptive variation: selection acting on existing (standing) genetic variation and selection acting on *de novo* mutations. Standing adaptive genetic variation can either be inherited from an ancestral population or species, or, alternatively, be introduced into a species (recipient) from another (donor) via introgression (Hedrick 2013; Jamie and Meier 2020).

Adaptation from standing genetic variation is usually assumed to be more plausible or faster, as the higher initial frequency of adaptive genetic variants from standing variation may result in higher probability of fixation (Jamie and Meier 2020). Adaptive genetic variants by *de novo* mutation take longer to arise, in the sense that in order for a beneficial mutation to occur and then rise in frequency in the population takes much longer than an allele that is already present (Hedrick 2013; Jamie and Meier 2020). Even when *de novo* mutations arise, they first have low initial frequency and thus may not necessarily persist in the population due to random genetic drift, or because their potential contribution to increased fitness is not immediate. In circumstances in which the same genomic regions are targeted by selection, as the result of species sharing the same selective regimes, the probability of recurrent mutation (i.e. same *de novo* mutations in different species) being the target of selection and maintained (convergence) is higher, specifically when fitness traits associated with such polymorphisms have a simple genetic basis (i.e. non-complex traits associated with single or few mutations) (Hedrick 2013; Jamie and Meier 2020).

Empirical evidence suggests that, for evolutionary time scales spanning same-species populations, closely related species and groups of species from the same genus, selection acting on inherited ancestral standing genetic variation is more likely than *de novo* mutation (Hedrick 2013; Jamie and Meier 2020). For example, the ability of deer mice (*Peromyscus maniculatus*) to cope in hypoxic High-altitude environments in North-America has been described as the result of selection acting on standing genetic variation in globin genes (Bedford and Hoekstra 2015). Nelson and Cresko 2018 also demonstrated that ancestral standing genetic variation dating back 15 Ma enabled marine populations of stickleback (*Gasterosteus aculeatus*) to repeatedly adapt to freshwater environments (Nelson and Cresko 2018). Adaptive ancestral polymorphism thus seems to persist after species divergence and across long evolutionary time scales. Inversion polymorphisms constitute some of the most spectacular cases of ancestral polymorphisms. Inversions are often on the origin of supergenes, which are sets of genes or clusters of linked genetic loci (haploblocks) that jointly affect the expression of complex

phenotypes. Hager et al. 2021 discovered a 41Mb inversion that is strongly linked to variation in tail length and coat color in forest and prairie ecotypes of deer mice, and that has helped maintained highly divergent adaptive alleles in strong linkage disequilibrium (Hager et al. 2021). This inversion originated 160 kya before the divergence of the two ecotypes, thus illustrating a classic case in which potentially adaptive ancestral genetic variation has evolved prior to the origin of the populations or species being studied. Shared ancestral inversion polymorphisms have also been linked to the evolution of mimicry in swallowtail butterfly (*Papilio polytes)* (Zhang et al. 2017), and to the evolution of polymorphic social strategies in *Solenopsis* fire ants (Yan et al. 2020).

While inherited ancestral variation may be a more plausible source for shared adaptive evolution compared to *de novo* mutation, shared standing genetic variation can also be the result of introgressive hybridization events. Within the last decade, there has been increasing evidence that introgression is both a frequent phenomenon in nature and frequent source of adaptive variation in many plants and animals, and at both recent and deep evolutionary time scales (Whitney et al. 2006, 2010, 2015; Abbott et al. 2013; Zhang et al. 2016; Schweizer et al. 2018; Suarez-Gonzalez et al. 2018; Jones et al. 2020; Owens et al. 2020; Todesco et al. 2020). Classic examples of adaptive introgression in plants include introgression of genes involved in herbivore resistance and flowering time in species of the genus *Helianthus sp.*(Whitney et al. 2006, 2010, 2015; Owens et al. 2020), and tolerance to flooded environments in *Iris sp.* (Martin et al. 2006). Among the most emblematic documented cases of introgressed traits in animals is the adaptive introgression of protective warning wing color variation between *Heliconius* species (Zhang et al. 2016; Enciso-Romero et al. 2017; Jay et al. 2018; Edelman et al. 2019), the introgression of adaptive brown winter coats from black-tailed jackrabbits into snowshoe hares inhabiting regions with low snow cover/mild-winters (Jones et al. 2018; Mills et al. 2018; Giska et al. 2019), the introgression of adaptive toxic resistance in species of killifish (Oziolor et al. 2019), and the introgression of black color in some populations of North American gray wolves from dogs (Anderson et al. 2009; Schweizer et al. 2018). Notably, in the case of *Heliconius* species, introgressed genes involved in color-wing were linked to structural rearrangements such as inversions (Jay et al. 2018; Edelman et al. 2019).

Adaptive introgression also seems to be a particularly frequent source of adaptive variation in extreme terrestrial environments. Studies undertaken after those that found EPAS1 gene to be under positive selection in human Tibetans and the Tibetan Mastiff dog breed, found strong evidence that this gene was in fact introgressed through interbreeding with formerly adapted ancestral counterparts, as opposed to a simple model of selection from de novo mutation or from ancestral standing variation. In the case of human Tibetans the donor of EPAS1 was attributed to a Denisovan-like ancestor (Huerta-Sánchez et al. 2014; Zhang et al. 2020), while in Tibetan mastiff the donor was presumed to be the

Tibetan gray wolf (Miao et al. 2017; von Holdt et al. 2017). Furthermore, the most recent study on canid high-altitude adaptation compared the genomes of Tibetan dogs and Himalayan/Tibetan wolves with several lowland wolf and dog lineages and found evidence that EPAS1 was in fact introgressed into both Tibetan dogs and wolves from an unknown and presumably extinct canid lineage (Wang et al. 2020a). This type of introgression from unknown lineages that have not yet been documented, nor can be attributed to a described fossil species, is often named as 'ghost introgression' (Ottenburghs 2020). Similarly, follow-up studies on Greenlandic Inuit adaptation to the High-Arctic found that two genes previously found to be putatively under selection, TBX15 and WARS2, were introgressed into the Inuit from an unknown archaic population related to Denisovans (Racimo et al. 2016). Adaptive introgression thus seem particularly relevant under scenarios of colonization of new/extreme environments by recipient species, which at some point received genetic variants that had already been targeted by natural selection in a donor species which may or may not have become extinct (Abbott et al. 2013).

Taken together, these studies have shown that introgression may occur in different evolutionary time scales and across speciation events, between currently extinct and extant species (Huerta-Sánchez et al. 2014; Racimo et al. 2016), from unknown donors (Wang et al. 2020a), and may involve just a few genes (Jones et al. 2018; Wang et al. 2020a), or structural rearrangements harboring highly divergent adaptive alleles in strong linkage disequilibrium (Jay et al. 2018; Edelman et al. 2019).

1.3. Foxes as models to study adaptation

The genus *Vulpes* includes all members of the Canidae family (order CARNIVORA) that, together with the raccoon dog (*Nyctereutes procyonoides*) and the bat-eared fox (*Otocyon megalotis*), form an ecologically diverse monophyletic clade that diverged from wolf-like canids and south American fox-like canids approximately 9-10 Mya (Figure 1.1, Figure 1.2) (Sillero-Zubiri et al. 2004; Lindblad-Toh et al. 2005; Wayne and Ostrander 2007). The genus *Vulpes* likely originated in the Late Miocene in North America, from where it radiated to other continents. The oldest records of *Vulpes*, comprising North American fossils ca. 8.2 Ma (*Vulpes stenognathus* and *V. kernensis*), are followed by records of a 7 Myr-old species found in Chad, North-Africa (*V. riffautae*) (Bartolini Lucenti and Madurell-Malapeira 2020). This puts this genus among the first members of the Canidae to have entered the 'Old World' (Bartolini Lucenti and Madurell-Malapeira 2020).

Currently, the genus *Vulpes* comprise 12 extant species that can be found distributed all over the major biomes of the Northern Hemisphere (Figure 1.2), occupying a great variety of habitats, including short-grass and mixed-grass prairies (*V. velox*, Swift fox) and semi-arid short open grasslands (*V. bengalensis*, Bengal fox), arid (*V. macrotis,* kit fox) and hyper-arid deserts (*V. zerda,* fennec fox; *V. rueppellii*, Rueppell's fox), extreme cold dry/sub-humid regions at high-altitudes (*V. ferrilata*, Tibetan sand

fox), cold open grassy semi-arid steppes at low altitudes (*V. corsac*, corsac fox), hot steppe regions and grasslands (*V. pallida*, Pale fox; *V. chama*, Cape fox), mountain steppes (*V. cana*, Blanford's fox), the extreme cold Arctic and alpine tundra (*V. lagopus,* arctic fox) (Sillero-Zubiri et al. 2004), and even urban environments (*V. vulpes*, red fox). Among fox species, the red fox (*V. vulpes*) is the largest in size and has the widest range, with a geographic distribution that spans North America, Europe, Asia and parts of North-Africa (Sillero-Zubiri et al. 2004) (Figure 1.2).

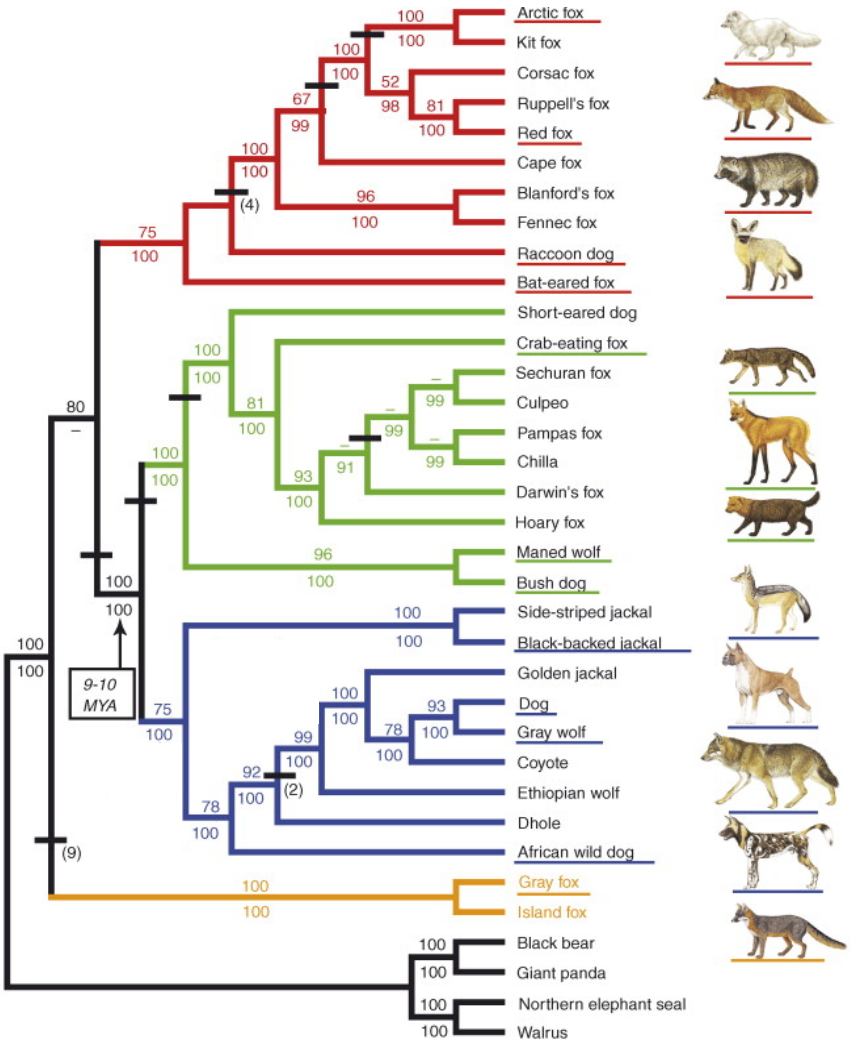


Figure 1.1. Phylogeny of the family Canidae based on 14kb DNA sequence from 12 exons and four introns (adapted from from Lindblad-Toh et al. 2005). The numbers above and below each branching point indicate two different measures of branching point support (bootstrapping and Bayesian posterior probability values). The three main clades of canids are indicated by red (red fox-like canids), green (South American fox-like canids) and blue (wolf-like canids) colors. Example species (underlined) are as follows: Arctic fox (*Vulpes lagopus*), red fox (*Vulpes vulpes*), raccoon dog (*Nyctereutes procyonoides*) and bat-eared fox (*Otocyon megalotis*), crab-eating fox (*Cerdocyon thous*), maned wolf (*Chrysocyon brachyurus*), bush dog (*Speothos venaticus*), domestic dog (*Canis familaris*), gray wolf (*Canis lupus*), golden jackal (*Canis aureus*), coyote (*Canis latrans*), African jackal (*Canis mesomelas*) and gray fox (*Urocyon cinereoargenteus*).

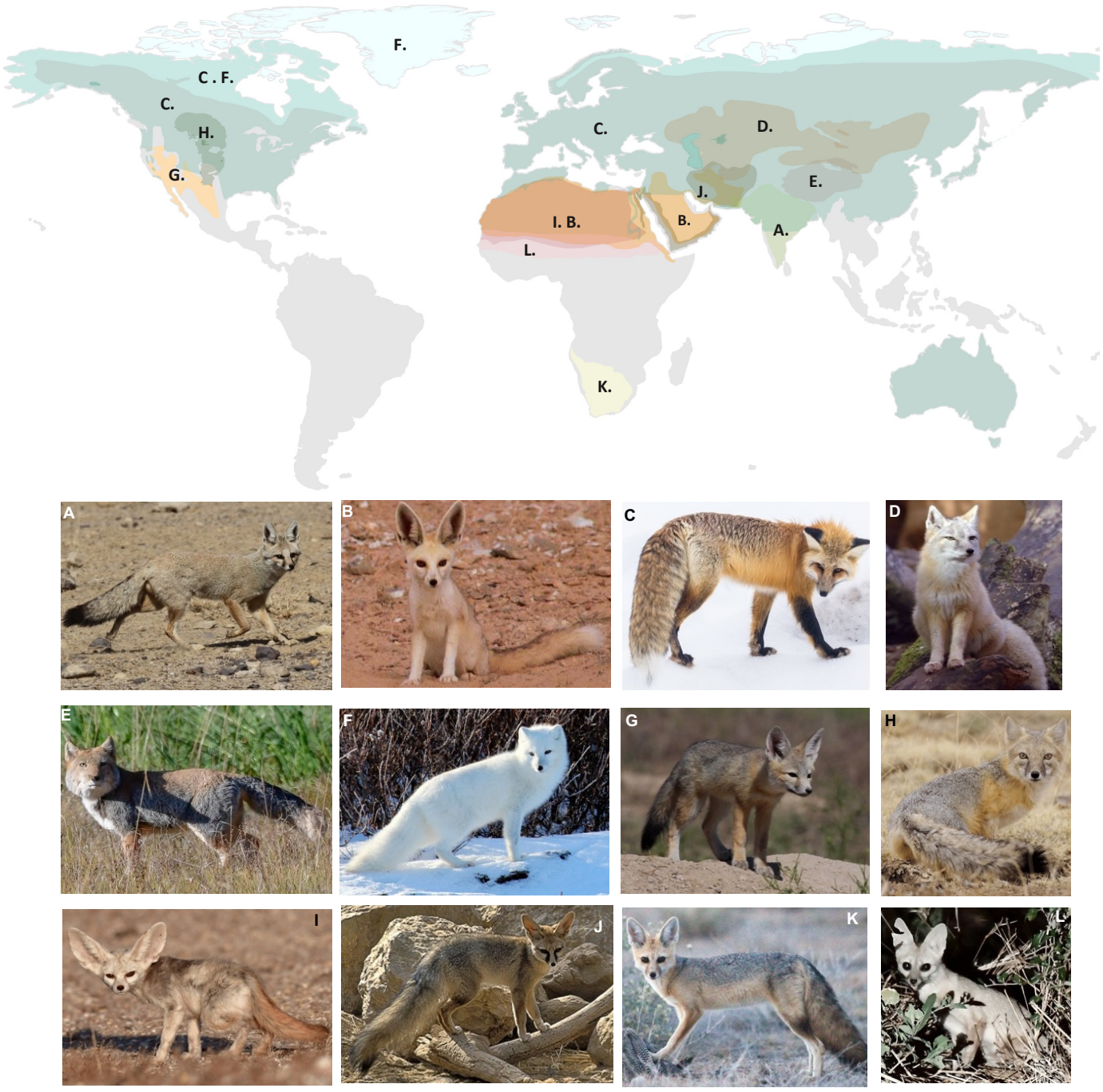


Figure 1.2. Distribution range for the 12 extant species of the genus *Vulpes*: **(A)** Bengal fox, *V. bengalensis (By Mikael Bauer on iNaturalist.org*); **(B)** Rueppell's fox, *V. rueppellii (By Mickaël Villemagne, iNaturalist.org);* **(C)** Red fox, *V. vulpes (By Yellowstone National Park on Flickr. Public domain);* **(D)** Corsac fox, *V. corsac (By Wilm Stan, Wikimedia commons. Public Domain);* **(E)** Tibetan fox, *V. ferrilata (By pfaucher, iNaturalist.org);* **(F)** Arctic fox, *V. lagopus (By Emma, Flickr);* **(G)** Kit fox, *V. macrotis (By Seth Topham. Public domain);* **(H)** Swift fox, *V. velox (By leptim, iNaturalist.org);* **(I)** Fennec fox, *V. zerda (By pfaucher, iNaturalist.org);* **(J)** Blanford's fox *V. cana (Eyal Bartov, Wikimedia);* **(K)** *V. chama (By Yathin S Krishnappa, Wikimedia);* **(L)** Pale fox, *V. pallida (By pfaucher, iNaturalist.*org). Distribution ranges were downloaded from IUCN.

According to molecular phylogenetics, the red fox is sister species to Rueppell's fox (Lindblad-Toh et al. 2005; Leite et al. 2015; Sacks et al. 2018) (Figure 1.1, Figure 1.3). Rueppell's and red fox share a most recent common ancestor with Asian bengal fox (*V. bengalensis*) (analyses performed under the scope of this thesis; Figure 1.3), and are outgrouped by Asian sister species tibetan and corsac fox (Lindblad-Toh et al. 2005; Zhao et al. 2016) (Figure 1.3). Before sharing a common ancestor with North-African fennec and Middle Eastern Blanford's fox, this clade first shares a common ancestor with arctic fox and kit/swift foxes (Lindblad-Toh et al. 2005; Perini et al. 2010; Zhao et al. 2016) (Figure 1.3).

The proposed divergence time between the arctic fox clade and the ancestor to both Asian foxes and the red/Rueppell's fox clade, around 2.9 Ma (Perini et al. 2010), marks the beginning of the Pleistocene, a geological epoch acknowledged by the repeated series of alternating cold glacial and warm interglacial periods that shaped the evolution of many Northern Hemisphere mammalian species, shifting species distribution ranges, and driving multiple speciation and introgressive hybridization events (Hewitt 2000; Stewart et al. 2010). During glacial periods ice sheets would expand and transform entire terrestrial biomes, while during interglacial periods ice sheets contracted (Colinvaux 1997; Hewitt 2000; deMenocal 2004; Stewart et al. 2010).

The deserts and arid regions of the Northern Hemisphere are among the biomes affected by the expansion and retraction of ice sheets (glacial-interglacial cycles), which largely influenced monsoon systems and led to climate oscillations between arid and humid conditions (Hewitt 2000; deMenocal 2004; Drake et al. 2011; Tierney and DeMenocal 2013; Skonieczny et al. 2019). Some of these regions are thought to have been replaced by vegetation cover and wetlands that harbored diverse species (humid periods) (Hewitt 2000; deMenocal 2004; Trauth et al. 2009; Tierney and DeMenocal 2013; Ehrmann et al. 2017; Kutzbach et al. 2020). As former suitable habitats to non-desert organisms were displaced by desertic habitats during dry periods, species would find themselves under increasingly arid conditions, either adapting or retracting in range to climatically suitable regions (Leite et al. 2015; Sarabia et al. 2020). This suggest that Pleistocene-diverged fox species (Perini et al. 2010; Zhao et al. 2016) inhabiting the high-arctic (arctic fox), hot-deserts (kit and Rueppell's fox), cold-arid lowlands (corsac fox) and highlands (Tibetan fox) represent several replications of a natural experiment in which a given common ancestor found itself in extreme conditions and became remarkably adapted.

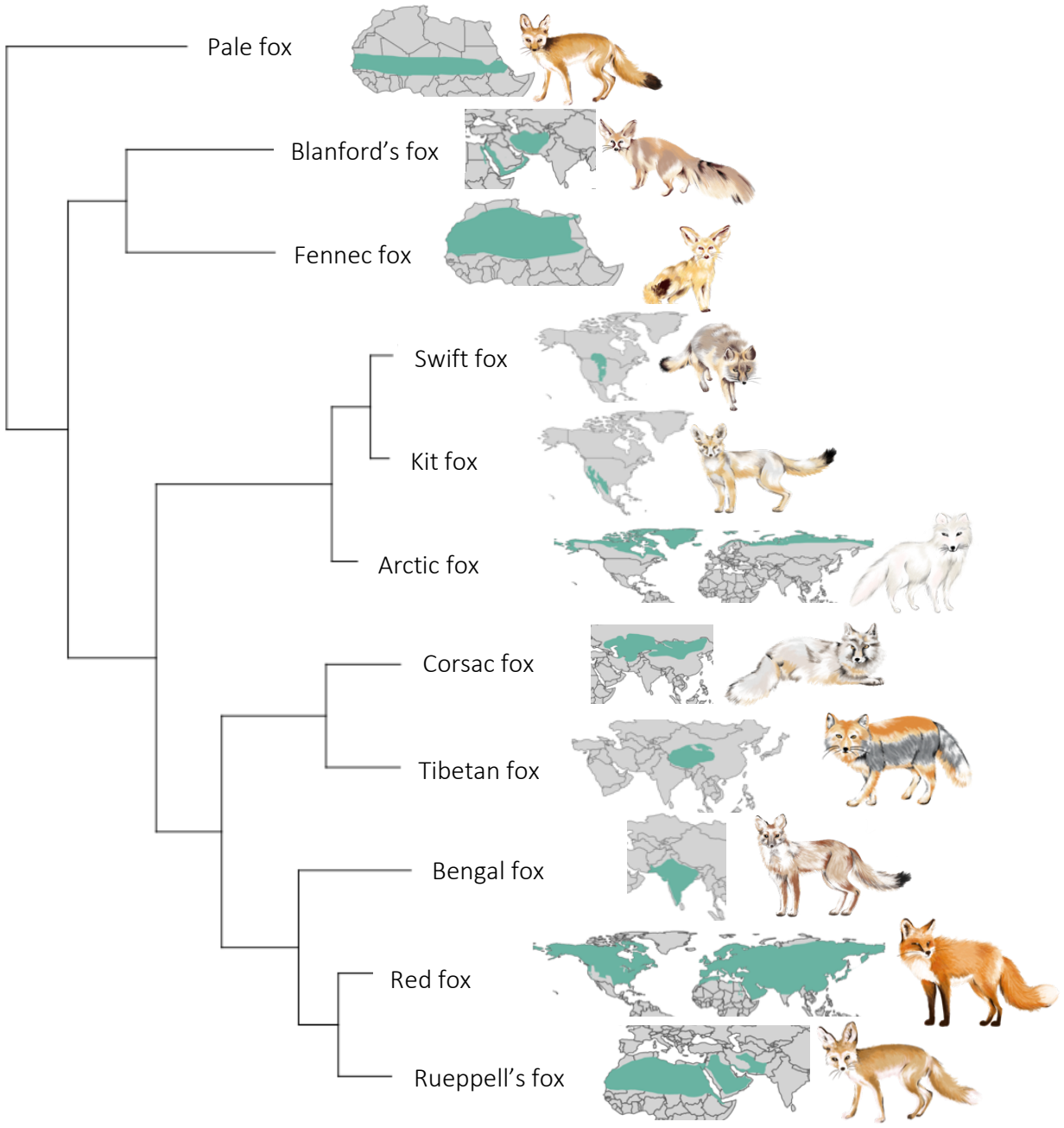


Figure 1.3. Illustration of *Vulpes* cytochrome b gene phylogeny and respective distribution range maps. Phylogenetic tree was inferred by using the Maximum Likelihood method and General Time Reversible model in MEGA. Sequences were downloaded from GenBank for each species. Distribution ranges were downloaded from IUCN (2020) and plotted in R and are followed by scientific illustration of each species (illustrations by *Margarida Tangerina*). Analyses performed under the scope of this thesis.

1.3.1. On the origins of Foxes

The existing gaps in fossil records and the absence of ancient DNA studies using fossil material make it hard to unequivocally pinpoint the geographic and climatic context of some of the recent common ancestors of different clades of extant foxes (Figure 1.3). Based on phylogenetic evidence, the North-African endemic Sahelian pale fox is the most basal outgroup to all extant *Vulpes* species (Leite et al. 2015). The pale fox diverged from the ancestral to all remainder foxes approximately 7 Ma (Leite et al. 2015). This divergence time matches the calibration for the earliest 'Old World' fossil records of Vulpes (*V. riffautae*) in Late Miocene Chad (Bartolini Lucenti and Madurell-Malapeira 2020), and is coincident with the estimated date for the first onset of the Sahara Desert (Zhang et al. 2014b). Another early African fossil species (*V. mathisoni*) described as related to pale fox , cape fox (*V. chama*) and Bengal fox (*V. bengalensis*) was found in Ethiopia 4 Ma (Bartolini Lucenti and Madurell-Malapeira 2020). North-African Fennec and Middle Eastern Blanford's fox are thought to have diverged from remainder foxes approximately 5 Ma (Perini et al. 2010).These findings suggest that 'Old World' *Vulpe*s species originated in Africa at the end of the Miocene before expanding into Eurasia in the Early Pliocene, and back to North America and Africa in the Middle to Late Pleistocene (Bartolini Lucenti and Madurell-Malapeira 2020).

The earliest Eurasian occurrence of *Vulpes sp.* is an Early Pliocene (3.6-5.08 Ma) arctic-fox-like fossil species from the Tibetan Plateau (*V. qiuzhudingi*), presumed to be ancestor of extant Arctic fox (Qiu 2014; Wang et al. 2014b; Bartolini Lucenti and Madurell-Malapeira 2020). Other examples of cold-adapted animals to have their ancestry traced back to Tibet include the wooly rhino (Deng et al. 2011), the snow leopard (Tseng et al. 2014), Pikas (Wang et al. 2020b), sheep (Wang et al. 2016), and currently extinct hypercarnivores such as the canid *Sinicuon dubius* (Wang et al. 2015), and the running hyena *Chasmaporthetes gangsriensis* (Tseng et al. 2013). As the arctic fox clade is the outgroup to other Eurasian species such as Corsac and Tibetan fox (Figure 1.2, Figure 1.3), this would suggest that adaptation to cold started in a cold-snowy Tibet - often called the 'third pole' (Yao et al. 2012) - and this is the ancestral to arctic fox and descending Eurasian fox species (Figure 1.4) that then spread to other regions of the world ('out of Tibet' hypothesis) (Qiu 2014; Wang et al. 2014b; Fuentes-Hurtado et al. 2016).

On the other hand, the morphological traits that make the cold-adapted extinct Tibetan fossil fox (*V. qiuzhudingi*) and arctic fox similar could have independently evolved in response to similar conditions (Qiu 2014). As most fox species dwell in dry sub-humid to arid regions (Figure 1.4), another plausible explanation for the multiple occurrences of desert-like foxes along the *Vulpes* phylogeny is that adaptation to both extreme cold polar and extreme hyper-arid deserts has evolved from semi-arid or arid ancestors. It has been suggested that the arctic fox, for example, has evolved from a kit fox-like ancestor (as proposed in Kumar et al. 2015) or from a temperate/dry sub-humid swift fox-like ancestor (suggested

in Stewart et al. 2010), instead of a cold-adapted Tibetan ancestor. Regrettably, the huge time gap in the *Vulpes* fossil record and absence of ancient DNA material has not yet allowed to explicitly test which scenario is correct (Qiu 2014). This is further hampered by uncertainties within the extant *Vulpes* phylogeny, which remains unresolved. For example, the phylogenetic position of the Cape fox varies across the literature, with some studies placing it as the extant ancestral to arctic fox (as in Figure 1.1) (Lindblad-Toh et al. 2005), while others place it as a basal *Vulpes* and the sister species to Pale fox (Geffen et al. 1992; Nyakatura and Bininda-Emonds 2012).

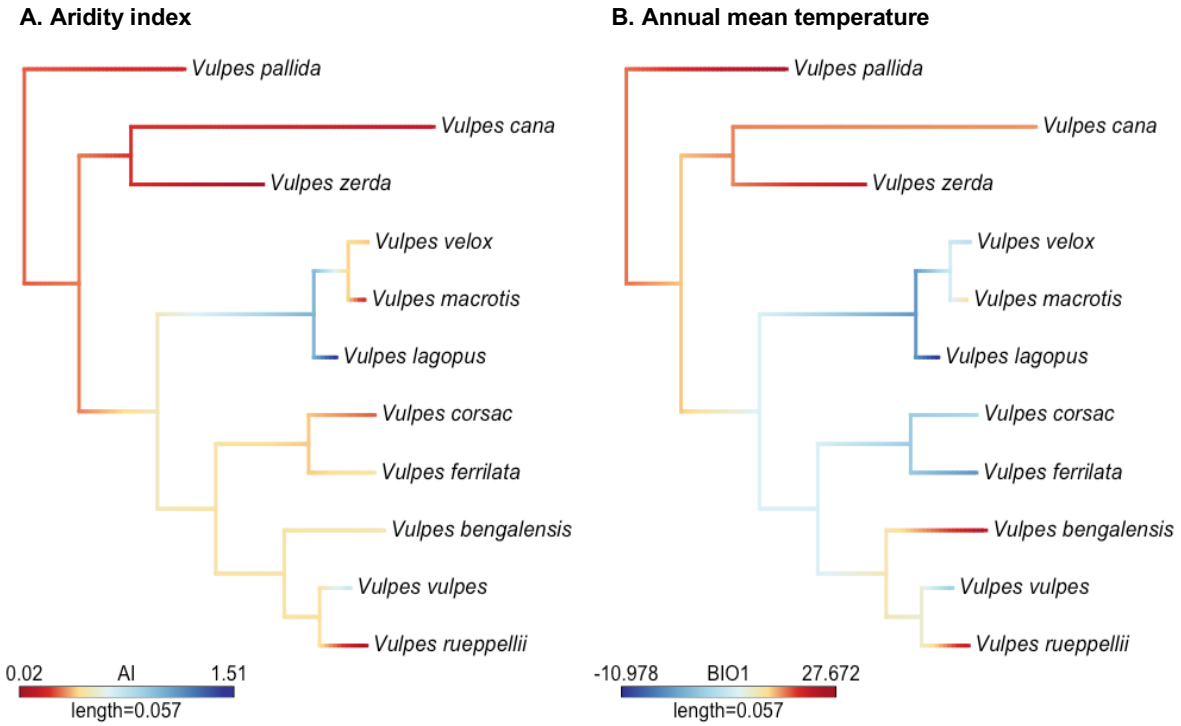


Figure 1.4. Representation of mean Annual aridity index (**A**) and Annual mean temperature (**B**) in a cytochrome b phylogenetic tree with 11 *Vulpes* species represented in Figure 1.3. Aridity index was obtained from the CGIAR-CSI Global Aridity database (Trabucco and Zomer 2009), and annual mean temperature (BIO1) from the WorldClim database (Fick and Hijmans 2017). Distribution ranges were downloaded from IUCN (2020) and Geographical Information System was used to calculate mean values for each environmental variable within the geographic range polygon of each species. The mean aridity index for the distribution range of different *Vulpes sp* can be classified into discrete categories as follows: hyper-arid (Rueppell's fox *V.rueppellii*; Fennec, *V. zerda*; dark red), arid (Kit fox, *V. macrotis*; Blanford's fox, *V. cana*; Pale fox *V. pallida*; red), semi-arid (Corsac fox, *V. corsac*; Swift fox, *V. velox*; orange-red), dry sub-humid (Tibetan fox, *V. ferrilata*; Bengal fox, *V. bengalensis*; yellow), humid (red fox, *V. vulpes*; Arctic fox, *V. lagopus*; blue). Ancestral states for mean AI and BIO1, estimated using maximum-likelihood methods (implemented with the R package 'phytools'), were projected onto the internal edges and nodes of the tree using a color gradient. These analyses were performed under the scope of this thesis and are presented to serve as an illustrative representation of species ecological contexts. Since trees species distribution ranges and environmental variables both rely on present-day information, assumptions on ancestral environments are inherently imprecise and should thus be interpreted with caution.

Despite these challenges, it is clear that the wide variety of different habitats occupied by foxes and the simultaneous presence of ecologically and morphologically distinct pairs of sister species living in different environmental settings throughout the *Vulpes* phylogeny (e.g. arctic/kit fox, corsac/Tibetan fox, red/Rueppell's fox*;* see Figure 1.3 and Figure 1.4), as well as ecologically and morphologically similar long-diverged species (e.g. desert-adapted Rueppell's, kit and fennec foxes; see Figure 1.3 and Figure 1.4), make them unique candidates to study adaptation and convergent evolution in extreme environments.

1.3.2. The intriguing evolutionary history of Red and Rueppell's foxes

Among fox species, Rueppell's and red foxes are of particular interest to study speciation, introgression and adaptation. The Rueppell's fox is a desert-dwelling species found widely distributed in arid and hyper-arid landscapes that extend from the North-African Sahara Desert and Arabian Empty Quarter to the Middle Eastern deserts and as far east as Pakistan (Sillero-Zubiri et al. 2004) (Figure 1.3, Figure 1.5). In North-Africa, Rueppell's live in close sympatry with long-diverged fennec fox (Sillero-Zubiri et al. 2004; Brito et al. 2009) (Figure 1.3; Figure 1.5).

Rueppell's and fennec foxes are physiologically, behaviorally and morphologically adapted to cope with life in extreme desert environments (Larivière and Seddon 2001; Sillero-Zubiri et al. 2004; Mallon, D., Murdoch, J.D. & Wacher 2015). Both species have enlarged ears, which are thought to provide an increased area for cooling the body without losing water by means of evaporation, hair on pads to protect from the hot sand temperatures, and pale coat colors to potentially reduce heat absorption whereas promoting background matching with a sandy desert (Larivière and Seddon 2001; Sillero-Zubiri et al. 2004; Mallon, D., Murdoch, J.D. & Wacher 2015). In the wild, they are nocturnal and use denning behavior during the day (Larivière and Seddon 2001; Sillero-Zubiri et al. 2004; Mallon, D., Murdoch, J.D. & Wacher 2015). Physiological studies in captive individuals showed that fennec and Rueppell's foxes have reduced mass-adjusted metabolic rates and non-renal water losses in comparison to canid species living in habitats with moderate or well-balanced supply of moisture, suggesting particularly efficiency at reducing cutaneous and/or respiratory water loss (Williams et al. 2002, 2004; Wilmer et al. 2005; Careau et al. 2007). Fennecs are also capable of producing highly concentrated urine, a record that is unknown in Rueppell's, but notoriously known as an adaptive trait for increased water-retention in many desert mice (Noll-Banholzer 1979; Maloiy et al. 1982).

Red foxes, on the other hand, can be found in diverse habitats throughout the Northern Hemisphere, from urban environments to latitudes as far north as the southern fringes of the arctic circle and as far south as the northern periphery of deserts (see Figure 1.2) (Sillero-Zubiri et al. 2004). The degree of morphological and ecological diversity seen in world-wide distributed red fox populations (Fig

1.6 A) makes this species an interesting candidate to investigate signatures of natural selection. The process of colonization of red foxes throughout the world was likely accompanied by local adaptation, competition and/or introgressive hybridization events with resident congeners. On the other hand, their distribution is limited to the outer edges of deserts and arctic (Figure 1.2), and thus they may not be genetically adapted to life in these environments, being highly phenotypically plastic instead. Only populations that have lived in extreme conditions over multiple generations are likely to have been subjected to selective pressures and have adapted physiologically and genetically to live in these environments. Despite overlapping in range with Rueppell's and fennec fox at the northern periphery of the Sahara, and with Rueppell's fox in the semi-arid steppes of the Middle East (Figure 1.5), it is generally presumed that red foxes are unable to survive at the core of deserts, or in arid regions without water availability (Sillero-Zubiri et al. 2004).

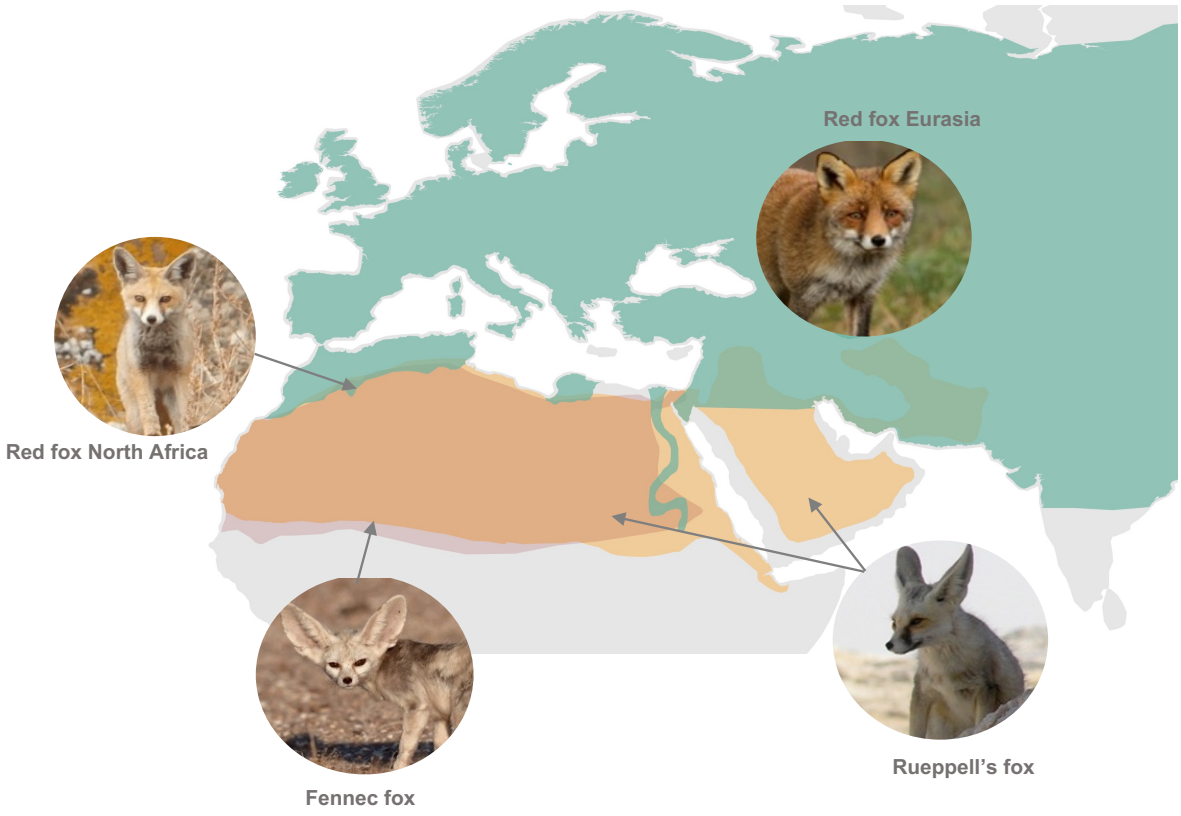


Figure 1.5. Regions with overlapping distribution between sister species Rueppell's (*V. rueppellii*; *photo by Jesper Särnesjö, Wikimedia*), red fox (*V. vulpes*; *photo by Paul Cools on iNaturalist.org*) and basal group Fennec fox (*V. zerda; photo by pfaucher, iNaturalist.org*). Rueppell's fox is found in the deserts of North-Africa and the Middle East (yellow). In North-Africa, Rueppell's shares most of the range of Fennec fox (orange) and overlaps with sister species red fox (photo by *Aissa Djamel Filali, iNaturalist.org*) at the Saharan borders and/or in regions with water availability. Distribution ranges were downloaded from IUCN and plotted in R.

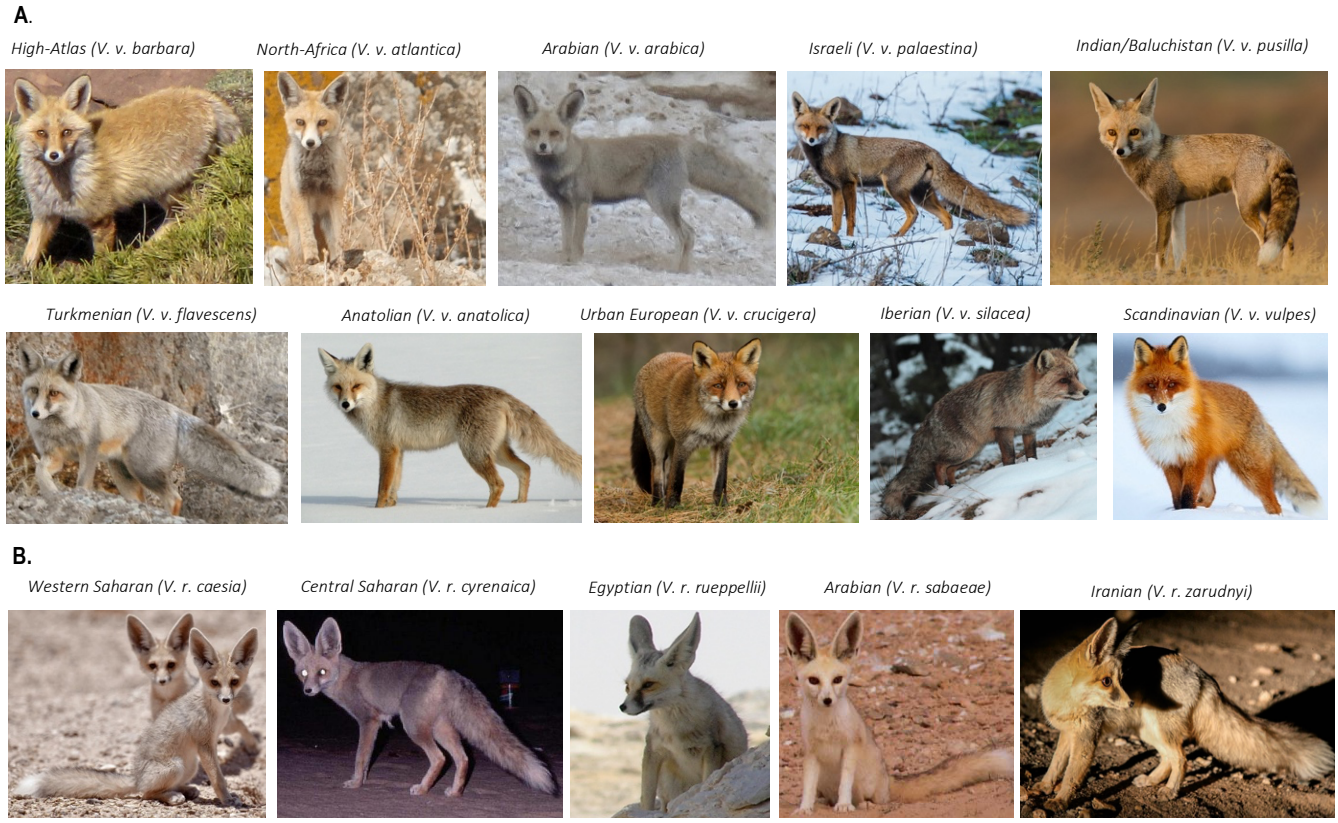


Figure 1.6. Illustration of 10 out of 44 morphological subspecies of red fox (**A**) and 5 subspecies of Rueppell's fox (**B**), though many are doubtful and do not translate genetic variability. Photo credits: *V. .v. barbara (Photo by abdellahbouazza on iNaturalist.org. Oukaïmeden, Morocco); V. v. atlantica (Photo by Aissa Djamel Filali on iNaturalist.org. Algeria); V. v. arabica (Photo by tbrooks on iNaturalist.org. Oman); V. v. palaestina (Photo by hagai_michaeli on iNaturalist.org. Golan, Israel); V. v. pusilla (Photo by Balaji Venkatesh Sivaramakrishnan on Flickr. India); V. v. flavescens (Photo by haniyeh_kiaee on iNaturalist.org. Iran); V. v. anatolica (Photo by A. Omer Karamollaoglu on Flickr. Turkey); V.v. crucigera (Photo by Paul Cools on iNaturalist.org); V. v. silacea (Photo by Agustín Povedano on Flickr. Spain); V. v. vulpes (Photo by Espen Bergersen in Nature Picture Library. Finland); V. r. caesia (Photo by pfaucher on iNaturalist.org); V. r. cyrenaica (Rabeil, T., Newby, J. and Harouna, A.); V. r. rueppellii (Photo by Jesper Särnesjö on Wikimedia); V. r. sabaeae (Photo by Mickaël Villemagne iNaturalist.org); V. r. zarudnyi (Photo by Bahman Izadi, Lut Desert, Iran).*

Several phylogeographic studies have placed the Middle East as cradle of modern red fox genetic diversity, which then expanded throughout Eurasia, and through the Beringian strait into North America (Aubry et al. 2009; Kutschera et al. 2013; Statham et al. 2014; Sacks et al. 2018; McDevitt et al. 2020). While these studies were able to fully resolve the geographic patterns of genetic diversity and history of European and North American colonization, how population dynamics in Eurasian red foxes relate to those currently inhabiting the African continent remains to elucidate to date. Both past and recent studies focused on red fox phylogeography either have failed to include samples from North-Africa or did not

have enough sample size and detailed geographic context for demographic inference (Statham et al. 2014; Sacks et al. 2018; McDevitt et al. 2020). Similarly, the Rueppell's fox (Figure 1.6 B) has not been the target of many population genetics or phylogeographic studies. As most outgroups to the Rueppell/red fox clade are of Eurasian origin (Figure 1.2), the two species likely diverged in the Middle East or south Asia and then spread into North-Africa at distinct points in time during the Pleistocene. However, such population dynamics remains poorly understood.

The only study to date focused in assessing the evolutionary history of genetic differentiation in Rueppell's and red foxes found high genetic similarity between the two species (Leite et al. 2015). In this study, phylogenetic analysis using short mitochondrial markers found that North-African Rueppell's are more closely related to North-African red foxes than the latter with Eurasian red foxes. This paraphyletic pattern has also been found in bears, with North American brown bears being more closely related to North American polar bears, than the later with European brown bears (Hailer et al. 2012; Liu et al. 2014). Population genetic analyses using 46 nuclear makers (microsatellites) were in discordance with mitochondrial evidence and found the two species to be monophyletic, though showing low level of genetic differentiation for two presumably distinct mammalian species (F_{ST}=0.14). Both markers were able to separate North-African from Eurasian red foxes (F_{ST}=0.05). Leite et al. 2015 explained nuclear-mtDNA discordance as either the result of past introgression of red fox into Rueppell's or as Ruppell's fox being a recently diverged desert ecotype of red fox (Leite et al. 2015).

The node age for all Rueppell and red foxes using mtDNA sequences was estimated to be at least 1 Mya, which matches a time frame during which global temperatures fluctuated between warm and cold phases, providing the environmental context for the evolution of many species' pairs with different ecological adaptations (Kumar et al. 2015; Leite et al. 2015). Notably, this time frame coincides with the divergence between arctic Fox and desert kit fox, between Asian Corsac and Tibetan fox (Figure 1.3) (Hailer et al. 2012; Liu et al. 2014; Kumar et al. 2015; Zhao et al. 2016). This suggests that desert-dwelling Rueppell's fox, and possibly red foxes inhabiting increasingly arid regions in North-Africa may have consecutively adapted to desertic conditions long-after desert-dwelling basal foxes such as the fennec fox (see Figure 1.3).

While intraspecific studies may offer a detailed perspective on recent or ongoing selection, and adaptation in short evolutionary time scales, comparative studies among closely related groups of species have the potential to elucidate the evolutionary routes underlying the origin of adaptive variation. Overall, the striking genetic similarity found between North-African red fox and Rueppell's fox, which sharply contrasts with their morphological (Figure 1.3), ecological (Figure 1.4) and behavioral distinctiveness offers an unprecedented study case to investigate signatures of selection and introgression in either species. The separation of red foxes in distinct North-African and Eurasian lineages

offers a model to investigate recent or ongoing local adaptation to divergent selective regimes. Finally, the evolutionary context of these foxes in North-Africa provides a unique additional opportunity to investigate whether adaptive changes found in Rueppell's fox, and potentially in North-African red foxes, were also selected in sympatric ancestral desert-dwelling fennec fox.

1.4. Objectives

Deserts are among the harshest environments on Earth. However, little is known about the mechanisms driving adaptive evolution in these terrestrial extremes. The major goal of this thesis was to advance our understanding of the mechanisms driving adaptation in deserts by reviewing the recent but growing body work on the genomics of mammalian desert adaptation, and by studying the demographic and adaptive history of North-African foxes. In particular, we aimed to answer the following questions:

1. Which genes and gene changes underlie desert adaptation in mammals?
2. How are genes related to function and past classical eco-physiological research?
3. Can desert climate help predict desert adaptive phenotypes?
4. What is the level of convergent evolution in deserts?
5. What is the role of past climatic oscillations in driving desert adaptation?
6. What is the source for adaptive genetic variation in desert environments?
7. Does the timescale for adaptation matter for the genomic patterns observed?

Taking advantage of a large sample collection of North-African fox species, namely the sister species Rueppell's fox (*Vulpes rueppellii*) and red fox (*V. vulpes*), and outgroup species fennec fox (*V. zerda*) and pale fox (*V. pallida*), we used a set of integrative approaches following phenotyping, and whole-genome and transcriptome sequencing, to investigate adaptation at different evolutionary times in these iconic model group of species.

1.5. Work outline

This thesis is divided into four chapters. Chapter I provides a general introduction to the main topics relating to the work developed in this thesis, which is then presented in Chapter II and III in the form of three scientific manuscripts (two published and one in preparation). The scientific outputs featured in Chapter II provide general knowledge, basis and motivation for Chapter III. Finally, the main findings and contributions of this dissertation are contextualized and discussed in Chapter IV.

In chapter II, we review current genomic work on the genetic basis of mammalian desert adaptation, answering questions 1 to 4. We find shared evidence of selection among different mammalian species in several genes, and functional classes of genes, that may underlie complex phenotypes associated with adaptation to water/food scarcity and extreme temperatures. We then explore our current understanding of the limits and future potential of using genomics to study adaptive evolution in deserts and provide explicit recommendations for future research that may help resolve outstanding questions. Our review also reflects on the striking paucity of studies focused on desert (non-model) organisms, particularly when contrasted with knowledge on adaptive evolution in other terrestrial extremes. This work is currently published at *Trends in Ecology and Evolution* (https://doi.org/10.1016/j.tree.2021.03.007).

Still in chapter II, we present the results of a meta-analysis that aims to understand how desert adaptive traits described over decades of eco-physiology research relate to environmental factors. We use phylogenetic comparative methods to explicitly test the hypothesis that adaptive phenotypes have repeatedly and independently evolved in different desert-dwelling mammals. We find that the mean annual aridity index of a species range largely predicts its urine-concentrating ability, suggesting that aridity may be a key selective pressure driving phenotypic evolution in desert environments (question 3). This work is published at *Mammal Review* (https://doi.org/10.1111/mam.12244).

In Chapter III we study the genetic basis of desert adaptation in North-African foxes (*Vulpes sp.*) and contribute with empirical evidence on the mechanisms shaping adaptive evolution at different evolutionary scales (questions 5 to 7). By studying the demographic and adaptive history of Rueppell's and red foxes, we find that past gene flow between the two sister species in North-Africa likely facilitated red fox recent adaptation to an expanding Sahara, after they split from Eurasian populations ~78 Kya. Further comparative analysis identified a highly unusual and divergent 25 Mb genomic region in Rueppell's fox that can be explained as the result of an ancestral hybridization from a fennec fox-like species. Scans for signatures of adaptation revealed signals at genes associated with water retention through reduced panting and sweating (SLC12A2), temperature perception (TRPA1) and adaptive thermogenesis (THADA) in North-African red fox, and genes affecting renal water homeostasis in Rueppell's (WFS1) and fennec fox (KIRREL1, CACNA1H), some of which were supported by expression and/or physiological differences. These discoveries underscore the complexity of canid evolution and suggest that both standing adaptive genetic variation and introgression with older resident desert species have been key drivers of desert adaptation in North-African foxes and are likely to be critical components of evolutionary responses to climate change. This article is currently being prepared for submission to a peer reviewed journal.

References

Abbott R., Albach D., Ansell S., Arntzen J.W., Baird S.J.E., Bierne N., Boughman J., Brelsford A., Buerkle C.A., Buggs R., Butlin R.K., Dieckmann U., Eroukhmanoff F., Grill A., Cahan S.H., Hermansen J.S., Hewitt G., Hudson A.G., Jiggins C., Jones J., Keller B., Marczewski T., Mallet J., Martinez-Rodriguez P., Möst M., Mullen S., Nichols R., Nolte A.W., Parisod C., Pfennig K., Rice A.M., Ritchie M.G., Seifert B., Smadja C.M., Stelkens R., Szymura J.M., Väinölä R., Wolf J.B.W., Zinner D. 2013. Hybridization and speciation. J. Evol. Biol. 26:229–246.

Alkorta-Aranburu G., Beall C.M., Witonsky D.B., Gebremedhin A., Pritchard J.K., Di Rienzo A. 2012. The Genetic Architecture of Adaptations to High Altitude in Ethiopia. PLoS Genet. 8.

Anderson T.M., VonHoldt B.M., Candille S.I., Musiani M., Greco C., Stahler D.R., Smith D.W., Padhukasahasram B., Randi E., Leonard J.A., Bustamante C.D., Ostrander E.A., Tang H., Wayne R.K., Barsh G.S. 2009. Molecular and evolutionary history of melanism in Norths American gray wolves. Science (80-.). 323:1339–1343.

Aubry K.B., Statham M.J., Sacks B.N., Perrine J.D., Wisely S.M. 2009. Phylogeography of the North American red fox: Vicariance in Pleistocene forest refugia. Mol. Ecol. 18:2668–2686.

Bartolini Lucenti S., Madurell-Malapeira J. 2020. Unraveling the fossil record of foxes: An updated review on the Plio-Pleistocene Vulpes spp. from Europe. Quat. Sci. Rev. 236.

Beall C.M., Cavalleri G.L., Deng L., Elston R.C., Gao Y., Knight J., Li C., Li J.C., Liang Y., McCormack M., Montgomery H.E., Pan H., Robbins P.A., Shianna K. V., Tam S.C., Tsering N., Veeramah K.R., Wang W., Wangdui P., Weale M.E., Xu Y., Xu Z., Yang L., Zaman M.J., Zeng C., Zhang L., Zhang X., Zhaxi P., Zheng Y.T. 2010. Natural selection on EPAS1 (HIF2α) associated with low hemoglobin concentration in Tibetan highlanders. Proc. Natl. Acad. Sci. U. S. A. 107:11459–11464.

Bedford N.L., Hoekstra H.E. 2015. Peromyscus mice as a model for studying natural variation. Elife. 4.

Bell E. 2012. Life at extremes : environments, organisms, and strategies for survival. CAB International.

Breed M.F., Harrison P.A., Blyth C., Byrne M., Gaget V., Gellie N.J.C., Groom S.V.C., Hodgson R., Mills J.G., Prowse T.A.A., Steane D.A., Mohr J.J. 2019. The potential of genomics for restoring ecosystems and biodiversity. Nat. Rev. Genet. 20:615–628.

Brito J.C., Acosta A.L., Álvares F., Cuzin F. 2009. 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.

Careau V., Morand-Ferron J., Thomas D. 2007. Basal Metabolic Rate of Canidae from Hot Deserts to Cold Arctic Climates. J. Mammal. 88:394–400.

Cho Y.S., Hu L., Hou H., Lee H., Xu J., Kwon S., Oh S., Kim H.M., Jho S., Kim S., Shin Y.A., Kim B.C.,

Kim H., Kim C.U., Luo S.J., Johnson W.E., Koepfli K.P., Schmidt-Küntzel A., Turner J.A., Marker L., Harper C., Miller S.M., Jacobs W., Bertola L.D., Kim T.H., Lee S., Zhou Q., Jung H.J., Xu X., Gadhvi P., Xu P., Xiong Y., Luo Y., Pan S., Gou C., Chu X., Zhang J., Liu S., He J., Chen Y., Yang L., Yang Y., He J., Liu S., Wang J., Kim C.H., Kwak H., Kim J.S., Hwang S., Ko J., Kim C.B., Kim S., Bayarlkhagva D., Paek W.K., Kim S.J., O'Brien S.J., Wang J., Bhak J. 2013. The tiger genome and comparative analysis with lion and snow leopard genomes. Nat. Commun. 4:443–270.

Colella J.P., Talbot S.L., Brochmann C., Taylor E.B., Hoberg E.P., Cook J.A. 2020. Conservation Genomics in a Changing Arctic. Trends Ecol. Evol. 35:149–162.

Colinvaux P.A. 1997. An arid Amazon? Trends Ecol. Evol. 12:318–319.

Corlett R.T. 2017. A Bigger Toolbox: Biotechnology in Biodiversity Conservation. Trends Biotechnol. 35:55–65.

Crawford J.E., Amaru R., Song J., Julian C.G., Racimo F., Cheng J.Y., Guo X., Yao J., Ambale-Venkatesh B., Lima J.A., Rotter J.I., Stehlik J., Moore L.G., Prchal J.T., Nielsen R. 2017. Natural Selection on Genes Related to Cardiovascular Health in High-Altitude Adapted Andeans. Am. J. Hum. Genet. 101:752–767.

Crawford J.E., Nielsen R. 2013. Detecting adaptive trait loci in nonmodel systems: divergence or admixture mapping? Mol. Ecol. 22:6131–6148.

deMenocal P.B. 2004. African climate change and faunal evolution during the Pliocene-Pleistocene. Earth Planet. Sci. Lett. 220:3–24.

Deng T., Wang X., Fortelius M., Li Q., Wang Y., Tseng Z.J., Takeuchi G.T., Saylor J.E., Säilä L.K., Xie G. 2011. Out of Tibet: Pliocene woolly rhino suggests high-plateau origin of ice age megaherbivores. Science (80-.). 333:1285–1288.

Drake N.A., Blench R.M., Armitage S.J., Bristow C.S., White K.H. 2011. Ancient watercourses and biogeography of the Sahara explain the peopling of the desert. Proc. Natl. Acad. Sci. U. S. A. 108:458–462.

Edelman N.B., Frandsen P.B., Miyagi M., Clavijo B., Davey J., Dikow R.B., García-Accinelli G., Van Belleghem S.M., Patterson N., Neafsey D.E., Challis R., Kumar S., Moreira G.R.P., Salazar C., Chouteau M., Counterman B.A., Papa R., Blaxter M., Reed R.D., Dasmahapatra K.K., Kronforst M., Joron M., Jiggins C.D., Owen McMillan W., Palma F. Di, Blumberg A.J., Wakeley J., Jaffe D., Mallet J. 2019. Genomic architecture and introgression shape a butterfly radiation. Science (80-.). 366:594–599.

Ehrmann W., Schmiedl G., Beuscher S., Krüger S. 2017. Intensity of African Humid Periods Estimated from Saharan Dust Fluxes. PLoS One. 12:e0170989.

Enciso-Romero J., Pardo-Díaz C., Martin S.H., Arias C.F., Linares M., McMillan W.O., Jiggins C.D.,

Salazar C. 2017. Evolution of novel mimicry rings facilitated by adaptive introgression in tropical butterflies. Mol. Ecol. 26:5160–5172.

Exposito-Alonso M., Exposito-Alonso M., Gómez Rodríguez R., Barragán C., Capovilla G., Chae E., Devos J., Dogan E.S., Friedemann C., Gross C., Lang P., Lundberg D., Middendorf V., Kageyama J., Karasov T., Kersten S., Petersen S., Rabbani L., Regalado J., Reinelt L., Rowan B., Seymour D.K., Symeonidi E., Schwab R., Tran D.T.N., Venkataramani K., Van de Weyer A.L., Vasseur F., Wang G., Wedegärtner R., Weiss F., Wu R., Xi W., Zaidem M., Zhu W., García-Arenal F., Burbano H.A., Bossdorf O., Weigel D., Burbano H.A., Bossdorf O., Nielsen R., Weigel D. 2019. Natural selection on the Arabidopsis thaliana genome in present and future climates. Nature. 573:126–129.

Fick S.E., Hijmans R.J. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol. 37:4302–4315.

Figueiró H. V., Li G., Trindade F.J., Assis J., Pais F., Fernandes G., Santos S.H.D., Hughes G.M., Komissarov A., Antunes A., Trinca C.S., Rodrigues M.R., Linderoth T., Bi K., Silveira L., Azevedo F.C.C., Kantek D., Ramalho E., Brassaloti R.A., Villela P.M.S., Nunes A.L.V., Teixeira R.H.F., Morato R.G., Loska D., Saragüeta P., Gabaldón T., Teeling E.C., O'Brien S.J., Nielsen R., Coutinho L.L., Oliveira G., Murphy W.J., Eizirik E. 2017. Genome-wide signatures of complex introgression and adaptive evolution in the big cats. Sci. Adv. 3:1–14.

Fuentes-Hurtado M., Hof A.R., Jansson R. 2016. Paleodistribution modeling suggests glacial refugia in Scandinavia and out-of-Tibet range expansion of the Arctic fox. Ecol. Evol. 6:170–180.

Fumagalli M., Moltke I., Grarup N., Racimo F., Bjerregaard P., Jørgensen M.E., Korneliussen T.S., Gerbault P., Skotte L., Linneberg A., Christensen C., Brandslund I., Jørgensen T., Huerta-Sánchez E., Schmidt E.B., Pedersen O., Hansen T., Albrechtsen A., Nielsen R. 2015. Greenlandic Inuit show genetic signatures of diet and climate adaptation. Science (80-.). 349:1343–1347.

Geffen E., Mercure A., Girman D.J., Macdonald D.W., Wayne R.K. 1992. Phylogenetic relationships of the fox-like canids: mitochondrial DNA restriction fragment, site and cytochrome *b* sequence analyses. J. Zool. 228:27–39.

Giska I., Farelo L., Pimenta J., Seixas F.A., Ferreira M.S., Marques J.P., Miranda I., Letty J., Jenny H., Hackländer K., Magnussen E., Melo-Ferreira J. 2019. Introgression drives repeated evolution of winter coat color polymorphism in hares. Proc. Natl. Acad. Sci. U. S. A. 116:24150–24156.

Gou X., Wang Z., Li N., Qiu F., Xu Z., Yan D., Yang S., Jia, Konga X., Wei Z., Lu S., Lian L., Wu C., Wang X., Li G., Teng M., Jiang Q., Zhao X., Yang J., Liu B., Wei D., Li H., Yang J., Yan Y., Zhao G., Dong X., Li M., Deng W., Leng J., Wei C., Wang C., Mao H., Zhang H., Ding G., Li Y. 2014. Whole-genome sequencing of six dog breeds from continuous altitudes reveals adaptation to high-

altitude hypoxia. Genome Res. 24:1308–1315.

Hager E.R., Harringmeyer O.S., Brock Wooldridge T., Theingi S., Gable J.T., McFadden S., Neugeboren B., Turner K.M., Hoekstra H.E. 2021. A chromosomal inversion drives evolution of multiple adaptive traits in deer mice. bioRxiv.:2021.01.21.427490.

Hailer F., Kutschera V.E., Hallström B.M., Klassert D., Fain S.R., Leonard J.A., Arnason U., Janke A. 2012. Nuclear genomic sequences reveal that polar bears are an old and distinct bear lineage. Science (80-.). 336:344–347.

Hedrick P.W. 2013. Adaptive introgression in animals: Examples and comparison to new mutation and standing variation as sources of adaptive variation. Mol. Ecol. 22:4606–4618.

Hedrick P.W., Fredrickson R. 2010. Genetic rescue guidelines with examples from Mexican wolves and Florida panthers. Conserv. Genet. 11:615–626.

Hewitt G. 2000. The genetic legacy of the quaternary ice ages. Nature. 405:907–913.

Hoffmann A.A., Miller A.D., Weeks A.R. 2020. Genetic mixing for population management: From genetic rescue to provenancing. Evol. Appl.

von Holdt B., Fan Z., Ortega-Del Vecchyo D., Wayne R.K. 2017. EPAS1 variants in high altitude Tibetan wolves were selectively introgressed into highland dogs. PeerJ. 2017:e3522.

Huerta-Sánchez E., DeGiorgio M., Pagani L., Tarekegn A., Ekong R., Antao T., Cardona A., Montgomery H.E., Cavalleri G.L., Robbins P.A., Weale M.E., Bradman N., Bekele E., Kivisild T., Tyler-Smith C., Nielsen R. 2013. Genetic Signatures Reveal High-Altitude Adaptation in a Set of Ethiopian Populations. Mol. Biol. Evol. 30:1877–1888.

Huerta-Sánchez E., Jin X., Asan, Bianba Z., Peter B.M., Vinckenbosch N., Liang Y., Yi X., He M., Somel M., Ni P., Wang B., Ou X., Huasang, Luosang J., Cuo Z.X.P., Li K., Gao G., Yin Y., Wang W., Zhang X., Xu X., Yang H., Li Y., Wang J., Wang J., Nielsen R. 2014. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature. 512:194–197.

Ilardo M., Nielsen R. 2018. Human adaptation to extreme environmental conditions. Curr. Opin. Genet. Dev. 53:77–82.

Jamie G.A., Meier J.I. 2020. The Persistence of Polymorphisms across Species Radiations. Trends Ecol. Evol. 35:795–808.

Jay P., Whibley A., Frézal L., Rodríguez de Cara M.Á., Nowell R.W., Mallet J., Dasmahapatra K.K., Joron M. 2018. Supergene Evolution Triggered by the Introgression of a Chromosomal Inversion. Curr. Biol. 28:1839-1845.e3.

Jones M.R., Mills L.S., Jensen J.D., Good J.M. 2020. The Origin and Spread of Locally Adaptive Seasonal Camouflage in Snowshoe Hares. Am. Nat. 196:316–332.

Jones M.R., Scott Mills L., Alves P.C., Callahan C.M., Alves J.M., Lafferty D.J.R., Jiggins F.M., Jensen

J.D., Melo-Ferreira J., Good J.M. 2018. Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. Science (80-.). 360:1355–1358.

Key F.M., Abdul-Aziz M.A., Mundry R., Peter B.M., Sekar A., D'Amato M., Dennis M.Y., Schmidt J.M., Andrés A.M. 2018. Human local adaptation of the TRPM8 cold receptor along a latitudinal cline. PLoS Genet. 14:1–22.

Kumar V., Kutschera V.E., Nilsson M.A., Janke A. 2015. Genetic signatures of adaptation revealed from transcriptome sequencing of Arctic and red foxes. BMC Genomics. 16.

Kutschera V.E., Lecomte N., Janke A., Selva N., Sokolov A.A., Haun T., Steyer K., Nowak C., Hailer F. 2013. A range-wide synthesis and timeline for phylogeographic events in the red fox (Vulpes vulpes). BMC Evol. Biol. 13.

Kutzbach J.E., Guan J., He F., Cohen A.S., Orland I.J., Chen G. 2020. African climate response to orbital and glacial forcing in 140,000-y simulation with implications for early modern human environments. Proc. Natl. Acad. Sci. U. S. A. 117:2255–2264.

Larivière S., Seddon P.J. 2001. Vulpes rueppelli. Mamm. Species. 678:1–5.

Leclère D., Obersteiner M., Barrett M., Butchart S.H.M., Chaudhary A., De Palma A., DeClerck F.A.J., Di Marco M., Doelman J.C., Dürauer M., Freeman R., Harfoot M., Hasegawa T., Hellweg S., Hilbers J.P., Hill S.L.L., Humpenöder F., Jennings N., Krisztin T., Mace G.M., Ohashi H., Popp A., Purvis A., Schipper A.M., Tabeau A., Valin H., van Meijl H., van Zeist W.J., Visconti P., Alkemade R., Almond R., Bunting G., Burgess N.D., Cornell S.E., Di Fulvio F., Ferrier S., Fritz S., Fujimori S., Grooten M., Harwood T., Havlík P., Herrero M., Hoskins A.J., Jung M., Kram T., Lotze-Campen H., Matsui T., Meyer C., Nel D., Newbold T., Schmidt-Traub G., Stehfest E., Strassburg B.B.N., van Vuuren D.P., Ware C., Watson J.E.M., Wu W., Young L. 2020. Bending the curve of terrestrial biodiversity needs an integrated strategy. Nature. 585:551–556.

Leite J.V., Álvares F., Velo-Antón G., Brito J.C., Godinho R. 2015. 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.

Li Y., Wu D.D., Boyko A.R., Wang G.D., Wu S.F., Irwin D.M., Zhang Y.P. 2014. Population variation revealed high-altitude adaptation of tibetan mastiffs. Mol. Biol. Evol. 31:1200–1205.

Lindblad-Toh K., Wade C.M., Mikkelsen T.S., Karlsson E.K., Jaffe D.B., Kamal M., Clamp M., Chang J.L., Kulbokas E.J., Zody M.C., Mauceli E., Xie X., Breen M., Wayne R.K., Ostrander E.A., Ponting C.P., Galibert F., Smith D.R., DeJong P.J., Kirkness E., Alvarez P., Biagi T., Brockman W., Butler J., Chin C.W., Cook A., Cuff J., Daly M.J., DeCaprio D., Gnerre S., Grabherr M., Kellis M., Kleber M., Bardeleben C., Goodstadt L., Heger A., Hitte C., Kim L., Koepfli K.P., Parker H.G., Pollinger J.P., Searle S.M.J., Sutter N.B., Thomas R., Webber C., Lander E.S. 2005. Genome sequence,

comparative analysis and haplotype structure of the domestic dog. Nature. 438:803–819.

Linnen C.R., Hoekstra H.E. 2009. Measuring natural selection on genotypes and phenotypes in the wild. Cold Spring Harb. Symp. Quant. Biol. 74:155–168.

Liu S., Lorenzen E.D., Fumagalli M., Li B., Harris K., Xiong Z., Zhou L., Korneliussen T.S., Somel M., Babbitt C., Wray G., Li J., He W., Wang Z., Fu W., Xiang X., Morgan C.C., Doherty A., O'Connell M.J., McInerney J.O., Born E.W., Dalén L., Dietz R., Orlando L., Sonne C., Zhang G., Nielsen R., Willerslev E., Wang J. 2014. Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. Cell. 157:785–794.

Lorenzo F.R., Huff C., Myllymäki M., Olenchock B., Swierczek S., Tashi T., Gordeuk V., Wuren T., Ge R.L., McClain D.A., Khan T.M., Koul P.A., Guchhait P., Salama M.E., Xing J., Semenza G.L., Liberzon E., Wilson A., Simonson T.S., Jorde L.B., Kaelin W.G., Koivunen P., Prchal J.T. 2014. A genetic mechanism for Tibetan high-altitude adaptation. Nat. Genet. 46:951–956.

Lynch V.J., Bedoya-Reina O.C., Ratan A., Sulak M., Drautz-Moses D.I., Perry G.H., Miller W., Schuster S.C. 2015. Elephantid Genomes Reveal the Molecular Bases of Woolly Mammoth Adaptations to the Arctic. Cell Rep. 12:217–228.

Mallon, D., Murdoch, J.D. & Wacher T. 2015. Vulpes rueppellii. IUCN Red List Threat. Species. 8235.

Maloiy G.M.O., Kamau J.M.Z., Shkolnik A., Meir M., Arieli R. 1982. Thermoregulation and metabolism in a small desert carnivore: the Fennec fox (Fennecus zerda) (Mammalia). J. Zool. 198:279–291.

Martin N.H., Bouck A.C., Arnold M.L. 2006. Detecting adaptive trait introgression between Iris fulva and I. brevicaulis in highly selective field conditions. Genetics. 172:2481–2489.

McDevitt A.D., Coscia I., Browett S.S., Ruiz-González A., Statham M.J., Ruczyńska I., Roberts L., Stojak J., Frantz A.C., Norén K., Ågren E.O., Learmount J., Basto M., Fernandes C., Stuart P., Tosh D.G., Sindicic M., Andreanszky T., Isomursu M., Panek M., Korolev A., Okhlopkov I.M., Saveljev A.P., Pokorny B., Flajšman K., Harrison S.W.R., Lobkov V., Ćirović D., Mullins J., Pertoldi C., Randi E., Sacks B.N., Kowalczyk R., Wójcik J.M. 2020. Next-generation phylogeography resolves post-glacial colonization patterns in a widespread carnivore, the red fox (Vulpes vulpes), in Europe. bioRxiv.:2020.02.21.954966.

Miao B., Wang Z., Li Y. 2017. 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.

Miller W., Schuster S.C., Welch A.J., Ratan A., Bedoya-Reinaa O.C., Zhao F., Kim H.L., Burhans R.C., Drautz D.I., Wittekindt N.E., Tomsho L.P., Ibarra-Laclette E., Herrera-Estrella L., Peacock E., Farley S., Sage G.K., Rode K., Obbard M., Montiel R., Bachmann L., Ingólfsson Ó., Aars J., Mailund T., Wiig Ø., Talbot S.L., Lindqvist C. 2012. Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. Proc. Natl. Acad. Sci. U. S. A.

109:E2382–E2390.

Mills L.S., Bragina E. V., Kumar A. V., Zimova M., Lafferty D.J.R., Feltner J., Davis B.M., Hackländer K., Alves P.C., Good J.M., Melo-Ferreira J., Dietz A., Abramov A. V., Lopatina N., Fay K. 2018. Winter color polymorphisms identify global hot spots for evolutionary rescue from climate change. Science (80-.). 359:1033–1036.

Nelson T.C., Cresko W.A. 2018. Ancient genomic variation underlies repeated ecological adaptation in young stickleback populations. Evol. Lett. 2:9–21.

Noll-Banholzer U. 1979. Body temperature, oxygen consumption, evaporative water loss and heart rate in the fennec. Comp. Biochem. Physiol. -- Part A Physiol. 62:585–592.

Nyakatura K., Bininda-Emonds O.R.P. 2012. Updating the evolutionary history of Carnivora (Mammalia): A new species-level supertree complete with divergence time estimates. BMC Biol. 10:12.

Ottenburghs J. 2020. Ghost Introgression: Spooky Gene Flow in the Distant Past. BioEssays. 42:2000012.

Owens G.L., Todesco M., Bercovich N., Légaré J.-S., Mitchell N., Whitney K.D., Rieseberg L.H. 2020. An unexpected donor in the adaptive introgression candidate Helianthus annuus subsp. texanus. bioRxiv.:2020.11.26.398933.

Oziolor E.M., Reid N.M., Yair S., Lee K.M., VerPloeg S.G., Bruns P.C., Shaw J.R., Whitehead A., Matson C.W. 2019. Adaptive introgression enables evolutionary rescue from extreme environmental pollution. Science (80-.). 364:455–457.

Perini F.A., Russo C.A.M., Schrago C.G. 2010. The evolution of South American endemic canids: A history of rapid diversification and morphological parallelism. J. Evol. Biol. 23:311–322.

Qiu J. 2014. Origins of Arctic fox traced back to Tibet. Nature.

Racimo F., Gokhman D., Fumagalli M., Ko A., Hansen T., Moltke I., Albrechtsen A., Carmel L., Huerta-Sánchez E., Nielsen R. 2016. Archaic adaptive introgression in *TBX15/WARS2*. Mol. Biol. Evol. 34:msw283.

Riddell E.A., Iknayan K.J., Hargrove L., Tremor S., Patton J.L., Ramirez R., Wolf B.O., Beissinger S.R. 2021. Exposure to climate change drives stability or collapse of desert mammal and bird communities. Science (80-.). 371:633–638.

Rinker D.C., Specian N.K., Zhao S., Gibbons J.G. 2019. Polar bear evolution is marked by rapid changes in gene copy number in response to dietary shift. Proc. Natl. Acad. Sci. U. S. A. 116:13446–13451.

Sacks B.N., Lounsberry Z.T., Statham M.J. 2018. Nuclear Genetic Analysis of the Red Fox Across its Trans-Pacific Range. J. Hered. 109:573–584.

Sarabia C., vonHoldt B., Larrasoaña J.C., Uríos V., Leonard J.A. 2020. Pleistocene climate fluctuations drove demographic history of African golden wolves (*Canis lupaster*). Mol. Ecol.:mec.15784.

Schweizer R.M., Durvasula A., Smith J., Vohr S.H., Stahler D.R., Galaverni M., Thalmann O., Smith D.W., Randi E., Ostrander E.A., Green R.E., Lohmueller K.E., Novembre J., Wayne R.K. 2018. Natural Selection and Origin of a Melanistic Allele in North American Gray Wolves. Mol. Biol. Evol. 35:1190–1209.

Schweizer R.M., Robinson J., Harrigan R., Silva P., Galverni M., Musiani M., Green R.E., Novembre J., Wayne R.K. 2016. Targeted capture and resequencing of 1040 genes reveal environmentally driven functional variation in grey wolves. Mol. Ecol. 25:357–379.

Sillero-Zubiri C., Hoffmann M., Macdonald D.W. 2004. Canids: Foxes, Wolves, Jackals and Dogs. Status survey and conservation action plan. .

Simonson T.S., Huff C.D., Witherspoon D.J., Prchal J.T., Jorde L.B. 2015. Adaptive genetic changes related to haemoglobin concentration in native high-altitude Tibetans. Exp. Physiol. 100:1263–1268.

Sinding M.H.S., Gopalakrishnan S., Ramos-Madrigal J., Manuel M. De, Pitulko V. V., Kuderna L., Feuerborn T.R., Frantz L.A.F., Vieira F.G., Niemann J., Castruita J.A.S., Carøe C., Andersen-Ranberg E.U., Jordan P.D., Pavlova E.Y., Nikolskiy P.A., Kasparov A.K., Ivanova V. V., Willerslev E., Skoglund P., Fredholm M., Wennerberg S.E., Heide-Jørgensen M.P., Dietz R., Sonne C., Meldgaard M., Dalén L., Larson G., Petersen B., Sicheritz-Pontén T., Bachmann L., Wiig Ø., Marques-Bonet T., Hansen A.J., Gilbert M.T.P. 2020. Arctic-adapted dogs emerged at the Pleistocene-Holocene transition. Science (80-.). 368:1495–1499.

Skonieczny C., McGee D., Winckler G., Bory A., Bradtmiller L.I., Kinsley C.W., Polissar P.J., De Pol-Holz R., Rossignol L., Malaizé B. 2019. Monsoon-driven Saharan dust variability over the past 240,000 years. Sci. Adv. 5:eaav1887.

Smith S.D., Pennell M.W., Dunn C.W., Edwards S. V. 2020. Phylogenetics is the New Genetics (for Most of Biodiversity). Trends Ecol. Evol. 35:415–425.

Stapley J., Reger J., Feulner P.G.D., Smadja C., Galindo J., Ekblom R., Bennison C., Ball A.D., Beckerman A.P., Slate J. 2010. Adaptation genomics: The next generation. Trends Ecol. Evol. 25:705–712.

Statham M.J., Murdoch J., Janecka J., Aubry K.B., Edwards C.J., Soulsbury C.D., Berry O., Wang Z., Harrison D., Pearch M., Tomsett L., Chupasko J., Sacks B.N. 2014. Range-wide multilocus phylogeography of the red fox reveals ancient continental divergence, minimal genomic exchange and distinct demographic histories. Mol. Ecol. 23:4813–4830.

Stewart J.R., Lister A.M., Barnes I., Dalén L. 2010. Refugia revisited: individualistic responses of

species in space and time. Proc. R. Soc. B Biol. Sci. 277:661–671.

Suarez-Gonzalez A., Lexer C., Cronk Q.C.B. 2018. Adaptive introgression: a plant perspective. Biol. Lett. 14:20170688.

Tierney J.E., DeMenocal P.B. 2013. Abrupt shifts in Horn of Africa hydroclimate since the last glacial maximum. Science (80-.). 342:843–846.

Todesco M., Owens G.L., Bercovich N., Légaré J.S., Soudi S., Burge D.O., Huang K., Ostevik K.L., Drummond E.B.M., Imerovski I., Lande K., Pascual-Robles M.A., Nanavati M., Jahani M., Cheung W., Staton S.E., Muños S., Nielsen R., Donovan L.A., Burke J.M., Yeaman S., Rieseberg L.H. 2020. Massive haplotypes underlie ecotypic differentiation in sunflowers. Nature. 584:602–607.

Trabucco A., Zomer R.. 2009. Global Aridity Index (Global-Aridity) and Global Potential Evapo-Transpiration (Global-PET) Geospatial Database. CGIAR Consort. Spat. Inf.

Trauth M.H., Larrasoaña J.C., Mudelsee M. 2009. Trends, rhythms and events in Plio-Pleistocene African climate. Quat. Sci. Rev. 28:399–411.

Tseng Z.J., Li Q., Wang X. 2013. A new cursorial hyena from Tibet, and analysis of biostratigraphy, paleozoogeography, and dental morphology of Chasmaporthetes (Mammalia, Carnivora). J. Vertebr. Paleontol. 33:1457–1471.

Tseng Z.J., Wang X., Slater G.J., Takeuchi G.T., Li Q., Liu J., Xie G. 2014. Himalayan fossils of the oldest known pantherine establish ancient origin of big cats. Proc. R. Soc. B Biol. Sci. 281:20132686.

Vale C.G., Brito J.C. 2015. Desert-adapted species are vulnerable to climate change: Insights from the warmest region on Earth. Glob. Ecol. Conserv. 4:369–379.

Wang G.-D., Fan R.-X., Zhai W., Liu F., Wang L., Zhong L., Wu H., Yang H.-C., Wu S.-F., Zhu C.-L., Li Y., Gao Y., Ge R.-L., Wu C.-I., Zhang Y.-P. 2014a. Genetic Convergence in the Adaptation of Dogs and Humans to the High-Altitude Environment of the Tibetan Plateau. Genome Biol. Evol. 6:2122–2128.

Wang M.S., Wang S., Li Y., Jhala Y., Thakur M., Otecko N.O., Si J.F., Chen H.M., Shapiro B., Nielsen R., Zhang Y.P., Wu D.D. 2020a. Ancient hybridization with an unknown population facilitated high-altitude adaptation of canids. Mol. Biol. Evol. 37:2616–2629.

Wang X., Li Q., Takeuchi G.T. 2016. Out of Tibet: an early sheep from the Pliocene of Tibet, Protovis himalayensis, genus and species nov. (Bovidae, Caprini), and origin of Ice Age mountain sheep. J. Vertebr. Paleontol. 36:e1169190.

Wang X., Li Q., Xie G. 2015. Earliest record of Sinicuon in Zanda Basin, southern Tibet and implications for hypercarnivores in cold environments. Quat. Int. 355:3–10.

Wang X., Liang D., Jin W., Tang M., Shalayiwu, Liu S., Zhang P. 2020b. Out of Tibet: Genomic

Perspectives on the Evolutionary History of Extant Pikas. Mol. Biol. Evol. 37:1577–1592.

Wang X., Tseng Z.J., Li Q., Takeuchi G.T., Xie G. 2014b. From “third pole” to north pole: A Himalayan origin for the arctic fox. Proc. R. Soc. B Biol. Sci. 281.

Wayne R.K., Ostrander E.A. 2007. Lessons learned from the dog genome. Trends Genet. 23:557–567.

Welch A.J., Bedoya-Reina O.C., Carretero-Paulet L., Miller W., Rode K.D., Lindqvist C. 2014. Polar bears exhibit genome-wide signatures of bioenergetic adaptation to life in the arctic environment. Genome Biol. Evol. 6:433–450.

Whitney K.D., Broman K.W., Kane N.C., Hovick S.M., Randell R.A., Rieseberg L.H. 2015. Quantitative trait locus mapping identifies candidate alleles involved in adaptive introgression and range expansion in a wild sunflower. Mol. Ecol. 24:2194–2211.

Whitney K.D., Randell R.A., Rieseberg L.H. 2006. Adaptive introgression of herbivore resistance traits in the weedv sunflower Helianthus annuus. Am. Nat. 167:794–807.

Whitney K.D., Randell R.A., Rieseberg L.H. 2010. Adaptive introgression of abiotic tolerance traits in the sunflower Helianthus annuus. New Phytol. 187:230–239.

Williams J.B., Lenain D., Ostrowski S., Tieleman B.I., Seddon P.J. 2002. Energy expenditure and water flux of Rüppell’s foxes in Saudi Arabia. Physiol. Biochem. Zool. 75:479–488.

Williams J.B., Muñoz-Garcia A., Ostrowski S., Tieleman B.I. 2004. A phylogenetic analysis of basal metabolism, total evaporative water loss, and life-history among foxes from desert and mesic regions. J. Comp. Physiol. B Biochem. Syst. Environ. Physiol. 174:29–39.

Wilmer P., Stone G., Johnston I. a. 2005. Environmental physiology of animals. .

Witt K.E., Huerta-Sánchez E. 2019. Convergent evolution in human and domesticate adaptation to high-altitude environments. Philos. Trans. R. Soc. B Biol. Sci. 374.

Yan Z., Martin S.H., Gotzek D., Arsenault S. V., Duchen P., Helleu Q., Riba-Grognuz O., Hunt B.G., Salamin N., Shoemaker D.W., Ross K.G., Keller L. 2020. Evolution of a supergene that regulates a trans-species social polymorphism. Nat. Ecol. Evol. 4:240–249.

Yao T., Thompson L.G., Mosbrugger V., Zhang F., Ma Y., Luo T., Xu B., Yang X., Joswiak D.R., Wang W., Joswiak M.E., Devkota L.P., Tayal S., Jilani R., Fayziev R. 2012. Third Pole Environment (TPE). Environ. Dev. 3:52–64.

Yi X., Liang Y., Huerta-Sanchez E., Jin X., Xi Ping Cuo Z., Pool J.E., Xu X., Jiang H., Vinckenbosch N., Korneliussen T.S., Zheng H., Liu T., He W., Li K., Luo R., Nie X., Wu H., Zhao M., Cao H., Zou J., Shan Y., Li S., Yang Q., Asan, Ni P., Tian G., Xu J., Liu X., Jiang T., Wu R., Zhou G., Tang M., Qin J., Wang T., Feng S., Li G., Huasang, Luosang J., Wang W., Chen F., Wang Y., Zheng X., Li Z., Bianba Z., Yang G., Wang X., Tang S., Gao G., Chen Y., Luo Z., Gusang L., Cao Z., Zhang Q., Ouyang W., Ren X., Liang H., Zheng H., Huang Y., Li J., Bolund L., Kristiansen K., Li Y., Zhang Y.,

Zhang X., Li R., Li S., Yang H., Nielsen R., Wang J., Wang J. 2010. Sequencing of Fifty Human Exomes Reveals Adaptations to High Altitude. Science (80-.). 329:75–78.

Zhang W., Dasmahapatra K.K., Mallet J., Moreira G.R.P., Kronforst M.R. 2016. Genome-wide introgression among distantly related Heliconius butterfly species. Genome Biol. 17:1–15.

Zhang W., Fan Z., Han E., Hou R., Zhang L., Galaverni M., Huang J., Liu H., Silva P., Li P., Pollinger J.P., Du L., Zhang X.Y., Yue B., Wayne R.K., Zhang Z. 2014a. Hypoxia Adaptations in the Grey Wolf (Canis lupus chanco) from Qinghai-Tibet Plateau. PLoS Genet. 10.

Zhang W., Westerman E., Nitzany E., Palmer S., Kronforst M.R. 2017. Tracing the origin and evolution of supergene mimicry in butterflies. Nat. Commun. 8:1–11.

Zhang X., Witt K., Ko A., Yuan K., Xu S., Nielsen R., Huerta-Sanchez E. 2020. The history and evolution of the denisovan-EPAS1 haplotype in tibetans. bioRxiv.:2020.10.01.323113.

Zhang Z., Ramstein G., Schuster M., Li C., Contoux C., Yan Q. 2014b. Aridification of the Sahara desert caused by Tethys Sea shrinkage during the Late Miocene. Nature. 513:401–404.

Zhao C., Zhang H., Liu G., Yang X., Zhang J. 2016. The complete mitochondrial genome of the Tibetan fox (Vulpes ferrilata) and implications for the phylogeny of Canidae. Comptes Rendus - Biol. 339:68–77.

Chapter II.

On genes and phenotypes underlying desert adaptation in mammals

PAPER 1.

Life in Deserts:
The genetic basis of mammalian desert adaptation

Joana L. Rocha[1,2], Raquel Godinho[1,2,3], José C. Brito[1,2] and Rasmus Nielsen[4,5]

[1]CIBIO/InBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, Campus de Vairão, 4485-661 Vairão, Portugal.
[2]Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, 4169-007 Porto, Portugal.
[3]Department of Zoology, University of Johannesburg, PO Box 534, Auckland Park 2006, South Africa.
[4]Department of Integrative Biology and Department of Statistics, University of California Berkeley, Berkeley, CA 94820, USA.
[5]Globe Institute, University of Copenhagen, DK-1165 Copenhagen, Denmark.

Correspondence:
Joana L. Rocha (joanarocha@cibio.up.pt)
Rasmus Nielsen (rasmus_nielsen@berkeley.edu)

Published: April 13th 2021, *Trends in Ecology and Evolution*
DOI: https://doi.org/10.1016/j.tree.2021.03.007

Keywords:
Desert, adaptation, physiology, genomics, mammals

Highlights

The genomics of adaptation to deserts is a rapidly growing research field which provides examples of adaptation at different time scales and on vastly different organisms facing shared challenges.

Mammals inhabiting deserts show remarkable adaptive traits that have evolved repeatedly and independently in different species across the globe, in response to similar selective pressures of extreme temperatures, aridity, and water and food deprivation.

Genomic studies have shown that there are shared patterns of adaptation at the genomic level involving fat-metabolism and insulin-signaling, as well as the arachidonic acid metabolism.

Understanding the mechanisms by which species in the past successfully adapted to the physical and climatic challenges of deserts is important for evaluating the possibility of evolutionary rescue of species currently challenged by increased desertification.

Abstract

Deserts are among the harshest environments on Earth. The multiple ages of different deserts and their global distribution provide a unique opportunity to study repeated adaptation at different time scales. Here, we summarize recent genomic research on the genetic mechanisms underlying desert adaptations in mammals. Several studies on different desert mammals show large overlap in functional classes of genes and pathways, consistent with the complexity and variety of phenotypes associated with desert adaptation to water and food scarcity and extreme temperatures. However, studies of desert adaptation are also challenged by a lack of accurate genotype-phenotype-environment maps. We encourage development of systems that facilitate functional analyses, but also acknowledge the need for more studies on a wider variety of desert mammals.

Deserts – natural laboratories for studies of adaptation

Deserts (see **Glossary**) are the driest environments on the planet and cover at least 33% of the Earth's land surface (Ward 2016). Though mainly characterized by aridity and water scarcity, deserts also experience daily and annual extreme thermal amplitudes, and intense UV radiation (Ward 2016). Deserts have long been seen as natural laboratories for investigating how biological design is challenged by different aspects of the environment, and how organisms have adapted to these challenges (Willmer et al. 2005; Ward 2016). They also offer unique opportunities to study **convergent evolution** at distinct points in time and space given their well-documented geological age and diverse geographic origins (Willmer et al. 2005; Ward 2016).

The main challenges to life in deserts are maintaining body temperature and preserving water (Willmer et al. 2005). This is particularly difficult for species that heavily rely on evaporative water loss mechanisms (e.g. sweating, panting and salivation) for dissipating heat (Willmer et al. 2005; Riddell et al. 2021). Exposure to heat stress combined with water shortage in hot deserts triggers several pathophysiological processes resulting in heat exhaustion, heat stroke, and kidney dysfunction (Glaser et al. 2016; Johnson et al. 2016; Ward 2016). Exposure to extreme low temperatures in cold-arid deserts, or at night/winter in hot deserts, can lead to hypothermia (Willmer et al. 2005). The organismal responses required to survive these extreme conditions for multiple generations are usually outside the range of the **plastic response** of most non-desert species. Thus, it is generally presumed that desert-adapted species likely experienced strong selection acting on various complex **phenotypes** related to metabolism and water-retention (Williams and Tieleman 2005; Willmer et al. 2005; Rymer et al. 2016; Ward 2016; Rocha et al. 2021).

While desert adaptation has been subject to much research in evolutionary ecology**,** there have

been relatively few investigations on the underlying genetic basis of desert adaptive traits. Newly emerging research on desert mammalian case studies, however, is now contributing to our understanding of the role of past climatic processes, such as desertification, in driving **adaptation**. This review focuses on this nascent research field. We review different methods and synthesize current genomic work in mammalian systems. We then explore the limits and future potential of using genomics to address desert adaptation and provide explicit recommendations for future research.

Box1. Endurer-evader-evaporator concept

Aiming to summarize and categorize decades of classic literature on desert adaptations in species occupying distinct ecological niches, Willmer et al. proposed a classification system relative to extreme temperatures and body size (Willmer et al. 2005). Animals that can evade extreme temperatures through behavior and are physiologically adapted to minimize water loss are called “evaders”. Mammalian evaders include small rodents that can avoid overheating by being nocturnal and hiding in burrows during the day, and are less reliant on evaporative cooling (Willmer et al. 2005). Evaders can additionally minimize water loss by concentrating highly hyperosmotic urine with low volume (Rocha et al. 2021) and utilize water derived from fat metabolism (metabolic water) well-beyond the abilities of non-desert counterparts (see Table I for examples) (Willmer et al. 2005; Donald and Pannabecker 2015).

Animals that cannot shelter behaviorally and instead rely on morphology and physiology to minimize water loss and withstand heat are called “endurers” (Table I) (Willmer et al. 2005). Endurers include large ungulates and marsupials able to store heat without rising body temperature (Willmer et al. 2005; Donald and Pannabecker 2015). For example, Dromedary camels (*Camelus dromedarius*) can endure temperatures exceeding 42°C, and water losses greater than 25% of total body weight which are fatal to non-desert mammals (Ali et al. 2019). Some endurers also have lower respiratory water losses (Rymer et al. 2016), and can store water in their rumen, gut or intestines for water-retention (Willmer et al. 2005; Donald and Pannabecker 2015). It has also been proposed that the Arabian Oryx (*Oryx leucocoryx*) use water derived from fat metabolism and this may account for 24% of its overall water needs (Johnson et al. 2016). Endurers may also have enlarged body tissues that can store localized fat (e.g. fat-tail in sheep, *Ovis aries,* and fat-filled humps of camels) presumed to act as energy reserves during starvation (Atti et al. 2004; Rymer et al. 2016).

A third category was additionally proposed for medium-sized animals that are not able to avoid extreme conditions as efficiently as evaders nor withstand heat as efficiently as endurers. These are called

"evaporators" and comprise lagomorphs, some marsupials, and medium-sized carnivores (Table I) (Willmer et al. 2005). Evaporator species have enlarged body extremities (e.g. large ears) to dissipate heat by conduction and use denning and nocturnal behavior to avoid extreme heat. They also have significantly low mass-adjusted non-renal water loss (Willmer et al. 2005; Donald and Pannabecker 2015; Ward 2016) and can minimize renal water loss by concentrating highly hyperosmotic urine significantly above the levels seen in non-desert counterparts (Willmer et al. 2005; Rocha et al. 2021). We note that some authors have criticized this categorization due to overlap between the categories and have proposed other frameworks to synthesize desert mammalian phenotypic adaptations (e.g. Rymer et al. 2016).

Table I. Classic physiological systems allowing mammalian survival in deserts. The phenotypes listed may themselves be adaptive or may be part of a correlated plastic response (see Scholander et al. 1958; Ali et al. 2012, 2019; Fuller et al. 2014; Donald and Pannabecker 2015; Rymer et al. 2016; Kordonowy et al. 2017; Bittner et al. 2021; Rocha et al. 2021 for more examples and Schwimmer and Haim 2009; Fuller et al. 2014; Donald and Pannabecker 2015; Rymer et al. 2016 for detailed reviews).

Environmental stressor	**Upstream Phenotype**	**Downstream Phenotype**
Heat + Water	Decreased thyroxin [En (4,5), Ev (2,5), H]	Low metabolism or energy expenditures [En (1,2,9), Ev (1,2,8), Evap (1), H]
Heat + Water	Low metabolism /energy expenditures	Reduced non-renal water loss [En (1,2,10), Ev (2,8), Evap (1,2), H], Low respiration rate [En (1,2), Ev (6), Evap (1)], Torpor [Ev (1,5)]
Cold	Non-shivering thermogenesis [Ev (1,8)]	Reduced shivering[H]
Heat	Reduced aldosterone [En (1)]	Low cardiac rate and low blood pressure [En (1)]
Water	Increased vasopressin [En (1,8), Ev (1,6,10)] Increased aldosterone [En (1,3,4), Ev (3,4)]	Increased urine osmolality/Increased water reabsorption from the kidney [En (1,2), Ev, Evap (1,4)]/Decreased urine production [Ev (1,5,7)]
Water	Reduced aldosterone [En (1)]	Increased urine sodium excretion [En (1)]
Food/Water	Increased fat storage in body tissues [En (1,7,7*,4)]	Energy reservation during food/water scarcity
Food/Water	Metabolism of fat [En (1,2,4,7), Ev (13,14)]	Enhanced use of metabolic water [En (1,2,7), Ev (1,2,8, 12,13,14), Evap (1,2)]
Food/Water	Higher plasma sodium [En (1)]	Tolerance to high salt diets [En (1,6,7*), Ev (1,4,8,9), Evap (5)]
Food/Water	Decreased insulin secretion without diabetes [Ev (1)]	Adaptive tolerance to dehydration and starvation
Water	Low glomerular filtration rate [En (1,3), Ev (2,5,7)]	Higher levels of plasma creatinine with no apparent kidney damage or renal impairment [En (1,2), Ev (5,7)]
Water	Higher urea [En (1,2), Ev (5,7)], glucose [En (1,7), potassium [Ev (7)], sodium [En (1)], chloride [Ev (5,7)]	Increased plasma osmolality [En (1,2), Ev (1, 5,7,10), Evap (5)]

Superscript codes and numbers designate: *En – Endurers (1*-Dromedary camel, *Camelus dromedarius; 2*-Arabian oryx, *Oryx leucoryx; 3*-Black Bedouin goat, *Capra aegagrus hircus; 4*-Awassi fat-tailed sheep, *Ovis aries; 5*-Marwari sheep*; 6*-Desert bighorn sheep, *O. canadensis nelsoni; 7*-domestic Bactrian camel, *Camelus bactrianus; 7**- wild Bactrian camel, *Camelus ferus; 8*-Ethiopian Somali goat, *Capra aegagrus hircus; 9*-Common wallaroo*, Osphranter robustus; 10*-Gemsbok, *Oryx gazella; 11*-Arabian sand gazelle*, Gazella marica); Ev – Evaders (1-Golden spiny mouse, Acomys russatus; 2*-Merriam's kangaroo rat, *Dipodomys merriami; 3*-Tarabul's gerbil, *Gerbillus tarabuli, 4*-Cairo spiny mouse*, Acomys cahirinus; 5*-Cactus mouse*, Peromyscus eremicus, 6*-Spinifex hopping mouse, *Notomys alexis; 7*-Sonoran desert mice, *Mus musculus; 8*-Mongolian gerbil, *Meriones unguiculatus; 9*-Fawn hopping mouse*, Notomys cervinus; 10*-Degu, Octodon degus*; 11*-Agile kangaroo rat*, Dipodomys agilis; 12*-Desert pocket mouse, *Chaetodipus penicillatus; 13*-Namib dune gerbil, *Gerbillurus tytonis; 14*-Fat sand rat, *Psammomys*

obesus); Evap – Evaporators (1-Fennec fox, *Vulpes zerda; 2*-Rueppell's fox, *Vulpes rueppellii; 3*-Kit fox, *Vulpes macrotis; 4*-Cape hare, *Lepus capensis; 5*-Tammar wallaby, *Macropus eugenii; 6*-Springhare*, Pedetes capensis); H*-Aboriginal Australians, *Homo sapiens.*

Detecting adaptation

The recurrence of diverse complex adaptive phenotypes across diverse desert organisms (see **Box 1** for examples) raises a fundamental interest on their underlying genetic basis. Enabled by increasingly cost-effective sequencing technologies and computational resources for data analysis, researchers may rely on numerous methods that were designed to identify regions associated with selection, the footprints of adaptation, from large genomic datasets (Nielsen 2005; Stern and Nielsen 2019). The choice of the sequencing and analytical approach should be guided by knowledge on the ecological and evolutionary context of the species, and by the possibility of biological sampling infrastructures in remote arid regions.

Population-specific desert adaptation

Methods in **population genomics** are designed to detect population-specific signatures of ongoing or very recent natural selection (see Nielsen 2005; Stern and Nielsen 2019 for detailed reviews). These methods require genome-wide sequencing or high-density SNP sampling to ensure statistical power to detect signatures of selection against a background of neutral variation (Nielsen 2005; Stern and Nielsen 2019). Some methods focus on finding genomic regions with high allele frequency differentiation between one or more populations (F_{ST}-scans) (Yi et al. 2010). These methods are useful when the background genome-wide differentiation between desert and non-desert populations is relatively low (e.g., global $F_{ST} < 0.2$; **Figure 1A)** (Malaspinas et al. 2016; Yang et al. 2016). Such evolutionary context is necessary to be able to detect variants targeted by selection that are still segregating in the desert population (**Figure 1A**). Other methods can be applied to a single desert population and identify genomic regions of reduced diversity, skewed allele frequency distribution, increased haplotype homozygosity (haplotype-based) (Voight et al. 2006; Sabeti et al. 2007), or combinations thereof, all of which could be signatures of **selective sweeps** (Nielsen 2005; Stern and Nielsen 2019). These approaches have been used to detect selection in different desert-dwelling mammals (Kim et al. 2016; Malaspinas et al. 2016; Yang et al. 2016; Ababaikeri et al. 2020; Tigano et al. 2020), and we highlight some specific examples below.

Following domestication around 8,000-9,000 years ago (yra), sheep (*Ovis aries*) rapidly adapted to diverse agroecological zones with extreme conditions (Yang et al. 2016), including the 5.3 million year-old cold-arid Taklamakan desert (NW China) (Sun and Liu 2006). To find signatures of selection in native

sheep of the Taklamakan and sheep from neighboring cold-arid zones, Yang et al (Yang et al. 2016) focused on genomic regions with reduced variability, haplotype-based methods, and F_{ST}-scans comparing them with sheep inhabiting humid environments. These combined approaches allowed the identification of putatively selected regions harboring genes from Renin-Angiotensin-Aldosterone system (RAAS), oxytocin- and arachidonic acid metabolism pathway in Taklamakan sheep (Yang et al. 2016). Similar methods on hot-desert fat-tail sheep breeds and goats (*Capra hircus*) identified many candidate regions spanning genes and/or pathways significantly enriched for fat metabolism, insulin-signaling, and kidney function, oxidative stress and DNA repair, response to heat and UV radiation, body size development and melanogenesis (Kim et al. 2016; Mwacharo et al. 2017).

Species of the genus *Peromyscus* have also been used in case studies in population genomics as examples of rapid environmental/ecological differentiation (Colella et al. 2020; Tigano et al. 2020). Different arid-dwelling species repeatedly evolved during the radiation of the genus in North America. An example is the Cactus mouse (*P. eremicus*) from southern California, which may have adapted to local desert conditions following the recent onset of southwestern hot-deserts approximately 10,000 yra (Pavlik 2008; Tigano et al. 2020), although it is not clear if its immediate ancestor was also desert adapted. Genomic scans for selective sweeps in *P. eremicus* identified hundreds of genetic variants significantly under selection upstream and/or downstream of genes involved in synthesis and degradation of proteins, sensory perception of bitter taste and lipid metabolism (Tigano et al. 2020).

Following the emergence of modern humans at least 100,000 yra, our species rapidly spread inside and out of Africa and came to inhabit some of the most inhospitable environments on Earth, including deserts (Veth et al. 2005; Ilardo and Nielsen 2018). Though life in deserts has clearly been facilitated by cultural changes, and by innovations such as clothing, fire, water/food storage, and hunting techniques (Brooks 2006; Sereno et al. 2008; Drake et al. 2011; Henn et al. 2012; Williams and Williams 2014; Janz et al. 2017; D'Atanasio et al. 2018), some human desert populations also appear to have adapted biologically to their environment (see **Box 2**).

Box 2 Genetic adaptations in humans

Australians

Until European settlement of Australia 200 years ago, Aboriginal people lived a hunter-gatherer lifestyle for at least 10,000-32,000 years (Williams and Williams 2014; Malaspinas et al. 2016). Early anthropological studies showed that desert-dwelling Aboriginal Australians can maintain reduced metabolic rates without shivering when temperatures drop below 0°C at night (Scholander et al. 1958). Compared to people of European descent, they also have lower levels of thyroid hormone thyroxine (T4)

and reduced affinity for T4-binding globulin, suggesting that selection has targeted thyroid-induced metabolism in Aboriginal Australians, resulting in increased adaptive thermal tolerance (Dick and Watson 1981; Mohr et al. 1987; Takeda et al. 1989; Qi et al. 2014).

Malspinas et al. (Malaspinas et al. 2016) used an F_{ST}-based test to find genome-wide signatures of selection in Aboriginal Australians as compared to other human populations and identified variants nearby genes NETO1, related to the thyroid system, and SLC2A12, associated with serum urate levels, among other genes. Further comparisons of desert-dwelling and wet-tropical Aboriginal Australians identified putatively selected variants in the desert group harboring genes involved in several functions of relevance for life in deserts, including regulation of thyroid levels (KCNJ2) and vasodilation, enhanced capillary permeability, diuresis and natriuresis (KNG1). These results raised the hypothesis that putatively selected variants changing expression patterns of thyroid-related genes (KCNJ2/NETO1 and possibly other genes) may underlie Aboriginal Australians' adaptive thermal tolerance (Malaspinas et al. 2016). Another hypothesis was that selected variants in SLC2A12 may underlie Aboriginal Australians' adaptive tolerance to dehydration (Malaspinas et al. 2016).

Native American

The survival of early Native American Andean populations living in the Atacama desert since at least 7,000 yra heavily relied on water bodies naturally contaminated with arsenic (Apata et al. 2017). This selective pressure has left selection signatures in the genomes of modern admixed descendants near variants that may play a role in immune defense against arsenic poisoning (Apata et al. 2017; Vicuña et al. 2019).

Kalahari peoples

To test for evidence of selection in present-day southern African Khoe-San relative to ancient south Africans, Skoglund et al. (Skoglund et al. 2017) estimated allele frequency differentiation for hundreds of functional categories of genes, and found that the most divergent category was *response to radiation* (Skoglund et al. 2017). It has then been suggested that groups that live a hunter-gather lifestyle in arid regions of the Kalahari basin may have genetically adapted to cope with prolonged exposure to sunlight (Skoglund et al. 2017). Haplotype-based scans (Voight et al. 2006; Sabeti et al. 2007) in specific Khoe-San groups additionally identified variants under selection nearby genes involved in fat metabolism and storage (Sugden et al. 2018; Schlebusch et al. 2020) and uptake of xenobiotic compounds (Schuster et

al. 2010; Schlebusch et al. 2020), among other functions. However, these may not necessarily reflect desert adaptation as some of these groups are spread in diverse climatic conditions.

Species-specific desert adaptation

When all living populations of a species descend from a desert-adapted ancestor, population-based methods may not be successful at identifying regions associated with desert adaptation. An alternative is to use methods in **comparative genomics,** which can detect selection specific to desert species by comparing them to divergent, but related, species that are not desert specialists (**Figure 1B**). Some of these methods focus on finding regions with an increased ratio of nonsynonymous-to-synonymous substitutions (dN/dS) in desert species, which can be signatures of positive selection (Yang 2007; Jirimutu et al. 2012; Marra et al. 2012, 2014; MacManes and Eisen 2014; Wu et al. 2014; Kordonowy and MacManes 2016; Chebii et al. 2020). Other studies focus on identifying changes in gene copy number, that may have been the result of natural selection, using models that take into account branch lengths on phylogenetic trees (De Bie et al. 2006; Jirimutu et al. 2012; Wu et al. 2014; Tigano et al. 2020).

Species of the genus *Camelus*, for example, are desert specialists that descend from a desert-dwelling ancestor that diverged from south American counterparts around 16.3 Mya and cattle 55-60 Mya (Wu et al. 2014). Jirimutu et al. (Jirimutu et al. 2012) found that Bactrian camels (*C. bactrianus/ferus*) show differences in copy number of CYP2 genes relative to cattle and other mammals that may contribute to the enhanced production of a vasodilator of renal preglomerular vessels that stimulates water reabsorption at the kidney (Jirimutu et al. 2012). This study further suggests that extra copies of CPY2J2, a gene associated with high-salt diets and regulation of blood pressure (Jirimutu et al. 2012; Wu et al. 2014; Ali et al. 2019), may explain camels' ability to endure large salt ingestion without developing hypertension (Ali et al. 2012, 2019) (**Box 1**). Evidence from dN/dS-based test and gene family evolution revealed complex features of adaptation in both Bactrian and Dromedary camels (*C. dromedarius*), including strong selection in genes from the insulin-signaling pathway, fat and water metabolism, stress responses to heat, ultraviolet radiation and airborne dust (Wu et al. 2014).

Association with desert phenotypes and climatic variables

In the sections above we highlighted approaches that were designed to explicitly target genomic footprints of natural selection (**Figure 1A-B**). For more examples of genomic studies of mammalian desert adaptation, see online Supplemental Information TableS1. In most studies, selection tests have

been combined with gene-function annotations and **gene ontology enrichment (GO) analysis** to make functional connections, which may or may not be related to some of the desert-related phenotypes that have been described over decades of eco-physiological research (**Box 1**, **Figure 1D**). These methods are phenotype-independent, as they do not require simultaneous sampling of both genomic and phenotypic data. They are, therefore, logistically simple to carry out, but the resulting findings are often difficult to relate directly to specific phenotypes.

To directly relate desert-adaptive phenotypes to variants at the genetic level, **genome-wide association studies (GWAS)** can be combined with selection studies to identify genetic variants affecting specific traits of interest (Tam et al. 2019). Similar approaches known as **genotype-environment associations (GEAs)** can also be used to test for correlation between variants putatively under selection in desert populations and specific climatic variables of interest (Coop et al. 2010; Frichot et al. 2013; Lv et al. 2014; Yang et al. 2016). For example, Yang *et al.* (Yang et al. 2016) used GEAs to test for correlation between variants of Taklamakan desert sheep and precipitation, and found significant associations near candidate genes also identified with selection scans (Yang et al. 2016).

Linking genotype to phenotype is a general challenge in selection studies. Comparative-based approaches, in particular, are challenged by the fact that the candidate mutations are not segregating within a population, making further genetic studies difficult (Smith et al. 2020). Population-based approaches, in this sense, have a distinct advantage in that they may identify candidate segregating variants that can be subject for further genetic analyses, including GWAS. However, GWAS can suffer from low power and be logistically challenging (Marigorta et al. 2018; Tam et al. 2019) and may not be feasible in many organisms (Smith et al. 2020). Even when GWAS is possible (e.g. humans and model organisms), it may still be difficult to directly associate variants with specific desert adaptive traits due to **pleiotropic effects.** For example, Aboriginal Australian show signatures of selection in gene KNG1 that is involved in different physiological functions, including blood coagulation, vasodilation, induction of hypotension, natriuresis and diuresis, blood glucose levels, and nociception (Malaspinas et al. 2016).

Gene expression differences and phenotypic plasticity

Transcriptomics can be complementary to both population and comparative genomics to identify genes associated with desert adaptation and, in contrast to GWAS, it only requires sampling of blood or tissues across few individuals (**Figure 1C**). Studies of desert transcriptomics have analyzed species-specific expression differences either across tissues (e.g. kidney versus spleen in banner-tailed kangaroo rat, *Dipodomys spectabilis* (Marra et al. 2012)) or distinct treatments (e.g. water-fed versus water-deprived states in *P. eremicus* (Kordonowy and MacManes 2017; MacManes 2017)). Another approach

(**Figure 1C-I**) has been to analyze pairwise differences in gene expression between wild-caught individuals from desert and non-desert adapted species (e.g. desert *D. spectabili*s, desert Bailey's pocket mouse/*Chaetodipus baileyi* and tropical forest spiny pocket mouse/*Heteromys desmarestianus* (Marra et al. 2014)) or populations (e.g. Patagonian olive mouse, *Abrothrix olivacea,* from forest and steppe (Giorello et al. 2018)). Because these studies do not control for phenotypic plasticity, they are limited in their ability to identify genetically encoded differences between populations, or species. Thus, they may not identify genes underlying adaptation.

Common garden experiments are needed in order to control for the confounding effects of environmental conditions and plastic responses to different treatments in transcriptomics (De Villemereuil et al. 2016). Bittner *et al*. (Bittner et al. 2021) compared desert and non-desert house mice (*Mus musculus*) populations in water-fed and water-deprived conditions (**Figure 1C-II**), while measuring phenotypic variation in the same individuals. These analyses identified a set of co-expressed genes involved in water-retention correlated with higher blood urea and creatinine (**Box 1**). Bittner et al. (Bittner et al. 2021) also demonstrated that expression changes in the plastic response of water-deprived non-desert mice were in the same direction as the genetically controlled difference between non-desert and desert mice. This points to a genetic basis for phenotypic differences in which plasticity precedes adaptation. Furthermore, the fact that desert mice have reduced expression plasticity could suggest that selection has acted to change the **norm of reaction**, hinting towards **genetic assimilation** as a possible evolutionary avenue for desert adaptation (Bittner et al. 2021).

An important consideration is that even in common garden conditions, the observed expression differences may have evolved neutrally and thus may not necessarily be adaptive. More importantly, it is not possible to identify genetic variation underlying expression differences through transcriptomics alone, as causal variants affecting gene expression may lie in distant enhancers or trans-acting factors (Evans 2015).

Nevertheless, expression studies can provide molecular phenotypes associated with adaptation for further studies and may help identify molecular pathways and mechanisms underlying adaptation (De Villemereuil et al. 2016; Bittner et al. 2021). They also provide an intermediate step between genotype and organismal phenotype that provide information about the functional effects of adaptive variants and help link the results of selection studies to phenotypic variation (e.g. BMP2 gene in (Pan et al. 2019)).

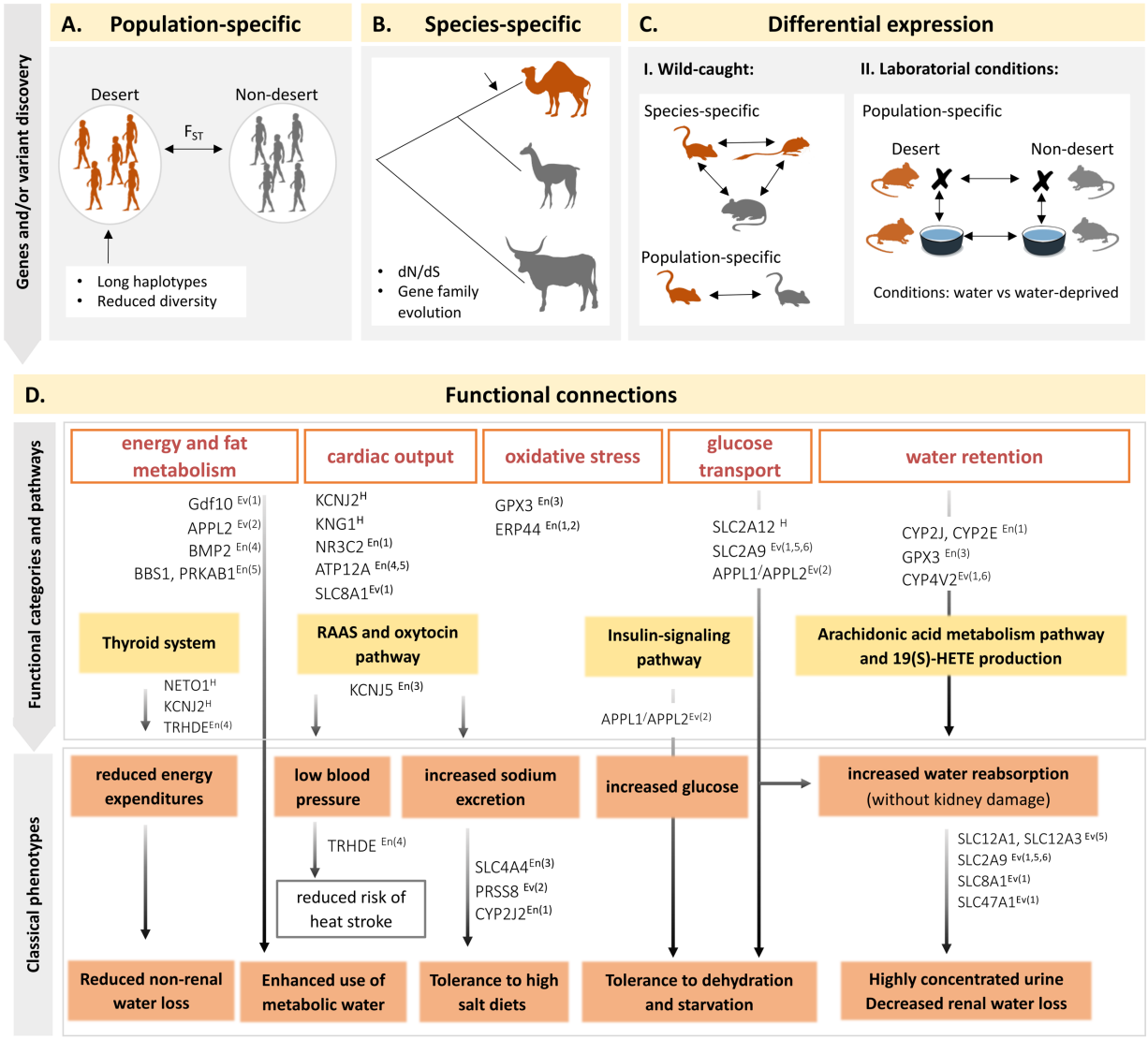


Figure 1. Genomic approaches used to identify genes underlying desert adaptation: A-population-based, B-comparative-based methods, and C-differential expression analyses in natural (I) and laboratory (II) conditions; A-C: different silhouettes represent distinct species, silhouettes in orange represent desert populations/species, and grey silhouettes represent non-desert populations/species under comparison and outgroups; double-headed arrows represent the comparison being made, one-headed arrows represent the group or lineage targeted by the method; X symbol represents a condition of experimental dehydration or water-deprivation, as opposed to animals being given water. D-schematic representation synthesizing and relating some of the recently discovered genes and functional gene classes to previously described phenotypes. Functional connections result from gene ontology enrichment analyses on sets of candidate genes and/or gene-function annotation (Gene Cards, OMIM, UniProt). Some of the highlighted genes identified with population-based and comparative-based methods on whole-genomes and DNA sequences of candidate genes found differentially expressed are represented with superscripts as follows: En – Endurers (1-Bactrian Camel, *Camelus bactrianus*, 2-Dromedary camel, *C. dromedarius*, 3-Taklamakan and 4- Barki sheep, *Ovis aries* 5-Barki goat, *Capra hircus*, 6-Arid-dwelling Asian *sheep*); 6-Ev – Evaders (1- Cactus mouse, *Peromyscus eremicus*, 2- Canyon mouse *P. citrinus*, 3- House mouse, *Mus musculus*, 4-Patagonian olive mouse, *Abrothrix olivacea*, 5-Banner-tailed kangaroo rat, *Dipodomys spectabilis*, 6-Bailey's pocket mouse, *Chaetodipus baileyi*); H – Humans (Aboriginal Australians).

Convergent evolution at the genetic level

Several studies on different desert mammals identified genes, and functional classes of genes and pathways, involved in fat metabolism (Wu et al. 2014; Kim et al. 2016; Sugden et al. 2018; Chebii et al. 2020; Colella et al. 2020), thyroid-induced metabolism (Malaspinas et al. 2016), salt metabolism and prevention of high blood pressure (Jirimutu et al. 2012; Yang et al. 2016; Ababaikeri et al. 2020), insulin-signaling (Wu et al. 2014; Kim et al. 2016; Chebii et al. 2020; Colella et al. 2020), and water-retention (Wu et al. 2014; Kim et al. 2016; Yang et al. 2016; Ababaikeri et al. 2020; Colella et al. 2020) which may provide a genetic basis for previously identified phenotypes. Such phenotypes include energy storage and enhanced use of metabolic water, reduced non-renal water loss and low energy expenditures, ability to cope with high-salt diets without developing hypertension, adaptive tolerance to starvation and dehydration, and decreased renal water loss (**Box 1; Figure 1D**). Some studies additionally identified functional categories of genes involved in stress responses (oxidative stress, DNA damage and repair, heat stress) (Wu et al. 2014; Yang et al. 2016; Tigano et al. 2020), response to radiation (Wu et al. 2014; Skoglund et al. 2017), dust ingestion (Wu et al. 2014), and toxic diets (perception of bitter taste (Tigano et al. 2020) and xenobiotic metabolism (Wu et al. 2014; Kim et al. 2016; Schlebusch et al. 2020)) as targets of natural selection, suggesting other biological adaptations that have not been explicitly target in phenotypic studies.

We synthesize the extent of overlap in gene sets and GO categories between 12 populations/species for which selection candidates have been previously identified, perhaps suggesting convergent evolution at the genetic level in **Figure 2**. For individual genes, significant overlap is mostly found among related species such as *C. bactrianus/dromedarius* and arid-dwelling Asian/Taklamakan *O. aries*, possibly as the result of selection acting on inherited standing genetic variation (**Figure 2A**). Nonetheless, divergent ungulates (e.g. Barki *O. aries*, *Cervus elaphus yarkandensis*/Yarkand deer, and *C. bactrianus*) share significant overlap in BMP2, a gene involved in fat-cell differentiation at several tissues, also associated to the fat-tail phenotype (Moioli et al. 2015; Yuan et al. 2017; Mastrangelo et al. 2019; Pan et al. 2019). This seems to support the role of localized fat as an important source of energy under scarce resources (**Box 1**). Even in divergent species without genome-wide significant overlap, there are some interesting observations. For example, *P. eremicus* shares with *D. spectabilis* and *C. baileyi* evidence of selection in a glucose and serum urate transporter (SLC2A9) (MacManes and Eisen 2014; Marra et al. 2014). Human Aboriginal Australians show signatures of selection at another glucose transporter (SLC2A12), presumably involved in adaptive tolerance to dehydration(Malaspinas et al. 2016).

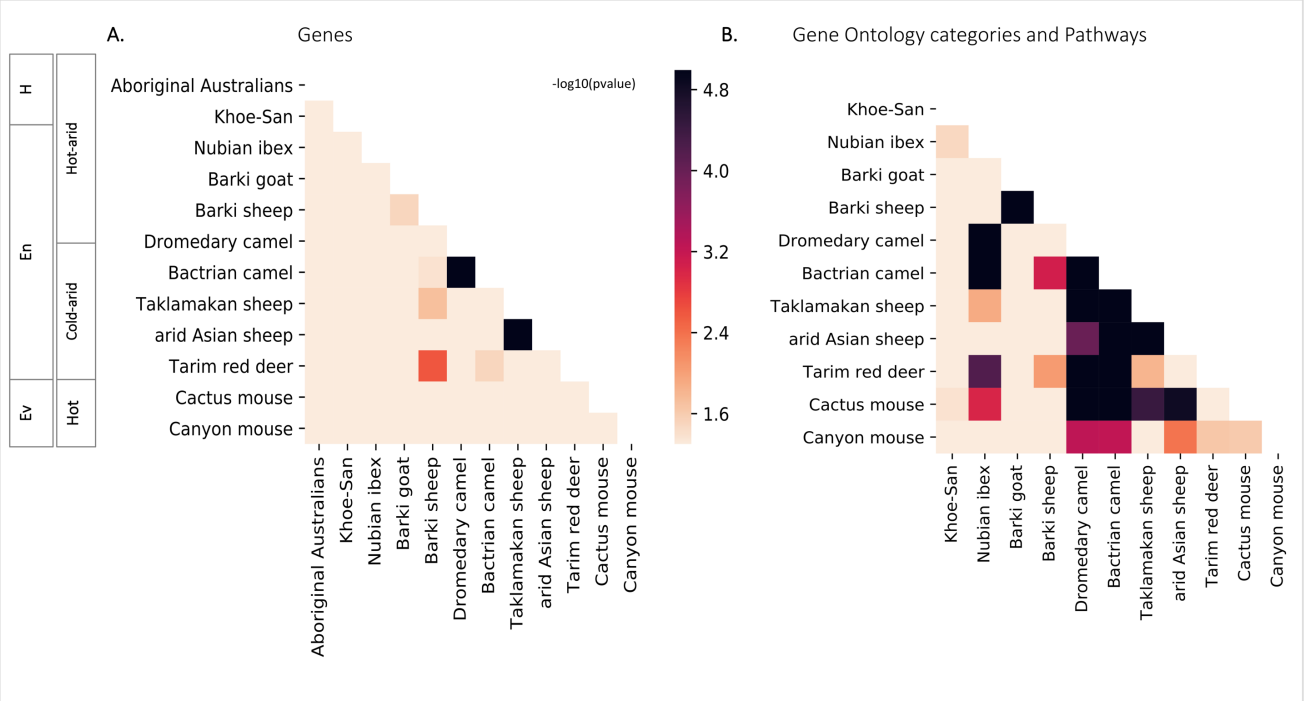


Figure 2. Shared evidence of selection in genes (A) and GO terms and pathways (B) obtained from comparative and population genomics in evader (Ev), endurer (En) and human (H) case studies. Examples reviewed here include: Cactus mouse (*Peromyscus eremicus*) (Tigano et al. 2020), Canyon mouse (*P. citrinus*) (Colella et al. 2020), Tarim red deer (*Cervus elaphus yarkandensis*) (Ababaikeri et al. 2020), arid-dwelling Asian and Taklamakan desert sheep (*Ovis aries*) (Yang et al. 2016), Bactrian camel (*Camelus bactrianus*) (Jirimutu et al. 2012; Wu et al. 2014), Dromedary camel (*C. dromedarius*) (Wu et al. 2014), North-African Barki sheep (*O. aries*) and goat (*Capra hircus*) (Kim et al. 2016), Nubian Ibex (*Capra nubiana*) (Chebii et al. 2020), southern African Khoe-San (Schlebusch et al. 2020) and Aboriginal Australians (Malaspinas et al. 2016) (*Homo sapiens*). Significance of gene and GO overlap was calculated using Fisher's exact test on a contingency table. Significant overlap is represented by colors standing out against the lightest background (marginal p-values > 0.05).

There is also a remarkably high degree of significant overlap in multiple functional categories and pathways between species with very different ages, occupying rather distinct ecological niches, with different strategies of endurance and avoidance, and in different arid regions (classified as hot-young and cold-old deserts (Ward 2016)), consistent with the previously reported high overlap of adaptive phenotypes (**Figure 2B**). The major mechanisms shared across mammals relate to phenotypes that have been associated to adaptation to starvation and dehydration, as well as water retention at the kidney. For instance, the highest overlap is in functional classes of genes associated with fat storage and fat metabolism, including genes that affect energy expenditures, and/or cold- and diet- induced thermogenesis (Wu et al. 2014; Kim et al. 2016; Sugden et al. 2018; Chebii et al. 2020; Colella et al. 2020). Changes in genes involved in fat metabolism may relate to diverse adaptive phenotypes that result in adaptive tolerance to heat/cold (through regulation of energy expenditures and thermogenesis),

dehydration (through enhanced use of metabolic water from fat) and food scarcity (energy stored in localized fat) (Johnson et al. 2016; Rymer et al. 2016) (**Figure 1D**).

In addition to fat metabolism, the strongest evidence of shared selection is in insulin-signaling/response and involves genes such as glucose and serum urate transporters. Some species also share evidence of selection in pathways involved in adipocytokine signaling (implicated in insulin resistance and sensitivity) and Type II diabetes (Wu et al. 2014; Kim et al. 2016), consistent with reported phenotypes of insulin resistance without diabetes (**Box 1**). Insulin resistance plays relevant roles in starvation and reducing energy demands and its evolutionarily well-preserved (Soeters and Soeters 2012), suggesting that it might be beneficial under selective regimes of water/food scarcity. Interestingly, desert-dwelling mammals for which changes in insulin-regulating or glucose transporter genes have been described, are clinically healthy under natural low caloric intake or low energy diets associated with desert life, but particularly prone to obesity and diabetes under induced diet shifts and high caloric foods (Schmidt-Nielsen et al. 1964; O'Dea 1991; Wu et al. 2014; Malaspinas et al. 2016; Hargreaves et al. 2017).

While past investigations of adaptive phenotypes underlying increased water-retention at the kidney have focused on endocrine systems (e.g. vasopressin/RAAS-mediated osmoregulation)(Bridges and James 1982; Schwimmer and Haim 2009; Sands and Layton 2014; Donald and Pannabecker 2015), significant shared evidence of selection across genomic studies (e.g. (Jirimutu et al. 2012; Kim et al. 2016; Yang et al. 2016; Ababaikeri et al. 2020; Tigano et al. 2020)) point to the arachidonic acid metabolism pathway as perhaps the most important adaptive water-retention mechanisms driving convergent evolution in desert mammals, and birds (Tian et al. 2020). Some genes from this pathway putatively under selection (see **Figure 1D**) may be involved in the production of vasodilators of renal preglomerular vessels (19(S)-HETE) that simulate water reabsorption at the kidney. These results hold exciting promise for further investigation of genetic changes underlying desert mammal's outstanding urine concentrating ability (Rocha et al. 2021).

Taken together, these studies not only show that adaptive changes in fat metabolism, insulin-response and arachidonic acid metabolism, may be pivotal for mammalian survival in different deserts of the world, but also that water and food deprivation, perhaps more than temperature, are likely the main drivers of convergent evolution in deserts.

Concluding remarks: Outstanding challenges and Future directions

Selection in desert adapted organisms appears to have targeted genes in multiple functional classes and pathways, consistent with the complexity and variety of phenotypes associated with desert

adaptation. This has led some authors to suggest that phenotypic adaptation in deserts may be polygenic (Mwacharo et al. 2017; Skoglund et al. 2017; Tigano et al. 2020). While it *a priori* seems likely that multiple complex desert adaptive phenotypes are affected by many genes, the hypothesis of selection acting on polygenic traits has not been directly tested in desert adapted mammals, as such tests require detailed knowledge of the genetic basis of the trait (Barghi et al. 2020). Unfortunately, this question, and many other questions regarding the genetic architecture underlying desert adaptation (see **Outstanding questions**), are limited by our lack of knowledge of the phenotypic effects of individual mutations. More research is needed on desert systems in which GWAS, other methods for genetic mapping, or functional studies can be used complementarily to selection studies for understanding how natural selection has affected phenotypes.

However, such systems (mostly model desert evaders) represent only a relative narrow set of desert phenotypes and there is also a need for more studies on a diverse set of organisms. We note that research on the genomic basis of desert adaptation in evaporator species is currently missing, despite past evidence of phenotypic adaptations in several promising biological systems (e.g. arid-dwelling Cape hare/*Lepus capensis* (Kronfeld and Shkolnik 1996), and desert-adapted species of the genus *Vulpes* (Williams et al. 2004); see **Box 1** for more examples). Since most desert phenotypes reported for evaporators are shared with other mammals, they may also share similar functional classes of genes and pathways. Future genomic studies in these and other organisms, and detailed phenotyping for sampled individuals, will help further elucidate convergent patterns of desert adaptation.

Combined genomic, transcriptomic, phenotypic, and/or climatic data can provide powerful approaches for linking genetic variation with expression levels, phenotypes and environmental conditions, and the choice of study-system might be guided by the possibility of obtaining such data (see **Box 3**). Decades of classic phenotypic research in diverse mammalian species, as well as previous evidence for **aridity index**, precipitation and temperature measures as predictors of certain desert phenotypes (Lovegrove 2000; Fristoe et al. 2015; Giorello et al. 2018; Rocha et al. 2021), will continue to inform genomic studies of desert adaptation.

Box 3. Functional integration in evolutionary studies

For model evaders that can be captively bred, phenotypic and transcriptomic data collection under common garden experiments (**Figure I-A**) is essential and can be integrated with genomic data from the same individuals in order to investigate causal genomic variants targeted by natural selection (**Figure I-B**). Additionally, in order to understand a wider range of phenotypes expression, data should not be limited to kidneys (e.g. Marra et al. 2014; MacManes 2017; Giorello et al. 2018; Bittner et al. 2021) but

extend to diverse tissues and cell types. In the context of expression studies, it will be particularly interesting to understand how the norm of reaction has evolved and if short-term adaptation proceeds by expanding the range of plastic responses or by a fixed shift in expression levels, and to explicitly test the hypothesis of genetic assimilation driving desert adaptation. Future research should also move beyond experimental dehydration when measuring plasticity and address the relative and combined effects of different environmental cues in driving differential expression of genes and phenotypes (**Figure I-A-C**). It will be particularly exciting if **reciprocal transplants** in natural conditions are possible (**Figure I-A-C**).

Providing a genotype-phenotype map for non-model organisms is challenging if sample sizes necessary for GWAS are unattainable. However, phenotypes expressed at the cellular level can be studied in cell culture and **CRISPR-Cas9** (**Figure I-C**) can be used to test explicit functional hypotheses (Knott and Doudna 2018). We expect such experiments as follow-up on genomic selection scans to become an important tool in evolutionary studies of non-model organisms. Another recent advance that promises to make the challenge of establishing genotype-phenotype maps easier, is the improvement in computational methods for variant effect prediction, including coding regions affecting protein structure (Senior et al. 2020) and noncoding regions affecting gene expression (Zhou and Troyanskaya 2015) and splicing (Jaganathan et al. 2019).

Even without specific genotype-phenotype maps, **phylogenetic comparative methods (PCMs)** are quite promising as a tool for generating new functional hypotheses, particular with respect to patterns of expression-level evolution, and can also be used directly to test hypotheses about phenotypic evolution (Smith et al. 2020). These methods can be used to test for associations between different variables on a phylogeny with multiple species, some of which are desert adapted and some of which are not. For example, correlated changes between gene expression and specific traits or aridity index of a species can be identified (Rohlfs and Nielsen 2015). Similarly, elevated dN/dS ratios in specific genes on lineages associated with switches to desert environments (Muntané et al. 2018; Hu et al. 2019), or with changes in climate (Pease et al. 2016), might suggest selection acting in these genes associated with desert adaptation.

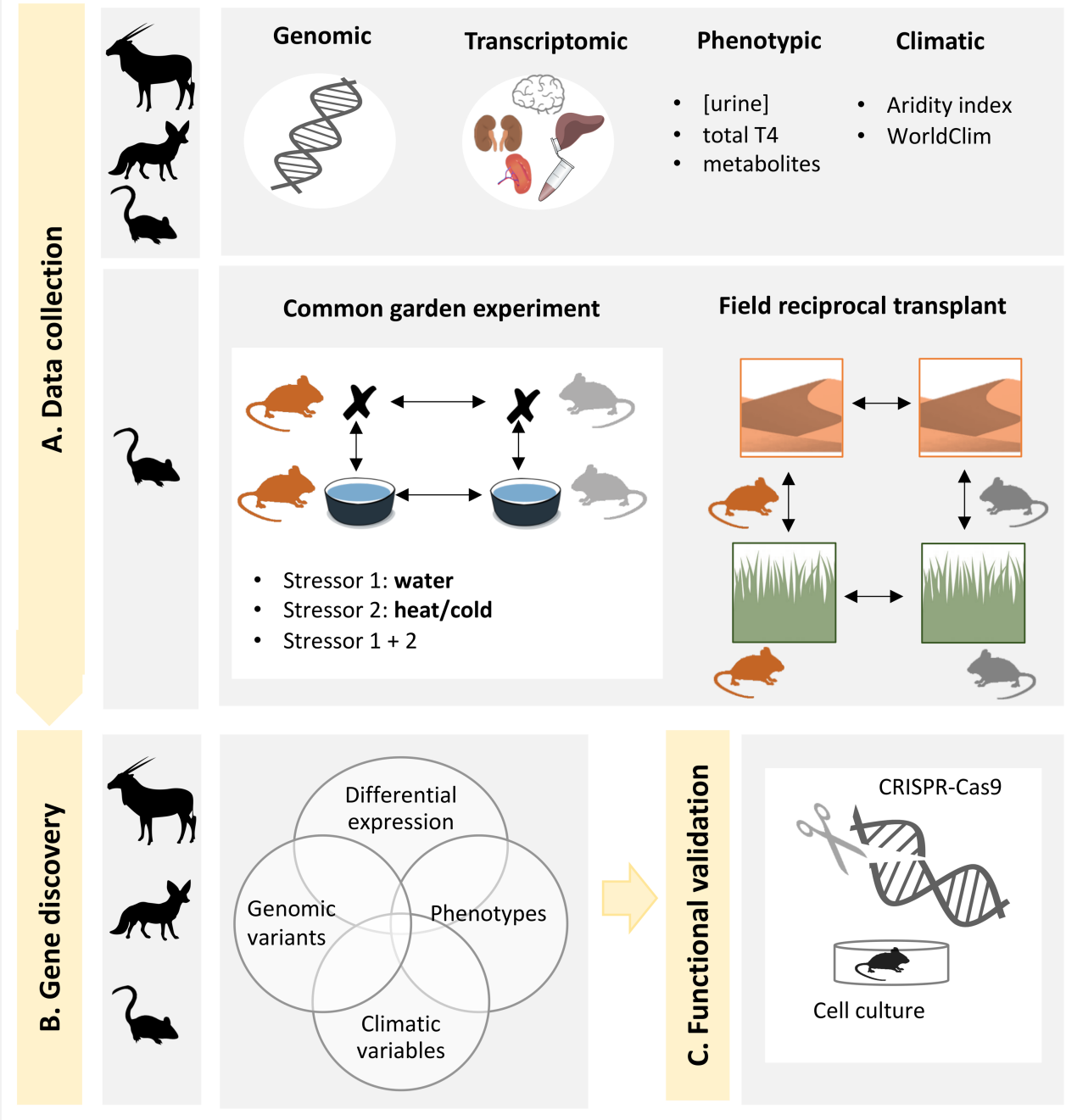


Figure I. The different steps for future research attempting to integrate available methodologies to address the genomic basis of desert adaptation. In part A we suggest simultaneous collection of genomic, transcriptomic (across different tissues and also blood), phenotypic (see Box1) and climatic data (e.g. WorldClim database). Data collection under common garden and field reciprocal transplant experiments are suggested for model organisms (e.g. *Mus* and *Peromyscus* sp). For common garden experiments we suggest different combinations of environmental stressors. Different silhouettes represent distinct species, silhouettes in orange represent desert populations/species, and grey silhouettes represent non-desert populations/species; double-headed arrows represent the comparison being made; X symbol represents a condition of experimental dehydration or water-deprivation, as opposed to animals being given water. In part B we illustrate the diverse ways in

which different datasets can be combined for methods in population and comparative genomics, depending on data availability and the evolutionary context of the study-system of choice. The strength of selection studies is that they are powerful for identifying candidates for further testing. In part C we suggest experimental validation of candidate genes under selection (CRISPR-Cas 9 and cell culture).

Future research should also focus on the tempo and mode of desert adaptation, to take advantage of the fact that deserts have vastly different ages and, therefore, provide an opportunity to study adaptation at different time scales (Ward 2016). The age of deserts varies between just a few thousand years (Pavlik 2008; Tierney and DeMenocal 2013) to millions of years (Van der Wateren and Dunai 2001; Sun and Liu 2006; Zhang et al. 2014) and throughout their lifetime some deserts undergone climate changes marked by alternating hyper-arid and humid periods (Sereno et al. 2008; Drake et al. 2011; Tierney and DeMenocal 2013; D'Atanasio et al. 2018), but little is known about the age of selective events in organisms living in these deserts. Similarly, despite shared evidence of selection across species (**Figure 2**), it remains unclear whether this was the result of similar selective pressures acting independently on new mutations, or on ancestral standing genetic variation, or is the result of **introgression**. Demographic modelling of adaptive variants, or of genomic regions putatively under selection, and methods designed to detect genome-wide introgression will add to our understanding of when and how adaptations can arise, and what are the factors driving adaptation and convergent evolution in deserts.

Knowledge of the genomic basis of desert adaptation will be instrumental for rescuing biodiversity currently threatened by increased desertification (Corlett 2017; Breed et al. 2019; Hoffmann et al. 2020). By establishing the link between phenotype, genotype and the environment it will be possible to predict if/how species can adapt to climate change and which functional traits may be the target of natural selection under increasingly arid conditions. Preserving and maintaining adaptive genetic variation will be an important first step to protect the adaptive potential to cope with ongoing environmental changes (Corlett 2017; Breed et al. 2019; Hoffmann et al. 2020). Gene editing for increased thermal tolerance or adaptive tolerance to dehydration/starvation and other genetic-rescue initiatives (Shapiro 2017; Knott and Doudna 2018) may help accelerate adaptive capacity in arid regions, although such measures remain technically challenging and controversial. Despite the challenges of establishing genotype-phenotype-environment maps, evolutionary studies of desert adaptation hold the promise to not only provide new insights into the process adaptation at multiple time scales, but also to provide a knowledgebase for future work on mitigating the effects of climate change.

Outstanding questions

Do desert adapted traits evolve as a consequence of selection acting on multiple genes of small effect or by few genes of large effects?

Is convergent evolution among different desert adapted species caused by introgression or independent selection? Can different levels of convergence be explained by the Endurer-evader-evaporator concept or by selective regimes associated to different types of deserts?

Does genetic adaptation proceed by expanding or narrowing plastic responses? Is genetic assimilation a key mechanism by which species adapt to deserts?

Is selection in protein coding regions or in non-coding/regulatory areas of the genome most important in desert adaptation? Does the timescale of adaptation affect the answer to this question?

What is the tempo and mode for desert adaptation? How often, how fast, and when can adaptations arise? Does the timescale for adaptation matter for the genomic patterns observed?

Will the continued development of methods in comparative and population genomics deliver on the promise of genetic rescue in species currently challenged by increased desertification?

Can phenotypes and/or genotypes be predicted by climate?

Future studies investigating association of diverse bioclimatic variables with desert adaptive phenotypes and genotypes are needed. Understanding the mechanisms by which species repeatedly adapted to deserts can help elucidate future adaptative responses to increasingly arid conditions.

Glossary

Adaptation (genetic or evolutionary): genetic change at the population or species level resulting in increased fitness in a particular environment. Often thought of as the evolutionary response to a specific selective pressure.

Aridity index (AI): mean annual precipitation divided by mean annual evapo-transpiration.

Common garden experiment: experiment to control for the environment when comparing two species (or lines) by moving them both from their native environments into a common environment. To fully control for environmental effects, the study organism, and possibly their parents, should be reared in the common garden.

Comparative genomics: comparison of genomes of different species (e.g. analyses of mutational differences, or differences in number of genes, gene families, or structural variants).

Convergent evolution: the independent evolution of similar adaptive traits in different species.

CRISPR-Cas9: gene editing technology that can be used for targeted modification of DNA sequences.

Deserts: all warm or cold hyper-arid regions with aridity index below 0.05, and arid regions with aridity index between 0.05 and 0.20, excluding polar regions.

Gene ontology (GO) enrichment analysis: statistical tests identifying enrichment of certain biological functions in a candidate gene set over that expected in a random set of genes.

Genetic assimilation: a reduction of phenotypic plasticity, where selection leads to expression of an ancestrally conditional phenotype in the absence of the original environmental cue.

Genome-wide association studies (GWAS): statistical tests of association between specific genetic variants (typically single-nucleotide polymorphisms) with particular phenotypes, including diseases. These studies rely on large samples of both genomic data and phenotypes.

Genotype: the combination of alleles found in an individual at a particular locus.

Genotype-environment association studies (GEAs): statistical tests of association between specific genetic variants with particular environmental variables in order to identify adaptive genes or genomic regions under environmental selection.

Introgression: gene flow from one species to another.

Norm of reaction: the predicted plastic response to different environmental stimuli.

Phenotype: a trait measurable at the individual level for an organism; may depend on the interaction of different genotypes as well as effects of the environment.

Phylogenetic comparative methods (PCMs): statistical tests of association between variables while accounting for the nonindependence of species due to common ancestry.

Plastic responses (phenotypic plasticity): phenotypic changes (including expression changes) in response to environmental stimuli.

Pleiotropy (pleiotropic): genes or variants that affect multiple traits.

Population genomics: the study and comparison of genetic variability at the genomic level within species or within populations.

Reciprocal transplant experiment: an experiment in which organisms from each of two environments are introduced into the other.

Selective sweeps: genomic footprint of a 'hitchhiking' effect when a selected variant increase in frequency and linked neutral alleles consequently also either increase or decrease in frequency depending on linkage pattern.

Transcriptomics: Genome-wide comparison of abundance of RNA, often to identify genes differentially expressed between individuals from different populations or species, and/or in response to different treatments.

Acknowledgments

J.L.R., J.C.B and R.G. were supported by the Portuguese Foundation for Science and Technology, FCT (SFRH/BD/116397/2016, CEECINST/00014/2018/CP1512/CT0001, and contract under DL57/2016, respectively). RN was supported by NIH grant R01GM138634. This work was partially supported by FCT project PTDC/BIA-EVL/31902/2017

References

Ababaikeri B., Abduriyim S., Tohetahong Y., Mamat T., Ahmat A., Halik M. 2020. Whole-genome sequencing of Tarim red deer (Cervus elaphus yarkandensis) reveals demographic history and adaptations to an arid-desert environment. Front. Zool. 17:31.

Ali A., Baby B., Vijayan R. 2019. From desert to medicine: A review of camel genomics and therapeutic products. Front. Genet. 10:1–20.

Ali M.A., Adem A., Chandranath I.S., Benedict S., Pathan J.Y., Nagelkerke N., Nyberg F., Lewis L.K., Yandle T.G., Nicholls G.M., Frampton C.M., Kazzam E. 2012. Responses to Dehydration in the One-Humped Camel and Effects of Blocking the Renin-Angiotensin System. PLoS One. 7:e37299.

Apata M., Arriaza B., Llop E., Moraga M. 2017. Human adaptation to arsenic in Andean populations of the Atacama Desert. Am. J. Phys. Anthropol. 163:192–199.

Atti N., Bocquier F., Khaldi G. 2004. Performance of the fat-tailed Barbarine sheep in its environment: adaptive capacity to alternation of underfeeding and re-feeding periods. A review. Anim. Res. 53:165–176.

Barghi N., Hermisson J., Schlötterer C. 2020. Polygenic adaptation: a unifying framework to understand positive selection. Nat. Rev. Genet. 21:769–781.

De Bie T., Cristianini N., Demuth J.P., Hahn M.W. 2006. CAFE: a computational tool for the study of gene family evolution. Bioinformatics. 22:1269–1271.

Bittner N.K.J., Mack K.L., Nachman M.W. 2021. Gene expression plasticity and desert adaptation in house mice. Evolution (N. Y).:evo.14172.

Breed M.F., Harrison P.A., Blyth C., Byrne M., Gaget V., Gellie N.J.C., Groom S.V.C., Hodgson R., Mills J.G., Prowse T.A.A., Steane D.A., Mohr J.J. 2019. The potential of genomics for restoring ecosystems and biodiversity. Nat. Rev. Genet. 20:615–628.

Bridges T.E., James N. V. 1982. 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.

Brooks N. 2006. Cultural responses to aridity in the Middle Holocene and increased social complexity. Quat. Int. 151:29–49.

Chebii V.J., Oyola S.O., Kotze A., Domelevo Entfellner J.-B., Musembi Mutuku J., Agaba M. 2020. Genome-Wide Analysis of Nubian Ibex Reveals Candidate Positively Selected Genes That Contribute to Its Adaptation to the Desert Environment. Animals. 10:2181.

Colella J.P., Tigano A., Dudchenko O., Omer A.D., Khan R., Bochkov I.D., Aiden E.L., Macmanes M.D. 2020. Multiple evolutionary pathways to achieve thermal adaptation in small mammals. bioRxiv.:2020.06.29.178392.

Coop G., Witonsky D., Di Rienzo A., Pritchard J.K. 2010. Using Environmental Correlations to Identify Loci Underlying Local Adaptation. Genetics. 185:1411–1423.

Corlett R.T. 2017. A Bigger Toolbox: Biotechnology in Biodiversity Conservation. Trends Biotechnol. 35:55–65.

D'Atanasio E., Trombetta B., Bonito M., Finocchio A., Di Vito G., Seghizzi M., Romano R., Russo G., Paganotti G.M., Watson E., Coppa A., Anagnostou P., Dugoujon J.M., Moral P., Sellitto D., Novelletto A., Cruciani F. 2018. The peopling of the last Green Sahara revealed by high-coverage resequencing of trans-Saharan patrilineages. Genome Biol. 19:20.

Dick M., Watson F. 1981. A possible variant of thyroxine-binding globulin in Australian Aborigines. Clin. Chim. Acta. 116:361–367.

Donald J., Pannabecker T.L. 2015. Osmoregulation in Desert-Adapted Mammals. Sodium and Water Homeostasis. Springer New York. p. 191–211.

Drake N.A., Blench R.M., Armitage S.J., Bristow C.S., White K.H. 2011. Ancient watercourses and biogeography of the Sahara explain the peopling of the desert. Proc. Natl. Acad. Sci. U. S. A. 108:458–462.

Evans T.G. 2015. Considerations for the use of transcriptomics in identifying the "genes that matter" for environmental adaptation. J. Exp. Biol. 218:1925–1935.

Frichot E., Schoville S.D., Bouchard G., François O. 2013. Testing for Associations between Loci and Environmental Gradients Using Latent Factor Mixed Models. Mol. Biol. Evol. 30:1687–1699.

Fristoe T.S., Burger J.R., Balk M.A., Khaliq I., Hof C., Brown J.H. 2015. Metabolic heat production and thermal conductance are mass-independent adaptations to thermal environment in birds and mammals. Proc. Natl. Acad. Sci. U. S. A. 112:15934–15939.

Fuller A., Hetem R.S., Maloney S.K., Mitchell D. 2014. Adaptation to heat and water shortage in large, arid-zone mammals. Physiology. 29:159–167.

Giorello F.M., Feijoo M., D'Elía G., Naya D.E., Valdez L., Opazo J.C., Lessa E.P. 2018. An association between differential expression and genetic divergence in the Patagonian olive mouse (*Abrothrix olivacea*). Mol. Ecol. 27:3274–3286.

Glaser J., Lemery J., Rajagopalan B., Diaz H.F., García-Trabanino R., Taduri G., Madero M., Amarasinghe M., Abraham G., Anutrakulchai S., Jha V., Stenvinkel P., Roncal-Jimenez C., Lanaspa M.A., Correa-Rotter R., Sheikh-Hamad D., Burdmann E.A., Andres-Hernando A., Milagres T., Weiss I., Kanbay M., Wesseling C., Sánchez-Lozada L.G., Johnson R.J. 2016. Climate change and the emergent epidemic of CKD from heat stress in rural communities: The case for heat stress nephropathy. Clin. J. Am. Soc. Nephrol. 11:1472–1483.

Hargreaves A.D., Zhou L., Christensen J., Marlétaz F., Liu S., Li F., Jansen P.G., Spiga E., Hansen M.T., Pedersen S.V.H., Biswas S., Serikawa K., Fox B.A., Taylor W.R., Mulley J.F., Zhang G., Heller R.S., Holland P.W.H. 2017. Genome sequence of a diabetes-prone rodent reveals a mutation hotspot around the ParaHox gene cluster. Proc. Natl. Acad. Sci. U. S. A. 114:7677–7682.

Henn B.M., Botigué L.R., Gravel S., Wang W., Brisbin A., Byrnes J.K., Fadhlaoui-Zid K., Zalloua P.A., Moreno-Estrada A., Bertranpetit J., Bustamante C.D., Comas D. 2012. Genomic Ancestry of North Africans Supports Back-to-Africa Migrations. PLoS Genet. 8:e1002397.

Hoffmann A.A., Miller A.D., Weeks A.R. 2020. Genetic mixing for population management: From genetic rescue to provenancing. Evol. Appl.

Hu Z., Sackton T.B., Edwards S. V., Liu J.S. 2019. Bayesian Detection of Convergent Rate Changes of Conserved Noncoding Elements on Phylogenetic Trees. Mol. Biol. Evol. 36:1086–1100.

Ilardo M., Nielsen R. 2018. Human adaptation to extreme environmental conditions. Curr. Opin. Genet. Dev. 53:77–82.

Jaganathan K., Kyriazopoulou Panagiotopoulou S., McRae J.F., Darbandi S.F., Knowles D., Li Y.I., Kosmicki J.A., Arbelaez J., Cui W., Schwartz G.B., Chow E.D., Kanterakis E., Gao H., Kia A., Batzoglou S., Sanders S.J., Farh K.K.H. 2019. Predicting Splicing from Primary Sequence with Deep Learning. Cell. 176:535-548.e24.

Janz L., Odsuren D., Bukhchuluun D. 2017. Transitions in Palaeoecology and Technology: Hunter-Gatherers and Early Herders in the Gobi Desert. J. World Prehistory. 30:1–80.

Jirimutu, Wang Z., Ding G., Chen G., Sun Y., Sun Z., Zhang H., Wang L., Hasi S., Zhang Y., Li J., Shi Y., Xu Z., He C., Yu S., Li S., Zhang W., Batmunkh M., Ts B., Narenbatu, Unierhu, Bat-Ireedui S., Gao H., Baysgalan B., Li Q., Jia Z., Turigenbayila, Subudenggerile, Narenmanduhu, Wang Z., Wang J., Pan L., Chen Y., Ganerdene Y., Dabxilt, Erdemt, Altansha, Altansukh, Liu T., Cao M., Aruuntsever, Bayart, Hosblig, He F., Zha-Ti A., Zheng G., Qiu F., Sun Z., Zhao L., Zhao W., Liu B.,

Li C., Chen Y., Tang X., Guo C., Liu W., Ming L., Temuulen, Cui A., Li Y., Gao J., Li J., Wurentaodi, Niu S., Sun T., Zhai Z., Zhang M., Chen C., Baldan T., Bayaer T., Li Y., Meng H. 2012. Genome sequences of wild and domestic bactrian camels. Nat. Commun. 3.

Johnson R.J., Stenvinkel P., Jensen T., Lanaspa M.A., Roncal C., Song Z., Bankir L., Sánchez-Lozada L.G. 2016. Metabolic and kidney diseases in the setting of climate change, water shortage, and survival factors. J. Am. Soc. Nephrol. 27:2247–2256.

Kim E.S., Elbeltagy A.R., Aboul-Naga A.M., Rischkowsky B., Sayre B., Mwacharo J.M., Rothschild M.F. 2016. Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. Heredity (Edinb). 116:255–264.

Knott G.J., Doudna J.A. 2018. CRISPR-Cas guides the future of genetic engineering. Science (80-.). 361:866–869.

Kordonowy L., Lombardo K.D., Green H.L., Dawson M.D., Bolton E.A., LaCourse S., MacManes M.D. 2017. Physiological and biochemical changes associated with acute experimental dehydration in the desert adapted mouse, Peromyscus eremicus. Physiol. Rep. 5:1–8.

Kordonowy L., MacManes M. 2017. Characterizing the reproductive transcriptomic correlates of acute dehydration in males in the desert-adapted rodent, Peromyscus eremicus. BMC Genomics. 18:1–20.

Kordonowy L.L., MacManes M.D. 2016. Characterization of a male reproductive transcriptome for Peromyscus eremicus (Cactus mouse). PeerJ. 2016:e2617.

Kronfeld N., Shkolnik A. 1996. Adaptation to Life in the Desert in the Brown Hare (Lepus capensis). J. Mammal. 77:171–178.

Lovegrove B.G. 2000. The zoogeography of mammalian basal metabolic rate. Am. Nat. 156:201–219.

Lv F.-H., Agha S., Kantanen J., Colli L., Stucki S., Kijas J.W., Joost S., Li M.-H., Ajmone Marsan P. 2014. Adaptations to Climate-Mediated Selective Pressures in Sheep. Mol. Biol. Evol. 31:3324–3343.

MacManes M.D. 2017. Severe acute dehydration in a desert rodent elicits a transcriptional response that effectively prevents kidney injury. Am. J. Physiol. - Ren. Physiol. 313:F262–F272.

MacManes M.D., Eisen M.B. 2014. Characterization of the transcriptome, nucleotide sequence polymorphism, and natural selection in the desert adapted mouse Peromyscus eremicus. PeerJ. 1:1–14.

Malaspinas A.S., Westaway M.C., Muller C., Sousa V.C., Lao O., Alves I., Bergström A., Athanasiadis G., Cheng J.Y., Crawford J.E., Heupink T.H., MacHoldt E., Peischl S., Rasmussen S., Schiffels S., Subramanian S., Wright J.L., Albrechtsen A., Barbieri C., Dupanloup I., Eriksson A., Margaryan A.,

Moltke I., Pugach I., Korneliussen T.S., Levkivskyi I.P., Moreno-Mayar J.V., Ni S., Racimo F., Sikora M., Xue Y., Aghakhanian F.A., Brucato N., Brunak S., Campos P.F., Clark W., Ellingvåg S., Fourmile G., Gerbault P., Injie D., Koki G., Leavesley M., Logan B., Lynch A., Matisoo-Smith E.A., McAllister P.J., Mentzer A.J., Metspalu M., Migliano A.B., Murgha L., Phipps M.E., Pomat W., Reynolds D., Ricaut F.X., Siba P., Thomas M.G., Wales T., Wall C.M.R., Oppenheimer S.J., Tyler-Smith C., Durbin R., Dortch J., Manica A., Schierup M.H., Foley R.A., Lahr M.M., Bowern C., Wall J.D., Mailund T., Stoneking M., Nielsen R., Sandhu M.S., Excoffier L., Lambert D.M., Willerslev E. 2016. A genomic history of Aboriginal Australia. Nature. 538:207–214.

Marigorta U.M., Rodríguez J.A., Gibson G., Navarro A. 2018. Replicability and Prediction: Lessons and Challenges from GWAS. Trends Genet. 34:504–517.

Marra N.J., Eo S.H., Hale M.C., Waser P.M., Dewoody J.A. 2012. A priori and a posteriori approaches for finding genes of evolutionary interest in non-model species: Osmoregulatory genes in the kidney transcriptome of the desert rodent Dipodomys spectabilis (banner-tailed kangaroo rat). Comp. Biochem. Physiol. - Part D Genomics Proteomics. 7:328–339.

Marra N.J., Romero A., Dewoody J.A. 2014. Natural selection and the genetic basis of osmoregulation in heteromyid rodents as revealed by RNA-seq. Mol. Ecol. 23:2699–2711.

Mastrangelo S., Moioli B., Ahbara A., Latairish S., Portolano B., Pilla F., Ciani E. 2019. Genome-wide scan of fat-tail sheep identifies signals of selection for fat deposition and adaptation. Anim. Prod. Sci. 59:835–848.

Mohr V.S., Barlow J.W., Topliss D.J., O'Dea K., Stockigt J.R. 1987. Evaluation of T4 and T3 Binding Kinetics in the Thyroxine Binding Globulin Abnormality of Australian Aborigines. Clin. Endocrinol. (Oxf). 26:531–540.

Moioli B., Pilla F., Ciani E. 2015. Signatures of selection identify loci associated with fat tail in sheep. J. Anim. Sci. 93:4660–4669.

Muntané G., Farré X., Rodríguez J.A., Pegueroles C., Hughes D.A., de Magalhães J.P., Gabaldón T., Navarro A. 2018. Biological Processes Modulating Longevity across Primates: A Phylogenetic Genome-Phenome Analysis. Mol. Biol. Evol. 35:1990–2004.

Mwacharo J.M., Kim E.S., Elbeltagy A.R., Aboul-Naga A.M., Rischkowsky B.A., Rothschild M.F. 2017. Genomic footprints of dryland stress adaptation in Egyptian fat-Tail sheep and their divergence from East African and western Asia cohorts. Sci. Rep. 7:1–10.

Nielsen R. 2005. Molecular Signatures of Natural Selection. Annu. Rev. Genet. 39:197–218.

O'Dea K. 1991. Westernisation, insulin resistance and diabetes in Australian Aborigines. Med. J. Aust. 155:258–264.

Pan Z., Li S., Liu Q., Wang Z., Zhou Z., Di R., An X., Miao B., Wang X., Hu W., Guo X., Lv S., Li F., Ding G., Chu M., Li Y. 2019. Rapid evolution of a retro-transposable hotspot of ovine genome underlies the alteration of BMP2 expression and development of fat tails. BMC Genomics. 20:1–15.

Pavlik B.M. 2008. The California deserts: An ecological rediscovery. University of California Press.

Pease J.B., Haak D.C., Hahn M.W., Moyle L.C. 2016. Phylogenomics Reveals Three Sources of Adaptive Variation during a Rapid Radiation. PLoS Biol. 14:e1002379.

Qi X., Chan W.L., Read R.J., Zhou A., Carrell R.W. 2014. Temperature-responsive release of thyroxine and its environmental adaptation in Australians. Proc. R. Soc. B Biol. Sci. 281:0–5.

Riddell E.A., Iknayan K.J., Hargrove L., Tremor S., Patton J.L., Ramirez R., Wolf B.O., Beissinger S.R. 2021. Exposure to climate change drives stability or collapse of desert mammal and bird communities. Science (80-.). 371:633–638.

Rocha J.L., Brito J.C., Nielsen R., Godinho R. 2021. Convergent evolution of increased urine-concentrating ability in desert mammals. Mamm. Rev.:mam.12244.

Rohlfs R. V., Nielsen R. 2015. Phylogenetic ANOVA: The Expression Variance and Evolution Model for Quantitative Trait Evolution. Syst. Biol. 64:695–708.

Rymer T.L., Pillay N., Schradin C. 2016. Resilience to droughts in mammals: A conceptual framework for estimating vulnerability of a single species. Q. Rev. Biol. 91:133–176.

Sabeti P.C., Varilly P., Fry B., Lohmueller J., Hostetter E., Cotsapas C., Xie X., Byrne E.H., McCarroll S.A., Gaudet R., Schaffner S.F., Lander E.S., Frazer K.A., Ballinger D.G., Cox D.R., Hinds D.A., Stuve L.L., Gibbs R.A., Belmont J.W., Boudreau A., Hardenbol P., Leal S.M., Pasternak S., Wheeler D.A., Willis T.D., Yu F., Yang H., Zeng C., Gao Y., Hu H., Hu W., Li C., Lin W., Liu S., Pan H., Tang X., Wang J., Wang W., Yu J., Zhang B., Zhang Q., Zhao H., Zhao H., Zhou J., Gabriel S.B., Barry R., Blumenstiel B., Camargo A., Defelice M., Faggart M., Goyette M., Gupta S., Moore J., Nguyen H., Onofrio R.C., Parkin M., Roy J., Stahl E., Winchester E., Ziaugra L., Altshuler D., Shen Y., Yao Z., Huang W., Chu X., He Y., Jin L., Liu Y., Shen Y., Sun W., Wang H., Wang Y., Wang Y., Xiong X., Xu L., Waye M.M.Y., Tsui S.K.W., Xue H., Wong J.T.F., Galver L.M., Fan J.B., Gunderson K., Murray S.S., Oliphant A.R., Chee M.S., Montpetit A., Chagnon F., Ferretti V., Leboeuf M., Olivier J.F., Phillips M.S., Roumy S., Sallée C., Verner A., Hudson T.J., Kwok P.Y., Cai D., Koboldt D.C., Miller R.D., Pawlikowska L., Taillon-Miller P., Xiao M., Tsui L.C., Mak W., You Q.S., Tam P.K.H., Nakamura Y., Kawaguchi T., Kitamoto T., Morizono T., Nagashima A., Ohnishi Y., Sekine A., Tanaka T., Tsunoda T., Deloukas P., Bird C.P., Delgado M., Dermitzakis E.T., Gwilliam R., Hunt S., Morrison J., Powell D., Stranger B.E., Whittaker P., Bentley D.R., Daly

M.J., De Bakker P.I.W., Barrett J., Chretien Y.R., Maller J., McCarroll S., Patterson N., Pe'Er I., Price A., Purcell S., Richter D.J., Saxena R., Sham P.C., Stein L.D., Krishnan L., Smith A.V., Tello-Ruiz M.K., Thorisson G.A., Chakravarti A., Chen P.E., Cutler D.J., Kashuk C.S., Lin S., Abecasis G.R., Guan W., Li Y., Munro H.M., Qin Z.S., Thomas D.J., McVean G., Auton A., Bottolo L., Cardin N., Eyheramendy S., Freeman C., Marchini J., Myers S., Spencer C., Stephens M., Donnelly P., Cardon L.R., Clarke G., Evans D.M., Morris A.P., Weir B.S., Johnson T.A., Mullikin J.C., Sherry S.T., Feolo M., Skol A., Zhang H., Matsuda I., Fukushima Y., MacEr D.R., Suda E., Rotimi C.N., Adebamowo C.A., Ajayi I., Aniagwu T., Marshall P.A., Nkwodimmah C., Royal C.D.M., Leppert M.F., Dixon M., Peiffer A., Qiu R., Kent A., Kato K., Niikawa N., Adewole I.F., Knoppers B.M., Foster M.W., Clayton E.W., Watkin J., Muzny D., Nazareth L., Sodergren E., Weinstock G.M., Yakub I., Birren B.W., Wilson R.K., Fulton L.L., Rogers J., Burton J., Carter N.P., Clee C.M., Griffiths M., Jones M.C., McLay K., Plumb R.W., Ross M.T., Sims S.K., Willey D.L., Chen Z., Han H., Kang L., Godbout M., Wallenburg J.C., L'Archevêque P., Bellemare G., Saeki K., Wang H., An D., Fu H., Li Q., Wang Z., Wang R., Holden A.L., Brooks L.D., McEwen J.E., Guyer M.S., Wang V.O., Peterson J.L., Shi M., Spiegel J., Sung L.M., Zacharia L.F., Collins F.S., Kennedy K., Jamieson R., Stewart J. 2007. Genome-wide detection and characterization of positive selection in human populations. Nature. 449:913–918.

Sands J.M., Layton H.E. 2014. Advances in Understanding the Urine-Concentrating Mechanism. Annu. Rev. Physiol. 76:387–409.

Schlebusch C.M., Sjödin P., Breton G., Günther T., Naidoo T., Hollfelder N., Sjöstrand A.E., Xu J., Gattepaille L.M., Vicente M., Scofield D.G., Malmström H., de Jongh M., Lombard M., Soodyall H., Jakobsson M. 2020. Khoe-San Genomes Reveal Unique Variation and Confirm the Deepest Population Divergence in Homo sapiens. Mol. Biol. Evol. 37:2944–2954.

Schmidt-Nielsen K., Haines H.B., Hackel D.B. 1964. Diabetes mellitus in the sand rat induced by standard laboratory diets. Science (80-.). 143:689–690.

Scholander P.F., Hammel H.T., Hart J.S., Lemmusurier D.H., Steen J. 1958. Cold adaptation in Australian aborigines. J. Appl. Physiol. 13:211–218.

Schuster S.C., Miller W., Ratan A., Tomsho L.P., Giardine B., Kasson L.R., Harris R.S., Petersen D.C., Zhao F., Qi J., Alkan C., Kidd J.M., Sun Y., Drautz D.I., Bouffard P., Muzny D.M., Reid J.G., Nazareth L. V., Wang Q., Burhans R., Riemer C., Wittekindt N.E., Moorjani P., Tindall E.A., Danko C.G., Teo W.S., Buboltz A.M., Zhang Z., Ma Q., Oosthuysen A., Steenkamp A.W., Oostuisen H., Venter P., Gajewski J., Zhang Y., Pugh B.F., Makova K.D., Nekrutenko A., Mardis E.R., Patterson N., Pringle T.H., Chiaromonte F., Mullikin J.C., Eichler E.E., Hardison R.C., Gibbs R.A., Harkins

T.T., Hayes V.M. 2010. Complete Khoisan and Bantu genomes from southern Africa. Nature. 463:943–947.

Schwimmer H., Haim A. 2009. Physiological adaptations of small mammals to desert ecosystems. Integr. Zool. 4:357–366.

Senior A.W., Evans R., Jumper J., Kirkpatrick J., Sifre L., Green T., Qin C., Žídek A., Nelson A.W.R., Bridgland A., Penedones H., Petersen S., Simonyan K., Crossan S., Kohli P., Jones D.T., Silver D., Kavukcuoglu K., Hassabis D. 2020. Improved protein structure prediction using potentials from deep learning. Nature. 577:706–710.

Sereno P.C., Garcea E.A.A., Jousse H., Stojanowski C.M., Saliège J.-F., Maga A., Ide O.A., Knudson K.J., Mercuri A.M., Stafford T.W., Kaye T.G., Giraudi C., N'siala I.M., Cocca E., Moots H.M., Dutheil D.B., Stivers J.P. 2008. Lakeside Cemeteries in the Sahara: 5000 Years of Holocene Population and Environmental Change. PLoS One. 3:e2995.

Shapiro B. 2017. Pathways to de-extinction: how close can we get to resurrection of an extinct species? Funct. Ecol. 31:996–1002.

Skoglund P., Thompson J.C., Prendergast M.E., Mittnik A., Sirak K., Hajdinjak M., Salie T., Rohland N., Mallick S., Peltzer A., Heinze A., Olalde I., Ferry M., Harney E., Michel M., Stewardson K., Cerezo-Román J.I., Chiumia C., Crowther A., Gomani-Chindebvu E., Gidna A.O., Grillo K.M., Helenius I.T., Hellenthal G., Helm R., Horton M., López S., Mabulla A.Z.P., Parkington J., Shipton C., Thomas M.G., Tibesasa R., Welling M., Hayes V.M., Kennett D.J., Ramesar R., Meyer M., Pääbo S., Patterson N., Morris A.G., Boivin N., Pinhasi R., Krause J., Reich D. 2017. Reconstructing Prehistoric African Population Structure. Cell. 171:59-71.e21.

Smith S.D., Pennell M.W., Dunn C.W., Edwards S. V. 2020. Phylogenetics is the New Genetics (for Most of Biodiversity). Trends Ecol. Evol. 35:415–425.

Soeters M.R., Soeters P.B. 2012. The evolutionary benefit of insulin resistance. Clin. Nutr. 31:1002–1007.

Stern A.J., Nielsen R. 2019. Detecting Natural Selection. Handbook of Statistical Genomics. Wiley. p. 397–40.

Sugden L.A., Atkinson E.G., Fischer A.P., Rong S., Henn B.M., Ramachandran S. 2018. Localization of adaptive variants in human genomes using averaged one-dependence estimation. Nat. Commun. 9.

Sun J., Liu T. 2006. The age of the Taklimakan Desert. Science (80-.). 312:1621.

Takeda K., Mori Y., Sobieszczyk S., Seo H., Dick M., Watson F., Flink I.L., Seino S., Bell G., Refetoff S. 1989. Sequence of the Variant Thyroxine-binding Globulin of Australian Aborigines. J. Clin. Invest.

83:1344–1348.

Tam V., Patel N., Turcotte M., Bossé Y., Paré G., Meyre D. 2019. Benefits and limitations of genome-wide association studies. Nat. Rev. Genet. 20:467–484.

Tian S., Zhou X., Phuntsok T., Zhao N., Zhang D., Ning C., Li D., Zhao H. 2020. Genomic Analyses Reveal Genetic Adaptations to Tropical Climates in Chickens. iScience. 23:101644.

Tierney J.E., DeMenocal P.B. 2013. Abrupt shifts in Horn of Africa hydroclimate since the last glacial maximum. Science (80-.). 342:843–846.

Tigano A., Colella J.P., MacManes M.D. 2020. Comparative and population genomics approaches reveal the basis of adaptation to deserts in a small rodent. Mol. Ecol. 29:1300–1314.

Veth P., Smith M., Hiscock P. 2005. Desert Peoples. Oxford, UK: Blackwell Publishing Ltd.

Vicuña L., Fernandez M.I., Vial C., Valdebenito P., Chaparro E., Espinoza K., Ziegler A., Bustamante A., Eyheramendy S. 2019. Adaptation to Extreme Environments in an Admixed Human Population from the Atacama Desert. Genome Biol. Evol. 11:2468–2479.

De Villemereuil P., Gaggiotti O.E., Mouterde M., Till-Bottraud I. 2016. Common garden experiments in the genomic era: New perspectives and opportunities. Heredity (Edinb). 116:249–254.

Voight B.F., Kudaravalli S., Wen X., Pritchard J.K. 2006. A Map of Recent Positive Selection in the Human Genome. PLoS Biol. 4:e72.

Ward D. 2016. The biology of deserts. Oxford University Press.

Van der Wateren F.M., Dunai T.J. 2001. Late Neogene passive margin denudation history - Cosmogenic isotope measurements from the central Namib desert. Glob. Planet. Change. 30:271–307.

Williams J.B., Muñoz-Garcia A., Ostrowski S., Tieleman B.I. 2004. A phylogenetic analysis of basal metabolism, total evaporative water loss, and life-history among foxes from desert and mesic regions. J. Comp. Physiol. B Biochem. Syst. Environ. Physiol. 174:29–39.

Williams J.B., Tieleman B.I. 2005. Physiological adaptation in desert birds. Bioscience. 55:416–425.

Williams M., Williams M. 2014. Adaptations to life in deserts. Climate Change in Deserts. Cambridge University Press. p. 37–56.

Willmer P., Stone G., Johnston I.A. 2005. Environmental physiology of animals. Blackwell Pub.

Wu H., Guang X., Al-Fageeh M.B., Cao J., Pan S., Zhou H., Zhang L., Abutarboush M.H., Xing Y., Xie Z., Alshanqeeti A.S., Zhang Y., Yao Q., Al-Shomrani B.M., Zhang D., Li J., Manee M.M., Yang Z., Yang L., Liu Y., Zhang J., Altammami M.A., Wang S., Yu L., Zhang W., Liu S., Ba L., Liu C., Yang X., Meng F., Wang S., Li L., Li E., Li X., Wu K., Zhang S., Wang J., Yin Y., Yang H., Al-Swailem A.M., Wang J. 2014. Camelid genomes reveal evolution and adaptation to desert environments.

Nat. Commun. 5:10.

Yang J., Li W.R., Lv F.H., He S.G., Tian S.L., Peng W.F., Sun Y.W., Zhao Y.X., Tu X.L., Zhang M., Xie X.L., Wang Y.T., Li J.Q., Liu Y.G., Shen Z.Q., Wang F., Liu G.J., Lu H.F., Kantanen J., Han J.L., Li M.H., Liu M.J. 2016. Whole-Genome Sequencing of Native Sheep Provides Insights into Rapid Adaptations to Extreme Environments. Mol. Biol. Evol. 33:2576–2592.

Yang Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol. Biol. Evol. 24:1586–1591.

Yi X., Liang Y., Huerta-Sanchez E., Jin X., Xi Ping Cuo Z., Pool J.E., Xu X., Jiang H., Vinckenbosch N., Korneliussen T.S., Zheng H., Liu T., He W., Li K., Luo R., Nie X., Wu H., Zhao M., Cao H., Zou J., Shan Y., Li S., Yang Q., Asan, Ni P., Tian G., Xu J., Liu X., Jiang T., Wu R., Zhou G., Tang M., Qin J., Wang T., Feng S., Li G., Huasang, Luosang J., Wang W., Chen F., Wang Y., Zheng X., Li Z., Bianba Z., Yang G., Wang X., Tang S., Gao G., Chen Y., Luo Z., Gusang L., Cao Z., Zhang Q., Ouyang W., Ren X., Liang H., Zheng H., Huang Y., Li J., Bolund L., Kristiansen K., Li Y., Zhang Y., Zhang X., Li R., Li S., Yang H., Nielsen R., Wang J., Wang J. 2010. Sequencing of Fifty Human Exomes Reveals Adaptations to High Altitude. Science (80-.). 329:75–78.

Yuan Z., Liu E., Liu Z., Kijas J.W., Zhu C., Hu S., Ma X., Zhang L., Du L., Wang H., Wei C. 2017. Selection signature analysis reveals genes associated with tail type in Chinese indigenous sheep. Anim. Genet. 48:55–66.

Zhang Z., Ramstein G., Schuster M., Li C., Contoux C., Yan Q. 2014. Aridification of the Sahara desert caused by Tethys Sea shrinkage during the Late Miocene. Nature. 513:401–404.

Zhou J., Troyanskaya O.G. 2015. Predicting effects of noncoding variants with deep learning-based sequence model. Nat. Methods. 12:931–934.

Table S1. Recent case studies using genomics and transcriptomics in desert mammals (excluding humans).

Taxa	Category	Data collection	Species-specific	Population-based	Functional connection	Refs
Egyptian Barki sheep *Ovis aries*	Hot desert endurer	Genomic, Phenotypic	NA	F_{ST}, reduced diversity, haplotype-based	GO, KEGG pathway, GWAS	(Elbeltagy et al. 2016; Kim et al. 2016; Mwacharo et al. 2017)
Egyptian Barki goat *Capra hircus*	Hot desert endurer	Genomic, Phenotypic	NA	F_{ST}, reduced diversity, haplotype-based	GO, GWAS	(Elbeltagy et al. 2016; Kim et al. 2016; Mwacharo et al. 2017)
Taklimakan/ Arid Asian *Ovis aries*	Cold desert endurer	Genomic, Climatic	NA	F_{ST}, reduced diversity, haplotype-based	GO, PANTHER pathway, GEAs	(Yang et al. 2016)
Tarim Red deer *Cervus elaphus yarkandensis*	Cold desert endurer	Genomic	NA	F_{ST}, reduced diversity, haplotype-based	GO, KEGG pathway	(Ababaikeri et al. 2020)
Bactrian camel *Camelus ferus/bactrianus*	Cold desert endurer	Genomic	dN/dS; gene family evolution	NA	GO, KEGG pathway	(Jirimutu et al. 2012)
Bactrian camel *Camelus bactrianus*	Cold desert endurer	Genomic, Transcriptomic	dN/dS; gene family evolution;	NA	GO	(Wu et al. 2014)
Dromedary *Camelus dromedarius*	Hot desert endurer	Genomic	dN/dS; gene family evolution	NA	GO	(Wu et al. 2014)
Nubian Ibex *Capra nubiana*	Hot desert endurer	Genomic	dN/dS	NA	GO	(Chebii et al. 2020)
Banner-tailed Kangaroo rat, *Dipodomys spectabilis;* Bailey's pocket mouse, *Chaetodipus baileyi*	Hot desert evader	Transcriptomic	DE analysis; dN/dS on transcripts	NA	GO	(Marra et al. 2014)
Olive mouse *Abrothrix olivacea*	Cold desert evader	Transcriptomic, Phenotypic	NA	DE; F_{ST}, GCN	GO, GCN-phenotype	(Giorello et al. 2018)
House mouse *Mus musculus*	Hot desert evader	Transcriptomic, Phenotypic	NA	DE; GCN	GO, GCN - phenotype	(Bittner et al. 2020)
Cactus mouse *Peromyscus eremicus*	Hot desert evader	Genomic Transcriptomic	Gene family evolution; dN/dS on transcripts	SweepFinder2; DE analysis	GO	(MacManes and Eisen 2014; Kordonowy

						and MacManes 2016, 2017; MacManes 2017; Tigano et al. 2020)
Canyon mouse *Peromyscus citrinus*	Hot desert evader	Genomic	Gene family evolution	SweepFinder2	GO	(Colella et al. 2020)

Abbreviations are as follows: GO – Gene ontology enrichment analysis; GCN – gene clustering co-expression network analyses; dN/dS – methods based on the ratio of non-synonymous to synonymous substitutions; F_{ST} – methods based on extreme allele frequency differentiation; DE – Differential expression; KEGG – Kyoto Encyclopedia of Genes and Genomes.

Supplemental References

SAbabaikeri B., Abduriyim S., Tohetahong Y., Mamat T., Ahmat A., Halik M. 2020. Whole-genome sequencing of Tarim red deer (Cervus elaphus yarkandensis) reveals demographic history and adaptations to an arid-desert environment. Front. Zool. 17:31.

Bittner N.K.J., Mack K.L., Nachman M.W. 2020. Plasticity in gene expression facilitates invasion of the desert environment in house mice. bioRxiv Prepr. 2011:1–9.

Chebii V.J., Oyola S.O., Kotze A., Domelevo Entfellner J.-B., Musembi Mutuku J., Agaba M. 2020. Genome-Wide Analysis of Nubian Ibex Reveals Candidate Positively Selected Genes That Contribute to Its Adaptation to the Desert Environment. Animals. 10:2181.

Colella J.P., Tigano A., Dudchenko O., Omer A.D., Khan R., Bochkov I.D., Aiden E.L., Macmanes M.D. 2020. Multiple evolutionary pathways to achieve thermal adaptation in small mammals. bioRxiv.:2020.06.29.178392.

Elbeltagy A.R., Kim E.-S., Rischkowsky B., Aboul-naga A.M., Mwacharo J.M., Rothschild M.F. 2016. Genome-wide Analysis of Small Ruminant Tolerance to Grazing Stress Under Arid Desert. Iowa State Univ. Anim. Ind. Rep. 13.

Giorello F.M., Feijoo M., D'Elía G., Naya D.E., Valdez L., Opazo J.C., Lessa E.P. 2018. An association between differential expression and genetic divergence in the Patagonian olive mouse (*Abrothrix olivacea*). Mol. Ecol. 27:3274–3286.

Jirimutu, Wang Z., Ding G., Chen G., Sun Y., Sun Z., Zhang H., Wang L., Hasi S., Zhang Y., Li J., Shi Y., Xu Z., He C., Yu S., Li S., Zhang W., Batmunkh M., Ts B., Narenbatu, Unierhu, Bat-Ireedui S.,

Gao H., Baysgalan B., Li Q., Jia Z., Turigenbayila, Subudenggerile, Narenmanduhu, Wang Z., Wang J., Pan L., Chen Y., Ganerdene Y., Dabxilt, Erdemt, Altansha, Altansukh, Liu T., Cao M., Aruuntsever, Bayart, Hosblig, He F., Zha-Ti A., Zheng G., Qiu F., Sun Z., Zhao L., Zhao W., Liu B., Li C., Chen Y., Tang X., Guo C., Liu W., Ming L., Temuulen, Cui A., Li Y., Gao J., Li J., Wurentaodi, Niu S., Sun T., Zhai Z., Zhang M., Chen C., Baldan T., Bayaer T., Li Y., Meng H. 2012. Genome sequences of wild and domestic bactrian camels. Nat. Commun. 3.

Kim E.S., Elbeltagy A.R., Aboul-Naga A.M., Rischkowsky B., Sayre B., Mwacharo J.M., Rothschild M.F. 2016. Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. Heredity (Edinb). 116:255–264.

Kordonowy L., MacManes M. 2017. Characterizing the reproductive transcriptomic correlates of acute dehydration in males in the desert-adapted rodent, Peromyscus eremicus. BMC Genomics. 18:1–20.

Kordonowy L.L., MacManes M.D. 2016. Characterization of a male reproductive transcriptome for Peromyscus eremicus (Cactus mouse). PeerJ. 2016:e2617.

MacManes M.D. 2017. Severe acute dehydration in a desert rodent elicits a transcriptional response that effectively prevents kidney injury. Am. J. Physiol. - Ren. Physiol. 313:F262–F272.

MacManes M.D., Eisen M.B. 2014. Characterization of the transcriptome, nucleotide sequence polymorphism, and natural selection in the desert adapted mouse Peromyscus eremicus. PeerJ. 1:1–14.

Marra N.J., Romero A., Dewoody J.A. 2014. Natural selection and the genetic basis of osmoregulation in heteromyid rodents as revealed by RNA-seq. Mol. Ecol. 23:2699–2711.

Mwacharo J.M., Kim E.S., Elbeltagy A.R., Aboul-Naga A.M., Rischkowsky B.A., Rothschild M.F. 2017. Genomic footprints of dryland stress adaptation in Egyptian fat-Tail sheep and their divergence from East African and western Asia cohorts. Sci. Rep. 7:1–10.

Tigano A., Colella J.P., MacManes M.D. 2020. Comparative and population genomics approaches reveal the basis of adaptation to deserts in a small rodent. Mol. Ecol. 29:1300–1314.

Wu H., Guang X., Al-Fageeh M.B., Cao J., Pan S., Zhou H., Zhang L., Abutarboush M.H., Xing Y., Xie Z., Alshanqeeti A.S., Zhang Y., Yao Q., Al-Shomrani B.M., Zhang D., Li J., Manee M.M., Yang Z., Yang L., Liu Y., Zhang J., Altammami M.A., Wang S., Yu L., Zhang W., Liu S., Ba L., Liu C., Yang X., Meng F., Wang S., Li L., Li E., Li X., Wu K., Zhang S., Wang J., Yin Y., Yang H., Al-Swailem A.M., Wang J. 2014. Camelid genomes reveal evolution and adaptation to desert environments. Nat. Commun. 5:10.

Yang J., Li W.R., Lv F.H., He S.G., Tian S.L., Peng W.F., Sun Y.W., Zhao Y.X., Tu X.L., Zhang M., Xie

X.L., Wang Y.T., Li J.Q., Liu Y.G., Shen Z.Q., Wang F., Liu G.J., Lu H.F., Kantanen J., Han J.L., Li M.H., Liu M.J. 2016. Whole-Genome Sequencing of Native Sheep Provides Insights into Rapid Adaptations to Extreme Environments. Mol. Biol. Evol. 33:2576–2592.

PAPER 2.

Convergent evolution of increased urine-concentrating ability in desert mammals

Authors: Joana L. Rocha[1,2], José C. Brito[1,2], Rasmus Nielsen[3,4] and Raquel Godinho[1,2,5]

[1]CIBIO/InBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, Campus de Vairão, 4485-661 Vairão, Portugal.
[2]Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, 4169-007 Porto, Portugal.
[3]Department of Integrative Biology and Department of Statistics, University of California Berkeley, Berkeley, CA 94820, USA.
[4]Globe Institute, University of Copenhagen, DK-1165 Copenhagen, Denmark.
[5]Department of Zoology, University of Johannesburg, PO Box 534, Auckland Park 2006, South Africa.

Author for correspondence:
Joana L. Rocha, email: joanarocha@cibio.up.pt;
Raquel Godinho, email: rgodinho@cibio.up.pt;

Published: March 1st 2021, *Mammal Review*
DOI: **https://doi.org/10.1111/mam.12244**

Keywords:
adaptation, aridity index, convergent evolution, deserts, hyperosmotic urine, mammals, phylogenetic generalised least-squares (PGLS) models.

Graphical Abstract

Desert mammals are remarkably adapted in their physiology. Among the many physiological mechanisms allowing desert mammalian species to survive in extreme conditions of water scarcity is their ability to produce hyperosmotic urine. We review urine-concentrating ability in 121 mammal species, measured as maximum urine osmolality (mosmol/Kg), and use a phylogenetic comparative approach to explicitly test the hypothesis that desert-dwelling species have evolved increased ability to concentrate urine compared to non-desert species. We also demonstrate that the mean aridity index of a species' geographic range largely predicts its urine-concentrating ability. This suggests that, among the many physical and climatic challenges faced by desert species, aridity has been one of the main selective pressures leading to increasing maximum urine-concentrating ability and driving its repeated evolution in different desert mammalian lineages. Future comprehensive statistical studies investigating correlations of diverse bioclimatic variables with more desert adaptive phenotypes, will provide further elucidation on the extent to which climate is predictive for mammalian phenotypes.

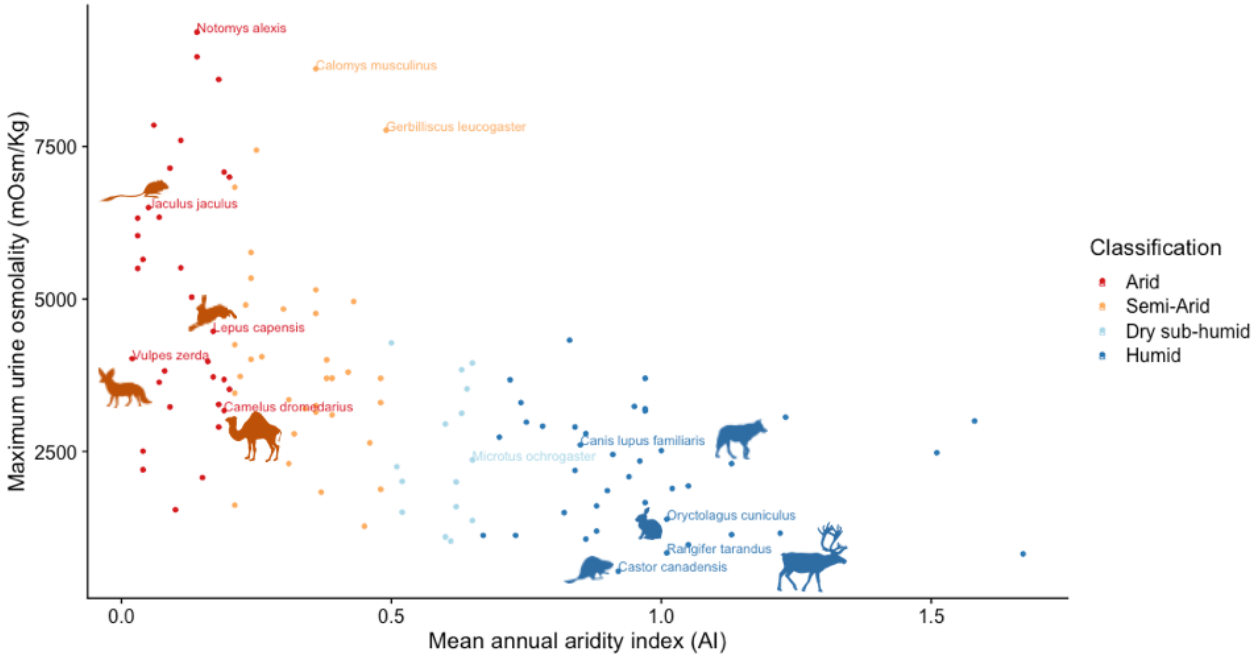


Figure I. Urine concentrating ability in 121 mammal species, measured as maximum urine osmolality (mosmol/Kg), along species distribution range mean annual aridity index.

Abstract

1. One of the most celebrated textbook examples of physiological adaptations to desert environments is the unique ability that desert mammals have to produce hyperosmotic urine. Commonly perceived as an adaptation mainly observed in small rodents, the extent to which urine-concentrating ability has evolved independently in distinct mammalian lineages has not previously been assessed using modern phylogenetic approaches.

2. We review urine-concentrating ability data from the literature in 121 mammalian species with geographic ranges encompassing varying climatic conditions. We explicitly test the general hypothesis that desert-dwelling mammals have evolved greater ability to concentrate urine than non-desert species, controlling for body mass, phylogenetic affinity and other covariates.

3. Ancestral state reconstruction across our dataset's phylogeny shows that the ability to produce hyperosmotic urine, measured as maximum urine osmolality, has evolved convergently in mammalian species with geographic ranges characterised by low mean annual aridity index.

4. Phylogenetic generalised least-squares (PGLS) models show that the mean annual aridity index of a species' geographic range largely predicts its urine-concentrating ability, even when accounting for body mass differences, phylogenetic correlations, the specific condition under which urine osmolality was measured, the method used to measure urine osmolality, and the species' diet.

5. In contrast, we find much weaker correlations between mass-adjusted basal metabolic rate and environmental variables when analysing 84 of the species included in the urine osmolality analysis.

6. Taken together, our results not only show that desert mammals effectively concentrate more urine than non-desert mammals, but further suggest that aridity is likely to have been one of the main selective pressures leading to increasing maximum urine-concentrating ability and driving its repeated evolution in different desert mammalian lineages.

Introduction

Deserts are defined as regions with an aridity index, a ratio of annual precipitation and potential evapo-transpiration, below 0.20 (Ward 2016). Desert species independently evolved striking adaptations to cope with the water scarcity and extreme climatic and physical conditions that characterise such habitats (Nagy 2004; Willmer et al. 2005; Schwimmer and Haim 2009; Pannabecker 2013; Fuller et al. 2014; Rymer et al. 2016). For the last fifty years, this has captured the attention of many eco-physiologists who turned to desert biology to study the physiological mechanisms that help desert animals to maintain their body temperature and retain water. Remarkably, most of the classic works on mammalian desert physiology, pioneered by Knut Schmidt-Nielsen (Schmidt-nielsen et al. 1948; Schmidt-Nielsen et al. 1964; Schmidt-nielsen 1965), have survived the test of time, as researchers have applied new perspectives and tools to test specific hypothesis regarding the evolution of adaptive traits (Walsberg 2000; Tracy and Walsberg 2002).

Among the classical findings of physiological adaptations that minimise water loss is the ability of some desert mammals to produce highly concentrated urine (Sands and Layton 2014). This phenotype is mostly associated with desert 'evaders', such as the Australian spinifex hopping mouse *Notomys alexis*, which holds the record for highest hyperosmotic urine (above 9000 mOsm/kg; Schwimmer & Haim 2009, Donald & Pannabecker 2015). Evaders, part of a classification system proposed by Willmer et al. (2005), are small animals that are able to evade extreme conditions through behaviour (Willmer et al. 2005). By contrast, large mammals unable to shelter from extreme climates that are forced to withstand heat, are called 'endurers', and medium-sized mammals unable to evade or withstand extremes as efficiently as evaders and endurers are called 'evaporators' (Willmer et al. 2005). Mammalian urine-concentrating ability is negatively correlated with body mass (Beuchat 1990), so the small desert evaders stand out in this capacity (Donald and Pannabecker 2015). However, when mass-adjusted, the capacity to produce highly concentrated urine seems to have evolved independently in desert mammals (Beuchat 1990, 1996).

Classical comparative physiology studies have provided evidence for convergent adaptive evolution of ecologically relevant phenotypes in multiple desert mammals. A textbook example includes basal metabolic rate (BMR), which is predicted by latitude and zonal climate, even when accounting for phylogeny and body mass (Lovegrove 2000; Fristoe et al. 2015). Desert mammals evolved lower production of metabolic heat to maintain body temperatures and reduce evaporative water loss (Elgar and Harvey 1987; Lovegrove 2000, 2003; Clarke et al. 2010; Fristoe et al. 2015). Surprisingly, and to the best of our knowledge, no study to date has attempted to identify continuous climatic or environmental variables that help explain variation in urine-concentrating ability, one of the most celebrated examples of mammalian desert adaptation. Moreover, previous studies concerning urine-concentrating ability have

ignored phylogenetic correlations when analysing associations between this trait and the environment (Beuchat 1990, 1996).

We explicitly test the hypothesis that the aridity index of a species' geographic range predicts its maximum urine-concentrating ability, even when accounting for body mass and phylogenetic relationship, and provide statistical evidence that the ability to avoid water loss by producing hyperosmotic urine has evolved independently in multiple phylogenetic lineages of desert mammals. Additionally, we use similar models to re-analyse mass-adjusted BMR for species in which urine osmolality was obtained, and test aridity index and temperature variables as predictors for metabolism in desert mammals.

Methods

Phenotypes were obtained by extensive web-based literature search and by taking advantage of large datasets from previously published revision studies (Beuchat 1990, 1996; Lovegrove 2000; White and Seymour 2003). A total of 108 mammalian species' maximum urine-concentrating abilities came from two previous studies by Beuchat (1990, 1996), in which the author recorded the specific condition (C) under which urine osmolality (mOsm) was measured, including: dehydrated (D), given salt (S) or protein (P) loading, combinations of these conditions, or treatment not specified (X); the method (M) used to measure urine mOsm, as well as the study-species' diet (Di; see Beuchat 1990, 1996 and Appendix S1 for more information). We searched Google Scholar (http://scholar.google.com) for the keywords 'urine osmolality', 'mammals', 'urine-concentrating ability', and recorded the highest mOsm values for 13 additional mammalian species from other studies (Nagy et al. 1976; Maloiy et al. 1979; Downs and Perrin 1991; Kronfeld and Shkolnik 1996; Diaz and Ojeda 1999; Ostrowski et al. 2006). For these studies we also noted the experimental conditions the study-cases were subjected to, the method used to measure urine mOsm, and species' diet type, following Beuchat (1990, 1996). For species with more than one record of urine mOsm, we opted to keep values taken from captive study-subjects under controlled or experimentally induced dehydration, as opposed to measurements of unknown condition taken in the field. Only one measurement was used per species. All species in our dataset were classified according to the endurer-evader-evaporator concept proposed by Willmer et al. 2005. A full detail of the conditions and studies can be found in Appendix S1.

As mOsm is a plastic trait, the experimental conditions under which this parameter is measured can greatly affect estimates, particularly the degree of hydration of study-subjects. Therefore, we curated two datasets: 1) a larger dataset including all 121 reported measurements of mOsm (see Appendix S1); and 2) a subset containing only 87 observations in which the study-subject, in captivity, was dehydrated and not given food or other treatments. However, we note that there is still some residual variation in the number of days under water-deprivation for an animal to be considered fully dehydrated, and variation in

the method used for measuring mOsm. Also, for the 121-species dataset, 24 records came from studies that did not report the procedures used to measure mOsm. Downstream analyses were performed on both data sets.

The geographic range polygons of each species for which physiological data were available were downloaded from the website of the International Union for Conservation of Nature (IUCN 2019, 2020). We also downloaded the following environmental variables at about 1 km spatial resolution: annual aridity index (AI) from the CGIAR-CSI Global Aridity database (Trabucco and Zomer 2009), and annual mean temperature (BIO1), maximum temperature of the warmest month (BIO5), minimum temperature of the coldest month (BIO6), and temperature annual range (BIO7) from the WorldClim database (Fick and Hijmans 2017). We then used a Geographical Information System to calculate mean values for each environmental variable within the geographic range polygon of each species. For species with global geographic ranges, and for which phenotypes under review were sampled from a specific region, we partitioned the original distribution polygon accordingly and calculated mean values for the sampled zone. A full detail of the mean annual aridity index (AI) and mean temperature variables (BIOS) for each species' geographic range can be found in Appendix S1. To avoid the effects of collinearity, we tested for correlations between pairs of variables using a Spearman correlation test before testing them together as predictors for mOsm and BMR in downstream analyses. This test was performed using the 'ggscatter' function implemented in the R package 'ggpubr' with the correlation method option set to 'Spearman'. We used a previously published mammalian phylogenetic tree estimated using both nuclear and mitochondrial genes (Upham et al. 2019), and pruned the tree to retain only the species included in our study. For visualisation of trait evolution on the tree, we projected states for maximum mOsm and temperature variables, estimated using maximum-likelihood methods, onto the internal edges and nodes of the tree using a colour gradient. The estimation was done using the 'fastAnc' option in the 'contMap' function implemented in the R package 'phytools' (Revell 2012).

Phylogenetic generalised least squares (PGLS) as implemented in 'phytools' were performed using the 'gls' function with the model of Pagel (1999) to test different linear models with increasing complexity. Briefly, the model of Pagel adjusts the off-diagonal elements of variance-covariance matrix in a Brownian evolution model with a multiplicative factor, λ, such that when λ = 0 a star phylogeny is obtained representing no phylogenetic signal, while when λ = 1 a standard Brownian motion process on the reference tree is obtained. As residuals of maximum mOsm are not normally distributed (Fig. 1), we performed PGLS analyses on log10(maximum mOsm) using log10(body mass) as covariate. Mean AI and/or mean temperature variables were used in PGLS models as potential predictors of urine-concentrating ability. Condition (C), method (M) and diet (Di) were also included as covariates. For the dataset including only dehydrated individuals, the same PGLS models were tested with AI converted to

a categorical variable using the following binning (defined by Trabucco & Zomer 2009): humid for AI > 0.65, dry sub-humid for 0.5 < AI < 0.65, semi-arid for 0.2 < AI < 0.5, and arid for AI < 0.2. To test for phylogenetic signal for specific traits, we calculated Pagel's λ on urine-concentrating ability, body mass and AI using the function 'phylosig' as implemented in 'phytools'. All statistics were performed in R v. 4.0.1 and RStudio v. 1.3.1073. Detailed code implementing these analyses is available at https://github.com/joanocha/GIS_PGLS_MetaAnalysis.

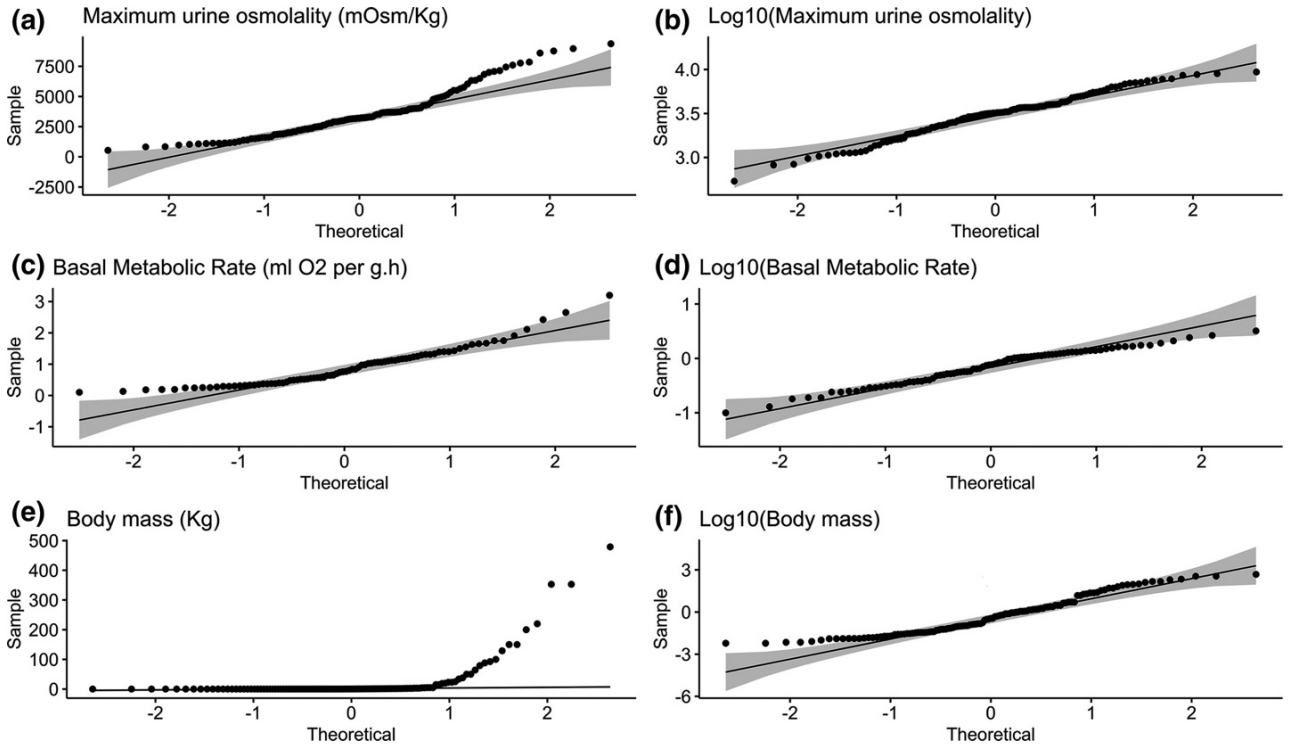


Figure 1. Quantile-quantile plots for the normal distribution of variables (A, C, E) and log10 of variables (B, D, F) used in this study of urine-concentrating ability of desert mammals.

Results

We analysed the dataset with information on maximum urine-concentrating abilities, mean annual aridity index (AI) and WorldClim temperatures (BIOS) for a total of 121 mammal species with available DNA sequence information. Ancestral state reconstruction across our dataset's phylogeny shows that the ability to produce hyperosmotic urine has evolved multiple times in mammalian species with geographic ranges with low mean annual aridity index (Fig. 2).

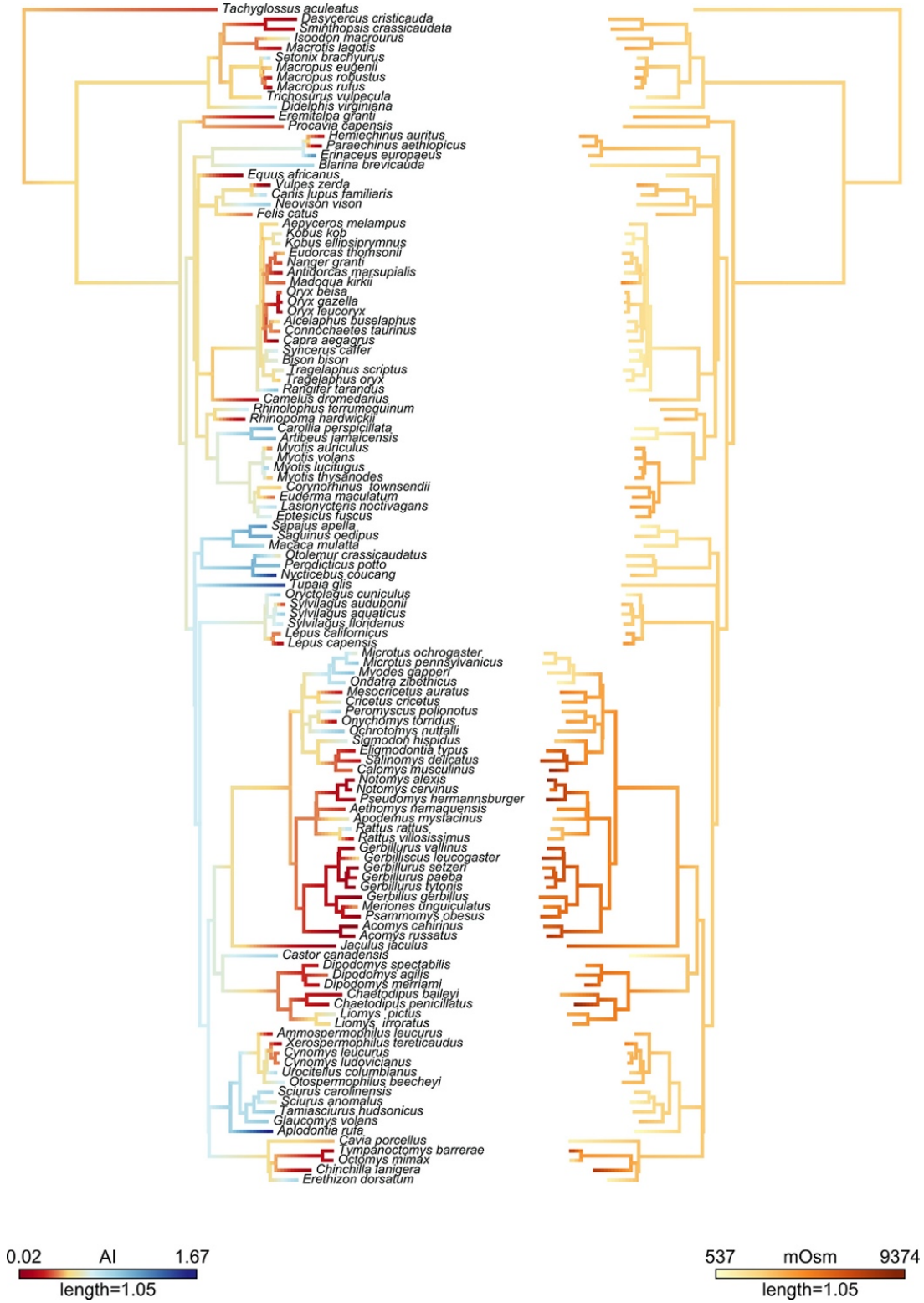


Figure 2. Ancestral state reconstruction for mean annual aridity index (AI, left) and maximum urine osmolalities (mOsm/kg; right) in a phylogenetic tree with 121 mammalian species.

When testing for phylogenetic signal using Pagel's λ, we found significant phylogenetic signal for most variables, validating the relevance of adjusting the linear models for phylogenetic covariance (Table 1). All PGLS models tested in this study point to mean AI as a significant and important predictor of mammalian maximum mOsm (Table 2; see also Appendix S2 for all PGLS models tested). The lower the mean annual aridity index, the more efficiently mammals are able to concentrate urine (Fig. 3), even when accounting for the differences in body mass, diet, experimental condition, methodology and phylogeny (Table 2, Appendix S2). Overall, species inhabiting arid and semi-arid environments can maximise urine concentration well above the levels of their dry sub-humid and humid counterparts (Fig. 3). In congruence with Beuchat (1990), our results also show that log-transformed body mass is also negatively correlated with log-transformed maximum mOsm, though not to the same extent as mean annual aridity index (Table 2, Fig. 4). We also find that urine-concentrating ability is significantly lower in species with a frugivorous diet than in strictly carnivorous species (Table 2). Mean annual aridity index or mean AI correlates with temperature variables, except mean BIO7 (Appendix S3). Overall, no strong correlations were found between urine-concentrating ability and temperature variables (see Appendix S2 for PGLS using temperature variables as covariates, and Appendix S4 for visual illustration of the relationship between mOsm and WorldClim temperatures). When experimental condition and methodology were used as co-variates in the PGLS analysis, their role in predicting maximum mOsm was not significant (Table 2, Appendix S2).

Table 1. Phylogenetic signal (λ) for variables with significant *P*-values. Variables for which λ is close to 1, and with *P*-values < 0.05, have significant phylogenetic signals.

Variable	**Pagel's λ (*P*-value)**
Log10(maximum urine osmolality)	0.988 (3.26e-12)
Log10 (body mass)	1.000 (1.26e-41)
Mean AI (annual aridity index)	0.854 (0.0002)
Mean BIO1 (annual mean temperature)	0.668 (0.0026)
Mean BIO6 (minimum temperature of the coldest month)	0.772 (0.0002)
Mean BIO7 (temperature annual Range)	0.840 (4.90 e-06)

We tested similar PGLS models on the subset of the data that only included dehydrated study-subjects, resulting in a smaller dataset including 87 species (18 arid, 26 semi-arid, 13 dry sub-humid, and 30 humid). The results from this analysis (illustrated in Table 3) are similar to those of the larger dataset conditions (Table 2). When using AI as a discrete variable, we found that mOsm was significantly higher in species from arid environments than in species from semi-arid, dry sub-humid and humid regions

(Table 3). In all analyses, the mean AI of the species' geographic range is a stronger predictor of maximum mOsm than body mass and diet, and the condition of study-subjects and methodology used to measure this trait are not significant predictors (Tables 2 and 3; see also Appendix S2).

Table 2. PGLS models for predicting mammalian log10(maximum urine osmolality). Significant *P*-values are highlighted in bold; AIC = Akaike information criterion, s.e. = standard error.

Model	*AIC*	*Variable*	*Coefficient*	*s.e.*	*t-value*	*P-value*
AI + ***BM*** *(best model)*	-74.4	Intercept	3.632	0.059	61.64	**0.0000**
		AI	-0.373	0.040	-9.20	**0.0000**
		BM	-0.092	0.014	-6.43	**0.0000**
AI + ***BM*** + *C* + *M* +***Di***	-16.3	intercept	3.703	0.097	37.99	**0.0000**
		AI	-0.368	0.043	-8.47	**0.0000**
		BM	-0.082	0.015	-5.54	**0.0000**
		$DF_{condition}$	0.023	0.080	0.28	0.7763
		$DP_{condition}$	-0.081	0.108	-0.75	0.4537
		$DPS_{condition}$	0.020	0.178	0.11	0.9104
		$DS_{condition}$	-0.047	0.106	-0.45	0.6573
		$P_{condition}$	-0.006	0.077	-0.08	0.9341
		$S_{condition}$	0.049	0.065	0.75	0.4536
		$SP_{condition}$	0.148	0.148	1.00	0.3202
		$X_{condition}$	-0.049	0.047	-1.06	0.2911
		A_{method}	-0.068	0.153	-0.44	0.6592
		F_{method}	0.016	0.039	0.41	0.6801
		FV_{method}	0.058	0.097	0.60	0.5530
		V_{method}	0.106	0.055	1.93	0.0570
		CI_{diet}	-0.158	0.141	-1.12	0.2649
		Fr_{diet}	-0.533	0.149	-3.57	**0.0005**
		G_{diet}	0.013	0.103	0.13	0.8980
		H_{diet}	-0.132	0.088	-1.50	0.1366
		I_{diet}	-0.130	0.107	-1.22	0.2260
		O_{diet}	-0.055	0.095	-0.58	0.5634

Model notation refers to the variables used as predictors: mean AI = annual aridity index, BM = log10 body mass (kg); C = condition (fed-F, given salt load-S, protein load-P, unknown condition-X, and combinations of these, all relative to dehydrated-D); M = method (addition of solutes-A, freezing point-F, vapour pressure-V, and combinations of these, all relative to unknown method-?); Di = diet (frugivorous-Fr, granivorous-G, herbivorous-H, insectivorous-I, omnivorous-O, and CI-both carnivorous and insectivorous, all relative to strictly carnivorous-C).

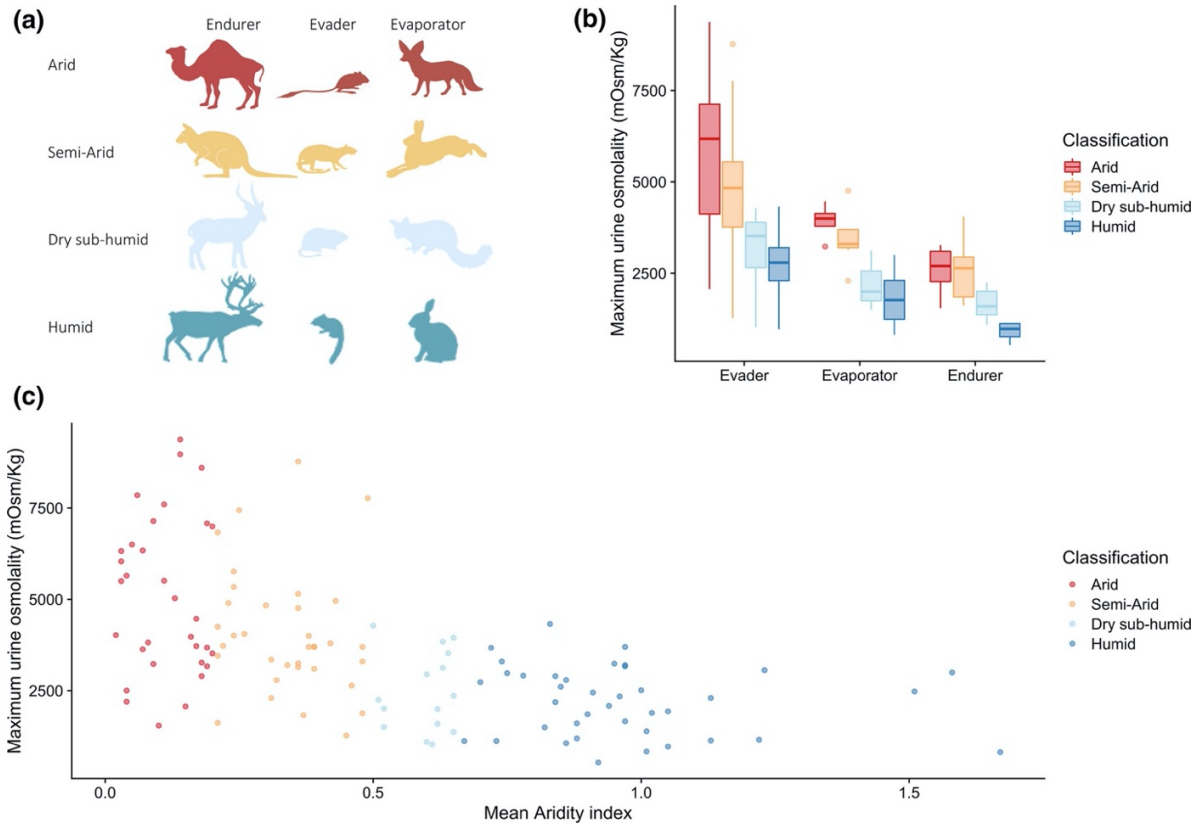


Figure 3. Maximum urine osmolality (mOsm/kg) in mammals and mean aridity index (AI). (A) Schematic illustration of the endurer-evader-evaporator concept proposed by Willmer et al. (2005) and classification of mammals into arid, semi-arid, dry sub-humid and humid habitat occupiers according to Trabucco and Zomer (2009). (B) Boxplot for maximum mOsm in 121 endurers, evaders, and evaporators with ranges with different AIs. (C) Scatterplot for maximum mOsm in 121 mammal species.

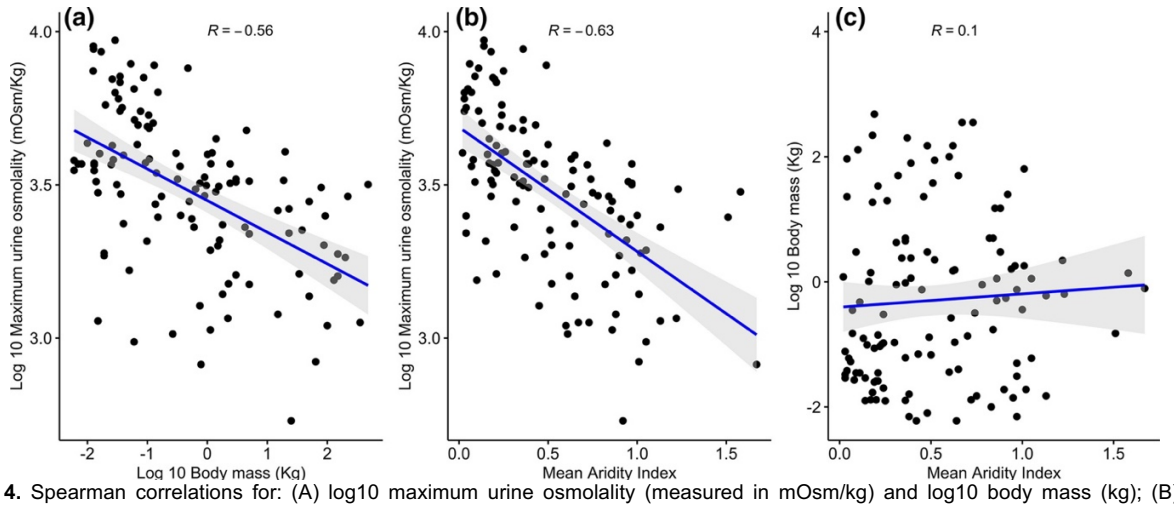


Figure 4. Spearman correlations for: (A) log10 maximum urine osmolality (measured in mOsm/kg) and log10 body mass (kg); (B) log10 maximum urine osmolality (measured in mOsm/kg) and mean annual aridity index; and (C) log10 body mass (kg) and mean annual aridity index. *P*-values are not presented, as the plots do not take phylogenetic affinity into account.

Table 3. PGLS models for predicting mammalian log10 maximum urine osmolality (mOsm/kg) in the dataset including only dehydrated individuals. Significant *P*-values are highlighted in bold; AIC = Akaike information criterion, s.e. = standard error.

Model	*AIC*	*Variable*	*Coefficient*	*s.e.*	*t-value*	*P-value*
AI + ***BM***	-62.3	Intercept	3.572	0.167	21.33	**0.0000**
(best model)		AI	-0.309	0.047	-6.51	**0.0000**
		BM	-0.091	0.022	-4.13	**0.0004**
AI + ***BM*** + *M* + ***Di***	-34.7	intercept	3.679	0.132	27.83	**0.0000**
		AI	-0.344	0.043	-7.96	**0.0000**
		BM	-0.067	0.016	-4.18	**0.0001**
		A_{method}	-0.107	0.141	-0.75	0.4528
		F_{method}	-0.007	0.042	-0.16	0.8702
		FV_{method}	0.096	0.104	0.93	0.3563
		V_{method}	0.029	0.070	0.42	0.6758
		CI_{diet}	-0.098	0.165	-0.60	0.5536
		Fr_{diet}	-0.406	0.178	-2.28	**0.0254**
		G_{diet}	0.090	0.142	0.63	0.5281
		H_{diet}	-0.106	0.135	-0.79	0.4333
		I_{diet}	-0.049	0.140	-0.35	0.7261
		O_{diet}	0.002	0.139	0.02	0.9873
cAI + ***BM*** + *M* +***Di***	-24.99	intercept	3.663	0.137	26.71	**0.0000**
		Dry sub-humid	-0.242	0.051	-4.72	**0.0000**
		Humid	-0.342	0.045	-7.58	0.0000
		Semi-Arid	-0.105	0.045	-2.37	0.0207
		BM	-0.068	0.016	-4.32	0.0000
		A_{method}	-0.154	0.143	-1.08	0.2829
		F_{method}	-0.035	0.042	-0.83	0.4094
		FV_{method}	0.084	0.105	0.80	0.4246
		V_{method}	0.034	0.068	0.50	0.6208
		CI_{diet}	-0.116	0.167	-0.70	0.4880
		Fr_{diet}	-0.438	0.179	-2.45	**0.0166**
		G_{diet}	0.105	0.144	0.73	0.4653
		H_{diet}	-0.083	0.137	-0.61	0.5454
		I_{diet}	-0.011	0.143	-0.08	0.9384
		O_{diet}	0.036	0.140	0.26	0.7971

Model notation refers to the combination of variables used as predictors while taking phylogeny into account: mean AI = annual aridity index; BM = log10 body mass (kg); C = condition (fed-F, given salt load-S, protein load-P, unknown condition-X, and combinations of these, all relative to dehydrated-D); M = method (addition of solutes-A, freezing point-F, vapour pressure-V, and combinations of these, all relative to unknown

method-?); Di = diet (frugivorous-Fr, granivorous-G, herbivorous-H, insectivorous-I, omnivorous-O, and CI-both carnivorous and insectivorous, all relative to strictly carnivorous-C).

We obtained mass-adjusted BMR for 84 of the 121 species from our dataset. BMR was weakly correlated with bioclimatic variables (Appendix S5), and variation in this parameter is better explained by log10 body mass (Table 4).

Table 4. PGLS models for predicting log10 mass-adjusted basal metabolic rate (ml of O_2 per g per day) ranked by AIC (Akaike information criterion). Significant variables are highlighted in bold (*P*-value < 0.05). When λ is zero, the covariance between species is zero and the linear model approaches that of a non-phylogenetic regression, while when λ is 1, the evolution of the residual error follows a Brownian model.

Model	AIC	Pagel's λ
BM	-27.28	0.7601046
BM + AI	-23.61	0.7503346
BM + BIO1 + BIO5 + BIO6 + BIO7	-6.51	0.6825869
BM + AI + BIO1 + BIO5 + BIO6 + BIO7	-1.69	0.6759448
BM + AI + Di	-1.68	0.6288732
BM + **BIO1** + Di	1.73	0.5731028
BM + BIO5 + Di	1.77	0.6274593
BM + **BIO6** + Di	1.91	0.5383454
BM + BIO7 + Di	4.05	0.5487208
BM + **cAI** + Di	5.69	0.5982846
BM + **BIO1** + BIO5 + BIO6 + BIO7 + Di	12.28	0.4437627

Model notation refers to the combination of variables used as predictors while taking phylogeny into account. Variables are BM = log10 body mass (g); mean AI = annual aridity index (continuous), cAI = categorical AI (arid, semi-arid, dry sub-humid, humid), Di = diet (carnivorous-C, frugivorous-Fr, granivorous-G, herbivorous-H, insectivorous-I, omnivorous-O, carnivorous and insectivorous-CI); and mean temperature values for BIO1 = annual mean temperature, BIO5= maximum temperature of warmest month, BIO6 = minimum temperature of coldest month, BIO7 = temperature annual range.

Discussion

The extent to which different mammalian species from desert environments have significantly higher urine-concentrating abilities than mammals from other environments, thus exhibiting phenotypes enabling better water retention in extreme conditions of water-deprivation and extreme temperature ranges, has not previously been subject to rigorous statistical analyses that account for phylogeny (Beuchat 1990, 1996). We performed ancestral state reconstruction on a molecular phylogeny to show

that increased ability to concentrate urine, measured as maximum mOsm, has evolved convergently in many desert mammals.

PGLS analysis on a multi-species dataset revealed that desert-dwelling species effectively concentrate more urine than non-desert species, even when accounting for both body mass and ancestry, and, furthermore, that differences in mammalian maximum mOsm can be predicted, to some degree, by the aridity of species' range. Our measurement of environmental aridity is inherently imprecise as it does not account for spatial and temporal changes in the environmental conditions or in the species' geographic range, and it implicitly assumes a uniform density distribution across the species' range (Trabucco and Zomer 2009; Fick and Hijmans 2017). Still, we find it more rigorous than the often-used approach of classifying species as 'mesic' or 'xeric' based on habitat descriptions or information on whether the species' survival depends on standing water (Beuchat 1996; Lovegrove 2000). Despite the challenges, and while acknowledging additional noise added by varying experimental conditions and methodologies, we find a strongly significant correlation between maximum mOsm and mean AI. This observation is similar to what has been previously reported in other desert-adaptive phenotypes, such as BMR (Lovegrove 2000; Fristoe et al. 2015). However, we find a much weaker correlation for BMR and environmental variables when analysing 84 of the same species as in the mOsm analysis, suggesting that, in contrast to our results for mOsm, differences in BMR among species are mostly explained by body mass differences and are only weakly predicted by environmental variables (Genoud et al. 2018).

A contentious debate in the fields of ecological and evolutionary physiology is whether phenotypic differences between species and populations inhabiting contrasting environments are due to genetic adaptations or due to plastic responses, and difference in the ability to concentrate urine is no exception. Although all mammals, when water-deprived, are able to increase urine concentration in relation to their optimal hydrated states (Sands and Layton 2014; Knepper et al. 2015), our results, from a dataset in which subjects were all captive and water-deprived, provide strong evidence that increased maximal urine concentration is an adaptation to high aridity and has evolved multiple times in different phylogenetic lineages. Notably, increased urine-concentrating ability correlates with increased renal medullary thickness in desert mammals (Beuchat 1990; Schwimmer and Haim 2009).

Our PGLS analyses also suggest that, among the many physical and climatic challenges faced by desert species, aridity, more than temperature, has been one of the main selective pressures leading to increasing maximum urine-concentrating ability and driving its repeated evolution in different desert mammalian lineages. Future comprehensive statistical studies investigating correlations of bioclimatic variables with more of the classical desert phenotypes, using multiple covariates including phylogeny, will provide further elucidation on the extent to which climate is predictive for mammalian phenotypes. As population and comparative genomics studies in desert environments continue to search for the genetic

basis of desert adaptation (e.g. Wu et al. 2014, Kim et al. 2016, Yang et al. 2016, Bittner et al. 2020), maximum urine-concentrating ability might be one of the primary physiological traits that lends itself to more detailed genetic analyses.

Acknowledgments

We thank Pedro Tarroso and Antigoni Kaliontzopoulou for discussions that helped us conceive this study. J.L.R., J.C.B and R.G. were supported by the Portuguese Foundation for Science and Technology (SFRH/BD/116397/2016, CEECINST/00014/2018, and DL57/2016, respectively). This work was partially supported by FCT project PTDC/BIA-EVL/31902/2017.

Author's contributions

J.L.R contributed to the study's design, data analysis, and drafted the manuscript. R.G, J.C.B, R.N. contributed to the study's conception and design, data interpretation, editing and critical revision. All authors approve publication and agree to be accountable for all aspects of the work.

References

Beuchat C.A. 1990. Body size, medullary thickness, and urine concentrating ability in mammals. Am. J. Physiol. - Regul. Integr. Comp. Physiol. 258.

Beuchat C.A. 1996. Structure and concentrating ability of the mammalian kidney: Correlations with habitat. Am. J. Physiol. - Regul. Integr. Comp. Physiol. 271.

Bittner N.K.J., Mack K.L., Nachman M.W. 2020. Plasticity in gene expression facilitates invasion of the desert environment in house mice. bioRxiv Prepr. 2011:1–9.

Clarke A., Rothery P., Isaac N.J.B. 2010. Scaling of basal metabolic rate with body mass and temperature in mammals. J. Anim. Ecol. 79:610–619.

Diaz G.B., Ojeda R.A. 1999. Kidney structure and allometry of Argentine desert rodents. J. Arid Environ. 41:453–461.

Donald J., Pannabecker T.L. 2015. Osmoregulation in Desert-Adapted Mammals. Sodium and Water Homeostasis. Springer New York. p. 191–211.

Downs C.T., Perrin M.R. 1991. Urinary concentrating ability of four Gerbillurus species of southern African arid regions. J. Arid Environ. 20:71–81.

Elgar M.A., Harvey P.H. 1987. Basal Metabolic Rates in Mammals: Allometry, Phylogeny and Ecology. Funct. Ecol. 1:25.

Fick S.E., Hijmans R.J. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land

areas. Int. J. Climatol. 37:4302–4315.

Fristoe T.S., Burger J.R., Balk M.A., Khaliq I., Hof C., Brown J.H. 2015. Metabolic heat production and thermal conductance are mass-independent adaptations to thermal environment in birds and mammals. Proc. Natl. Acad. Sci. U. S. A. 112:15934–15939.

Fuller A., Hetem R.S., Maloney S.K., Mitchell D. 2014. Adaptation to heat and water shortage in large, arid-zone mammals. Physiology. 29:159–167.

Genoud M., Isler K., Martin R.D. 2018. Comparative analyses of basal rate of metabolism in mammals: data selection does matter. Biol. Rev. 93:404–438.

IUCN. 2019. International Union for Conservation of Nature. The IUCN Red List of Threatened Species. Version 2019-1. .

IUCN. 2020. International Union for Conservation of Nature. The IUCN Red List of Threatened Species. Version 2020-1. .

Kim E.S., Elbeltagy A.R., Aboul-Naga A.M., Rischkowsky B., Sayre B., Mwacharo J.M., Rothschild M.F. 2016. Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. Heredity (Edinb). 116:255–264.

Knepper M.A., Kwon T.H., Nielsen S., Ingelfinger J.R. 2015. Molecular physiology of water balance. N. Engl. J. Med. 372:1349–1358.

Kronfeld N., Shkolnik A. 1996. Adaptation to Life in the Desert in the Brown Hare (Lepus capensis). J. Mammal. 77:171–178.

Lovegrove B.G. 2000. The zoogeography of mammalian basal metabolic rate. Am. Nat. 156:201–219.

Lovegrove B.G. 2003. The influence of climate on the basal metabolic rate of small mammals: A slow-fast metabolic continuum. J. Comp. Physiol. B Biochem. Syst. Environ. Physiol. 173:87–112.

Maloiy G., Macfarlane W., Shkolnik A. 1979. Mammalian herbivores. In: Maloiy G.M.O., editor. London, United Kingdom: Academic Press.

Nagy K.A. 2004. Water economy of free-living desert animals. Int. Congr. Ser. 1275:291–297.

Nagy K.A., Shoemaker V.H., Costa W.R. 1976. Water, Electrolyte, and Nitrogen Budgets of Jackrabbits (Lepus californicus) in the Mojave Desert. Physiol. Zool. 49:351–363.

Ostrowski S., Williams J.B., Mésochina P., Sauerwein H. 2006. Physiological acclimation of a desert antelope, Arabian oryx (Oryx leucoryx), to long-term food and water restriction. J. Comp. Physiol. B Biochem. Syst. Environ. Physiol. 176:191–201.

Pagel M. 1999. Inferring the historical patterns of biological evolution. Nature. 401:877–884.

Pannabecker T.L. 2013. Comparative physiology and architecture associated with the mammalian urine concentrating mechanism: Role of inner medullary water and urea transport pathways in the rodent medulla. Am. J. Physiol. - Regul. Integr. Comp. Physiol. 304:488–503.

Revell L.J. 2012. phytools: An R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. 3:217–223.

Rymer T.L., Pillay N., Schradin C. 2016. Resilience to droughts in mammals: A conceptual framework for estimating vulnerability of a single species. Q. Rev. Biol. 91:133–176.

Sands J.M., Layton H.E. 2014. Advances in Understanding the Urine-Concentrating Mechanism. Annu. Rev. Physiol. 76:387–409.

Schmidt-nielsen B., Schmidt-Nielsen K., Brokaw A., Schneiderman H. 1948. Water conservation in desert rodents. J. Cell. Comp. Physiol. 32:331–360.

Schmidt-nielsen K. 1965. Desert animals. Physiological problems of heat and water. Oxford University Press.

Schmidt-Nielsen K., Haines H.B., Hackel D.B. 1964. Diabetes mellitus in the sand rat induced by standard laboratory diets. Science (80-.). 143:689–690.

Schwimmer H., Haim A. 2009. Physiological adaptations of small mammals to desert ecosystems. Integr. Zool. 4:357–366.

Trabucco A., Zomer R.. 2009. Global Aridity Index (Global-Aridity) and Global Potential Evapo-Transpiration (Global-PET) Geospatial Database. CGIAR Consort. Spat. Inf.

Tracy R.L., Walsberg G.E. 2002. Kangaroo rats revisited: Re-evaluating a classic case of desert survival. Oecologia. 133:449–457.

Upham N.S., Esselstyn J.A., Jetz W. 2019. Inferring the mammal tree: Species-level sets of phylogenies for questions in ecology, evolution, and conservation. PLoS Biol. 17:e3000494.

Walsberg G.E. 2000. Small mammals in hot deserts: Some generalizations revisited. Bioscience. 50:109–120.

Ward D. 2016. The biology of deserts. Oxford University Press.

White C.R., Seymour R.S. 2003. Mammalian basal metabolic rate is proportional to body mass2/3. Proc. Natl. Acad. Sci. U. S. A. 100:4046–4049.

Willmer P., Stone G., Johnston I.A. 2005. Environmental physiology of animals. Blackwell Pub.

Wu H., Guang X., Al-Fageeh M.B., Cao J., Pan S., Zhou H., Zhang L., Abutarboush M.H., Xing Y., Xie Z., Alshanqeeti A.S., Zhang Y., Yao Q., Al-Shomrani B.M., Zhang D., Li J., Manee M.M., Yang Z., Yang L., Liu Y., Zhang J., Altammami M.A., Wang S., Yu L., Zhang W., Liu S., Ba L., Liu C., Yang X., Meng F., Wang S., Li L., Li E., Li X., Wu K., Zhang S., Wang J., Yin Y., Yang H., Al-Swailem A.M., Wang J. 2014. Camelid genomes reveal evolution and adaptation to desert environments. Nat. Commun. 5:10.

Yang J., Li W.R., Lv F.H., He S.G., Tian S.L., Peng W.F., Sun Y.W., Zhao Y.X., Tu X.L., Zhang M., Xie X.L., Wang Y.T., Li J.Q., Liu Y.G., Shen Z.Q., Wang F., Liu G.J., Lu H.F., Kantanen J., Han J.L., Li

M.H., Liu M.J. 2016. Whole-Genome Sequencing of Native Sheep Provides Insights into Rapid Adaptations to Extreme Environments. Mol. Biol. Evol. 33:2576–2592.

Supplementary Material

Appendix S1- Data collection from past studies concerning 121 species listed alphabetically by scientific name, followed by family and order. For each species, the following variables are reported: category according to the endurer-evader-evaporator concept, C = condition (dehydrated-D, fed-F, given salt load-S, given protein load-P, unknown-x and combinations of these), M = Method (unknown-?, addition of solutes-A, Freezing point-F, Vapor pressure-V, and combinations of these), Di = Diet (Carnivorous-C, Fructivorous-Fr, Granivorous-G, Herbivorous-H, Insectivorous-I, Omnivorous-O, and combinations of these), BM = body mass (kg), mOsm = maximum urine osmolality (mosmol/Kg H_2O), cAI = categorical annual aridity index (Arid, Semi-Arid, Dry sub-humid, Humid), mean AI = annual aridity index (continuous), and mean temperature values for BIO1 = annual mean temperature, BIO5= maximum temperature of warmest month, BIO6 = minimum temperature of coldest month, BIO7 = temperature annual range (BIO5-BIO6). Mass-adjusted Basal Metabolic Rates (ml of O_2 per g per day) are also reported (BMR). References for mOsm and BMR are also reported after the measurement.

Species_Family_Order	Category	C	M	Di	BM	mOsm	cAI	AI	BIO1	BIO5	BIO6	BIO7	BMR
Acomys_cahirinus_MURIDAE_RODENTIA	Evader	D	V	O	0.033	6039[1]	Arid	0.03	23.3	37.99	7.45	30.54	1.1[2,3]
Acomys_russatus_MURIDAE_RODENTIA	Evader	x	?	O	0.053	7850[4]	Arid	0.06	23.99	36.9	9.98	26.92	0.77[2,3]
Aepyceros_melampus_BOVIDAE_CETARTIODACTYLA	Endurer	D	F	H	38	2250[1]	Dry sub-humid	0.51	21.99	31.46	10.8	20.66	
Aethomys_namaquensis_MURIDAE_RODENTIA	Evader	D	V	O	0.107	4835[1]	Semi-Arid	0.3	19.81	31.6	5.41	26.19	0.84[2,3]
Alcelaphus_buselaphus_BOVIDAE_CETARTIODACTYLA	Endurer	D	F	H	88	2010[1]	Dry sub-humid	0.52	23.27	33.59	11.94	21.65	
Ammospermophilus_leucurus_SCIURIDAE_RODENTIA	Evader	D	F	O	0.093	3730[1]	Semi-Arid	0.22	11.73	32.64	-5.39	38.03	0.98[2,3]
Antidorcas_marsupialis_BOVIDAE_CETARTIODACTYLA	Endurer	x	F	H	34	1620[4]	Semi-Arid	0.21	19.41	31.92	4.16	27.77	
Aplodontia_rufa_APLODONTIIDAE_RODENTIA	Evaporator	D	F	H	0.785	820[1]	Humid	1.67	7.96	24.34	-3.26	27.59	0.52[2,3]
Apodemus_mystacinus_MURIDAE_RODENTIA	Evader	x	?	O	0.036	2950[4]	Dry sub-humid	0.6	11.17	28.89	-3.7	32.59	1.67[2,3]

Artibeus_jamaicensis_PHYLLOSTOMIDAE_CHIROPTERA	Evader	D	V	Fr	0.06	972[1]	Humid	1.05	23.97	31.99	15.76	16.23	1.25[2,3]
Bison_bison_BOVIDAE_CETARTIODACTYLA	Endurer	D	?	H	353	1124[1]	Humid	0.67	-2.04	22.08	-26.12	48.2	
Blarina_brevicauda_SORICIDAE_EULIPOTYPHLA	Evader	D	F	O	0.019	1857[1]	Humid	0.9	6.61	27.23	-14.89	42.11	3.2[2,3]
Calomys_musculinus_CRICETIDAE_RODENTIA	Evader	x	V	O	0.0127	8773[5]	Semi-Arid	0.36	15.27	29.24	1.96	27.28	1.63[2,3]
Camelus_dromedarius_CAMELIDAE_CETARTIODACTYLA	Endurer	D	F	H	479	3170[1]	Arid	0.19	22.51	37.81	6.91	30.9	0.1[2,3]
Canis_lupus_familiaris_CANIDAE_CARNIVORA	Evaporator	x	F	C	15	2608[1]	Humid	0.85	0.18	22.62	-21.02	43.64	0.33[2,3]
Capra_aegagrus_BOVIDAE_CETARTIODACTYLA	Endurer	D	?	H	23	2200[1]	Arid	0.04	13.93	32.96	-3.39	36.34	
Carollia_perspicillata_PHYLLOSTOMIDAE_CHIROPTERA	Evader	D	V	Fr	0.015	1137[1]	Humid	1.13	24.36	31.52	16.84	14.68	2.11[2,3]
Castor_canadensis_CASTORIDAE_RODENTIA	Endurer	x	F	H	25	537[1]	Humid	0.92	3.22	24.2	-17.29	41.49	
Cavia_porcellus_CAVIIDAE_RODENTIA	Evader	D	F	H	0.75	1275[1]	Semi-Arid	0.45	12.1	21.34	0.93	20.42	0.55[2,3]
Chaetodipus_baileyi_HETEROMYIDAE_RODENTIA	Evader	DS	?	G	0.025	3680[1]	Arid	0.19	20.28	37.93	3.26	34.67	1.31[2,3]
Chaetodipus_penicillatus_HETEROMYIDAE_RODENTIA	Evader	x	F	G	0.017	8600[1]	Arid	0.18	19.65	37.93	2.71	35.22	1.31[2,3]
Chinchilla_lanigera_CHINCHILLIDAE_RODENTIA	Evader	D	F	H	0.475	7599[1]	Arid	0.11	12.93	22.62	4.26	18.36	1.4[2,3]
Connochaetes_taurinus_BOVIDAE_CETARTIODACTYLA	Endurer	D	?	H	200	1832[1]	Semi-Arid	0.37	21.08	32.2	6.64	25.56	0.7[2,3]
Corynorhinus_townsendii_VESPERTILIONIDAE_CHIROPTERA	Evader	D	F	I	0.008	3700[1]	Semi-Arid	0.48	11.32	30.39	-5.24	35.63	0.29[2,3]
Cricetus_cricetus_CRICETIDAE_RODENTIA	Evader	D	F	O	0.108	3840[1]	Dry sub-humid	0.63	4.93	26.17	-14.75	40.92	0.64[2,3]

Cynomys_leucurus_SCIURIDAE_RODENTIA	Evaporator	D	F	H	0.962	3145[1]	Semi-Arid	0.36	5.54	27.84	-13.1	40.94	
Cynomys_ludovicianus_SCIURIDAE_RODENTIA	Evaporator	D	F	H	1.145	3704[1]	Semi-Arid	0.39	10.84	31.64	-9.08	40.72	0.38[2,3]
Dasycercus_cristicauda_DASYURIDAE_DASYUROMORPHIA	Evaporator	D	A	O	3	3231[1]	Arid	0.09	22.36	38.3	5.64	32.66	0.52[2,3]
Didelphis_virginiana_DIDELPHIDAE_DIDELPHIMORPHIA	Evaporator	DP	F	O	5	1497[1]	Humid	0.82	15.9	31.59	0.07	31.52	0.33[2,3]
Dipodomys_agilis_HETEROMYIDAE_RODENTIA	Evader	D	?	G	0.061	5150[1]	Semi-Arid	0.36	14.34	31.37	0.88	30.5	1.05[2,3]
Dipodomys_merriami_HETEROMYIDAE_RODENTIA	Evader	DP	V	G	0.035	6832[1]	Semi-Arid	0.21	18.08	35.21	1.42	33.79	1.18[2,3]
Dipodomys_spectabilis_HETEROMYIDAE_RODENTIA	Evader	D	F	G	0.1	4900[4]	Semi-Arid	0.23	15.21	33.46	-2.8	36.26	
Eligmodontia_typus_CRICETIDAE_RODENTIA	Evader	D	V	G	0.02	5763[1)	Semi-Arid	0.24	7.75	18.54	-5.8	24.34	
Eptesicus_fuscus_VESPERTILIONIDAE_CHIROPTERA	Evader	D	F	I	0.013	3675[1]	Humid	0.72	10.29	28.84	-7.92	36.76	2.42[2,3]
Equus_africanus_EQUIDAE_PERISSODACTYLA	Endurer	D	F	H	129	1545[4]	Arid	0.1	27.48	37.16	18.32	18.84	
Eremitalpa_granti_CHRYSOCHLORIDAE_AFROSORICIDA	Evader	D	F	I	0.027	3820[4]	Arid	0.08	18.81	29.88	8.13	21.75	0.49[2,3]
Erethizon_dorsatum_ERETHIZONTIDAE_RODENTIA	Evaporator	P	F	H	15	1195[1]	Humid	0.88	1.74	23.46	-19.38	42.84	0.25[2,3]
Erinaceus_europaeus_ERINACEIDAE_EULIPOTYPHLA	Evader	P	F	O	0.637	3062[1]	Humid	1.23	7.34	22.56	-5.18	27.74	0.4[2,3]
Euderma_maculatum_VESPERTILIONIDAE_CHIROPTERA	Evader	D	F	I	0.016	4000[1]	Semi-Arid	0.38	9.79	30.25	-7.41	37.66	
Eudorcas_thomsonii_BOVIDAE_CETARTIODACTYLA	Endurer	D	?	H	23	2640[6]	Semi-Arid	0.46	20.63	28.89	12.61	16.28	
Felis_catus_FELIDAE_CARNIVORA	Evaporator	D	?	C	5	3250[1]	Semi-Arid	0.36	20.08	33.98	5.87	28.11	

Gerbilliscus_leucogaster_MURIDAE_RODENTIA	Evader	D	V	G	0.13	7767[1]	Semi-Arid	0.49	21.5	31.7	8.81	22.9	1.02[2,3]
Gerbillurus_paeba_MURIDAE_RODENTIA	Evader	DF	F	G	0.026	6997[7]	Arid	0.2	19.99	32.46	4.92	27.54	1.03[2,3]
Gerbillurus_setzeri_MURIDAE_RODENTIA	Evader	DF	F	G	0.038	5648[7]	Arid	0.04	20.24	28.24	11.18	17.06	0.8[2,3]
Gerbillurus_tytonis_MURIDAE_RODENTIA	Evader	D	V	G	0.029	6324[1]	Arid	0.03	18.64	29.16	7.81	21.36	1.06[2,3]
Gerbillurus_vallinus_MURIDAE_RODENTIA	Evader	DF	F	G	0.035	7144[7]	Arid	0.09	19.05	33.26	3.68	29.57	0.9[2,3]
Gerbillus_gerbillus_MURIDAE_RODENTIA	Evader	x	?	G	0.077	5500[4]	Arid	0.03	24.68	39.93	8.01	31.92	1.06[2,3]
Glaucomys_volans_SCIURIDAE_RODENTIA	Evader	D	F	O	0.75	3200[1]	Humid	0.97	12.75	30.07	-5.14	35.21	1.2[2,3]
Hemiechinus_auritus_ERINACEIDAE_EULIPOTYPHLA	Evader	P	F	O	0.301	4010[4],	Semi-Arid	0.24	8.48	30.96	-13.47	44.43	0.38[2,3]
Isoodon_macrourus_PERAMELIDAE_PERAMELEMORPHIA	Evaporator	D	F	I	1.551	3126[1]	Dry sub-humid	0.63	23	33.24	10.86	22.38	0.37[2,3]
Jaculus_jaculus_DIPODIDAE_RODENTIA	Evader	x	?	G	0.06	6500[1]	Arid	0.05	23.76	39.42	7.41	32	1.23[2,3]
Kobus_ellipsiprymnus_BOVIDAE_CETARTIODACTYLA	Endurer	D	?	H	100	1098[1]	Dry sub-humid	0.6	24.04	33.54	14.14	19.41	0.27[2,3]
Kobus_kob_BOVIDAE_CETARTIODACTYLA	Endurer	D	?	H	150	1594[4]	Dry sub-humid	0.62	26.25	36.04	16.91	19.14	
Lasionycteris_noctivagans_VESPERTILIONIDAE_CHIROPTERA	Evader	D	F	I	0.01	4325[1]	Humid	0.83	7.47	27.3	-11.76	39.06	
Lepus_californicus_LEPORIDAE_LAGOMORPHA	Evaporator	DF	V	H	2.4	3700[8]	Semi-Arid	0.39	13.8	32.54	-3.48	36.02	0.57[2,3]
Lepus_capensis_LEPORIDAE_LAGOMORPHA	Evaporator	S	V	H	1.4	4470[9]	Arid	0.17	22.42	36.38	7.76	28.63	
Liomys_irroratus_HETEROMYIDAE_RODENTIA	Evader	D	V	G	0.068	4280[1]	Dry sub-humid	0.5	19.26	31.13	6.42	24.71	1.12[2,3]

Liomys_pictus_HETEROMYIDAE_RODENTIA	Evader	D	?	G	0.04	3950[1]	Dry sub-humid	0.65	22.8	33.88	11.16	22.72	
Macaca_mulatta_CERCOPITHECIDAE_PRIMATES	Evaporator	D	F	O	3	1607[1]	Humid	0.88	19.08	32.21	3.96	28.25	
Macropus_eugenii_MACROPODIDAE_DIPROTODONTIA	Evaporator	SP	F	H	3	3301[1]	Semi-Arid	0.48	15.82	27.52	6.69	20.84	0.29[2,3]
Macropus_robustus_MACROPODIDAE_DIPROTODONTIA	Endurer	D	F	H	19.8	4054[1]	Semi-Arid	0.26	22.57	36.38	7.69	28.69	0.19[2,3]
Macropus_rufus_MACROPODIDAE_DIPROTODONTIA	Endurer	D	F	H	18.7	3270[1]	Arid	0.18	22.51	36.99	7.13	29.86	0.18[2,3]
Macrotis_lagotis_THYLACOMYIDAE_PERAMELEMORPHIA	Evaporator	D	F	CI	1.011	3976[1]	Arid	0.16	25.53	39.29	9.32	29.97	0.35[2,3]
Madoqua_kirkii_BOVIDAE_CETARTIODACTYLA	Evaporator	D	?	H	4.5	4762[1]	Semi-Arid	0.36	22.38	30.92	13.18	17.74	0.5[2,3]
Meriones_unguiculatus_MURIDAE_RODENTIA	Evader	D	F	G	0.07	4958[4]	Semi-Arid	0.43	1.43	25.02	-23.97	48.99	1.15[2,3]
Mesocricetus_auratus_CRICETIDAE_RODENTIA	Evader	D	F	O	0.105	5340[1]	Semi-Arid	0.24	17.26	37.73	0.42	37.31	1.5[2,3]
Microtus_ochrogaster_CRICETIDAE_RODENTIA	Evader	D	F	H	0.04	2362[1]	Dry sub-humid	0.65	7.97	28.98	-13.34	42.32	1.749[2,3]
Microtus_pennsylvanicus_CRICETIDAE_RODENTIA	Evader	D	F	H	0.05	1663[1]	Humid	0.97	-0.2	21.97	-22.05	44.02	2.65[2,3]
Myodes_gapperi_CRICETIDAE_RODENTIA	Evader	D	F	H	0.019	1893[1]	Humid	1.02	1.19	22.76	-20.69	43.45	1.75[2,3]
Myotis_auriculus_VESPERTILIONIDAE_CHIROPTERA	Evader	D	F	I	0.007	3700[1]	Semi-Arid	0.38	17.26	32.71	1.43	31.27	
Myotis_lucifugus_VESPERTILIONIDAE_CHIROPTERA	Evader	D	F	I	0.007	3700[1]	Humid	0.97	3.15	24.09	-17.51	41.6	1.43[2,3]
Myotis_thysanodes_VESPERTILIONIDAE_CHIROPTERA	Evader	D	F	I	0.006	3800[1]	Semi-Arid	0.42	12.85	31.22	-3.21	34.43	
Myotis_volans_VESPERTILIONIDAE_CHIROPTERA	Evader	D	?	I	0.006	3525[1]	Dry sub-humid	0.64	8.45	27.56	-8.04	35.59	

Nanger_granti_BOVIDAE_CETARTIODACTYLA	Endurer	D	?	H	50	2789[1]	Semi-Arid	0.32	24.19	32.23	16.78	15.45	
Neovison_vison_MUSTELIDAE_CARNIVORA	Evaporator	S	F	C	1.6	2087[4]	Humid	0.94	1.48	23.05	-19.48	42.53	0.74[2,3]
Notomys_alexis_MURIDAE_RODENTIA	Evader	D	FV	G	0.029	9374[4]	Arid	0.14	22.77	37.88	6.88	30.99	1.4[2,3]
Notomys_cervinus _MURIDAE_RODENTIA	Evader	S	FV	G	0.035	5510[1]	Arid	0.11	23.8	38.76	7.16	31.6	1.33[2,3]
Nycticebus_coucang_LORISIDAE_PRIMATES	Evaporator	D	?	O	1.38	3000[1]	Humid	1.58	25.33	30.45	20.64	9.82	0.24[2,3]
Ochrotomys_nuttalli_CRICETIDAE_RODENTIA	Evader	D	F	G	0.031	3166[1]	Humid	0.97	15.92	32.05	-0.87	32.92	1.39[2,3]
Octomys_mimax_OCTODONTIDAE_RODENTIA	Evader	x	V	H	0.09832	2071[5]	Arid	0.15	15.97	30.18	0.72	29.47	0.97[2,3]
Ondatra_zibethicus_CRICETIDAE_RODENTIA	Evaporator	D	F	H	1.124	1063[1]	Humid	0.86	0.94	23.67	-21.06	44.73	0.64[2,3]
Onychomys_torridus_CRICETIDAE_RODENTIA	Evader	DPS	F	CI	0.026	4250[1]	Semi-Arid	0.21	16.87	35.89	0.21	35.67	1.55[2,3]
Oryctolagus_cuniculus_LEPORIDAE_LAGOMORPHA	Evaporator	D	F	H	1.8	1390[4]	Humid	1.01	10.06	24.1	-1.01	25.11	0.57[2,3]
Oryx_beisa_BOVIDAE_CETARTIODACTYLA	Endurer	D	?	H	79	3100[6]	Semi-Arid	0.39	23.2	31.06	15.83	15.24	
Oryx_gazella_BOVIDAE_CETARTIODACTYLA	Endurer	x	?	H	220	2900[6]	Arid	0.18	20.21	32.88	4.68	28.19	
Oryx_leucoryx_BOVIDAE_CETARTIODACTYLA	Endurer	D	F	H	92.5	2504[10]	Arid	0.04	27.47	41.97	11.21	30.76	0.19[2,3]
Otolemur_crassicaudatus_GALAGIDAE_PRIMATES	Evaporator	D	F	O	1.5	2000[1]	Dry sub-humid	0.62	22.08	30.91	11.65	19.26	0.43[2,3]
Otospermophilus_beecheyi_SCIURIDAE_RODENTIA	Evader	D	V	O	0.318	3300[1]	Humid	0.74	12.26	29.93	-0.55	30.48	0.53[2,3]
Paraechinus_aethiopicus_ERINACEIDAE_EULIPOTYPHLA	Evader	P	F	O	0.352	3634[4]	Arid	0.07	23.96	39.9	7.88	32.02	0.25[2,3]

Perodicticus_potto_LORISIDAE_PRIMATES	Evaporator	D	F	O	1.125	1935[1]	Humid	1.05	25.9	33.18	19.7	13.48	0.37[2,3]
Peromyscus_polionotus_CRICETIDAE_RODENTIA	Evader	D	F	O	0.014	3240[1]	Humid	0.95	17.8	32.66	2.03	30.63	1.17[2,3]
Procavia_capensis_PROCAVIIDAE_HYRACOIDEA	Evaporator	D	F	H	2.4	3200[4]	Semi-Arid	0.34	24.78	35.94	12.84	23.11	0.4[2,3]
Psammomys_obesus_MURIDAE_RODENTIA	Evader	D	F	H	0.15	6340[1]	Arid	0.07	22.09	38.32	6.02	32.3	
Pseudomys_hermannsburgensis_MURIDAE_RODENTIA	Evader	D	F	G	0.0126	8970[1]	Arid	0.14	22.77	37.74	6.92	30.82	1.91[2,3]
Rangifer_tarandus_CERVIDAE_CETARTIODACTYLA	Endurer	D	F	H	64	837[1]	Humid	1.01	-8.42	17.57	-32.09	49.65	0.31[2,3]
Rattus_rattus_MURIDAE_RODENTIA	Evader	D	F	O	0.172	2900[1]	Humid	0.84	17.81	31.1	4.48	26.62	
Rattus_villosissimus_MURIDAE_RODENTIA	Evader	D	?	O	0.141	3457[1]	Semi-Arid	0.21	24.15	37.82	8.45	29.37	0.58[2,3]
Rhinolophus_ferrumequinum_RHINOLOPHIDAE_CHIROPTERA	Evader	D	F	I	0.015	2980[1]	Humid	0.75	12.42	28.99	-3.19	32.18	
Rhinopoma_hardwickii_RHINOPOMATIDAE_CHIROPTERA	Evader	D	F	I	0.013	3722[1]	Arid	0.17	24.76	39.16	9.25	29.91	
Saguinus_oedipus_CALLITRICHIDAE_PRIMATES	Evader	D	F	O	0.6	2300[1]	Humid	1.13	27.03	32.67	21.73	10.94	
Salinomys_delicatus_CRICETIDAE_RODENTIA	Evader	x	V	O	0.0125	7440[5]	Semi-Arid	0.25	19.36	33.75	3.85	29.89	
Sapajus_apella_CEBIDAE_PRIMATES	Evaporator	D	F	O	2.2	1160[1]	Humid	1.22	25.85	32.66	19.26	13.4	
Sciurus_anomalus_SCIURIDAE_RODENTIA	Evader	x	?	O	0.265	1031[1]	Dry sub-humid	0.61	12.5	30.23	-2.51	32.74	
Sciurus_carolinensis_SCIURIDAE_RODENTIA	Evader	D	F	O	0.55	2450[1]	Humid	0.91	10.78	29.42	-8.66	38.08	0.84[2,3]
Setonix_brachyurus_MACROPODIDAE_DIPROTODONTIA	Evaporator	DS	F	H	5	2188[1]	Humid	0.84	15.44	26.63	7.02	19.61	0.3[2,3]

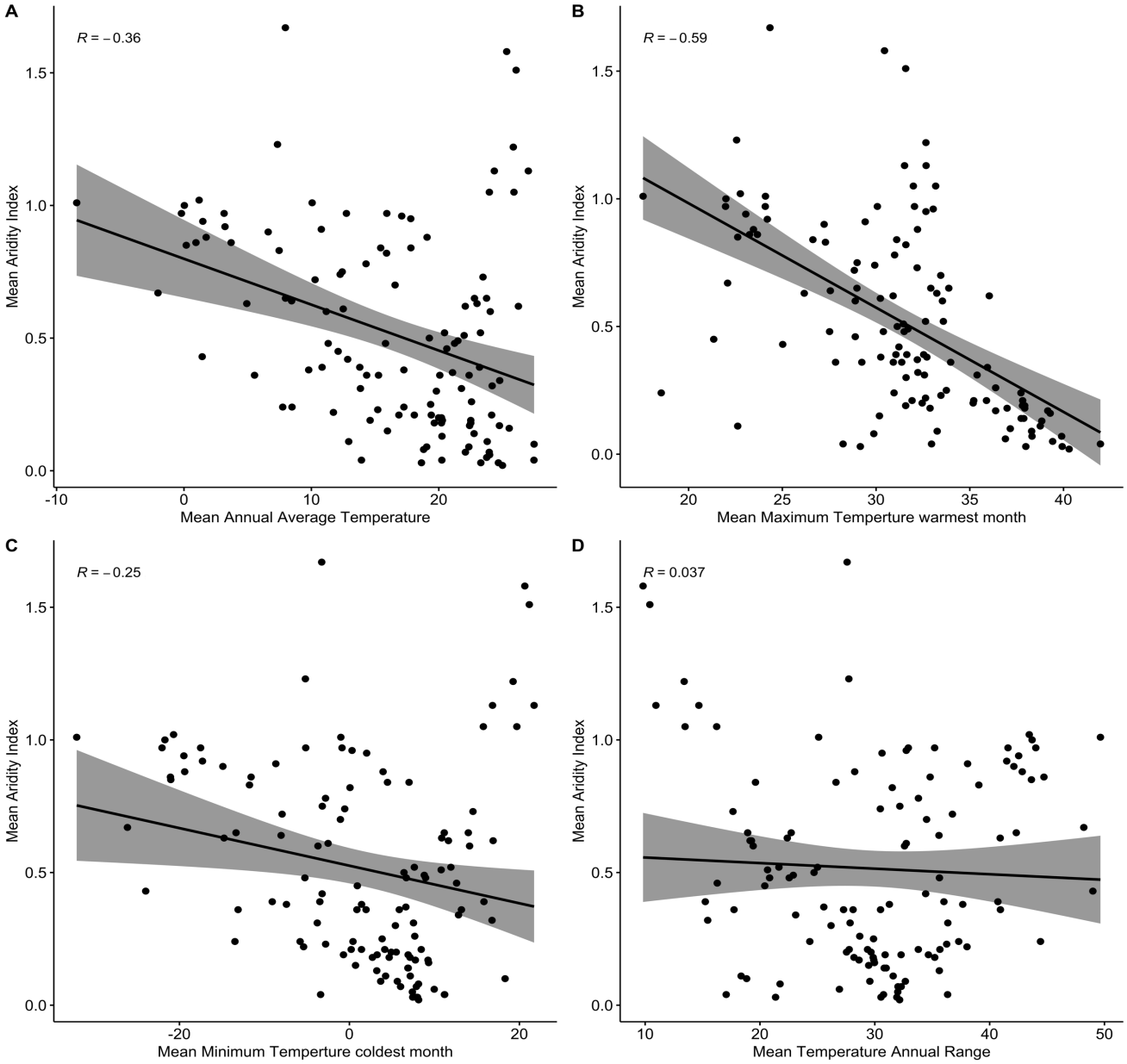
Sigmodon_hispidus_CRICETIDAE_RODENTIA	Evader	D	F	O	0.136	2734[1]	Humid	0.7	16.57	33.45	-1.06	34.51	1.654[2,3]
Sminthopsis_crassicaudata_DASYURIDAE_DASYUROMORPHIA	Evader	D	F	CI	0.013	3519[4]	Arid	0.2	20.21	35.18	5.51	29.67	1.29[2,3]
Sylvilagus_aquaticus_LEPORIDAE_LAGOMORPHA	Evaporator	S	F	H	1.8	2344[1]	Humid	0.96	17.09	33.05	0.31	32.74	
Sylvilagus_audubonii_LEPORIDAE_LAGOMORPHA	Evaporator	S	F	H	0.9	3349[1]	Semi-Arid	0.31	13.85	32.58	-3.78	36.36	0.65[2,3]
Sylvilagus_floridanus_LEPORIDAE_LAGOMORPHA	Evaporator	S	F	H	0.9	2914[1]	Humid	0.78	14.29	30.99	-2.81	33.8	
Syncerus_caffer_BOVIDAE_CETARTIODACTYLA	Endurer	D	?	H	353	1124[4]	Humid	0.73	23.47	32.19	14.54	17.65	
Tachyglossus_aculeatus_TACHYGLOSSIDAE_MONOTREMATA	Evaporator	D	F	I	4.25	2300[1]	Semi-Arid	0.31	21.78	35.39	7.54	27.85	0.13[2,3]
Tamiasciurus_hudsonicus_SCIURIDAE_RODENTIA	Evader	D	F	O	0.36	2513[1]	Humid	1	0.03	21.99	-21.71	43.7	1.12[2,3]
Tragelaphus_oryx_BOVIDAE_CETARTIODACTYLA	Endurer	D	F	H	150	1881[1]	Semi-Arid	0.48	21.24	31.5	8.96	22.55	0.24[2,3]
Tragelaphus_scriptus_BOVIDAE_CETARTIODACTYLA	Endurer	D	?	H	50	1369[1]	Dry sub-humid	0.65	23.75	32.92	13.99	18.93	
Trichosurus_vulpecula_PHALANGERIDAE_DIPROTODONTIA	Evaporator	D	F	H	2.25	1504[1]	Dry sub-humid	0.52	20.45	32.64	7.64	25.01	0.32[2,3]
Tupaia_glis_TUPAIIDAE_SCANDENTIA	Evader	D	F	O	0.15	2480[4]	Humid	1.51	26.07	31.58	21.17	10.41	0.76[2,3]
Tympanoctomys_barrerae_OCTODONTIDAE_RODENTIA	Evader	x	V	H	0.0868	7080[5)	Arid	0.19	14.61	31.58	-0.7	32.29	1.08[2,3]
Urocitellus_columbianus_SCIURIDAE_RODENTIA	Evader	D	FV	O	0.5	2792[1]	Humid	0.86	3.7	23.25	-11.57	34.82	
Vulpes_zerda_CANIDAE_CARNIVORA	Evaporator	D	F	O	1.2	4022[1]	Arid	0.02	25	40.3	8.14	32.16	0.48[2,3]
Xerospermophilus_tereticaudus_SCIURIDAE_RODENTIA	Evader	D	F	H	0.125	5030[1]	Arid	0.13	20.25	38.84	3.22	35.62	0.56[2,3]

Appendix S2 – PGLS models for predicting log10 mammalian maximum urine osmolality (mOsm/kg) ranked by AIC (Akaike information criterion)**.** Model notation refers to the combination of variables used as predictors while taking phylogeny into account. When **λ** is zero the covariance between species is zero and the linear model approaches that of a non-phylogenetic regression, while when **λ** is 1 the evolution of the residual error follows a Brownian model. Highlighted are significant variables (p-value < 0.05).

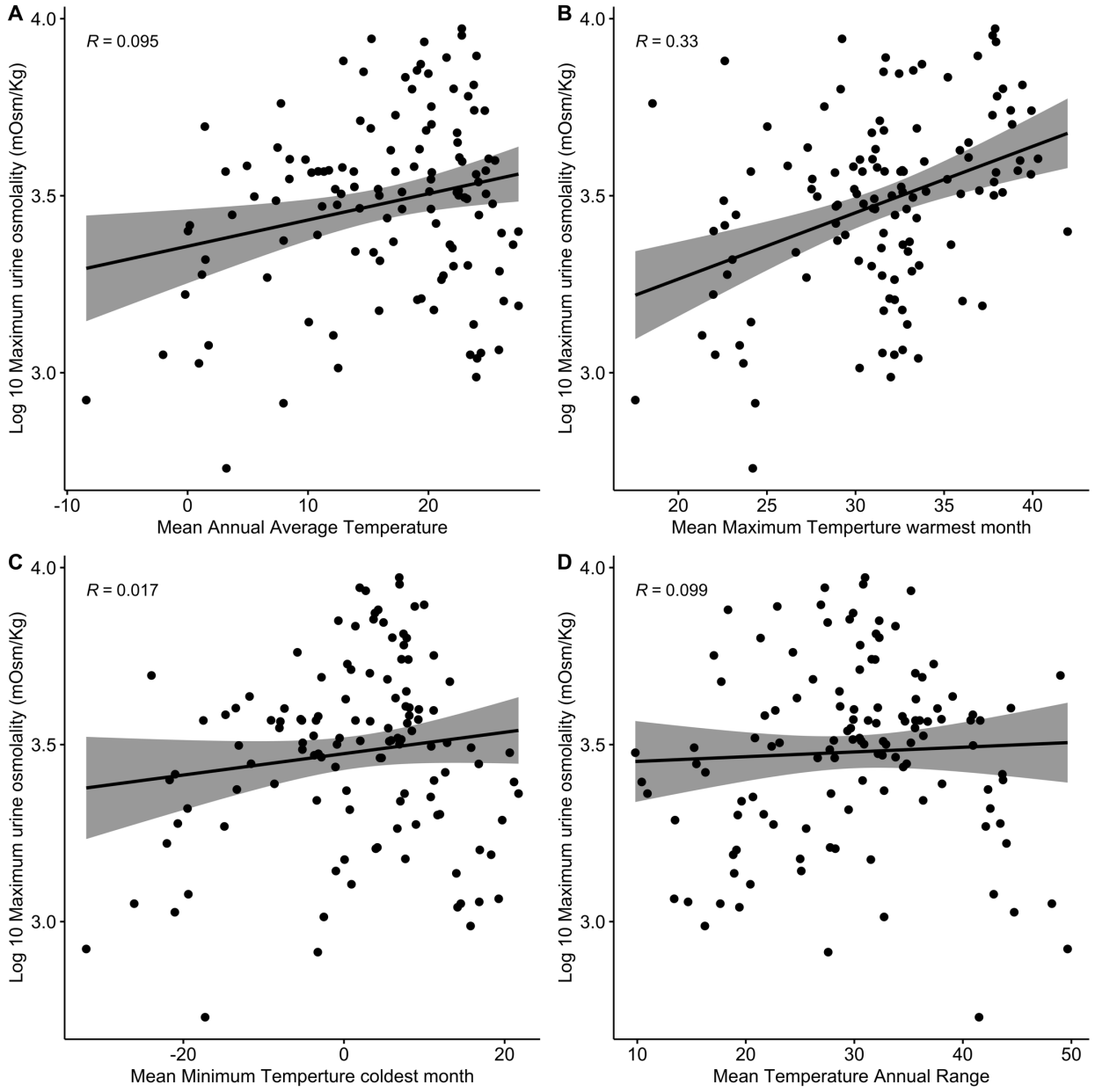
Dataset	Model	AIC	Pagel's λ
Different conditions (121 species)	**AI** + **BM**	-74.4	0.2219936
	AI + **BM** + **Di**	-66.3	-0.0146275
	AI	-60.2	0.8561328
	AI + **BM** + M	-54.6	0.1146517
	BIO5 + **BM**	-49.4	0.955564
	AI + **BM** + BIO1 + BIO5 + BIO6 + BIO7	-47.7	0.2364178
	BIO1 + **BM**	-45.6	0.9632151
	BIO6 + **BM**	-39.6	0.9679175
	AI + **BM** + C	-38.6	0.158529
	BIO5	-37.4	0.9680426
	BIO1	-35.2	0.9748651
	AI +BIO1 + BIO5 + BIO6 + BIO7	-34.5	0.8533049
	BIO5 + **BM** + **Di**	-34.4	0.8907402
	BIO1 + **BM** + **Di**	-31.3	0.9043004
	BIO5 + **BM** + M	-30.9	0.9517756
	BIO6	-30.0	0.9796061
	AI + **BM** + C + D	-29.9	-0.013942
	BIO1 + **BM** + M	-26.5	0.9606075
	BIO6 + **BM** + **Di**	-24.2	0.9170073
	BIO7	-22.2	0.9845628
	BIO6 + **BM** + M	-20.8	0.9655426
	AI + **BM** + C + M	-19.1	0.08665056
	AI + **BM** + C + M + **Di**	-16.3	-0.0235275
	BIO5 + **BM** + C	-14.7	0.9561875
	BIO1 + **BM** + C	-11.5	0.9658072
	BIO6 + **BM** + C	-6.72	0.9712682
	AI + **BM** + C + M + **Di** + BIO1 + BIO5 + BIO6 + BIO7	10.02	-0.0165843
	BIO5 + **BM** + C + M + **Di**	12.57	0.0488838
	BIO1 + **BM** + C + M + **Di**	18.32	0.8811903
	BIO6 + **BM** + C + M + Di	24.33	0.9169933
Dehydrated only (87 species)	**AI** + **BM**	-62.3	0.9294967
	AI	-54.4	0.9436382
	AI + **BM** + **Di**	-52.3	-0.0325474
	cAI + **BM**	-48.9	0.9504767
	AI + **BM** + M	-42.8	0.898354
	cAI	-42.8	0.9528914
	cAI + **BM** + **Di**	-42.7	-0.0327938
	BIO1 + **BM**	-40.7	0.99556
	BIO5 + **BM**	-39.5	0.985862
	AI + **BM** + BIO1 + BIO5 + BIO6 + BIO7	-39.1	0.9680209
	BIO6 + **BM**	-38.0	1.000221
	AI + **BM** + M + **Di**	-34.7	-0.0227454
	BIO1	-34.0	0.9969119
	BIO7 + **BM**	-33.4	1.002295
	BIO5	-32.3	0.9876165
	BIO6	-32.3	1.001864
	AI + BIO1 + BIO5 + BIO6 + BIO7	-30.5	0.9716893
	cAI + **BM** + M	-29.5	0.9392195
	BIO7	-28.6	1.004338
	BIO1 + **BM** + **Di**	-27.5	0.9877439
	BIO5 + **BM** + **Di**	-26.6	0.9676973

	cAI + **BM** + M + **Di**	-25.0	-0.0193277
	cAI + **BM** + BIO1 + BIO5 + BIO6 + BIO7	-24.2	0.9664614
	BIO6 + **BM** + **Di**	-23.9	0.9957233
	AI + **BM** + M + **Di** + BIO7	-23.9	-0.0339047
	BIO5 + **BM** + M	-20.6	0.9851374
	BIO1 + **BM** + M	-20.4	0.995414
	BIO7 + **BM** + **Di**	-17.6	0.9999706
	BIO6 + **BM** + M	-17.5	0.999279
	cAI + BIO1 + BIO5 + BIO6 + BIO7	-17.2	0.9684298
	BIO7 + **BM** + M	-13.4	1.001486
	cAI + **BM** + M + D + BIO7	-10.8	-0.0192356
	AI + **BM** + M + **Di** + BIO1 + BIO5 + BIO6 + BIO7	-8.97	-0.0326146
	BIO1 + **BM** + M + **Di**	-8.06	0.9871385
	BIO5 + **BM** + M + **Di**	-5.79	-0.0216553
	BIO6 + **BM** + M + **Di**	-4.36	0.9942829
	cAI + **BM** + M + **Di** + BIO1 + BIO5 + BIO6 + BIO7	0.40	-0.0227543
	BIO7 + **BM** + M+ **Di**	1.47	0.9991637

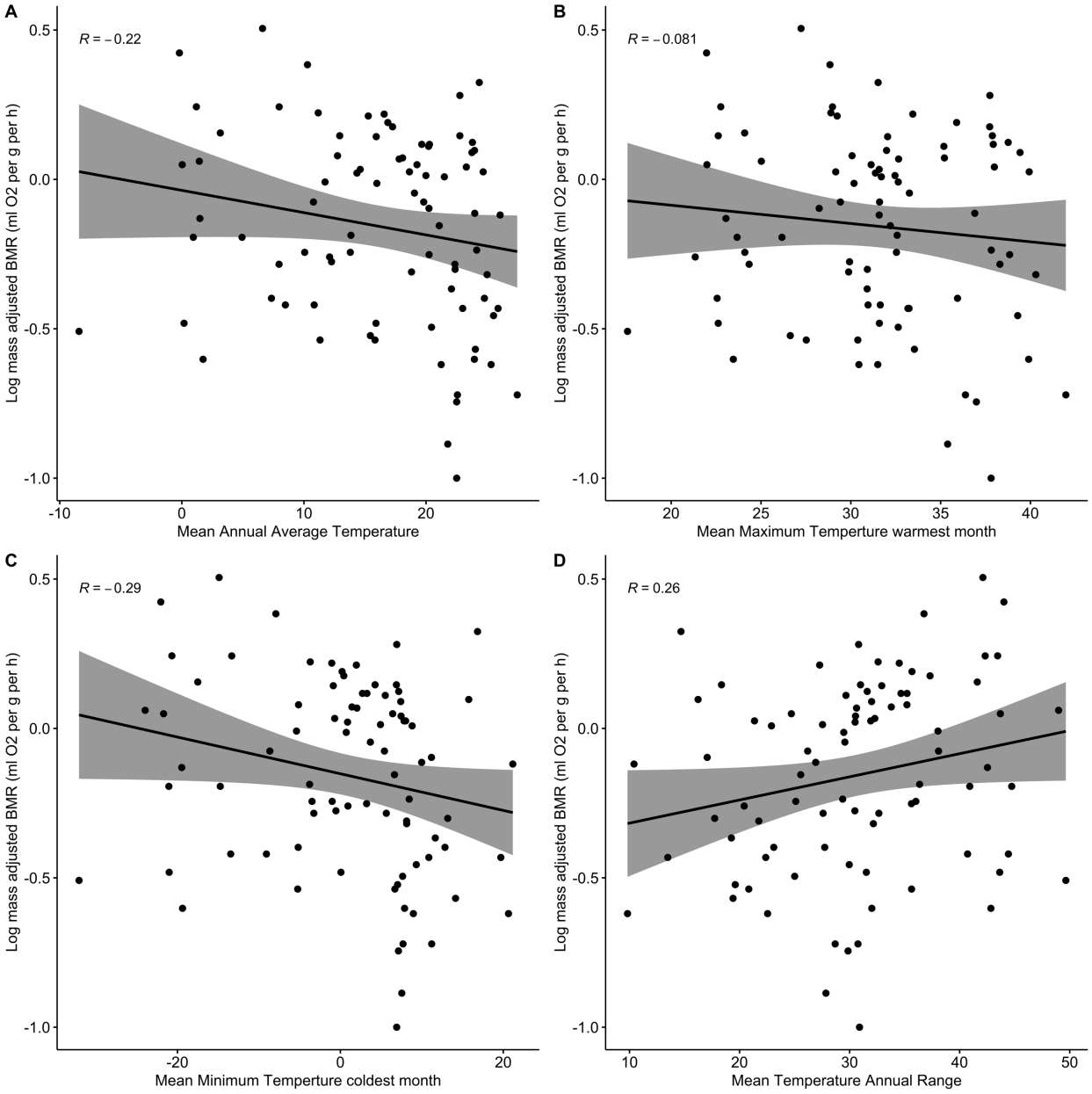
Variables are mean AI = annual aridity index (continuous), cAI = categorical AI (Arid, Semi-Arid, Dry sub-humid, Humid), BM = log10 body mass (kg), C = condition (dehydrated-D, fed-F, given salt load-S, given protein load-P, unknown-x and combinations of these), M = Method (unknown-?, solutes-A, Freezing point-F, Vapor pressure-V, and combinations of these), Di = Diet (Carnivorous-C, Frugivorous-Fr, Granivorous-G, Herbivorous-H, Insectivorous-I, Omnivorous-O, Carnivorous and Insectivorous-CI), and mean values for BIO1 = annual mean temperature, BIO5= maximum temperature of warmest month, BIO6 = minimum temperature of coldest month, BIO7 = temperature annual range.



Appendix S3. Spearman correlation for mean annual aridity index and four measurements of temperature variation. P-values are not presented as the plots do not take spatial correlation into account.



Appendix S4. Spearman correlation for log10 maximum urine osmolality (measured in mOsm/Kg) and four measurements of temperature variation. P-values are not represented as the plots do not take phylogenetic affinity into account.



Appendix S5. Spearman correlation for log10 mass-adjusted Basal Metabolic Rate (ml of O_2 per g per day) and four measurements of temperature variation. P-values are not represented as plots do not take phylogenetic affinity into account.

References

1.Beuchat CA (1990) Body size, medullary thickness, and urine concentrating ability in mammals. *American Journal of Physiology* 258: 298-308.

2. Lovegrove BG (2000) The zoogeography of mammalian basal metabolic rate. *American Naturalist* 156: 201–219.

3. White CR, Seymour RS (2003) Mammalian basal metabolic rate is proportional to body mass2/3. *Proceedings of the National Academy of Sciences of the United States of America* 100: 4046–4049.

4. Beuchat CA (1996) Structure and concentrating ability of the mammalian kidney: correlations with habitat. *American Journal of Physiology* 271: 157-179.

5. Diaz GB, Ojeda RA (1999) Kidney structure and allometry of Argentine desert rodents. *Journal of Arid Environments* 41: 453–461.

6. Maloiy G, Macfarlane W, Shkolnik A (1979) Mammalian herbivores. In: Maloiy G (ed.) *Comparative Physiology of Osmoregulation in Animals*, vol. 2, 185-209. Academic Press, London, UK.

7. Downs CT, Perrin MR (1991) Urinary concentrating ability of four *Gerbillurus* species of southern African arid regions. *Journal of Arid Environments* 20: 71–81.

8. Nagy KA, Shoemaker VH, Costa WR (1976) Water, electrolyte, and Nitrogen budgets of jackrabbits (*Lepus californicus*) in the Mojave Desert. *Physiological Zoology* 49: 351–363.

9. Kronfeld N, Shkolnik A (1996) Adaptation to life in the desert in the brown hare (*Lepus capensis*). *Journal of Mammalogy* 77: 171–178.

10. Ostrowski S, Williams JB, Mésochina P, Sauerwein H (2006) Physiological acclimation of a desert antelope, Arabian oryx (*Oryx leucoryx),* to long-term food and water restriction. *Journal of Comparative Physiology B: Biochemical, Systemic, and Environmental Physiology* 176: 191–201.

Chapter III.

Genomic insights into the adaptive history of North-African foxes

PAPER 3.

North-African fox genomes reveal signatures of ancient introgression and adaptation to life in deserts

Rocha J. et al.
(manuscript in preparation for submission)

Abstract

North-African desert-dwelling foxes (*Vulpes spp.*) have lived for a long time in the extreme conditions of the Sahara Desert, including high thermal amplitudes and water and food scarcity. We applied a genomic approach to investigate the genetic basis underlying the ability of different fox species to thrive in desert environments. We show that past gene flow between North-African red foxes (*Vulpes vulpes*), and desert-specialist Rueppell's fox (*V. rueppelli*), likely facilitated red fox recent adaptation to an expanding Sahara, after they split from Eurasian populations ca. 78 Kya. Further comparative analysis identified a highly unusual and divergent 25 Mb genomic region in Rueppell's fox that can be explained as the result of an ancient introgression from another fox related to the fennec (*V. zerda*). Scans for signatures of adaptation revealed signals at genes associated with water retention through reduced panting and sweating, temperature perception and adaptive thermogenesis in North-African red fox, and genes affecting renal water homeostasis in Rueppell's and Fennec fox, some of which were supported by expression and/or physiological differences. These discoveries suggest that both standing adaptive genetic variation and introgression with older resident desert species have been key drivers of desert adaptation in North-African foxes and are likely to be critical components of evolutionary responses to climate change.

Main

Despite multiple studies on the adaptation of dogs and wolves to local environments, including cold climates and high altitudes (Sinding et al. 2020; Wang et al. 2020), our understanding of the genetic basis of canid adaptation to hot desert environments remains limited. Foxes of the genus *Vulpes* are the only group of canids with geographic distributions that fully overlap with some of the most extreme deserts on Earth (Sillero-Zubiri et al. 2004). The Rueppell's fox (*V. rueppellii*) is a desert-

adapted species widely distributed in arid and hyper-arid regions of North-Africa and the Middle East (**Figure 1A**). In North-Africa, Rueppell's fox fully overlaps in range with a more distantly related congener, the fennec (*V. zerda*), and both species are remarkably adapted at coping with extreme temperatures and water scarcity (Sillero-Zubiri et al. 2004). The sister species to 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 (Leite et al. 2015). Rueppell's and red foxes are likely to be of Eurasian origin and overlap in range at the outer edges of the Middle Eastern deserts and at the northern fringes of the Sahara (**Figure 1A**) (Brito et al. 2009). Though the two sister species differ fundamentally in ecology, morphology, behavior and physiology, red fox desert morphs have been described and a previous study has alluded to past introgressive-hybridization events with Rueppell's fox in North-Africa (Sillero-Zubiri et al. 2004; Leite et al. 2015). 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.

Genomic dataset, phylogeny and population structure

To address the genetic basis underlying adaptation to hot desert environments, we sequenced the complete genomes of 82 foxes (~4.5x coverage), including 48 red foxes (North-Africa, *n*=30; Eurasia, *n*=18) and 24 Rueppell's (North-Africa) (**Figure 1A**). We additionally sequenced genomes for two North-African outgroup species, the fennec (*n*=5) and the pale fox (*V. pallida*; *n*=5) and used a suite of statistical methods targeted for low to medium coverage data (see supplementary Methods and **Table S1**) (Korneliussen et al. 2014).The arctic fox (*V. lagopus*) (Hasselgren et al. in press) and dog (*Canis lupus familiaris*) genomes were used as reference and outgroups (**Figure S1).**

Whole-genome phylogenetic analyses (**Figure 1B**) and mtDNA-based ancestral range reconstruction for several *Vulpes* (**Table S2, Figures S2-5, Tables S4-5)** supported a Eurasian origin for Rueppell's and red foxes with independent North-African colonization. These analyses further show that the two sister species share common ancestors with Asian foxes (**Figures S3-4, TableS5**) and the arctic fox (**Figure 1B**) after divergence from fennec and pale fox. Principal component and admixture analyses first separated Eurasian and North-African red foxes, and then Eurasian populations into Europe and Asia (**Figure 1C**, **Figures S6-7**). This suggests that gene flow has been more limited between North-Africa and Eurasian red foxes than among Eurasian populations, consistent with red fox isolation in North-Africa (**Figure 1**). We did not find evidence for geographic structure within North-African Rueppell's fox, suggesting genetic continuity throughout the Sahara (**Figures S6-7, Figure 1**).

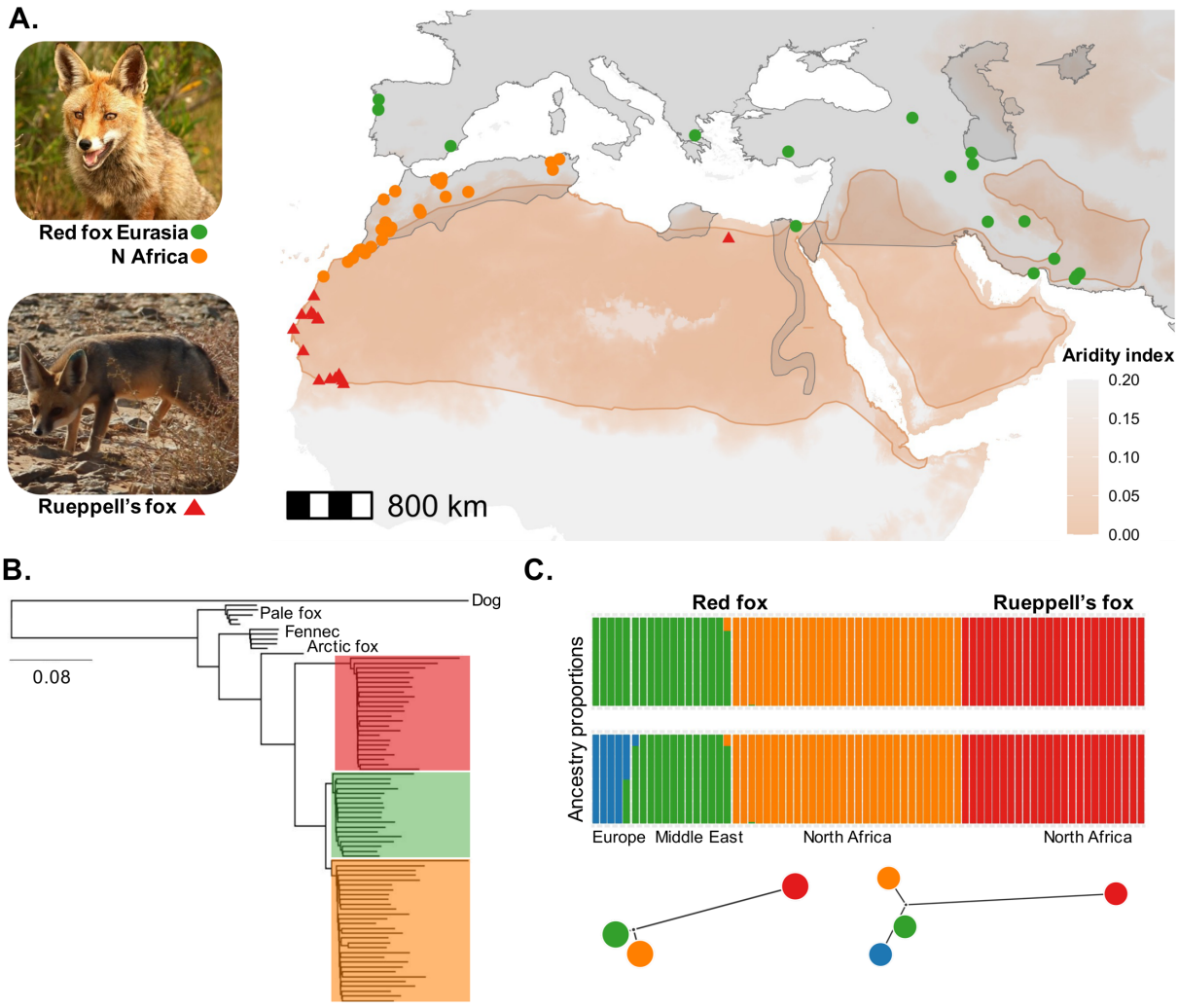


Figure 1. Sampling localities and overall species relationships. **(A)** Geographic location of the samples (points) and distribution ranges (lines) for Rueppell's and Red fox used for population genetic analyses (see **Table S1** for sampling details and **Figure S2**) [Photo credit: Ricardo Rocha, Monia Nakamura]. North-African red fox individuals were sampled across an aridity gradient ranging from dry-sub humid to hyper-arid regions. We used a color range for Aridity Index (AI) from 0 to 0.2, to define arid (0.05 < AI <0.2) and hyper-arid regions (AI<0.05) and demark them from non-desert areas (AI > 0.2). Dark grey lines delimit the distribution range of red foxes, and beige line border the distribution range of Rueppell's fox (**B**) RaxML whole-genome tree constructed with genetic distances among 83 foxes based on genotype likelihoods using the Dog as outgroup. Genetic distances were calculated with ngsDist with genome-wide polymorphic sites, filtering for a minor allele frequency of 0.05. This was performed through concatenation of genomic data, which typically leads to long external and short internal branches within species. (**C**) Admixture proportions inferred with Ohana structure analysis and trees for K=3 (top structure plot, left tree) and K=4 (bottom structure plot, right tree). Each bar shows the inferred ancestry fraction for an individual. Labels above and below bars identify the species and geographical location.

Introgressive-hybridization events

Previous studies have shown that introgressive-hybridization events are pervasive in canids, and may extend far deeper than between pairs of closely related species (Gopalakrishnan et al. 2018; Wang et al. 2020). We used *D* statistics (Soraggi et al. 2018) and TreeMix admixture graphs (Pickrell

and Pritchard 2012) (**Figures S8-9**) to test for introgression among North-African foxes. We found significant excess of allele sharing between North-African red foxes and Rueppell's foxes relative to Eurasian red foxes, providing evidence of past interspecific gene flow (**Table S6)**. Past introgression between North-African Rueppell's and red foxes was also supported by inference of their joint demographic history using *GADMA* (Noskova et al. 2020) and *dadi* (Gutenkunst et al. 2009), which showed nonzero two-way asymmetric migration after their divergence ~576 kya (**Figure S10, Table S7**). Remarkably, D statistics and admixture graphs further suggested introgression between Rueppell's fox and the highly divergent fennec fox (**Table S6, Figure S8**). Introgression in arid regions could have facilitated evolutionary responses to climate change, as well as the long-term maintenance of adaptive variation among ecologically similar and geographically overlapping desert-dwelling species (**Figures 1A, S2A**). To identify potentially introgressed regions, we calculated $f\hat{}_d$ (Martin et al. 2015) and Z(D) in non-overlapping windows of 200 kb (**Figure 2, Figures S11-12**). This revealed several genomic regions shared between North-African Rueppell's and red foxes (**Figure 2A**), and a large ~25 Mb region shared between Rueppell's fox and fennec (**Figure 2B**). 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 $f\hat{}_d$ (above the 99th percentile; **Figure 2A, Tables S8-9, Figures S13-S15**) and for the 25Mb region (**Figure 2B, Table S10-S11; Figure S16**).

In contrast to the topology observed for the genome-wide tree, nearly all North-African red fox individuals appeared more closely related to Ruppell's fox than to other red foxes at SLC6A16, which shows the second strongest signal of introgression (**Figure 2A, Figures S14-15**). Though the gene-species tree discordance at SLC6A16 is consistent with significant evidence for genome-wide introgression, it could also be explained by incomplete lineage sorting (Hedrick 2013; Marques et al. 2019). To disentangle these hypotheses, we compared levels of genome-wide differentiation (F_{ST}) and genetic divergence (d_{xy}) between Rueppell's and North-African red foxes with those estimated at introgressed regions (**Figures S17-19**). Consistent with an introgression scenario, F_{ST} and d_{xy} levels across SLC6A16 were remarkably lower between North-African red and Rueppell's foxes (below 0.5% of the genome-wide empirical distribution) (**Figures S17-19**). SLC6A16 is a member of a family of neurotransmitter transporters involved in multiple biological processes and has been implicated in chronic kidney disease and cellular response to oxidative stress (Pramod et al. 2013; Ainsworth et al. 2017). 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 d_{xy}. (**Figure 2A,C; Figures S14-15, Figures S17-19**). Unexpectedly, we found a very rare topology at the CRADD gene (top outlier in **Figure 2A**), in which some Eurasian red foxes and the arctic fox formed a highly divergent outgroup to *Vulpes* (**Figure S13**). The introgression signal observed between Ruppell's fox and North-African red foxes can therefore be explained by the presence of deeply divergent alleles predating the origin of all *Vulpes* at the CRADD

region and reflects a complex evolutionary history in Eurasian red foxes and arctic fox. CRADD is associated with differences in body size in mice and humans (Kimura et al. 2008). Likely, the arctic fox and affected Eurasian red foxes could have received their divergent CRADD haplotypes via ancient introgression from a highly divergent unknown canid lineage (**Figure S13**). Alternatively, the observed gene-species tree discordance in Eurasian red foxes could reflect an ancestral trans-species polymorphism maintained by balancing selection. However, this hypothesis would require multiple losses of one of the haplotypes in different species and strong recombination suppression in the region (e.g. by means of a structural rearrangement such as an inversion) (Jamie and Meier 2020).

Next, we explored the genomic region shared between fennec and Rueppell's foxes (**Figure 2B**). Enrichment analysis on the entire 25 Mb outlier region identified several gene ontology (GO) terms that were overrepresented (**Table S11;** FDR p-value < 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 behavior (NMU), and pigmentation (KIT). The gene tree for this region unambiguously grouped Rueppell's foxes as sister species to the fennec (**Figure 2C, Figure S16**). However, the pairwise divergence between the two desert-dwelling species is not much different from the genome-wide average **(Figures S20-21**), suggesting that introgression did not happen directly between fennec and Rueppell's fox but, instead, from a fennec-related donor species. Sliding window analyses at the affected 25 Mb region further revealed a complex pattern in which Rueppell's and red fox revert to being sister taxa at the KIT gene (**Figures S22-23**). KIT is a well-known case of an autosomal gene with paralogues located on red foxes supernumerary chromosomes (Bs), which may vary in number across individuals (Makunin et al. 2014). Consistent with this, we found that red and Rueppell's fox individuals carry different numbers of copies of KIT that were strongly correlated with the copy number variation pattern of other autosomal genes (**Figure S22**). Under a scenario of introgression from another species we would expect KIT to be highly polymorphic in Rueppell's foxes, as the result of divergent autosomal alleles and Rueppell's unaffected supernumerary alleles mapping in the same region. Consistent with this hypothesis, Rueppell's showed a peak in genetic diversity in KIT not found in other foxes (**Figure S23**). Past correspondence analysis using chromosome painting map suggest the affected 25Mb region (**Figure 2B**) to chromosomes 2p in red fox, 12p in arctic fox and part of 9 in fennec (Graphodatsky et al. 2008). A major structural rearrangement, such as the translocation of a chromosomal arm, could therefore explain how divergent loci at this region were maintained in strong linkage disequilibrium. Additional *Vulpes* long-read-based genome assemblies are needed to resolve the evolutionary history of this large chromosomal region.

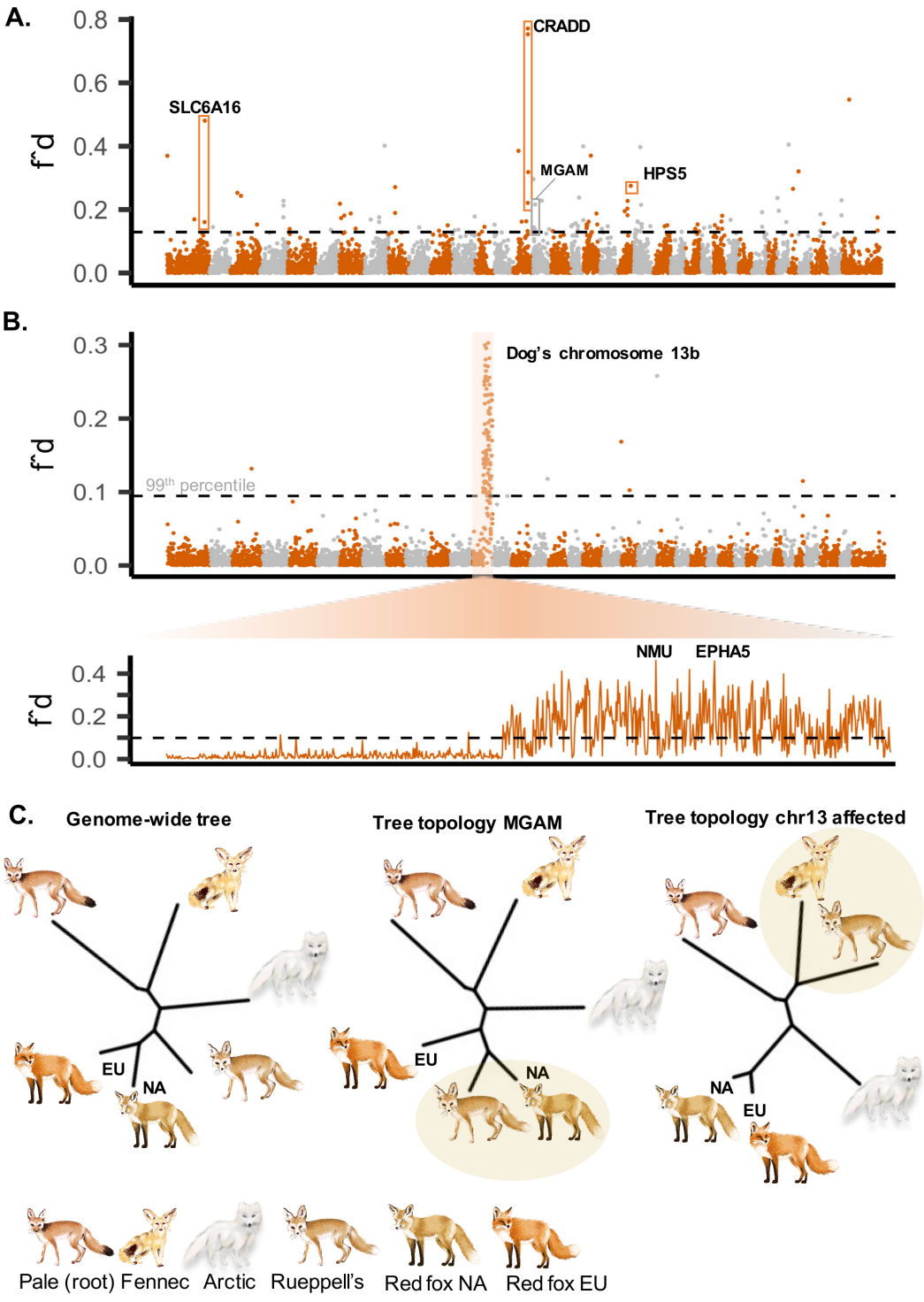


Figure 2. Introgression in North African foxes. Manhattan plot of fraction of admixture ($f\hat{}_d$) on non-overlapping windows of 200kb across all chromosomes (**A**) Eurasian red foxes were used as H1, North African red foxes as H2 and Rueppell's fox as H3, with the arctic fox as outgroup; (**B**) All red foxes were used as H1, Rueppell's foxes as H2, and fennec fox as H3, with the pale fox as outgroup; this is followed by a zoom-in view on 50kb windows (some examples highlighted). Windows from the same chromosome have the same color. Dark line threshold separates regions above the 99th percentile. We note that other genes not displayed in the figure can overlap the outlier regions. The arctic fox was used as reference genome and scaffolds were ordered with respect to the dog (**Figure S1**). (**C**) Estimated tree topologies for *Vulpes*: whole-genome tree (left), 200kb region harboring gene MGAM (center), and 4 scaffolds in the arctic fox genome (scaffold 47,155,126, 200) above the 99th percentile of the empirical distribution (right). Regions marked with a rectangle were selected for further investigations of gene-species tree discordance.

North-African colonization and adaptation in red foxes

After removing introgressed regions, we sought to model the joint demographic history of North-African and Eurasian red foxes. The model with highest composite likelihood found using *GADMA* (Noskova et al. 2020) and *dadi* (Gutenkunst et al. 2009) dates the divergence of the two populations to ~78 kya after the end of a warm interglacial of the Marine Isotope Stage 5 (MIS5), which spanned 130-80 kya (**Figure S10, Table S7**). In North-Africa, MIS5 has been associated with a "green Sahara" period that enabled human expansion inside the continent (Osborne et al. 2008; Hoffmann et al. 2016). Such a humid period could have also 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 Eurasia and introgression with North-African Rueppell's. The model also supports asymmetrical two-way migration after divergence, suggesting migration in subsequent African humid periods (**Figure S10**). The absence of significant admixed individuals in structure plots (**Figure S7**) suggests this is not a recent or ongoing phenomenon and, thus, red foxes likely remained isolated ever since the Sahara attained its current desert profile.

Red foxes living in increasingly arid environments at the Saharan northern edge for numerous generations may have faced unique selective pressures (**Figure 1**). To investigate signatures of positive selection specific to North-African red foxes, we looked for genomic regions either highly differentiated (Population branch statistic, PBS) and/or with reduced diversity (*Alpha*, *α*) in this population relative to Eurasian red foxes and Rueppell's foxes. A sliding window analysis revealed several genomic regions with high PBS and -log10(*α*), hinting at selection in North-African red foxes (**Figure 2**). We then focused on the most extreme values of the empirical distribution (above the 99.95% percentile) in both methods and identified overlapping genes (**Table S12 to S14**). Then, we looked for expression differences across blood samples (*n*=3 Rueppell's, *n*=2 red foxes North-Africa, *n*=5 Iberia) for major outlier genes (**Table S3, Table S15**, **Figure S24 to S26**).

While some of the selection signals unique to North-African red foxes may be related to other environmental factors, such as local pathogens (e.g. CD163, ranked 1st in -log10(α), **Figure 3, Figure S26**), top hits overlapped with candidate genes associated with relevant traits for life in arid regions. For instance, the second largest increase in -log10(α), was SLC12A2, a sodium chloride transporter that has been implicated in dry mouth disease in humans, and lack of saliva and sweat in knockout mice (**Figure 3**) (Evans et al. 2000; Stödberg et al. 2020). The strong selection signal in SLC12A2 may be related to selection for changes in expression as this gene is significantly less expressed in North-African red foxes and Rueppell's fox than Eurasian red foxes (**Figure S26;** log2fold = -1.13, p-adjust$_{(Africa\ vs\ Iberia)}$= 0.00113). Though this could also reflect plastic responses to different environmental cues, we hypothesize that selection acted on SLC12A2 in North-African red foxes to reduce the amount of evaporative water loss through panting (salivary water loss), the main mechanism for cooling in canids, or sweating. Consistent with this, enrichment analysis on top regions

with reduced diversity identified several biological processes in North-African red foxes 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 p-values > 0.05; see Table S14 and Tables S16-18**).

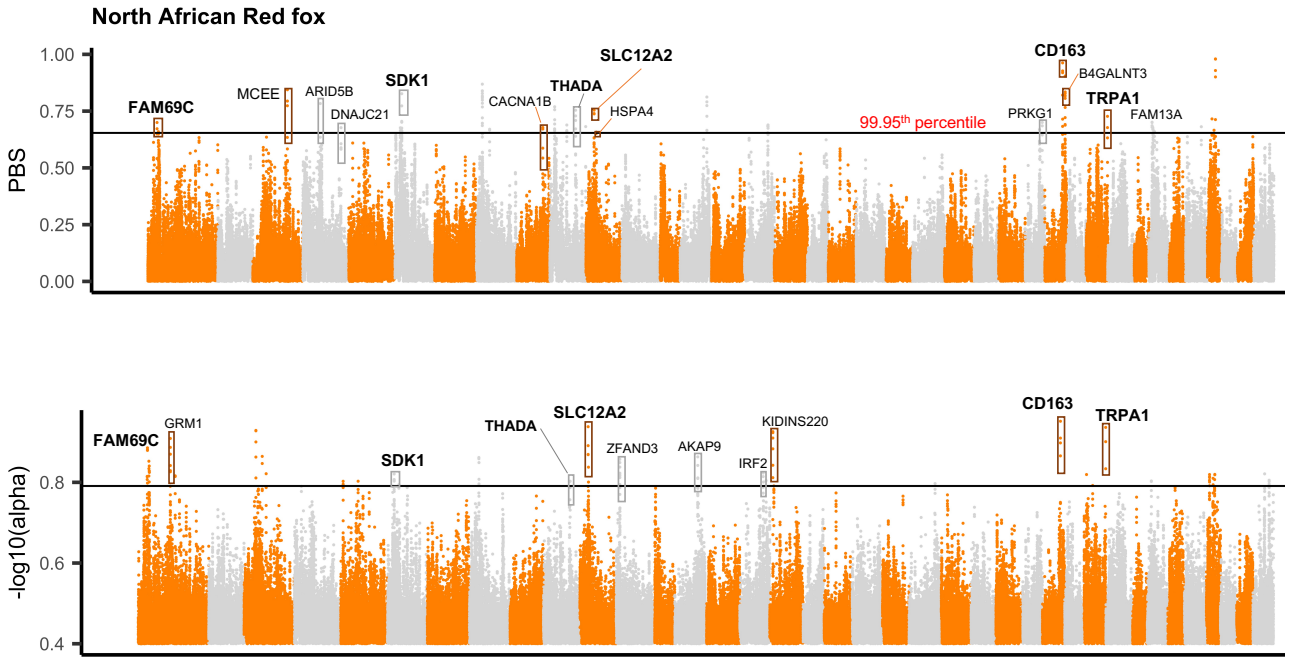


Figure 3. Selection specific to North-African Red foxes. Genome-wide selection scans on windows of 50 kb with a 10kb slide across autosomal chromosomes (x-axis) using North African Red foxes as focal group, Eurasian red foxes as closest ingroup and Rueppell's fox as outgroup. Dashed lines show the 99.95th percentile of the empirical distribution population branch statistics (PBS) and negative logarithm of alpha (α). Windows from the same chromosome have the same color. 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 (**Figure S1**). We note that other genes not displayed in the figure can overlap the outlier regions; a full list can be found in **Tables S12-S13**.

Though arid regions in North-Africa are classified as 'hot', they may experience extreme daily temperature oscillations with drops of temperature at night that can go down to -0.5°C (Ward 2016). Another highly differentiated and low diversity gene (ranked 3rd in -log10(α)) is TRPA1, encoding for a transient receptor potential (thermoTRP) channel protein that is gated by noxious cold or heat and also acts as sensor for environmental irritants (Sinica et al. 2019; Xu et al. 2019) (**Figure 3**). Several thermoTRP sensors have been reported under selection in arctic sled dogs (Sinding et al. 2020) and mammoths (Lynch et al. 2015) and associated with altered temperature sensation to increase cold tolerance in high-latitude environments. North-African foxes are nocturnal, with denning behavior during the day (Sillero-Zubiri et al. 2004). We hypothesize that reduced thermosensitivity to both extreme cold and heat can be adaptive in arid regions with high isothermality, and North-African red

foxes' genetic variation has likely been targeted by selection on phenotypes linked to bidirectional thermoception. Notably, the gene coding for thyroid adenoma-associated protein, THADA, involved in heat production and adaptive thermogenesis, and widely associated with climate adaptation in humans (Moraru et al. 2017), is also under selection in North-African red foxes (ranked 14th in PBS, **Tables S12-13**). As the top SNPs with highest PBS values for candidate genes lie in non-coding regions (**Figure S28**), we suggest that selection in North-African red foxes targeted variants affecting the expression of genes involved in functions beneficial to physical activity at low temperatures at night. Interestingly, we found a region harboring ZNF516, a gene involved in cold-induced thermogenesis (Dempersmier et al. 2015), to be under selection in Eurasian red foxes (**Figure S27, Table S16**).

Signatures of selection in Rueppell's fox and the fennec

Our results suggest that, in addition to potential adaptive introgression from Rueppell's fox, North-African red foxes have unique adaptations related to tolerance of extreme temperatures and water retention that may have increased survival chances in the increasingly arid regions at the Saharan fringes over the past ca. 78 kya. 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 much longer time periods such as Rueppell's fox (476-788 kya, **Table S5**) and the fennec (4.8-6.2 Mya, **Table S5**). To address this, we performed a genome-wide selection scan using a PBS-like method implemented in *Ohana* (Cheng et al. 2019) that can be applied to several ancestral components. We run *Ohana* on a sliding window to detect regions that strongly deviate in Rueppell's 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) (**Figure 4**). We found signals of selection in genes related to renal water homeostasis in both fennec (KIRREL1, CACNA1H) and Rueppell's (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 UV radiation (PTPRK) in Rueppell's (**Figure 4A, Tables S19-21**).

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 (Kurimoto et al. 2021). Remarkably, WFS1 not only is under selection in Rueppell's (ranked 2nd) but is also more highly expressed than in red foxes (log2fold change = 1.83, p-adjust$_{(\text{Rueppell's vs Red foxes})}$= 0.0013) (**Figure 4A,C**). To investigate further, we sought to test if Rueppell's had significant water-retention abilities compared to red foxes from North-Africa and Eurasia by looking into physiological parameters on blood samples and urine concentrating abilities (**Table S3, Tables S22-23, Figure S29-31, Text S1**). Consistent with this, we found that levels of copeptin, a biomarker for AVP, were significantly higher in Rueppell's compared to red foxes (**Table S23, Figure S29**). We thus hypothesize that mutations changing expression of WFS1 may explain increased urine concentrating abilities in Rueppell's

(**Figure S31**). Notably, WFS1 gene is within the 25 Mb region introgressed into Rueppell's genome (**Figure 2B**), which suggests the species may have inherited increased capacity to retain water under dehydration from an unidentified desert species (**Figure 4C**). Fennecs also have higher AVP levels and seem to concentrate more urine than red foxes (**Tables S22-23, Figures S29-31**). The gene with the highest selection signal in fennec, KIRREL1 (**Figure 4A, Table S20-21**), encodes a protein produced in kidney podocytes that plays a significant role in the normal function of the glomerular permeability. In humans, mutant KIRREL1 proteins fail to localize the podocyte cell membrane, resulting in nephrotic syndrome (Solanki et al. 2019). KIRREL1 is therefore a likely candidate gene that may be related to fennecs' increased kidney function (Noll-Banholzer 1979).

Collectively, these results suggest that selection in older desert-dwelling species may have targeted renal water-retention and kidney function to a higher degree than what we observe in desert-edge dwelling North-African red foxes which, presumably, have more access to water. Rueppell's and fennec also had significantly lower levels of thyroid hormone thyroxine (T4) than red foxes (**Figure S29, Table S23**), which may potentially underlie previous evidence of reduced metabolism and non-renal water losses in these species (Maloiy et al. 1982; Williams et al. 2004) (**Text S2**). 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 (ranked 1st). We also noted that the 25 Mb region likely introgressed into Rueppell's was enriched for sulfotransferase genes (SULTs, FDR p-value= 0.017), one of which involved in thyroid hormone metabolism (**Table S11**). These, and many of the candidate genes for desert adaptation identified in this study, remain promising targets for future studies.

Conclusions

Despite decades of studies on canid evolution, North-African foxes constitute an entirely new system on adaptation to extreme climatic conditions not previously explored. We have shown that ancient introgressive events combined with independent selection on adaptive tolerance to extreme temperatures and increased capacity to preserve water have enabled North-African foxes' survival in desert regions at different evolutionary scales. 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 (Liu and Xue 2020). We expect ancestral standing adaptive genetic variation and introgression of alleles pre-tested by natural selection in core desert species to be crucial components of ongoing adaptation to rapidly changing climates in foxes living at the edge of deserts. Preserving adaptive variation in these charismatic umbrella species may help conservation efforts focused on rescuing biodiversity currently challenged by desertification.

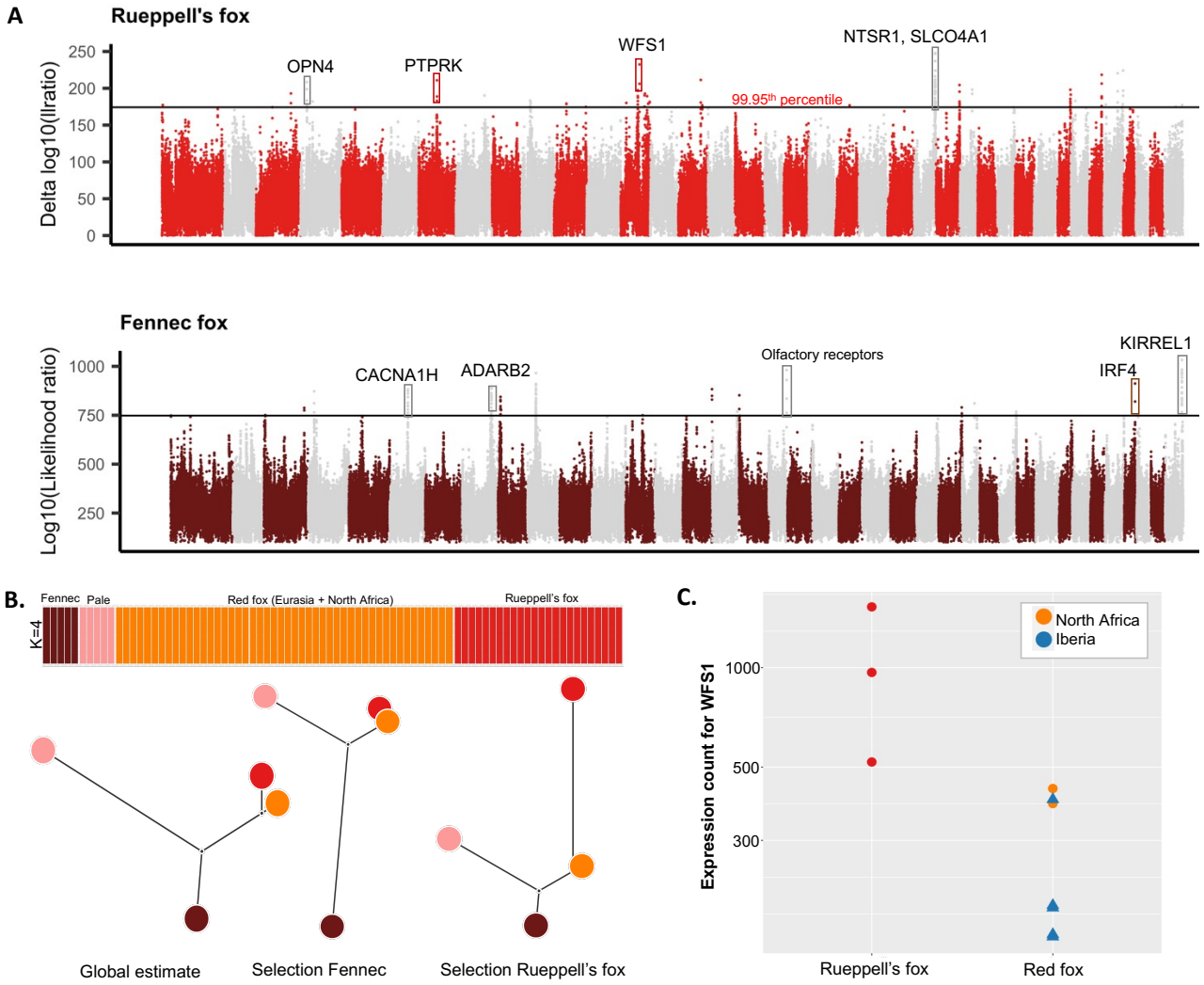


Figure 4. Genome-wide selection scans for Rueppell's and fennec fox using Ohana based on the structure and tree results from K = 4. (**A**) Manhattan plot of the log-likelihood ratio values for the top 500 SNPs on windows of 50 kb with a 10kb slide across chromosomes (x-axis) for Rueppell's fox and fennec fox (scaffolds are ordered with respect to the dog genome with each color representing a different chromosome). To account for selection specific to Rueppell's and not in their common ancestral with red foxes we subtracted the log-likelihood ratio of Rueppell's and Red fox for the same windows (delta log-likelihood ratio). Dashed 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 **Tables S19-20)**. (**B**) Representation of the selection study performed in (A): structure plot corresponding to four *Vulpes*, followed by the global tree and representation of the likelihood ratio test implemented in *Ohana*. 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). (**C**) WFS1 mRNA expression levels in blood of Rueppell's (*n*=3) and red foxes (*n*=2 North Africa, *n*=5 Iberia).

References

Ainsworth H.C., Langefeld C.D., Freedman B.I. 2017. Genetic epidemiology in kidney disease. Nephrol. Dial. Transplant. 32:ii159–ii169.

Brito J.C., Acosta A.L., Álvares F., Cuzin F. 2009. 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.

Cheng J.Y., Racimo F., Nielsen R. 2019. Ohana: Detecting selection in multiple populations by modelling ancestral admixture components. bioRxiv.:546408.

Dempersmier J., Sambeat A., Gulyaeva O., Paul S.M., Hudak C.S.S., Raposo H.F., Kwan H.Y., Kang C., Wong R.H.F., Sul H.S. 2015. Cold-inducible Zfp516 activates UCP1 transcription to promote browning of white fat and development of brown fat. Mol. Cell. 57:235–246.

Evans R.L., Park K., Turner R.J., Watson G.E., Nguyen H.-V., Dennett M.R., Hand A.R., Flagella M., Shull G.E., Melvin J.E. 2000. Severe Impairment of Salivation in Na+/K+/2Cl− Cotransporter (NKCC1)-deficient Mice. J. Biol. Chem. 275:26720–26726.

Gopalakrishnan S., Sinding M.H.S., Ramos-Madrigal J., Niemann J., Samaniego Castruita J.A., Vieira F.G., Carøe C., Montero M. de M., Kuderna L., Serres A., González-Basallote V.M., Liu Y.H., Wang G.D., Marques-Bonet T., Mirarab S., Fernandes C., Gaubert P., Koepfli K.P., Budd J., Rueness E.K., Heide-Jørgensen M.P., Petersen B., Sicheritz-Ponten T., Bachmann L., Wiig Ø., Hansen A.J., Gilbert M.T.P. 2018. Interspecific Gene Flow Shaped the Evolution of the Genus Canis. Curr. Biol. 28:3441-3449.e5.

Graphodatsky A.S., Perelman P.L., Sokolovskaya N. V., Beklemisheva V.R., Serdukova N.A., Dobigny G., O'Brien S.J., Ferguson-Smith M.A., Yang F. 2008. Phylogenomics of the dog and fox family (Canidae, Carnivora) revealed by chromosome painting. Chromosom. Res. 16:129–143.

Gutenkunst R.N., Hernandez R.D., Williamson S.H., Bustamante C.D. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet. 5:1000695.

Hedrick P.W. 2013. Adaptive introgression in animals: Examples and comparison to new mutation and standing variation as sources of adaptive variation. Mol. Ecol. 22:4606–4618.

Hoffmann D.L., Rogerson M., Spötl C., Luetscher M., Vance D., Osborne A.H., Fello N.M., Moseley G.E. 2016. Timing and causes of North African wet phases during the last glacial period and implications for modern human migration. Sci. Rep. 6:1–7.

Jamie G.A., Meier J.I. 2020. The Persistence of Polymorphisms across Species Radiations. Trends Ecol. Evol. 35:795–808.

Kimura R., Ohashi J., Matsumura Y., Nakazawa M., Inaoka T., Ohtsuka R., Osawa M., Tokunaga K. 2008. Gene Flow and Natural Selection in Oceanic Human Populations Inferred from Genome-Wide SNP Typing. Mol. Biol. Evol. 25:1750–1761.

Korneliussen T.S., Albrechtsen A., Nielsen R. 2014. ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics. 15:356.

Kurimoto J., Takagi H., Miyata T., Hodai Y., Kawaguchi Y., Hagiwara D., Suga H., Kobayashi T., Sugiyama M., Onoue T., Ito Y., Iwama S., Banno R., Tanabe K., Tanizawa Y., Arima H. 2021. Deficiency of WFS1 leads to the impairment of AVP secretion under dehydration in male mice.

Pituitary. 1:3.

Leite J.V., Álvares F., Velo-Antón G., Brito J.C., Godinho R. 2015. 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.

Liu Y., Xue Y. 2020. Expansion of the Sahara Desert and shrinking of frozen land of the Arctic. Sci. Rep. 10:1–9.

Lynch V.J., Bedoya-Reina O.C., Ratan A., Sulak M., Drautz-Moses D.I., Perry G.H., Miller W., Schuster S.C. 2015. Elephantid Genomes Reveal the Molecular Bases of Woolly Mammoth Adaptations to the Arctic. Cell Rep. 12:217–228.

Makunin A.I., Dementyeva P. V., Graphodatsky A.S., Volobouev V.T., Kukekova A. V., Trifonov V.A. 2014. Genes on B chromosomes of vertebrates. Mol. Cytogenet. 7:1–10.

Maloiy G.M.O., Kamau J.M.Z., Shkolnik A., Meir M., Arieli R. 1982. Thermoregulation and metabolism in a small desert carnivore: the Fennec fox (Fennecus zerda) (Mammalia). J. Zool. 198:279–291.

Marques D.A., Meier J.I., Seehausen O. 2019. A Combinatorial View on Speciation and Adaptive Radiation. Trends Ecol. Evol. 34:531–544.

Martin S.H., Davey J.W., Jiggins C.D. 2015. Evaluating the Use of ABBA–BABA Statistics to Locate Introgressed Loci. Mol. Biol. Evol. 32:244–257.

Moraru A., Cakan-Akdogan G., Strassburger K., Males M., Mueller S., Jabs M., Muelleder M., Frejno M., Braeckman B.P., Ralser M., Teleman A.A. 2017. THADA Regulates the Organismal Balance between Energy Storage and Heat Production. Dev. Cell. 41:72-81.e6.

Noll-Banholzer U. 1979. Water balance and kidney structure in the fennec. Comp. Biochem. Physiol. -- Part A Physiol. 62:593–597.

Noskova E., Ulyantsev V., Koepfli K.P., O'brien S.J., Dobrynin P. 2020. GADMA: Genetic algorithm for inferring demographic history of multiple populations from allele frequency spectrum data. Gigascience. 9:1–18.

Osborne A.H., Vance D., Rohling E.J., Barton N., Rogerson M., Fello N. 2008. A humid corridor across the Sahara for the migration of early modern humans out of Africa 120,000 years ago. Proc. Natl. Acad. Sci. U. S. A. 105:16444–16447.

Pickrell J.K., Pritchard J.K. 2012. Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. PLoS Genet. 8:1002967.

Pramod A.B., Foster J., Carvelli L., Henry L.K. 2013. SLC6 transporters: Structure, function, regulation, disease association and therapeutics. Mol. Aspects Med. 34:197–219.

Sillero-Zubiri C., Hoffmann M., Macdonald D.W. 2004. Canids: Foxes, Wolves, Jackals and Dogs. Status survey and conservation action plan. .

Sinding M.H.S., Gopalakrishnan S., Ramos-Madrigal J., Manuel M. De, Pitulko V. V., Kuderna L., Feuerborn T.R., Frantz L.A.F., Vieira F.G., Niemann J., Castruita J.A.S., Carøe C., Andersen-

Ranberg E.U., Jordan P.D., Pavlova E.Y., Nikolskiy P.A., Kasparov A.K., Ivanova V. V., Willerslev E., Skoglund P., Fredholm M., Wennerberg S.E., Heide-Jørgensen M.P., Dietz R., Sonne C., Meldgaard M., Dalén L., Larson G., Petersen B., Sicheritz-Pontén T., Bachmann L., Wiig Ø., Marques-Bonet T., Hansen A.J., Gilbert M.T.P. 2020. Arctic-adapted dogs emerged at the Pleistocene-Holocene transition. Science (80-.). 368:1495–1499.

Sinica V., Zimova L., Barvikova K., Macikova L., Barvik I., Vlachova V. 2019. Human and Mouse TRPA1 Are Heat and Cold Sensors Differentially Tuned by Voltage. Cells. 9:57.

Solanki A.K., Widmeier E., Arif E., Sharma S., Daga A., Srivastava P., Kwon S.H., Hugo H., Nakayama M., Mann N., Majmundar A.J., Tan W., Gee H.Y., Sadowski C.E., Rinat C., Becker-Cohen R., Bergmann C., Rosen S., Somers M., Shril S., Huber T.B., Mane S., Hildebrandt F., Nihalani D. 2019. Mutations in KIRREL1, a slit diaphragm component, cause steroid-resistant nephrotic syndrome. Kidney Int. 96:883–889.

Soraggi S., Wiuf C., Albrechtsen A. 2018. Powerful inference with the D-statistic on low-coverage whole-genome data. G3 Genes, Genomes, Genet. 8:551–566.

Stödberg T., Magnusson M., Lesko N., Wredenberg A., Martin Munoz D., Stranneheim H., Wedell A. 2020. SLC12A2 mutations cause NKCC1 deficiency with encephalopathy and impaired secretory epithelia . Neurol. Genet. 6:e478.

Wang M.S., Wang S., Li Y., Jhala Y., Thakur M., Otecko N.O., Si J.F., Chen H.M., Shapiro B., Nielsen R., Zhang Y.P., Wu D.D. 2020. Ancient hybridization with an unknown population facilitated high-altitude adaptation of canids. Mol. Biol. Evol. 37:2616–2629.

Ward D. 2016. The biology of deserts. Oxford University Press.

Williams J.B., Muñoz-Garcia A., Ostrowski S., Tieleman B.I. 2004. A phylogenetic analysis of basal metabolism, total evaporative water loss, and life-history among foxes from desert and mesic regions. J. Comp. Physiol. B Biochem. Syst. Environ. Physiol. 174:29–39.

Xu M., Zhang Y., Wang M., Zhang H., Chen Y., Adcock I.M., Chung K.F., Mo J., Zhang Y., Li F. 2019. TRPV1 and TRPA1 in lung inflammation and airway hyperresponsiveness induced by fine particulate matter (PM2.5). Oxid. Med. Cell. Longev. 2019.

Supporting information

Material and 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, Middle East and Europe (see **Table S1** and **Table S3** for details). 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 in Morocco, and Dakhla-Oued Ed-Dahab, Western Sahara. 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 frequented by foxes to set camera-traps, namely through personal observations of live and/or road-killed animals, compatible dens, tracks and feces.

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. The three types of traps were combined to maximize capture success and selected based on its efficiency and selectivity in capturing foxes while ensuring the safety of captured animals. All traps were set using odor-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 pre-baited for one night before being activated in order to encourage animals to explore the traps without being captured the first time (https://www.youtube.com/watch?v=Bj3nsLKdbic). Both Collarum and padded foothold traps were set coupled to remote satellite trap-alarms, which send a GSM (Global System Mobile) message to a predefined telephone each time the trap is activated. 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 (**Table S1**, **Table S3**).

A total of 36 animals (free-ranging or captive-held; **Table S3**) were anesthetized by intramuscular injection of a mixture of ketamine (Imalgene®, Merial, Lyon, France) and medetomidine (Domitor®, Merial, Lyon, France) by a certified wildlife veterinary. The animals were weighted, sexed, and measured for age (juvenile <1-year vs adult > 1-year), body, hindfoot, tail, and ear length. The health status of the individual was checked every 10 minutes during the whole procedure. Blood was recovered by venipuncture of the arm veins 30 minutes after administration of the anesthetic and preserved in both clotting and EDTA tubes. Serum/plasma was separated from blood cells through centrifugation at 1500g for 10 minutes 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 approximately 8-72h after captures. If the body mass of the animal was sufficiently high, additional whole-blood was collected and stored in PAXgene® Blood RNA tubes for RNA extraction (Iberian red foxes, *n*=5; North-African red foxes, *n*=2; Rueppell's fox, *n*=4). These tubes were stored at room temperature (RT) for no more than 2 days, followed by refrigerated storage at -20 and then -80 degrees Celsius (°C). Free-ranging (wild) individuals were additionally ear punched using an 8mm 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 to control for potential recapture, and individuals were safely released into their environment. Chemical immobilization was reversed by the intramuscular injection of atipamezole (Revertor®, Merial, Lyon, France) after administration of an anti-inflammatory.

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 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 (**Table S1**).

The geographical location of all samples included in this study were recorded with a global positioning system (GPS) on the WGS84 datum (**Figure 1A; Figure S2A**). 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 Center 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 re-capture of two individuals. All sample exportation/importation was performed under declaration of Institute for the Conservation of Nature and Forests (ICNF), and CITES listed specimens were accompanied with an appropriate CITES permit.

Phenotyping

In total, 34 individual *Vulpes* samples were phenotyped in this study (*n*=13, Rueppell's foxes; *n*=11, red foxes; *n*= 10, fennecs; **Table S3**). This comprised 21 males (61.7%) and 13 (38.2 %) females evenly distributed within species, most of which were adults (**Table S3**). Blood samples were analyzed 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, chloride), protein (albumin) and fat (cholesterol), plasma and urine osmolality (mOsm/Kg). Additionally, a Sandwich Canine Copeptin ELISA Kit (MBS007635) 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 up-regulation of this hormone stimulates the insertion of water channels (aquaporins) into the collecting tubule of the kidney, increasing water reabsorption and reducing urine volume (Sands and Layton 2014; Knepper et al. 2015; Xu and Wang 2016). 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 blood stream by pre-vasopressin 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 (Morgenthaler et al. 2008; Morgenthaler 2010). 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 (**Table S1).** These comprised sister species Rueppell's (*n*=29) and red foxes (*n*=50), as well as two outgroup species, the fennec (*n*=5) and pale fox (*n*=5) (**Table S1**). Genomic DNA was extracted using Qiagen DNeasy Blood & Tissue Kit (QIAGEN), following the manufacturer's protocol. Prior to the incubation step of the extraction protocol, 4 uL of 100mg/mL 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 1mL of PBS buffer to clean the tissues and deplete them from 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 Kits for DNA cleanup, 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. Genomic DNA extracts were quantified using a Qubit dsDNA High-Sensitivity Assay for blood sample extracts and Qubit dsDNA Broad-range Assay for

tissue extracts. Each sample was run on a 0.8% agarose gel to assess extraction quality and identify patterns of DNA fragmentation.

RNA was extracted using PAXgene® Blood RNA Kit (QIAGEN) for 11 fox specimens of Rueppell's (*n*=4) and North-African and Iberian red fox (*n*=5 Iberia; *n*=2 North-Africa; **Table S3**). All samples had sufficiently high concentration and RIN scores (> 8.2), with the exception of one individual (one Rueppell's fox sample, VR03; RIN = 6.8; **Table S3**),

Library Preparation and sequencing

Aliquots of 100 uL containing 0.5 to 2.5 ug of DNA were sheared to fragment sizes ranging from 200 to 600 bp using a Diagenode Bioruptor. Sonication protocols varied according to DNA quality. Fresh tissue and blood samples were sonicated with 7 cycles of 15 seconds of shearing (on) followed by 90 seconds of pause(off). Road-kill samples received the same timing of on and off, but varying cycles from 1 to 6 based on the degree of fragmentation. After shearing, samples were selected for an average fragment size of 350bp using AMPure XP beads 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 bellow 200 bp using low fragment single size selection with 1.6X of AMPure XP bead suspension. Sequencing libraries were built in batches of 20 samples at a time using the KAPA Hyper Prep Kit PRCR-free protocol (KAPA Biosystems, Boston, USA) following the manufacturer's instructions, adjusting overall volumes of each reaction step to half. These comprised a total of 89 libraries consisting of 29 North-African Rueppell's foxes, 32 North-African and 18 Eurasian red foxes, and additional 5 fennecs and 5 pale foxes to be used as outgroups in downstream analyses (**Table S1**). Each library preparation session included a negative control. Libraries were run on a 2% agarose gel to identify potential adapter dimers, which were removed through 1X bead-clean up, quantified under Kapa Library Quantification Kit qPCR protocol and pooled equimolar based on qPCR results into a total of two pools. Pool volumes were then bead-cleaned up with 1X ratio and quantified using Qubit dsDNA High-Sensitivity Assay. The quality, size distribution and concentration of pooled libraries was checked using a DNA 1000 chip on an Agilent Bioanalyzer 2100 instrument. Due to the Truseq style Y-shaped adapters, part of unamplified single stranded libraries could not be accurately run on an agarose gel or Bioanalyzer. To more accurately determine sizing, 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 v2 Nano kit 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,

California, USA). Samples were sequenced to an expected average raw coverage of 5x, 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 Hou et al. (Hou et al. 2015), a method that is strand-specific (i.e. the sequencing reads maintain the strand orientation of the transcripts) and uses oligo-dT enrichment (Hou et al. 2015). Sequencing was performed on one lane of an Illumina HiSeq X 10 platform (Macrogen Inc.; Seoul, Republic of Korea). As for whole-genome sequencing, reads were 150bp Paired-End.

Whole-genome sequencing data quality control

Read Processing

Adapter 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 adapter 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 adapter 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 (Martin 2011), which trims reads bellow a quality threshold of 20 before adapter removal with the specification that high-quality G bases at the 3'end are ignored. Trimmomatic v.0.38 (Bolger et al. 2014) was then used to perform a sliding window trimming, cutting bases once the average Phred quality within a 4bp 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 (*Vulpes lagopus*) reference genome (Hasselgren et al. in press, Mol Ecol) using the Burrows-Wheeler Aligner (BWA) version 0.7.17 (Li 2013) with the "mem" option and processed with samtools version 1.9 (Li et al. 2009) to remove unmapped reads, reads with an unmapped mate, reads failing quality checks, reads whose alignment was not primary and reads with mapping quality bellow 15. Read pairs were also independently aligned to the domestic dog genome (*Canis lupus familiaris*) publicly available at NCBI (CanFam 3.1) using the same methods. 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 (**Table S1**).

FastQC results identified severe bacterial and adapter contaminations in 4 Rueppell's fox road-killed samples and as result, the final coverage was bellow 1x in these individuals after all quality control steps. These individuals were excluded from downstream analysis (**Table S1**). Given that our remaining sample set included samples from road-killed individuals for which post-mortem DNA damage is a possibility, we followed the conventions of ancient DNA research and used MapDamage v.2.09 (Jónsson et al. 2013) which allows both the quantification of damage patterns and the downscaling of the quality score of potential damaged bases in such samples. Post-mortem 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 (**Table S2**). The arctic fox and the dog assemblies were aligned with nucmer in Mummer v.3. (Kurtz et al. 2004) 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 (**Figure S1**).

Site Quality filtering and SNP calling

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 1 sequencing read to retain the site, 2) sites with read depth above 50x in each individual (i.e. 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 (min p-value of 1×10^{-6}) were considered potential mapping errors and excluded, 6) if one of the two alternative alleles was biased with respect to the read base quality (minimum p-value of 1 x 10-10), mapping quality (minimum p-value of 1×10^{-4}), or distance from the end of the read (minimum p-value of 1×10^{-4}), the site was excluded. The subset of retained sites (positions) from all samples will, hereafter, be designated as all^{sites}. Similar analyses were performed for

a subset of the data comprising only ingroup foxes (red fox and Rueppell's fox; *ingroup*^{sites}) or using our dataset with additional canid genomes to be used as outgroups in phylogenetic analysis (**Table S2**).

For SNP calling and downstream population genetic analyses, we used the probabilistic framework based on genotype likelihoods implemented in ANGSD version 0.933 (Korneliussen et al. 2014) 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 1e-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 (hereafter SFS) of each group (determined by unsupervised clustering analysis). Briefly, ANGSD computes posterior probabilities of Sample Allele Frequency (SAF) for each site, which is then used to estimate the SFS. The shape of the SFS was used as proxy for the adequacy of data quality control measurements, and subsequent estimates of the SFS served as input for downstream analyses (namely F_{ST}, θ-based selection scans, and demography). We examined the SFS using *all*^{snps} estimated with -SNP_pval 1e-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 cut-off of SNP_pval 1e-5 for the majority of our analysis, unless otherwise noticed.

For mitochondrial DNA reads, consensus sequences were called for positions with a minimum coverage of 10 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 (**Table S2**). The combined datasets were aligned using MAFFT v7.407 (Katoh and Standley 2013) 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 only the cyb region (~1130bp) so that we could perform downstream analysis with other available fox sequences (**Table S2**). MtDNA alignments were eye-checked using SeaView v.4.7 (Gouy et al. 2010).

Relatedness

We ran NGSrelate (Korneliussen and Moltke 2015) 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 data. NGSrelate allows estimation of relatedness coefficients from genotype 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 (i.e -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 analyzed 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 (i.e. the higher the k0). Pairs of individuals with a k0 bellow 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 EM for optimization, allowing up to 500,000 EM iterations for each pair and a stopping criterion of 1e-12 difference in likelihood between two consecutive EM-iterations. This was performed 10 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 (**Table S1**). 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 (**Table S1**). 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 (Dobin et al. 2013). The reference genomes used were the domestic dog (canFam3.1), the red fox (vulVul2.2), downloaded from NCBI, and the arctic fox (*Vulpes lagopus)* (Hasselgren et al. in press, Mol Ecol) (**Figure S24**, **Table S15**). 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 (Shumate and Salzberg 2020) and the resulting GTF annotation was then used for overlap counting in STAR. Mapping results were inspected with Qualimap (Okonechnikov et al. 2015) and MultiQC (Ewels et al. 2016). 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 (Li 2011). The resulting genotypes were filtered with VCFtools (Danecek et al. 2011) to keep only bi-allelic loci with a minimum coverage of 4 reads, MAF of 0.05 and missing at most in 3 samples. The dataset was then pruned for loci in high linkage disequilibrium (variants with r^2 >0.2 in 50kb windows and step size of 10 variants) and genotypes were used in a Principal Components Analysis (PCA) in PLINK 1.90 (Chang et al. 2015) (**Figure S25 A**). Fragment counts (i.e. counts of read pairs overlapping annotated genes) from STAR were imported and processed using R 4.0.3 3 (R Core Team 2020). Only genes with at least 10 overlapping fragments (between all samples) were considered for the downstream analyses. Normalization, expression estimation and differential expression were performed with the DESeq2 package (Love et al. 2014). For exploratory analyses, namely PCAs based on gene expression, gene counts were transformed using the rlog transformation (**Figure S25 B-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 vs Red foxes; all North-African samples vs all Iberian samples; North-African Red foxes vs Rueppell's fox; all Rueppell's vs Iberian red foxes; North-African vs Iberian red foxes (**Table S15**).

Mitochondrial DNA Phylogenetic analysis

Maximum-likelihood trees for mitogenomes and cytb alignments including representatives from all fox species with available DNA information (**Table S2, Figure S3A, Figure S4**) were generated with MEGA X 10.1.5 using the general time reversible model with 1000 bootstrap samples (Kumar et al. 2018). The consensus phylogenies were used as inputs for ancestral range reconstruction under a maximum likelihood framework with the R package BioGeoBEARS (Matzke 2013, 2014). Species were assigned to North-Africa, Eurasia and North America, according to their current geographic distribution range (accessed from IUCN, 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 (Statham et al. 2014; Sacks et al. 2018). Furthermore, our red fox samples are from Eurasian and North-African origin. We run 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 tested with and without founder-event speciation (j parameter; Matzke 2014b) (**Figure S4**). We then chose the best fitting model (DEC, DEC+J, DIVALIKE, DIVALIKE+J, BAYAREALIKE and BAYAREALIKE+J) based on likelihood ratio test (LRT), the Akaike Information criterion (AIC) and corrected AIC (AICc) (**Table S4**).

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 v1.10.4

(Drummond and Rambaut 2007) (**Figure S3B**). We used the general time reversible model with a site heterogeneity model of gamma and invariant sites (GTR+G+I), as suggested by JModelTest (Posada 2008). We ran the settings with an uncorrelated relaxed log-normal clock model to allow rate variation among lineages. The Yule tree prior assumes a (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 any point in time every lineage can undergo speciation or go extinct and, given the observed cases of extinction in the Canidae family and *Vulpes* genus fossil record (Bartolini Lucenti and Madurell-Malapeira 2020), we assumed this tree prior 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 (Perini et al. 2010; Kumar et al. 2015; Zhao et al. 2016). Fossil calibrations were set using a gamma prior. The t_{MRCA} for *Vulpes* species and raccoon dog was set to 8.2 Mya, which coincides with the age estimate for the oldest fossil *Vulpes sp*. The oldest fossil assigned to Old World *Vulpes* is *V. riffautae* around 7 Mya (Bartolini Lucenti and Madurell-Malapeira 2020). We assumed 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 Mya. The t_{MRCA} for arctic fox and reminder species was set to 3.17 ± 0.09 Mya (Kumar et al. 2015). The t_{MRCA} for corsac/tibetan fox (*V. corsac/V. ferrilata*) and reminder *Vulpes* was set to 1.8 ± 0.5 Mya, and t_{MRCA} for corsac and tibetan fox to 1.02 ± 0.3 Mya (Perini et al. 2010). 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 iteration. We used Tracer v.1.7.1 (Drummond and Rambaut 2007) and effective sample size values to investigate convergence (**Table S5**). We next constructed a median-joining network for whole mitogenomic and cytb haplogroups using PopART v.1 to assess the relationships between the different haplotypes at a population level (http://popart.otago.ac.nz/links.shtml) (**Figure S5, Table S2**).

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 **Table S1** for information on average coverage per individual and geographic distribution). Principal Component Analysis (PCA) were executed in *PCangsd* using separate Beagle-format files generated in ANGSD with -sites all^{snps} for all 82 foxes

(**Figure S2B**), and for 48 red foxes (**Figure S6A)** and 24 Rueppell's foxes (**Figure S6B)**, 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 in order to obtain a PCA plot. Global F_{ST} estimates based on the joint Site Frequency Spectrum (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 and North African red foxes (Leite et al. 2015). The extent of genetic differentiation seen between North-African and Eurasian red foxes, and between Rueppell's and Red foxes was used to assess the most appropriate methods to detect signatures of selection. As input for the different site frequency spectrums we used a set of SNPs called with all individuals (*all*^{snps}).

In order to infer global admixture proportions and the covariance structure of allele frequencies in Rueppell's and red foxes (ingroups), we used the structure model implemented in the *Ohana* tool suite by assuming 2 to 6 ancestral populations (K, **Figure S7**). For each value of *K* in *Ohana* (Cheng et al. 2019) we used 32 independent executions with different random seeds and report the ones that reached the best likelihood. This was done using as input a beagle format file generated in ANGSD with Rueppell's 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 1x 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 downstream analysis on a sliding window (**Figure S2C**).

We also estimated whole-genome pairwise genetic distances between individuals using *ngsDist*, a program that takes the uncertainty of genotype's assignation into account (v.1.0.9; (Vieira et al. 2016). We run *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 run FastME (Lefort et al. 2015) to infer a tree for each of the resulting matrixes and then RaxML (Stamatakis 2014) to place support on the main tree (**Figure 1**).

Detection of ancient admixture events

To elucidate possible patterns of gene flow between Rueppell's and red fox we performed an ABBA-BABA test (D-statistics) as implemented in -doAbbababa2 of ANGSD (Soraggi et al. 2018). 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 foxes 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 (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 5Mbp followed by the Rscript DSTAT as implemented in ANGSD to obtain genome-wide level D statistic, Z score and p-values for significance (**Table S6**).

For comparisons in which significant genome-wide introgression was found, the ABBA-BABA test was performed on 200kb, 100kb and 50kb windows in order to find potentially introgressed outlier regions or genes. D statistic and Z scores were computed on windows along the length of the genome. To quantify deviations from the genome-wide pattern, we used the test statistic from a Pearson's chi-square test (https://github.com/joanocha/ngsIntrogression/) in a 2x2 contingency table comparing counts of ABBA and BABA sites genome-wide to that of the local window. We note that because of LD between SNPs within each window, a standard chi-square test of homogeneity is not appropriate for calculating p-values and we, therefore, just report the raw chi2 values without a probabilistic interpretation (**Figure S11 to S12**). Since spurious outliers may also arise in regions of reduced diversity, we adapted ABBA-BABA on windows to calculate f statistics (Martin et al. 2015). 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, $f\hat{}_{hom} = \text{numerator}_{(H1,H2,H3,H4)} / \text{numerator}_{(H1,H3,H3,H4)}$, and fraction admixture $f\hat{}_{d} = numerator_{(H1,H2,H3,H4)} / numerator_{(H1,HD,HD,H4)}$, where D=H2 or D=H3, depending on which maximized the frequency of the derived allele (https://github.com/joanocha/ngsIntrogression/). Genomic windows above the 99th percentile of the empirical distribution of $f\hat{}_{d}$ were considered to be outlier candidates for introgression between H2 and H3 (**Figure 2** in Main Text). These were then presented as ranked lists from highest to lowest value (**Table S8, Table S10**). The GTF annotation resulting from mapping the dog and arctic fox assembly with Liftoff 1.5.1 (Shumate and Salzberg 2020) was then used for overlap with genomic windows.

TreeMix v1.13 (Pickrell and Pritchard 2012) was also used to investigate the phylogenetic placement of Rueppell's and red foxes relative to the arctic fox, fennec fox 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 foxes (*n*=30), Eurasian red

foxes (*n*=18), Rueppell's fox (*n*=24), arctic fox (*n*=1), fennec foxes (*n*=5) and pale foxes (*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 based on 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 two times 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 0 to 3 migration events, with the pale fox as outgroup (**Figures S8 to S9**).

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 (i.e. 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 covered individual per species and within Eurasian and North African red foxes; **Table S1**). Selected bams were executed in ANGSD with -doHaplo 1 -doCounts 1 -minMinor 1 in order to call pseudo-haploids based on 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 (**Figure 2C** in main text). We then compared this tree topology to that estimated using *ngsDist* (**Figure S2C**). This workflow was also used to generate maximum-likelihood trees for potentially introgressed genes, or genomic regions, with the difference that ANGSD was run all the individuals from our dataset (**Figure 3C**, **Figure S13A**, **Figure S14, Figure S16A-B, Figure S23**) and with the -r option to provide the coordinates of the reference genome for the candidate gene. ANGSD was also run with -doHaplo1 and -r with additional available genomes (**Table S2**). To augment the phylogenetic analyses, we performed Principal Component Analysis (PCA) using a single read sampling approach in ANGSD using SNPs that were called for potentially introgressed regions based on genotype likelihoods (**Figure S15, Figure S16C**).

Finally, we used allele frequency files generated in ANGSD to perform scans of genetic divergence (d_{XY}) between species and red fox populations in 50kb windows with a 10kb slide and 25kb windows with a 5kb 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 joint Site Frequency Spectrum (jSFS) between different fox lineages (see section *Population Genetic analysis*) 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 (**Figures S17 to S21, Figure S23C**). As input for the different site frequency spectrums we used a set of SNPs called with all individuals (*all*^{snps}).

Demographic analyses

The jSFS (or 2D-SFS) of North-African Rueppell's and red foxes 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* (Gutenkunst et al. 2009). 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) (Noskova et al. 2020). 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 score (**Table S7, Figure S10A**). For all demographic analyses we assumed a generation time of 2, and a mutation rate similar to wolves ($\sim 4.5 \times 10^{-9}$) (Koch et al. 2019). Thus far, the conventional generation time used to model the demographic history of red fox has been 1 (Kutschera et al. 2013; Statham et al. 2014). However, this is unlikely to match the generation time in natural populations. For instance, the generation time is presumed to be 3 for Rueppell's fox (IUCN, 2021) and 3.4 for kit fox (Lonsinger et al. 2018). 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 (**Table S5**), and evidence from whole-genome analyses, including ABBA-BABA tests (**Table S6**) and admixture graphs (**Figure S8**). 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 and red foxes (**Table S7, Figure S10B).** European red fox samples were not analyzed due the degree of divergence from Middle Eastern red foxes.

Genome-wide signatures of selection in North African red foxes

To detect signatures of positive selection in North-African red foxes we used the population branch statistic (hereafter PBS) (Yi et al. 2010), 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 foxes we used Eurasian red foxes as the closest ingroup, and the Rueppell's fox as outgroup. To test the hypothesis of selection on Eurasian red foxes, 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). Based on considerations of the rate of decay of LD in the populations (**Figure S2C**), we chose to work on windows of 50 kilo-base pairs (kb) with a 10 kb slide (genome-wide), and 25kb with a 5kb slide (for a zoom-in view of candidate genes).

We then used ANGSD to compute genome-wide estimates of average number of pairwise differences (π) on windows of 50kb with a 10 kb slide, and 25kb with a 5kb 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 -log10(α) 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's reference genomes. We focused on candidate genes with extreme outlier regions under both methods (PBS and -log10(α)) (ranked in **Tables S12 to S13, Tables S16 to S17**) and tested for expression differences between groups (see *RNA-seq quality control and Differential expression analyses*) (**Figure S24-S27**, **Table S15)**. We then run PBS on individual SNPs to detect the top variants under selection for each of the top candidate genes for reads mapped to the dog's reference genome and inspected their location using the UCSC Genome Browser (**Figure S28**) (Kent et al. 2002).

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

PBS and α are powerful methods to find signatures of selection in red foxes because global estimates of F_{ST} between North-Africa and Eurasian red foxes was sufficiently low (Global F_{ST} ~ 0.13) and we had data from their immediate outgroup, the Rueppell's fox. While PBS has been proved effective for

identifying population-specific selection in humans and canid populations adapted to extreme environmental conditions (Yi et al. 2010; Malaspinas et al. 2016; Sinding et al. 2020), 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; Crawford and Nielsen, 2013). In sharp contrast to past F_{ST} estimates using microsatellites (Global F_{ST} ~ 0.14 between Rueppell's and North African red foxes, and Global F_{ST} ~ 0.13 between Rueppell's and Eurasian red foxes; Leite *et al.*, 2015), we found Rueppell's and red foxes to be highly differentiated using whole-genome data (Global F_{ST} ~ 0.52 between Rueppell's and North African red foxes, and Global F_{ST} ~ 0.51 between Rueppell's and Eurasian red foxes). 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 and fennec.

Multi-species Ohana

We ran *Ohana* (Cheng, Racimo and Nielsen, 2019) to obtain admixture plots and trees for our multi-species dataset (*n*=24 Rueppell's, *n*=48 red foxes, *n*=5 fennecs, *n*=5 pale foxes), using as input a beagle file with SNPs called from all individuals (*all*^{snps}). Using the results from *K*=4, we run 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 pre-defined populations, and can incorporate more than 3 branches in a tree, thus having greater power of analysis when more than 3 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 foxes and Rueppell's foxes.

For each SNP we introduced a scalar variable (h; see 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 genome-wide covariance pattern, and another that allows larger change in focal populations (i.e. higher variance components) than expected from 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 50kb, sliding every 10kb (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 (**Figure 4** in main text, **Table S19-20)**. We denote the log likelihood ratio calculated for fennec as LLR_F, Rueppell's as LLR_{Ru} and the one calculated for the red fox lineage as $LLR_{Red.}$.

We noted that many of the outlier regions putatively under selection were common to both Rueppell's (LLR_{Ru}) and red foxes (LLR_{Red}), nominally suggesting that selection has been strong on both species and targeted inherited ancestral standing genetic variation. In reality, this is likely an artifact caused by the fact that the nearest outgroup to Rueppell's and red foxes is quite divergent, and for that reason, it is difficult to distinguish between selection on the red fox branch and Rueppell's branch. For genes truly under selection in, say, red foxes, if not allowing selection in the red fox branch, an increase 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: $\delta_{LR} = LLR_{Ru} - LLR_{Red}$, where a positive value indicates more evidence that the selection is in Rueppell's fox and a negative value indicates that the selection most likely was in the red Fox. We note that when calculated for a single SNP, and not in windows, $Ratio_{LR}$ can be interpreted as Bayes factor. We then focused on the 99.95th percentile of the empirical distribution for δ_{LR} in Rueppell's foxes (**Table S19)**. Finally we looked if candidate genes were differentially expressed between Rueppell's and red foxes (see *RNA-seq quality control and Differential expression analyses*) (**Figure 4** in main text).

Intra-specific Polymorphism to Inter-specific divergence

We calculated the test statistic for a chi-square test of homogeneity akin to the Hudson-Aguade-Kreitman test (HKA) (Hudson et al. 1987; Zhai et al. 2009), as an alternative method for identifying signatures of selection specifically on Rueppell's and on fennec foxes, following a similar approach to that of Liu et al (2014). 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 of these tests provide an opportunity to detect selection acting on different time scales. 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;
D=number of fixed differences between red fox and both focal species and Dog;

This analysis was repeated with both the fennec fox 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 bellow 0.03 in the other. We further explore 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 foxes). 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 2x2 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 -log10(P-value) as a raw score that should not be interpreted probabilistically, similarly to Liu et al. (2014). Then windows were ranked based on the Homogeneity test score, filtering for windows in which:

- C was higher than A
- ratio of C/A was always greater than the ratio of D/B
- the ratio of C/A was always higher than $C_{genome\text{-}wide}/A_{genome\text{-}wide}$.
- the ratio of D/B was always lower than $D_{genome\text{-}wide}/B_{genome\text{-}wide}$.
- HKA chi2 value higher in the focal group (Rueppell's or fenenc) than red foxes
- HKA score in red foxes < 1.3

The rationale behind these conditions, also used for Polar bears (Liu et al. 2014), was to ensure we were targeting outlier windows specific to the Rueppell's or the fennec fox 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 after applying the same filtering (**Table S21**).

Functional enrichment analysis

We computed enrichment of gene ontology (GO) terms (http://geneontology.org/) for putatively introgressed regions between North-African foxes (i.e. above the 99.9th percentile for ABBA-BABA on windows) using biological process or molecular function. We also computed enrichment of gene ontology (GO) terms in the regions of the genome that displayed high -log10(α) values when using North-African foxes or Eurasian red foxes as focal populations, and with high δ_{LR} and LLR_{Fennec} (i.e. above the 99.95th percentile). The GO terms were considered to be significantly enriched if false discovery rate (FDR) p-value < 0.05. Both p-values and corrected (FDR) p-values were reported (**Tables S9, S11, S14, S18**). We repeated these analyses with 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 desert and non-desert foxes

To further explore evidence for desert adaptation in desert-dwelling foxes we tested for species-specific significant differences in physiological parameters between Rueppell's, fennec foxes and red foxes (see *Phenotyping* in Supplementary Material and Methods). Descriptive statistics of the data were performed in R, including summary statistics for each parameter and box plots (**Table S22, Figure S29**). 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 foxes, 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 meter). 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 with weight and other body measurements (hindfoot and tail length) (**Figure S30**). The linear model was implemented in R using the lm function (**Table S23)**. A Shapiro-Wilk test of normality was performed, and QQ-plots inspected to determine if the residuals were normally distributed. To teste for homogeneity of variances both Levene's and Barttlet's test were performed in R. For the parameters in which the Shapiro-Wilk test did not led 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. We also used certain physiological parameters as covariates in order to account for putative correlated responses

(**Table S23**). Finally, we were not able to collect enough urine samples for inferential statistics. As such, we compared the maximum values for urine concentrating ability obtained in this study with 121 mammalian species published in a recent review (Rocha et al. 2021) by plotting these values along an aridity gradient. This allowed to us to gain a general perspective on where our values lie in the general distribution of mammalian maximum urine osmolalities (**Figure S31**).

Supplementary text

Text S1. *Physiological differences between core desert foxes and red foxes*

We looked for differences in 12 physiological measurements in 34 individual foxes (*n*=13 Rueppell's foxes, *n*=11 red foxes, and *n*=10 fennec foxes; **Table S22**) in order to determine potential adaptive phenotypes specific to the two species. Box plots for different blood parameters showed clear visual differences between red foxes and core desert species, particularly for water- retention hormones such as aldosterone and vasopressin (AVP, here measured with copeptin), as well as thyroxin (T4), a metabolism-related hormone (**Figure S29**). Pairwise comparisons using *t*-tests with Bonferroni correction showed significant differences in the three hormones between red foxes and Rueppell's ($P_{Copeptin}$= 0.00037; $P_{Aldosterone}$=0.0034; P_{T4}= 8.8e-05) and between red foxes and fennec (*(*$P_{Copeptin}$=0.02273;$P_{Aldosterone}$ =0.0027; PT4=1.0e-06). No significant differences were found between Rueppell's and fennec ($P_{Copeptin}$= 0.55044; $P_{Aldosterone}$ =1.00; P_{T4}= 0.16). For each of these three parameters, Welch's two-sample t-tests also did not show statistically significant differences between North African and European red foxes ($P_{Copeptin}$=0.7087; $P_{Aldosterone}$ =0.7769; P_{T4}=0.2591), nor between captive and wild-caught individuals ($P_{Copeptin}$=0.9811; $P_{Aldosterone}$ =0.3522; P_{T4}=0.3867). This seems to suggest that, at least for these three hormones, there is no evidence of significant intraspecific differences in red foxes (*n*=4 North Africa and *n*=6 Iberia), and also that the differences observed are not due to the fact that some of our sampled individuals were free ranging while others had been held captive, though this should be stated with caution given that sample size is rather small. No significant differences between species were found in levels of uric acid, electrolytes, albumin, cholesterol and plasma osmolality, though we found outliers for creatinine and urea levels in both core desert-dwelling foxes, and for chloride and uric acid in Rueppell's fox (**Figure S29**).

There are several confounders that can affect the results of our pairwise comparisons. To account for these, we specifically tested for differences in hormone levels between fox species using a linear model that included sex, weight, body mass index and whether the individuals were caught free-ranging or captive as covariates (**Table S23**). With the exception of T4, both copeptin and aldosterone were log-

transformed in order to meet the assumptions of normality and homogeneity of variances (**Table S23**). After correcting for the potentially confounding factors, we still found the same pattern of significant differences between red fox and Rueppell's fox, and between red fox and fennec (**Table S23**). These results suggest 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.

Thyroid hormone thyroxine (T4) levels were significantly higher in red foxes compared to desert foxes (**Figure S29**). Lower levels of T4 have been associated to lower metabolisms in many desert organisms (Scott, Irene; Yousef, Mohamed K., Johnson 1976; Kamau and Maloiy 1983; Hulbert et al. 1985a, 1985b; Chaouacha-Chekir et al. 1997; Hulbert 2000; Banta and Holcombe 2002; Ishikawa and Kitano 2012). The selective advantages of lower metabolism in deserts include reduced energy demand, low energy expenditures, lower production of metabolic heat, and subsequent reduced need for evaporative cooling and water loss. This could therefore explain past observations of reduced evaporative water loss and basal metabolism in core desert-dwelling foxes (Williams et al. 2002, 2004; Careau et al. 2007).

Copeptin levels (biomarker for AVP) were significantly higher in core desert-dwelling foxes than red foxes, indicating that the former seem better able to retain water than the latter. AVP plays critical roles in water-reabsorption at the kidneys, resulting in increased urine concentrating ability and reduced urine volume, and species-specific differences in this hormone have been associated with potential adaptations related to water-retention in desert-dwelling mammals. In particular, plasma vasopressin levels in some desert-dwellers are higher than in non-desert counterparts, and correlate with urine concentrating abilities (Schwimmer and Haim 2009; Donald and Pannabecker 2015). For example, the neurohypophysis of the Australian hopping mouse – the species with highest record of maximum urine osmolality – contains three times the amount of vasopressin of laboratory rats (per body unit) (Bridges and James 1982; Donald and Pannabecker 2015). It is, therefore, very likely that selection targeted vasopressin or vasopressin-related pathways, in core desert-dwelling foxes to a different extent than in red foxes. In the particular case of Rueppell's, this could potentially be explained by selection acting on the gene WFS1 (see **Figure 4** in main text). We were able to measure urine concentrating ability in three individuals of Rueppell's (n=3) and red foxes (n=3). Our maximum urine osmolality observed for one Rueppell's was of 5126 mOsm/kg, while minimum urine osmolality observed was of 3224 mOsm/kg in another individual. Maximum urine osmolality observed for Iberian red foxes was of 3027 mOsm/kg, while minimum urine osmolality observed was of 2358 mOsm/kg. As we did not have sample size for inferential statistics, we plotted the maximum values obtained for Rueppell's and Iberian red foxes alongside other mammalian species, including fennec foxes, and re-plotted the dataset from Rocha et al. 2021. While

acknowledging the inherit limitation that different mammal species were captured under different circumstances (some dehydrated others water-fed and others unknown) and urine osmolality was measured with different methods, we note that the highest urine concentrating ability for one Rueppell's is similar to that of a Kangaroo rat and slightly above the maximum recorded value for the fennec fox (4022 mOsm/kg). The higher levels of copeptin seen in both Rueppell's and fennec could potentially explain the increased urine concentrating ability in core desert-dwelling fox species (**Figure S31**).

Plasma levels of aldosterone in core desert foxes were lower than the red fox. Aldosterone is a major hormone of the RAAS regulating sodium (and water) reabsorption at the kidney, as well as other life-determinant processes such as blood volume and pressure. Different levels of this hormone have been found in comparative physiological studies targeting other desert-adapted species (Schwimmer and Haim 2009; Stockand 2012; Donald and Pannabecker 2015). While cases with higher plasma aldosterone have been interpreted as contributing to increased water-retention abilities alongside vasopressin, lower levels of aldosterone have been explained in the light of other observed phenotypes, such as increased plasma sodium, increased urine sodium excretion, low cardiac rate and blood pressure, and increased vasodilation (Schwimmer and Haim 2009; Stockand 2012; Donald and Pannabecker 2015). Dromedaries, wild Bactrian camels and desert spiny mice, for example, tolerate a high dietary intake of salt, yet they do not develop hypertension, maintaining both cardiac rate and blood pressures at remarkably low levels (Dickinson et al. 2007; Schwimmer and Haim 2009; Ali et al. 2012, 2019; Donald and Pannabecker 2015). In the absence of severe pathologies the selective advantages of aldosterone-induced low blood pressure would include lower risk for heat stroke, kidney and heart disease (Ali et al. 2019).

Collectively, these results suggest that some aspects of evolutionary adaptation in North-African foxes is context dependent, with core desert-specialist species seemingly being more able to retain water at the kidney than the generalist red fox.

Supplementary Figures

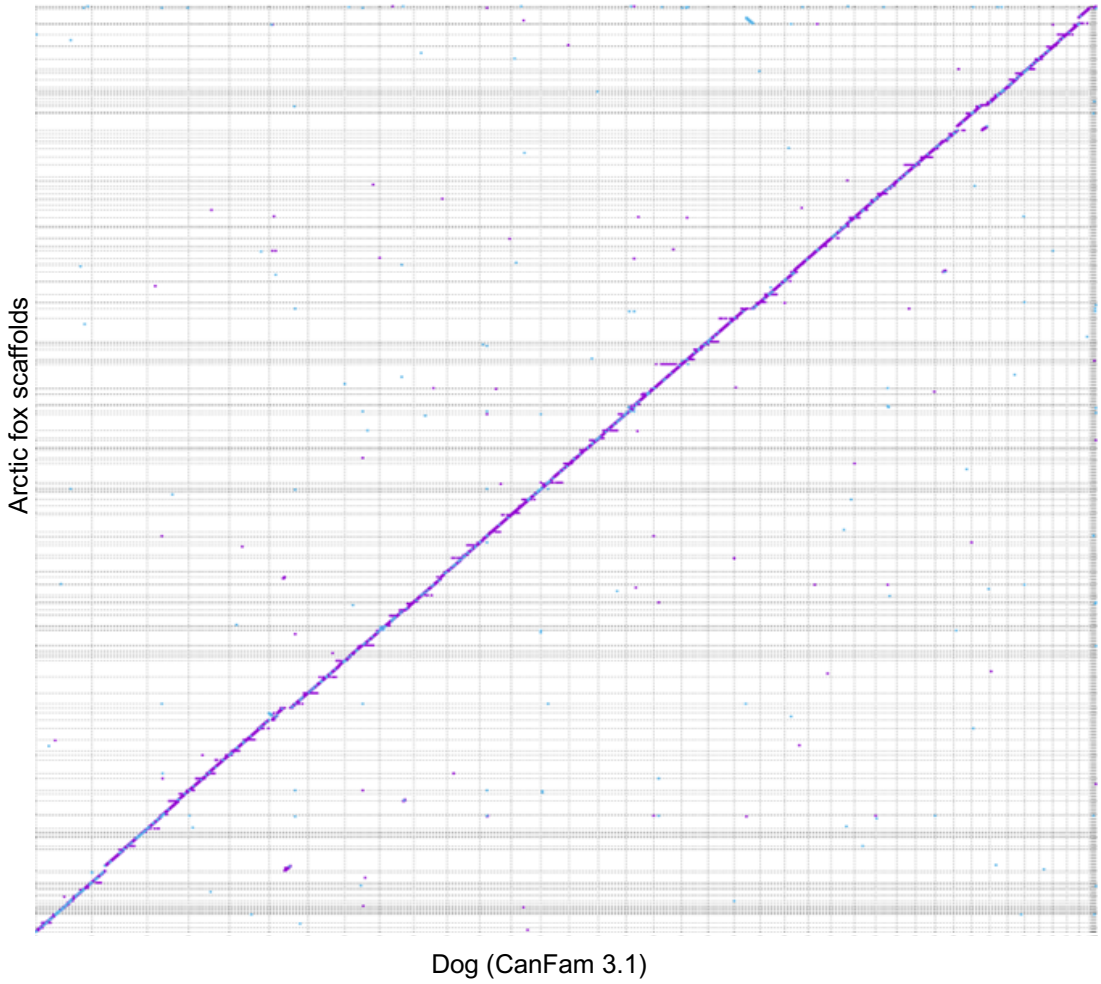


Figure S1. Reference genome mapping. Mummer plot obtained from mapping the arctic fox reference genome to the dog reference genome. Dot's represent matches with identity higher than 90%.

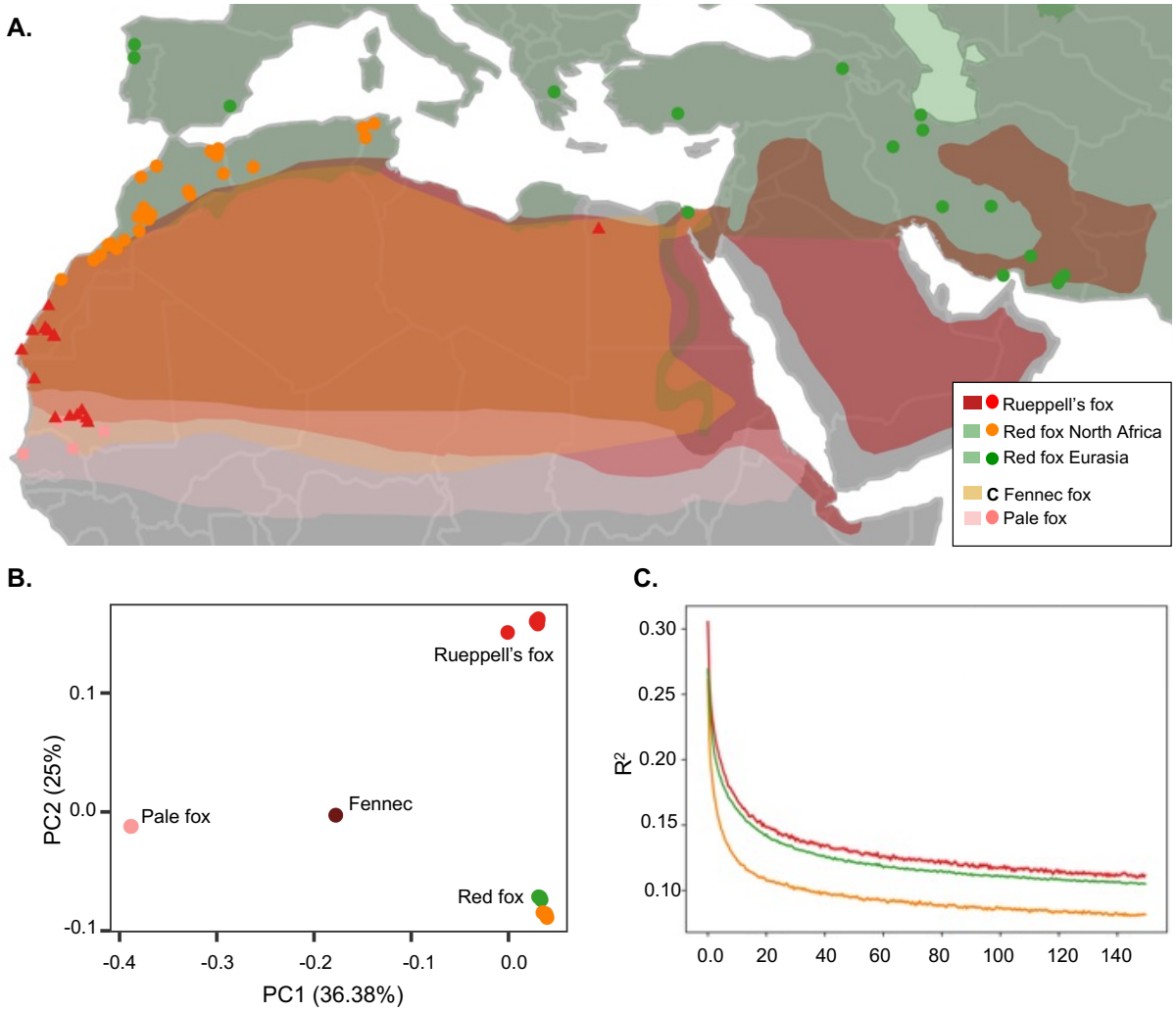


Figure S2. Geographic location of samples, species distribution range and overall species relationships using whole-genome data (related to Figure 1). (**A**) Species distribution range (shape files downloaded from IUCN, 2021) and sampling areas (points; see Table S1 for details; C in fennec stands for captive); (**B**) Principal component analyses (PCA) of 82 whole-genomes of *Vulpes sp*. PC1 separates North African outgroups pale fox (*V. pallida*) and fennec fox (*V. zerda*) used for comparative analysis from ingroups red fox and Rueppell's fox (*V.vulpes/rueppellii*) used in population genomics. PC2 separates Eurasian and North African red foxes (*V. vulpes*) from Rueppell's fox (PC2); (**C**) Linkage disequilibrium (LD) distribution of Rueppell's fox (red), North African red foxes (orange) and Eurasian red foxes (green).

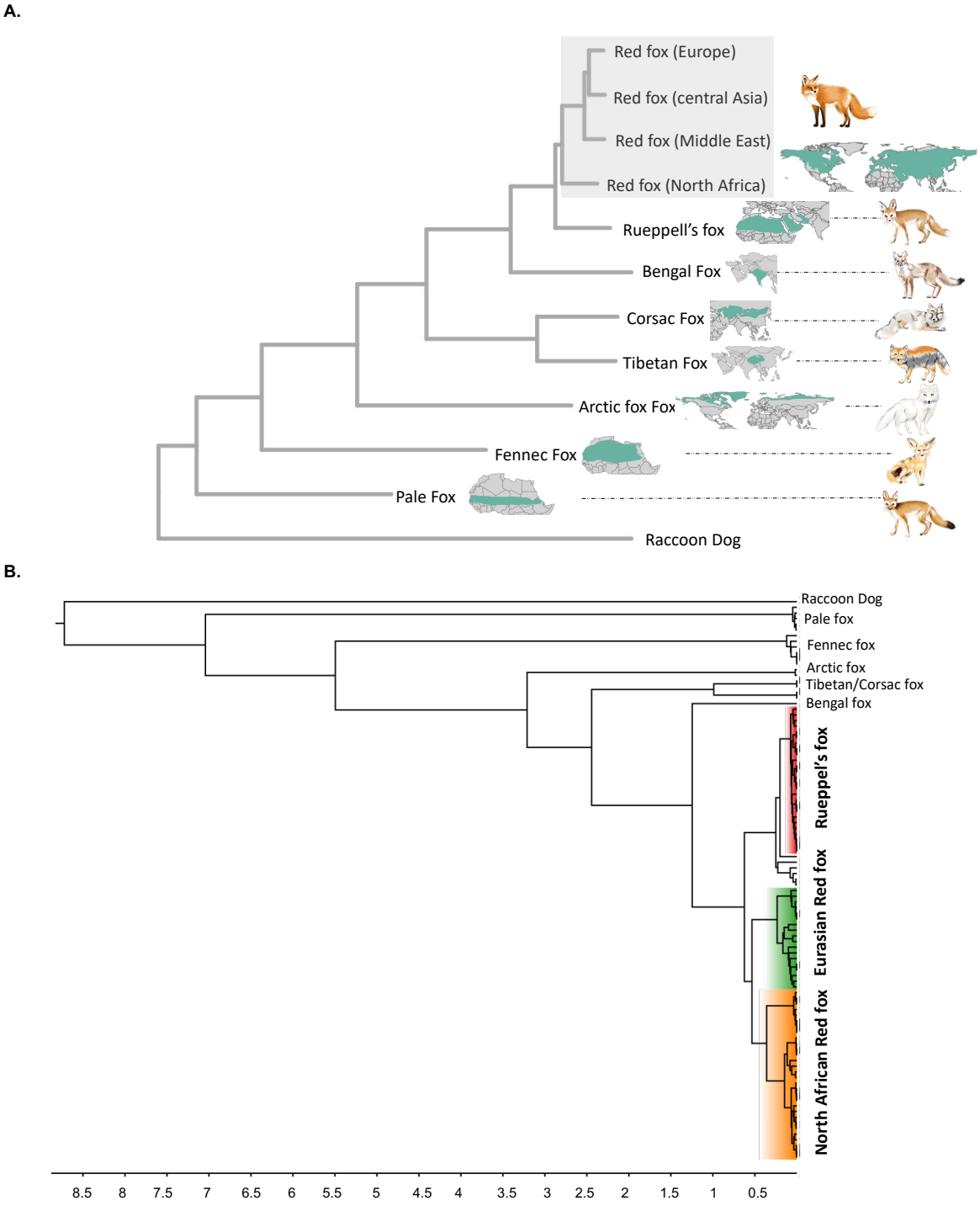


Figure S3. Mitogenomics of *Vulpes* using the raccoon dog as an outgroup (A) Maximum-likelihood tree followed by species distribution range downloaded from IUCN (2021). (**B)** Bayesian phylogenetic tree and time to the most recent common ancestor estimates; time scale in Mya represented bellow (see **Table S5** for the median time estimates and 95% HPD for each node). Accession numbers of published sequences are available in **Table S2**. North African red foxes are colored in orange, Eurasian red foxes green and Rueppell's fox in red. Rueppell's foxes are outgrouped by a non-colored clade of 6 red foxes (*n*=1 North-Africa, *n*= 6 Middle East), revealing a paraphyletic pattern in red foxes.

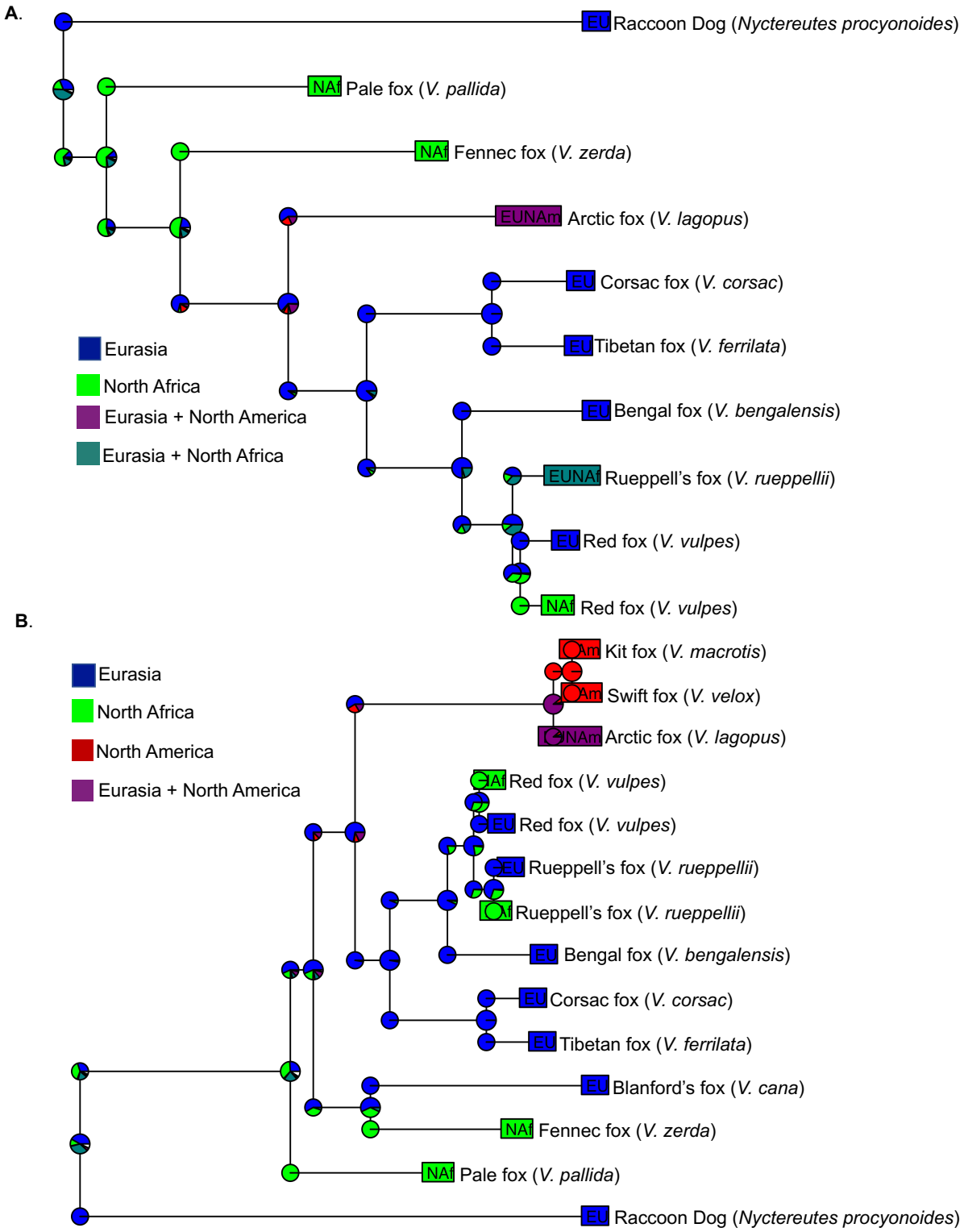


Figure S4. Ancestral range reconstruction for *Vulpes* species under the DEC+J model implemented in BioGeoBEARS. Maximum-likelihood trees for **A)** complete mtDNA genomes and **B)** cytb. Pie charts show the relative probability of each potential range. Squares at the tips represent the current range of extant species.

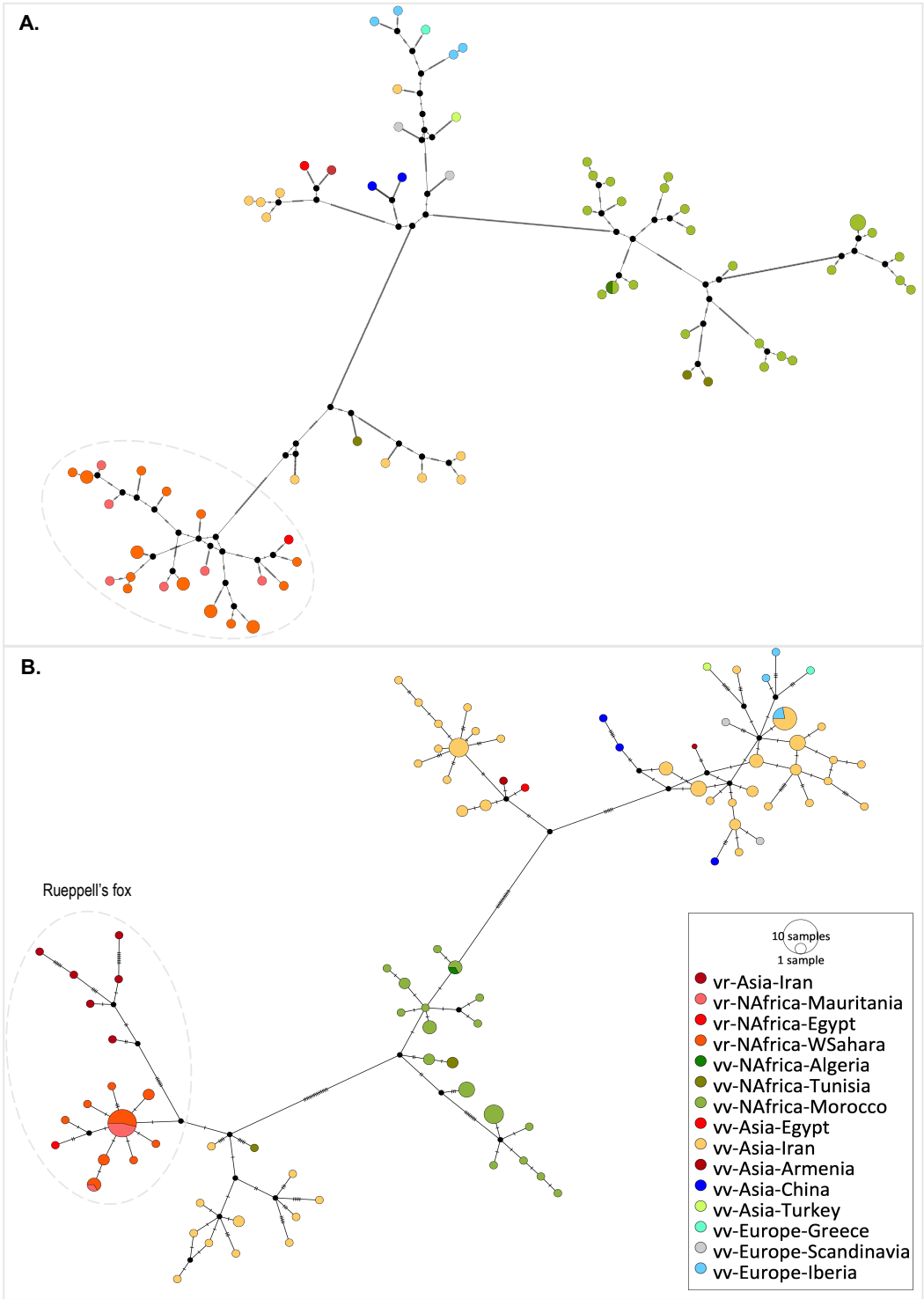


Figure S5. MtDNA haplotype networks. (**A**) Network for complete mitochondrial genomes from North African Rueppell's (*V. rueppellii*) and North-African and Eurasian red foxes (*V. vulpes*) from our dataset; **B**) mtDNA network for a 978 bp sequence alignment containing cytochrome b with additional 6 Asian Rueppell's and 58 red foxes from Iran (see Table **S2**). Rueppell's fox haplotypes are highlighted with a circle.

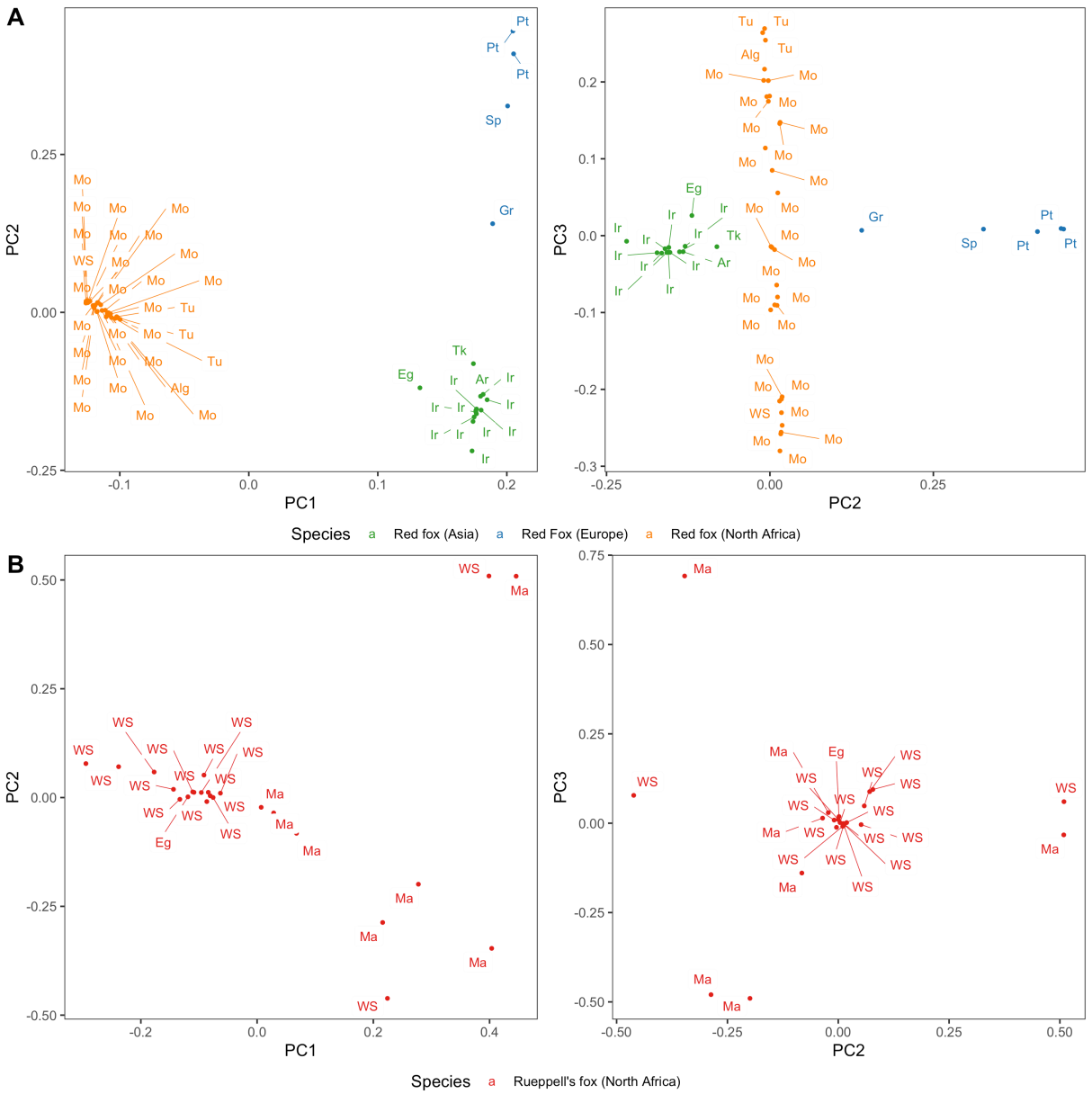


Figure S6. Principal component Analysis. (A) PCA of 48 whole-genomes of red foxes (*Vulpes vulpes*), which can be separated into North African and Eurasian (PC1, 14.64%), and Asian (Middle Eastern) and European (Iberia and Greece) (PC2, 3.13%). Consistent with this, the levels of genome-wide differentiation are greater between North-Africa and the Middle East (Global F_{ST} = 0.13), and between North-Africa and Europe (Global F_{ST} =0.19), than between Middle East and Europe (Global F_{ST} = 0.09). North African red foxes show isolation by distance along a latitudinal gradient from northern to southern regions in North Africa (PC3, 2. 49%). (B) PCA of 24 whole-genomes of Rueppell's foxes (*V. rueppellii*; PC1 = 7.72%, PC2=4.65%, PC3=4.59%) do not show strong geographic structure. Abbreviations are as follows: Alg- Algeria, Mo-Morocco, Ma-Mauritania, Tu-Tunisia, Eg-Egypt, Ir-Iran, Ar-Armenia, Tk-Turkey, Sp-Spain, Pt-Portugal, Gr-Greece, WS-Western Sahara.

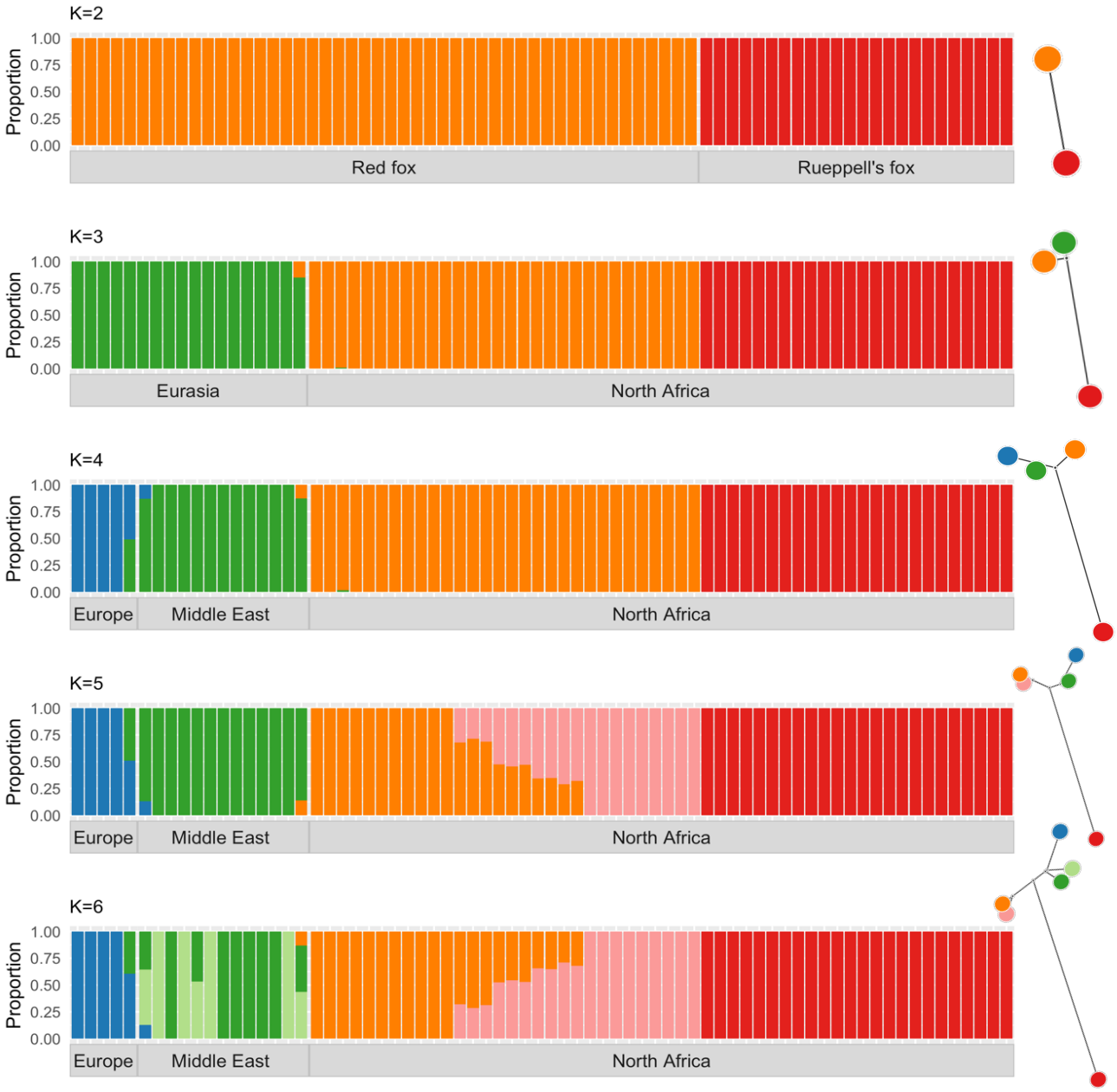


Figure S7. Ohana admixture results. Ancestry proportions and trees estimates inferred with Ohana from K=2 until K=6 for a dataset containing 72 individuals from ingroup species red fox (*V. vulpes*) and Rueppell's fox (*V. rueppellii*). On the left, ancestry proportions are shown, and each bar shows the inferred ancestry fraction for an individual. Individual colors illustrate the inferred admixture components, Ks values are indicated at the top. Species and geographic clusters are detailed at the bottom inside grey bars. On the right, population trees for the ancestry components that are most compatible with the inferred covariance matrix for each respective value of K. Branch length is an indication of drift.

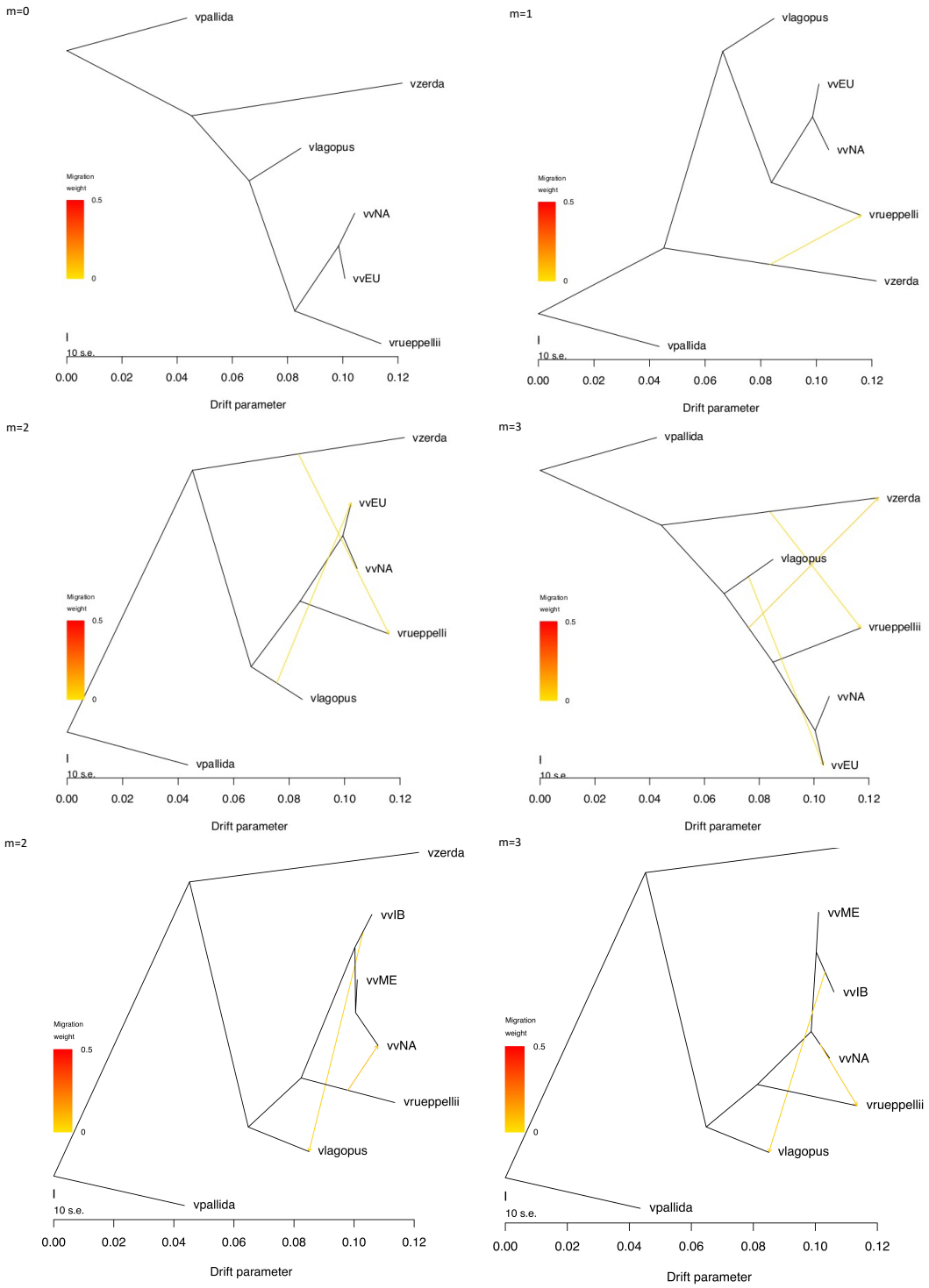


Figure S8. Phylogenetic placement of different *Vulpes* species using TreeMix admixture graphs on 83 individuals. TreeMix relationship of major lineages in our genomic dataset. Admixture graphs computed using TreeMix first for 6 lineages from 0 to 3 migration edges, and then 7 lineages with 2 to 3 migration edges. The pale fox (*V. pallida*) was used as outgroup. Pale fox – vpallida; fennec fox – vzerda; Arctic fox – vlagopus; Rueppell's fox – vrueppellii; North African red fox – vvNA; Eurasian red fox – vvEU; European red fox (Iberia and Greece) – vvIB; Asian red fox (Middle East) – vvME.

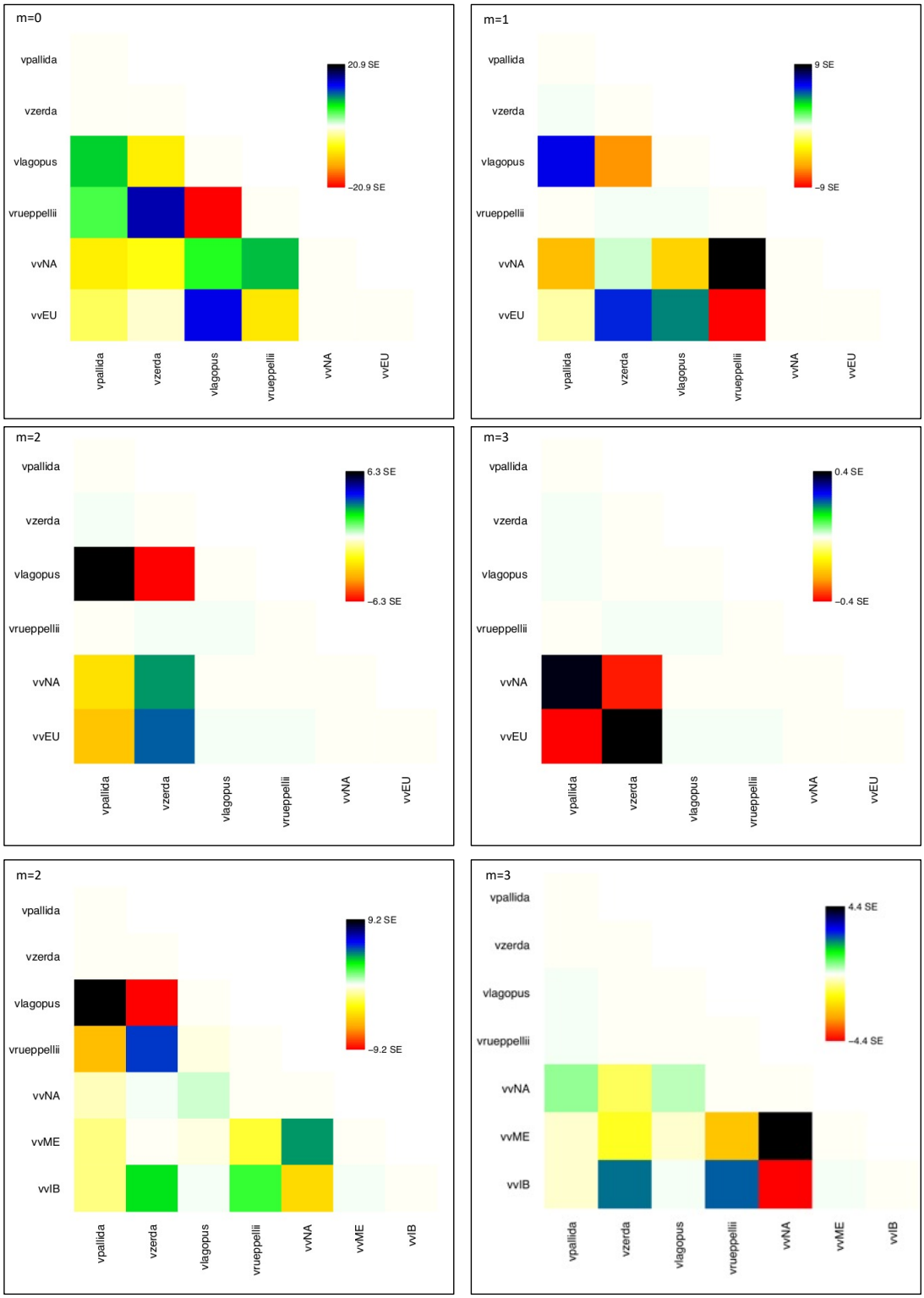


Figure S9. Residuals for the TreeMix admixture graphs presented in Figure S8. The pale fox (*V. pallida*) was used as outgroup. Pale fox – vpallida; fennec fox – vzerda; Arctic fox – vlagopus; Rueppell's fox – vrueppellii; North African red fox – vvNA; Eurasian red fox – vvEU; European red fox (Iberia and Greece) – vvIB; Asian red fox (Middle East) – vvME.

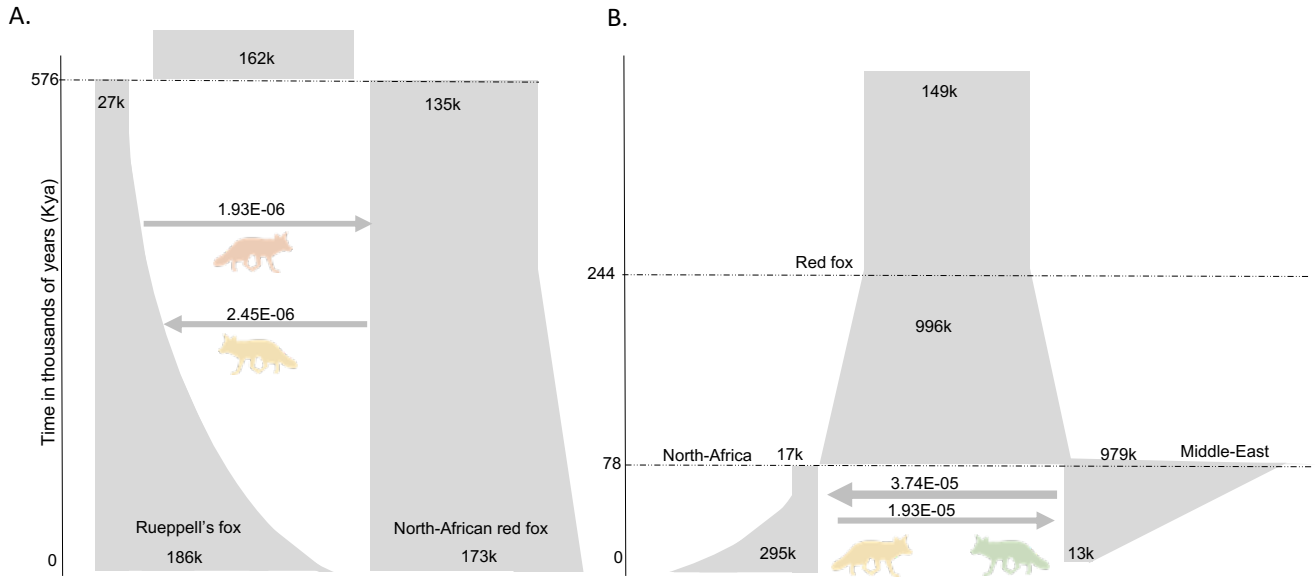


Figure S10. Demographic inference with *dadi* and using GADMA. The joint demographic history of Rueppell's and red foxes in North Africa **(A)** and the joint demographic history of North-African and Eurasian red foxes **(B)**, was inferred using diffusion approximation to the site frequency spectrum (SFS) as implemented in *dadi*. These analyses were performed on autosomes, removing potentially introgressed outliers (highlighted in Figure 2). Divergence times are represented in thousands of years (kya). The estimated effective population sizes and the migration rates are also indicated. Despite the inherent uncertainty in the genome-wide mutation rate estimate (set to be the same as wolves, $\mu = 4.5 \times 10^{-9}$) and variance in generation time (here assumed to be of 2 years), we find divergence time estimates between the Rueppell's fox and the red fox (A) to be in accordance with consensual earliest fossil records of red fox (Bartolini Lucenti and Madurell-Malapeira 2020). The best model in GADMA supports a population bottleneck in the red fox North-African population after divergence from Eurasian populations, followed by a recent exponential expansion (B). This model also supports the maintenance of the ancestral population size in Eurasian populations after divergence from North-African red foxes, followed by a linear population decline and smaller long-term effective population size (B).

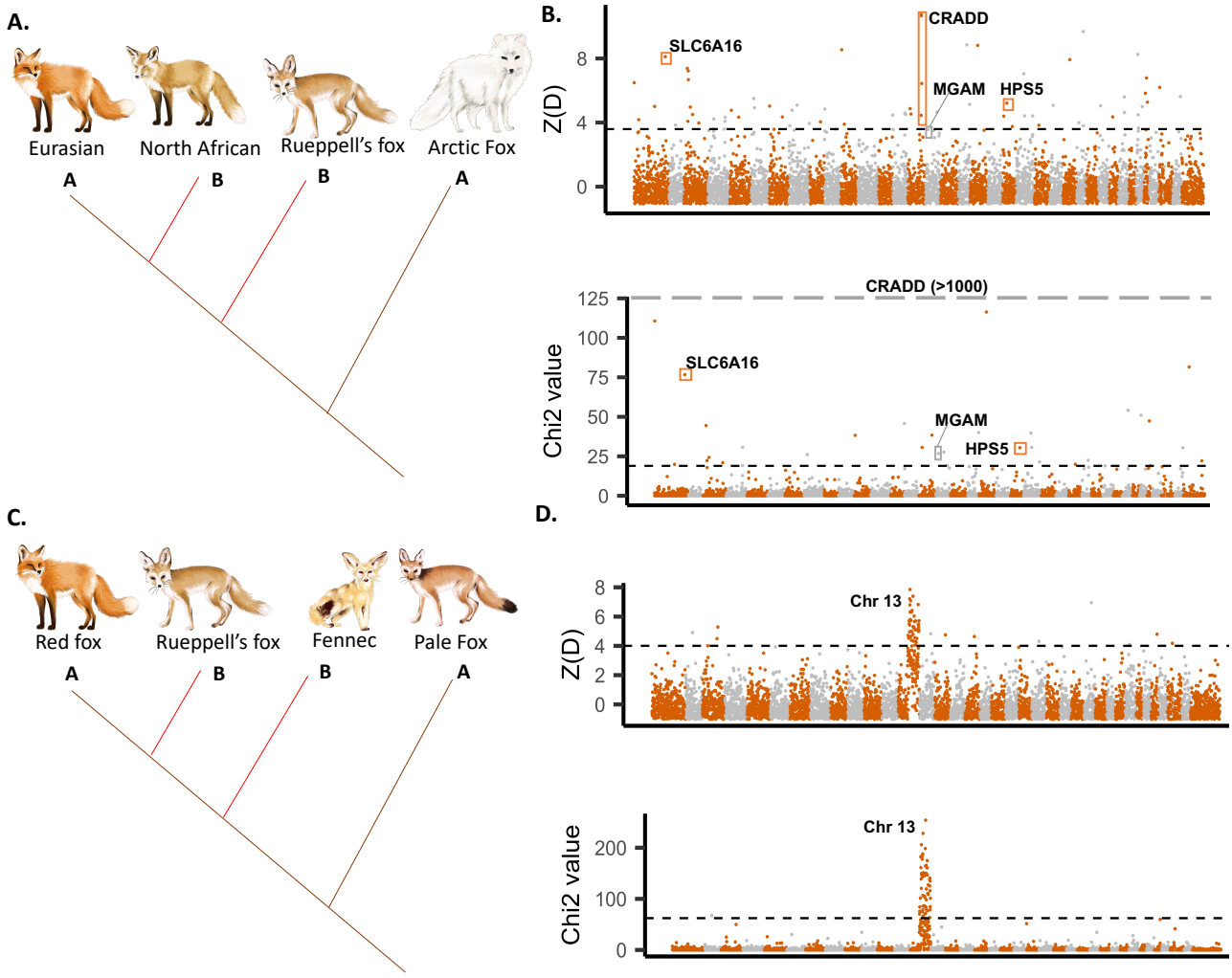


Figure S11. ABBA-BABA on windows using the arctic fox as reference genome (related to Figure 2). D statistics between Rueppell's fox (*Vulpes rueppellii*) and North African red fox (*V. vulpes*) (**A, B**) and between Rueppell's fox and fennec fox (*V. zerda*) (**C, D**) after mapping reads to the Dog (*Canis lupus familiaris*) reference genome. The pattern of ABBA/BABA in (**A**) is as follows: red fox Eurasia - state A or H1 - and North African red foxes – state B or H2 have a recent common ancestor that diverged from the ancestral of the Rueppell's fox (state B or H3), using the arctic fox as outgroup (state A or H4). The pattern of ABBA/BABA in (**C**): all red foxes - state A or H1 - and Rueppell's foxes – state B or H2 have a recent common ancestor that diverged from the ancestral of the fennec fox (state B or H3), using the pale fox as outgroup (state A or H4). (**B, D**) Manhattan plot of the Z-transformed D statistic and Chi2 value for non-overlapping windows of 200 kb across all chromosomes, according to pattern described in A and C, respectively. Each dot represents a window and windows from the same chromosome have the same color. Line threshold represents regions with Z score and Chi2 values above the 99.5th (**B**) and 99th (**D**) percentiles of the empirical distribution.

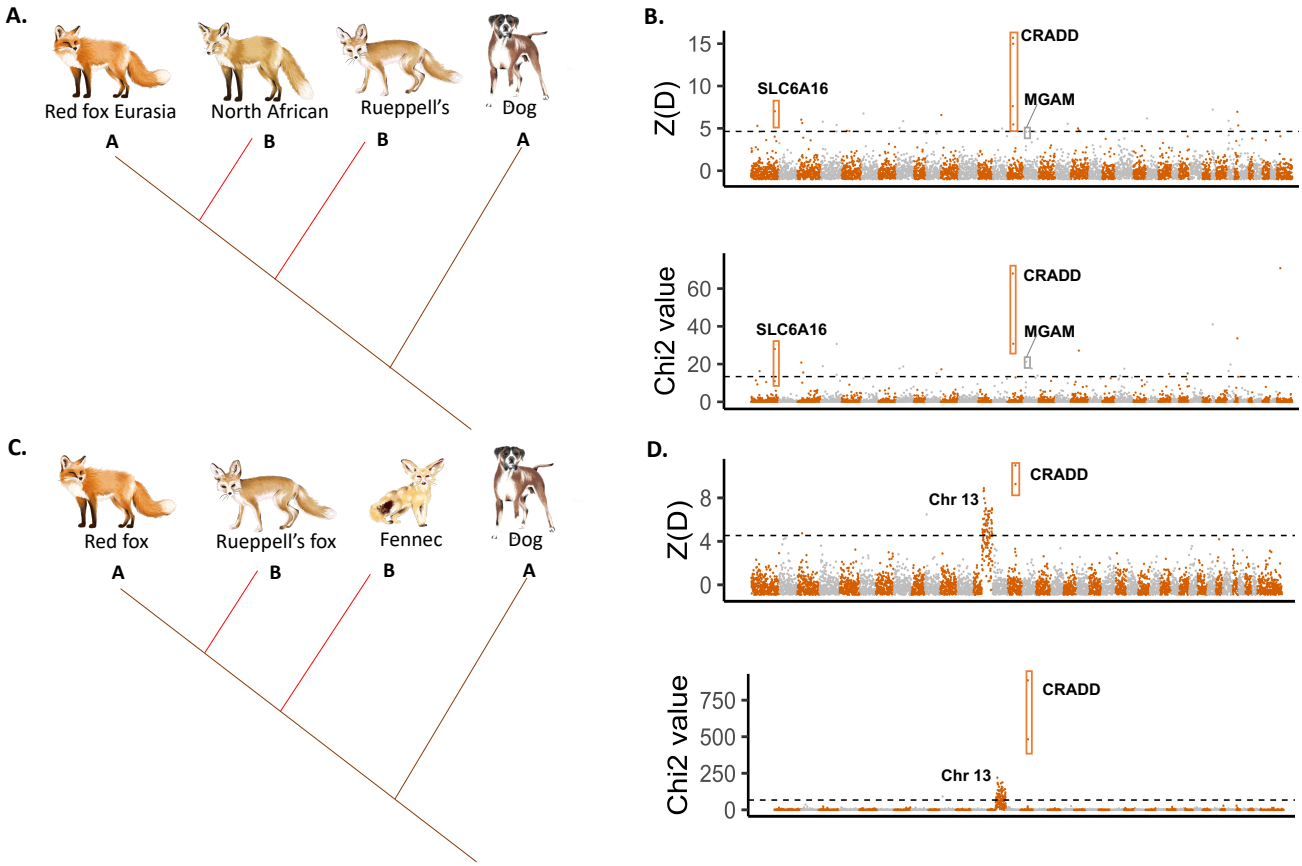


Figure S12. ABBA-BABA on windows using the domestic dog as reference genome and outgroup (related to Figure 2). D statistics between Rueppell's fox (*Vulpes rueppellii*) and North African red fox (*V. vulpes*) (**A, B**) and between Rueppell's fox and fennec fox (*V. zerda*) (**C, D**) after mapping reads to the Dog (*Canis lupus familiaris*) reference genome. The pattern of ABBA/BABA in (**A**) is as follows: red fox Eurasia - state A or H1 - and North African red foxes – state B or H2 have a recent common ancestor that diverged from the ancestral of the Rueppell's fox (state B or H3). The pattern of ABBA/BABA in (**C**): all red foxes - state A or H1 - and Rueppell's foxes – state B or H2 have a recent common ancestor that diverged from the ancestral of the fennec fox (state B or H3); the dog (state A or H4) was used as the outgroup. (**B, D**) Manhattan plot of the Z-transformed D statistic and Chi2 value for non-overlapping windows of 200 kb across all chromosomes, according to pattern described in A and C, respectively. Each dot represents a window and windows from the same chromosome have the same color. Line threshold represents regions with Z score and Chi2 values above the 99.5th (**B**) and 99th (**D**) percentiles of the empirical distribution.

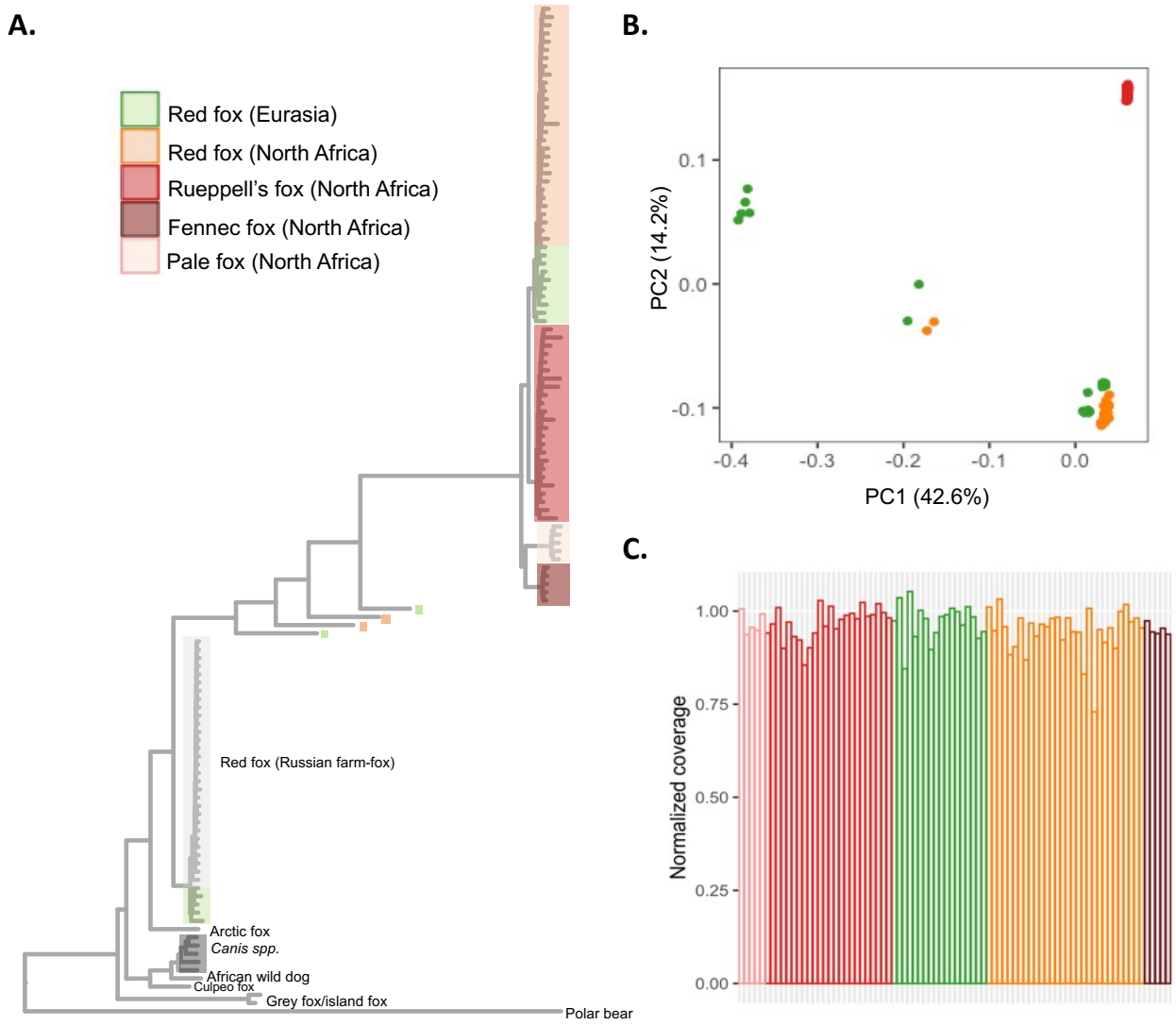


Figure S13. CRADD gene – top outlier above the 99th percentile of the empirical distribution for $f\hat{}_d$ and Z(D) in ABBA-BABA on windows (related to Figure 2A). (A) Maximum-likelihood tree estimated with IQtree for a sequence alignment of pseudo-haploids comprising the CRADD gene (~180kb) based on sampling of a random base (samples from our dataset are highlighted in the legend of the figure, see **TableS2** for additional genomes with legend in the tree). **(B)** PCA with 48 Red foxes and 25 Rueppell's using 23,016 SNPs called from a 1.3Mbp 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.

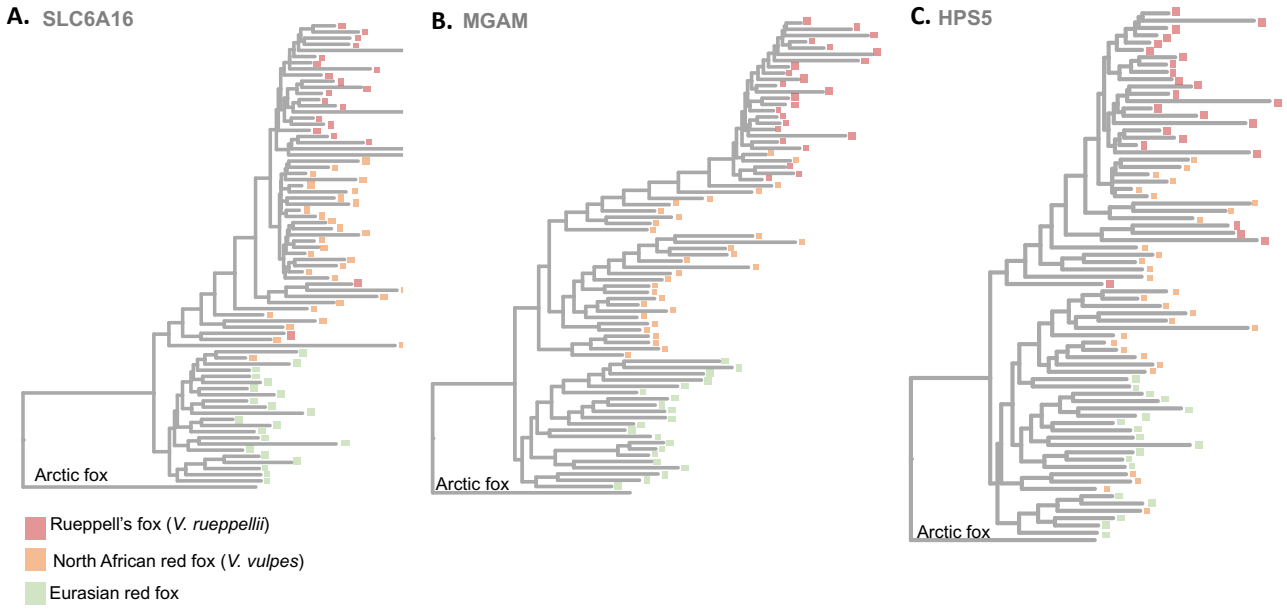


Figure S14 – **Tree topologies for SLC6A16 (ranked 2nd) and two other genes (MGAM, HPS5) in outlier regions above the 99th percentile of the empirical distribution for $f\hat{}_d$ in ABBA-BABA on windows (related to Figure 2A). (A-D)** Maximum-likelihood tree topologies were estimated with IQtree for a haploid sequence called using random based sampling with ANGSD (-doHaplo 1).

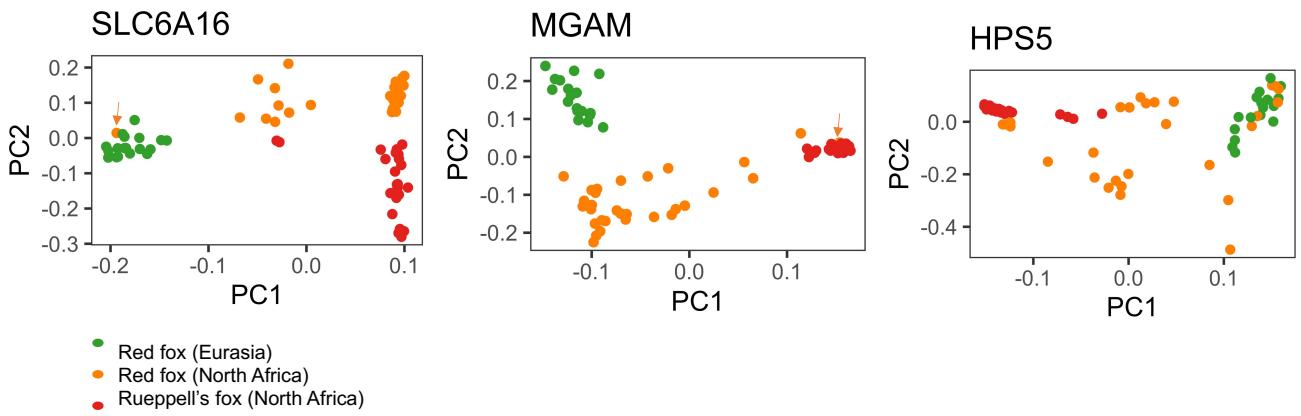


Figure S15. Principal component analysis for SLC6A16 (ranked 2nd) and two other genes in Figure S14. PCA based on sampling of a single read at each site for 4 genes above the 99th percentile of $f\hat{}_d$ empirical distribution (related to Figure 2A). North African red foxes individuals unambiguously cluster close to Rueppell's than other red foxes. Arrows highlight hidden North-African red fox individuals clustering with Eurasian red foxes (SLC6A16) and Rueppell's fox (MGAM).

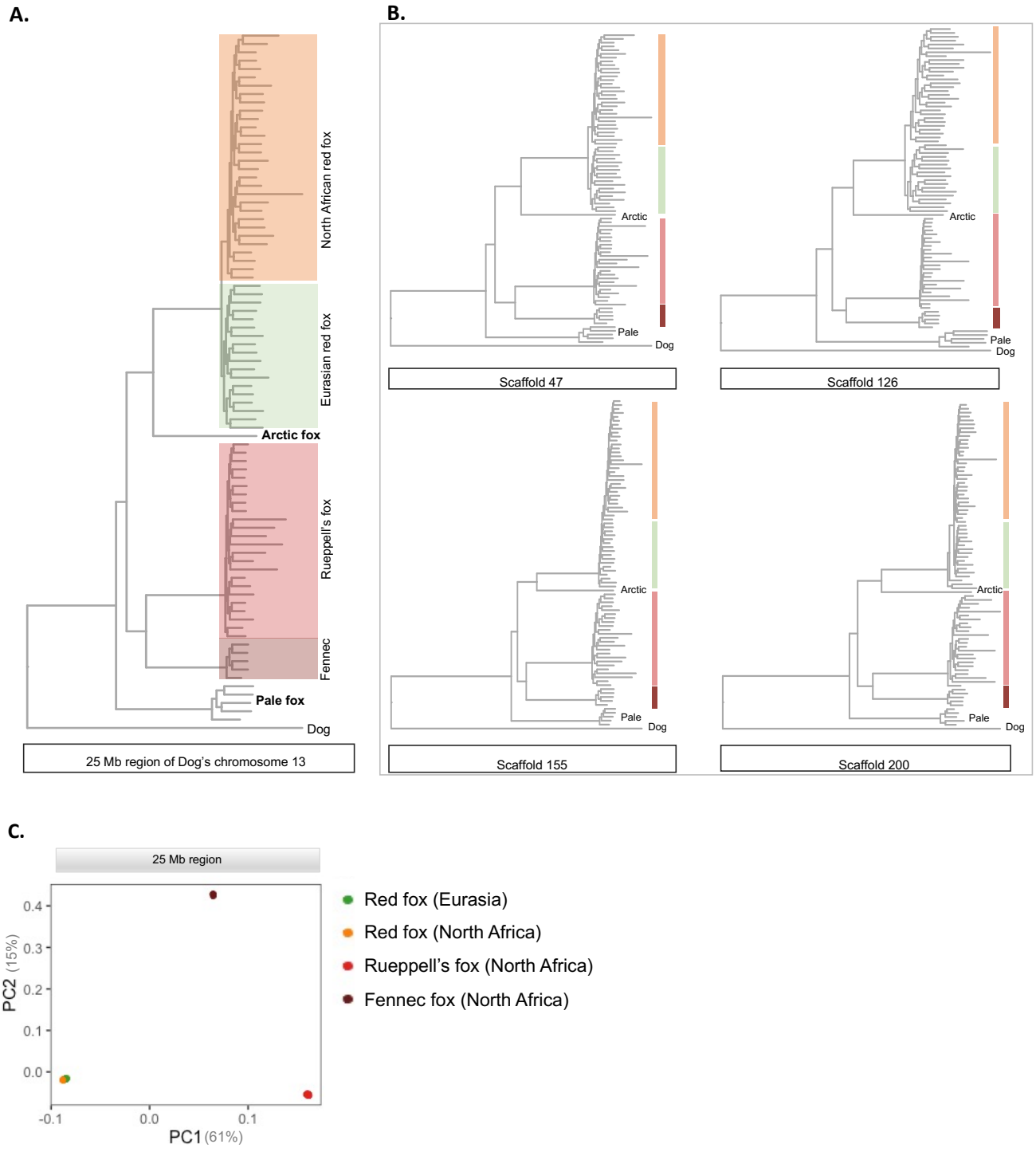


Figure S16. Tree topologies and Principal Component Analysis on the ~25Mb outlier region above the 99th percentile of the empirical distribution for $f^{\wedge}_{d}$ in ABBA-BABA on windows (related to Figure 2B). Maximum-likelihood tree estimated with IQtree for the affected chromosomal region called 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 25Mb region in 48 red foxes, 24 Rueppell's fox 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.

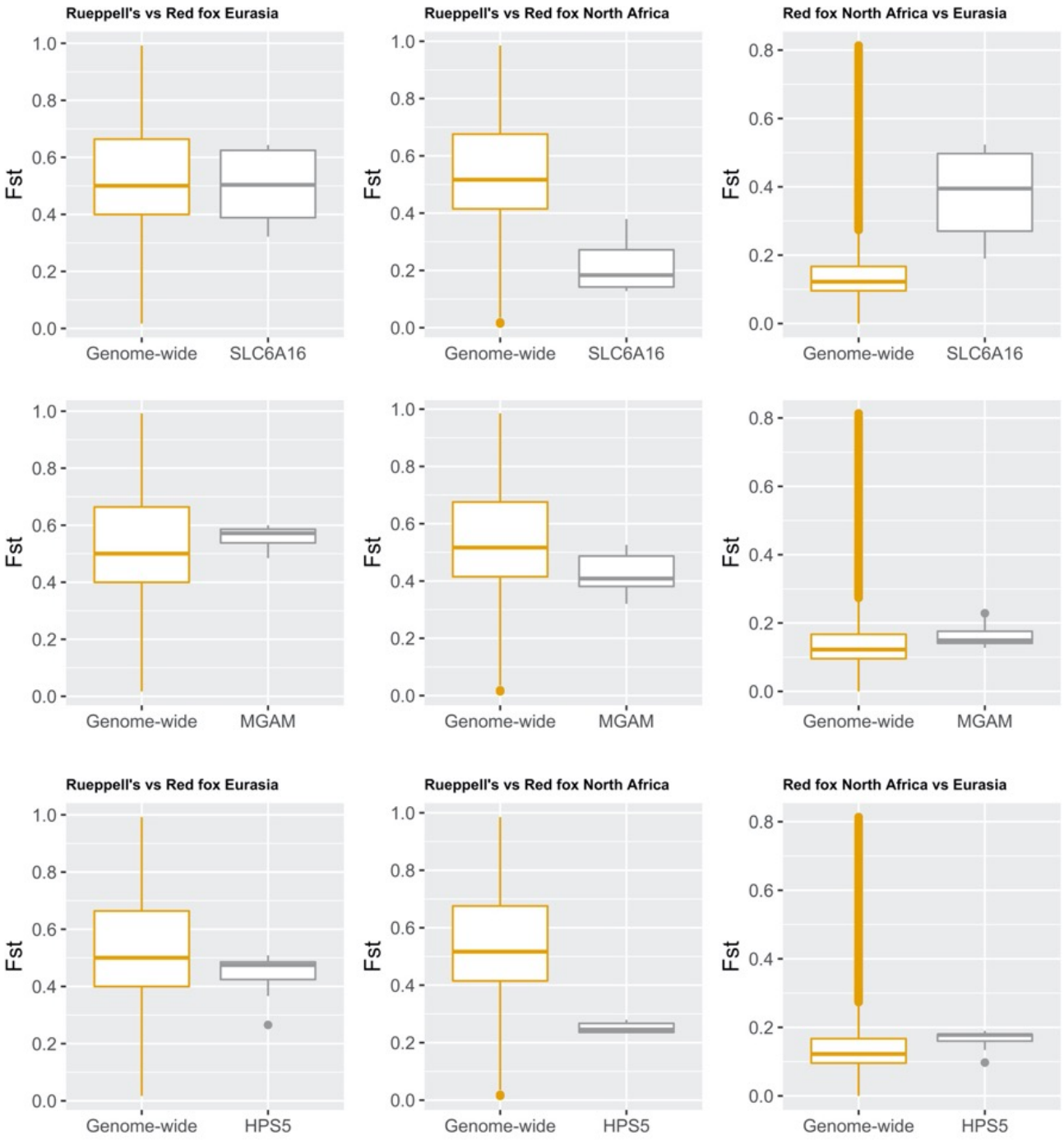


Figure S17. Genetic differentiation for SLC6A16 (ranked 2nd) and two other genes (MGAM, HPS5) above the 99th percentile of the empirical distribution for $f\hat{}_d$ in ABBA-BABA on windows (related to Figure 2A). Estimated F_{ST} between different pairs of species on sliding windows of 25kb overlapped every 5kb for the candidate genes and for all of the autosomes (excluding the CRADD gene).

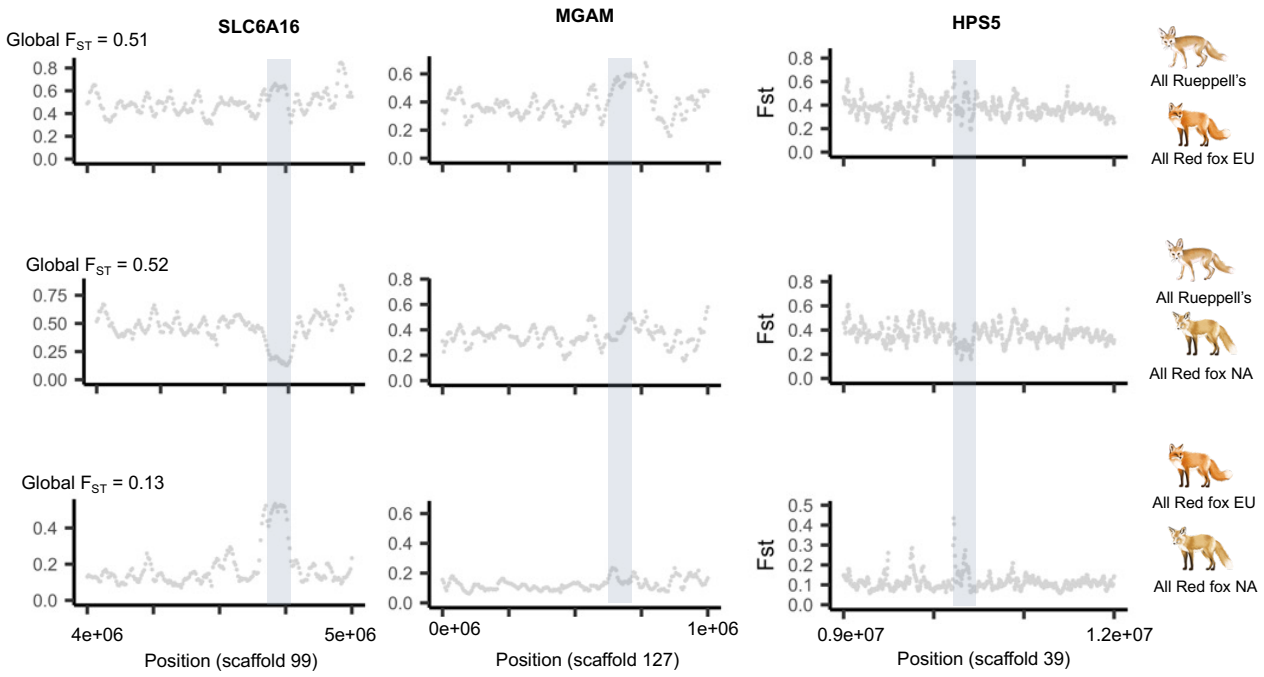


Figure S18. Zoom in view of SLC6A16 (ranked 2nd) and two other genes (HPS5, MGAM) above the 99th percentile of the empirical distribution for $f\hat{}_d$ in ABBA-BABA on windows (related to Figure 2A). Genetic differentiation (F_{ST}) between different pairs of populations and species on sliding windows of 25kb overlapped every 5kb for scaffolds harboring genes SLC6A15, MGAM and HPS5.

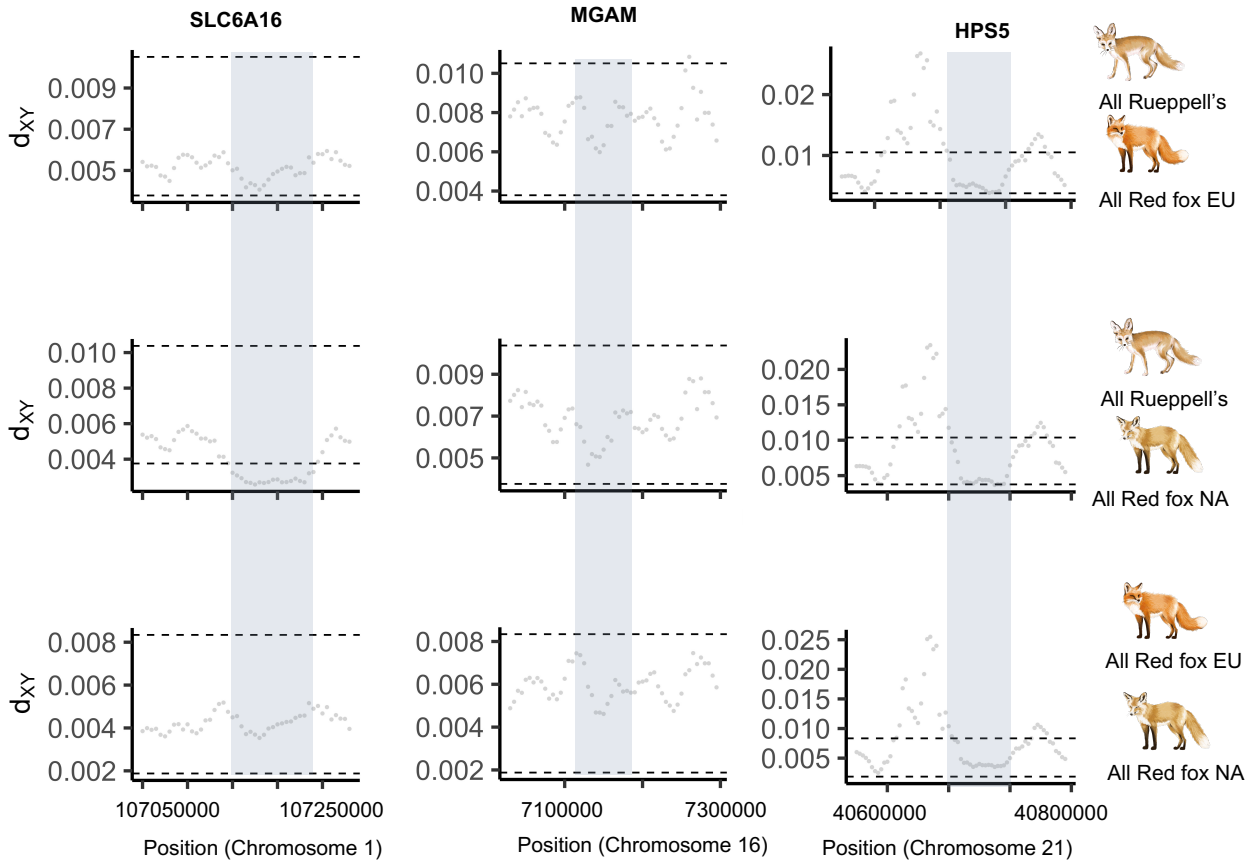


Figure S19. Zoom in view of SLC6A16 (ranked 2nd) and two other genes (HPS5, MGAM) above the 99th percentile of the empirical distribution for *f*^d in ABBA-BABA on windows (related to Figure 2A). Genetic divergence (d_{XY}) between different pairs of populations and species on sliding windows of 25kb overlapped every 5kb for scaffolds harboring genes SLC6A16, MGAM and HPS5. Threshold lines mark the 95% confidence interval of the genome-wide empirical distribution for d_{XY}.

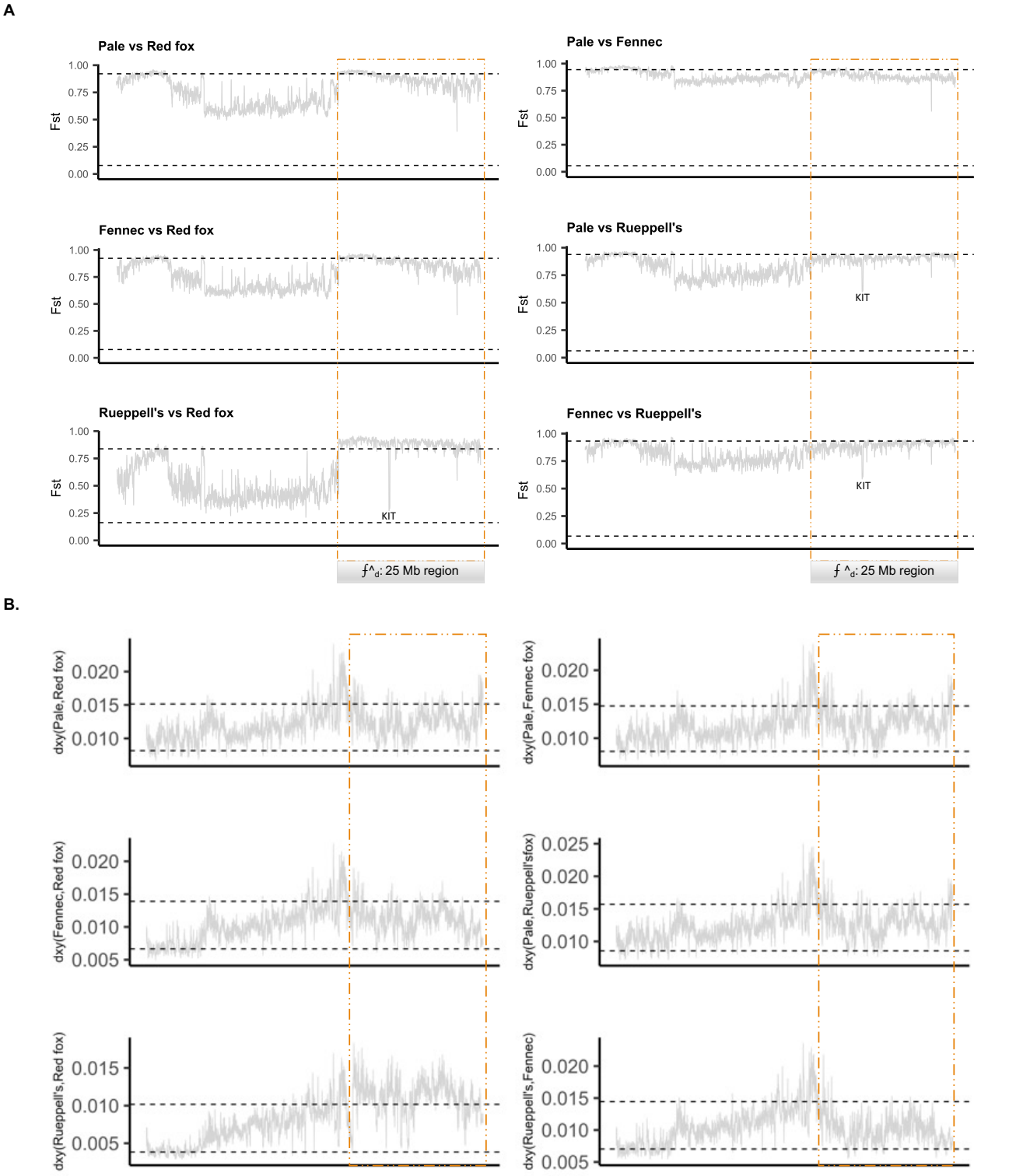


Figure S20. Zoom in view of the 25Mb region above the 99th percentile of the empirical distribution for $f^{\wedge}_d$ in ABBA-BABA on windows (related to Figure 2B). **(A)** Genetic differentiation (F_{ST}) and **(B)** divergence (d_{XY}) between different pairs of *Vulpes* species on sliding windows of 50kb overlapped every 10kb. Threshold lines mark the 95% and 5% of the genome-wide empirical distribution.

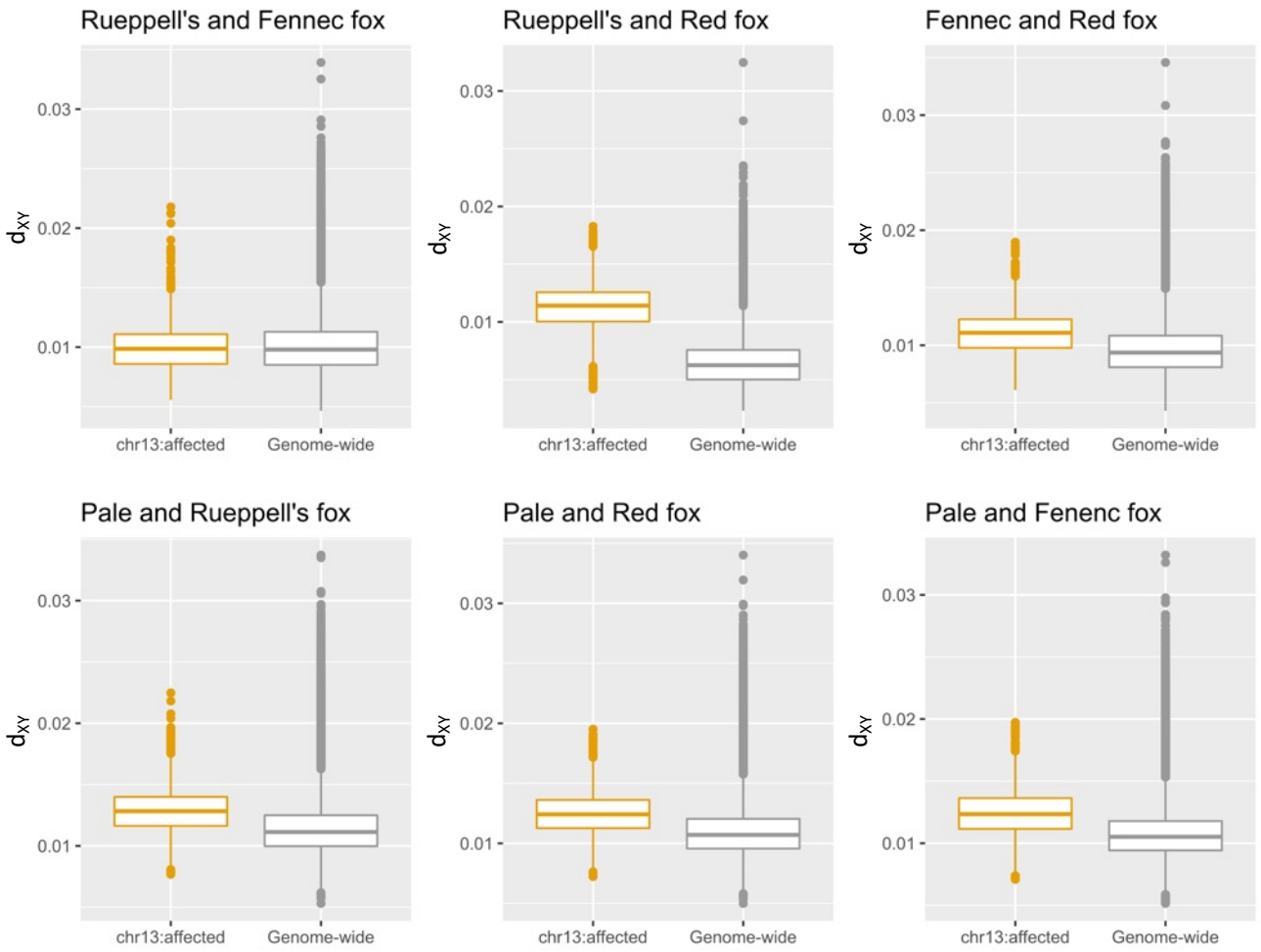


Figure S21. Genetic Divergence for the 25Mb region above the 99th percentile of the empirical distribution for $f\hat{}_d$ in ABBA-BABA on windows (related to Figure 2B). Estimated genetic divergences (d_{XY}) between different pairs of *Vulpes* species on sliding windows of 50 kb overlapped every 10 kb for the affected chromosomal region and for all of the autosomes (excluding the CRADD gene).

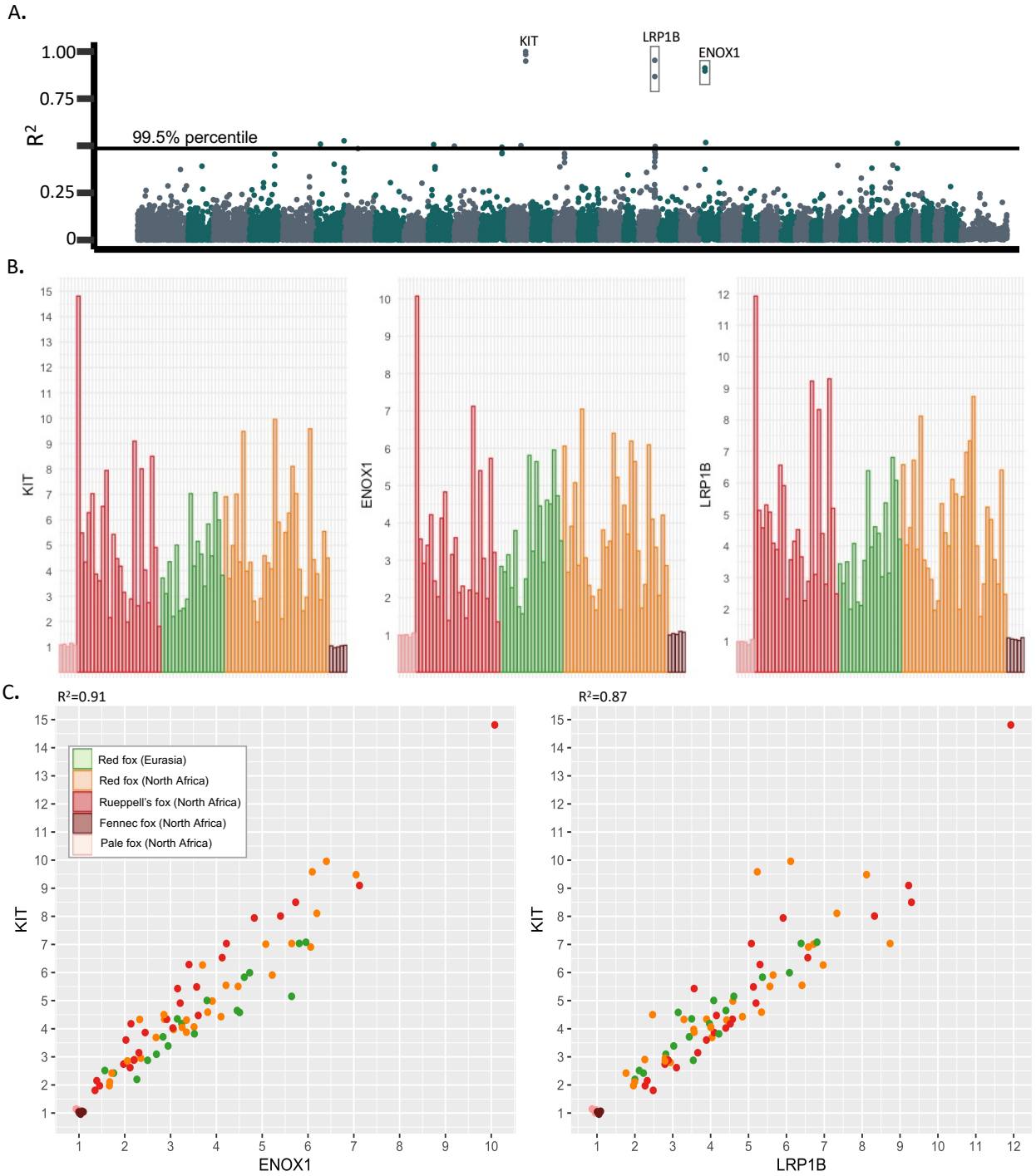


Figure S22. Read depth-based estimate of KIT copy number in 82 foxes and correlation with other genes (related to figure 2B). (A) Manhattan plot of the coefficient of correlation between depth of KIT and remainder regions of the genome on non-overlapping windows of 100kb; (**B**) Read depth-based estimate for KIT, ENOX1 and LRP1B per individual (depth for each gene divided by the genome-wide coverage per individual); (**C**) Scatterplot correlating estimated depth for LRP1B and ENOX1 with that of KIT. Both genes are located in distinct chromosomes and are strongly correlated with KIT, suggesting location in B chromosomes.

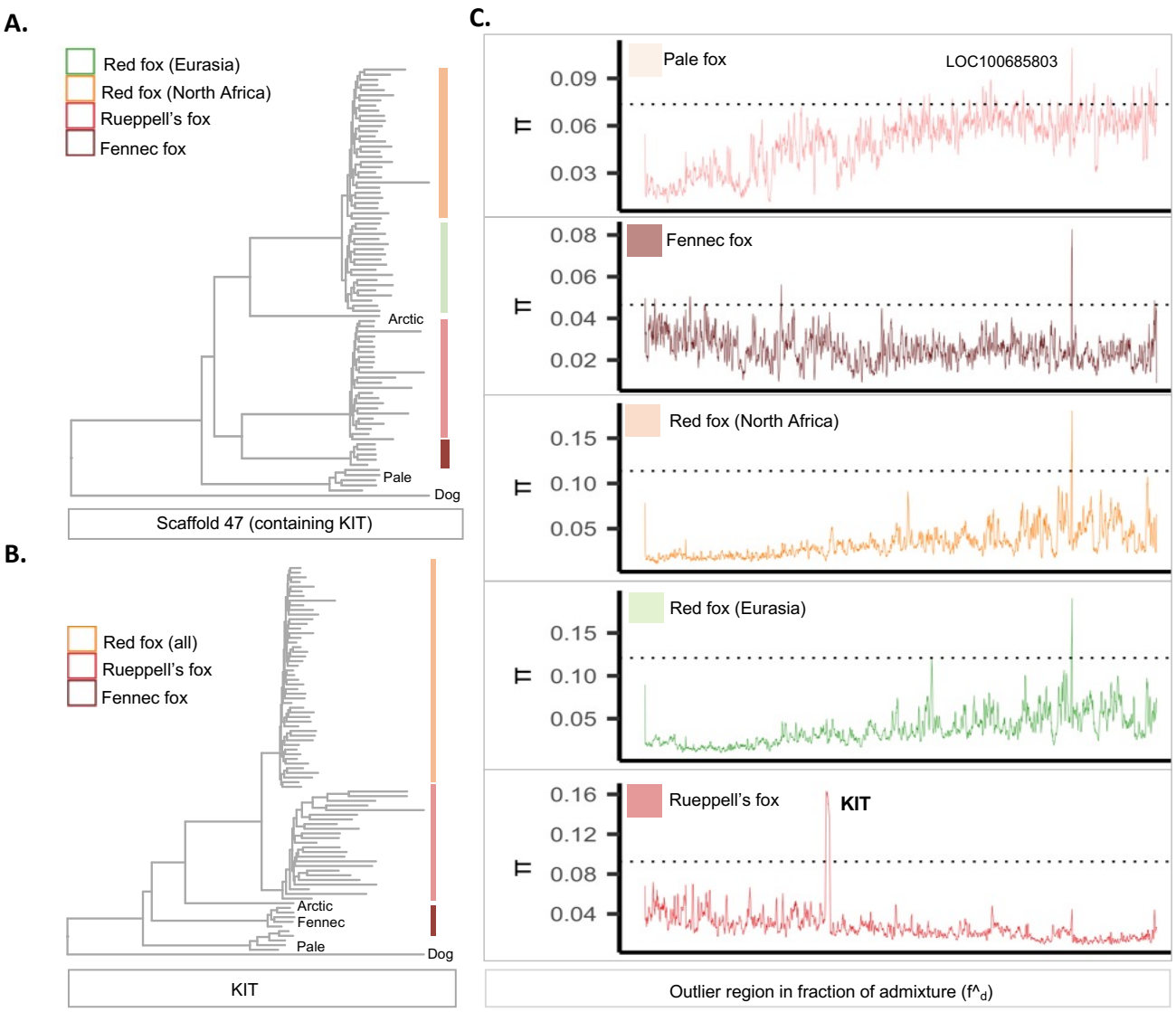
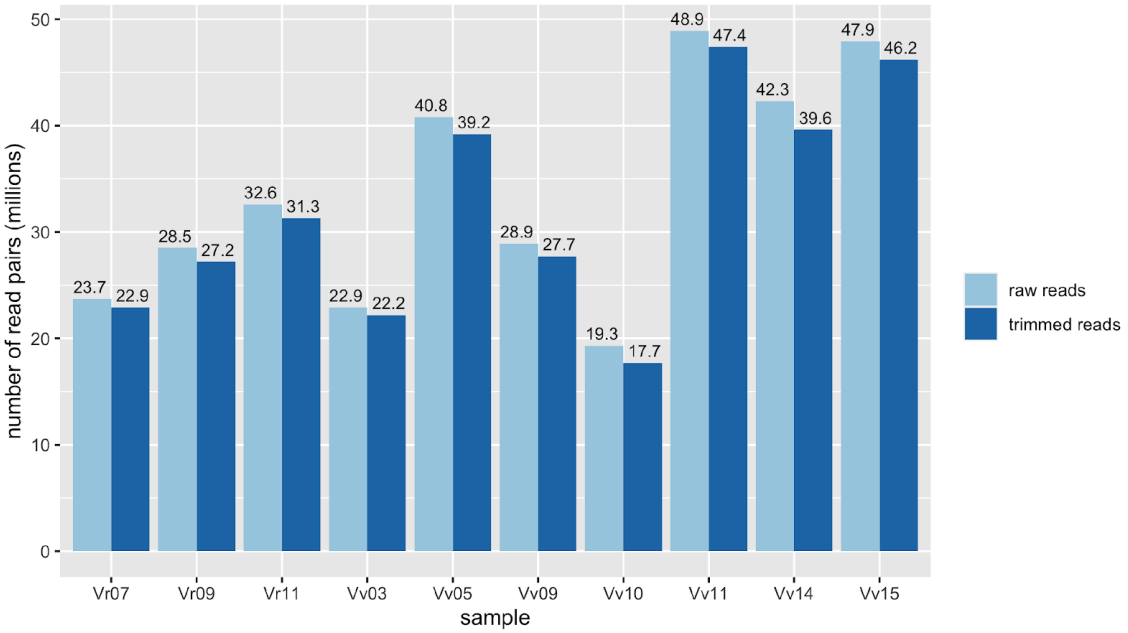


Figure S23. Evidence for introgression of the 25 Mb region supported by KIT (related to Figure 2B). (A) Tree topology for one of the four scaffolds comprising the 25 Mb region above the 99th percentile of the empirical distribution for $f\hat{}_d$ in ABBA-BABA on windows; fennec and Rueppell's are sister species. **(B)** Tree topology for KIT; Rueppell's and red foxes become sister species, as in the genome-wide tree. **(C)** Nucleotide diversity for the ~25Mb 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. 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.

A.



B.

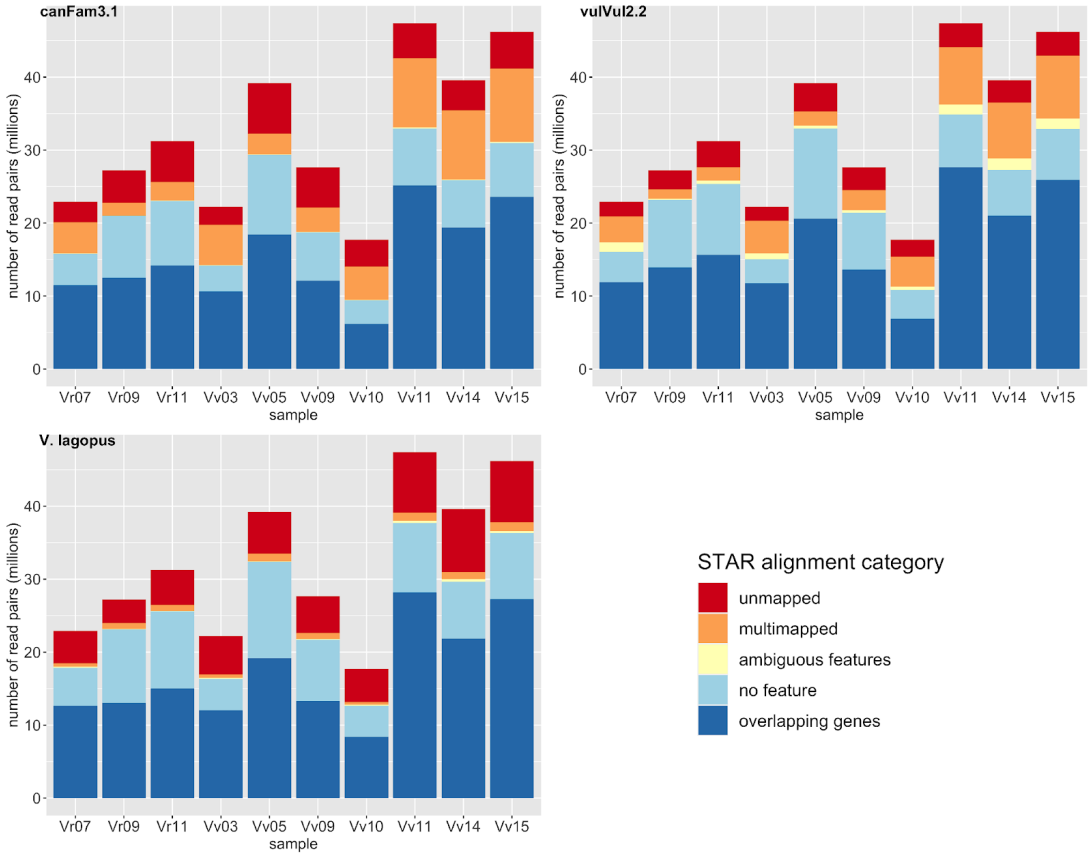


Figure S24. RNA seq quality control. (**A**) Number of read pairs for each sample used in, before and after trimming. (**B)** Number of alignments to each reference genome, as reported by STAR. canFam3.1: dog; vulVul2.2: red fox; *V. lagopus*: arctic fox.

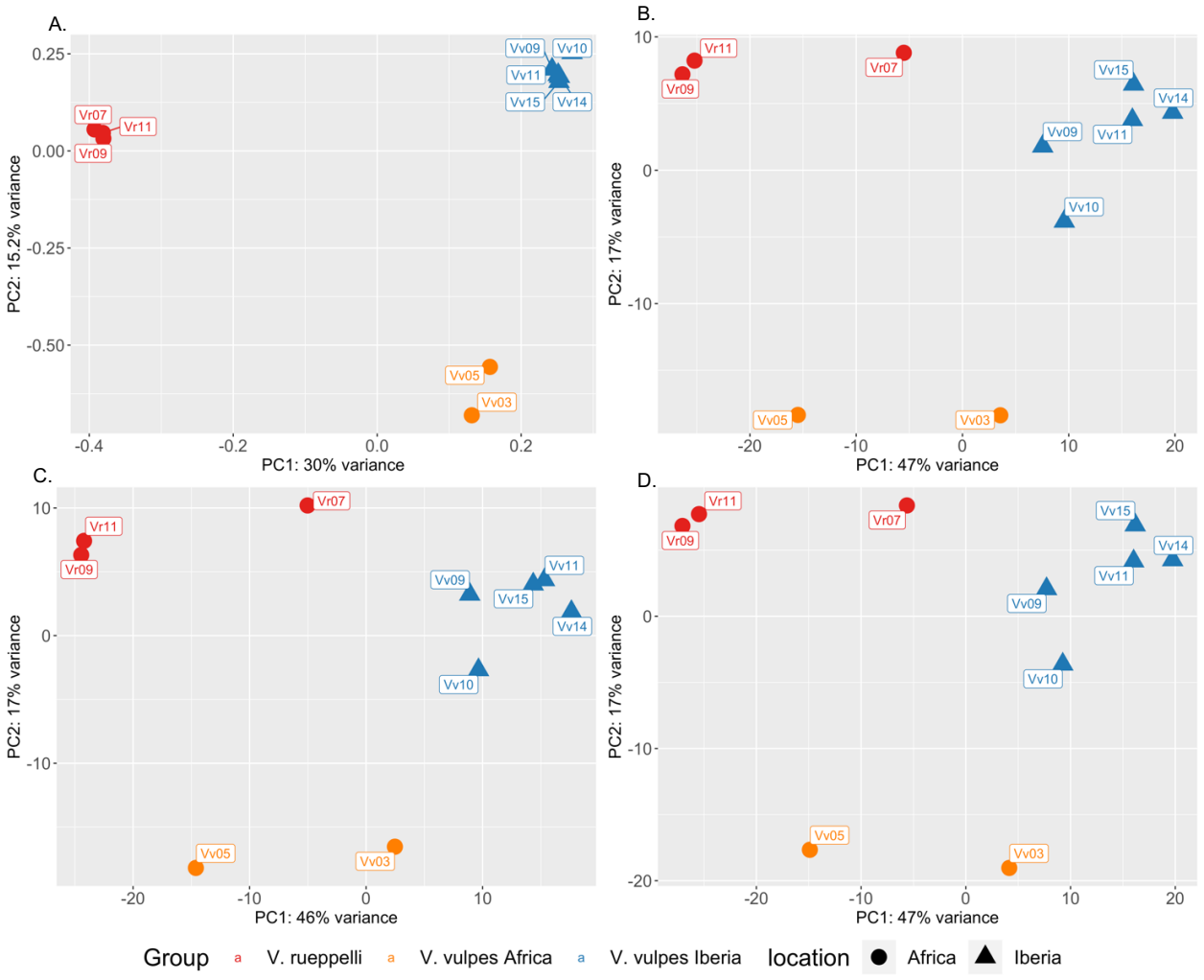


Figure S25. Principal Component Analysis for RNAseq data. (A) PCA based on called genotypes; (**B-D**) PCA based on read counts mapping reads to the transcriptomes of the following references: **(B)** dog (canFam3.1), **(C)** red fox (vulVul2.2) and **(D)** arctic fox (V. lagopus).

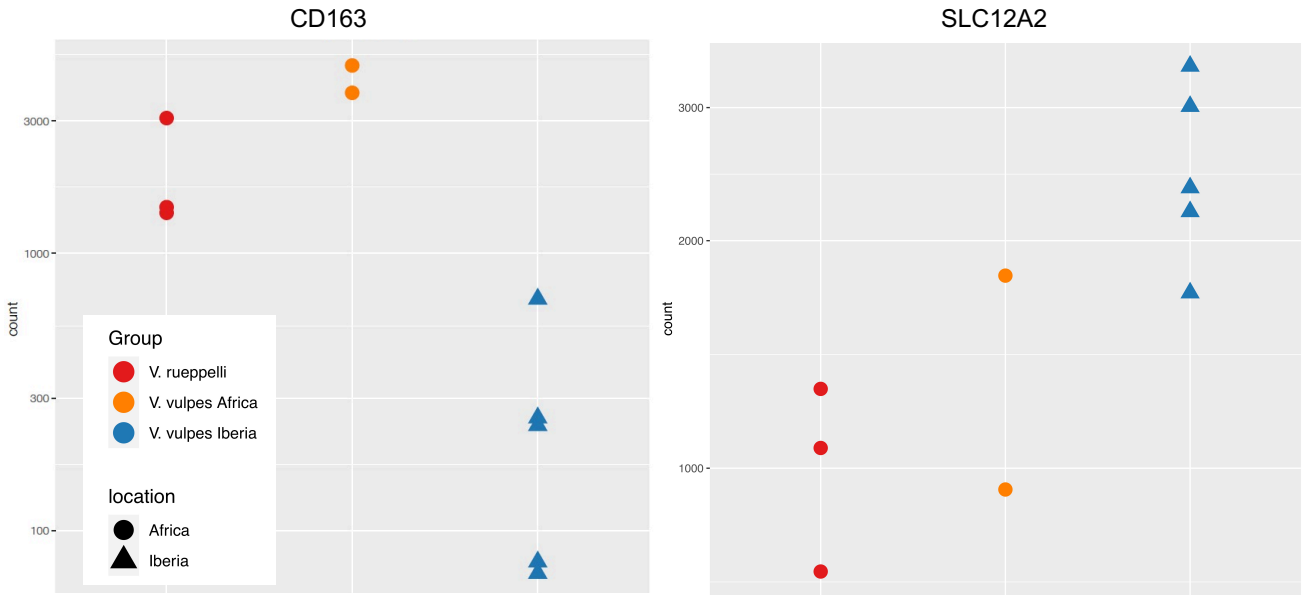


Figure S26. Gene expression differences between North-African and Eurasian red foxes and Rueppell's foxes. The plots represent top candidate genes putatively under selection in North African red foxes (CD163, SLC12A2) for which differential expression was significant between North-African and Eurasian 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.

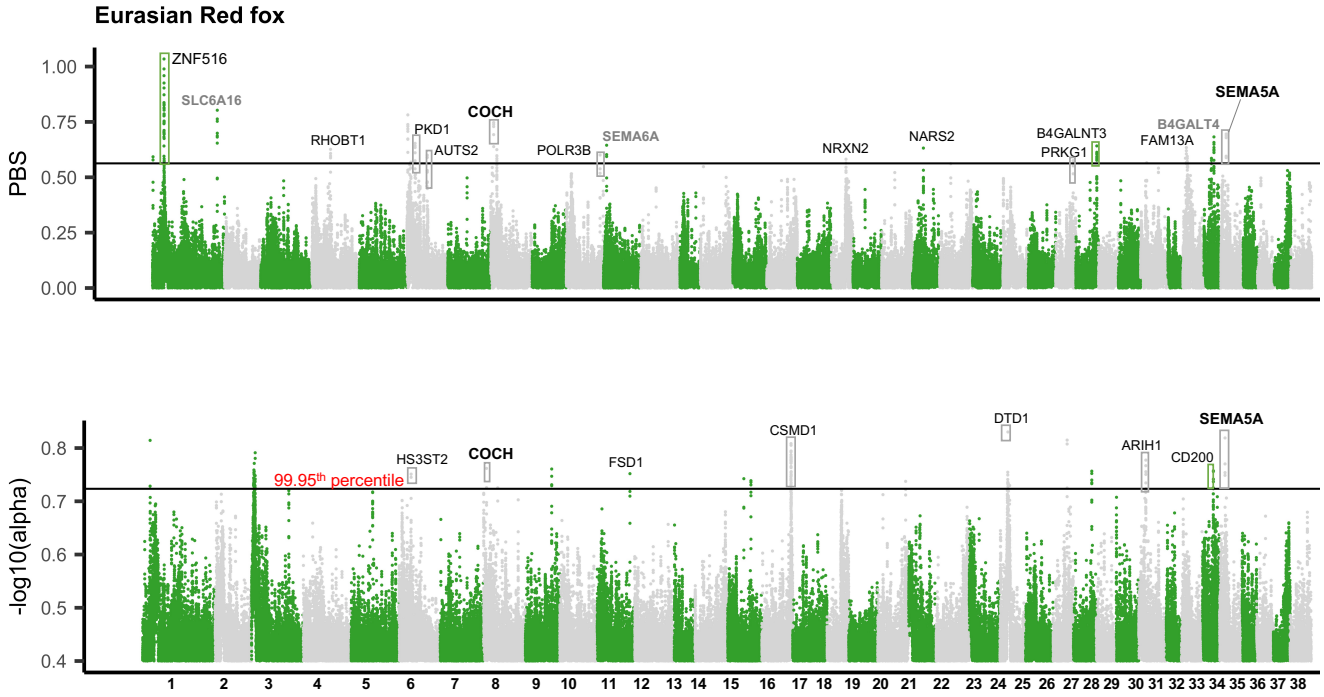


Figure S27. Selection specific to Eurasian red foxes. Genome-wide selection scans on windows of 50 kb with a 10kb slide across autosomal chromosomes, displayed as in the dog genome (x-axis): population branch statistics (PBS) and alpha (negative logarithm of alpha) were estimated using Eurasian red foxes 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 25Mb region identified in Figure 2 in main text were removed from the analysis. Dashed 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 **Tables S16-S17**. The arctic fox (*V. lagopus*) was used as reference genome.

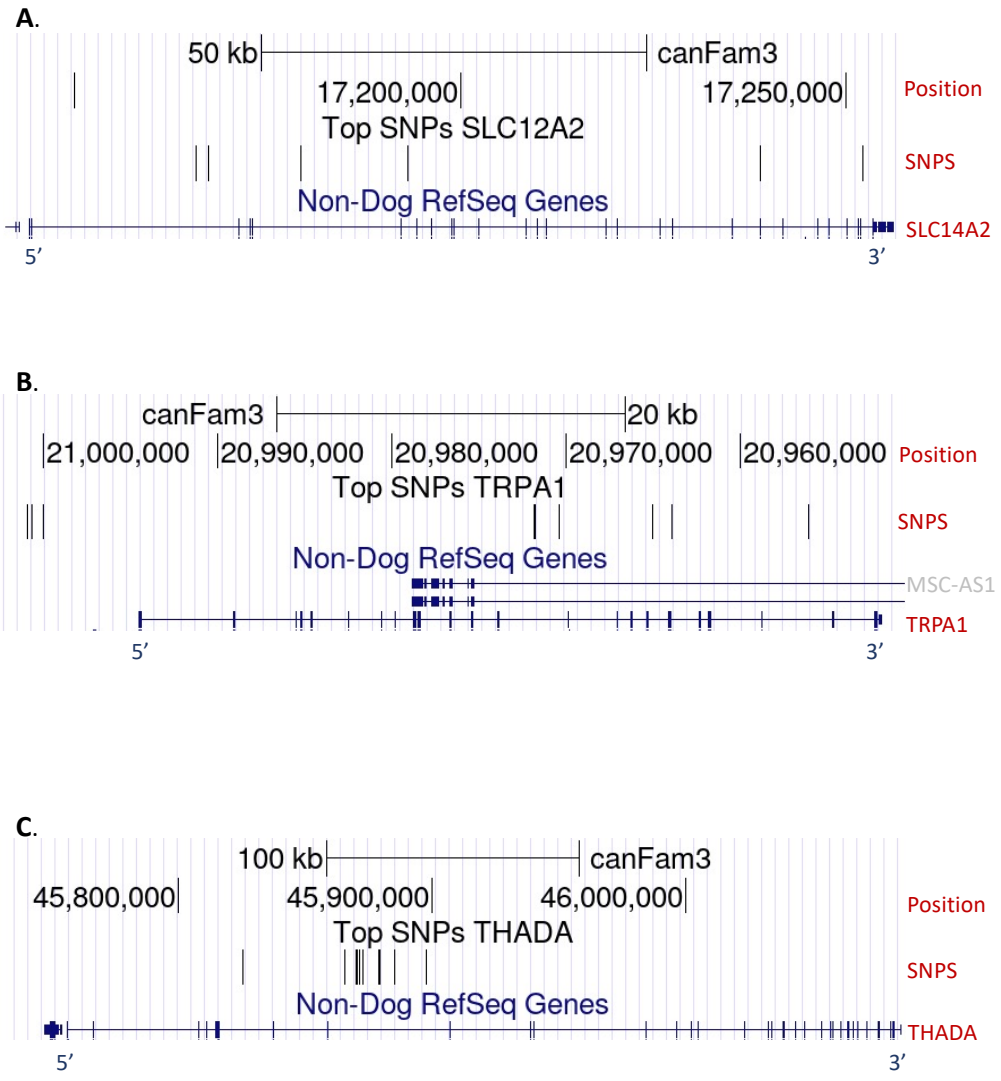


Figure S28 . **Zoomed-in view of the top SNPs harboring regions with highest PBS value for top candidate genes in North-African red foxes (A-C).** The top SNPs with highest PBS value in genes SLC12A2 (**A**), TRPA1 (**B**), and THADA (**C**) are visualized in the UCSC Genome Browser. All variants putatively under selection lie outside non-coding regions.

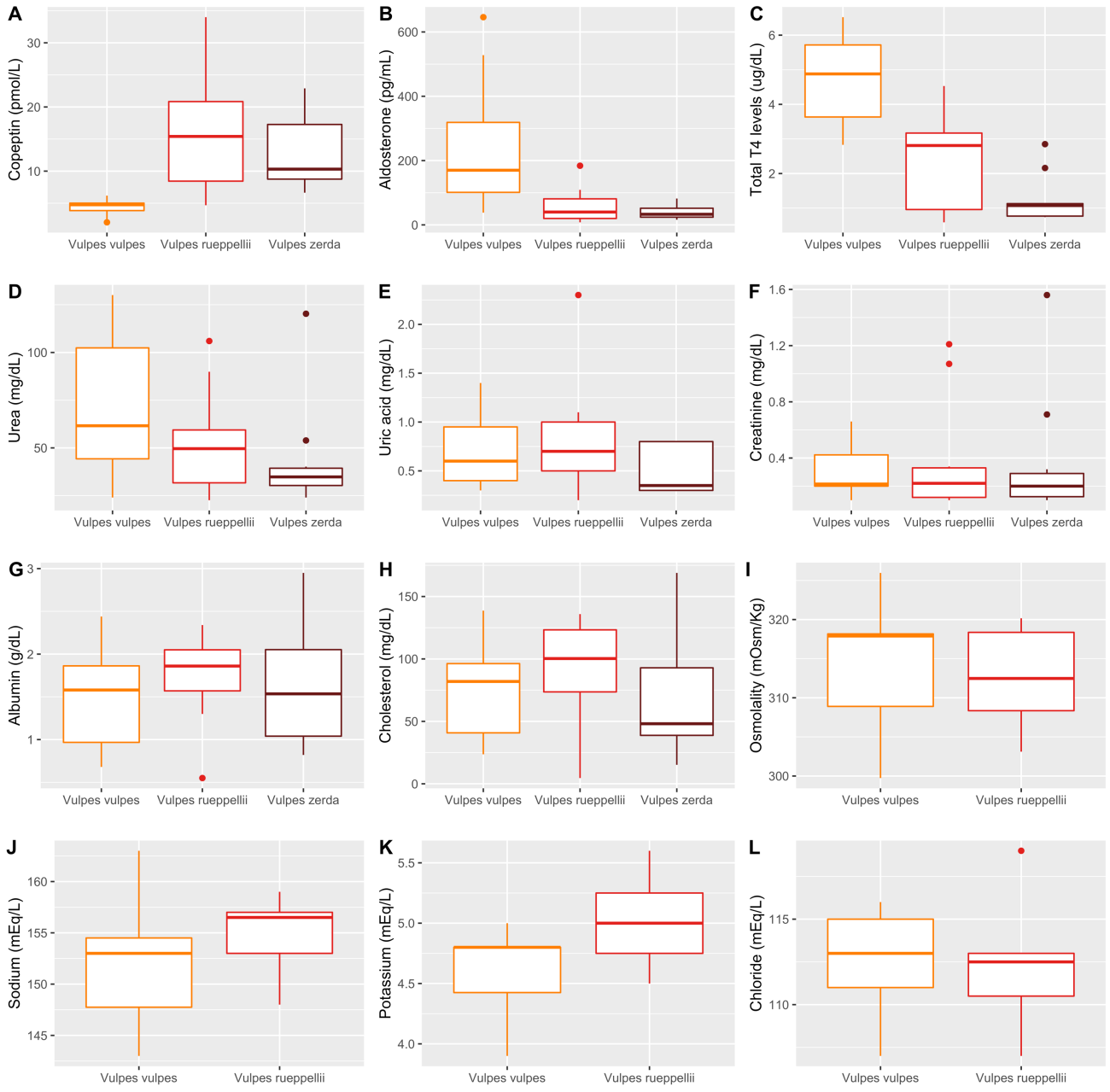


Figure S29. Physiological differences between three species of *Vulpes* (relative to Tables S22-23). Boxplots of 12 physiological parameters measured from the serum/plasma of red fox (*V. vulpes*; n=10), desert-dwelling Rueppell's fox (*V. rueppellii*; n=13), and fennec fox (*V. zerda*; n=10): (A) Copeptin (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. No significant differences between species were found except for A-C (pairwise comparisons using t-tests with Bonferroni correction, $p_{Copeptin}$= 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). 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.

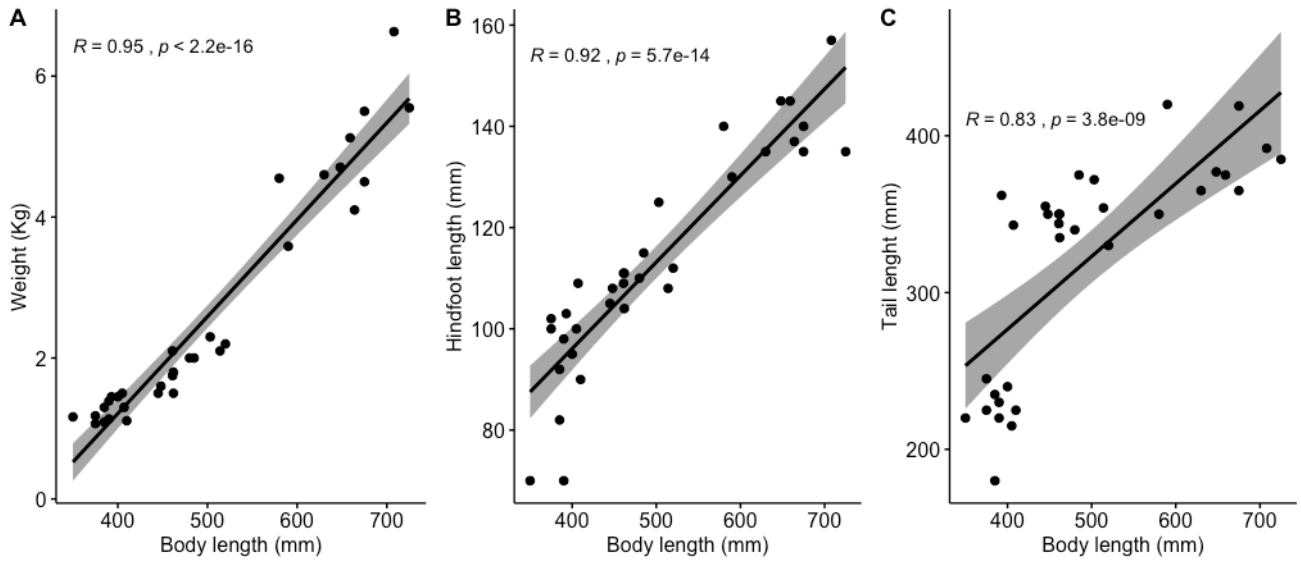


Figure S30. Plots measuring the degree of correlation between body length and weight (A), hindfoot length (B) and tail length (C) in Rueppell's fox, fennec fox and red fox (**Table S3**). For each comparison, the correlation coefficient (R) and associated p-value are shown.

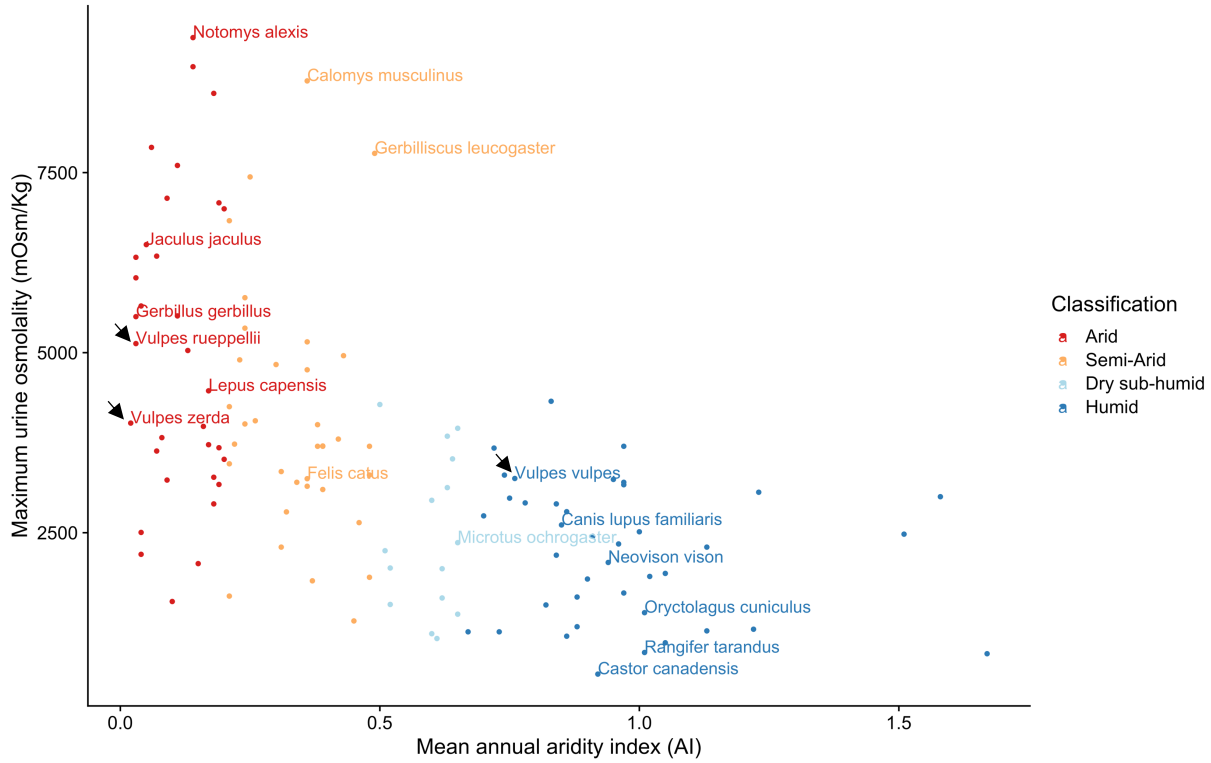


Figure S31. Maximum urine osmolality (mOsm/kg) in mammals and mean aridity index (AI). The values provided in this study for North-African Rueppell's fox (*V. rueppellii*) and red fox from Northern Iberia peninsula (*V. vulpes*) are plotted and highlighted with the fennec fox (*V. zerda*) and 120 mammalian species reviewed in Rocha et al. (2021). Both desert dwelling *Vulpes* species (*V. rueppellii/zerda*) show urine concentrating abilities similar to desert rodents and higher than red foxes.

Supplementary Tables

DATASETS

Table S1. **Whole genome sequencing data**. Information on the whole genome sequenced samples used in this study. The samples average coverage refers to final coverage after mapping to arctic fox genome and quality control.

ID	Species	Continent	Region	Status	Coverage	Featured analysis
84	*V. rueppellii*	North Africa	Mauritania	Roadkill	6.2	all
131	*V. rueppellii*	North Africa	Mauritania	Roadkill	3.5	all
542	*V. rueppellii*	North Africa	Western Sahara	Roadkill	5.0	all
785	*V. rueppellii*	North Africa	Western Sahara	Roadkill	5.7	all
795	*V. rueppellii*	North Africa	Western Sahara	Roadkill	4.9	ngsRelate
2782	*V. rueppellii*	North Africa	Western Sahara	Roadkill	3.9	all
3213	*V. rueppellii*	North Africa	Mauritania	Roadkill	4.0	all
3625	*V. rueppellii*	North Africa	Mauritania	Roadkill	failed	none
4964	*V. rueppellii*	North Africa	Mauritania	Roadkill	4.5	all
5704	*V. rueppellii*	North Africa	Egypt (W. Desert)	Roadkill	3.7	all
6010	*V. rueppellii*	North Africa	Mauritania	Roadkill	5.5	all
7367	*V. rueppellii*	North Africa	Mauritania	Roadkill	4.4	all
7902	*V. rueppellii*	North Africa	Mauritania	Roadkill	2.5	all
8391	*V. rueppellii*	North Africa	Algeria	Roadkill	failed	none
13667	*V. rueppellii*	North Africa	Mauritania	Roadkill	failed	none
VR01	*V. rueppellii*	North Africa	Western Sahara	Captured	9.8	all
VR02	*V. rueppellii*	North Africa	Western Sahara	Captured	5.7	all
VR03	*V. rueppellii*	North Africa	Western Sahara	Captured	4.6	all
VR04	*V. rueppellii*	North Africa	Western Sahara	Captured	5.2	ngsRelate
VR05	*V. rueppellii*	North Africa	Western Sahara	Captured	3.0	all
VR06	*V. rueppellii*	North Africa	Western Sahara	Captured	5.1	all
VR07	*V. rueppellii*	North Africa	Western Sahara	Captured	4.6	all
VR08	*V. rueppellii*	North Africa	Western Sahara	Captured	4.1	all
VR09	*V. rueppellii*	North Africa	Western Sahara	Captured	3.7	all
VR10	*V. rueppellii*	North Africa	Western Sahara	Captured	4.5	all
VR11	*V. rueppellii*	North Africa	Western Sahara	Captured	4.8	all
VR12	*V. rueppellii*	North Africa	Western Sahara	Captured	3.5	all
VR13	*V. rueppellii*	North Africa	Western Sahara	Captured	3.9	all
VR14	*V. rueppellii*	North Africa	Western Sahara	Roadkill	3.9	all
305	*V. vulpes*	North Africa	Tunisia	Roadkill	4.8	all
707	*V. vulpes*	North Africa	Morocco	Roadkill	5.1	all
1092	*V. vulpes*	North Africa	Morocco	Roadkill	8.3	all
1446	*V. vulpes*	North Africa	Morocco	Roadkill	3.2	all
1515	*V. vulpes*	North Africa	Algeria	Captured	3.9	all
2790	*V. vulpes*	Europe	Spain	Roadkill	4.2	all but demography
2794	*V. vulpes*	Europe	Greece	Roadkill	5.1	all but demography
3642	*V. vulpes*	North Africa	Morocco	Roadkill	4.8	all
3643	*V. vulpes*	North Africa	Morocco	Roadkill	3.8	all
5627	*V. vulpes*	Middle East	Turkey	Captured	5.3	all
5701	*V. vulpes*	North Africa	Morocco	Roadkill	4.7	all
5705	*V. vulpes*	Middle East	Egypt (Sinai)	Roadkill	3.6	all

5706	*V. vulpes*	Middle East	Armenia	Roadkill	5.2	all
5710	*V. vulpes*	North Africa	Morocco	Roadkill	4.7	all
6588	*V. vulpes*	North Africa	Morocco	Roadkill	4.2	all
6589	*V. vulpes*	North Africa	Morocco	Roadkill	3.4	all
6820	*V. vulpes*	North Africa	Tunisia	Roadkill	4.2	all
6824	*V. vulpes*	North Africa	Tunisia	Roadkill	5.1	all
6891	*V. vulpes*	North Africa	Morocco	Roadkill	4.2	all
7024	*V. vulpes*	North Africa	Morocco	Roadkill	4.3	all
8030	*V. vulpes*	North Africa	Morocco	Roadkill	3.1	all
8047	*V. vulpes*	Middle East	Iran	Roadkill	5.0	all
8050	*V. vulpes*	Middle East	Iran	Roadkill	5.0	all
8079	*V. vulpes*	North Africa	Morocco	Roadkill	3.9	all
8522	*V. vulpes*	North Africa	Morocco	Roadkill	3.3	all
8645	*V. vulpes*	Middle East	Iran	Roadkill	4.5	all
8953	*V. vulpes*	Middle East	Iran	Roadkill	3.8	all
9053	*V. vulpes*	North Africa	Morocco	Roadkill	3.6	all
9055	*V. vulpes*	North Africa	Morocco	Roadkill	4.3	all
10185	*V. vulpes*	North Africa	Morocco	Roadkill	3.1	all
10470	*V. vulpes*	Middle East	Iran	Roadkill	4.4	all
10476	*V. vulpes*	Middle East	Iran	Roadkill	5.2	all
10608	*V. vulpes*	North Africa	Morocco	Roadkill	4.2	all
11464	*V. vulpes*	Middle East	Iran	Roadkill	3.7	all
11466	*V. vulpes*	Middle East	Iran	Roadkill	4.0	all
11469	*V. vulpes*	Middle East	Iran	Roadkill	5.0	all
11470	*V. vulpes*	Middle East	Iran	Roadkill	5.4	all
12146	*V. vulpes*	North Africa	Morocco	Roadkill	4.7	all
12238	*V. vulpes*	North Africa	Morocco	Roadkill	4.0	all
13451	*V. vulpes*	North Africa	Morocco	Roadkill	3.3	all
VV01	*V. vulpes*	North Africa	Morocco	Captured	3.7	all
VV02	*V. vulpes*	North Africa	Morocco	Captive	3.8	ngsRelate
VV03	*V. vulpes*	North Africa	Morocco	Captive	4.1	all
VV04	*V. vulpes*	North Africa	Morocco	Captive	3.9	all
VV05	*V. vulpes*	North Africa	Morocco	Captive	4.0	all
VV07	*V. vulpes*	Europe	Portugal	Captured	5.2	all but demography
VV10	*V. vulpes*	Europe	Portugal	Captive	4.7	all but demography
VV11	*V. vulpes*	Europe	Portugal	Captive	5.7	all but demography
VV12	*V. vulpes*	North Africa	Western Sahara	Roadkill	3.8	all
VV13	*V. vulpes*	North Africa	Western Sahara	Roadkill	failed	none
VZ01	*V. zerda*	North Africa	Morocco	Captive	3.3	selection, D statistics
VZ02	*V. zerda*	North Africa	Morocco	Captive	3.2	selection, D statistics
VZ03	*V. zerda*	North Africa	Morocco	Captive	4.1	selection, D statistics
VZ05	*V. zerda*	North Africa	Morocco	Captive	3.2	selection, D statistics
VZ07	*V. zerda*	North Africa	Morocco	Captive	3.3	selection, D statistics
468	*V. pallida*	North Africa	Senegal	Roadkill	3.6	selection, D statistics
968	*V. pallida*	North Africa	Mauritania	Roadkill	9.0	selection, D statistics
1001	*V. pallida*	North Africa	Mauritania	Roadkill	5.9	selection, D statistics
3483	*V. pallida*	North Africa	Mauritania	Roadkill	3.7	selection, D statistics
10358	*V. pallida*	North Africa	Senegal	Roadkill	3.9	selection, D statistics

Table S2. Whole genome and mitochondrial sequencing data obtained from NCBI.

ID	Species	Region	Data Type	Featured analysis
KU378369	*V. rueppellii*	Iran	mtDNA, cytb	Haplotype network
KU378368	*V. rueppellii*	Iran	mtDNA, cytb	Haplotype network
KU378373	*V. rueppellii*	Iran	mtDNA, cytb	Haplotype network
KU378372	*V. rueppellii*	Iran	mtDNA, cytb	Haplotype network
KU378370	*V. rueppellii*	Iran	mtDNA, cytb	Haplotype network
KU378371	*V. rueppellii*	Iran	mtDNA, cytb	Haplotype network, BioGeoBEARS
KU378543	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378531	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378551	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378550	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378553	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378554	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378558	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378533	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378538	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378524	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378536	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378530	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378559	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378505	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378504	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378503	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378557	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378563	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378528	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378555	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378556	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378564	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378506	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378513	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378508	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378509	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378510	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378512	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378549	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378547	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378548	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378542	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378546	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378545	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378552	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network

KU378571	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378572	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378573	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378574	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378577	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378578	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378579	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378507	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378511	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378514	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378520	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378521	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378522	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378523	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378561	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378500	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378560	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378562	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378529	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378527	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378526	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
KU378525	*V. vulpes*	Iran	mtDNA, cytb	Haplotype network
GQ374180	*V. vulpes*	China	mtDNA, complete	Haplotype network
KF387633	*V. vulpes*	Tibet	mtDNA, complete	Haplotype network
JN711443	*V. vulpes*	South Korea	mtDNA, complete	Haplotype network
MN122913	*V. vulpes*		mtDNA, complete	Haplotype network
NC_008434	*V. vulpes*		mtDNA, complete	Haplotype network
KU378587.1	*V. cana*		cytb	BioGeoBEARS
L11734.2	*V. velox*		cytb	BioGeoBEARS
JF489127.1	*V. macrotis*		cytb	BioGeoBEARS
KT033906	*V. ferrilata*	Tibet	mtDNA, complete	BEAST, BioGeoBEARS
NC_027935	*V. ferrilata*	Tibet	mtDNA, complete	BEAST
KJ140137	*V. corsac*	Mongolia	mtDNA, complete	BEAST
NC_023958	*V. corsac*	Mongolia	mtDNA, complete	BEAST, BioGeoBEARS
NC_045533	*V. bengalensis*	India	mtDNA, complete	BEAST, BioGeoBEARS
KJ603240	*V. zerda*		mtDNA, complete	BEAST, BioGeoBEARS
NC_026529.1	*V. lagopus*		mtDNA, complete	BEAST, BioGeoBEARS
KP342451.1	*V. lagopus*		mtDNA, complete	BEAST
NC_013700	*Nyctereutes procyonoides*		mtDNA, complete	BEAST, BioGeoBEARS
SRR7704840	*V. lagopus*		genome	ngsDist, gene trees
SRR5280501	*V. vulpes*	Russian farm	genome	CRADD gene tree
SRR5280502	*V. vulpes*	Russian farm	genome	CRADD gene tree
SRR5280503	*V. vulpes*	Russian farm	genome	CRADD gene tree

SRR5280504	*V. vulpes*	Russian farm	genome	CRADD gene tree
SRR5280505	*V. vulpes*	Russian farm	genome	CRADD gene tree
SRR5280506	*V. vulpes*	Russian farm	genome	CRADD gene tree
SRR5280507	*V. vulpes*	Russian farm	genome	CRADD gene tree
SRR5280508	*V. vulpes*	Russian farm	genome	CRADD gene tree
SRR5280509	*V. vulpes*	Russian farm	genome	CRADD gene tree
SRR5280510	*V. vulpes*	Russian farm	genome	CRADD gene tree
SRR5280486	*V. vulpes*	Russian farm	genome	CRADD gene tree
SRR5280487	*V. vulpes*	Russian farm	genome	CRADD gene tree
SRR5280488	*V. vulpes*	Russian farm	genome	CRADD gene tree
SRR5280489	*V. vulpes*	Russian farm	genome	CRADD gene tree
SRR5280490	*V. vulpes*	Russian farm	genome	CRADD gene tree
SRR5280491	*V. vulpes*	Russian farm	genome	CRADD gene tree
SRR5280492	*V. vulpes*	Russian farm	genome	CRADD gene tree
SRR5280493	*V. vulpes*	Russian farm	genome	CRADD gene tree
SRR5280494	*V. vulpes*	Russian farm	genome	CRADD gene tree
SRR5280495	*V. vulpes*	Russian farm	genome	CRADD gene tree
SRR5280476	*V. vulpes*	Russian farm	genome	CRADD gene tree
SRR5280477	*V. vulpes*	Russian farm	genome	CRADD gene tree
SRR5280478	*V. vulpes*	Russian farm	genome	CRADD gene tree
SRR5280479	*V. vulpes*	Russian farm	genome	CRADD gene tree
SRR5280480	*V. vulpes*	Russian farm	genome	CRADD gene tree
SRR5280481	*V. vulpes*	Russian farm	genome	CRADD gene tree
SRR5280482	*V. vulpes*	Russian farm	genome	CRADD gene tree
SRR5280483	*V. vulpes*	Russian farm	genome	CRADD gene tree
SRR5280484	*V. vulpes*	Russian farm	genome	CRADD gene tree
SRR5280485	*V. vulpes*	Russian farm	genome	CRADD gene tree
SRR1138360	*Canis familiaris*	Tibetan Mastiff	genome	ngsDist, gene trees
SRR8049195	*Canis lupus*	Greenland	genome	CRADD gene tree
SRR8049186	*Canis latrans*	Alaska	genome	CRADD gene tree
SRR8049196	*Canis anthus*	Africa	genome	CRADD gene tree
SRR2149878	*Canis aureus*	Eurasia	genome	CRADD gene tree
SRR7874817	*Lycaon pictus*	Zimbabwe	genome	CRADD gene tree
SRR1066703	*Lycalopex culpaeus*	South America	genome	CRADD gene tree
SRR5198016 SRR5198017 SRR5198018	*Urocyon littoralis*	USA	genome	CRADD gene tree
SRS1937014	*Urocyon cinereoargenteus*	USA	genome	CRADD gene tree
SRR942310	*Ursus maritimus*		genome	CRADD gene tree

Table S3. **Phenotypic and Transcriptomic data collected in this study.** Abbreviations are as follows: F- female, M – Male, NA – not applicable.

ID	Species	Status	Age	Sex	Trap	Region	Territory	Selected
VR01	*V. rueppellii*	wild	Adult	F	Tomahawk	North Africa	Morocco	Phenotyping
VR02	*V. rueppellii*	wild	Adult	M	Leghold	North Africa	Morocco	Phenotyping
VR03	*V. rueppellii*	wild	Adult	F	Tomahawk	North Africa	Morocco	Phenotyping RNA-seq*
VR04	*V. rueppellii*	wild	Adult	M	Tomahawk	North Africa	Morocco	Phenotyping
VR05	*V. rueppellii*	wild	Adult	F	Tomahawk	North Africa	Morocco	Phenotyping
VR06	*V. rueppellii*	wild	Adult	M	Tomahawk	North Africa	Morocco	Phenotyping
VR07	*V. rueppellii*	wild	Adult	F	Tomahawk	North Africa	Morocco	Phenotyping RNA-seq
VR08	*V. rueppellii*	wild	Adult	M	Leghold	North Africa	Morocco	Phenotyping
VR09	*V. rueppellii*	wild	Adult	M	Leghold	North Africa	Morocco	Phenotyping RNA-seq
VR10	*V. rueppellii*	wild	Adult	M	Tomahawk	North Africa	Morocco	Phenotyping
VR11	*V. rueppellii*	wild	Adult	M	Tomahawk	North Africa	Morocco	Phenotyping RNA-seq
VR12	*V. rueppellii*	wild	Adult	M	Tomahawk	North Africa	Morocco	Phenotyping
VR13	*V. rueppellii*	wild	Adult	F	Tomahawk	North Africa	Morocco	Phenotyping
VV01	*V. vulpes*	wild	Juvenile	M	Collarum	North Africa	Morocco	Phenotyping
VV02	*V. vulpes*	captive	Adult	F	NA	North Africa	Morocco	Phenotyping
VV03	*V. vulpes*	captive	Adult	M	NA	North Africa	Morocco	Phenotyping RNA-seq
VV04	*V. vulpes*	captive	Adult	M	NA	North Africa	Morocco	Phenotyping
VV05	*V. vulpes*	captive	Juvenile	M	NA	North Africa	Morocco	Phenotyping RNA-seq
VV06	*V. vulpes*	wild	Adult	F	Leghold	Iberia	Portugal	Phenotyping
VV07	*V. vulpes*	wild	Adult	F	Leghold	Iberia	Portugal	Phenotyping
VV08	*V. vulpes*	wild	Adult	M	Leghold	Iberia	Portugal	Phenotyping
VV09	*V. vulpes*	captive	Adult	F	NA	Iberia	Portugal	Phenotyping RNA-seq
VV10	*V. vulpes*	captive	Adult	F	NA	Iberia	Portugal	Phenotyping RNA-seq
VV11	*V. vulpes*	captive	Adult	F	NA	Iberia	Portugal	Phenotyping RNA-seq
VV14	*V. vulpes*	wild	Adult	NA	Leghold	Iberia	Spain	RNA-seq
VV15	*V. vulpes*	wild	Adult	NA	Leghold	Iberia	Spain	RNA-seq
VZ01	*V. zerda*	captive	Adult	M	NA	North Africa	Morocco	Phenotyping
VZ02	*V. zerda*	captive	Adult	M	NA	North Africa	Morocco	Phenotyping
VZ03	*V. zerda*	captive	Adult	F	NA	North Africa	Morocco	Phenotyping
VZ04	*V. zerda*	captive	Adult	M	NA	North Africa	Morocco	Phenotyping
VZ05	*V. zerda*	captive	Adult	M	NA	North Africa	Morocco	Phenotyping
VZ06	*V. zerda*	captive	Adult	M	NA	North Africa	Morocco	Phenotyping
VZ07	*V. zerda*	captive	Adult	F	NA	North Africa	Morocco	Phenotyping
VZ08	*V. zerda*	captive	Adult	M	NA	North Africa	Morocco	Phenotyping
VZ09	*V. zerda*	captive	Adult	M	NA	North Africa	Morocco	Phenotyping
VZ10	*V. zerda*	captive	Adult	M	NA	North Africa	Morocco	Phenotyping

*Failed due to low RIN score.

PHYLOGENETIC ANALYSES

Table S4. Ancestral Range reconstruction in BioGeoBEARS using complete and cytb mtDNA maximum-likelihood trees. Model comparison for dispersal-extinction-cladogenesis (DEC), dispersal-vicariance-like (DIVAKIKE), and Bayesian Inference of historical biogeography for discrete areas (BAYAREALIKE), with and without founder-event speciation (with or without +J).

Models	**Ln likelihood**	**Number of Parameters**	**d**	**e**	**j**	**AIC**	**AICc**
Complete mitochondrial genomes with 8 *Vulpes*							
DEC	-18.06	2	3.48	1.00E-12	0	40.13	41.84
DEC+J	**-14.40**	**3**	**2.21**	**1.00E-12**	**0.15**	**34.80**	**38.8**
DIVALIKE	-18.03	2	4.53	1.00E-12	0	40.07	41.78
DIVALIKE+J	-14.55	3	2.48	1.00E-12	0.12	35.12	39.12
BAYAREALIKE	-21.20	2	2.88	5	0	46.40	48.12
BAYAREALIKE + J	-14.56	3	2.48	1.00E-12	0.122	35.12	39.12
Cytochrome B-based with 11 *Vulpes*							
DEC	-24.74	2.00	1.76	0.00	0.00	53.48	54.57
DEC+J	**-16.87**	**3.00**	**0.49**	**0.00**	**0.18**	**39.74**	**42.14**
DIVALIKE	-25.71	2.00	2.79	0.00	0.00	55.42	56.51
DIVALIKE+J	-18.63	3.00	0.63	0.00	0.21	43.25	45.65
BAYAREALIKE	-33.48	2.00	1.25	5.00	0.00	70.95	72.04
BAYAREALIKE + J	-18.63	3.00	0.63	0.00	0.21	43.25	45.65

Abbreviations are as follows: d, rate of dispersal; e, rate of extinction; j, likelihood of founder event speciation at cladogenesis; AIC, Akaike's information criterion; AICc, Akaike's information criterion, with correction for sample size.

Table S5. **Posterior estimates of the time to the most recent common ancestor (T_{MRCA}) of *Vulpes* mtDNA.** The median node age for *Vulpes sp.* based on complete mitochondrial genomes is followed by the 95% highest posterior density intervals.

Node	Median	95% HPD
T_{MRCA} (*V. pallida* + other *Vulpes sp.*)	7.06 mya	7.0-7.2 mya
T_{MRCA} (*V. zerda* + other *Vulpes sp.*)	5.50 mya	4.8-6.2 mya
T_{MRCA} (*V. lagopus* + other *Vulpes sp.*)	3.21 mya	3.0-3.4 mya
T_{MRCA} (*V. ferrilata* + *V. corsac* + other *Vulpes sp.*)	2.44 mya	2.1-2.8 mya
T_{MRCA} (*V. bengalensis* + *V. rueppellii* + *V. vulpes*)	1.24 mya	927 kya-1.6 mya
T_{MRCA} (*V. ferrilata* + *V. corsac*)	0.99 mya	689 kya-1.3 mya
T_{MRCA} (*V. rueppellii* + *V. vulpes*)	625 kya	476-788 kya

INTROGRESSION AND DEMOGRAPHY

Table S6. D-statistics as implemented in ANGSD. All D-statistics were computed using either arctic fox (*Vulpes lagopus*) or dog (*Canis lupus familiaris*) as reference genomes and with different outgroups. Only significant cases for significant allele sharing between H2 and H3 are presented. Abbreviations are as follows: vvNA – red fox North Africa, vvEU – red fox Eurasia, vr – Rueppell's fox, vz – fennec fox, vp – pale fox.

***D* statistic**	**Z score**	**p-value**	**nABBA**	**nBABA**	**nBlocks**	**H1 (A)**	**H2 (B)**	**H3 (B)**	**H4 (A)**	**Ref**
0.033774	39.63	0.000001	899736	840945	4090	vvEU	vvNA	vr	arctic	arctic
0.031359	59.15	0.000000	889746	835639	4035	vvEU	vvNA	vr	vz	arctic
0.029141	54.28	0.000000	901861	850787	4065	vvEU	vvNA	vr	vp	arctic
0.028694	42.53	0.000000	1015207	958571	1465	vvEU	vvNA	vr	vz	dog
0.024829	36.89	0.000000	1056349	1005163	1615	vvEU	vvNA	vr	vp	dog
0.008460	8.64	0.000000	1222508	1201997	1772	vvEU	vvNA	vr	dog	dog
0.015214	10.08	0.000000	593273	575491	3780	vv	vr	vz	vp	arctic
0.012988	6.02	0.000000	689425	671746	1390	vv	vr	vz	vp	dog
0.010753	5.91	0.000000	940562	920550	1471	vv	vr	vz	dog	dog

Table S7. Demographic Inference. Joint demographic parameters inferred with *dadi* using GADMA for the history of divergence between Rueppell's fox and red fox, and between North-Africa and Middle Eastern red foxes. The mutation rate (μ) was set to 4.5 x 10^{-9} and generation time (gt) of 2 years (divergence time estimates are also provided for generation time of 3 years).

Model parameters	**Rueppell's and red fox North Africa**	**Red fox North-Africa and Middle East**
Loglikelihood	-9570062.64*	-1545461.83
Ancestral size	162,186	149,241
Time of ancestral size change (kya) according to generation time (gt)	NA	gt_2=244 gt_3=366
Ancestral size change	NA	995,897
Divergence time (kya) according to generation time (gt)	gt_2=576 gt_3=863	gt_2=78 gt_3=117
Population size after divergence of ancestral	vr: 27,053 vvNA: 135,132	vvME: 979,010 vvNA: 16,887
Migration Rates (migrants/generation)	vvNA to vr: 2.46E-06 vr to vvNA: 1.93E-06	vvNA to vvME: 1.93E-05 vvME to vvNA: 3.74E-05
Current population size	vr: 186,048 vvNA: 173,135	vvME: 12,730 vvNA: 295,225

*The Logelikelihood of a model with no migration is of -170222606.78

Table S8 – Genomic regions above the 99th percentile of empirical distribution of the fraction of admixture (f^d) on non-overlapping windows of 200 kb using red fox (*V. vulpes*) Eurasia as H1, red fox North Africa as H2 and Rueppell's fox (*V. rueppellii*) as H3, outgrouped by arctic fox (*V. lagopus*).

Dog chr	Arctic	start	end	f^hom	f^d	overlapping genes
Chr15	scaffold_3	30400000	30599999	0.77	0.77	CRADD
X	scaffold_69	12200000	12399999	0.55	0.55	GPM6B,OFD1,RAB9A,TCEANC,TRAPPC2
Chr1	scaffold_99	4600000	4799999	0.48	0.48	C1H19orf73,CD37,HRC,PPFIA3,SLC6A16,TEAD2,TRPM4
Chr8	scaffold_63	200000	399999	0.40	0.40	LOC100855826
Chr18	scaffold_61	13400000	13599999	0.40	0.40	COR4C1P,COR4C31,COR4C32,COR4C38,OR14D11
Chr22	scaffold_38	21000000	21199999	0.40	0.40	ALG11,ATP7B,NEK5
Chr19	scaffold_14	2000000	2199999	0.37	0.37	AMMECR1L,NDUFA, ISOC2,POLR2D,WDR33
Chr21	scaffold_39	10200000	10399999	0.28	0.28	HPS5,SAA1,SAAL1
Chr9	scaffold_59	13800000	13999999	0.27	0.27	LOC100686518
Chr33	scaffold_25	8600000	8799999	0.27	0.27	B4GALT4,C33H3orf30,IGSF11,UPK1B
Chr3	scaffold_15	23000000	23199999	0.25	0.25	LOC100686296
Chr3	scaffold_121	2600000	2799999	0.24	0.24	GABRG3
Chr18	scaffold_27	25200000	25399999	0.24	0.24	LOC102157067,LOC106557501,LOC111091656,LOC111097868,SEPT14
Chr38	scaffold_136	600000	799999	0.23	0.23	CD1A6,CD1B,CD1C,CD1E,OR10R3B,OR10R3C,OR10R3,OR10T1,OR10T2,OR10K2
Chr16	scaffold_30	14600000	14799999	0.23	0.23	KCNU1
Chr4	scaffold_36	7800000	7999999	0.23	0.23	NPR3
Chr21	scaffold_39	0	199999	0.23	0.23	NLRP14,OR10F05,ZNF214,ZNF215
Chr15	scaffold_3	30200000	30399999	0.22	0.22	PLXNC1
Chr7	scaffold_7	1200000	1399999	0.22	0.22	DNAH14
Chr16	scaffold_127	600000	799999	0.22	0.22	CLEC5A,MGAM,MGAM2,TAS2R38
Chr22	scaffold_6	2000000	2199999	0.21	0.21	LOC111091646
Chr32	scaffold_46	7600000	7799999	0.20	0.20	HSD17B11,HSD17B13,KLHL8
Chr21	scaffold_22	21600000	21799999	0.19	0.19	GDPD4,LOC111091524,PAK1
Chr30	scaffold_5	25000000	25199999	0.19	0.19	FAM227B,FGF7,GALK2
Chr7	scaffold_7	36400000	36599999	0.19	0.19	NEK7
Chr21	scaffold_168	1400000	1599999	0.18	0.18	OR2D3
Chr7	scaffold_7	15400000	15599999	0.18	0.18	ANKRD45,CENPL,DARS2,KLHL20,SLC9C2,TEX50
Chr27	scaffold_107	1000000	1199999	0.18	0.18	LMO3,LOC102152174
Chr12	scaffold_95	6200000	6399999	0.18	0.18	.
Chr8	scaffold_142	2800000	2999999	0.18	0.18	LOC611720
X	scaffold_101	1200000	1399999	0.18	0.18	LOC100856306
Chr7	scaffold_7	5200000	5399999	0.17	0.17	CATSPERE,DESI2
Chr1	scaffold_106	3000000	3199999	0.17	0.17	C1H9orf64,HNRNPK,KIF27,MIR7-1,RMI1
Chr8	scaffold_24	2400000	2599999	0.17	0.17	.
Chr28	scaffold_64	11200000	11399999	0.17	0.17	MARCH8,WASHC2C,ZFAND4
Chr15	scaffold_3	25200000	25399999	0.16	0.16	ACTR6,DEPDC4,SCYL2,UHRF1BP1L
Chr22	scaffold_38	11600000	11799999	0.16	0.16	KBTBD7,MTRF1,NAA16
Chr27	scaffold_74	11800000	11999999	0.16	0.16	MUC19
Chr15	scaffold_3	10400000	10599999	0.16	0.16	LOC111098920
Chr19	scaffold_41	15800000	15999999	0.16	0.16	LOC111091096
Chr8	scaffold_142	2600000	2799999	0.22	0.16	ADAM20,ADAM21,MAD2L1
Chr32	scaffold_55	13400000	13599999	0.16	0.16	LOC102151449,LOC102153596,LOC102154615
Chr1	scaffold_99	3800000	3999999	0.16	0.16	CARD8,CCDC114,LIG1,ZNF114,ZSWIM9
Chr22	scaffold_38	12600000	12799999	0.16	0.16	AKAP11
Chr18	scaffold_27	14000000	14199999	0.16	0.16	NAMPT
Chr37	scaffold_16	10800000	10999999	0.16	0.16	LOC102154178
Chr22	scaffold_6	4200000	4399999	0.15	0.15	LOC111091751

Chr25	scaffold_48	6200000	6399999	0.15	0.15	DOCK10
Chr19	scaffold_41	9000000	9199999	0.15	0.15	LOC102157193
Chr3	scaffold_45	17400000	17599999	0.15	0.15	LOC111095398
Chr14	scaffold_0	0	199999	0.15	0.15	SLC13A1
Chr18	scaffold_27	4000000	4199999	0.15	0.15	LOC111090859
Chr27	scaffold_67	3400000	3599999	0.15	0.15	CSRNP2,HIGD1C,LETMD1,POU6F1,SLC11A2,TFCP2
Chr19	scaffold_41	11400000	11599999	0.15	0.15	PCDH10
Chr23	scaffold_72	600000	799999	0.15	0.15	LOC111091977,PLOD2,PLSCR4
Chr11	scaffold_52	4800000	4999999	0.15	0.15	INIP,KIAA1958,LOC111089963
Chr18	scaffold_27	21600000	21799999	0.15	0.15	CACNA2D1
Chr22	scaffold_6	32200000	32399999	0.15	0.15	.
Chr23	scaffold_56	14200000	14399999	0.15	0.15	LOC100687988,LOC111092988
Chr18	scaffold_27	0	199999	0.15	0.15	ABCA13,C18H7orf57,SUN3
Chr20	scaffold_1	24800000	24999999	0.15	0.15	APPL1,ASB14,DNAH12
Chr26	scaffold_187	400000	599999	0.15	0.15	LOC100683217,LOC100856527,LOC102155737
Chr13	scaffold_43	14600000	14799999	0.15	0.15	MRPL13,MTBP,SNTB1
Chr2	scaffold_53	12200000	12399999	0.14	0.14	CCDC7
Chr10	scaffold_4	7400000	7599999	0.14	0.14	CCT4,COMMD1
Chr16	scaffold_117	1800000	1999999	0.14	0.14	CNTNAP2
Chr26	scaffold_84	7400000	7599999	0.14	0.14	BCL7A,IL31,LOC100683019,MLXIP,WDR66
Chr27	scaffold_67	13000000	13199999	0.14	0.14	LOC102157056,LOC102157106
Chr7	scaffold_8	11400000	11599999	0.14	0.14	EMILIN2,SMCHD1
Chr20	scaffold_1	5600000	5799999	0.14	0.14	MBD3L1
Chr25	scaffold_119	3200000	3399999	0.14	0.14	HMBOX1,INTS9,LOC111092294
Chr27	scaffold_107	600000	799999	0.14	0.14	.
Chr5	scaffold_17	6200000	6399999	0.14	0.14	DDX10
Chr16	scaffold_158	0	199999	0.14	0.14	AGA,LOC102151763,NEIL3
Chr16	scaffold_117	4400000	4599999	0.14	0.14	CASP2,CLCN1,EPHA1,FAM131B,TAS2R41,TAS2R60,ZYX
Chr12	scaffold_12	20000000	20199999	0.14	0.14	IL17A,IL17F,LOC100687519,MCM3,PAQR8
Chr22	scaffold_38	5200000	5399999	0.14	0.14	DIAPH3
Chr6	scaffold_9	9200000	9399999	0.14	0.14	AGL,MFSD14A,SASS6,SLC35A3
Chr18	scaffold_27	21000000	21199999	0.14	0.14	LOC100687935
Chr12	scaffold_12	7600000	7799999	0.13	0.13	DNAH8,GLO1
X	scaffold_101	1600000	1799999	0.13	0.13	.
Chr36	scaffold_166	800000	999999	0.13	0.13	GALNT13
Chr11	scaffold_29	15200000	15399999	0.13	0.13	.
Chr17	scaffold_54	200000	399999	0.13	0.13	.
Chr25	scaffold_119	2800000	2999999	0.13	0.13	FBXO16,FZD3
Chr15	scaffold_3	29600000	29799999	0.13	0.13	NDUFA12,NR2C1,TMCC3
Chr25	scaffold_119	1400000	1599999	0.13	0.13	STMN4
Chr22	scaffold_38	20800000	20999999	0.13	0.13	ATP7B,CCDC70,WDFY2
Chr25	scaffold_48	10800000	10999999	0.13	0.13	SP100,SP140
Chr13	scaffold_43	6200000	6399999	0.13	0.13	ADCY8,LOC111098451

Table S9 – Gene ontology enrichment based on the regions in the top 99th percentile of the empirical distribution of the fraction of admixture (f^d) calculated using ABBA-BABA on windows for potential introgression between Rueppell's fox and North-African red fox. FDR p-values are non-significant.

Description of bilogical process	Gene count	P-value	FDR p-value
Gene Ontology using a ranked list in http://geneontology.org/			
renal system process (GO:0003014)	2	1.77E-03	1.00E+00
superior temporal gyrus development (GO:0071109)	1	1.96E-03	1.00E+00
inorganic ion transmembrane transport (GO:0098660)	4	3.50E-03	1.00E+00
protein localization to juxtaparanode region of axon (GO:0071205)	1	3.92E-03	1.00E+00
membrane depolarization during bundle of His cell action potential (GO:0086048)	1	3.92E-03	1.00E+00
anterior head development (GO:0097065)	1	3.92E-03	1.00E+00
osteoclast proliferation (GO:0002158)	1	4.89E-03	1.00E+00
bundle of His cell action potential (GO:0086043)	1	4.89E-03	1.00E+00
bundle of His cell to Purkinje myocyte signaling (GO:0086028)	1	4.89E-03	1.00E+00
inorganic anion transmembrane transport (GO:0098661)	2	5.23E-03	1.00E+00
regulation of calcium ion transmembrane transport via high voltage-gated calcium channel (GO:1902514)	1	5.87E-03	1.00E+00
protein localization to axon (GO:0099612)	1	5.87E-03	1.00E+00
observational learning (GO:0098597)	1	5.87E-03	1.00E+00
imitative learning (GO:0098596)	1	5.87E-03	1.00E+00
vocal learning (GO:0042297)	1	5.87E-03	1.00E+00
protein O-linked glycosylation via threonine (GO:0018243)	1	6.84E-03	1.00E+00
learned vocalization behavior or vocal learning (GO:0098598)	1	7.82E-03	1.00E+00
protein O-linked glycosylation via serine (GO:0018242)	1	8.79E-03	1.00E+00
positive regulation of urine volume (GO:0035810)	1	8.79E-03	1.00E+00
auditory behavior (GO:0031223)	1	8.79E-03	1.00E+00
membrane depolarization during cardiac muscle cell action potential (GO:0086012)	1	8.79E-03	1.00E+00
renal water homeostasis (GO:0003091)	1	9.76E-03	1.00E+00
protein localization to endosome (GO:0036010)	1	9.76E-03	1.00E+00
mechanosensory behavior (GO:0007638)	1	9.76E-03	1.00E+00

There are no GO terms with significant enrichment empirical p-values for GOrilla enrichment analysis.

Table S10 – Genomic regions above the 99th percentile of the empirical distribution of the fraction of admixture (f^d) on non-overlapping windows of 200 kb with red fox (*V. vulpes*) as H1, Rueppell's fox (*V. rueppellii*) as H2 and fennec (*V. zerda*) as H3, outgrouped by pale fox (*V. pallida*).

Dog chr	Arctic	start	end	f^hom	f^d	overlapping genes
Chr13	scaffold_47	11000000	11199999	0.32	0.30	.
Chr13	scaffold_47	1800000	1999999	0.30	0.30	LOC111098609
Chr13	scaffold_47	5400000	5599999	0.30	0.30	LNX1
Chr13	scaffold_47	14400000	14599999	0.29	0.28	.
Chr13	scaffold_47	3600000	3799999	0.28	0.28	FRYL
Chr13	scaffold_47	2800000	2999999	0.28	0.27	ATP10D,CORIN
Chr13	scaffold_47	4600000	4799999	0.26	0.26	USP46
Chr22	scaffold_6	41000000	41199999	0.26	0.26	.
Chr13	scaffold_47	4000000	4199999	0.26	0.26	CWH43,DCUN1D4
Chr13	scaffold_47	14200000	14399999	0.26	0.26	LOC612669
Chr13	scaffold_47	8600000	8799999	0.25	0.25	.
Chr13	scaffold_47	400000	599999	0.25	0.25	GNPDA2,GUF1,YIPF7
Chr13	scaffold_47	14600000	14799999	0.25	0.25	TECRL
Chr13	scaffold_47	800000	999999	0.25	0.25	LOC100687522,LOC111098466,LOC111098608
Chr13	scaffold_47	6200000	6399999	0.24	0.24	.
Chr13	scaffold_126	1200000	1399999	0.23	0.23	ANKRD17,COX18
Chr13	scaffold_126	1800000	1999999	0.25	0.23	ADAMTS3,NPFFR2
Chr13	scaffold_47	2200000	2399999	0.23	0.23	GABRA4,GABRB1
Chr13	scaffold_47	4200000	4399999	0.23	0.23	LRRC66,SGCB,SPATA18
Chr13	scaffold_126	2000000	2199999	0.23	0.23	NPFFR2
Chr13	scaffold_47	12000000	12199999	0.22	0.22	ADGRL3,LOC102153251
Chr13	scaffold_47	1400000	1599999	0.22	0.22	GABRG1
Chr13	scaffold_47	16400000	16599999	0.21	0.21	.
Chr13	scaffold_47	15400000	15599999	0.22	0.21	EPHA5
Chr13	scaffold_47	16600000	16799999	0.20	0.20	LOC100684630
Chr13	scaffold_126	3000000	3199999	0.20	0.20	ENAM,GRSF1,JCHAIN,RUFY3,UTP3
Chr13	scaffold_47	11600000	11799999	0.20	0.19	LOC111098628
Chr13	scaffold_47	1600000	1799999	0.19	0.19	GABRA2
Chr13	scaffold_126	2600000	2799999	0.20	0.19	SLC4A4
Chr13	scaffold_47	5600000	5799999	0.19	0.18	LOC111098427,LOC111098617
Chr13	scaffold_47	9400000	9599999	0.19	0.18	LOC100683616
Chr13	scaffold_200	200000	399999	0.18	0.18	LOC111098461,LOC111098462,LOC111098463
Chr13	scaffold_47	9600000	9799999	0.18	0.18	LOC111098627
Chr21	scaffold_22	0	199999	0.17	0.17	LOC111091471
Chr13	scaffold_47	15200000	15399999	0.16	0.16	.
Chr13	scaffold_47	4800000	4999999	0.16	0.16	RASL11B,SCFD2
Chr13	scaffold_155	400000	599999	0.16	0.15	GRXCR1
Chr13	scaffold_47	5800000	5999999	0.15	0.15	CHIC2,GSX2,LOC100856597
Chr13	scaffold_47	6000000	6199999	0.15	0.15	PDGFRA
Chr13	scaffold_47	200000	399999	0.15	0.15	.
Chr13	scaffold_155	800000	999999	0.14	0.14	ATP8A1,SHISA3
Chr13	scaffold_47	4400000	4599999	0.14	0.14	LOC102152757
Chr13	scaffold_47	600000	799999	0.14	0.14	LOC111098465,LOC111098608
Chr13	scaffold_155	1000000	1199999	0.14	0.14	BEND4,SLC30A9
Chr13	scaffold_47	7200000	7399999	0.14	0.14	NMU
Chr13	scaffold_47	16800000	16999999	0.15	0.14	.
Chr13	scaffold_47	1200000	1399999	0.13	0.13	LOC111098467,LOC111098468
Chr3	scaffold_13	31200000	31399999	0.13	0.13	CHRNA7
Chr13	scaffold_47	14800000	14999999	0.13	0.12	LOC100684472,LOC102153987
Chr16	scaffold_158	1800000	1999999	0.12	0.12	LOC102153584
Chr33	scaffold_25	19400000	19599999	0.11	0.11	LOC102156904

Chr13	scaffold_47	0	199999	0.11	0.11	KCTD8
Chr13	scaffold_47	6800000	6999999	0.12	0.11	CLOCK,SRD5A3,TMEM165
Chr13	scaffold_47	1000000	1199999	0.11	0.11	.
Chr13	scaffold_155	0	199999	0.11	0.11	LOC100684895,LOC102156700
Chr13	scaffold_126	600000	799999	0.11	0.11	CXCL8,PPBP
Chr13	scaffold_47	3200000	3399999	0.10	0.10	LOC106559633,TEC,TXK
Chr21	scaffold_168	0	199999	0.10	0.10	COR56A20,COR56A24, cOR56A13P
Chr13	scaffold_155	1200000	1399999	0.10	0.10	PHOX2B,TMEM33
Chr13	scaffold_47	8000000	8199999	0.10	0.10	ARL9,HOPX,SRP72,THEGL
Chr13	scaffold_47	9200000	9399999	0.10	0.10	LOC102153092
Chr10	scaffold_71	3200000	3399999	0.10	0.10	HMGA2,LLPH,TMBIM4
Chr13	scaffold_47	8400000	8599999	0.09	0.09	IGFBP7,POLR2B

Table S11 – Gene ontology enrichment analysis for the ~25Mb chromosomic region detected in ABBA-BABA analysis on windows for potential introgression from a fennec-like species into Rueppell's fox.

Molecular function	p-value	FDR p-value	Genes	Function of individual genes
Gene Ontology using a ranked list in http://geneontology.org/				
aryl sulfotransferase activity	1.21E-05	1.70E-02	SULT1B1	Sulfotransferase enzyme that sulfates dopamine, small phenols such as 1-naphthol and p-nitrophenol and thyroid hormones, including thyroxine. Involved in thyroid hormone metabolic process.
			SULT1D1	Sulfotransferase involved in catecholamine metabolic process. Sult1d1 in mice is directly induced by glucocorticoids and may attenuate elevated catecholamine activity during the stress response and metabolic fasting.
			SULT1E1	Sulfotransferase involved in estrogen metabolic process
Neuromedin, U receptor binding	1.32E-04	4.66E-02	NMU	The encoded protein is a precursor that is proteolytically processed to generate a biologically active neuropeptide that plays a role in pain, stress, immune-mediated inflammatory diseases and feeding regulation.
protein tyrosine kinase activity	1.79E-05	1.89E-02	KIT	This gene encodes a receptor tyrosine kinase. c-KIT activation has been shown to have oncogenic activity in gastrointestinal stromal tumors, melanomas, lung cancer, and other tumor types. It is also associated with melanocyte differentiation and pigmentation.
			KDR	This gene encodes a tyrosine kinase receptor that plays an essential role in the regulation of angiogenesis, vascular development, vascular permeability, and embryonic hematopoiesis.
			EPHA5	Receptor tyrosine kinase involved in the ephrin receptor signaling pathway. Eph receptors are implicated in many processes, particularly cancer development and progression.
			PDGFRA	Tyrosine-protein kinase that plays a role in organ development, wound healing, and tumor progression.
			TEC	Non-receptor tyrosine kinase that is involved in regulation of the adaptive immune response

			TXK	Non-receptor tyrosine kinase that plays a redundant role with ITK in regulation of the adaptive immune response.
transmembrane receptor protein tyrosine kinase activity	1.54E-04	4.65E-02	KIT, KDR, EPHA5, PDGFRA	
ligand-gated ion channel activity	4.08E-05	1.91E-02	GABRB1 GABRA4 GABRG1 GABRA2	Components of the heteropentameric receptor for GABA, the major inhibitory neurotransmitter in the vertebrate brain. GABA mediates neuronal inhibition by binding to the GABA/benzodiazepine receptor and opening an integral chloride channel.
ligand-gated channel activity	4.08E-05	2.15E-02	CNGA1	Subunit of the rod cyclic GMP-gated cation channel, which is involved in the final stage of the phototransduction pathway. Involved in visual perception.
GABA receptor activity	4.67E-06	9.86E-03	GABRB1, GABRA4, GABRG1,GABRA2	
GABA-A receptor activity	2.66E-06	1.12E-02		
GABA-gated chloride ion channel activity	3.61E-05	2.18E-02	GABRA4, GABRG1,GABRA2	
benzodiazepine receptor activity	2.85E-05	2.40E-02		
inhibitory extracellular ligand-gated ion channel activity	7.98E-05	3.37E-02		
ligand-gated anion channel activity	1.29E-04	4.94E-02		
CXCR chemokine receptor binding	2.85E-05	2.00E-02	CXCL8, CXCL7	Major mediators of the inflammatory response
interleukin-8 receptor binding	1.32E-04	4.30E-02	CXCL8	Diseases associated with CXCL8 include Melanoma and Adult Respiratory Distress Syndrome.
GOrilla results using two unranked lists of genes (target and background)				

GABA-gated chloride ion channel activity	7.47E-07	3.41E-03	GABRG1, GABRB1, GABRA2, GABRA4
ligand-gated anion channel activity	3.12E-06	7.13E-03	GABRG1, GABRB1, GABRA2, GABRA4
GABA-A receptor activity	3.94E-06	5.99E-03	GABRG1, GABRB1, GABRA2, GABRA4
GABA receptor activity	7.34E-06	8.37E-03	GABRG1, GABRB1, GABRA2, GABRA4
benzodiazepine receptor activity	3.07E-05	2.80E-02	GABRG1 , GABRA2, GABRA4
inhibitory extracellular ligand-gated ion channel activity	8.33E-05	6.33E-02	GABRG1, GABRA2, GABRA4

SELECTION IN NORTH AFRICAN RED FOXES

Table S12. Genomic regions above the 99.95th percentile of the PBS empirical distribution for North African red fox, using Eurasian red foxes as closely related ingroup and Rueppell's fox as outgroup (removing the CRADD region and the 25 Mb chromosomic introgressed region). For overlapping windows the window with highest PBS value is represented. The arctic fox was used as reference. The same outliers are obtained when using the dog as reference (not shown).

Dog Chr	Arctic ID	start	end	pbs	overlapping genes
Chr35	scaffold_31	15610000	15660000	0.980	.
Chr27	scaffold_80	2300000	2350000	0.962	CD163
Chr27	scaffold_80	7230000	7280000	0.904	B4GALNT3
Chr8	scaffold_75	9210000	9260000	0.869	DHRS1,RPL27A,LTB4R2,NOP9,RABGGTA, TGM1
Chr3	scaffold_13	30260000	30310000	0.844	MCEE,MPHOSPH10
X	scaffold_35	6830000	6880000	0.830	.
Chr6	scaffold_115	1220000	1270000	0.827	SDK1
Chr14	scaffold_0	51270000	51320000	0.812	.
Chr4	scaffold_23	23310000	23360000	0.783	ARID5B
X	scaffold_62	2900000	2950000	0.774	OR13H2P, OR13K7P
Chr14	scaffold_0	51260000	51310000	0.774	.
Chr10	scaffold_71	9360000	9410000	0.769	.
Chr11	scaffold_29	3920000	3970000	0.754	SLC12A2
Chr10	scaffold_4	23870000	23920000	0.754	THADA
Chr29	scaffold_37	21000000	21050000	0.727	TRPA1
Chr35	scaffold_31	8890000	8940000	0.716	ATXN1
X	scaffold_35	12860000	12910000	0.713	ATP6AP2
X	scaffold_146	1850000	1900000	0.703	SH3BGRL
Chr32	scaffold_46	6220000	6270000	0.701	FAM13A
Chr26	scaffold_96	2850000	2900000	0.700	PRKG1
Chr1	scaffold_28	17710000	17760000	0.700	.
Chr10	scaffold_4	6810000	6860000	0.689	EHBP1
Chr16	scaffold_92	7080000	7130000	0.684	SUMO2
Chr10	scaffold_4	43430000	43480000	0.682	MAFF,TMEM184B
Chr34	scaffold_19	18090000	18140000	0.681	SI
Chr32	scaffold_46	7630000	7680000	0.678	HSD17B11,HSD17B13
Chr9	scaffold_57	550000	600000	0.678	CACNA1B
Chr29	scaffold_37	20990000	21040000	0.677	.
Chr4	scaffold_36	6490000	6540000	0.677	BRIX1,DNAJC21,RAD1
Chr6	scaffold_151	770000	820000	0.676	TRRAP
X	scaffold_76	2210000	2260000	0.675	.
Chr16	scaffold_92	7550000	7600000	0.675	IRF2
Chr10	scaffold_71	9270000	9320000	0.672	SRP14
Chr8	scaffold_24	8110000	8160000	0.670	.
Chr32	scaffold_46	8150000	8200000	0.670	PTPN13
Chr26	scaffold_129	1220000	1270000	0.669	SDA1, ZNF74
Chr35	scaffold_31	12850000	12900000	0.666	HIVEP1
Chr32	scaffold_46	10770000	10820000	0.660	.
Chr1	scaffold_28	20420000	20470000	0.659	FAM69C
X	scaffold_62	9430000	9480000	0.659	TENM1
Chr11	scaffold_29	8120000	8170000	0.657	HSPA4

Table S13. Genomic regions above the 99.95th percentile of the negative logarithm of alpha for North African red fox as focal group, using Eurasian red foxes as closely related ingroup and Rueppell's fox as outgroup (removing CRADD and the 25MB region). For overlapping windows the window with highest value is represented. The arctic fox was used as reference. The same outliers are obtained when using the dog as reference (not shown).

Dog chr	Arctic	start	end	-log10(alpha)	overlapping genes
Chr27	scaffold_80	2300001	2350001	0.95	CD163,
Chr11	scaffold_29	3920001	3970001	0.94	SLC12A2
Chr29	scaffold_37	21000001	21050001	0.94	TRPA1
Chr17	scaffold_54	890001	940001	0.93	KIDINS220
Chr1	scaffold_2	32530001	32580001	0.91	GRM1
Chr3	scaffold_121	2810001	2860001	0.86	GABRG3
Chr14	scaffold_0	43530001	43580001	0.86	AKAP9
Chr8	scaffold_24	550001	600001	0.86	.
Chr12	scaffold_12	6880001	6930001	0.86	ZFAND3
Chr1	scaffold_28	20420001	20470001	0.85	FAM69C
X	scaffold_35	12850001	12900001	0.84	.
Chr1	scaffold_28	21130001	21180001	0.83	TSHZ1
Chr16	scaffold_92	7080001	7130001	0.82	LOC100682962
X	scaffold_164	960001	1010001	0.82	.
Chr3	scaffold_13	5760001	5810001	0.82	FGFR3,LETM1
Chr38	scaffold_10	22510001	22560001	0.82	TTC17
Chr6	scaffold_115	1140001	1190001	0.82	SDK1
Chr35	scaffold_31	16120001	16170001	0.82	LDHB
Chr35	scaffold_31	6020001	6070001	0.82	.
Chr29	scaffold_40	4650001	4700001	0.82	.
Chr1	scaffold_2	41890001	41940001	0.82	PDE4D
X	scaffold_112	2200001	2250001	0.82	DACH2
X	scaffold_156	340001	390001	0.81	CENPN
Chr1	scaffold_28	17170001	17220001	0.81	RBMS1,RTTN
Chr35	scaffold_31	12850001	12900001	0.81	HIVEP1
X	scaffold_138	460001	510001	0.81	.
Chr8	scaffold_24	780001	830001	0.81	.
Chr16	scaffold_118	4040001	4090001	0.80	.
Chr32	scaffold_46	7640001	7690001	0.80	HSD17B13
Chr5	scaffold_66	3000001	3050001	0.80	ALOX12B,ALOXE3,TRNAK-UUU,TRNAL-UAG,TRNAQ-CUG
Chr5	scaffold_17	4610001	4660001	0.80	GUCY1A2,DPY30
Chr10	scaffold_4	23870001	23920001	0.80	THADA
Chr16	scaffold_92	7550001	7600001	0.80	IRF2
Chr22	scaffold_6	28930001	28980001	0.80	GNB1,TM9SF2
Chr38	scaffold_10	31170001	31220001	0.80	CCDC73,SETSIP
Chr6	scaffold_128	1250001	1300001	0.79	COG7,SCNN1B
Chr29	scaffold_40	16460001	16510001	0.79	SGK3
X	scaffold_35	10870001	10920001	0.79	RPGR

Table S14. Gene ontology enrichment based on the regions in the top 99.95 percentile of the empirical distribution of the negative logarithm of alpha, with North African red fox as the focal group.

GO term	Description of biological process	Gene count	P-value	FDR p-value
Gene Ontology using a ranked list in http://geneontology.org/				
GO:0030104	water homeostasis	3	6.94E-05	2.44E-01
GO:0051121	hepoxilin metabolic process	2	5.62E-05	2.64E-01
GO:0043651	linoleic acid metabolic process	2	9.62E-05	2.71E-01
GO:0019372	lipoxygenase pathway	2	1.20E-04	2.82E-01
GO:0051122	hepoxilin biosynthetic process	2	5.62E-05	3.96E-01
GO:0042759	long-chain fatty acid biosynthetic process	2	2.08E-04	4.18E-01
GO:0048878	chemical homeostasis	7	2.62E-04	4.62E-01
GO:0050878	regulation of body fluid levels	4	3.89E-04	5.48E-01
GO:0070633	transepithelial transport	2	3.60E-04	5.64E-01
GO:0050891	multicellular organismal water homeostasis	3	4.98E-05	7.01E-01
GO:0061436	establishment of skin barrier	2	6.66E-04	8.53E-01
GO:0055065	metal ion homeostasis	5	7.63E-04	8.95E-01
GO:0033561	regulation of water loss via skin	2	8.53E-04	9.24E-01
GO:0055078	sodium ion homeostasis	2	1.14E-03	1.00E+00
GO:0065008	regulation of biological quality	12	1.22E-03	1.00E+00
GO:0055080	cation homeostasis	5	1.31E-03	1.00E+00
GO:0098771	inorganic ion homeostasis	5	1.41E-03	1.00E+00
GO:0019369	arachidonic acid metabolic process	2	1.46E-03	1.00E+00
GO:0050801	ion homeostasis	5	1.89E-03	1.00E+00
GO:0019233	sensory perception of pain	2	1.92E-03	1.00E+00
GOrilla results using two unranked lists of genes (target and background)				
GO:0019233	sensory perception of pain	3	9.23E-05	4.67E-01
GO:0007600	sensory perception	6	1.47E-04	5.58E-01
GO:0051121	hepoxilin metabolic process	2	8.17E-05	6.21E-01
GO:0050891	multicellular organismal water homeostasis	2	2.70E-04	8.19E-01
GO:0019372	lipoxygenase pathway	2	4.04E-04	8.78E-01
GO:0030104	water homeostasis	3	3.53E-04	8.93E-01
GO:0048878	chemical homeostasis	8	6.17E-04	9.37E-01
GO:0061436	establishment of skin barrier	2	8.03E-04	9.38E-01
GO:0042759	long-chain fatty acid biosynthetic process	2	8.82E-04	9.57E-01
GO:0055065	metal ion homeostasis	6	6.97E-04	9.63E-01
GO:0033561	regulation of water loss via skin	2	9.65E-04	9.77E-01
GO:0032501	multicellular organismal process	14	7.75E-04	9.81E-01
GO:0043651	linoleic acid metabolic process	2	5.23E-04	9.92E-01
GO:0050877	nervous system process	7	5.88E-04	9.92E-01
GO:0051122	hepoxilin biosynthetic process	2	8.17E-05	1.00E+00

Table S15. Number of genes differentially expressed in the different comparisons, for three reference genomes (canFam3.1: dog; vulVul2.2: red fox; *V. lagopus*: arctic fox).

Comparison	**canFam3.1 (19 144 genes)**	**vulVul2.2 (16 525 genes)**	***V. lagopus* (19 618 genes)**
North Africa vs Iberia	3719 (19.4%)	3644 (22%)	3756 (19.1%)
Red fox North Africa vs Iberia	1947 (10%)	3524 (21.3%)	3654 (18.6%)
Rueppell's vs red fox Iberia	3578 (18.7%)	3524 (21.3%)	3654 (18.6%)
Rueppell's vs red fox Africa	524 (2.7%)	524 (3.2%)	524 (2.7%)
Rueppell's vs red fox	1474 (7.7%)	1522 (9.2%)	1502 (7.7%)

SELECTION IN EURASIAN RED FOXES

Table S16. Genomic regions above the 99.95th percentile of the PBS empirical distribution for Eurasian red fox as focal group, using North African red foxes as closely related ingroup and Rueppell's fox as outgroup (removing the CRADD and the 25Mb region). For overlapping windows the window with highest PBS value is represented. The arctic fox was used as reference. The same outliers are obtained when using the dog as reference (not shown).

Dog	Arctic	start	end	pbs	overlapping genes
Chr1	scaffold_28	21600000	21650000	1.034	ZNF516
Chr1	scaffold_99	4710000	4760000	0.803	SLC6A16
Chr6	scaffold_86	1150000	1220000	0.782	FIS1,IFT22,CLDN15,PLOD3,ZNHIT1
Chr8	scaffold_75	3230000	3300000	0.747	COCH
Chr13	scaffold_47	10400000	10450000	0.711	.
Chr34	scaffold_85	7240000	7290000	0.695	SEMA5A
Chr33	scaffold_25	13290000	13340000	0.682	CCDC191,GRAMD1C, ZDHHC23
Chr6	scaffold_115	3530000	3580000	0.653	.
Chr11	scaffold_100	4700000	4750000	0.645	SEMA6A
Chr27	scaffold_80	7210000	7260000	0.642	B4GALNT3
Chr32	scaffold_46	6380000	6430000	0.634	FAM13A
Chr21	scaffold_22	20560000	20610000	0.632	NARS2
Chr4	scaffold_23	24340000	24390000	0.626	RHOBTB1
Chr8	scaffold_75	9270000	9320000	0.625	CHMP4A,GMPR2,IPO4,RPL37,MDP1,NEDD8,REC8,TM9SF1, TSSK4
Chr10	scaffold_4	37530000	37580000	0.601	POLR3B
Chr6	scaffold_178	930000	980000	0.600	PKD1,TSC2
Chr32	scaffold_46	8150000	8200000	0.594	PTPN13, SLC10A6
Chr1	scaffold_28	910000	960000	0.593	.
Chr4	scaffold_23	24320000	24370000	0.588	.
Chr32	scaffold_46	4600000	4650000	0.586	CCSER1
Chr33	scaffold_25	8600000	8650000	0.586	B4GALT4,UPK1B
Chr18	scaffold_61	1820000	1870000	0.582	NRXN2
Chr26	scaffold_96	2850000	2900000	0.575	PRKG1
Chr1	scaffold_28	21940000	21990000	0.574	.
Chr6	scaffold_86	6950000	7000000	0.568	AUTS2

Table S17. Genomic regions above the 99.95th percentile of the negative logarithm of alpha for Eurasian red fox as focal group (removing the CRADD region and large chromosomic introgressed region). For overlapping windows the window with highest alpha value is represented. The arctic fox was used as reference. The same outliers are obtained when using the dog as reference (not shown).

Dog chr	Arctic	start	end	-log10(alpha)	overlapping genes
Chr24	scaffold_44	14890001	14940001	0.830	DTD1
Chr34	scaffold_85	7240001	7290001	0.819	SEMA5A
Chr26	scaffold_187	460001	510001	0.815	PNRC1,RPL7L1,IGLV1-40
Chr1	scaffold_28	13970001	14020001	0.814	.
Chr16	scaffold_135	680001	730001	0.808	CSMD1
Chr3	scaffold_15	6770001	6820001	0.791	.
Chr30	scaffold_5	4360001	4410001	0.777	ARIH1
Chr3	scaffold_15	5000001	5050001	0.771	.
Chr8	scaffold_75	3220001	3270001	0.762	COCH,STRN3
Chr9	scaffold_26	25130001	25180001	0.761	SGSM2,MNT,TSR1
Chr3	scaffold_15	4380001	4430001	0.759	.
Chr33	scaffold_25	14630001	14680001	0.757	CD200
Chr27	scaffold_80	1990001	2040001	0.757	LOC111092727
Chr24	scaffold_44	14600001	14650001	0.755	DZANK1
Chr27	scaffold_80	1970001	2020001	0.752	SLC2A3
Chr11	scaffold_52	10760001	10810001	0.752	FSD1
Chr3	scaffold_15	4570001	4620001	0.751	.
Chr6	scaffold_128	1640001	1690001	0.751	HS3ST2
Chr3	scaffold_15	3880001	3930001	0.744	FBXL17
Chr15	scaffold_3	13220001	13270001	0.742	MND1
Chr3	scaffold_15	4580001	4630001	0.739	EFNA5
Chr15	scaffold_3	26500001	26550001	0.739	APAF1
Chr3	scaffold_15	5010001	5060001	0.739	.
Chr20	scaffold_1	52320001	52370001	0.737	CAND2,RPL32
Chr33	scaffold_25	14590001	14640001	0.737	LOC487974
Chr24	scaffold_44	14680001	14730001	0.736	RBBP9,SEC23B
Chr24	scaffold_44	14410001	14460001	0.734	KAT14,PET117
Chr24	scaffold_44	15610001	15660001	0.734	SLC24A3
Chr3	scaffold_15	5520001	5570001	0.732	.
Chr3	scaffold_15	5510001	5560001	0.729	.
Chr24	scaffold_44	15110001	15160001	0.728	.
Chr8	scaffold_24	8900001	8950001	0.725	DNAAF2,MGAT2,POLE2,RPL36AL
Chr33	scaffold_25	14470001	14520001	0.725	ATG3,SLC35A5

Table S18. Gene ontology enrichment based on the regions in the top 99.95 percentile of the empirical distribution of the negative logarithm of alpha, with Eurasian red fox as the focal group.

GO term	Description of bilogical process	Gene count	P-value	FDR p-value
Gene Ontology using a ranked list in http://geneontology.org/				
GO:0008643	carbohydrate transport	3	9.74E-05	1.00E+00
GO:1904659	glucose transmembrane transport	2	8.85E-04	1.00E+00
GO:0008645	hexose transmembrane transport	2	9.46E-04	1.00E+00
GO:0015749	monosaccharide transmembrane transport	2	1.01E-03	1.00E+00
GO:0034219	carbohydrate transmembrane transport	2	1.14E-03	1.00E+00
GO:0000086	G2/M transition of mitotic cell cycle	2	1.35E-03	1.00E+00
GO:0044839	cell cycle G2/M phase transition	2	1.50E-03	1.00E+00
GO:0050919	negative chemotaxis	2	1.90E-03	1.00E+00
GO:0098656	anion transmembrane transport	4	2.61E-03	1.00E+00
GO:1990708	conditioned place preference	1	2.94E-03	1.00E+00
GO:1904465	negative regulation of matrix metallopeptidase secretion	1	2.94E-03	1.00E+00
GO:0150072	positive regulation of arginase activity	1	2.94E-03	1.00E+00
GO:0150070	regulation of arginase activity	1	2.94E-03	1.00E+00
GO:0048588	developmental cell growth	2	3.51E-03	1.00E+00
GO:0016049	cell growth	2	3.74E-03	1.00E+00
GO:0010265	SCF complex assembly	1	4.40E-03	1.00E+00
GO:0150074	positive regulation of protein-glutamine gamma-glutamyltransferase activity	1	4.40E-03	1.00E+00
GO:0150073	regulation of protein-glutamine gamma-glutamyltransferase activity	1	4.40E-03	1.00E+00
GO:0045089	positive regulation of innate immune response	2	4.61E-03	1.00E+00
GO:0061387	regulation of extent of cell growth	2	5.42E-03	1.00E+00

There are no GO terms with significant enrichment empirical p-values for GOrilla enrichment analysis.

SELECTION IN CORE DESERT-DWELLING SPECIES: RUEPPELL'S AND FENNEC FOXES

Table S19. Genomic regions above the 99.95th percentile of the delta δ log-likelihood ratio (LLR) empirical distribution for Rueppell's fox (for K=4). The arctic fox genome was used as reference.

Dog	Arctic	start	end	LLR_{Ru}	LLR_{Red}	Max	δ LLR	genes
Chr24	scaffold_134	1120000	1170000	589.13	341.37	589.13	247.76	NTSR1, SLCO4A1
Chr13	scaffold_155	1750000	1800000	1400.24	1167.83	1400.24	232.41	WFS1
Chr34	scaffold_19	29650000	29700000	594.69	370.47	594.69	224.21	.
Chr24	scaffold_134	1360000	1410000	542.54	318.49	542.54	224.05	LAMA5
Chr34	scaffold_19	17740000	17790000	600.31	379.92	600.31	220.39	.
Chr33	scaffold_25	21700000	21750000	695.64	477.30	695.64	218.33	.
Chr24	scaffold_134	640000	690000	568.78	352.57	568.78	216.20	CHRNA4, COL20A1, KCNQ2
Chr15	scaffold_3	31580000	31630000	473.86	262.51	473.86	211.35	LOC111090035
Chr7	scaffold_7	39960000	40010000	588.52	377.69	588.52	210.83	LGR6, PTPRK
Chr24	scaffold_134	710000	760000	668.20	459.60	668.20	208.59	ARFGAP1, BIRC7,NKAIN4
Chr24	scaffold_134	1200000	1250000	545.66	340.74	545.66	204.92	MIR1-1, MIR133C
Chr25	scaffold_175	200000	250000	521.86	317.45	521.86	204.42	ANKMY1, GPC1, MIR149
Chr4	scaffold_23	2750000	2810000	522.06	323.25	522.06	198.80	LDB3,OPN4
Chr26	scaffold_187	500000	550000	458.00	259.77	458.00	198.23	
Chr31	scaffold_110	4730000	4780000	696.16	498.14	696.16	198.02	FAM207A
Chr13	scaffold_148	2150000	2200000	506.31	308.90	506.31	197.42	LOC102155929, TSNARE1
Chr3	scaffold_45	6450000	6500000	487.24	294.23	487.24	193.01	EVC
Chr13	scaffold_47	11260000	11310000	1166.35	973.59	1166.35	192.76	LOC111098476
Chr8	scaffold_51	14620000	14670000	596.24	406.02	596.24	190.22	.
Chr13	scaffold_47	15930000	15980000	1052.80	863.51	1052.80	189.29	.
Chr20	scaffold_1	26240000	26290000	342.75	155.42	342.75	187.33	C20H3orf67
Chr10	scaffold_83	1990000	2040000	694.03	510.78	694.03	183.26	FAM19A5
Chr32	scaffold_46	9500000	9550000	619.50	436.65	619.50	182.84	ARHGAP24
Chr4	scaffold_23	15440000	15490000	704.72	522.68	704.72	182.04	.
Chr24	scaffold_134	1790000	1840000	400.14	218.22	400.14	181.92	CDH4
Chr13	scaffold_126	2940000	2990000	766.11	584.64	766.11	181.47	LOC111095751

								MOB1B
Chr13	scaffold_109	3120000	3170000	448.67	268.57	448.67	180.10	LOC102156107
Chr10	scaffold_83	2820000	2870000	774.33	594.48	774.33	179.85	LOC102154327
Chr11	scaffold_29	14490000	14540000	614.15	435.04	614.15	179.11	KDM4C
Chr13	scaffold_47	18370000	18420000	432.89	253.87	432.89	179.02	
Chr25	scaffold_175	500000	550000	535.41	357.03	535.41	178.38	CROCC2
Chr34	scaffold_85	1530000	1580000	654.30	476.55	654.30	177.76	.
Chr1	scaffold_28	1540000	1590000	500.54	323.26	500.54	177.28	.
Chr38	scaffold_136	1670000	1720000	476.24	298.98	476.24	177.26	FCRL6, SLAMF8
Chr4	scaffold_23	3160000	3210000	441.01	264.03	441.01	176.98	GRID1
Chr15	scaffold_3	34710000	34760000	447.45	270.54	447.45	176.91	LOC102153513
Chr21	scaffold_168	1260000	1310000	304.38	127.48	304.38	176.90	
Chr38	scaffold_10	24400000	24450000	460.00	284.59	460.00	175.41	LRRC4C

Table S20. Genomic regions above the 99.95th percentile of the log-likelihood ratio (LLR) empirical distribution for fennec fox (for K=4). The arctic fox genome was used as reference.

Dog	Arctic	start	end	LLR	genes
Chr38	scaffold_136	330000	380000	1032.50	KIRREL1
Chr18	scaffold_61	13630000	13680000	981.18	COR4C30, OR4C28
Chr38	scaffold_136	210000	260000	970.35	DISP1,SUSD4,TLR5
Chr10	scaffold_83	1890000	1940000	965.37	.
Chr18	scaffold_61	13620000	13670000	929.85	LOC483582
Chr35	scaffold_31	23850000	23900000	912.12	DUSP22,IRF4
Chr15	scaffold_3	46770000	46820000	883.12	EDDM3B, RNASE4
Chr6	scaffold_181	360000	410000	882.49	CACNA1H
Chr4	scaffold_23	2370000	2420000	872.37	GPRIN2,LOC489080
Chr8	scaffold_51	16310000	16360000	862.00	ADARB2
Chr10	scaffold_83	2000000	2050000	855.46	FAM19A5
Chr17	scaffold_114	600000	650000	851.98	SNTG2
Chr9	scaffold_102	5310000	5360000	843.98	BAIAP2
Chr6	scaffold_181	270000	320000	814.04	SSTR5
Chr26	scaffold_269	0	50000	810.62	LOC111092524
Chr3	scaffold_89	190000	240000	787.60	CPLX1,GAK,
Chr6	scaffold_181	240000	290000	779.29	SOX8
Chr9	scaffold_102	6700000	6750000	777.11	B3GNTL1,TBCD
Chr8	scaffold_63	700000	750000	773.59	PACS2
Chr16	scaffold_188	580000	630000	767.90	LOC491446,MYOM2
Chr25	scaffold_175	510000	560000	759.15	CROCC2, SNED1
Chr8	scaffold_51	16780000	16830000	752.56	DIP2C
Chr36	scaffold_18	1990000	2040000	752.26	ITGAV

Table S21. Top-20 regions ranked by Homogeneity Test for reduction of polymorphic sites (low polymorphism-to-divergence ratio) in fennec foxes relative to red foxes and the dog. The dog genome was used as reference.

Chr Dog	ID	start	end	Homogeneity Test Score	Overlapping genes
38	NC_006620.3	23520000	23570000	143.7	KIRREL1
4	NC_006586.3	34090000	34140000	130.3	GRID1
38	NC_006620.3	23700000	23750000	127.8	DISP1,TLR5
2	NC_006584.3	32250000	32300000	124.4	PFKP
10	NC_006592.3	18130000	18180000	121.8	FAM19A5
2	NC_006584.3	56890000	56940000	115.6	.
3	NC_006585.3	76130000	76180000	114.9	.
30	NC_006612.3	34000000	34050000	113.7	.
36	NC_006618.3	28740000	28790000	106.7	ZC3H15
36	NC_006618.3	29750000	29800000	105.6	.
3	NC_006585.3	60510000	60560000	105.3	.
38	NC_006620.3	22170000	22220000	105	IGSF9,LOC106558344,SLAMF9
26	NC_006608.3	32170000	32220000	104.7	.
17	NC_006599.3	1910000	1960000	104.3	EIPR1
39	NC_006617.3	3600000	3650000	103.6	SLC22A23
16	NC_006598.3	15610000	15660000	103.4	PRKAG2
27	NC_006609.3	35490000	35540000	102.3	LOC111092750
4	NC_006586.3	77920000	77970000	101.2	.
9	NC_006591.3	50800000	50850000	101.2	COL5A1
13	NC_006595.3	28170000	28220000	101	.

PHYSIOLOGICAL ADAPTATIONS IN RUEPPELL'S AND FENNEC FOXES

Table S22. Summary statistics describing physiological parameters collected from blood of three *Vulpes sp*. (red fox, *V. vulpes*; Rueppell's fox, *V. rueppellii*; fennec fox, *V. zerda*).

Plasma contents	Parameter	Species	Mean	s.d.	min-max
Hormones	Copeptin (pmol/L)	*V. vulpes*	4.41	1.32	2.02-6.18
		V. rueppellii	16.86	9.33	4.68 -34
		V. zerda	13.02	5.73	6.64- 22.9
	Aldosterone(pg/ml)	*V. vulpes*	241.8	206.14	38-646
		V. rueppellii	55.62	49.31	8-184
		V. zerda	40.2	24.11	82-66
	Total T4 (ug/dL)	*V. vulpes*	4.73	1.38	2.83- 6.52
		V. rueppellii	2.26	1.33	0.59-4.53
		V. zerda	1.24	0.7	0.74-2.85
Metabolites	Urea (mg/dL)	*V. vulpes*	71.35	35.7	24-130.1
		V. rueppellii	51.33	24.39	22.6-106.1
		V. zerda	43.46	28.2	24-120.3
	Uric acid (mg/dL)	*V. vulpes*	0.7	0.39	0.3-1.4
		V. rueppellii	0.81	0.54	0.2-2.3
		V. zerda	0.51	0.25	0.3-0.8
	Creatinine (mg/dL)	*V. vulpes*	0.32	0.18	0.1-0.66
		V. rueppellii	0.34	0.37	0.1-1.21
		V. zerda	0.37	0.46	0.1-1.56
Protein	Albumin (g/dL)	*V. vulpes*	1.49	0.61	0.68-2.44
		V. rueppellii	1.74	0.47	0.55-2.34
		V. zerda	1.62	0.72	0.82-2.95
Fat	Cholesterol (mg/dL)	*V. vulpes*	76.32	39.66	23.6-138.7
		V. rueppellii	94.73	37	4.6-135.9
		V. zerda	65.12	47.41	15.2-168.9
Electrolytes	Sodium (mEq/L)	*V. vulpes*	152.17	7.05	143-163
		V. rueppellii	154.83	4.07	148-159
	Potassium (mEq/L)	*V. vulpes*	4.6	0.41	3.9-5
		V. rueppellii	5.02	0.4	4.5-5.6
	Chloride (mEq/L)	*V. vulpes*	112.5	3.45	107-116
		V. rueppellii	112.33	3.98	107-119
Osmolality	Osmolality (mOsm/Kg)	*V. vulpes*	314.27	9.57	299.74-325.96
		V. rueppellii	312.59	6.84	303.11-320.16

Table S23. ANOVA table generated for a linear model where Y is a given physiological parameter using linear models that included species, sex, weight, body mass index and status as covariates. The status of the individuals means weather they were caught in the wild or had been previously held captive. Some of the variables were log-transformed to meet the assumptions of normality and homogeneity in linear models.

	DF	Sum of Squares	Mean Sum of squares	F value	Pr(>F)
Logarithm of Copeptin					
Species	2	9.551	4.775	18.779	8.99e-06
Weight	1	0.333	0.333	1.309	0.263
BMI	1	0.000	0.000	0.000	0.985
Sex	1	0.033	0.033	0.131	0.720
Status	1	0.059	0.059	0.231	0.635
Residuals	26	6.612	0.254		
Logarithm of Aldosterone					
Species	2	16.351	8.175	10.795	0.000386
Weight	1	0.947	0.947	1.251	0.273633
BMI	1	0.130	0.130	0.172	0.681549
Sex	1	0.090	0.090	0.119	0.733120
Status	1	0.865	0.865	1.143	0.294921
Residuals	26	19.692	0.757		
Thyroxine(T4)					
Species	2	65.01	32.50	24.815	9.37e-07
Weight	1	0.26	0.26	0.201	0.6579
BMI	1	4.95	4.95	3.778	0.0628
Sex	1	0.00	0.00	0.000	0.9854
Status	1	3.70	3.70	2.823	0.1049
Residuals	26	34.06	1.31		
Logarithm of Copeptin using T4 and Aldosterone as two additional covariates					
Species	2	9.551	4.775	17.719	1.88e-05
T4 (ug/dL)	1	0.050	0.050	0.185	0.671
Aldosterone (pg/mL)	1	0.028	0.028	0.104	0.749
Weight	1	0.328	0.328	1.216	0.281
BMI	1	0.007	0.007	0.027	0.870
Sex	1	0.036	0.036	0.135	0.717
Status	1	0.119	0.119	0.442	0.513
Residuals	24	6.468	0.270		
Logarithm of Aldosterone using T4 and Copeptin as two additional covariates					
Species	2	16.351	8.175	11.374	0.000335
T4 (ug/dL)	1	0.772	0.772	1.075	0.310217
Copeptin (pmol/L)	1	0.398	0.398	0.554	0.463945
Weight	1	0.599	0.599	0.833	0.370484
BMI	1	0.506	0.506	0.704	0.409563
Sex	1	0.129	0.129	0.179	0.675713
Status	1	2.070	2.070	2.880	0.102619
Residuals	24	17.250	0.719		
T4 using Aldosterone and Copeptin as two additional covariates					
Species	2	65.01	32.50	26.045	9.7e-07
Aldosterone(pg/mL)	1	1.08	1.08	0.862	0.3623
Copeptin (pmol/L)	1	0.96	0.96	0.767	0.3899
Weight	1	0.18	0.18	0.147	0.7047
BMI	1	5.11	5.11	4.098	0.0542

Sex	1	0.14	0.14	0.111	0.7424
Status	1	5.54	5.54	4.442	0.0457
Residuals	24	29.95	1.25		

Supplementary References

Ali A., Baby B., Vijayan R. 2019. From desert to medicine: A review of camel genomics and therapeutic products. Front. Genet. 10:1–20.

Ali M.A., Adem A., Chandranath I.S., Benedict S., Pathan J.Y., Nagelkerke N., Nyberg F., Lewis L.K., Yandle T.G., Nicholls G.M., Frampton C.M., Kazzam E. 2012. Responses to Dehydration in the One-Humped Camel and Effects of Blocking the Renin-Angiotensin System. PLoS One. 7:e37299.

Banta M.R., Holcombe D.W. 2002. The effects of thyroxine on metabolism and water balance in a desert-dwelling rodent, Merriam's kangaroo rat (Dipodomys merriami). J. Comp. Physiol. B Biochem. Syst. Environ. Physiol. 172:17–25.

Bartolini Lucenti S., Madurell-Malapeira J. 2020. Unraveling the fossil record of foxes: An updated review on the Plio-Pleistocene Vulpes spp. from Europe. Quat. Sci. Rev. 236.

Bolger A.M., Lohse M., Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30:2114–2120.

Bridges T.E., James N. V. 1982. 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.

Careau V., Morand-Ferron J., Thomas D. 2007. Basal Metabolic Rate of Canidae from Hot Deserts to Cold Arctic Climates. J. Mammal. 88:394–400.

Chang C.C., Chow C.C., Tellier L.C.A.M., Vattikuti S., Purcell S.M., Lee J.J. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 4:7.

Chaouacha-Chekir R. Ben, Leloup J., Lachiver F. 1997. Influence of thyroid status on water metabolism and survival of normal and dehydrated desert rodents Meriones libycus. Gen. Comp. Endocrinol. 105:1–8.

Cheng J.Y., Racimo F., Nielsen R. 2019. Ohana: Detecting selection in multiple populations by modelling ancestral admixture components. bioRxiv.:546408.

Crawford J.E., Nielsen R. 2013. Detecting adaptive trait loci in nonmodel systems: divergence or admixture mapping? Mol. Ecol. 22:6131–6148.

Danecek P., Auton A., Abecasis G., Albers C.A., Banks E., DePristo M.A., Handsaker R.E., Lunter G.,

Marth G.T., Sherry S.T., McVean G., Durbin R., 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. Bioinformatics. 27:2156–2158.

Dickinson H., Moritz K., Wintour E.M., Walker D.W., Kett M.M. 2007. A comparative study of renal function in the desert-adapted spiny mouse and the laboratory-adapted C57BL/6 mouse: Response to dietary salt load. Am. J. Physiol. - Ren. Physiol. 293.

Dobin A., Davis C.A., Schlesinger F., Drenkow J., Zaleski C., Jha S., Batut P., Chaisson M., Gingeras T.R. 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 29:15–21.

Donald J., Pannabecker T.L. 2015. Osmoregulation in Desert-Adapted Mammals. Sodium and Water Homeostasis. Springer New York. p. 191–211.

Drummond A.J., Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7:214.

Ewels P., Magnusson M., Lundin S., Käller M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 32:3047–3048.

Gouy M., Guindon S., Gascuel O. 2010. SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. Mol. Biol. Evol. 27:221–224.

Gutenkunst R.N., Hernandez R.D., Williamson S.H., Bustamante C.D. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet. 5:1000695.

Hou Z., Jiang P., Swanson S.A., Elwell A.L., Nguyen B.K.S., Bolin J.M., Stewart R., Thomson J.A. 2015. A cost-effective RNA sequencing protocol for large-scale gene expression studies. Sci. Rep. 5:1–5.

Hudson R.R., Kreitman M., Aguadé M. 1987. A Test of Neutral Molecular Evolution Based on Nucleotide Data. Genetics. 116.

Hulbert A.J. 2000. Thyroid hormones and their effects: A new perspective. Biol. Rev. 75:519–631.

Hulbert A.J., Hinds D.S., MacMillen R.E. 1985a. Minimal metabolism, summit metabolism and plasma thyroxine in rodents from different environments. Comp. Biochem. Physiol. -- Part A Physiol. 81:687–693.

Hulbert A.J., Hinds D.S., MacMillen R.E. 1985b. Minimal metabolism, summit metabolism and plasma thyroxine in rodents from different environments. Comp. Biochem. Physiol. -- Part A Physiol. 81:687–693.

Ishikawa A., Kitano J. 2012. Ecological Genetics of Thyroid Hormone Physiology in Humans and Wild Animals. Thyroid Horm.

Jónsson H., Ginolhac A., Schubert M., Johnson P.L.F., Orlando L. 2013. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics. 29:1682–

1684.

Kamau J.M.Z., Maloiy G.M.O. 1983. Changes in thyroxine, triiodothyronine and metabolic rate in the dik-dik antelope (Rhynchotragus kirkii): effects of temperature and dehydration. J. Comp. Physiol. B. 152.

Katoh K., Standley D.M. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol. Biol. Evol. 30:772–780.

Kent W.J., Sugnet C.W., Furey T.S., Roskin K.M., Pringle T.H., Zahler A.M., Haussler a. D. 2002. The Human Genome Browser at UCSC. Genome Res. 12:996–1006.

Knepper M.A., Kwon T.H., Nielsen S., Ingelfinger J.R. 2015. Molecular physiology of water balance. N. Engl. J. Med. 372:1349–1358.

Koch E.M., Schweizer R.M., Schweizer T.M., Stahler D.R., Smith D.W., Wayne R.K., Novembre J. 2019. De Novo Mutation Rate Estimation in Wolves of Known Pedigree. Mol. Biol. Evol. 36:2536–2547.

Korneliussen T.S., Albrechtsen A., Nielsen R. 2014. ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics. 15:356.

Korneliussen T.S., Moltke I. 2015. NgsRelate: A software tool for estimating pairwise relatedness from next-generation sequencing data. Bioinformatics. 31:4009–4011.

Kumar S., Stecher G., Li M., Knyaz C., Tamura K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol. Biol. Evol. 35:1547–1549.

Kumar V., Kutschera V.E., Nilsson M.A., Janke A. 2015. Genetic signatures of adaptation revealed from transcriptome sequencing of Arctic and red foxes. BMC Genomics. 16.

Kurtz S., Phillippy A., Delcher A.L., Smoot M., Shumway M., Antonescu C., Salzberg S.L. 2004. Versatile and open software for comparing large genomes. Genome Biol. 5:R12.

Kutschera V.E., Lecomte N., Janke A., Selva N., Sokolov A.A., Haun T., Steyer K., Nowak C., Hailer F. 2013. A range-wide synthesis and timeline for phylogeographic events in the red fox (Vulpes vulpes). BMC Evol. Biol. 13.

Lefort V., Desper R., Gascuel O. 2015. FastME 2.0: A Comprehensive, Accurate, and Fast Distance-Based Phylogeny Inference Program: Table 1. Mol. Biol. Evol. 32:2798–2800.

Leite J.V., Álvares F., Velo-Antón G., Brito J.C., Godinho R. 2015. 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.

Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 27:2987–2993.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. .

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 25:2078–2079.

Liu S., Lorenzen E.D., Fumagalli M., Li B., Harris K., Xiong Z., Zhou L., Korneliussen T.S., Somel M., Babbitt C., Wray G., Li J., He W., Wang Z., Fu W., Xiang X., Morgan C.C., Doherty A., O'Connell M.J., McInerney J.O., Born E.W., Dalén L., Dietz R., Orlando L., Sonne C., Zhang G., Nielsen R., Willerslev E., Wang J. 2014. Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. Cell. 157:785–794.

Lonsinger R.C., Adams J.R., Waits L.P. 2018. Evaluating effective population size and genetic diversity of a declining kit fox population using contemporary and historical specimens. Ecol. Evol. 8:12011–12021.

Love M.I., Huber W., Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15:550.

Malaspinas A.S., Westaway M.C., Muller C., Sousa V.C., Lao O., Alves I., Bergström A., Athanasiadis G., Cheng J.Y., Crawford J.E., Heupink T.H., MacHoldt E., Peischl S., Rasmussen S., Schiffels S., Subramanian S., Wright J.L., Albrechtsen A., Barbieri C., Dupanloup I., Eriksson A., Margaryan A., Moltke I., Pugach I., Korneliussen T.S., Levkivskyi I.P., Moreno-Mayar J.V., Ni S., Racimo F., Sikora M., Xue Y., Aghakhanian F.A., Brucato N., Brunak S., Campos P.F., Clark W., Ellingvåg S., Fourmile G., Gerbault P., Injie D., Koki G., Leavesley M., Logan B., Lynch A., Matisoo-Smith E.A., McAllister P.J., Mentzer A.J., Metspalu M., Migliano A.B., Murgha L., Phipps M.E., Pomat W., Reynolds D., Ricaut F.X., Siba P., Thomas M.G., Wales T., Wall C.M.R., Oppenheimer S.J., Tyler-Smith C., Durbin R., Dortch J., Manica A., Schierup M.H., Foley R.A., Lahr M.M., Bowern C., Wall J.D., Mailund T., Stoneking M., Nielsen R., Sandhu M.S., Excoffier L., Lambert D.M., Willerslev E. 2016. A genomic history of Aboriginal Australia. Nature. 538:207–214.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 17:10.

Martin S.H., Davey J.W., Jiggins C.D. 2015. Evaluating the Use of ABBA–BABA Statistics to Locate Introgressed Loci. Mol. Biol. Evol. 32:244–257.

Matzke N.J. 2013. Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. Front. Biogeogr. 5.

Matzke N.J. 2014. Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. Syst. Biol. 63:951–970.

Morgenthaler N.G. 2010. Copeptin: A Biomarker of Cardiovascular and Renal Function. Congest. Hear. Fail. 16:S37–S44.

Morgenthaler N.G., Struck J., Jochberger S., Dünser M.W. 2008. Copeptin: clinical use of a new

biomarker. Trends Endocrinol. Metab. 19:43–49.

Noskova E., Ulyantsev V., Koepfli K.P., O'brien S.J., Dobrynin P. 2020. GADMA: Genetic algorithm for inferring demographic history of multiple populations from allele frequency spectrum data. Gigascience. 9:1–18.

Okonechnikov K., Conesa A., García-Alcalde F. 2015. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics. 32:292–294.

Perini F.A., Russo C.A.M., Schrago C.G. 2010. The evolution of South American endemic canids: A history of rapid diversification and morphological parallelism. J. Evol. Biol. 23:311–322.

Pickrell J.K., Pritchard J.K. 2012. Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. PLoS Genet. 8:1002967.

Posada D. 2008. jModelTest: Phylogenetic Model Averaging. Mol. Biol. Evol. 25:1253–1256.

Rocha J.L., Brito J.C., Nielsen R., Godinho R. 2021. Convergent evolution of increased urine-concentrating ability in desert mammals. Mamm. Rev.:mam.12244.

Sacks B.N., Lounsberry Z.T., Statham M.J. 2018. Nuclear Genetic Analysis of the Red Fox Across its Trans-Pacific Range. J. Hered. 109:573–584.

Sands J.M., Layton H.E. 2014. Advances in Understanding the Urine-Concentrating Mechanism. Annu. Rev. Physiol. 76:387–409.

Schwimmer H., Haim A. 2009. Physiological adaptations of small mammals to desert ecosystems. Integr. Zool. 4:357–366.

Scott, Irene; Yousef, Mohamed K., Johnson H.D. 1976. Plasma thyroxine levels of mammals:desert and mountain. Life Sci. 19:807–812.

Shumate A., Salzberg S.L. 2020. Liftoff: accurate mapping of gene annotations. Bioinformatics.:btaa1016.

Sinding M.H.S., Gopalakrishnan S., Ramos-Madrigal J., Manuel M. De, Pitulko V. V., Kuderna L., Feuerborn T.R., Frantz L.A.F., Vieira F.G., Niemann J., Castruita J.A.S., Carøe C., Andersen-Ranberg E.U., Jordan P.D., Pavlova E.Y., Nikolskiy P.A., Kasparov A.K., Ivanova V. V., Willerslev E., Skoglund P., Fredholm M., Wennerberg S.E., Heide-Jørgensen M.P., Dietz R., Sonne C., Meldgaard M., Dalén L., Larson G., Petersen B., Sicheritz-Pontén T., Bachmann L., Wiig Ø., Marques-Bonet T., Hansen A.J., Gilbert M.T.P. 2020. Arctic-adapted dogs emerged at the Pleistocene-Holocene transition. Science (80-.). 368:1495–1499.

Soraggi S., Wiuf C., Albrechtsen A. 2018. Powerful inference with the D-statistic on low-coverage whole-genome data. G3 Genes, Genomes, Genet. 8:551–566.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30:1312–1313.

Statham M.J., Murdoch J., Janecka J., Aubry K.B., Edwards C.J., Soulsbury C.D., Berry O., Wang Z., Harrison D., Pearch M., Tomsett L., Chupasko J., Sacks B.N. 2014. Range-wide multilocus phylogeography of the red fox reveals ancient continental divergence, minimal genomic exchange and distinct demographic histories. Mol. Ecol. 23:4813–4830.

Stockand J.D. 2012. The role of the epithelial Na + channel (enac) in high AVP but low aldosterone states. Front. Physiol. 3 JUL:1–5.

Vieira F.G., Lassalle F., Korneliussen T.S., Fumagalli M. 2016. Improving the estimation of genetic distances from Next-Generation Sequencing data. Biol. J. Linn. Soc. 117:139–149.

Williams J.B., Lenain D., Ostrowski S., Tieleman B.I., Seddon P.J. 2002. Energy expenditure and water flux of Rüppell's foxes in Saudi Arabia. Physiol. Biochem. Zool. 75:479–488.

Williams J.B., Muñoz-Garcia A., Ostrowski S., Tieleman B.I. 2004. A phylogenetic analysis of basal metabolism, total evaporative water loss, and life-history among foxes from desert and mesic regions. J. Comp. Physiol. B Biochem. Syst. Environ. Physiol. 174:29–39.

Xu M.M., Wang D.H. 2016. Water deprivation up-regulates urine osmolality and renal aquaporin 2 in Mongolian gerbils (Meriones unguiculatus). Comp. Biochem. Physiol. -Part A Mol. Integr. Physiol. 194:37–44.

Yi X., Liang Y., Huerta-Sanchez E., Jin X., Xi Ping Cuo Z., Pool J.E., Xu X., Jiang H., Vinckenbosch N., Korneliussen T.S., Zheng H., Liu T., He W., Li K., Luo R., Nie X., Wu H., Zhao M., Cao H., Zou J., Shan Y., Li S., Yang Q., Asan, Ni P., Tian G., Xu J., Liu X., Jiang T., Wu R., Zhou G., Tang M., Qin J., Wang T., Feng S., Li G., Huasang, Luosang J., Wang W., Chen F., Wang Y., Zheng X., Li Z., Bianba Z., Yang G., Wang X., Tang S., Gao G., Chen Y., Luo Z., Gusang L., Cao Z., Zhang Q., Ouyang W., Ren X., Liang H., Zheng H., Huang Y., Li J., Bolund L., Kristiansen K., Li Y., Zhang Y., Zhang X., Li R., Li S., Yang H., Nielsen R., Wang J., Wang J. 2010. Sequencing of Fifty Human Exomes Reveals Adaptations to High Altitude. Science (80-.). 329:75–78.

Zhai W., Nielsen R., Slatkin M. 2009. An Investigation of the Statistical Power of Neutrality Tests Based on Comparative and Population Genetic Data. Mol. Biol. Evol. 26:273–283.

Zhao C., Zhang H., Liu G., Yang X., Zhang J. 2016. The complete mitochondrial genome of the Tibetan fox (Vulpes ferrilata) and implications for the phylogeny of Canidae. Comptes Rendus - Biol. 339:68–77.

Chapter IV.

General Discussion

This dissertation presents a window into the genetic basis of desert adaptation. By synthesizing methods and current genomic research on different mammalian species we were able to identify shared patterns of adaptation at the genetic, pathway and/or phenotypic level in biological systems with different evolutionary ages, occupying distinct ecological niches and at distinct stages of adaptive evolution in different deserts of the world (Chapter II, Paper 1). By integrating eco-physiological data obtained from the literature with bioclimatic variables and phylogenetic comparative methods we also demonstrated that, among the many physical and climatic challenges faced by desert species, aridity has been one of the main selective pressures driving phenotypic evolution in different desert mammalian lineages (Chapter II, Paper 2). Tough these meta-analyses offer important insights into the most relevant genes and primary physiological traits involved in mammalian adaptation to arid regions, as well as on key environmental factors driving phenotypic evolution, the evolutionary mechanisms at play during adaptive evolution remain poorly understood. We could not draw major patterns on the nature of desert adaptive genetic variation, nor on the sources by which adaptive genetic variants may arise (Chapter II, Paper 1). Furthermore, the majority of selection studies struggled to identify links between genes, phenotypes and desert environments, either through direct association testing or, indirectly, by looking for changes in gene expression and/or physiology. Our comprehensive scan of the existing literature ultimately raises more questions than provides answers, owing to lack of segregating phenotypic variation, lack of biological sampling infrastructures and expense of researching in remote desert regions.

In Chapter III, we defied logistical challenges and set off to the Sahara Desert to study the evolutionary mechanisms driving adaptation in North-African foxes. By combining state-of-the-art genomics with eco-physiological and expression data, we were able to address outstanding questions concerning adaptive evolution in arid regions, and on the evolutionary history of these iconic models. Our study reveals a complex history of past interspecific gene flow among geographically overlapping sister species and highly divergent clades that lends additional evidence on the pervasive nature of introgression in canids. In addition, this work underscores the important roles of introgressive-hybridization events and ancestral standing genetic variation as sources for adaptive genetic variation at different spatiotemporal scales. Finally, we provide important insights on the genes and gene changes potentially underlying adaptation in an entirely new system, including decreased renal and

non-renal water losses and adaptive tolerance to extreme temperatures. The next few sections will be devoted to discussions on how studying North-African foxes has increased our general understanding of the evolutionary history of the genus, as well as the evolutionary processes shaping life in deserts, and how complex ancestral hybridization events may have facilitated evolutionary responses to environmental change, but also confound our understanding of evolutionary history. Lastly, I discuss outstanding challenges and perspectives.

4.1. On the origins and evolution of North-African *Vulpes*

The most extraordinary study-systems to investigate the genetic mechanisms underlying rapid and recent adaptation to novel environments, or lifestyles, are pairs of closely related and recently diverged groups of organisms that differ fundamentally in ecology, morphology, behavior or physiology (Li et al. 2014; Liu et al. 2014; Zhang et al. 2014; Sinding et al. 2020). On the other hand, the independent colonization and local adaptation to similar environments by distant lineages provides a unique opportunity to study mechanisms underlying shared adaptations and the evolution of similar solutions to the same adaptive problem (Ferreira et al. 2021). North-African desert-dwelling foxes such as the fennec fox (*V. zerda*) and Rueppell's fox (*V. rueppellii*) constitute remarkable examples of independent colonization and adaptation to the same desert environment by distant lineages of the same genus. Up until this study, little had been known about the geographical origins of some pairs of closely related species of *Vulpes*. For instance, previous studies focused on the evolutionary history of Rueppell's fox (*V. rueppellii*) and red fox (*V. vulpes*) could not convincingly discriminate if the two sister species differentiated in Eurasia, followed by independent colonization of North-Africa, or if they originated and migrated out of Africa (Leite et al. 2015).

Phylogenetic analysis and ancestral range reconstruction with mitochondrial genomes in Paper 3 show that Rueppell's fox and red fox likely differentiated in Eurasia. We further show that the two sister species are outgrouped by the Bengal fox (*V.* bengalensis; **Figure 4.1A**), a species from scrub and semi-arid grassland habitats, and avoider of dense forests and 'true deserts' (Sillero-Zubiri et al. 2004). Canid specialists refer to Rueppell's as desert specialists and red foxes' as generalists not found in extreme deserts and with a natural habitat defined as dry, mixed landscape with abundant scrub and woodland (Sillero-Zubiri et al. 2004). As such, while Rueppell's differentiated from red foxes ~576 kya and evolved to inhabit some of the most extreme hot deserts on Earth, including the heart of the Lut desert of Iran, the Rub Al-Khali of the Arabian Peninsula, and the North-African Sahara, being considered 'true desert-dwellers', red foxes radiated from Eurasia into different regions of the globe and came to occupy different ecological niches (Sillero-Zubiri et al. 2004). With this study, we show that North-African red foxes have diverged from Eurasian populations ~78 kya, and since then successfully adapted to life in increasingly arid and hyper-arid regions at the northern Saharan fringes (Brito et al. 2009), providing a prime example of recent adaptation to rapidly changing climates.

Furthermore, we show that separation from Eurasia and local adaptation in North-Africa is related to the paleoclimatic history of the Sahara, offering an unprecedented opportunity to understand how climatic oscillations over the last ~78 kya promoted and shaped adaptive evolution. North-African foxes (fennec, Rueppell's fox and red fox) can therefore be viewed as the current outcome of consecutive replications of a natural experiment in which different ancestral *Vulpes* lineages were able to independently colonize and adapt to life in hot arid and hyper-arid regions, providing an exciting opportunity to study the processes and sources underlying shared and/or independent adaptive evolution at different spatiotemporal scales.

4.2. Introgression in canids and implications to desert life

Introgressive hybridization is an important and pervasive source of genetic variation in many groups of canid species, as attested by evidence of recurrent interspecific gene flow between wolves and dogs (Miao et al. 2017; von Holdt et al. 2017; Schweizer et al. 2018; Signore et al. 2019; Wang et al. 2020), coyotes and wolves (VonHoldt et al. 2016), Eurasian golden jackals and the gray wolf/dog clade (Koepfli et al. 2015), gray and Ethiopian wolves and the African golden wolf (Gopalakrishnan et al. 2018). Phylogenomic analyses of the *Canis* diversification further revealed complex histories of ancient hybridization among the ancestors of extant species, and between highly divergent and unidentified lineages of canids (called 'ghost lineages'). For instance, coyotes and gray wolves (Gopalakrishnan et al. 2018) and Tibetan/Himalayan wolves (Wang et al. 2020) likely carry genetic material introgressed from unidentified lineages. These results have shown that introgression can extend well-beyond recent or ongoing hybridization between extant and/or closely related pairs of species and may occur as deep in a phylogeny as the origin of a genus.

Our work with different North-African foxes underscores an equally complex history of interspecific gene flow at different evolutionary scales, including significant genome-wide introgression between more closely related North-African Rueppell's and red foxes, and distantly related Rueppell's and fennec foxes (Chapter III). Subsequent attempts to explore these results led us to propose that Rueppell's fox may carry a large chromosomal region (25 Mb) likely introgressed from an as-yet-undefined fennec-like fox. Future phylogenomic analyses with all extant members of *Vulpes* are needed to appraise the extent of interspecific gene flow throughout the genus. For instance, with high-quality long-read-based genomes for all fox species we would be able to resolve the structure of the putatively introgressed 25 Mb region – which we speculate might be the result of a translocation – and test if it is phylogenetically closer to the fennec or its sister species, the Blanford's fox (*Vulpes cana*), which overlaps in range with Rueppell's fox in the Middle East (**Figure 4.1B**). We would also be able to accurately reconstruct haplotypes and explore the extent of recombination suppression at this region. Future studies should also ask if the phylogenetic placement of Bengal's fox, Tibetan fox (*V. ferrilata*) and corsac fox (*V. corsac*) in the tree topology for

the 25 Mb region is consistent with overall species relationships. For example, if the 25 Mb region has been specifically introgressed into Rueppell's from a species closely related to the fennec, then the three Asian species should be phylogenetically closer to red foxes at this region, and outgrouped by the arctic fox clade (*V. lagopus*, *V. macrotis*, *V. velox*). On the other hand, if they are phylogenetically placed with Rueppell's fox, this would reflect a more complex history in which the affected 25 Mb region would concern the red fox and arctic fox instead (**Figure 4.1)**.

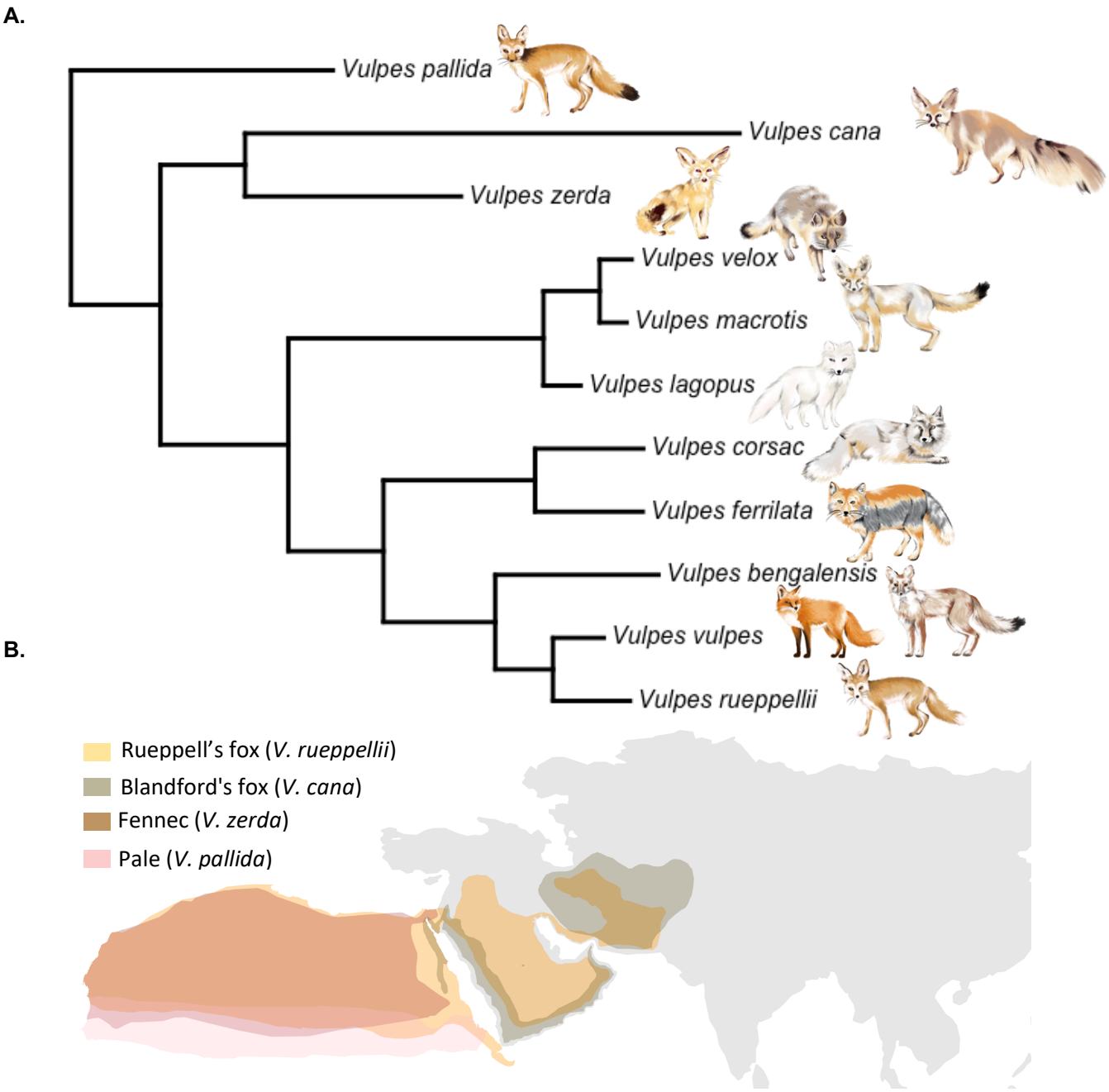


Figure 4.1. Phylogenetic relationships of *Vulpes* spp. (A) and geographically overlapping species with Rueppell's fox other than the red fox (B) (related to Figure 2B in Main Text of Chapter III).

The fact that introgressive-hybridization events observed in this work concerned geographically overlapping species inhabiting arid and hyper-arid regions for multiple generations, facing shared challenges related to extreme temperatures and scarce resources, suggests that introgression may have played an important role in facilitating desert adaptation. Introgression has the potential of introducing different combinations of alleles, involving one or multiple genes, that have already been pre-tested in nature. Rueppell's carry a putatively introgressed 25 Mb region significantly enriched for genes involved in for sulfotransferase activity (SULT1B1, SULT1D1, SULT1E1), neuromedin receptor binding (NMU), protein tyrosine kinase activity (KIT, KDR, EPHA5, PDGFRA, TEC, TXK), and GABA-related activity (GABRB1, GABRA4, GABRG1,GABRA2), and CXCR chemokine receptor binding (CXCL7, CXCL8). Functions of individuals genes include thyroid hormone metabolic process (SULT1B1; GeneCards/UniProt), attenuation of catecholamine activity during stress response and metabolic fasting (SULT1D1; GeneCards/UniProt), energy homeostasis, positive regulation of smooth-muscle contraction and heart rate, temperature homeostasis, stress responses, negative regulation of eating behavior, and regulation of circadian sleep/wake cycle (NMU, Brighton et al. 2004), all of which could be relevant to life in extreme desert conditions (Chapter II, Paper 1). Genome-wide scans for signatures of adaptation further revealed that the gene with second strongest signal of selection, WFS1, is within this 25 Mb region and appears related to Rueppell's fox increased ability to produce hyperosmotic urine. Though we could not directly assess the functional significance for all these genes, we have indirect evidence that, under water deprivation, knockout mice for WFS1 suffer from an impairment of vasopressin (AVP) that prevents them to concentrate urine (Kurimoto et al. 2021). Therefore, variants affecting the expression of this gene in Rueppell's fox could underlie increased abilities to concentrate urine via increased AVP production.

We were also able to link more recent introgression between Rueppell's fox and North-African red foxes to a still poorly investigated gene, SLC6A16, that may be involved in chronic kidney disease and oxidative stress (Pramod et al. 2013; Ainsworth et al. 2017). While the functional relevance of this gene to life in arid environments is not entirely clear, introgressed variants have nearly become fixed in the North-African population, suggesting some adaptive significance. Adaptive introgression of SLC6A16 seems to be supported by high allele frequency differences between the North-African red fox population and Eurasian red foxes, and remarkably low allele frequency differences and divergence between Rueppell's and North-African red foxes. Our analyses also suggested potential introgression of genes with functions that appeared related with adaptation to life in North-African arid environments, such as associated with starch digestion and glycogen storage (MGAM, Axelsson et al. 2013), and pigmentation (HPS5, Pavan and Sturm 2019). However, because introgressed variants at these genes are present in fewer North-African red fox individuals, we could not argue that they may have been targeted by selection, nor rule out a scenario of incomplete lineage sorting. Functional

analyses are therefore needed to understand if or how these candidate genes may have contributed to North-African red foxes desert adaptation.

Overall, providing convincing evidence that introgression has an adaptive significance is extremely challenging when we lack detailed knowledge on the traits involved, and are not able to explicitly demonstrate that introgressed variants have fitness effects in the recipient species. We followed the approach of first detecting introgressed regions and then using selection scans to find signatures of adaptive advantage. The statistics we used rely on the premise that selection is expected to increase linkage disequilibrium and create shifts in allelic frequencies (Nielsen 2005). The problem with this approach is that introgression can in itself create changes in allele frequency and confound tests for detecting positive selection that rely on allele frequency differences. Future analyses using coalescent simulations are thus needed to provide significance for introgression, and functional analysis would be required for additional clarity on its adaptiveness.

Despite widespread evidence of both ancestral hybridization and recent introgression in canid species, introgression has rarely been directly linked to environmental adaptation (Anderson et al. 2009; Schweizer et al. 2018; Wang et al. 2020). So far, the only documented cases in canids are from studies in which introgression was detected in regions previously suggested to be under positive selection (Miao et al. 2017; Wang et al. 2020). For instance, F_{ST}-based selection scans first discovered EPAS1 alleles to be highly differentiated between lowland and highland dogs (Li et al. 2014), and between lowland wolves and highland wolves (Zhang et al. 2014). Then, follow-up studies found that these adaptive alleles were deeply divergent to an extent that could not be explained by selection alone and, therefore, were likely evolved in a presently unknown diverged lineage, having persisted in both Tibetan dogs and wolves due to their adaptive significance to life in high altitudes (Wang et al. 2020). Although a hypoxia-related gene is an obvious candidate for adaptive introgression in high altitudes, evidence for the adaptive significance of introgression in these studies was indirect, without further functional studies that assess the fitness effects of introgressed variants. On the other hand, there is compelling evidence from humans to support both arguments of ancient hybridization with an extinct lineage and of adaptive significance, as EPAS1 was introgressed into Tibetans from a Denisovan-like population, and association studies were able to link variation at this gene with adaptation to high-altitude hypoxia in humans (Yi et al. 2010; Huerta-Sánchez et al. 2014). Although our study does not consistently provide direct functional significance of introgression, it is reasonable to suggest that interspecific gene flow played an important role in the adaptive evolution of desert foxes, particularly when taking into account the increasing body of literature supporting the potential for introgressive hybridization to promote adaptation and contribute with adaptive genetic variation in different spatiotemporal contexts (Liu et al. 2015; Jones et al. 2018; Edelman et al. 2019; Giska et al. 2019; Marques et al. 2019; Oziolor et al. 2019; Witt and Huerta-Sánchez 2019; Jamie and Meier 2020; Cao et al. 2021; Ferreira et al. 2021).

4.3. Insights into complex evolutionary histories

A general lesson from different studies in canids is that introgressive-hybridization events are complex and often confound our understanding of evolutionary history. The narrative around canid adaptation at the Tibetan plateau is a remarkable illustration of this: what was first explained as the result of adaptive introgression from wolves into dogs (Miao et al. 2017; von Holdt et al. 2017) has recently been revised as an ancient introgressive hybridization event with a ghost lineage (Wang et al. 2020). Our work revealed a striking signal for shared genetic variation between Rueppell's foxes and North-African red foxes compared to Eurasian red foxes at the CRADD gene. While this pattern would suggest introgression between Rueppell's and Red foxes, phylogenetic analysis at the CRADD region revealed that this actually resulted from the fact that some Eurasian red foxes and one arctic fox share deeply divergent alleles that predate the origin of all extant members of the *Vulpes* genus. By analyzing our dataset with other members of the Canidae family, and also including published genomes for red foxes from the Russian farm experiment, which are of North American origin (Kukekova et al. 2018) we were able to confirm the occurrence of deeply divergent alleles at the CRADD gene within red foxes. Although we have no means to directly assess the functional significance of CRADD, this gene has been associated with changes in body size in mice, and has been associated to body height, heel bone mineral density and BMI-adjusted hip circumference ratio in humans (GWAS catalog, GeneCards). This gene has also been found to be under selection in arctic sled dogs (Sinding et al. 2020) and humans (Kimura et al. 2008).

The occurrence of such deeply divergent alleles at a locus within a species could reflect: 1) two introgression events involving an extinct canid species that predates the origin of all extant members of *Vulpes* and both arctic fox and Eurasian foxes, 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; 2) the maintenance of ancestral polymorphisms tightly linked in strong linkage disequilibrium across multiple speciation events throughout the *Vulpes* radiation by strong balancing selection on an inversion. Both hypotheses are incredibly complex and difficult to convincingly explain. First, we don't have data on red fox populations that currently overlap in range with the arctic fox to be able to appraise the geographical extent of this gene-species tree discordance in red foxes. The only evidence we have is that affected Eurasian red fox individuals that form a divergent clade with the arctic fox for CRADD either come from the Middle East or include farm-bred foxes that trace back their origin to North America (Kukekova et al. 2018). The southern European foxes included in this study, on the other hand, do not carry the divergent CRADD haplotype. As such we may only speculate that red foxes and arctic foxes may have once overlapped in range with an ancestral canid species. Secondly, we only have one arctic fox individual and therefore cannot appraise whether a highly divergent CRADD haplotype is fixed in this species or shows the same

paraphyletic pattern as red foxes, with some individuals clustering as expected in the *Vulpes* phylogeny, while others remaining as outgroups to the genus. There is a growing body of empirical evidence that shows that ancestral polymorphisms (or trans-species polymorphisms) can be maintained for millions of years and across multiple speciation events by balancing selection and may even predate the first speciation events in a clade (Marques et al. 2019; Jamie and Meier 2020). On the other hand, balancing selection should leave a very narrow trace. Since a large region harboring CRADD can't be explained by balancing selection alone unless there is also recombination suppression, we hypothesize that this would have to involve a structural rearrangement such as an inversion. Phylogenomics studies comprising all extant members of *Vulpes* are needed to dissect the complex evolutionary history of this genomic region. Future sampling efforts should also include a larger number of red fox and arctic fox individuals from diverse localities across their geographic distribution in Eurasia and North America, in order to allow a large-scale comparative study.

4.4. A gradual understanding of the mechanisms underlying desert adaptation

To date, the different sources and nature of adaptive variation in deserts remain largely understudied (Chapter II, Paper 1). Empirical evidence from other extreme-environmental settings, such as the arctic and high altitudes, suggests that both standing genetic variation and introgression are critical components of adaptation (Bedford and Hoekstra 2015; Witt and Huerta-Sánchez 2019; Sinding et al. 2020), and sometimes with equally important contributions to the same evolutionary history (Fumagalli et al. 2015; Racimo et al. 2016). Our work with North-African foxes is concordant with those findings, showing that, though introgressive hybridization events may have, to a certain extent, facilitated desert adaptation by extending the pool of variation on which selection can act upon, standing genetic variation has also been critical for desert adaptation at different evolutionary scales.

In line with what has been reported for other desert mammalian species, selection in desert-dwelling foxes appears to have targeted genetic variation involved in very distinct functions, consistent with the complexity and variety of phenotypes they had to evolve in order to cope with life in extreme conditions of water deprivation, food scarcity and extreme thermal amplitudes (Chapter I, Paper 1). For instance, we find that North-African red foxes recently adapted to their current ecological circumstances of increased aridity via reduced non-renal water losses in mechanisms such as panting and sweating, consistent with what has been reported in other desert mammals (Chapter II, Paper 1). North-African red foxes also adapted to cope with life in extreme temperatures as attested by selection on genes related to both temperature perception and thermoregulation through adaptive thermogenesis. Notably, genes involved in energy metabolism, heat production and adaptive thermogenesis have also been extensively described in different desert mammals (Chapter II, Paper 1). In contrast with other desert-dwelling foxes such as Rueppell's and fennec, we did not find

signatures of selection in North-African red foxes related to kidney function. Instead, the evidence of introgressed genetic variation from Rueppell's fox into almost all North-African red foxes at a gene, SLC6A16, which may be implicated in chronic kidney disease (Ainsworth et al. 2017). This could suggest a complementary role of interspecific gene flow as a rapid source for adaptive variation, as the Sahara continues to expand (Liu and Xue 2020). These results also suggest that selective pressures on water retention mechanisms may have been more constrained at the core of deserts, where water is barely available, (Chapter II). Further investigations of the top candidate genes putatively under selection in North-African red foxes (SLC12A2, THADA, TRPA1) show that most variants targeted by selection lie outside coding regions. This suggests that changes in noncoding or regulatory areas of the genomes affecting the expression of genes are likely to play a more important role in driving North-African red foxes recent desert adaptation than changes affecting protein structure. Future analyses are needed to investigate the nature of genetic variants targeted by selection in core desert-dwelling species adapted to life in hyper-arid regions for longer evolutionary times. This will allow us to appraise the extent to which the nature of desert adaptation relates to evolutionary age (Chapter II).

4.5. Challenges in dissecting the genetic basis of traits in non-model organisms

I started this PhD thesis with the preliminary knowledge from microsatellite data that Rueppell's and red foxes were much more closely related than what would be expected for two divergent species (F_{ST} =0.14), and with a possible explanatory scenario that Rueppell's fox could constitute a desert ecotype evolved from red foxes (Leite et al. 2015). We had hoped that with such low genetic differentiation between the two species we could use a population genomics approach to perform a selection study oriented towards dissecting the genetic basis of desert related traits, including textbook examples such as urine concentrating ability (Chapter II, Paper 2). To this end, we performed a comprehensive literature search for phenotypic traits potentially relevant to life in deserts (Chapter II) and attempted the sampling of different physiological parameters (Chapter III). However, whole-genome analysis revealed that Rueppell's and red foxes were highly differentiated (F_{ST}=0.5), undermining the possibility of linking genotypes with phenotypes directly. First, the majority of population-based methods lack power to disentangle selection from random genetic drift when average genome-wide differentiation is high in natural populations (Crawford and Nielsen 2013). Secondly, it is not possible to establish genotype-phenotype associations if traits of interest are not found segregating in nature, and only very closely related groups of organisms with low genome-wide differentiation are likely to display trait polymorphism (Crawford and Nielsen 2013). Lastly, capturing foxes in the field proved to be a difficult task, and we lacked the time and the infrastructures to obtain the required sample sizes for genotype-phenotype associations. These and many other

challenges extensively explored in our review paper (Chapter II) underscore why our understanding of the genetic basis for phenotypic diversity is concentrated around a narrow range of species and/or very specific traits. As such, while our work with North-African foxes adds to our review (Chapter II) by shedding light on some of the evolutionary mechanisms underlying desert adaptation at different spatiotemporal scales, our review adds to this empirical work (Chapter III) by providing a set of perspectives that may help guide future and ongoing work.

4.6. Concluding Remarks and Perspectives

Deserts make up a third of the planet's land surface, comprising alien landscapes characterized by never-ending tracts of sand or rocks, as well as extreme harsh climatic conditions apparently too hostile to sustain life. And yet, life finds a way. I write this from the personal experience of someone who had the privilege of witnessing deserts extraordinary scenery and intriguing life forms, and with the hope that the scientific outputs presented in this dissertation may contribute to a better understanding of the evolutionary mechanisms underlying desert life.

The first part of this thesis offers a general perspective on the state-of-the-art and integrates knowledge from classical studies in desert eco-physiology with emerging research on the genomics of adaptation to deserts (Paper 1, Chapter II). This chapter also contributes to the establishment of the general hypothesis that aridity is one of the main selective pressures driving phenotypic evolution in different desert mammalian lineages (Paper 1, Chapter II). Future studies investigating association of diverse bioclimatic variables with desert phenotypes and genotypes are needed, as they hold the potential to elucidate future adaptative responses to increasingly arid conditions. In addition, we present a valuable discussion on the importance of working with a diverse set of desert-dwelling organisms and provide explicit recommendations on how future research could be guided to help circumvent outstanding challenges and help rescue biodiversity threatened by desertification (Paper 1, Chapter II). The second part of this dissertation (Chapter III) provides an outstanding empirical example on how different species adapted to extreme desert conditions throughout different evolutionary times. Studying the genetic basis of desert adaptation in North-African foxes not only allowed us to scratch the surface of the many outstanding questions concerning the evolutionary mechanisms driving adaptation in desert environments, but also provided a glimpse into the origins and evolution of some of the most elusive desert-dwelling animals.

Despite the challenges of establishing genotype-phenotype associations, we were able to generate the most comprehensive genomic dataset for *Vulpes* to date, including whole genomes for 82 foxes belonging to 4 distinct species, and successfully sampled many individuals for which we could gather expression data to be used as intermediate steps between genes and phenotypes and help inform the functional effects of adaptive genetic variation. With this dataset we were able to study the evolutionary history of these charismatic species in terms of possible introgressive-hybridization

events, divergence time, demographic history and selection, and therefore significantly contribute with new insights on the evolutionary mechanisms at play during desert adaptation. In particular, we have shown that ancient introgressive events combined with independent selection on adaptive tolerance to extreme temperatures and increased capacity to preserve water have enabled North-African foxes' survival in desert regions at different evolutionary scales. Ultimately, we open a window into North-African foxes' adaptive genetic variation and show that the joint role of introgression and selection on ancestral standing variation may extend well-beyond extant sister species and may involve genes associated with complex structural rearrangements.

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 (Vale and Brito 2015; Iknayan and Beissinger 2018; Liu and Xue 2020). Given the role of aridity as a major selective pressure driving phenotypic evolution, we expect ancestral standing adaptive genetic variation and potential introgression of alleles pre-tested by natural selection in core hyper-arid dwelling species to be crucial components of ongoing adaptation to rapidly changing climates in foxes living at the increasingly arid edges of deserts. Though preserving adaptive variation in these charismatic umbrella species may help conservation efforts focused on rescuing biodiversity currently challenged by desertification, we still require a deeper understanding of the direct relationships between genes, phenotypes, the environment and how these relationships translate into fitness. Our inability to provide genotype-phenotype-fitness maps for North-African foxes lends itself to the many challenges of working with elusive and poorly studied organisms in remote arid regions. As follow-up to our selection scans, and considering the future directions offered in the first part of this thesis, we hope to be able to validate and test functional hypothesis about our candidate genes by resorting to novel techniques such as CRISPR-Cas9 or cell culture. We also hope to gather high-quality *Vulpes* genomes not only to elucidate the role of potential structural rearrangements as drivers of adaptive evolution, but also to take advantage of phylogenetic comparative methods to test for associations between different variables on a phylogeny with several *Vulpes* species.

In sum, this dissertation comprises a group of diverse contributions - including a review, a meta-analysis and empirical research - that will hopefully be of interest to desert researchers, and integrative biologists that wish to address the genetic basis of desert adaptation. Finally, I would like to highlight the importance of going into the desert when studying deserts and not thinking so much about nature and biodiversity as "systems" more or less suitable to address questions, or amenable to the available methodologies. In the course of this work, I have learned from foxes in the Sahara that one only truly understands the things that one captivates.

References

Ainsworth H.C., Langefeld C.D., Freedman B.I. 2017. Genetic epidemiology in kidney disease. Nephrol. Dial. Transplant. 32:ii159–ii169.

Anderson T.M., VonHoldt B.M., Candille S.I., Musiani M., Greco C., Stahler D.R., Smith D.W., Padhukasahasram B., Randi E., Leonard J.A., Bustamante C.D., Ostrander E.A., Tang H., Wayne R.K., Barsh G.S. 2009. Molecular and evolutionary history of melanism in Norths American gray wolves. Science (80-.). 323:1339–1343.

Axelsson E., Ratnakumar A., Arendt M.L., Maqbool K., Webster M.T., Perloski M., Liberg O., Arnemo J.M., Hedhammar Å., Lindblad-Toh K. 2013. The genomic signature of dog domestication reveals adaptation to a starch-rich diet. Nature. 495:360–364.

Bedford N.L., Hoekstra H.E. 2015. Peromyscus mice as a model for studying natural variation. Elife. 4.

Brighton P.J., Szekeres P.G., Willars G.B. 2004. Neuromedin U and its receptors: Structure, function, and physiological roles. Pharmacol. Rev. 56:231–248.

Brito J.C., Acosta A.L., Álvares F., Cuzin F. 2009. 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.

Cao Y.H., Xu S.S., Shen M., Chen Z.H., Gao L., Lv F.H., Xie X.L., Wang X.H., Yang H., Liu C. Bin, Zhou P., Wan P.C., Zhang Y.S., Yang J.Q., Pi W.H., Hehua Ee., Berry D.P., Barbato M., Esmailizadeh A., Nosrati M., Salehian-Dehkordi H., Dehghani-Qanatqestani M., Dotsev A. V., Deniskova T.E., Zinovieva N.A., Brem G., Štěpánek O., Ciani E., Weimann C., Erhardt G., Mwacharo J.M., Ahbara A., Han J.L., Hanotte O., Miller J.M., Sim Z., Coltman D., Kantanen J., Bruford M.W., Lenstra J.A., Kijas J., Li M.H. 2021. Historical Introgression from Wild Relatives Enhanced Climatic Adaptation and Resistance to Pneumonia in Sheep. Mol. Biol. Evol. 38:838–855.

Crawford J.E., Nielsen R. 2013. Detecting adaptive trait loci in nonmodel systems: divergence or admixture mapping? Mol. Ecol. 22:6131–6148.

Edelman N.B., Frandsen P.B., Miyagi M., Clavijo B., Davey J., Dikow R.B., García-Accinelli G., Van Belleghem S.M., Patterson N., Neafsey D.E., Challis R., Kumar S., Moreira G.R.P., Salazar C., Chouteau M., Counterman B.A., Papa R., Blaxter M., Reed R.D., Dasmahapatra K.K., Kronforst M., Joron M., Jiggins C.D., Owen McMillan W., Palma F. Di, Blumberg A.J., Wakeley J., Jaffe D., Mallet J. 2019. Genomic architecture and introgression shape a butterfly radiation. Science (80-.). 366:594–599.

Ferreira M.S., Jones M.R., Callahan C.M., Farelo L., Tolesa Z., Suchentrunk F., Boursot P., Mills

L.S., Alves P.C., Good J.M., Melo-Ferreira J. 2021. The Legacy of Recurrent Introgression during the Radiation of Hares. Syst. Biol. 70:593–607.

Fumagalli M., Moltke I., Grarup N., Racimo F., Bjerregaard P., Jørgensen M.E., Korneliussen T.S., Gerbault P., Skotte L., Linneberg A., Christensen C., Brandslund I., Jørgensen T., Huerta-Sánchez E., Schmidt E.B., Pedersen O., Hansen T., Albrechtsen A., Nielsen R. 2015. Greenlandic Inuit show genetic signatures of diet and climate adaptation. Science (80-.). 349:1343–1347.

Giska I., Farelo L., Pimenta J., Seixas F.A., Ferreira M.S., Marques J.P., Miranda I., Letty J., Jenny H., Hackländer K., Magnussen E., Melo-Ferreira J. 2019. Introgression drives repeated evolution of winter coat color polymorphism in hares. Proc. Natl. Acad. Sci. U. S. A. 116:24150–24156.

Gopalakrishnan S., Sinding M.H.S., Ramos-Madrigal J., Niemann J., Samaniego Castruita J.A., Vieira F.G., Carøe C., Montero M. de M., Kuderna L., Serres A., González-Basallote V.M., Liu Y.H., Wang G.D., Marques-Bonet T., Mirarab S., Fernandes C., Gaubert P., Koepfli K.P., Budd J., Rueness E.K., Heide-Jørgensen M.P., Petersen B., Sicheritz-Ponten T., Bachmann L., Wiig Ø., Hansen A.J., Gilbert M.T.P. 2018. Interspecific Gene Flow Shaped the Evolution of the Genus Canis. Curr. Biol. 28:3441-3449.e5.

von Holdt B., Fan Z., Ortega-Del Vecchyo D., Wayne R.K. 2017. EPAS1 variants in high altitude Tibetan wolves were selectively introgressed into highland dogs. PeerJ. 2017:e3522.

Huerta-Sánchez E., Jin X., Asan, Bianba Z., Peter B.M., Vinckenbosch N., Liang Y., Yi X., He M., Somel M., Ni P., Wang B., Ou X., Huasang, Luosang J., Cuo Z.X.P., Li K., Gao G., Yin Y., Wang W., Zhang X., Xu X., Yang H., Li Y., Wang J., Wang J., Nielsen R. 2014. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature. 512:194–197.

Iknayan K.J., Beissinger S.R. 2018. Collapse of a desert bird community over the past century driven by climate change. Proc. Natl. Acad. Sci. U. S. A. 115:8597–8602.

Jamie G.A., Meier J.I. 2020. The Persistence of Polymorphisms across Species Radiations. Trends Ecol. Evol. 35:795–808.

Jones M.R., Scott Mills L., Alves P.C., Callahan C.M., Alves J.M., Lafferty D.J.R., Jiggins F.M., Jensen J.D., Melo-Ferreira J., Good J.M. 2018. Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. Science (80-.). 360:1355–1358.

Kimura R., Ohashi J., Matsumura Y., Nakazawa M., Inaoka T., Ohtsuka R., Osawa M., Tokunaga K. 2008. Gene Flow and Natural Selection in Oceanic Human Populations Inferred from Genome-Wide SNP Typing. Mol. Biol. Evol. 25:1750–1761.

Koepfli K.P., Pollinger J., Godinho R., Robinson J., Lea A., Hendricks S., Schweizer R.M., Thalmann O., Silva P., Fan Z., Yurchenko A.A., Dobrynin P., Makunin A., Cahill J.A., Shapiro B., Álvares F., Brito J.C., Geffen E., Leonard J.A., Helgen K.M., Johnson W.E., O'Brien S.J.,

Van Valkenburgh B., Wayne R.K. 2015. Genome-wide evidence reveals that African and Eurasian golden jackals are distinct species. Curr. Biol. 25:2158–2165.

Kukekova A. V., Johnson J.L., Xiang X., Feng S., Liu S., Rando H.M., Kharlamova A. V., Herbeck Y., Serdyukova N.A., Xiong Z., Beklemischeva V., Koepfli K.P., Gulevich R.G., Vladimirova A. V., Hekman J.P., Perelman P.L., Graphodatsky A.S., O'Brien S.J., Wang X., Clark A.G., Acland G.M., Trut L.N., Zhang G. 2018. Red fox genome assembly identifies genomic regions associated with tame and aggressive behaviours. Nat. Ecol. Evol. 2:1479–1491.

Kurimoto J., Takagi H., Miyata T., Hodai Y., Kawaguchi Y., Hagiwara D., Suga H., Kobayashi T., Sugiyama M., Onoue T., Ito Y., Iwama S., Banno R., Tanabe K., Tanizawa Y., Arima H. 2021. Deficiency of WFS1 leads to the impairment of AVP secretion under dehydration in male mice. Pituitary. 1:3.

Leite J.V., Álvares F., Velo-Antón G., Brito J.C., Godinho R. 2015. 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.

Li Y., Wu D.D., Boyko A.R., Wang G.D., Wu S.F., Irwin D.M., Zhang Y.P. 2014. Population variation revealed high-altitude adaptation of tibetan mastiffs. Mol. Biol. Evol. 31:1200–1205.

Liu K.J., Steinberg E., Yozzo A., Song Y., Kohn M.H., Nakhleh L. 2015. Interspecific introgressive origin of genomic diversity in the house mouse. Proc. Natl. Acad. Sci. U. S. A. 112:196–201.

Liu S., Lorenzen E.D., Fumagalli M., Li B., Harris K., Xiong Z., Zhou L., Korneliussen T.S., Somel M., Babbitt C., Wray G., Li J., He W., Wang Z., Fu W., Xiang X., Morgan C.C., Doherty A., O'Connell M.J., McInerney J.O., Born E.W., Dalén L., Dietz R., Orlando L., Sonne C., Zhang G., Nielsen R., Willerslev E., Wang J. 2014. Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. Cell. 157:785–794.

Liu Y., Xue Y. 2020. Expansion of the Sahara Desert and shrinking of frozen land of the Arctic. Sci. Rep. 10:1–9.

Marques D.A., Meier J.I., Seehausen O. 2019. A Combinatorial View on Speciation and Adaptive Radiation. Trends Ecol. Evol. 34:531–544.

Miao B., Wang Z., Li Y. 2017. 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.

Nielsen R. 2005. Molecular Signatures of Natural Selection. Annu. Rev. Genet. 39:197–218.

Oziolor E.M., Reid N.M., Yair S., Lee K.M., VerPloeg S.G., Bruns P.C., Shaw J.R., Whitehead A., Matson C.W. 2019. Adaptive introgression enables evolutionary rescue from extreme environmental pollution. Science (80-.). 364:455–457.

Pavan W.J., Sturm R.A. 2019. The Genetics of Human Skin and Hair Pigmentation. Annu. Rev. Genomics Hum. Genet. 20:41–72.

Pramod A.B., Foster J., Carvelli L., Henry L.K. 2013. SLC6 transporters: Structure, function,

regulation, disease association and therapeutics. Mol. Aspects Med. 34:197–219.

Racimo F., Gokhman D., Fumagalli M., Ko A., Hansen T., Moltke I., Albrechtsen A., Carmel L., Huerta-Sánchez E., Nielsen R. 2016. Archaic adaptive introgression in *TBX15/WARS2*. Mol. Biol. Evol. 34:msw283.

Schweizer R.M., Durvasula A., Smith J., Vohr S.H., Stahler D.R., Galaverni M., Thalmann O., Smith D.W., Randi E., Ostrander E.A., Green R.E., Lohmueller K.E., Novembre J., Wayne R.K. 2018. Natural Selection and Origin of a Melanistic Allele in North American Gray Wolves. Mol. Biol. Evol. 35:1190–1209.

Signore A. V, Yang Y.-Z., Yang Q.-Y., Qin G., Moriyama H., Ge R.-L., Storz J.F. 2019. Adaptive Changes in Hemoglobin Function in High-Altitude Tibetan Canids Were Derived via Gene Conversion and Introgression. Mol. Biol. Evol. 36:2227–2237.

Sillero-Zubiri C., Hoffmann M., Macdonald D.W. 2004. Canids: Foxes, Wolves, Jackals and Dogs. Status survey and conservation action plan. .

Sinding M.H.S., Gopalakrishnan S., Ramos-Madrigal J., Manuel M. De, Pitulko V. V., Kuderna L., Feuerborn T.R., Frantz L.A.F., Vieira F.G., Niemann J., Castruita J.A.S., Carøe C., Andersen-Ranberg E.U., Jordan P.D., Pavlova E.Y., Nikolskiy P.A., Kasparov A.K., Ivanova V. V., Willerslev E., Skoglund P., Fredholm M., Wennerberg S.E., Heide-Jørgensen M.P., Dietz R., Sonne C., Meldgaard M., Dalén L., Larson G., Petersen B., Sicheritz-Pontén T., Bachmann L., Wiig Ø., Marques-Bonet T., Hansen A.J., Gilbert M.T.P. 2020. Arctic-adapted dogs emerged at the Pleistocene-Holocene transition. Science (80-.). 368:1495–1499.

Vale C.G., Brito J.C. 2015. Desert-adapted species are vulnerable to climate change: Insights from the warmest region on Earth. Glob. Ecol. Conserv. 4:369–379.

VonHoldt B.M., Cahill J.A., Fan Z., Gronau I., Robinson J., Pollinger J.P., Shapiro B., Wall J., Wayne R.K. 2016. Whole-genome sequence analysis shows that two endemic species of North American Wolf are admixtures of the coyote and gray Wolf. Sci. Adv. 2:1–14.

Wang M.S., Wang S., Li Y., Jhala Y., Thakur M., Otecko N.O., Si J.F., Chen H.M., Shapiro B., Nielsen R., Zhang Y.P., Wu D.D. 2020. Ancient hybridization with an unknown population facilitated high-altitude adaptation of canids. Mol. Biol. Evol. 37:2616–2629.

Witt K.E., Huerta-Sánchez E. 2019. Convergent evolution in human and domesticate adaptation to high-altitude environments. Philos. Trans. R. Soc. B Biol. Sci. 374.

Yi X., Liang Y., Huerta-Sanchez E., Jin X., Xi Ping Cuo Z., Pool J.E., Xu X., Jiang H., Vinckenbosch N., Korneliussen T.S., Zheng H., Liu T., He W., Li K., Luo R., Nie X., Wu H., Zhao M., Cao H., Zou J., Shan Y., Li S., Yang Q., Asan, Ni P., Tian G., Xu J., Liu X., Jiang T., Wu R., Zhou G., Tang M., Qin J., Wang T., Feng S., Li G., Huasang, Luosang J., Wang W., Chen F., Wang Y., Zheng X., Li Z., Bianba Z., Yang G., Wang X., Tang S., Gao G., Chen Y., Luo Z., Gusang L., Cao Z., Zhang Q., Ouyang W., Ren X., Liang H., Zheng H., Huang Y., Li J., Bolund L.,

Kristiansen K., Li Y., Zhang Y., Zhang X., Li R., Li S., Yang H., Nielsen R., Wang J., Wang J. 2010. Sequencing of Fifty Human Exomes Reveals Adaptations to High Altitude. Science (80-.). 329:75–78.

Zhang W., Fan Z., Han E., Hou R., Zhang L., Galaverni M., Huang J., Liu H., Silva P., Li P., Pollinger J.P., Du L., Zhang X.Y., Yue B., Wayne R.K., Zhang Z. 2014. Hypoxia Adaptations in the Grey Wolf (Canis lupus chanco) from Qinghai-Tibet Plateau. PLoS Genet. 10.