UNIVERSITY OF CALIFORNIA
RIVERSIDE

Use of Genomic Data to Resolve Evolutionary Questions in Kangaroo Rats (Genus *Dipodomys*)

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Evolution, Ecology, and Organismal Biology

by

Yuwei Cui

September 2023

Dissertation Committee:
Dr. Leonard Nunney, Chairperson
Dr. Alan Brelsford
Dr. Mark Springer

The Dissertation of Yuwei Cui is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

Acknowledgements

I thank my advisor, Dr. Leonard Nunney, for his great support and inspiring mentoring on my research. His honesty, intelligence, and kindness helped me a lot in my study and life at UCR. I thank my lab mates, Kevin Thai, and Dr. Derek O'Meara, for their company and support.

I thank the professors on my guidance committee, oral examination committee, and dissertation committee. To Dr. Alan Brelsford, thank you for your instruction on genomics. To Dr. Mark Springer, thank you for holding the meetings to discuss the phylogeny papers. To Dr. Thomas Girke, thank you for your guidance on bioinformatics. To Dr. Richard Stouthamer and Dr. Paul Rugman-Jones, thank you for providing the lab resources and instructions. I thank the other faculty and staff at EEOB Department and the school for your help.

I thank my student volunteers Grace Li and Samantha Hom for helping with trapping and sampling the animals, and Jennifer Hoffman and Tara Dadafarin in the biological monitoring at Western Riverside County Regional Conservation Authority and Joseph E Messin at UCR reserves for providing the training and helping with sample collection.

I thank my cohort and school mates, for your friendship and the good times we spent together. I thank my parents and friends, for your support through my graduate school time.

ABSTRACT OF THE DISSERTATION

Use of Genomic Data to Resolve Evolutionary Questions in Kangaroo Rats (Genus *Dipodomys)*

by

Yuwei Cui

Doctor of Philosophy, Graduate Program in Evolution, Ecology and Organismal Biology
University of California, Riverside, September 2023
Dr. Leonard Nunney, Chairperson

The development of sequencing techniques and increased availability of genome assemblies have enabled comparative studies on genome evolution in the non-model organisms to better understand genomic diversity and adaptation. Using genomic data, we investigated evolutionary questions in the kangaroo rats (genus *Dipodomys*), a group of New World rodents adapted to arid and semi-arid habitats, and we revealed examples of genomic signatures of their speciation and adaptive evolution. We first resolved the taxonomic issue of whether the formerly conspecific species with different karyotypes, the agile kangaroo rat (AKR, *D, agilis*, 2N=62) and the Dulzura kangaroo rat (DKR, *D. simulans*, 2N=60) are distinct species. The results supported their species level status with no evidence of sympatry or interspecific gene flow. We estimated their divergence time as ~0.7 Mya in middle Pleistocene, followed by a ~5-fold decline in the DKR's effective population size (N_e) to ~20,000 at ~0.1 Mya, recently recovering to ~50,000, converging with that of AKR. AKR is generally distributed to higher elevations than

DKR, and we investigated the inter- and intra-specific effects of elevation. The mean length of runs of homozygosity (ROHs) was negatively correlated with the elevation, suggested more structured populations at higher elevations in the two species, yet the levels of nucleotide diversity (π) were constant. AKR showed more selective sweeps than DKR, apparently associated with adaptation to high elevation, e.g., thermoregulation and response to hypoxic stress, and intraspecific comparisons among populations also revealed selective sweeps associated with local adaptation linked to elevation. The changes identified appeared to be regulatory. Finally, we employed data-driven and hypothesis-driven approaches to identify genes under positive selection in arid adapted heteromyids (kangaroo rats and mice and pocket mice) and convergently in unrelated arid adapted rodents including jerboa, gerbil, and cactus mouse. Notable among the candidate genes were those involved in energy production and hemostasis, and in response to osmotic stress.

Table of Contents

Chapter 1: The use of genomic data in the analysis of speciation and demography of two formerly conspecific kangaroo rats, the agile kangaroo rat (*D. agilis*) the Dulzura kangaroo rat (*D. simulans*)

Abstract 1

Introduction 2

Methods. 5

Results 12

Discussion 14

Conclusion 18

Figures 19

Chapter 2: Genomic evidence of adaptation linked to elevation in two previously conspecific kangaroo rats, the agile kangaroo rat (*D. agilis*) and the Dulzura kangaroo rat (*D. simulans*)

Abstract 24

Introduction 24

Methods. 28

Results 31

Discussion 39

Conclusion 49

Figures 50

Chapter 3: Detection of genomic adaptation and convergence in arid-adapted rodents with a focus on the Heteromyidae.

Abstract 54

Introduction 54

Methods. 59

Results 68

Discussion 73

Conclusion 78

Figures 79

Synthesis 84

Reference 85

List of Figures

Chapter 1.

Figure 1.1. Sampling sites and the location of the salvaged animal. The elevations and samples sizes (n) are shown in parentheses below the sample site abbreviations. For full names of sample sites, see Methods. The shading defines major mountain ranges, and black line defines Santa Ana River.

Figure 1.2. Major genetic clusters defined by the genomes of individuals sampled along the North-South transect. K defines the number of clusters fitted, with K=2 separating DKR and AKR. Sample sites are shown at the bottom, with genetic composition of each of the n individuals sampled at a site defined by a vertical bar.

Figure 1.3. Principal component analysis (PCA) on genome-wide SNPs in samples from all sites. The AKR samples (unfilled symbols) are separated from the DKR samples (filled symbols) by PC1.

Figure 1.4. ML phylogenic of six kangaroo rat species (Genus *Dipodomys*). PPM was used as an outgroup to root the tree. The branch lengths scaled by the rate of substitutions are shown above the branches with the estimated divergence times below. All nodes have 100% bootstrap support.

Figure 1.5. Historical effective population sizes of AKR and DKR. The green and red lines represent the changes in historical N_e of AKR and DKR, respectively.

Chapter 2.

Figure 2.1. AKR and DKR sampling sites. (a) AKR sites. (b) DKR sites. The site of the sample for Pacbio reference genome (SCP) is label as a filled square and others as filled triangles. The isohypse lines were obtained on http://contourmapcreator.urgr8.ch/.

Figure 2.2. Mean sizes of ROHs in AKR and DKR populations. The box plots show the mean sizes of ROHs in all sampling sites, and the sites for the same species are ordered by elevation from high to low. The round dots and horizontal lines represent mean and median values respectively. The error bars represent the 95% confidence intervals, and the diamonds show outliners.

Figure 2.3. The numbers of windows with top 1% Fst and reduction in π on different contigs in AKR. Two contigs, ptg000074l and ptg000045l possess relatively more regions under selective sweeps than the other contigs.

Figure 2.4. Principal component analysis (PCA) using genome-wide SNPs for each species considered separately. (a) AKR samples showing clear separation of sample sites. (b) DKR samples showing that BXSL separate from the other sites.

Chapter 3.

Figure 3.1. Phylogenetic tree of rodent taxa used to identify adaptation in the Heteromyidae. The branch lengths were estimated in PAML using synonymous substitution rates in the concatenate genes but were not used in the analyses. The tree is rooted using human sequences (not shown).

Figure 3.2. Phylogenic tree of 4 pairs of rodent taxa used for convergence study. The branch lengths were estimated in PAML using synonymous substitution rates in the concatenate genes but were not used in the analyses. The tree is rooted using human sequences (not shown). The branches in red are arid adapted, the one in blue is aquatic adapted, and the ones in black are not adapted to special environments.

Figure 3.3. Best fit models for the candidate genes in Heteromyidae detected using the data-driven approach. The branches with the same color had the same ω in the best fit models, with red defining the larger ω and black. In the 18 candidate genes, 3 showed (a), 2 showed (b), 2 showed (c), and 11 showed (d).

Figure 3.4. Phylogenic tree of the species used in hypothesis-driven approach. The branch lengths were estimated in PAML using synonymous substitution rates in the concatenated genes but were not used in the analyses. The tree is rooted using human sequences (not shown).

Figure 3.5. Best fit models for the candidate genes in Heteromyidae detected using the hypothesis-driven approach. The branches with the same color had the same ω in the best fit models, with red defining the largest ω, then orange, and then black.

List of Tables

Chapter 1.

Table 1.1. Genome-wide Fst between AKR and DKR populations. AKR sites are underlined.

Chapter 2.

N/A

Chapter 3.

Table 3.1. Maximum likelihood models and hierarchical analyses to detect positive selection in heteromyids. The models are defined as in PAML (M0 or M2) followed by the number of dn/ds (=ω) values estimated.

Table 3.2. Maximum likelihood models and hierarchical analyses to detect convergent evolution in the arid adapted rodents.

Table 3.3. Maximum likelihood models and analyses to detect gene under positive selection in each arid adapted rodent.

Table 3.4. The candidate genes for adaptation with significantly elevated ωs in heteromyids. The Chi squares and p values are for comparing M2ω2 to M0ω1. The ωs

are the ones in the best fit models (Mbf) defined in Figure 3. ω1 is for the branches under selection, and ω0 is for the background branches.

Table 3.5. The candidate genes for convergent evolution with significantly elevated ωs in arid adapted rodents. The Chi squares and p values are for comparing M2ω2 to M0ω1. The ωs are the ones in the M2ω2. ω-arid is for the arid-adapted rodents, and ω-nonarid is for the background branches.

Table 3.6. The candidate genes with significantly elevated ωs in one of the arid species. The Genes listed show a significantly elevated ω in the target species arid (ω-arid) species, relative to the non-arid control and their ancestral branch (ω-nonarid), tested by comparing model M2ω3 to M2ω2.

Table 3.7. The candidate genes with elevated ωs in heteromyids from selected GO terms. The Chi squares and p values are for comparing M2ω2 to M0ω1. The ωs are the ones in the best fit models (Mbf).

Chapter 1: The use of genomic data in the analysis of speciation and demography of two formerly conspecific kangaroo rats, Dulzura kangaroo rat (*D. simulans*) and the agile kangaroo rat (*D. agilis*)

Abstract

Karyotypes differences within the Pacific kangaroo rat *Dipodomys agilis*, combined with other minor differences, resulted in its division in 1997 into two species, the agile kangaroo rat (AKR, *D. agilis*, 2N=62) in the north of its range in California, and the Dulzura kangaroo rat (DKR, *D. simulans,* 2N=60) in the south, with a possible zone of sympatry immediately south of the San Gabriel and San Bernardino Mountains. To test this revision, we sampled along a ~120km transect from north of the San Gabriel mountains to SW Riverside County. The genomic data supported their species-level status. It showed marked genetic differentiation (Fst>0.2) with no evidence of hybridization or of sympatry, even though AKR was found at the southern edge of the mountains, precluding a mountain barrier. Adding 4 additional kangaroo rat genomes, we estimated divergence times back to ~5mya. AKR and DKR diverged from Stephen's kangaroo rat (*D. stephensi*) ~2.5mya, and from each other ~0.7mya. After their split, DKR's effective population size (N_e) declined to ~20,000 0.1mya, ~5x lower than AKR's. More recently their N_e converged at ~50,000, although runs of homozygosity indicated AKR has a smaller neighborhood size; however, nucleotide diversity was the same in both species.

Introduction

Species are mainly defined by their geographical distribution and their morphological traits but other features ranging from behavior to karyotype may also be used. However, species delineation can be complicated by historical gene flow and current hybridization between taxa. In the past, gene introgression and hybridization were generally difficult to detect; however, with the development of molecular and genomic tools, cases have been found in a wide variety of species (Baack & Rieseberg 2007), including birds (Burri et al 2015; Han et al 2017), insects (Pardo-Diaz et al 2012), plants (Suarez-Gonzalez et al 2018; Wong, Hiscock, & Filatov 2022). Similarly, recent genomic studies suggest that there are more hybrid zones and gene introgression between mammalian taxa than previously acknowledged (Beysard et al 2012; Shurtliff et al 2013; Fontsere et al 2019; Kubiak et al 2020; Ge et al 2022). Such studies emphasize the importance of genomic data in confirming ambiguous species delineations.

One factor that can minimize gene flow and promote reproductive isolation and speciation is the development of karyotype differences within a species (Castiglia 2014). Karyotype differences can inhibit hybridization if there is a fitness loss in chromosomal heterozygotes due to meiotic incompatibilities. For example, in the western house mouse (*Mus musculus domesticus*), the reproductive isolation between the Poschiavo race (2N = 26) and the Upper Valtellina (2N=24) race is likely a result of reduced fertility in hybrids (Capanna 1982; Capanna et al 1984; 1985; Lavrenchenko & Bulatova 2016).

However, such fitness loss is not inevitable. Karyotype polymorphism resulting from a Robertsonian fusion (the joining of two acrocentric chromosomes at their

centromeric region) is widespread in mammals suggesting chromosomal heterozygotes suffer little or no fitness loss (Dobigny et al 2015). For example, in the shrews of the *Sorex araneus* group, microsatellite data showed that their chromosomal rearrangements had no effect on gene flow (Horn et al 2012), and there was no indication of a fitness loss due to extensive karyotype polymorphism in the Brazilian marsh rat *Holochilus brasiliensis* (Nachman 1992). Such observations of karyotype polymorphism indicate that karyotype differences may not always be a good indication of reproductive isolation or species level status, again emphasizing the importance of genomic data in determining genetic isolation and species delineation.

These complex issues of defining taxonomic status in the absence of genomic data are apparent in the kangaroo rats (*Dipodomys* spp.). Intraspecific karyotype polymorphism is found in several species of kangaroo rat, *D. panamintinus*, *D. spectabilis*, and *D. microps* (Stock 1974; Csuti 1979), but a karyotype difference among subspecies of the Pacific kangaroo rat (PacKR, *D. agilis*) was the primary basis for redefining PacKR as two species, *D. agilis*, now given the common name of the agile kangaroo rat (AKR), and *D. simulans*, the Dulzura kangaroo rat (DKR) (Sullivan and Best 1997).

Prior to 1997, PacKR was considered a single species divided into a number of geographically separated subspecies with a range extending from Tulare County in California to Magdalena Plain of southern Baja California (Best 1978; 1983). However, chromosomal analyses showed karyotype differences between two of the northern

subspecies (subspp. *agilis* and *perplexus*) with 2N = 62 (Csuti 1971) and two of the southern subspecies (subspp. *simulans* and *plectilis*) with 2N = 60 (Stock 1974).

The morphological analysis of Best (1983) across the range of PacKR failed to find clear species-level differences between populations or variation associated with karyotype differences; however, subsequent morphological and allozyme analyses indicated that individuals from northern populations (2N=62) were generally larger and clustered together genetically, although the differences found did not suggest species-level separation (Best et al 1986). Sullivan and Best (1997) provided a more extensive analysis of the morphological differences and concluded that the chromosomal variants of PacKR represented two distinct species,

Sullivan and Best (1997) also examined animals of unknown karyotype and classified them by morphology. Their results suggested that the two species had a broad zone of sympatry from the southern foothills of the San Gabriel and San Bernardino mountains south to the Santa Ana Mountains. This raised the possibility of ongoing or historical hybridization and genetic introgression, possibly at a level inconsistent with the species designations.

This subdivision of the Pacific kangaroo rat into the two species, AKR and DKR, has been generally accepted; however, there has been no genetic examination of the differences between these two species, and there is no other evidence of significant reproductive isolation. While the karyotype difference could indicate a species-level difference between the taxa, the supporting data remain relatively weak. To evaluate the degree of genetic isolation between the taxa, we used genomic data from individuals

sampled along a north-south transect from the northern edge of the San Gabriel Mountains to the lower elevation lands to the south in the western Riverside County.

Sullivan and Best (1997) hypothesized that AKR are distributed at higher elevations (>800m) in the San Gabriel, San Bernardino, and coastal Santa Monica mountains while DKR occurs at lower elevations (<800m) south from the foothills of those mountains, suggesting that elevation may be a major environmental factor influencing the distribution of these two species in the area of potential sympatry. To test this hypothesis, we also sampled within the hypothesized sympatric area from a low elevation and high elevation site in the Box Springs Mountains (BXS).

To determine whether AKR and DKR are distinct species, we estimated the genome-wide differentiation and divergence between them. We also examined whether the divergence of the two taxa has been associated with any notable bottlenecks or other changes in their historical population sizes (N_e).

Methods

Field sampling and sample preservation

Six sites from along a ~120 km transect crossing the San Gabriel mountains and the foothills to the south were selected for sampling (Figure 1.1). Based on the conclusions of Sullivan and Best (1997), two sites within the San Gabriel mountains were chosen that were expected to be exclusively AKR (Phelan, PHL, Applewhite Camp, AWC, both >1000m elevation), with one site exclusively DKR in SW Riverside County (Aguanga, AGA, at 640m elevation). Between them, four sites were within the area of sympatry proposed by Sullivan and Best (1997) (Lytle Creek, LTC at the southern edge

of the mountains, at 630m; two sites within the Box Springs Mountains, Box Springs high (BXSH, 740m elevation) and Box Springs low (BXSL, 400m elevation); and the UC Motte Rimrock Reserve, MRR, 580m). The two Box Springs Mountains sites were sampled because Sullivan and Best (1997) had proposed that the two species might be separated by elevation.

Sherman live traps were used to capture the rodents, and ~10 mg ear slices were cut from both ears of the targeted kangaroo rats. The samples were stored on dry ice in the field and then put into -80°C freezer. Trapping and sampling was approved under California Dept. of Fish and Wildlife Scientific Collecting Permit SC-11898 and UC Riverside Institutional Animal Care and Use Committee (IACUC) Animal Use Protocol 20140001 and 20180008.

Reference genome sequencing and assembly

A dead animal, identified as DKR, was salvaged at the Sycamore Canyon Park (SCP) and was used to assemble a reference genome. The whole animal was stored at -80°C. We defrosted and dissected the animal and sent the whole heart (~250 mg) to UC Davis genome center, where the high molecular weight DNA (HMW DNA) was extracted, and quality check was performed prior to Pacific BioSciences (PacBio) sequencing (Eid et al 2009). The sequencing was performed using 5 single-molecule real-time (SMRT) cells on Sequel II sequencer to generate High-Fidelity (HiFi) Long Reads. The adaptors used in the technique have a hairpin structure, which enables the circular sequencing of the long (10-15 kb) DNA molecules for about 4 rounds and Circular-Consensus-Sequencing (CCS) analysis that results in 99.9% accuracy. Also, the error rate of Pacbio is unbiased

(Delahaye & Nicolas 2021). The HiFi reads of PacBio were assembled using hifiasm (Cheng et al 2021) to obtain a *de novo* genome of this non-model organism, which was used as a reference for our population genomics studies. The basic statistics of this assembly were obtained using a python script (https://github.com/PacificBiosciences/pb-assembly/blob/master/scripts/get_asm_stats.py). To get the reading depth, the Pacbio reads were aligned to the assembled contigs using minimap2 v2.24 (Li 2018), and samtools v1.14 (Li et al 2009) was used to estimate the coverage. The completeness of the assembly was checked by BUSCO (Benchmarking Universal Single-Copy Orthologs) (Simão et al 2015) using the mammalia_odb10 lineage dataset, which looked for 9226 conserved single-copy orthologs in mammals.

Genome re-sequencing, alignment and SNP calling

Genomic DNA was extracted from one ear slice of each sampled animal using Qiagen DNeasy Blood & Tissue Kits (except for one DKR sample - see below). The whole genome re-sequencing was performed at the UC Davis genome center, where the samples were pooled with 23 samples for other projects (60 samples in total) and barcoded for library preparation and sequenced on 2 lanes of an Illumina NovaSeq S4 sequencer for 150bp paired end (PE) reads. For one of the AKR samples captured at AWC, we targeted at a higher coverage to get a better-quality genome representing this species, so 9 libraries were prepared for this sample while only one library for each of the other samples in this study. One DKR captured at AGA was sequenced differently. This sample was outsourced to HudsonAlpha Institute for Biotechnology (https://hudsonalpha.org/) for 10X genomics sequencing

(https://www.10xgenomics.com/). The short reads generated on a lane of an Illumina sequencer were used for the same population genomics analyses as the other samples.

The reads were trimmed by TrimGalore (https://github.com/FelixKrueger/TrimGalore) with the pass quality set as default (Phred score >=20) and adapters automatically detected. The forward and reverse reads were merged using PEAR v0.9.6 (Paired-end read merger; Zhang et al 2014) if the paired reads overlap each other and aligned to our DKR PacBio assembly using Burrows-Wheeler Aligner v 0.7.17 (BWA; Li & Durbin 2009). The alignment was processed using samtools v1.16 (Li et al 2009), and SNP calling was performed by bcftools mpileup function (Danecek et al 2021). Then, vcftools v 0.1.16-18 (Danecek et al 2011) was used to filter the variants, for which the minor allele frequency (MAF) was set as 0.05 and the maximum missing data as 20%. In addition, an AKR assembly was obtained by calling consensus sequence from the alignment using samtools and converted into fasta file using Seqtk (Li 2013).

Population structure

To detect the population structure and potential admixtures, we used Admixture v1.3.0 (Alexander, Novembre, & Lange 2009) to estimate the global ancestry of the genomes of the samples. To accommodate the reference assembly of a non-model organism, admixturePipeline v3.0 (Mussmann et al 2020, https://github.com/stevemussmann/admixturePipeline) was used to process the variants and run Admixture. The number of clusters (K) tested ranged from 2 to 10. CLUMPAK (Cluster Markov Packager Across K, Kopelman et al 2015,

http://clumpak"http://clumpak.tau.ac.il/) was used to process the output of the main pipeline and identify the optimal clustering. The best K was estimated using an ad hoc statistic, Delta K, which measures how much log probability changes between successive Ks (Evanno, Regnaut, & Goudet, 2005).

To visualize the genetic variation, we performed principal component analysis (PCA) using Plink v1.90b6.25 (Purcell et al 2007), which transforms the genetic variation into principal components to identify population clusters.

Genomic differentiation

To estimate the level of differentiation, we estimated genome wide Fst. The small samples from MRR and AGA were combined based on their similarity in the PCA test results to avoid bias due to small sample sizes and increase statistical power. Fst or fixation index (Wright 1949) measures the excess of homozygosity among populations due to differentiation (the Wahlund effect). Weir and Cockerham (1984) developed a modified method to estimate Fst that control for the sample sizes with multiple loci and alleles. The Weir & Cockerham weighted Fst values between pairwise populations were estimated using vcftools v 0.1.16-18.

Divergence time

To estimate the divergence time between AKR and DKR, we created maximum likelihood phylogenic trees from four-fold degenerate sites in the genomes of the kangaroo rat species that are available and estimated the branch length in PAUP (Swofford 2002). The assemblies of the Stephens' kangaroo rat (SKR, *D. stephensi*) (GCA_004024685.1), the Merriam's kangaroo rat (MKR, *D. merriami*)

(GCA_024711535.1), the Banner-tailed kangaroo rat (BKR, *D. spectabilis*) (GCF_019054845.1), the Ord's kangaroo rat (OKR, *D. ordii*) (GCA_000151885.2), and the pacific (little) pocket mouse (PPM, *Perognathus longimembris pacificus*) (GCA_004363475.1) were downloaded from NCBI. Single copy orthologs from the assemblies were retrieved using BUSCO and the dataset mammalia_odb10. To get accurate codon-by codon alignments, amino acid sequences were aligned using clustalw (Sievers et al 2011) and used as reference for aligning nucleotide sequences using pal2nal (Suyama et al 2006). The nucleotides at the four-fold degenerate sites were extracted and concatenated using a custom R script (Suppl. Material).

To get the initial parameters for PAUP, the best model and rates of nucleotide substitution were estimated by ModelTest-NG (Darriba et al 2020), and Akaike Information Criterion (AIC) tests were used to compare the models. In PAUP, the trees were generated using heuristic search for random addition sequences and running 5 repetitions of TBR (tree bisection and reconnection) branch-swapping. The significance of the nodes was tested using 100 bootstraps. The tree was visualized using Evolview (Subramanian et al 2019). The branch lengths in substitution rate (l) were estimated in PAUP. The generation time (g) was set as 1.7 years, which was estimated for the banner-tailed kangaroo rat (Swanson 2001). The mutation rate (μ) was set as 5×10^{-9} per bp per generation, the middle of the range of $3.5\text{-}6.5\times10^{-9}$ estimated for mice (Uchimura et al 2015; Lindsay et al 2019). Given the assumed neutrality of those sites, time since divergence (t) should have a linear relationship with the branch length, which can be calculated using $t= l/(\mu/g)$.

Demographic history

The AKR sample from AWC that was re-sequenced at a higher coverage and the DKR sample from TMC that was sequenced using 10X Genomic technology were used to get estimates of historical effective populations sizes (N_e) of each taxon using pairwise sequential Markovian coalescent (PSMC) model (Li & Durbin 2011). The PSMC method uses the heterogeneity in heterozygosity (i.e., SNPs) along the genome to define recombinant blocks. The coalescence time of the blocks is estimated by the divergence, and fewer regions coalescing at a certain time in the history implies a larger population size at that time. The parameters are set as -N80 -t15 -r1.33 -b -p "2*4+18*2+4+6" which were used for *Rattus norvegicus* in Deinum et al (2015). The number of iterations (N) was increased to 80 from the default 25 to encourage the convergence of history inferences, and the number of free parameters (p) was reduced from the default (4+25*2+4+6) to prevent overfitting (Deinum et al 2015). The initial mutation/recombination rate ratio *r* was set at 1.33 based on estimates from laboratory rats and mice (Jensen-Seaman et al 2004; Deinum et al 2015). The other parameters were set as default. For estimating N_e, a generation time (g) of 1.7 yrs and mutation rate (μ) of 5×10^{-9}) were used (see above).

Genetic variance

Pixy v1.2.7 (Korunes & Samuk 2021) was used to analyze the unfiltered vcf files and estimate the genome wide nucleotide diversity (π) in each population. This software uses a modified algorithm based on equation (1) to estimate π avoiding the downward bias caused by missing data, where *L* is the length of the reference genomic sequence to

which at least one sample aligned, and n_i is the count of haplotypes called in all samples at position i, and c_{i0} and c_{i1} are the counts of two alternative nucleotides at position i.

$$\pi = \frac{\sum_{i=1}^{L} c_{i0} c_{i1}}{\sum_{i=1}^{L} \binom{n_i}{2}} \quad (1)$$

The runs of homozygosity (ROHs) were estimated for all samples except for the DKR sequenced using 10X Genomics, which uses linked reads that could affect the estimation. Since ROH are determined by the local mating patterns, we tested whether the population structure was different in the two taxa, by comparing the individual mean ROH lengths using a t-test.

Results

Reference genome assembly

The size of the DKR reference genome Pacbio assembly was 3.398 Gb, with 3569 contigs and a N50 of 10.88M bp, with an average coverage of 23.589. BUSCO retrieved 8778 (95.1%) of the genes in the database mammalia_odb10, including 160 (1.7%) duplicated genes and 95 (1.0%) fragmented genes, while 353 (3.9%) genes were missing. This individual was from SCP (Figure 1.1).

Species delineation

The results from the Admixture analysis showed that AKR and DKR are genetically distinct with no evidence of gene flow (Figure 1.2, K=2). This division was the most strongly supported partitioning of the data, with Delta K = 1.15×10^{10} for K=2, while all other transitions ranged from 0.26 to 1.42. For K>2, the AKR/DKR partition was maintained with no evidence that samples from any site included both taxa (i.e., no evidence of sympatry) and it remained the case that there was no evidence of

hybridization between the two taxa, suggesting their long-term isolation (Figure 1.2). Based on these results, we determined that the populations on both the north (PHL), central (AWC), and south (LTC) sides of the San Gabriel Mountains were AKR, and the populations further south (BXSH/L, MRR, AGA) were DKR.

The PCA results were consistent with the finding of Admixture (Figure 1.3), separating AKR and DKR samples on the axis of PC1 (the first principal component), which explained 30.1% of the variance. The Fst values further supported the species-level divergence between AKR and DKR with ($0.2129 \leq Fst \leq 0.2580$), while the values between the populations within the two species were close to an order of magnitude smaller, ranging from 0.0233 to 0.0396 (Table 1.1).

Table 1.1. Genome-wide Fst between AKR and DKR populations. AKR sites are underlined.

Fst	PHL	AWC	LTC	BXSH	BXSL	MRR+AGA
PHL	-					
AWC	0.0238	-				
LTC	0.0322	0.0395	-			
BXSH	0.2285	0.2343	0.2129	-		
BXSL	0.2510	0.2580	0.2383	0.0396	-	
MRR+AGA	0.2354	0.2426	0.2223	0.0233	0.0362	-

Historical estimates

The divergence between DKR and AKR was estimated at roughly 0.5~1 mya (mean 0.69 mya; Figure 1.4), and from that time until relatively recently AKR appears to have had a consistently higher effective populations sizes (N_e) than DKR, although by about ten thousand years ago, both species had a similar N_e of about 50,000 (Figure 1.5). Before this time, the effective size of DKR appears to have decreased rapidly around the time of speciation and remained low in the range of 20,000-40,000 until about twenty thousand

years ago when it began to increase. In contrast, in the same period, the estimates indicate that the N_e of AKR oscillated around 100,000 until it began a long-term decline around fifty thousand years ago.

Discussion

Genomic data validates the decision of Sullivan and Best (1997) to divide PacKR into two distinct species, AKR and DKR, a distinction that Stock (1974) had also thought possible based on their karyotype difference. We established that these two taxa appear to have diverged approximately 0.5~1 mya and have accumulated significant genetic differentiation (Fst>0.2) with no evidence of hybridization, despite there being no obvious geological barrier between them. The two species have similar levels of nucleotide diversity, even though for a long period following speciation DKR had a much smaller effective population size.

A divergence time of roughly 0.5-1mya is intermediate between the two previous estimates. Stock (1974) suggested that DKR and AKR (at that time still considered subspecies) were separated very recently, after the Wisconsin Pluvial (~13,000 yrs ago). In contrast, Sullivan and Best (1997) proposed a much earlier separation perhaps due to climate change causing marine transgressions or from the San Gorgonio Pass becoming a barrier about 2.5 mya during early-Pleistocene or perhaps even earlier. Under this scenario, they suggested that the ancestors of DKR were isolated in Baja California and later spread north to the Los Angeles Basin, becoming sympatric with AKR. Our genomic estimate is between these two time periods and does not support either, instead

indicating a role of some geological or climatic events in the mid- to late Pleistocene event causing their separation.

Another pair of kangaroo rat sister species, the Panamint kangaroo rat (PKR, *D. panamintinus*) and Stephens' kangaroo rat (SKR, *D. stephensi*) are broadly sympatric with AKR and DKR, respectively. This clade separated from AKR and DKR about 2.5 mya (estimates range from ~2-3 mya; Figure 1.4), and it is believed that PKR and SKR separated about 1.5 to 2 mya, as a result of the uplift of the San Gabriel and San Bernardino mountains (Metcalf, Nunney, & Hyman 2001), since neither species is found at higher elevations. In contrast, we found that AKR is found at higher elevations from north to south across the San Gabriel mountains, so that these mountains were not the geological barrier that created the AKR/DKR split. It is possible that the Santa Ana River served and could still serve as a geological barrier separating AKR and DKR today. Samples from the north side of Santa Ana River are needed to test the possibility that AKR (and not DKR) is found there; however, the habitat shift at the base of the mountains seems more likely to be a significant barrier.

A similar example is provided by the Brazilian rodent, *Oecomys catherinae*, which shows a karyotype polymorphism in geographically separated populations in the Amazonian biome (2N=62) and in the Atlantic Forest (2N=60) (Malcher et al 2017). Although the two forms exhibit very minor morphological and genetic (cytochrome b) differences, the authors proposed that they may represent two cryptic species resulting from chromosomal speciation. However, as in the AKR/DKR example, it is difficult to

detangle the effects of karyotype difference, allopatry and environmental factors in their divergence.

Based on morphological measurements, Sullivan and Best (1997) had suggested that the species were sympatric over a large area south of the mountains, but with AKR expected at higher elevations. To test this hypothesis, we sampled a high and low elevation site in the Box Springs Mountains (740m vs 400m); however, only DKR was found at both sites even though the high elevation site was higher than the AKR site of LTC (630m).

Consistent with our finding of an apparent lack of sympatry was the complete absence of any indication of genetic exchange between the two species. It is not known if the lack of introgression is a result of allopatry alone or, if there are some areas of sympatry, is due to mating isolation. The divergence time 0.5~1 mya for AKR and DKR is a short period to create reproductive isolation between mammal species. Fitzpatrick (2004) noted that the typical time for mammals to develop hybrid inviability is 2-4 M yrs; however, the fitness loss due to chromosomal incompatibility can promote reproductive isolation and speed up speciation. Franchini et al (2008) used microsatellite markers to estimate the level of reproductive isolation between two chromosomal races (2N = 22 & 24) of the house mouse, *Mus musculus domesticus*, and revealed no gene flow between them despite the absence of geological or ecological barriers. These races are believed to have originated about 10,000–3000 years ago (Cucchi et al 2005).

It is parsimonious to conclude that the karyotype 2N=60 found in DKR is ancestral, since *D. venustus* and *D. elephantinus*, close relatives of AKR and DKR (Alexander &

Riddle 2005), both have 2N=60. In contrast, Stock (1974) suggested that 2N=62 was more probably ancestral because centric fusion was more common than fission in the intraspecific variants in kangaroo rats; however, this hypothesis requires 2N=60 to have become fixed at least twice, but certainly cannot be ruled out.

Kangaroo rats typically have limited dispersal. For example, the individual dispersal distance of SKR (*D. stephensi*), a species commonly sympatric with DKR in the study area, is generally low (<40m), although occasionally individuals disperse much farther (>400m) (Price, Kelly, & Goldingay 1994). This pattern suggests that habitat discontinuity may severely limit movement, and this may be the case in the more mountainous habitat of the AKR. The distance separating the three AKR sample sites was much less than the distance between DKR sites (Figure 1.1) and yet the genetic differentiation among sites was essentially identical in the two species (Table 1.1), with the AKR site LTC showing a marked separation along PC2 (Figure 1.3). Supporting the view that dispersal is more limited in the mountains was the finding that the runs of homozygosity (ROH) were significantly longer in AKR. Long ROHs are long stretches of homozygous loci indicating IBD (identical by descent) genomic segments that are from a recent common ancestor and are an indicator of the level of recent inbreeding (Ceballos et al 2018; Brüniche-Olsen et al 2018). Although the larger sizes of ROHs in AKR indicates more inbreeding at these sites, the levels of nucleotide diversity (π) are the same in both species, consistent with AKR having a lower effective neighborhood size (N_n), which is a function of spatial structure caused by dispersal patterns, but this

generally has a minimal effect on the overall effective population size N_e (Nunney 2016), which we found to be generally higher in AKR than DKR in the past (Figure 1.5).

DKR had a long period of a relatively low historical N_e, which could be a result of geographic isolation in reduced and fragmented of this species in Baja California during middle to late Pleistocene due to climatic fluctuation, consistent with the isolated distribution of other small mammals with limited dispersal abilities in the Mexican fauna (Ceballos, Arroyo-Cabrales, & Ponce 2010). Also, three species, *D. peninsularis*, *D. paralius*, and *D. antiquarius*, in Baja California were later identifies as subspecies of DKR (Stock 1974; Best 1983), supporting its subdivision into isolated groups in the past. However, the genomic diversity and the estimates of N_e for DKR give no indication of past severe bottlenecking.

Conclusion

In summary, we have shown that genomic analysis shows clearly that AKR and DKR are two genetically isolated species, with no evidence of genetic introgression. In doing so, we showed that genomic data can overcome the problem of similar morphology in taxonomic studies, revealing, in this case, the absence of interspecific gene flow, differences in population structure and historical N_e. The rapidly increasing abundance of genomic data in non-model organisms can enrich our knowledge about their evolutionary history and help predict the trends of population changes under climate change.

Chapter 1 Figures

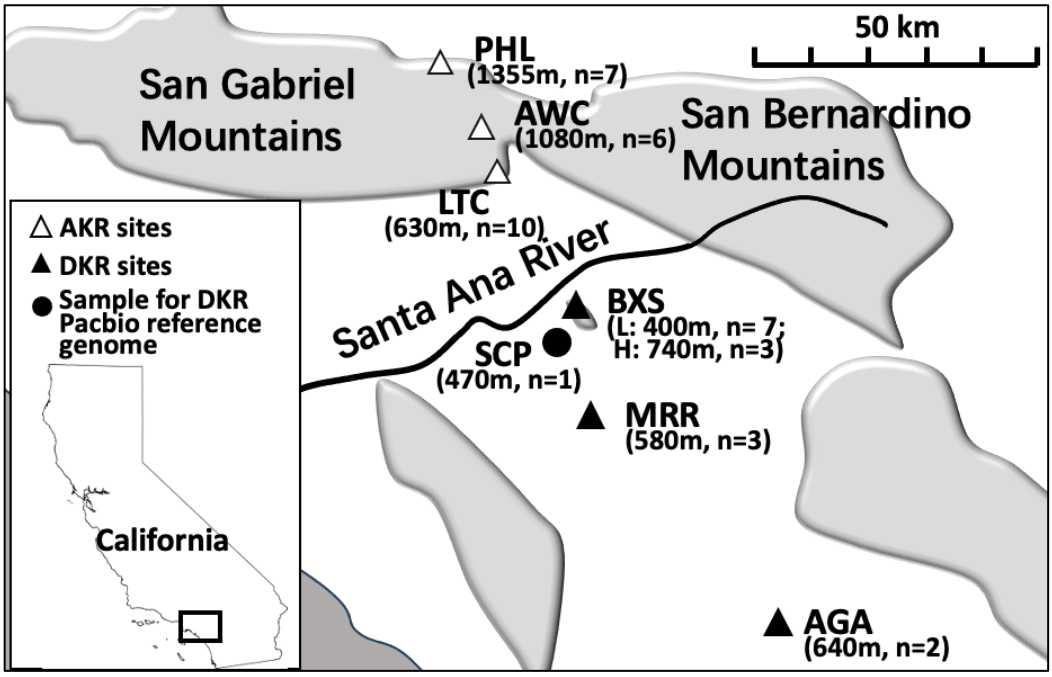

Figure 1.1. Sampling sites and the location of the salvaged animal. The elevations and samples sizes (n) are shown in parentheses below the sample site abbreviations. For full names of sample sites, see Methods. The shading defines major mountain ranges, and black line defines Santa Ana River.

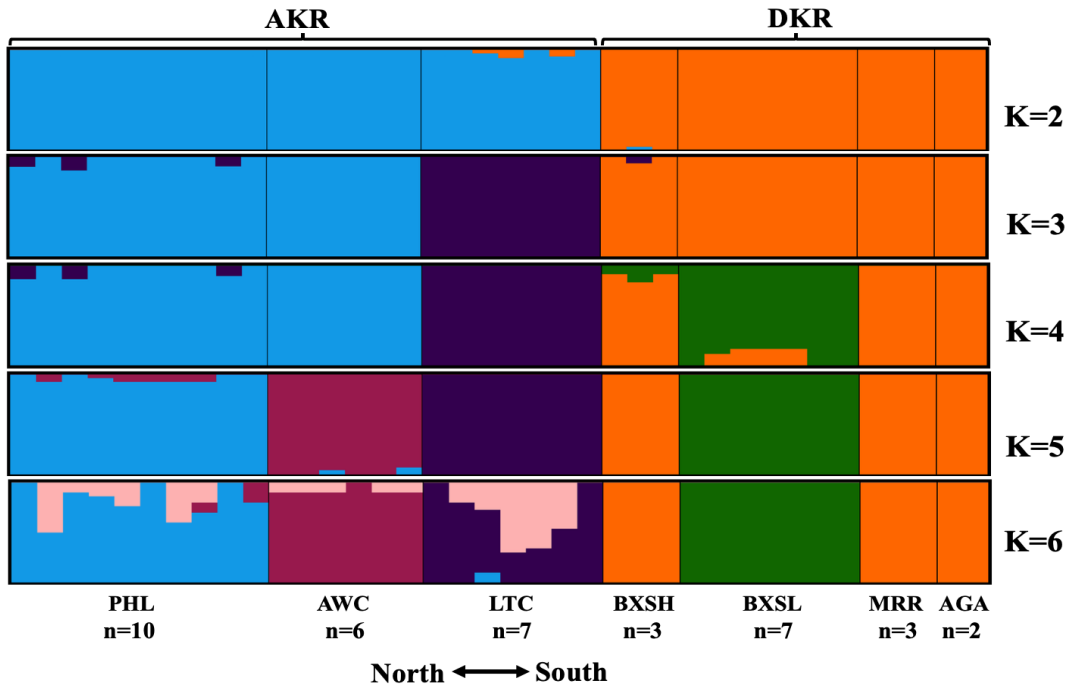

Figure 1.2. Major genetic clusters defined by the genomes of individuals sampled along the North-South transect. K defines the number of clusters fitted, with K=2 separating DKR and AKR. Sample sites are shown at the bottom, with genetic composition of each of the n individuals sampled at a site defined by a vertical bar.

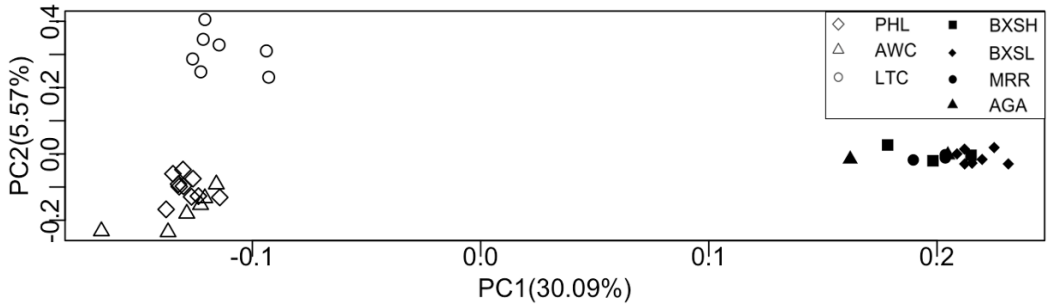

Figure 1.3. Principal component analysis (PCA) on genome-wide SNPs in samples from all sites. The AKR samples (unfilled symbols) are separated from the DKR samples (filled symbols) by PC1.

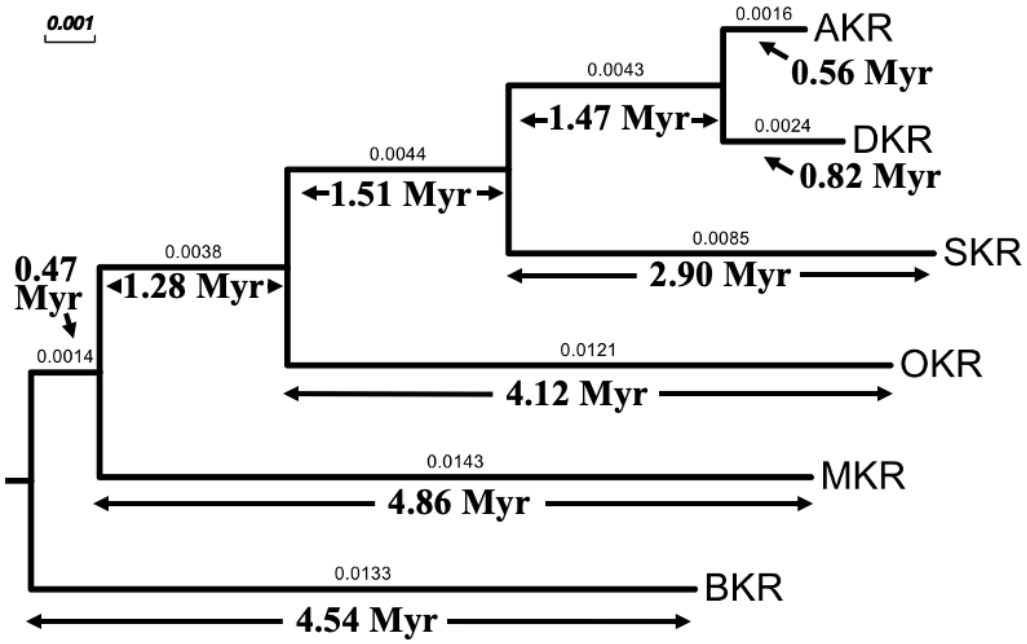

Figure 1.4. ML phylogenic of six kangaroo rat species (Genus *Dipodomys*). PPM was used as an outgroup to root the tree. The branch lengths scaled by the rate of substitutions are shown above the branches with the estimated divergence times below. All nodes have 100% bootstrap support.

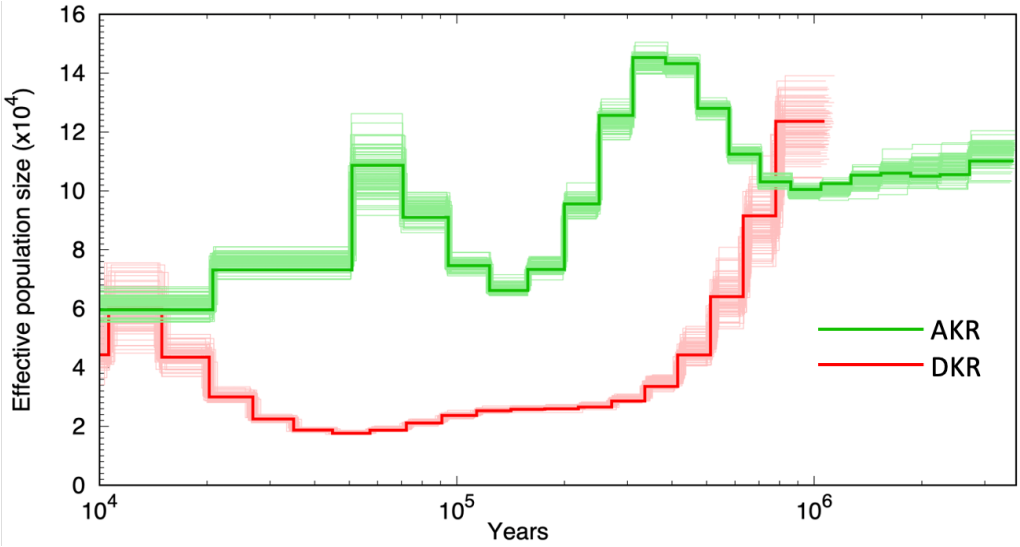

Figure 1.5. Historical effective population sizes of AKR and DKR. The green and red lines represent the changes in historical N_e of AKR and DKR, respectively.

Chapter 2. Genomic evidence of adaptation linked to elevation in two previously conspecific kangaroo rats, the agile kangaroo rat (*D. agilis*) and the Dulzura kangaroo rat (*D. simulans*)

Abstract

The previously conspecific sister species, the agile kangaroo rat (AKR, *Dipodomys agilis*) and the Dulzura kangaroo rat (DKR, *D. simulans*) show a north-south divide in Southern California, with AKR often occupying higher mountainous areas. Using genomic data, we investigated adaptation and its possible link to elevation both between and within the species. We sampled from three AKR sites (elevations from 630m to 1355m) across the San Gabriel Mountains and four DRK sites (elevations from 400m to 740m) from south of the mountains in western Riverside County. More signatures of selective sweeps were found in AKR than DKR, and several candidate genes in AKR were associated with thermoregulation and hypoxic stress, consistent with adaptation to higher elevations. Within each species, genomic differentiation was more closely linked to elevation than geographic distances. Grouping the intraspecific sites by elevations (high vs. low) we identified signatures of selective sweeps, finding more in the high elevation group of AKR where candidate genes were also associated with cold and hypoxic stress, while in DKR, the low elevation group showed more selection sweeps, but the candidate genes showed no obvious pattern with respect to biological processes.

Introduction

Local adaptation refers to the evolutionary changes that increases the fitness of populations in their local environments. Understanding such adaptative change and its

genetic basis is important in the conservation of species and of locally adapted populations. The loss of locally adapted populations can cause genetic diversity loss in a species that may affect its ability to adapt to environmental change, and failure to understand such adaptation can result in the failure of re-introduction efforts due to phenotype-environment mismatches (Garcia de Leaniz 2007; Exposito-Alonso 2023). Furthermore, understanding the genetic basis of local adaptation can facilitate the prediction of distribution ranges and population sizes under changing climates (Atkins & Travis 2010) and human disturbance (Vincze 2016).

In the past, the tools available limited studies of the genetic basis of local adaption to observable phenotypic polymorphisms with a genetically simple basis (Savolainen, Lascoux, & Merilä 2013). Examples include sickle cell trait and malaria (Allison 1954) and lactase persistence in pastoralist populations (Enattah et al 2002; Tishkoff et al 2007) in humans, armour plates number and pigmentation in three-spine sticklebacks (Cresko et al 2004; Miller et al 2007), selection on color patterns in the snail *Cepaea* (Jones 1977), melanism in peppered moth (van't Hof et al 2011) and the rock pocket mice (Nachman, Hoekstra, & D'Agostino 2003; Hoekstra, Drumm, & Nachman 2004; van't Hof et al 2011), and coloration in deer mice on different soils (Linnen et al 2009). Genomic data have enabled genome scans in non-model organisms for candidate genes under selection in their ecological environments (Tiffin & Ross-Ibarra 2014) and reveal the genomic changes underlying the polygenic complex traits and gene networks. E.g., studies have investigated genomic changes underlying the remarkable renal function in desert dwelling kangaroo rats (Marra et al 2012; Marra, Romero, & DeWoody 2014) and

genomic adaptation to arid environments in camel (Wu et al 2014), goats, and sheep (Kim et al 2016).

Understanding local adaptation is important to protect the ecosystems where current global warming is making them generally hotter and drier (Chen et al 2011; Kousari & Asadi Zarch 2011), and the temperature increase has caused plants and animals to migrate to higher latitudes and elevations (Li, Kräuchi, & Gao 2006). As a major determinant of local adaptation, elevation could affect temperature, humanity, oxygen availability, UV lights, etc., and it is among strongest agents of selection in humans (Bigham 2016). Studies have identified Epas1, Egln1 and other candidate genes in the HIF (hypoxia inducible factors) transcriptional pathway in response to hypoxia in Tibetan populations (Beall et al 2010; Bigham et al 2010; Xiang et al 2013; Lin et al 2023). Prkaa1 genes has been identified to affect birth weight and uterine artery diameter in Andean women to overcome altitude-associated intrauterine growth restriction (IUGR) (Bigham et al 2014). A variant of endothelial nitric oxide synthase (Nos3) was identified as a candidate gene in Andean population that increases nitric oxide in blood as an adaptation to hypoxia (Wang et al 2010). Genetic basis of adaptation to high elevations has also been characterized in deer mice (Storz et al 2007; Wilsterman et al 2023), and genomic signatures of adaptation had been found in American pika at alpine areas (Sjodin & Russello 2022).

The kangaroo rats (family *Dipodomys*) are a group that includes more than twenty recognized species with very little morphological variation (Alexander & Riddle 2005). Some species, e.g., the Merriam's kangaroo rat (*D. merriami*), have a large geographical

range, subdivided into many subspecies (Lidicker 1960), but in Southern California in there is a local diversification of kangaroo rat species presumably due to geological features fragmenting habitats near the San Andreas fault, with some species having a very limited distribution, e.g., Stephen's kangaroo rat (SKR, *D. stephensi*; Metcalf et al. 2001).

The agile kangaroo rat (AKR, *D. agilis*) and the Dulzura kangaroo rat (DKR, *D. simulans*) are two species in Southern California that were considered as conspecific subspecies with different karyotypes (2N=62 in AKR and 2N=60 in DKR) (Csuti 1971; Stock 1974; Best et al 1986) but were later re-classified as separate species (Sullivan and Best 1997). Recent genomic analyses confirmed their species-level status with an estimated divergence time as 0.5~1 Mya (Chapter 1).

Sullivan and Best (1997) hypothesized that AKR and DKR differ in their elevational range, with AKR distributed roughly above 800m and DKR under 800m, a difference supported by the occurrence of AKR throughout the San Gabriel Mountains (Chapter 1). This upland distribution is unusual in kangaroo rats that typically occur in arid and semi-arid environments and raises the possibility that AKR has evolved unique adaptation to high elevations. If such adaptation is an important component of fitness, then it is also possible that local intra-specific differences in elevation in both AKR and DKR could drive local adaptation.

To investigate whether geographic features such as elevation have affected the genomic differentiation in AKR and DKR, we sampled three AKR sites across an about ~ 23 km northwest to southeast transect across San Gabriel Mountains and four DKR sites across an about ~67 transect in the same direction across Western Riverside County for

genomic re-sequencing. To test the hypothesis of interspecific adaptation and intraspecific local adaptation to different elevations and habitats, we scanned their genomes for selective sweeps, using the criterion that newly arising advantageous mutations are expected to increase genetic differentiation (as measured by Fst) and reduce variation in the surrounding region (Booker, Jackson, & Keightley 2017).

Methods

Sampling sites and field collection

As mentioned in Chapter 1, We sampled AKR from three sites in the San Gabriel Mountains, Phelan (PHL, elevation=1355m) on the northern side, Applewhite Camp (AWC, elevation=1080m) in the center, and Lytle Creek (LTC, elevation=630m) on the southern side, and DKR from four sites, the top and the foot of the Box Spring Mountains (BXSH, elevation=740m; BXSL, elevation=400m), the UC Motte Rimrock Reserve (MRR, elevation=580m), and Aguanga (AGA, elevation=640m) (Figure 2.1). The trapping and sample preservation methods and permitting see Chapter 1.

Reference genome, genome re-sequencing, and data processing

As mentioned in Chapter 1, we sequenced and assembled a DKR genome using Pacbio (Eid et al 2009) and assembled using hifiasm (Cheng et al 2021). One DKR sampled from AGA was sequenced using 10X Genomics (https://www.10xgenomics.com/), and the other 37 samples were re-sequenced using Illumina as 150bp paired end (PE) reads.

The methods used for trimming, merging of paired end reads, alignments, SNP calling, and variant filtering are the same as Chapter 1 (Li et al 2009; Li & Durbin 2009; Danecek et al 2011; Danecek et al 2021; Zhang et al 2014).

Estimation of intraspecific genomic differentiation

The intraspecific genomic variation was estimated by principal component analysis (PCA) using Plink v1.90b6.25 (Purcell et al 2007) and visualized in R studio (Figure 2.2). Also, Weir & Cockerham weighted Fst (Weir and Cockerham 1984) between pairwise sites within each species were estimated using vcftools v 0.1.16-18 (MRR and ARR are combined based on the PCA test results).

Examination of genomic variation

The genome wide nucleotide diversity (π) within in each site was estimated using Pixy v1.2.7 (Korunes & Samuk 2021) using unfiltered vcf files. The runs of homozygosity (ROHs) were estimated using bcftools for each re-sequenced individual excluding the DKR sequenced using 10X Genomics. To test whether mean length of ROHs is associated with elevation, we used an ANCOVA test on all samples of AKR and DKR pooled together using two variables, species, and elevation.

Selective sweeps

To identify regions that were potentially subject to positive selection in only one of the two species, we adopted the approach of Nanaei et al (2023), and used two criteria evaluated using 50kb genomic windows (and a 25kb step size): 1) high Fst, indicating an inter-specific divergence in allele frequencies; and 2) high proportional difference in π, delta π (defined as $\Delta(\log\pi) = \log_2(\pi_{control}/\pi_{adapted})$; Nanaei et al 2023), indicating a loss of

variation in one of the species. A high Fst and high $\Delta\pi$ is potentially indicative of a selective sweep in the species with the lowest π. Windows with ≤ 3 variable sites were removed to avoid stochastic issues. Windows within both the top 1% of Fst and top 1% of the proportional difference in π were selected as regions of interest, with overlapping windows merged into a single region.

To identify candidate genes within the regions of interest, the dataset refseq_select_rna (https://www.ncbi.nlm.nih.gov/refseq/refseq_select/) was used. This dataset contains a listing of well supported transcripts from human and mouse. The sequence in each region of interest was individually blasted against the dataset using ncbi-blast 2.13.0+ (Camacho et al 2009). We kept genes with E-values smaller than 10^{-15} for further examination. To check whether hits were orthologous functional genes, the repetitive DNA was annotated and masked by RepeatModeler 2.0.3 and RepeatMasker 4.1.4 (Smit, Hubley & Green 2015), and we filtered out the genes that were masked as repetitive, fragmented as a small piece without the rest of the gene sequence nearby, with obvious frameshifts and/or other signatures of degeneration. We calculated the average Fst and $\Delta(\log\pi)$ of each candidate gene by averaging the values of all 50kb regions of interest that the gene spanned.

The segregating sites in all the regions of interest were pooled together, and the subgroups with top 5%, 10%, and 15% Fst were examined for their relative position to candidate genes.

Detection of selective sweeps

To test the hypothesis that local adaptation has caused increased intraspecific genomic differentiation. we grouped the samples from PHL and AWC to represent the high elevation group of AKR (AKRhigh), with LTC representing a low elevation population (AKRlow), based on the elevations and the differentiation levels showed by PCA and Fst analyses,. Similarly, BXSH, MRR and TMC defined the high elevation group of DKR (DKRhigh), with BXSL representing a low elevation group (DKRlow). We used the same methods mentioned above to look for genomic regions, candidate genes and SNPs potentially under selection.

In addition, because the intraspecific genome-wide Fst is low in both species (0.02 - 0.04), a fixed or near fixed site could be subject to selection or be close to the selected site. Where there were fixed or nearly fixed difference between the high and low elevation groups (Fst $\geq$ 0.9), we selected the 50kb region around the site (25k bp upstream and 25kp downstream) and calculated the average Fst and π of all variants in the region, and also we searched for genes in these regions using the approaches described above.

Results

The locations of the sampling sites and sample sizes are shown in Figure 2.1. The mean elevation of the AKR sites was 1022m (ranging from 630m to 1355m) and those of DKR was 590m (ranging from 400m to 740m. Because there are only 2 samples from AGA, we merged this site with MRR based on PCA analysis. All sites in both species showed similar levels of nucleotide diversity (π). In AKR, the πs for PHL, AWC, and

LTC are 0.00222, 0.00233, and 0.00235. In DKR, the πs for BXSH, BXSL, and MRR+AGA are 0.00236, 0.00220, and 0.00232. However, the elevation of sites appeared to reduce neighborhood size, as measured by an increase in the inbreeding measure of runs of homozygosity (ROHs). A difference between AKR and DKR in ROHs was previously noted (Chapter 1); however, including sample locations showed that this effect was driven by elevation ($F(df=1) = 23.62$, $p = 2.6\times10^{-5}$) while the species effect was no longer significant ($F(df=1) = 3.61$, $p = 0.066$), indicating that elevation is a major factor affecting population structure (Figure 2.2).

Interspecific genomic differentiation and adaptation

Regions with the signature of a selective sweeps were identified by scanning for 50kb windows that were in the top 1% of two measures: a local proportional decrease in variation (π) in the species experiencing the sweep combined with an increase in Fst. After filtering out windows with ≤3 variant sites, 101,863 remained. AKR, with 120 windows spanning 80 regions of interests (each ≥ 50kb), showed significantly more signatures of selective sweeps than DKR, with 29 windows spanning 26 regions of interests (for difference in window number, $\chi^2=54.40$, $df = 1$, $p = 1.6\times10^{-13}$). In AKR there were also significantly more candidate genes within those regions, with 45 candidate genes identified in 34 of the regions, while in DKR only one candidate gene was identified ($\chi^2=12.42$, $df = 1$, $p = 0.00042$).

Of the 26 regions identified in DKR, all but one consisted of only repetitive DNA, mostly short tandem repeats, while none of the regions in AKR contained only repetitive DNA. For the regions of interest in DKR, where the π values in DKR should be reduced,

those of AKR (mean π_{AKR}=0.00052) were also consistently well below the mean π in all windows after filtering (mean π_{AKRall}=0.00260, t = 24.058, df = 28.152, p-value < 2.2×10^{-16}), while for the regions identified in AKR, the π values in DKR were not reduced from the genome wide mean (mean π_{DKR}=0.00278, mean π_{DKRall}=0.00280. t = 0.32208, df = 119.99, p-value = 0.374), suggesting the signals of selection in DKR were likely driven by low π in both groups in repetitive DNA.

Two contigs, ptg000074l and ptg000045l, showed increased numbers of 50k windows associated with selective sweeps (Figure 2.3). There are 7 regions of interests that are close to each other and span the position 12725001-14400000 on ptg000074l, and 3 candidate genes, Chsy3, Minar2, and Adamts19, are in those regions. On ptg000045l, there are 7 regions of interests spanning position 100001-700000, 1025001-1075000, and 1975001-2100000, and 3 candidate genes, Kcnmb2, Pik3ca, and Zmat3, are in the first range, while none is in the other 2 ranges.

Among the regions of interest identified, within the top 15% of differentiated SNPs (based on Fst) there was no significant association between the level of differentiation and the presence of coding genes: in the 84 SNPs with top 5% Fst in AKR, 42 (50%) were in the regions with a candidate gene, and the numbers for the next two 5% groups are 39 (46.4%) and 35 (41.7%) (χ^2=1.182, df = 2, p = 0.5538). In all the SNPs with top 15% Fst that are in regions with candidate genes, 95 only fall into the intron regions, 14 only fall into downstream regions, and 7 are in the upstream regions of candidate genes, and 1 is causing a synonymous change. Of the upstream SNPs 5 were also in the introns or downstream regions of other candidate genes (Suppl. Material). We identified only one

SNP (Fst=0.817, group=top 10%-15%) that resulted in an amino acid alteration, Arg404His, in Slc39a4. This is a G/A substitution at the 2nd nucleotide of the codon, and the frequency of Arg(G) allele in AKR is 0.947 and that in DKR is 0.125. Because most of the SNPs with high Fst are not causing nonsynonymous substitutions, the adaptive changes could be mostly regulatory.

The 45 candidate genes identified in AKR showed no significantly enriched GO term categories. Because coldness and hypoxia were considered two major selective pressures linked to elevation, we looked for candidate genes involved in relevant biological process and identified 4 in thermoregulation (Rptor, Fkbp8, Sugct and Ncor1) and 2 in response to hypoxia (Arid1b and Cpsf1).

Intraspecific genomic differentiation linked to elevation

The PCA analysis of the three AKR sites showed that the low elevation site LTC is genetically differentiated from the two high elevation sites (AWC, PHL) by PC1 (Figure 2.4a). Similarly, among the DKR samples PC1 separated the low elevation site BXSL from the high elevation sites (BXSH, MRR, AGA) (Figure 2.4b). In both species, PC2 separated the high elevation sites; however, given the minimal separation of MRR and AGA samples of DKR, these samples were combined to increase the sample size for estimating Fst. The Fst results reinforced the PCA analysis. In AKR, Fst between the two high elevation sites PHL and AWC was 0.0238, while the Fst between those two sites and LTC was larger (0.0322 and 0.0395, respectively). Similarly, in DKR, the Fst between the two high elevation samples (BXSH and MRR+AGA) was 0.0233, while the

Fst between them and the low elevation BXSL was larger (0.0396 and 0.0362, respectively).

Selective sweeps in AKR

There were 105,943 50kb windows left after filtering out windows with ≤3 variant sites. In AKR, the high elevation populations (AKRhigh = PHL+AWC) showed significantly more regions with the signature of selective sweeps (117 windows spanning 77 regions of interest) than the low elevation population (AKRlow = site LTC) with only 27 windows spanning 21 regions (χ^2=55.04, df =1, p =1.2×10^{-13}). Also, there were significantly more candidate genes identified in a region of interest in AKRhigh (37 genes identified in 32 regions) than in AKRlow (2 genes identified in 2 regions) (χ^2=6.1264, df =1, p-value=0.01332). Furthermore, in AKRlow the regions of interest were generally characterized by repetitive DNA. There are 17 out of 21 regions of interest in AKRlow that consist of only repetitive DNA, while the number in AKRhigh was only 6 out of 77 (χ^2= 45.181, df = 1, p-value =1.797×10^{-11}). Like in the regions of interest in DKR in interspecific comparison that consisted of repetitive DNA, the mean πs in the regions of interests in AKRlow had a very low value relative to the overall mean of all windows analyzed (π_{AKRL}= 0.00069 vs $\pi_{AKRLall}$ =0.00259), consistent with the lowered diversity following a selective sweep. However, given such a selective sweep, the mean π in the same regions in AKRhigh should not be reduced, but it was found to be significantly lower than the overall mean (π_{AKRH}=0.00124 vs. $\pi_{AKRHall}$ = 0.00247; t = 5.78, df = 26.022, p = 2.2×10^{-06}). In contrast, the regions identified as candidates for a selective sweep in AKRhigh showed the expected pattern of low genomic diversity in AKRhigh

(π_{AKRH}=0.00156) but no reduction in the same regions in AKRlow (π_{AKRL}= 0.00273) relative to $\pi_{AKRLall}$ (t = -1.44, df = 116.51, p = 0.924).

Of the 220 SNPs with top 5% of Fst values in the regions of interest, in AKRhigh, only 84 (38.18%) are located in a region with one or more candidate genes, and the numbers in the next two 5% groups are 71 (32.27%) and 87 (39.55%). In the 242 SNPs in this group of the largest 15% of the Fst values that could be associated with one or more candidate genes, none resulted in a non-synonymous substitution. Instead, all were located in the non-coding regions (59 upstream only, 129 intron only, 48 downstream only, and 4 in both upstream and intron regions) of those genes except for 2 that cause synonymous substitutions (Suppl. Material). As with the interspecific results, this pattern suggests that the adaptation is mostly due to regulation of gene expression.

The adaptation to higher elevation may play an important role in the genomic evolution of AKR. A candidate gene, Dmd, identified in AKRhigh vs. AKRlow were also apparent in the AKR vs. DKR comparison, which encodes essential muscle protein dystrophin. Two of the candidate genes in AKRhigh, Ndufb2 and Rock1, were suggested to be involved in thermal adaptation. Also, Rock1 and Braf were suggested to be involved in response to hypoxic stress. However, no significantly enriched GO term categories in the candidate genes were found in AKRhigh.

In AKRlow, 16 out of 39 SNPs within the top 5% of Fst values are located in the introns of a single gene, Ift88 suggesting that this locus is under selection, and the gene is involved in cilium biogenesis. Only one of the top SNPs is downstream to the other

candidate gene Anapc4, while the others are in regions of interest without a candidate gene (Suppl. Material).

There were no fixed differences found between AKRhigh and AKRlow and only two SNPs with Fst ≥ 0.9, and these SNPs were not associated with any gene, at least within the 50kb windows surrounding these sites.

Selective sweeps in DKR

The selective sweeps showed a reversed pattern between the high and low populations in DKR compared to AKR, with more identified in the low elevation genomes (DKRlow = site BXSL). There were 107,189 50kb windows left after filtering for those with >3 SNPs. In DKRlow, 69 windows had an Fst in the top 1% with a proportional difference in π in the top 1% and these windows concatenated to 47 "regions of interest". In DKRhigh (= sites BXSH, MRR+AGA) only 20 windows, spanning 17 regions, satisfied the criterion, a highly significant difference (χ^2= 25.90, df = 1, p = 3.6×10^{-07}). Furthermore, in DKRlow these regions of interest were more likely to be associated with coding genes with 25 genes spanning 22 out of 47 regions compared to 2 genes spanning 2 out of 17 regions, in DKRhigh (χ^2 =5.1317, df = 1, p = 0.02349).

As was the case in AKR vs DKR, and AKRhigh vs AKRlow, the regions of interest identified in the group showing fewer signatures of selective sweeps were dominated by regions of repetitive DNA. A majority of the DKRhigh regions consist of only repetitive DNA, with 10 out of 17 regions in DKRhigh compared to 5 out of 47 regions in DKRlow (χ^2=13.58, df = 1, p = 0.00023). Another repeated feature of this pattern was that in the regions of interest identified for DKRhigh, nucleotide diversity (π) was low (as expected)

in those regions for DKRhigh (0.00059 vs. mean 0.00280) but in addition in these same regions DKRlow was also significantly below the genome-wide mean (0.00100 vs mean 0.00263; t= 4.506, df = 19.005, p = 0.00012). This was not seen in the regions identified for DKRlow, with DKRlow $\pi = 0.00181$ (as expected given a selective sweep), but for DKRhigh there was no reduction from the mean, with $\pi = 0.00320$ (t = -2.3787, df = 68.098, p = 0.9899).

The 25 candidate genes associated with the regions of interest in DKRlow showed no significantly enriched GO term categories. In the 118 SNPs defining the top 5% of Fst values in the regions of interest, 70 (59.32%) are located close to or within the ranges of candidate genes, which was not notably different from the proportion in the next two 5% groups (47.46% and 57.63%, respectively). Only one of these SNPs was within a gene, causing a synonymous change, while the remainder were in the upstream, intron and downstream of the candidate genes. No amino acid changes were found (Suppl Material).

In the regions associated with a selective sweep in DKRhigh, only 2 of the 30 SNPs in the top 15% of Fst values were close to one of the 2 candidate genes associated with the regions of interest in DKRhigh. They were both upstream to Or2l13, an olfactory receptor gene.

There were 6 fixed SNPs between DKRhigh and DKRlow across the whole genome. One each was in an intron of Fhod3 and Yap1, although the average Fst in the extended 50k bp regions surrounding these two SNPs was relatively low compared to other regions of interests (Fst=0.228-0.460), being 0.0407 and 0.0902, respectively, and the Fst in the

extended 50k bp regions without a candidate gene range from 0.0940 to 0.1424 (Suppl Material).

Discussion

High altitude adaptation to has been found in many taxa, including humans (Bigham 2016; Witt & Huerta-Sánchez 2019), domesticated animals (Witt & Huerta-Sánchez 2019; Friedrich & Wiener 2020), and a wide array of small mammals (Robertson & Wilsterman 2020), and cold and hypoxia were identified as two major selection pressures. However, this adaptation is generally investigated over a large elevational range. In the present study of two very closely related species of kangaroo rat, AKR and DKR, we found evidence of intraspecific genetic differences associated with local changes in elevation of only 200-350 m in DKR and 450-750m in AKR. Henry & Russello (2013) used AFLP based genome scan and identified the loci underlying local adaptation in American pika (*Ochotona princeps*) across 350 -1300m elevational transects, suggesting this level of difference in elevation can possibly cause adaptive divergence.

There is no evidence of sympatry between AKR and DKR, and our results showed the division of their habitats is likely due to local adaptation. An elevational difference was found in the sympatric Panamint kangaroo rat (PKR, *D. panamintinus*) and the Merriam's kangaroo rat (MKR, *D. merriami*) in the Mojave Desert (Price, Waser, & McDonald 2020). The authors concluded that interspecific competition might be an important factor. In contrast, there is no evidence of sympatry between AKR and DKR so direct competition could not be a factor, although it could have driven their elevational habitat separation in the past.

Analysis of runs of homozygosity (ROHs) indicated that elevation affected population structure. It was found that ROHs increased in length at higher elevation, indicating more inbreeding, probably due to a decrease in neighborhood size; however, this decrease did not reduce nucleotide diversity of the high elevation populations or the historical effective size of AKR.

To investigate if positive selection was differentiating AKR vs DKR and high- vs. low-elevation populations), we searched for "regions of interest" based on criteria that were both simple and easy to interpret, the highest 1% of Fst and the highest delta π, noting that since our data were not phased, linkage disequilibrium estimates were not available. This approach was used by Nanaei et al (2023) to identify genomic regions associated with local adaptation to hot and dry desert in Southwest Asia indigenous goats. The correlations between Fst and delta π were 0.3955 between the two species, 0.1558 within AKR, and 0.1141 within DKR. Some correlation was expected; however, Fst targets regions showing notable SNP frequency divergence, while delta π identifies those regions where this difference is associated with a loss of variation in one of the two groups. From this we inferred directionality: the group with the low genetic variation was subject to the selection that resulted in the increased Fst.

Adaptation in AKR relative to DKR

AKR was found to have 80 genomic regions with evidence of recent adaptation, while DKR had significantly fewer regions (= 26). Moreover 25 of those 26 regions consisted primarily of short tandem repeats and these regions showed reduced π in both species, whereas a selective sweep is expected to produce a π reduced below the genomic

average in one species. It is possible that the repetitive DNA created false positives involving the misalignment of identical short reads. Therefore, there is strong evidence suggesting that substantially more adaptive evolution has occurred in AKR.

One hypothesis to account for the positive selection in AKR is that the species has been adapting to the uplift of San Gabriel and San Bernardino Mountains, which has been continuous since the split of the two taxa. The need for such adaptation is suggested by the vicariant speciation of SKR and PKR noted earlier, given that neither of these species was able to adapt to the mountainous habitat, even though PKR can be found up to 1200m in parts of the Mojave Desert (Price, Waser, & McDonald 2000).

While it seems reasonable to assume that ecological factors associated with higher elevations, such as colder temperature, are likely to have played a role in the selection acting on AKR, it is notable that many of the regions identified by our screen do not contain coding genes, and when they do, there is an almost complete lack of nonsynonymous changes. Only one nonsynonymous change was found in our study: Arg(AKR)/ His(DKR) substitution in Slc39a4 (Zip4) (codon 404, Fst = 0.81713, within top 10%-15% range), a ZIP zinc transporter that is identified as a candidate gene in AKR. In humans, variation in this gene is hypothesized to reduce zinc uptake as a defense to pathogens (Engelken et al 2014). However, it seems unlikely that the prevalent allele Arg (G at the 2nd nucleotide of the codon) in AKR has driven a selective sweep since it is probably ancestral. This allele is found in all other kangaroo rat species (Ord's kangaroo rat, Stephen's kangaroo rat, and Merriam's kangaroo rat) except the banner-tailed kangaroo rat used in our phylogenetic study (Chapter 1). The outgroup, the Pacific pocket

mouse, also has the nucleotide G at this position but has the amino acid Gly, a pattern identical to the house mouse, suggesting that the G nucleotide is ancestral. No other SNP within the top 15% Fst values was identified near this candidate gene in DKR.

The deficit of amino acid changes suggests that most of the adaptive changes are regulatory. If so, then is it possible to identify the coding genes affected? Unfortunately moving beyond chromosomal regions that show the signature of selection to validating a candidate gene affected by a regulatory change is complex. For example, the mutations strongly associated with lactase persistence in pastoralist population in Europe and Africa were found at ~14k upstream of the lactase-phlorizin hydrolase (LPH) coding gene Lct and are in an intron of the adjacent minichromosome maintenance 6 (Mcm6) gene (Enattah et al 2002; Tishkoff et al 2007). Tishkoff et al (2007) showed signatures of selective sweeps in the surrounding region but used an expression essay to validate the mutation and its link to the gene. Another well-established example illustrates the large genomic distance that can separate a regulatory change from the target gene. Two enhancers of the En1 gene are located 400k – 500kb downstream of the gene, a gene that is essential for the formation of eccrine sweat glands in humans (Aldea et al 2023). Thus, even if a mutation under positive selection is found in a gene, that gene itself is not necessarily the causative factor of adaptation.

However, given that many regulatory elements are within a few thousand bases of the coding gene, candidate genes located within regions linked to selection were identified (Suppl. Material), since they are likely to be the genes being regulated. Some of these genes are known to be associated with adaptation to cold temperatures and

hypoxia. The candidate genes Rptor and Fkbp8 function in the rapamycin (mTOR) signaling pathway that is regulating the development of brown and beige adipocytes and heat production (Ye et al 2019). The expression of another candidate gene, Sugct, is reduced in mouse brown adipose tissue after cold exposure, and the expression of this gene may inhibit a thermogenic effector Ucp1 that is involved in non-shivering thermogenesis of brown and beige adipocytes (Chen 2021). Ncor1 is also a regulator of thermogenesis, and its downregulation improve fat oxidation and cold resistance in mice (Yamamoto et al 2011). Those four candidate genes were identified in separate regions of interests, and they all span or locate close to at least one SNP with an Fst in the top 5%. Arid1b encodes a component of the Swi/Snf chromatin remodeling complex that alter gene expression that regulates erythropoiesis under hypoxia (Azad et al 2022). Cpsf1 encodes cleavage and polyadenylation specificity factor-1 that promotes degradation of hypoxia-inducible factor 1-α (HIF-1α) though inhibition of ABL kinase (Mayro et al 2023).

Also, we have identified two contigs, ptg000074l and ptg000045l, that have more windows under selective sweeps than other contigs, and 3 candidate genes were identified on each of them. On ptg000074l, Chsy3 is glycosyltransferase that also carry out N-acetylgalactosaminyltransferase activities. Minar2 encodes membrane integral Notch2-associated receptor 2, the mutations in which causes deficiencies in stereocilia of hair cells and hearing loss in human and mice, and it is suggested to function in the inner ear (Bademci et al 2022). Adamts19 encodes an extracellular matrix metallopeptidase. On ptg000045l, Kcnmb2 encodes potassium calcium-activated channel subfamily M

regulatory beta subunit 2, and there are 2 SNPs (Fst = 0.969 & 0.937) downstream of this gene that highly differentiated between the species, with top 5 Fst values in all SNPs within regions of interests (Suppl Material). Pik3ca encodes phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha that is involved in glucose metabolic process and other pathways. Zmat3 regulates apoptotic process and DNA damage response.

The 4 loci that are known to be involved in adaptation to high elevations in human mentioned above, Epas1, Egln1, Prkaa1, Nos3, showed no signatures of selective sweeps.

The effects of elevation on the population structures in AKR and DKR

The levels of genomic differentiation among sites in both species was not consistent with isolation by distance. In AKR, LTC is most genetically unique site (Figure 2.4a), even though PHL is the most geographically distant site (Figure 2.1a). This same pattern is even more extreme in DKR, with BXSL being the most genetically unique site, even though it is very close to BXSH, while the sites AGA and MRR were very similar (Figure 2.4b) despite being separated by more than 40 km (Figure 2.1b). This pattern of the lowest elevation population in both species being the most genetically unique is consistent with local adaptation to elevation driving genetic divergence.

The populations of AKR and DKR are more structured at higher elevations (as indicated by the mean lengths of ROH; Figure 2.2). This could be a result of smaller and more spaced patches of habitats and/or lower population densities at higher elevations. Anecdotally, we found that finding good trapping locations was harder at AWC and

BSXH, where the terrain was steeper. The highest site, PHL, in contrast, is in a valley and very flat, but it is isolated by the surrounding steep mountain slopes.

The observations showed us that the dispersal of AKR and DKR is affected by geographic and ecological barriers, which can limit their ability to persist under climate changes and human disturbance. The vulnerability of kangaroo rat species is illustrated by SKR (*D. stephensi*), a species that is largely sympatric with DKR in western Riverside County and northern San Diego County. It is found in flat grassland habitats, and much of its habitat was lost due urbanization, with the remainder highly fragmented (Brock & Kelt 2004). The result was listing as a Federally endangered species which was recently modified to "threatened" due to the perceived success of its habitat conservation plan (Roach 2018). DKR was not as directly affected by human development, being distributed to adjacent hill slopes (Kelt et al 2008), but they are now increasingly restricted to vulnerable “islands” in the urbanized areas, and genetically unique sites, like BXSL, could be eliminated easily by habitat fragmentation and population fluctuations.

Local adaptation within AKR

Within AKR, the high elevation group (AKRhigh) showed more signatures of selective sweeps than AKRlow (77 vs 21 regions of interests; 37 vs 2 candidate genes), consistent with the hypothesis that the local adaptation in this group is primarily linked to selection at high elevations. In AKRhigh, 3 candidate genes have known functions that are also associated with cold adaptation and hypoxia. Ndufb2 and Rock1 were identified as candidate genes in fish that are associated with adaptation to coldness. Ndufb2, as a subunit of the NADH:ubiquinone oxidoreductase (complex I), is involved in the NADH

dehydrogenase activity and oxidoreductase activity was found to be under selection in polar fish (Hotaling et al 2023), and Rock1 was identified in related to migration timing in redband trout which largely depend on water temperatures (Chen & Narum 2021). Given that we previously identified 4 candidate genes related to thermogenesis in AKR compared to DKR, this may not be a coincidence, and the cold tolerance of those two species and intraspecific populations of AKR demands study. Multifunctionality of the genes make them possibility involved in adaptation to multiple ecological factors. Rock1 may also be associated with local adaptation to hypoxia, and it contributes to development of pulmonary hypertension under hypoxia stress that promotes hypoxemia (Penumatsa et al 2022) and was suggested to be an upstream regulator of HIF-1α pathway that is essential to hypoxic adaptation (Ohta et al 2012; Cheng et al 2021). Ndufb2 in the respiratory pathway mentioned above may play a role in the hypoxic adaptation as well. The oncogenic mutation V600E in another candidate gene, Braf, was found to increase HIF-1α expression (Kumar et al 2007; Zerilli et al 2010).

Using exome-wide population genomic analysis, Schweizer et al (2021) identified 992 candidate genes associated with local adaptation to high elevation in deer mice, and they reviewed 14 studies on 9 other vertebrates using population genomic data that identified a total of 3,983 unique genes involved in high elevation adaptation. However, only 163 (16.4%) of their 992 genes overlap with the genes identified in the other studies. Here we conclude that 1) given the large number of the candidate genes they identified, the criteria they used could be not stringent enough and prone to false positives in a data-driven study. 2) the loci involved in high elevation adaptation are variable across species.

4 of our candidate genes (10.8%) in AKRhigh, Ankhd1, Dmd, Rock1, and Fsd1l, are in their 992 genes, and Ankhd1 was also identified in Andean population of human (Eichstaedt et al 2014), which is thought to encode a scaffolding protein and negatively regulate apoptosis. In the interspecific comparison, 5 of the candidate genes (11.1%) in AKR, Strbp, Pik3ca, Zmat3, Dmd, And Rb1cc1, are in the 992 genes, and Strbp was also identified in Sakar Falcon (Pan et al 2017), which is an RNA binding protein predicted to be involved in mechanosensory behavior and spermatid development.

Though there are regions of interests identified in AKRlow as well, we think there is actually not much signatures of selection in this group. Similar to DKR in the interspecific comparison, the πs in those regions of interests in AKRlow are low for both groups, mostly driven by repetitive DNA, making the π ratio method more sensitive to the difference in π and prone to alignment error. As sister species, AKR is adapted to much higher mountainous areas than DKR, because AKRhigh are under more selective pressures, AKR may have migrated to and exploit these new habitats after their divergence from DKR, which cause more adaptative genomic changes in the groups at higher elevations. One of the two candidate genes, Ift88, in AKRlow showed high significance and encodes a ciliary protein in found in nephrons. Mutations in this gene cause polycystic kidney disease in mice (Lehman et al 2008; Hu et al 2021), consistent with the possibility that its regulation is altered in AKRlow, in response to limited water availability.

Local adaptation within DKR

In DKR, there is evidence of local adaptation in the low elevation group at BXSL (47 regions of interest vs 17 in DKRhigh; 25 candidate genes vs 2 in DKRhigh). There is no obvious patterns in the functions of the candidate genes, but we hypothesized that one of the environmental factors that could have shaped the genomic adaptation in this group is heat-related stresses because the low elevation areas are generally hotter and dryer. A candidate gene, Dnajc21, is associated with heat shock proteins Hsp40s, which are cochaperone proteins that bind to Hsp70 chaperones through the J domain of Dnajc proteins and regulate the chaperone activities by ATP hydrolysis (Qiu et al 2006). The Hsp70 family mediate the folding of peptides and refolding of denatured proteins, thus playing an essential role in the heat-induced stresses and damages (Baler, Welch & Voellmy 1992; Morano 2007). Two other genes, Lrp2 and Bcl11b are involved in negative regulation of apoptosis, which could be a response to environmental stresses, while the functions of the others range from developmental regulation to immune responses, suggesting the complexity of the gene networks in shaping adaptive traits.

The biological factors like predation may also play a role in the local adaptation in these kangaroo rats. A candidate gene identified in DKRlow, Slc45a2, is a membrane-associated transporter that mediates melanin synthesis, and its variants are found to be related to the pigmentation of human, chicken, Japanese quail, horse, and sheep breeds (Graf et al 2007; Gunnarsson et al 2007; Deng et al 2009; Bellone 2010). In the future, Slc45a2 could be a new candidate gene for local adaptation in rodents like the Mc1r genes. Although the coat colors of kangaroo rats are similar overall, slight color variation

between subspecies and sister species have been documented (Lidicker 1960), and Slc45a2 could be one of the genes underlying the color variants. There is no data linking the coat color to elevations in kangaroo rats so far, and we didn't notice obvious difference in the colors of the two groups, variation may exist as the surrounding vegetations are different to some degree, as there are more invasive grasses and denser plants at BXSL.

In DKRhigh, we identified 2 SNPs with top 15% upstream to an olfactory receptor, Or2l13, but none near the other candidate gene. Olfaction has been demonstrated to play a role in the kangaroo rat behaviors such as foraging, avoiding predators, and interspecific and intraspecific interactions (Rebar 1995; Mirsky 1971; Perri & Randall 1999; Randall 2013; Herman & Valone 2000).

Overall, from the interspecific and intraspecific tests, we found the selective sweeps are associated with high elevation in AKR and low elevation in DKR, the two extreme elevations in the two species, suggesting more selective pressures at these areas, e.g., coldness and hypoxia at high elevations and heat and drought at low elevations.

Conclusion

Using genomic data, we revealed the population structures, genetic variation, and selective sweeps in two kangaroo rat sister species, AKR and DKR. Our results showed that elevation plays an important role in the evolutionary changes in AKR and DKR. We also identified candidate genes that are potentially underlying the local adaptation in both species and could be applied to future studies in kangaroo rats as well as other non-model organism.

Chapter 2 Figures

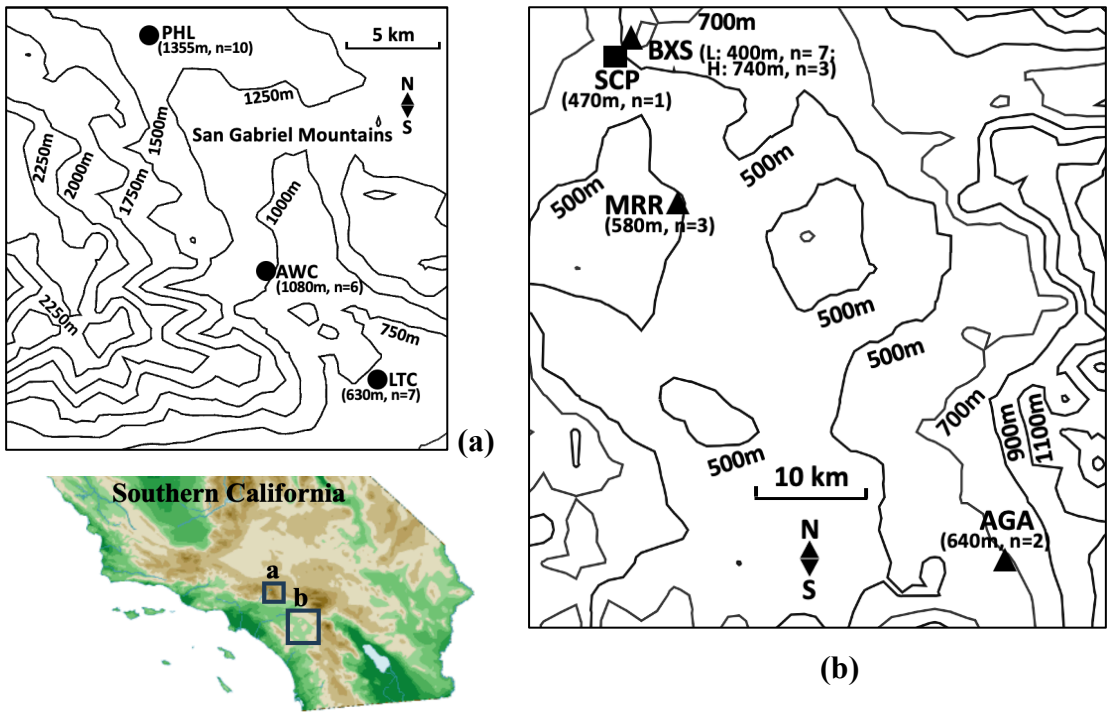

Figure 2.1. AKR and DKR sampling sites. (a) AKR sites. (b) DKR sites. The site of the sample for Pacbio reference genome (SCP) is label as a filled square and others as filled triangles. The isohypse lines were obtained on http://contourmapcreator.urgr8.ch/

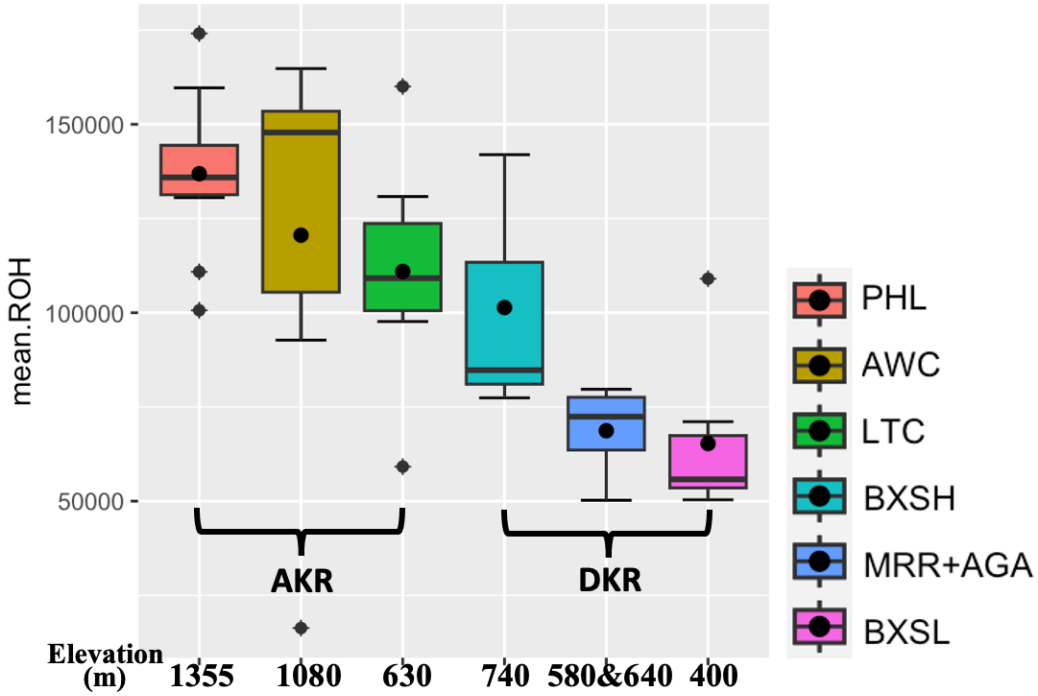

Figure 2.2. Mean sizes of ROHs in AKR and DKR populations. The box plots show the mean sizes of ROHs in all sampling sites, and the sites for the same species are ordered by elevation from high to low. The round dots and horizontal lines represent mean and median values respectively. The error bars represent the 95% confidence intervals, and the diamonds show outliners.

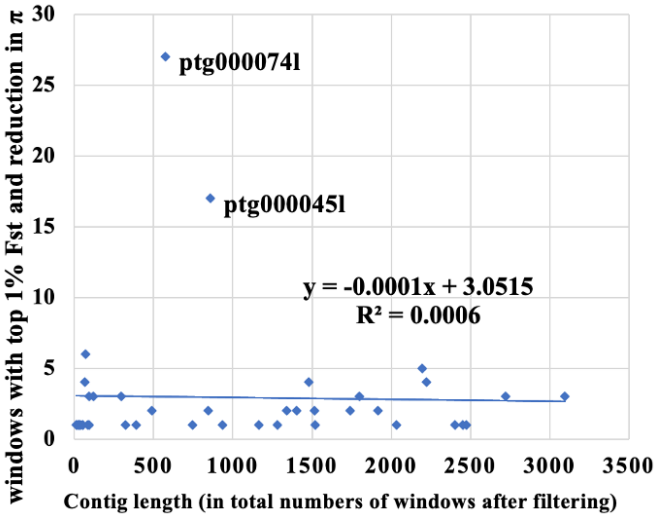

Figure 2.3. The numbers of windows with top 1% Fst and reduction in π on different contigs in AKR. Two contigs, ptg000074l and ptg000045l possess relatively more regions under selective sweeps than the other contigs.

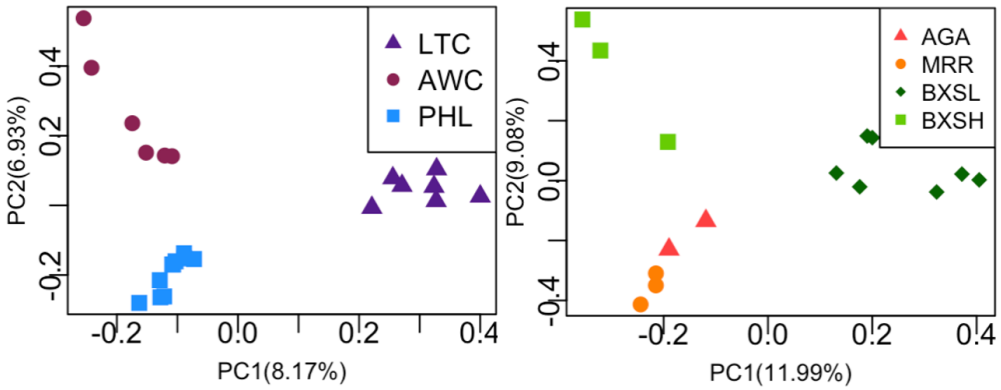

Figure 2.4. Principal component analysis (PCA) using genome-wide SNPs for each species considered separately. (a) AKR samples showing clear separation of sample sites. (b) DKR samples showing that BXSL separate from the other sites.

Chapter 3. Detection of genomic adaptation and convergence in arid-adapted rodents with a focus on the Heteromyidae.

Abstract

Rodents in the family Heteromyidae have evolved extraordinary adaptation to arid and semi-arid habitats. Genomic studies of this adaptation have been mostly focused on their kidney function and osmoregulation abilities. Taking advantage of the increasing availability of genomes of non-model organisms, we used both data- and hypothesis-driven approaches to look for genes subject to positive selection in kangaroo rats and a pocket mouse using a hierarchical testing method that put the search in a phylogenetic context. The identified candidate genes included several involved in energy production and maintaining homeostasis like glucose and osmosis pressure, apoptosis, plus various other functions. The selection on most of these genes was acting throughout the evolution of the heteromyid clade. Also, we investigated positive selection in unrelated arid-adapted rodents (jerboa, gerbil, and cactus mouse) that have evolved morphological and ecological similarity to heteromyids. We found evidence of convergence and linage specific adaptation. In the 7 candidate genes evolved convergently in the arid adapted rodents, there is one involved in glucose homeostasis, but we are not able to infer the selection agents of the others, and further studies on those genes may provide us new insights into arid adaptation.

Introduction

Adaptation refers to the evolutionary changes enabling species, subspecies, and populations to better survive and reproduce in their environments, changes that are driven

by natural selection. The development of whole genome sequencing has facilitated new in-depth studies of adaptive traits by combining sequence data with an understanding of natural history and special environmental conditions of the study species. One approach is to compare the genomic sequences of organisms having extreme adaptative features with their differently adapted close relatives. For example, comparative genomic studies have found evidence of the genes underlying adaption to some extreme environments in a variety of mammalian taxa: polar (Liu et al 2014); aquatic (Foote et al 2015); subterranean (Fang et al 2014; 2015); and arid (Wu et al 2014; Marra, Romero, & DeWoody 2014).

Next generation sequencing (NGS) and third-generation sequencing (that produces longer DNA reads; Bleidorn 2016) has boosted the genomic data available from non-model organism. These data have facilitated studies of adaptation that can be divided into two broad categories: data-driven and hypothesis-driven (Le Duc & Schöneberg 2016). Data-driven studies employ statistical methods based on population genetic theory to detect genes underlying adaptation. For example, Liu et al (2014) used genome re-sequencing data to find genes that contribute to cold adaption in polar bears using the closely related brown bear as an ecological contrast. Typically, this method involves scanning thousands of genes, so it has the disadvantage that the necessary correction for multiple testing greatly reduces statistical power. In contrast, the hypothesis-driven approach limits the multiple-testing problem by focusing on the small sample of genes involved in certain biological processes chosen *a priori*. This approach requires the

availability of lists of genes potentially involved in the adaptive phenotypic and/or ecological traits of the subject organisms.

Deserts are one of the most extreme environments on the earth, which are known for low water availability, high temperatures during daytime, and challenges for locomotion. Wu et al (2014) examined the genomic adaptation in arid-adapted camels using the alpaca, which is in the same family and lives at high altitudes, as a phylogenetic and ecological comparison. Another mammalian taxon with desert-adapted species are the rodents. They include species that have independently adapted to arid conditions in deserts all around the world.

The N. American kangaroo rats (family *Heteromyidae*) are seed-eating desert-adapted rodents in the genus *Dipodomys*. In this project, we selected kangaroo rats as focal species to look for genomic adaptation in heteromyids with an emphasis on arid adaptation using both "data-driven" and "hypothesis-driven" approaches.

Kangaroo rats have evolved a high level of drought tolerance and require little water intake (Nagy & Gruchacz 1994). In addition to the physiological adaptation for water conservation, kangaroo rats are nocturnal, even avoiding bright moonlight (Daly 1992), suggesting adaptations involving improved vision in dim light and hearing to reduce the risk of predation by owls and rattlesnakes (Webster & Webster 1971). For example, kangaroo rats have evolved enlarged auditory bullae (cavity for inner and middle ears) and low frequency hearing to avoid predators (Webster & Webster 1971; 1984), a feature also seen in the closely related group of pocket mice (Brown et al 1988).

Despite the well-known phenotypical arid adaptation of kangaroo rats and other heteromyids, little is known about the genetic basis of their arid adaptation. Marra et al (2012) investigated adaptation in the banner-tail kangaroo rat, employing both an *a priori* hypothesis-driven search of candidate genes known to be involved in kidney function in model rodents (*Mus musculus* and *Rattus norvegicus*) and an *a posteriori* data-driven search of the kidney transcriptome to identify genes upregulated (relative to the spleen). Marra, Romero, & DeWoody (2014) compared the kidney transcriptomes and orthologous sequences of the arid-adapted and non-adapted species of heteromyids, and they identified one gene, Slc2a9, putatively subject to positive selection (dn/ds>1), which is involved in urine concentration in humans and mutations in this gene cause renal hypouricemia type 2 (RHUC2) (Döring et al 2008; Kawamura et al 2011).

Following the data and hypothesis-driven approaches of Marra, Romero, & DeWoody (2014), first we looked for positive selection in kangaroo rats and more generally in heteromyids by comparing them to several outgroups as phylogenetic and ecological controls, and second we specifically investigated the genes involved in renal function and ion transportations as well as the genes related to rod cell development and inner ear and middle ear morphogenesis since they may have evolved to improve vision in dim light and hearing in the kangaroo rat.

Another approach to identifying gene adaptation is to look at convergent evolution among unrelated taxa subject to similar environments and lifestyles. Such convergence has been seen at all levels from phenotypes to DNA sequences (Buggiotti et al 2021; Foote et al 2015; Le Duc & Schöneberg 2016; Natarajan et al 2016; Parker et al 2013;

Thorpe et al 2015). For example, the genetic variants of a single gene, Mcr1, underlies the color adaptation of rodents (Hoekstra, Krenz, & Nachman 2005; Hoekstra et al 2006; Nachman et al 2003), lizards (Rosenblum et al 2004), and birds (Theron et al 2001), and it is well recognized that an increased hemoglobin-oxygen affinity is often involved in adaptation to the hypoxic environments (Storz 2016), such as high altitude (Chen et al 2016; Natarajan et al 2016; Signore et al 2019), aquatic (Tian et al 2016), and subterranean (Tomasco et al 2017; Pejo & Tomasco 2021) environments. Also, convergent amino acid substitutions are also found in genes involved in ammonia detoxification and CO_2-induced pain repression in subterranean mammals (Fang et al 2014).

Studies using genomic data for the detection of convergent gene evolution have involved both data-driven and hypothesis-driven approaches, with subject species selected because of their convergent phenotypical and ecological traits, e.g., dietary adaptations in carnivores and herbivores (Hecker, Sharma, & Hiller 2019), echolocating in bats and dolphins (Parker et al 2013), hypoxia adaptation in birds (Natarajan et al 2016), and morphological traits in anole lizards (Thorpe et al 2015). Foote et al (2015) conducted a comparative genomic study on marine mammals from three orders, (killer whale, walrus, and manatee), pairing each of them with terrestrial sister taxa in the same orders. This comparison revealed some examples of identical amino acid substitutions under positive selection in the marine species involved in bone and inner ear formation, hyperthyroidism, and cardiovascular regulation.

In rodents, a number of unrelated taxa dwell in desert habitats, including kangaroo rats, jerboas, and gerbils, and have evolved convergent phenotypic traits such as drought tolerance, bipedal locomotion, nocturnality and burrowing (Mares 1993). Of course, a convergent ecological adaptation may be attributed to the changes in the same or different features. For example, to improve hearing, kangaroo rats acquired enlarged auditory bullae, while jerboas evolved an enlarged ear pinna (Eisenberg 1975). However, in other cases there may be convergence at the morphological and gene level. To look for convergent genomic changes that are involved in arid adaptation in rodents, we used a data-driven approach to compare four unrelated desert-dwelling species (Dulzura kangaroo rat, lesser Egyptian jerboa, Mongolian gerbil, and cactus mouse), in each case using a non-arid adapted sister species as a control to confirm that the observed changes evolved in the arid-adapted branch.

Methods

Genomic Data

To identify positively selected genes in the kangaroo rat clade, and more broadly in the heteromyids, we used three heteromyid genomes, the Dulzura kangaroo rat (DKR, *Dipodomys simulans*) to represents the *agilis* group of kangaroo rats, the Banner-tailed kangaroo rat (BKR, *Dipodomys spectabilis*) which is in the *spectabilis* group of kangaroo rat (Alexander & Riddle 2005), and the little pocket mouse (LPM, *Perognathus longimembris*). The genomes of the BKR (GCF_019054845.1), and LPM (GCA_024363575.1) were downloaded from NCBI, and we used our genome sequence of DKR (Chapter 1). The genomes of three non-arid rodent species were used as controls:

the Botta's pocket gopher (*Thomomys bottae*) (GCA_024803745.1) that is within the same superfamily *Geomyoidea* with heteromyids but not arid adapted, the North American beaver (*Castor canadensis*) (GCA_009822645.1) that is within in the same suborder *Castorimorpha* with heteromyids but adapted to an aquatic environment, and two outgroups, the hoary bamboo rat (*Rhizomys pruinosus*) (GCA_009823505.1) that is in the suborder *Myomorpha* and the Malayan porcupine (*Hystrix brachyura*) (GCA_016801275.1) in the suborder *Hystricomorpha* that diverged from the other two suborders earlier.

To detect the genes involved in convergent arid adaptation of rodents, we examined four pairs of arid adapted and non-arid adapted species, with each pair being as closely related as genome availability in NCBI permitted. For this purpose, we paired DKR and the Botta's pocket gopher, and added three other species found in arid environments plus a non-arid partner for each: the lesser Egyptian jerboa (*Jaculus jaculus*) (GCF_020740685.1) and the meadow jumping mouse (*Zapus hudsonius*) (GCA_004024765.1); the cactus mouse (*Peromyscus eremicus*) (GCA_949786415.1) and the muskrat (*Ondatra zibethicus*) (GCA_004026605.1); and the Mongolian gerbil (*Meriones unguiculatus*) (GCA_002204375.1) and the house mouse (*Mus musculus*) (GCF_000001635.27).

Our data-driven approach to detect genes under positive selection in the arid-adapted rodents was based on the standard set of genes identified by BUSCO (Benchmarking Universal Single-Copy Orthologs) (Simão et al 2015) using the mammalia_odb10 lineage dataset. This dataset includes the nucleotide and amino acid sequences for 9226 complete

single-copy orthologs. Because using clustalw (Sievers et al 2011) to directly align the nucleotide sequences often produces misalignments and frameshifting caused by nucleotide substitutions, missing sequences, and non-homologous sequences at intron-exon boundaries, we used it to align the amino acid sequences of those genes and used the alignments as reference to align nucleotide sequences using pal2nal (Suyama et al 2006). Then, the suspected misalignments were trimmed using custom R scripts (Suppl. Material). After testing, the genes with significant evidence of selection were inspected for obvious misalignment that could have caused false positive, and the ones with bad alignment qualities were improved manually or removed.

To detect genes under positive selection in heteromyids and kangaroo rats, we selected outgroups from different clades to estimate the background ω to the largest extent keep the independency between the data-driven analyses and hypothesis-driven analyses, and for the hypothesis driven approach, we only selected those with annotated genomes in order to download coding sequences of the annotated genes in selected GO terms directly from NCBI. The species trees for the PAML tests were based on Blanga-Kanfi et al (2009), and the branch lengths of ds trees (derived from synonymous sites) were estimated on the concatenated orthologs in PAML (Yang 2017).

Detecting positive selection in heteromyids

We used the 7 species mentioned above (Figure 3.1) to detect the genes under adaptive positive selection in the heteromyids using the criterion that one or more dn/ds (= ω) values within the clade were significantly higher than those of the other taxa. For each gene that gave significant support to this *a priori* hypothesis, we employed the

phylogenetic approach of Nunney and Schuenzel (2006) to detect any further notable heterogeneity in ω within the tree. The testing procedure is summarized in Table 3.1, where models are defined by M and ω (e.g., model M0ω1), with M0 etc., referring to the model setting in PAML and ω1 etc., referring to the number of estimated values of dn/ds.

The criterion used to detect a gene with probable adaptive gene change in the heteromyids was a significantly higher ω in the branches of the heteromyid clade relative to the rest of the tree. To this end, we used codeml in PAML (Yang 2007) to compare the fit of a model with a single ω for all the branches (model M0ω1) to a model with one ω for the heteromyid clade and another for the remaining branches of the tree (model M2ω2). When the ω estimated for the heteromyid clade was greater than the background estimate, the significance of the difference was evaluated using the likelihood ratio test, LRT ($\chi^2 = -2[\ln L(H_0) - \ln L(H_a)]$). In this case, H_0 is assumes no heterogeneity in all branches (model M0ω1), and H_a assumes the heteromyid clade has a different ω than the other branches (model M2ω2). To correct for the testing all of the identified BUSCO genes, significance levels were adjusted using the Holm–Bonferroni correction (Holm 1979) to keep the family-wise error rate (FWER) under the 5% significance level. In this correction, the genes were ranked according to their significance level determined by the LRT in ascending order, and these p-values are multiplied by (n – rank + 1), where n is the total number of tests and rank is the rank of the gene. This was a 1-tailed test (higher ω within the heteromyids), the difference in ω was considered significant given a "corrected p" < 0.1.

Genes with significantly elevated ω in the heteromyid branches were tested for residual heterogeneity in ω within the heteromyid clade by comparing the model M2ω2 to model M2ω5 that estimated an ω for each heteromyid branch (except the two short kangaroo rat branches, estimated at 4.5 to 5.5 mya (see Chapter 1), which were still assigned the same ω) using an LRT (using unadjusted significance levels). If heterogeneity was found (i.e., model M2ω5 provided a significantly better fit), then following Nunney and Schuenzel (2006), we applied a series of hierarchical tests to identify its source (Table 3.1). We first modified model M2ω2 by adding a third ω by giving the basal branch of the heteromyids its own value (model M2ω3). Comparison of these two models tested the possibility that the short basal branch had a lower ω than the terminal heteromyid branches.

Following this test, if heterogeneity in ω was still apparent (i.e., M2ω5 provided a significantly better fit than M2ω3), then M2ω3 was modified to incorporate a separate ω for the pocket mouse branch (model M2ω4). Following a test of whether the ω of the pocket mouse branch was distinct (using an LRT comparing M2ω4 with M2ω3), heterogeneity was again assessed, and if there was, then heterogeneity among the 3 remaining outgroup branches (see Figure 3.1, bearing in mind the tree was analyzed as an unrooted tree) was assessed by assigning the ancestral branch of kangaroo rat a distinct ω (model M2ω5). These models and tests are summarized in Table 3.1. Finally, we ran the "best fit" model (Mbf) created by collapsing the ωs that were not significantly different and contiguous within the tree.

Table 3.1. Maximum likelihood models and hierarchical analyses to detect positive selection in heteromyids. The models are defined as in PAML (M0 or M2) followed by the number of dn/ds (=ω) values estimated.

(a) Maximum likelihood models tested

Model	Model description
M0ω1	Single ω for the whole tree
M2ω2	Single ω for the heteromyids, and single ω for the non-heteromyids
M2ω3	Root of heteromyids separated from the rest of the clade
M2ω4	Pocket mouse separated from the other heteromyids
M2ω5	Root of kangaroo rats separated from the kangaroo rats

(b) Hierarchical maximum likelihood analysis

Likelihood ratio test	Test of Null Hypothesis for ωs
M2ω2-M0ω1(df=1)	heteromyids ≤ non-heteromyids
M2ω5-M0ω2(df=3)	Homogeneity among the heteromyids
M2ω3-M2ω2(df=1)	Root of heteromyids = the heteromyids
M2ω4-M2ω3(df=1)	Pocket mouse = kangaroo rat
M2ω5-M2ω4(df=1)	Root of kangaroo rats = the kangaroo rats

Convergent evolution in arid adapted rodents

To detect genes that have undergone convergent evolution in arid-adapted species, single copy orthologs retrieved by BUSCO in 4 unrelated arid-adapted rodent species and their 4 non-arid partners, we used a standard rodent phylogeny (Figure 3.2; Blanga-Kanfi et al 2009). To detect convergent evolution, we used an approach similar to that outlined in the previous section, comparing model M2ω2 (one ω for the terminal branches of the 4 arid-adapted taxa, and another ω for the rest of the tree) to the null model M0ω1 (one ω for all branches), and determining if any genes showed a significantly higher ω in the arid adapted taxa. The basal branches were assumed as non-arid adapted, so they together with the non-arid adapted sister species provide a background ω.

These significant genes were then examined for residual heterogeneity by comparing M2ω2 to the model that assigns a different ω to every branch (M1ωall). If no

heterogeneity was present, then the gene was considered convergently selected. If heterogeneity was present, because the background levels of selection could be different in the 4 pairs of sister species and their ancestors, rather than locating the heterogeneity in ωs among the arid-adapted branches or other branches, we conducted a comparison between the ωs on the sister species pairs as defined in M1ωall. If every arid adapted species has a higher ω than its non-adapted sister taxon, the gene was selected as a candidate gene. The models and tests are listed in Table 3.2.

Table 3.2. Maximum likelihood models and hierarchical analyses to detect convergent evolution in the arid adapted rodents.

(a) Maximum likelihood models tested	
Model	Model description
M0ω1	Single ω for the whole tree
M1ωall	ω free to vary independently in all branches
M2ω2	Single ω for the arid adapted rodents, and single ω for the other branches
(b) Hierarchical maximum likelihood analysis	
Models Compared	Test of Null Hypothesis for ωs
M2ω2-M0ω1(df=1)	Arid-adapted rodent ≤ non-adapted rodent
M1ωall-M2ω2(df=11)	Homogeneity across arid adapted rodents and across rest of the tree

Positive selection in each of the arid adapted rodents

To look for genes under positive selection unique to each of the four arid adapted rodents used in the previous test, we compared the estimated ω between each arid-adapted species and its non-arid-adapted sister species (plus their ancestral branch) using the same alignments used in the convergence study mentioned above and ignoring the ωs in the rest of the tree. For the null model, we used an M2ω2 model, estimating one ω for the clade of the sister species and their ancestral branch, and one for the rest of the tree

(outgroups). The alternative model (M2ω3) assigned split the branch to the arid-adapted species from the non-arid-adapted species and their ancestral branch. If M2ω3 was significantly better than M2ω2 after correction for multiple testing (and given a 1-tailed test) and the ω of the arid species branch was the greater of the pair, the gene was kept. Then, we tested whether there was heterogeneity of ω between the non-adapted sister species and the ancestor using a model that separate the ancestral branch (M2ω4). If M2ω4 was significantly better than M2ω3, the gene was considered as a candidate gene if the non-adapted species ω was lower than the ancestor's. We did not consider heterogeneity in ω in the outgroups. The models and tests are listed in Table 3.3.

Finally, we checked whether the candidate genes identified in one species or in the analysis focused on the heteromyid clade had elevated ωs in any of the other arid-adapted species, noting that in the initial analysis some positively selected genes may not have been detected because of the multiple-test correction.

Table 3.3. Maximum likelihood models and analyses to detect gene under positive selection in each arid adapted rodent.

(a) Maximum likelihood models tested

Model	Model description
M2ω2	Single ω for the pair of sister species and their ancestor, and single ω for the others
M2ω3	Arid-adapted rodent separated from the non-adapted sister species and their ancestral root
M2ω4	Non-adapted sister species separated from the ancestral root

(b) Hierarchical maximum likelihood analysis

Models Compared	Test of Null Hypothesis for ωs
M2ω3-M2ω2(df=1)	Arid-adapted rodent ≤ non-adapted sister species and the ancestor
M2ω4-M2ω3(df=1)	Non-adapted sister species = the ancestor

Identifying candidate genes in heteromyids using the hypothesis driven analyses

To look for candidate genes related to nocturnality in the kangaroo rats, we selected several GO (gene ontology) term categories that are relevant to night vision and hearing suggested by Le Duc & Schöneberg (2016) in their investigation of nocturnal birds in their review, which are "retinal rod cell development" (GO: 0046548), "retinal cone cell development" (GO: 0046549), "sensory perception of sound" (GO: 0007605), "inner ear morphogenesis" (GO: 0042472), and "middle ear morphogenesis" (GO: 0042474). Also, we investigated whether positive selection has acted on the candidate genes in the arid-adaptation related GO category "renal system process" (GO:0003014) (http://geneontology.org/). For this analysis we used the available annotated heteromyid genomes: BKR, Ord's kangaroo rat (OKR, *Dipodomys ordii*, GCF_000151885.1), and the Pacific (little) pocket mouse. PPM, *P. l. pacificus*) (GCF_023159225.1), which is less completed than the LPM assembly but was annotated. The outgroups used were the beaver, the most closely related outgroup since the pocket gopher genomes was not annotated, plus the house mouse, naked mole-rat (*Heterocephalus glaber*, GCF_000247695.1) and the Arctic ground squirrel (*Urocitellus parryii*, GCF_003426925.1). These outgroups are all adapted to lifestyles very different from kangaroo rats and represent clades that were not included in the previous tests. They were downloaded from NCBI using E-utilities (Sayers 2010). The naked mole rat was excluded from the vision related genes. The house mouse was chosen because its genome is among the most completed and well annotated ones.

The hypotheses and the models tested are the same with those used for identifying positive selection in BUSCO genes in heteromyids (Table 3.1). For candidate genes identified, we also estimated the best fit models (Mbf).

Results

Positive selection in heteromyids

The phylogenetic tree of species used for detecting positive selection in heteromyids with branches lengths estimated in PAML is shown in Figure 3.1.

We identified 6069 of the 9026 BUSCO genes in all 7 genomes, and 18 genes were found with significantly elevated ωs in heteromyids (Table 3.4). The 4 types of best fit models (Mbf) for those genes are shown in Figure 3.3 (a) to (d). Of the 18 candidate genes identified, 11 were linked to heteromyid clade (Figure 3.3d), 5 to the kangaroo rats (Figures 3.3a & b) and 2 to the pocket mouse (Figure 3.3c).

In the 5 genes linked to kangaroo rats, Sesn3 (pattern Fig 3.3b) and C2cd5 (pattern Fig 3.3a) are involved in insulin regulation and glucose homeostasis, Cherp (pattern Fig 3.3a) functions in intracellular calcium ion homeostasis, Wdr48 (pattern Fig 3.3b) regulates response to DNA damage, and Xab2 (pattern Fig 3.3a) is a splicing factor that involved in transcription-coupled DNA repair. The two genes linked to pocket mouse, Aebp1 and Dsp, are both involved in extracellular structural constitution and wound healing. The 11 genes linked to heteromyids carry out various functions, suggesting different aspects of adaptation and complicity of gene networks e.g., Arhgef40 is involved in small GTPase mediated signal transduction (Suppl. Material). Notably, Raf1 is also involved in insulin secretion in response to glucose, Prkag2 is involved in muscle

cell glucose metabolism, and Tsc22d2 is involved in response to osmotic stress, which might be an important factor in arid adaptation.

Table 3.4. The candidate genes for adaptation with significantly elevated ωs in heteromyids. The Chi squares and p values are for comparing M2ω2 to M0ω1. The ωs are the ones in the best fit models (Mbf) defined in Figure 3.3. ω1 is for the branches under selection, and ω0 is for the background branches.

Mbf	Gene	Chi square	p-value	ω1	ω0	Major functions
Krats (Fig 3.3a)	Xab2	25.99	3.43×10^{-07}	0.1642	0.0106	mRNA splicing
	C2cd5	20.23	6.88×10^{-06}	0.2580	0.0511	response to insulin
	Cherp	18.54	1.67×10^{-05}	0.0811	0.0141	calcium ion homeostasis
Krats + root (Fig 3.3b)	Sesn3	45.99	1.19×10^{-11}	0.3080	0.0497	glucose homeostasis
	Wdr48	30.02	4.29×10^{-08}	0.0954	0.0136	response to DNA damage
Pmouse (Fig 3.3c)	Aebp1	27.20	1.84×10^{-07}	0.2658	0.0649	cell differentiation and extracellular matrix constitution
	Dsp	18.84	1.42×10^{-05}	0.1090	0.0372	intermediate filament anchoring
Het clade (Fig 3.3d)	Rbm15b	69.39	8.09×10^{-17}	0.0950	0.0171	regulation of alternative mRNA splicing
	Fam83b	48.34	3.59×10^{-12}	0.2163	0.0553	phosphatidylinositol 3-kinase binding
	Lingo1	32.74	1.05×10^{-08}	0.0298	0.0061	epidermal growth factor receptor binding
	Tsc22d2	25.91	3.58×10^{-07}	0.2341	0.0470	response to osmotic stress
	Prkag2	24.84	6.22×10^{-07}	0.1282	0.0339	cellular energy status regulation
	Wwc1	22.92	1.69×10^{-06}	0.1717	0.0777	regulation of transcription
	Shf	22.84	1.76×10^{-06}	0.2115	0.0586	apoptotic process
	Mtx2	20.79	5.13×10^{-06}	0.2550	0.0285	mitochondrial transport
	Raf1	20.57	5.75×10^{-06}	0.1009	0.0282	apoptotic process
	Arhgef40	18.84	1.42×10^{-05}	0.3264	0.1735	regulation of GTPase signal transduction
	Pknox1	18.54	1.67×10^{-05}	0.1409	0.0418	regulation of transcription

Convergent evolution in arid adapted rodents

The phylogeny of the four pairs of sister species used in the analysis is shown in Figure 3.2.

In the 3694 genes that were identified in the genomes of all 8 species, we found 7 genes that showed evidence of convergent evolution, having significantly elevated ω-values in the arid adapted relative to the non-adapted rodents (Table 3.5). One of those genes, Pik3c3, is involved in cellular response to glucose starvation. The other genes have various functions from structural to regulatory (Table 3.5 & Suppl. Materials).

Table 3.5. The candidate genes for convergent evolution with significantly elevated ωs in arid adapted rodents. The Chi squares and p values are for comparing M2ω2 to M0ω1. The ωs are the ones in the M2ω2. ω-arid is for the arid-adapted rodents, and ω-nonarid is for the background branches.

Gene	Chi-square	p-value	ω-arid	ω-non-arid	Major functions
Mns1	24.53	7.31×10^{-07}	0.2958	0.0940	nuclear skeletal protein used in meiosis
Ccdc166	22.20	2.45×10^{-06}	0.3770	0.1131	unknown functions
Pigm	21.04	4.50×10^{-06}	0.1292	0.0383	cell surface anchor protein synthesis
Pik3c3	20.88	4.90×10^{-06}	0.0610	0.0193	response to glucose starvation
Copb1	19.15	1.21×10^{-05}	0.0553	0.0141	Golgi vesicle-mediated transport
Ino80e	18.32	1.87×10^{-05}	0.2671	0.0487	chromatin remodeling and DNA repair
Cacnb3	17.71	2.57×10^{-05}	0.0579	0.0112	calcium ion transport

Positive selection in each of the arid adapted rodents

We identified four genes with elevated ω in DKR, two in the jerboa, three in the gerbil, and one in cactus mouse relative to their non-arid partner and the common

ancestor (Table 3.6). In those genes, Cherp was identified to be under selection linked to kangaroo rat in the previous tests using the heteromyid phylogeny. Of the other 3, Prpf8 was missing in the alignments and not analyzed, and the other 2 were not significant in the heteromyid analyses because the ω-values were relatively low in pocket mouse. In the other 4 genes with elevated ω in heteromyids and selection linked to kangaroo rats (Figure 3.3a & b) that showed no significance here, C2cd5 and Xab2 are missing from the alignments for this comparison, while the other 2 showed relatively higher ωs in pocket gophers than in the other outgroups in the previous tests.

Table 3.6. The candidate genes with significantly elevated ωs in one of the arid species. The Genes listed show a significantly elevated ω in the target species arid (ω-arid) species, relative to the non-arid control and their ancestral branch (ω-nonarid), tested by comparing model M2ω3 to M2ω2.

Gene	Species	Chi-square	p value	ω-arid	ω-non-arid	Major functions
Cherp	DKR	33.77	6.20×10^{-09}	0.1222	0.0117	calcium ion homeostasis
Kiaa0930	DKR	28.25	1.07×10^{-07}	0.2182	0.0159	unknown functions
Sytl1	DKR	20.01	7.71×10^{-06}	0.3706	0.0476	exocytosis
Prpf8	DKR	17.81	2.44×10^{-05}	0.0443	0.0040	mRNA splicing
Crh	jerboa	22.63	1.96×10^{-06}	2.8668	0.0111	corticotropin-releasing factor for response to stress
P2ry1	jerboa	18.88	1.39×10^{-05}	0.2527	0.0368	G-protein coupled receptor
Mcrs1	gerbil	27.40	1.65×10^{-07}	0.6260	0.0327	chromatin remodeling and DNA repair
Ddx54	gerbil	26.26	2.98×10^{-07}	0.2977	0.0509	RNA helicase activity
Mrpl15	gerbil	18.09	2.10×10^{-05}	0.5018	0.0413	mitochondrial translation
Nbr1	cactus mouse	19.30	1.12×10^{-05}	0.6062	0.1253	macroautophagy

One gene (Crh) was found to have ω > 1 in the jerboa (Table 3.6), and when we tested this gene in the other 3 arid adapted species, this gene also showed a markedly elevated ω in cactus mouse (ω = 1.0568). Crh encodes a corticotropin-releasing factor that responds to stress. Two genes identified in the Heteromyid analysis also showed significance in a second species (Suppl. Materials). Dsp with selection linked to pocket mouse and Arhgef40 linked to heteromyid clade (Table 3.4) have elevated ω in jerboa.

Positive selection in heteromyids within selected GO terms

The phylogeny of the annotated species with branch lengths estimated in PAML is shown in Figure 3.4.

43 genes in "renal system process" (GO:0003014) and the 53 genes in "inner ear morphogenesis" (GO: 0042472) were analyzed since they were identified in the genomes of all selected species. Of these, two of the 43 renal associated genes had a significantly elevated ω in the heteromyids, plus one of the inner ear genes (Table 3.7). Each showed a slightly different pattern of selection in heteromyids (Fig 3.5a to Fig 3.5c). Two renal related genes, Tbc1d8b (pattern Fig 3.5a) and Oxtr (pattern Fig 3.5b) encode a GTPase-activating protein related to nephrotic syndrome and oxytocin receptor, respectively. The inner ear-related gene, Gata2, encodes a transcription factor regulating inner ear morphology and showed the pattern in Fig 3.5c. No candidate gene was identified in the other GO term categories, "retinal rod cell development" (GO: 0046548), "retinal cone cell development" (GO: 0046549), "sensory perception of sound" (GO: 0007605), and "inner ear morphogenesis" (GO: 0042472).

Table 3.7. The candidate genes with elevated ωs in heteromyids from selected GO terms. The Chi squares and p values are for comparing M2ω2 to M0ω1. The ωs are the ones in the best fit models (Mbf).

Gene	Chi-square	P value	Mbf	ωs in Mbf	GO term & major functions
Tbc1d8b	15.48	3.59×10^{-3}	Het clade (Fig 3.5a)	0.1621 (extant heteromyids) 0.4016 (heteromyid ancestor) 0.1456 (outgroups)	renal system process (GO:0003014) Rab GTPase activity; mutations associated with nephrotic syndrome
Oxtr	11.52	0.0289	Het clade - no root (Fig 3.5b)	0.0569 (extant heteromyids) 0.0088 (others)	renal system process (GO:0003014) Oxytocin receptor
Gata2	11.78	0.0305	Het clade – no PM (Fig 3.5c)	0.0545 (kangaroo rats & heteromyid ancestor) 0.0010 (pocket mouse) 0.0146 (outgroups)	inner ear morphogenesis (GO: 0042472) transcription factor

Discussion

The genetic basis of adaptation to extreme environments and specialized lifestyles in no-model organism was poorly understood: The era of genomic sequencing has enabled the use of new methods to study the genetic basis of adaptation. In a study of adaptation to an arid environment, we used both data-driven and hypothesis driven approaches to identify genes that showed a strong signal of positive selection, using estimates of dn/ds (= ω). Using the data driven approach, we identified 18 genes as candidates for adaptive evolution to arid conditions in three heteromyids species, two kangaroo rats and a pocket mouse. We also broadened the analysis to detect convergent evolution across 4 unrelated

arid-adapted rodent taxa, and we identified 7 genes with functions ranging from structural to behavioral, with several involved in nutrient homeostasis and response to environmental stresses. We also used the hypothesis-driven approach to identify candidate genes underlying adaptation in the targeted heteromyid species and found two genes associated with renal functions and one with inner ear development.

Adaptation in kangaroo rats and heteromyids

Osmoregulation is considered as one of the most important adaptations to arid environments. In the 18 candidate genes identified using the data-driven approach, one, Tsc22d2, is linked to the response to osmotic stress. For example, the gene is upregulated in cultured mouse kidney cells exposed to hypertonicity and increases cell survival (Fiol, Mak, & Kültz 2007). This gene was linked to selection across the Heteromyid clade (Figure 3.2d) consistent the arid adaptation being an ancestral trait of heteromyids..

Energy production/regulation is another important factor for the survival of the desert species during droughts and the scarcity of food, and Wu et al (2014) identified genes involved in insulin response and glucose homeostasis being more rapidly evolved in camels than cattle. Among our candidate genes, Sesn3, which was linked to selection in kangaroo rats and their ancestor (Figure 3.3b), is suggested to regulate hepatic insulin sensitivity through mTORC2-Akt Signaling pathway (Tao et al 2015). C2cd5 was linked to selection in kangaroo rats (Figure 3.3a) and has been found necessary for the function of melanocortin-4 receptor (Mc4r) that regulate insulin secretion (Gavini et al 2022; Ni et al 2022). In addition, Raf1 was linked to selection throughout the heteromyids (Figure

3.3d) and regulates basal apoptosis and insulin secretion in human islets (Alejandro et al 2011).

Stress response in general is very important for surviving in arid environments, and Wu et al (2014) also found genes involved in DNA damage and repair and apoptosis under selection in camels. Except for Raf1 that is involved in stress responses and apoptosis, another candidate gene linked to selection in heteromyids (Figure 3.3d), Shf, was suggested to be an adapter in PDGF-receptor signaling, and its overexpression significantly reduced apoptosis in cultured mouse cells (Lindholm et al 2000). Similarly, Wdr48 linked to selection in kangaroo rats and their ancestors (Figure 3.3b) was found to interact with ubiquitin specific peptidase 1 (Usp1) and regulate DNA damage repair (Maria Goncalves et al 2017).

Using the hypothesis driven approach, we identified two genes in the "renal system process" (GO:0003014). Tbc1d8b was linked primarily to selection in the ancestor of heteromyids (Figure 3.5a), and the loss of function mutants of this gene causes nephrotic syndrome by affecting vesicular trafficking (Dorval et al 2019). Another renal-system gene, Oxtr, was linked to selection later in the heteromyids (Figure 3.5b). It encodes an oxytocin receptor that is involved in various signaling pathways that regulate from physiological to behavioral functions (Devost, Wrzal, & Zingg 2008). Slc2a9 as a candidate gene identified by Marra, Romero, & DeWoody (2014) that is involved in urine concentration, was missing in "renal system process" (GO:0003014) or our other alignments, but we previously repeated their test on this gene with our manually trimmed

alignment yet found no evidence of positive selection, suggesting the significant results might be caused by misalignment.

One gene in the "inner ear morphogenesis" (GO: 0042472), Gata2, was identified as a candidate for selection in kangaroo rats and the ancestor of heteromyids but not pocket mouse (Figure 3.5c). Webster (1962) proposed that the large perilymphatic space in kangaroo rats reduces damping and improves resonance to make more peaked auditory spectrum, which could make them more sensitive to the sounds made by their predators, and Gata2 is a zinc finger transcription factor that is essential for the develop of semicircular ducts and perilymphatic space of the inner ear (Haugas et al 2010).

Convergent evolution in arid adapted rodents

In the 7 candidate genes identified in all four of the unrelated arid adapted rodents examined, Pik3c3 is a phosphatidylinositol 3 kinase that regulates macroautophagy in response to glucose starvation (Yuan, Russell, & Guan 2013), consistent with the idea that nutrient deprivation could be a common environmental stress in arid habitats. Another candidate gene, Ino80e, is involved in DNA damage repair. While the known functions of the other 5 genes, as previously mentioned, could not be linked to certain selective agents, and their roles in arid adaptation or other adaptations demands further investigation.

Crh was identified as a candidate gene in jerboa with a very high ω of 2.8668, and it was then tested to have elevated ω of 1.0568 in cactus mouse. It encodes corticotropin releasing hormone, which is a major hormone in response to stress and regulates angiogenesis and inflammation (Nezi, Mastorakos, & Mouslech 2015).

Ambiguity of the Heteromyidae/Geomyidae phylogeny

Heteromyidae and Geomyidae are the two extant families of the superfamily Geomyoidea, with the Heteromyidae further divided into 3 subfamilies, Dipodomyinae (kangaroo rats and mice), Heteromyinae (spiny pocket mice), and Perognathinae (pocket mice), while Geomyidae includes pocket gophers. However, the phylogenic relationship of those taxa has been under debate. Asher et al (2019) used DNA, indels, and morphological data to put kangaroo rats and pocket gophers in a monophyletic group (Bayesian posterior probability=0.93) but with pocket mice and spiny pocket mice in another monophyletic group (Bayesian posterior probability=0.69). Swanson, Oliveros, & Esselstyn (2019) used ultraconserved element (UCE) sequences and found a different grouping, with a spiny pocket mouse and a pocket gopher in a monophyletic group (bootstrap support=0.48) but with a kangaroo rat paraphyletic to them. Previously, Alexander & Riddle (2005) generated a phylogeny of heteromyids using two mtDNA loci but wasn't able to resolve the phylogenic relationship between the three subfamilies, suggesting spiny pocket mice and pocket mice may not be much closer related to each other than to kangaroo rats, and combining the results from Swanson, Oliveros, & Esselstyn (2019), the phylogenetic relationships of these three groups and pocket gophers is under questions.

The species trees we used for PAML tests are based on the current taxonomy of those species, which is, assuming pocket gophers diverged from heteromyids before the divergence of kangaroo rats and pocket mice. Assuming this is true, based on the recent contradictory phylogenies, the ancestral branch of heteromyids from the divergence of

the Geomyid pocket gopher to the split of the kangaroo rats and pocket mice should be very short, which is consistent with the results of our data-driven tests on heteromyid branches, which found no significant difference ω isolated to this branch.

Conclusion

By applying data-driven and hypothesis-driven approaches to genomic data, we have identified genes selected in heteromyids and other arid adapted rodent species. Using hierarchical tests, we were able to link selection to specific lineages, which could help us to better understand the evolution processes of the organisms.

Chapter 3 Figures

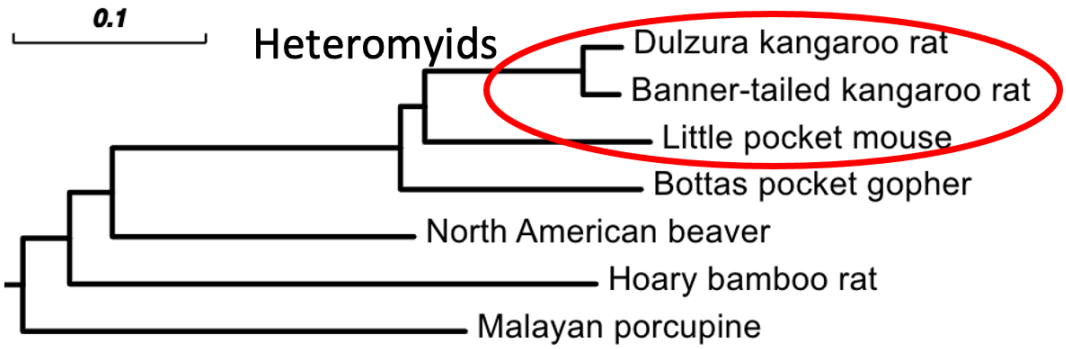

Figure 3.1. Phylogenetic tree of rodent taxa used to identify adaptation in the Heteromyidae. The branch lengths were estimated in PAML using synonymous substitution rates in the concatenate genes but were not used in the analyses. The tree is rooted using human sequences (not shown).

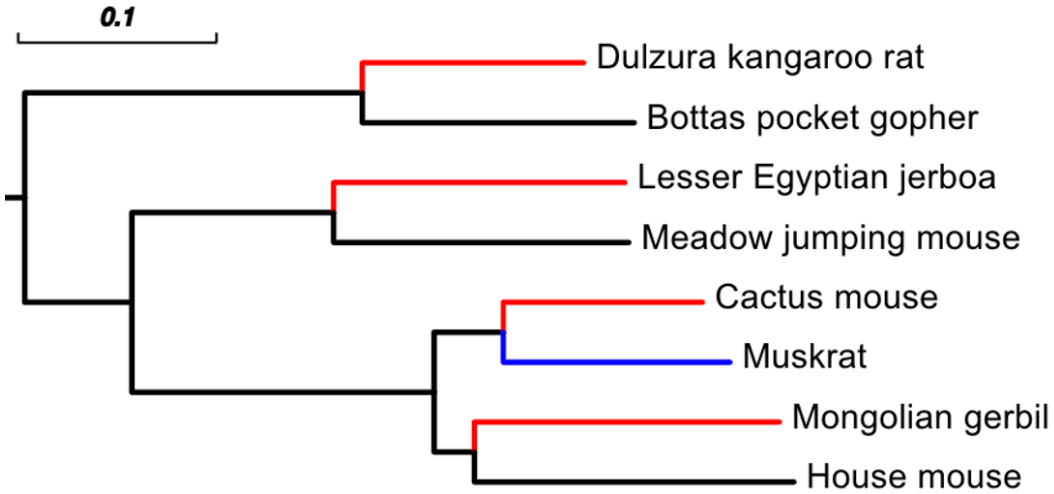

Figure 3.2. Phylogenic tree of 4 pairs of rodent taxa used for convergence study. The branch lengths were estimated in PAML using synonymous substitution rates in the concatenate genes but were not used in the analyses. The tree is rooted using human sequences (not shown). The branches in red are arid adapted, the one in blue is aquatic adapted, and the ones in black are not adapted to special environments.

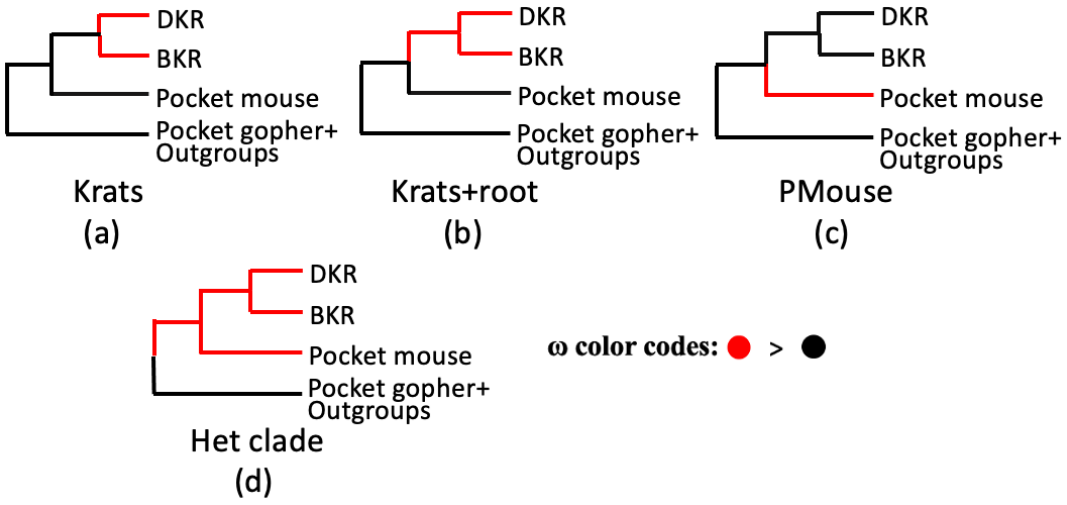

Figure 3.3. Best fit models for the candidate genes in Heteromyidae detected using the data-driven approach. The branches with the same color had the same ω in the best fit models, with red defining the larger ω and black. In the 18 candidate genes, 3 showed (a), 2 showed (b), 2 showed (c), and 11 showed (d).

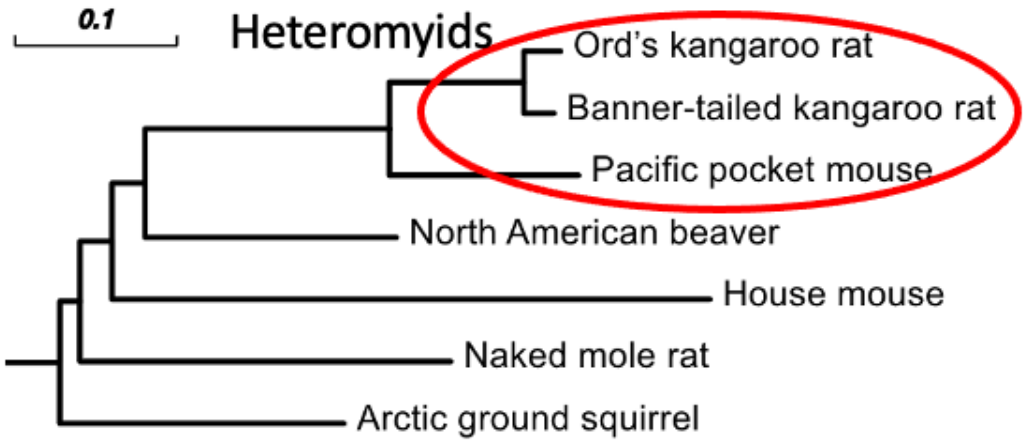

Figure 3.4. Phylogenic tree of the species used in hypothesis-driven approach. The branch lengths were estimated in PAML using synonymous substitution rates in the concatenated genes but were not used in the analyses. The tree is rooted using human sequences (not shown).

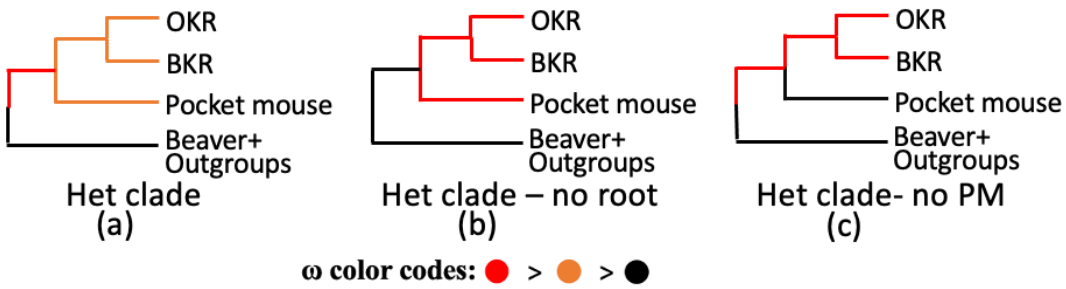

Figure 3.5. Best fit models for the candidate genes in Heteromyidae detected using the hypothesis-driven approach. The branches with the same color had the same ω in the best fit models, with red defining the largest ω, then orange, and then black.

Synthesis

Whole genome sequencing has enabled comparative studies to reveal in detail the nature of the genetic divergence between species and among populations, demographic features, and adaptive changes. Using genomic data, we were able to confirm the species-level separation of a pair of formerly conspecific kangaroo rats, AKR and DKR, with different karyotypes. They showed genetic isolation despite morphological similarities. Using thousands of 4-fold degenerate sites, we estimated their divergence time at ~0.7 Mya. Also, we estimated the historical changes in the effective populations sizes (N_e) and genome wide nucleotide diversity (π). We identified adaptation to elevation as a major factor in their genomic differentiation. Our identification of the signatures of selective sweeps indicated that selection was stronger in AKR compared to DKR and in the high elevation populations of AKR, likely associated with adaptation to cold temperature and hypoxia, while it was stronger in the low elevation in DKR. These adaptive changes were mostly regulatory. We also performed a genome scan for coding genes under positive selection in the heteromyids and in other unrelated arid-adapted rodents using the criterion of an elevated dn/ds ratio, identifying candidate genes involved in glucose homeostasis, response to osmotic stress, and various other functions. In addition, we looked for positive selection on the genes involved in the adaptive traits in the arid adapted heteromyids, including renal function, vision, and hearing, and we identified two candidate genes related to renal function and one to inner ear development.

Reference

Aldea, D., Kokalari, B., Atsuta, Y., Dingwall, H. L., Zheng, Y., Nace, A., ... & Kamberov, Y. G. (2023). Differential modularity of the mammalian Engrailed 1 enhancer network directs sweat gland development. *PLoS genetics*, 19(2), e1010614.

Alejandro, E. U., Lim, G. E., Mehran, A. E., Hu, X., Taghizadeh, F., Pelipeychenko, D., ... & Johnson, J. D. (2011). Pancreatic β-cell Raf-1 is required for glucose tolerance, insulin secretion, and insulin 2 transcription. *The FASEB Journal*, 25(11), 3884.

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome research*, 19(9), 1655-1664.

Alexander, L. F., & Riddle, B. R. (2005). Phylogenetics of the New World rodent family Heteromyidae. *Journal of Mammalogy*, 86(2), 366-379.

Allison, A. C. (1954). Protection afforded by sickle-cell trait against subtertian malarial infection. *British medical journal,* 1(4857), 290.

Asher, R. J., Smith, M. R., Rankin, A., & Emry, R. J. (2019). Congruence, fossils and the evolutionary tree of rodents and lagomorphs. *Royal Society open science*, 6(7), 190387.

Atkins, K. E., & Travis, J. M. J. (2010). Local adaptation and the evolution of species' ranges under climate change. *Journal of Theoretical Biology*, 266(3), 449-457.

Azad, P., Caldwell, A. B., Ramachandran, S., Spann, N. J., Akbari, A., Villafuerte, F. C., ... & Haddad, G. G. (2022). ARID1B, a molecular suppressor of erythropoiesis, is essential for the prevention of Monge's disease. Experimental & molecular medicine, 54(6), 777-787.

Baack, E. J., & Rieseberg, L. H. (2007). A genomic view of introgression and hybrid speciation. *Current opinion in genetics & development*, 17(6), 513-518.

Bademci, G., Lachgar-Ruiz, M., Deokar, M., Zafeer, M. F., Abad, C., Yildirim Baylan, M., ... & Tekin, M. (2022). Mutations in MINAR2 encoding membrane integral NOTCH2-associated receptor 2 cause deafness in humans and mice. *Proceedings of the National Academy of Sciences*, 119(26), e2204084119.

Baler, R., Welch, W. J., & Voellmy, R. (1992). Heat shock gene regulation by nascent polypeptides and denatured proteins: hsp70 as a potential autoregulatory factor. *The Journal of cell biology*, 117(6), 1151-1159.

Beall, C. M., Cavalleri, G. L., Deng, L., Elston, R. C., Gao, Y., Knight, J., ... & Zheng, Y. T. (2010). Natural selection on EPAS1 (HIF2α) associated with low hemoglobin concentration in Tibetan highlanders. *Proceedings of the National Academy of Sciences*, 107(25), 11459-11464.

Bellone, R. R. (2010). Pleiotropic effects of pigmentation genes in horses. Animal genetics, 41, 100-110.

Best, T. L. (1978). Variation in kangaroo rats (genus *Dipodomys*) of the heermanni group in Baja California, Mexico. *Journal of Mammalogy*, 59(1), 160-175.

Best, T. L. (1983). Intraspecific variation in the agile kangaroo rat (*Dipodomys agilis*). *Journal of mammalogy*, 64(3), 426-436.

Best, T. L., Sullivan, R. M., Cook, J. A., & Yates, T. L. (1986). Chromosomal, genic, and morphologic variation in the agile kangaroo rat, *Dipodomys agilis* (Rodentia: Heteromyidae). *Systematic Biology*, 35(3), 311-324.

Beysard, M., Perrin, N., Jaarola, M., Heckel, G., & Vogel, P. (2012). Asymmetric and differential gene introgression at a contact zone between two highly divergent lineages of field voles (Microtus agrestis). *Journal of Evolutionary Biology*, 25(2), 400-408.

Bigham, A., Bauchet, M., Pinto, D., Mao, X., Akey, J. M., Mei, R., ... & Shriver, M. D. (2010). Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data. *PLoS genetics*, 6(9), e1001116.

Bigham, A. W., Julian, C. G., Wilson, M. J., Vargas, E., Browne, V. A., Shriver, M. D., & Moore, L. G. (2014). Maternal PRKAA1 and EDNRA genotypes are associated with birth weight, and PRKAA1 with uterine artery diameter and metabolic homeostasis at high altitude. Physiological Genomics, 46(18), 687-697.

Bigham, A. W. (2016). Genetics of human origin and evolution: high-altitude adaptations. *Current opinion in genetics & development*, 41, 8-13.

Blanga-Kanfi, S., Miranda, H., Penn, O., Pupko, T., DeBry, R. W., & Huchon, D. (2009). Rodent phylogeny revised: analysis of six nuclear genes from all major rodent clades. *BMC evolutionary biology*, 9(1), 1-12.

Bleidorn, C. (2016). Third generation sequencing: technology and its potential impact on evolutionary biodiversity research. Systematics and biodiversity, 14(1), 1-8.

Booker, T. R., Jackson, B. C., & Keightley, P. D. (2017). Detecting positive selection in the genome. *BMC biology*, 15, 1-10.

Brock, R. E., & Kelt, D. A. (2004). Keystone effects of the endangered Stephens' kangaroo rat (Dipodomys stephensi). *Biological conservation*, 116(1), 131-139.

Brown, J. S., Kotler, B. P., Smith, R. J., & Wirtz, W. O. (1988). The effects of owl predation on the foraging behavior of heteromyid rodents. *Oecologia, 76*(3), 408-415.

Brüniche-Olsen, A., Kellner, K. F., Anderson, C. J., & DeWoody, J. A. (2018). Runs of homozygosity have utility in mammalian conservation and evolutionary studies. *Conservation Genetics*, 19(6), 1295-1307.

Buggiotti, L., Yurchenko, A. A., Yudin, N. S., Vander Jagt, C. J., Vorobieva, N. V., Kusliy, M. A., ... & Larkin, D. M. (2021). Demographic history, adaptation, and NRAP convergent evolution at amino acid residue 100 in the world northernmost cattle from Siberia. *Molecular biology and evolution*, 38(8), 3093-3110.

Burri, R., Nater, A., Kawakami, T., Mugal, C. F., Olason, P. I., Smeds, L., ... & Ellegren, H. (2015). Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of Ficedula flycatchers. *Genome research*, 25(11), 1656-1665.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC bioinformatics*, 10, 1-9.

Capanna, E. (1982). *Reproductive isolation between two chromosomal races of Mus musculus in the Rhaetian Alps (Northern Italy).*

Capanna, E., Corti, M., Mainardi, D., Parmigiani, S., & Brain, P. F. (1984). Karyotype and intermale aggression in wild house mice: ecology and speciation. *Behavior genetics*, 14(3), 195-208.

Capanna, E., Corti, M., Nascetti, G., & Bullini, L. (1985). Pre-and post-mating isolating mechanisms in the speciation of the European long-tailed house mouse *Mus musculus domesticus*. *Acta Zoologica Fennica*, 170, 115-120.

Castiglia, R. (2014). Sympatric sister species in rodents are more chromosomally differentiated than allopatric ones: implications for the role of chromosomal rearrangements in speciation. *Mammal Review*, 44(1), 1-4.

Ceballos, G., Arroyo-Cabrales, J., & Ponce, E. (2010). Effects of Pleistocene environmental changes on the distribution and community structure of the mammalian fauna of Mexico. *Quaternary research*, 73(3), 464-473.

Ceballos, F. C., Joshi, P. K., Clark, D. W., Ramsay, M., & Wilson, J. F. (2018). Runs of homozygosity: windows into population history and trait architecture. *Nature Reviews Genetics*, 19(4), 220-234.

Chen, C. (2021, January). Investigation of Potential Inhibitors of Brown Adipose Tissue Induced Thermogenesis. *In 2021 11th International Conference on Bioscience, Biochemistry and Bioinformatics* (pp. 52-60).

Chen, F., Huang, W., Jin, L., Chen, J., & Wang, J. (2011). Spatiotemporal precipitation variations in the arid Central Asia in the context of global warming. *Science China Earth Sciences*, 54, 1812-1821.

Chen, Z., & Narum, S. R. (2021). Whole genome resequencing reveals genomic regions associated with thermal adaptation in redband trout. *Molecular Ecology*, 30(1), 162-174.

Chen, Z., Qiao, F., He, Y., Xie, H., & Qi, D. (2016). Evidence for positive selection on α and β globin genes in pikas and zokor from the Qinghai-Tibetan Plateau. *Gene & Translational Bioinformatics*, 2.

Cheng, H., Concepcion, G. T., Feng, X., Zhang, H., & Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods*, 18(2), 170-175.

Cheng, C., Xu, D. L., Liu, X. B., Bi, S. J., & Zhang, J. (2021). MicroRNA-145-5p inhibits hypoxia/reoxygenation-induced apoptosis in H9c2 cardiomyocytes by targeting ROCK1. *Experimental and Therapeutic Medicine*, 22(2), 1-8.

Cresko, W. A., Amores, A., Wilson, C., Murphy, J., Currey, M., Phillips, P., ... & Postlethwait, J. H. (2004). Parallel genetic basis for repeated evolution of armor loss in Alaskan threespine stickleback populations. *Proceedings of the National Academy of Sciences*, 101(16), 6050-6055.

Csuti, B. A. (1971). Karyotypes of kangaroo rats from southern California. *Journal of Mammalogy*, 52(1), 202-206.

Csuti, B. A. (1979). Current status of the chisel-toothed kangaroo rat in Joshua Tree National Monument. *The Southwestern Naturalist*, 24(3), 530-533.

Cucchi, T., Vigne, J. D., & Auffray, J. C. (2005). First occurrence of the house mouse (*Mus musculus domesticus* Schwarz & Schwarz, 1943) in the Western Mediterranean: a zooarchaeological revision of subfossil occurrences. Biological Journal of the Linnean Society, 84(3), 429-445.

Daly, M., Behrends, P. R., Wilson, M. I., & Jacobs, L. F. (1992). Behavioural modulation of predation risk: moonlight avoidance and crepuscular compensation in a nocturnal desert rodent, Dipodomys merriami. *Animal behaviour*, 44(1), 1-9.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156-2158.

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., ... & Li, H. (2021). Twelve years of SAMtools and BCFtools. *Gigascience*, 10(2), giab008.

Darriba, D., Posada, D., Kozlov, A. M., Stamatakis, A., Morel, B., & Flouri, T. (2020). ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. *Molecular biology and evolution*, 37(1), 291-294.

Deinum, E. E., Halligan, D. L., Ness, R. W., Zhang, Y. H., Cong, L., Zhang, J. X., & Keightley, P. D. (2015). Recent evolution in Rattus norvegicus is shaped by declining effective population size. *Molecular biology and evolution*, 32(10), 2547-2558.

Delahaye, C., & Nicolas, J. (2021). Sequencing DNA with nanopores: Troubles and biases. *PLoS One*, 16(10), e0257521.

Deng, W. D., Shu, W., Yang, S. L., Shi, X. W., & Mao, H. M. (2009). Pigmentation in Black-boned sheep (Ovis aries): association with polymorphism of the MC1R gene. *Molecular Biology Reports*, 36, 431-436.

Devost, D., Wrzal, P., & Zingg, H. H. (2008). Oxytocin receptor signalling. *Progress in brain research*, 170, 167-176.

Dobigny, G., Britton‐Davidian, J., & Robinson, T. J. (2017). Chromosomal polymorphism in mammals: an evolutionary perspective. *Biological Reviews*, 92(1), 1-21.

Döring, A., Gieger, C., Mehta, D., Gohlke, H., Prokisch, H., Coassin, S., ... & Meisinger, C. (2008). SLC2A9 influences uric acid concentrations with pronounced sex-specific effects. *Nature genetics*, 40(4), 430-436.

Dorval, G., Kuzmuk, V., Gribouval, O., Welsh, G. I., Bierzynska, A., Schmitt, A., ... & Antignac, C. (2019). TBC1D8B loss-of-function mutations lead to X-linked nephrotic syndrome via defective trafficking pathways. *The American Journal of Human Genetics*, 104(2), 348-355.

Eichstaedt, C. A., Antao, T., Pagani, L., Cardona, A., Kivisild, T., & Mormina, M. (2014). The Andean adaptive toolkit to counteract high altitude maladaptation: genome-wide and phenotypic analysis of the Collas. *PloS one*, 9(3), e93314.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., ... & Turner, S. (2009). Real-time DNA sequencing from single polymerase molecules. *Science*, 323(5910), 133-138.

Eisenberg, J. F. (1975). The behavior patterns of desert rodents. *In Rodents in desert environments* (pp. 189-224). Dordrecht: Springer Netherlands.

Enattah, N. S., Sahi, T., Savilahti, E., Terwilliger, J. D., Peltonen, L., & Järvelä, I. (2002). Identification of a variant associated with adult-type hypolactasia. *Nature genetics,* 30(2), 233-237.

Engelken, J., Carnero-Montoro, E., Pybus, M., Andrews, G. K., Lalueza-Fox, C., Comas, D., ... & Bosch, E. (2014). Extreme population differences in the human zinc transporter ZIP4 (SLC39A4) are explained by positive selection in Sub-Saharan Africa. *PLoS genetics*, 10(2), e1004128.

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular ecology*, 14(8), 2611-2620.

Exposito-Alonso, M. (2023). Understanding local plant extinctions before it is too late: bridging evolutionary genomics with global ecology. *New Phytologist,* 237(6), 2005-2011.

Gavini, C. K., White, C. R., Mansuy-Aubert, V., & Aubert, G. (2022). Loss of C2 Domain Protein (C2CD5) Alters Hypothalamic Mitochondrial Trafficking, Structure, and Function. *Neuroendocrinology*, 112(4), 324-337.

Fang, X., Nevo, E., Han, L., Levanon, E. Y., Zhao, J., Avivi, A., ... & Wang, J. (2014). Genome-wide adaptive complexes to underground stresses in blind mole rats Spalax. *Nature communications*, 5(1), 1-11.

Fiol, D. F., Mak, S. K., & Kültz, D. (2007). Specific TSC22 domain transcripts are hypertonically induced and alternatively spliced to protect mouse kidney cells during osmotic stress. *The FEBS journal*, 274(1), 109-124.

Fitzpatrick, B. M. (2004). Rates of evolution of hybrid inviability in birds and mammals. *Evolution*, 58(8), 1865-1870.

Fontsere, C., de Manuel, M., Marques-Bonet, T., & Kuhlwilm, M. (2019). Admixture in Mammals and How to Understand Its Functional Implications: On the Abundance of

Gene Flow in Mammalian Species, Its Impact on the Genome, and Roads into a Functional Understanding. *BioEssays*, 41(12), 1900123.

Foote, A. D., Liu, Y., Thomas, G. W., Vinař, T., Alföldi, J., Deng, J., ... & Gibbs, R. A. (2015). Convergent evolution of the genomes of marine mammals. *Nature genetics*, 47(3), 272-275.

Franchini, P., Castiglia, R., & Capanna, E. (2008). Reproductive isolation between chromosomal races of the house mouse Mus musculus domesticus in a parapatric contact area revealed by an analysis of multiple unlinked loci. *Journal of Evolutionary Biology*, 21(2), 502-513.

Friedrich, J., & Wiener, P. (2020). Selection signatures for high‐altitude adaptation in ruminants. *Animal Genetics*, 51(2), 157-165.

Garcia de Leaniz, C., Fleming, I. A., Einum, S., Verspoor, E., Jordan, W. C., Consuegra, S., ... & Quinn, T. P. (2007). A critical review of adaptive genetic variation in Atlantic salmon: implications for conservation. *Biological reviews*, 82(2), 173-211.

Ge, D., Feijó, A., Wen, Z., Lissovsky, A., Zhang, D., Cheng, J., ... & Yang, Q. (2022). Ancient introgression underlying the unusual mito-nuclear discordance and coat phenotypic variation in the Moupin pika. *Diversity and Distributions*.

Graf, J., Voisey, J., Hughes, I., & Van Daal, A. (2007). Promoter polymorphisms in the MATP (SLC45A2) gene are associated with normal human skin color variation. *Human mutation*, 28(7), 710-717.

Gunnarsson, U., Hellström, A. R., Tixier-Boichard, M., Minvielle, F., Bed'Hom, B., Ito, S. I., ... & Andersson, L. (2007). Mutations in SLC45A2 cause plumage color variation in chicken and Japanese quail. *Genetics*, 175(2), 867-877.

Han, F., Lamichhaney, S., Grant, B. R., Grant, P. R., Andersson, L., & Webster, M. T. (2017). Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin's finches. *Genome research*, 27(6), 1004-1015.

Haugas, M., Lilleväli, K., Hakanen, J., & Salminen, M. (2010). Gata2 is required for the development of inner ear semicircular ducts and the surrounding perilymphatic space. *Developmental Dynamics*, 239(9), 2452-2469.

Hecker, N., Sharma, V., & Hiller, M. (2019). Convergent gene losses illuminate metabolic and physiological changes in herbivores and carnivores. *Proceedings of the National Academy of Sciences*, 116(8), 3036-3041.

Henry, P., & Russello, M. A. (2013). Adaptive divergence along environmental gradients in a climate-change-sensitive mammal. *Ecology and Evolution*, 3(11), 3906-3917.

Herman, C. S., & Valone, T. J. (2000). The effect of mammalian predator scent on the foraging behavior of *Dipodomys merriami*. *Oikos*, 91(1), 139-145.

Hoekstra, H. E., Drumm, K. E., & Nachman, M. W. (2004). Ecological genetics of adaptive color polymorphism in pocket mice: geographic variation in selected and neutral genes. Evolution, 58(6), 1329-1341.

Hoekstra, H. E., Krenz, J. G., & Nachman, M. W. (2005). Local adaptation in the rock pocket mouse (*Chaetodipus intermedius*): natural selection and phylogenetic history of populations. *Heredity*, 94(2), 217-228.

Hoekstra, H. E., Hirschmann, R. J., Bundey, R. A., Insel, P. A., & Crossland, J. P. (2006). A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science*, 313(5783), 101-104.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. Scandinavian journal of statistics, 65-70.

Horn, A., Basset, P., Yannic, G., Banaszek, A., Borodin, P. M., Bulatova, N. S., ... & Hausser, J. (2012). (*Sorex araneus*). Evolution: International Journal of Organic Evolution, 66(3), 882-889.

Hotaling, S., Desvignes, T., Sproul, J. S., Lins, L. S., & Kelley, J. L. (2023). Pathways to polar adaptation in fishes revealed by long-read sequencing. *Molecular Ecology*, 32(6), 1381-1397.

Hu, C., Lakshmipathi, J., Binning, E., Hyndman, K. A., Stuart, D., & Kohan, D. E. (2021). Sex-dependent effects of Nephron Ift88 disruption on BP, renal function, and cystogenesis. *Journal of the American Society of Nephrology*, 32(9), 2210-2222.

Jensen-Seaman, M. I., Furey, T. S., Payseur, B. A., Lu, Y., Roskin, K. M., Chen, C. F., ... & Jacob, H. J. (2004). Comparative recombination rates in the rat, mouse, and human genomes. *Genome research*, 14(4), 528-538.

Jones, J. S. J. S., Leith, B. H. & P Rawlings, P. (1977) Polymorphismin Cepaea: a problem with too many solutions? Ann. Rev. EcoL Syst. 8, 109-143

Kawamura, Y., Matsuo, H., Chiba, T., Nagamori, S., Nakayama, A., Inoue, H., ... & Shinomiya, N. (2011). Pathogenic GLUT9 mutations causing renal hypouricemia type 2 (RHUC2). *Nucleosides, Nucleotides and Nucleic Acids*, 30(12), 1105-1111.

Kelt, D. A., Wilson, J. A., Konno, E. S., Braswell, J. D., & Deutschman, D. (2008). Differential responses of two species of kangaroo rat (Dipodomys) to heavy rains: a humbling reappraisal. *Journal of Mammalogy,* 89(1), 252-254.

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*, 116(3), 255-264.

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular ecology resources*, 15(5), 1179-1191.

Korunes, K. L., & Samuk, K. (2021). pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Molecular ecology resources*, 21(4), 1359-1368.

Kousari, M., & Asadi Zarch, M. (2011). Minimum, maximum, and mean annual temperatures, relative humidity, and precipitation trends in arid and semi-arid regions of Iran. *Arabian Journal of Geosciences*, 4.

Kubiak, B. B., Kretschmer, R., Leipnitz, L. T., Maestri, R., de Almeida, T. S., Borges, L. R., ... & de Freitas, T. R. O. (2020). Hybridization between subterranean tuco-tucos (Rodentia, Ctenomyidae) with contrasting phylogenetic positions. *Scientific reports*, 10(1), 1-13.

Kumar, S. M., Yu, H., Edwards, R., Chen, L., Kazianis, S., Brafford, P., ... & Xu, X. (2007). Mutant V600E BRAF increases hypoxia inducible factor-1α expression in melanoma. *Cancer research*, 67(7), 3177-3184.

Lavrenchenko, L. A., & Bulatova, N. S. (2016). The role of hybrid zones in speciation: a case study on chromosome races of the house mouse *Mus domesticus* and common shrew *Sorex araneus*. *Biology Bulletin Reviews*, 6(3), 232-244.

Le Duc, D., & Schöneberg, T. (2016). Adaptation to nocturnality–learning from avian genomes. *BioEssays*, 38(7), 694-703.

Lehman, J. M., Michaud, E. J., Schoeb, T. R., Aydin-Son, Y., Miller, M., & Yoder, B. K. (2008). The Oak Ridge Polycystic Kidney mouse: modeling ciliopathies of mice and men. Developmental dynamics: an official publication of the American Association of Anatomists, 237(8), 1960-1971.

Li, H. (2013). Seqtk: a fast and lightweight tool for processing FASTA or FASTQ sequences. *Github*.

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18), 3094-3100.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. bioinformatics, 25(14), 1754-1760.

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, 475(7357), 493-496.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078-2079.

Lidicker, W. Z. (1960). *An Analysis of Intraspecific Variation in the Kangaroo Rat Dipodomys Merriami*. University of California Press.

Lin, Z., Lu, Y., Yu, G., Teng, H., Wang, B., Yang, Y., ... & Tian, P. (2023). Genome-wide DNA methylation landscape of four Chinese populations and epigenetic variation linked to Tibetan high-altitude adaptation. *Science China Life Sciences*, 1-16.

Lindholm, C. K., Frantz, J. D., Shoelson, S. E., & Welsh, M. (2000). Shf, a Shb-like adapter protein, is involved in PDGF-α-receptor regulation of apoptosis. *Biochemical and Biophysical Research Communications*, 278(3), 537-543.

Lindsay, S. J., Rahbari, R., Kaplanis, J., Keane, T., & Hurles, M. E. (2019). Similarities and differences in patterns of germline mutation between mice and humans. *Nature communications*, 10(1), 1-12.

Linnen, C. R., Kingsley, E. P., Jensen, J. D., & Hoekstra, H. E. (2009). On the origin and spread of an adaptive allele in deer mice. Science, 325(5944), 1095-1098.

Liu, S., Lorenzen, E. D., Fumagalli, M., Li, B., Harris, K., Xiong, Z., ... & Wang, J. (2014). Population genomics reveal recent speciation and rapid evolutionary adaptation in polar bears. *Cell*, 157(4), 785-794.

Malcher, S. M., Pieczarka, J. C., Geise, L., Rossi, R. V., Pereira, A. L., O'Brien, P. C. M., ... & Nagamachi, C. Y. (2017). *Oecomys catherinae* (Sigmodontinae, Cricetidae): Evidence for chromosomal speciation? *PloS one*, 12(7), e0181434.

Mares, M. A. (1993). Desert rodents, seed consumption, and convergence. *BioScience*, 43(6), 372-379.

Maria Goncalves, J., Mariela Rodriguez Cordeiro, M., & Riet Correa Rivero, E. (2017). The role of the complex USP1/WDR48 in differentiation and proliferation processes in cancer stem cells. *Current stem cell research & therapy*, 12(5), 416-422.

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). *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*, 7(4), 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. Molecular Ecology, 23(11), 2699-2711.

Mayro, B., Hoj, J. P., Cerda-Smith, C. G., Hutchinson, H. M., Caminear, M. W., Thrash, H. L., ... & Pendergast, A. M. (2023). ABL kinases regulate the stabilization of HIF-1α and MYC through CPSF1. Proceedings of the National Academy of Sciences, 120(16), e2210418120.

Metcalf, A. E., Nunney, L., & Hyman, B. C. (2001). Geographic Patterns of Genetic Differentiation Within the Restricted Range of the Endangered Stephens' Kangaroo Rat Dipodomys Stephensi. *Evolution* 55(6), 1233-1244.

Miller, C. T., Beleza, S., Pollen, A. A., Schluter, D., Kittles, R. A., Shriver, M. D., & Kingsley, D. M. (2007). cis-Regulatory changes in Kit ligand expression and parallel evolution of pigmentation in sticklebacks and humans. *Cell*, 131(6), 1179-1189.

Mirsky, E. N. (1971). Olfactory discrimination between *Dipodomys merriami* and *Dipodomys panamintinus* (Doctoral dissertation, California State University, Northridge).

Morano, K. A. (2007). New tricks for an old dog: the evolving world of Hsp70. *Annals of the New York Academy of Sciences*, 1113(1), 1-14.

Mussmann, S. M., Douglas, M. R., Chafin, T. K., & Douglas, M. E. (2020). A dmix P ipe: population analyses in A dmixture for non-model organisms. *BMC bioinformatics*, 21(1), 1-9.

Nachman, M. W. (1992). Geographic patterns of chromosomal variation in South American marsh rats, *Holochilus brasiliensis* and *H. vulpinus*. *Cytogenetic and Genome Research*, 61(1), 10-16.

Nachman, M. W., Hoekstra, H. E., & D'Agostino, S. L. (2003). The genetic basis of adaptive melanism in pocket mice. *Proceedings of the National Academy of Sciences*, 100(9), 5268-5273.

Nagy, K. A., & Gruchacz, M. J. (1994). Seasonal water and energy metabolism of the desert-dwelling kangaroo rat (Dipodomys merriami). *Physiological Zoology,* 67(6), 1461-1478.

Nanaei, H., Cai, Y., Alshawi, A., Wen, J., Hussain, T., Fu, W. W., ... & Jiang, Y. (2023). Genomic analysis of indigenous goats in Southwest Asia reveals evidence of ancient adaptive introgression related to desert climate. *Zoological research*, 44(1), 20-29.

Natarajan, C., Hoffmann, F. G., Weber, R. E., Fago, A., Witt, C. C., & Storz, J. F. (2016). Predictable convergence in hemoglobin function has unpredictable molecular underpinnings. *Science*, 354(6310), 336-339.

Nezi, M., Mastorakos, G., & Mouslech, Z. (2015). Corticotropin releasing hormone and the immune/inflammatory response.

Ni, Z., Wang, Y., Shi, C., Zhang, X., Gong, H., & Dong, Y. (2022). Islet MC4R Regulates PC1/3 to Improve insulin secretion in T2DM mice via the cAMP and β-arrestin-1 pathways. *Applied Biochemistry and Biotechnology*, 194(12), 6164-6178.

Nunney, L. (2016). The effect of neighborhood size on effective population size in theory and in practice. *Heredity*, 117(4), 224-232.

Nunney, L., & Schuenzel, E. L. (2006). Detecting natural selection at the molecular level: a reexamination of some "classic" examples of adaptive evolution. *Journal of molecular evolution*, 62(2), 176-195.

Pan, S., Zhang, T., Rong, Z., Hu, L., Gu, Z., Wu, Q., ... & Zhan, X. (2017). Population transcriptomes reveal synergistic responses of DNA polymorphism and RNA expression to extreme environments on the Qinghai–Tibetan Plateau in a predatory bird. *Molecular ecology*, 26(11), 2993-3010.

Pardo-Diaz, C., Salazar, C., Baxter, S. W., Merot, C., Figueiredo-Ready, W., Joron, M., ... & Jiggins, C. D. (2012). Adaptive introgression across species boundaries in Heliconius butterflies. *PLoS Genet*, 8(6), e1002752.

Parker, J., Tsagkogeorga, G., Cotton, J. A., Liu, Y., Provero, P., Stupka, E., & Rossiter, S. J. (2013). Genome-wide signatures of convergent evolution in echolocating mammals. *Nature*, 502(7470), 228-231.

Pejo, M., & Tomasco, I. H. (2021). Adaptive evolution of β-globin gene in subterranean in South America octodontid rodents. *Gene*, 772, 145352.

Penumatsa, K. C., Singhal, A. A., Warburton, R. R., Bear, M. D., Bhedi, C. D., Nasirova, S., ... & Toksoz, D. (2022). Vascular smooth muscle ROCK1 contributes to hypoxia-induced pulmonary hypertension development in mice. *Biochemical and Biophysical Research Communications*, 604, 137-143.

Perri, L. M., & Randall, J. A. (1999). Behavioral mechanisms of coexistence in sympatric species of desert rodents, *Dipodomys ordii* and *D. merriami*. *Journal of Mammalogy*, 80(4), 1297-1310.

Price, M. V., Kelly, P. A., & Goldingay, R. L. (1994). Distances moved by Stephens' kangaroo rat (*Dipodomys stephensi* Merriam) and implications for conservation. Journal of *Mammalogy*, 75(4), 929-939.

Price, M. V., Waser, N. M., & McDonald, S. A. (2000). Elevational distributions of kangaroo rats (genus Dipodomys): long-term trends at a Mojave Desert site. *The American Midland Naturalist*, 144(2), 352-361.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., ... & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American journal of human genetics*, 81(3), 559-575.

Ohta, T., Takahashi, T., Shibuya, T., Amita, M., Henmi, N., Takahashi, K., & Kurachi, H. (2012). Inhibition of the Rho/ROCK pathway enhances the efficacy of cisplatin through the blockage of hypoxia-inducible factor-1α in human ovarian cancer cells. *Cancer biology & therapy*, 13(1), 25-33.

Qiu, X. B., Shao, Y. M., Miao, S., & Wang, L. (2006). The diversity of the DnaJ/Hsp40 family, the crucial partners for Hsp70 chaperones. *Cellular and Molecular Life Sciences CMLS*, 63, 2560-2570.

Randall, J. A. (2013). olfactory communication in the Merriam's kangaroo rat, *Dipodomys*. *Chemical Signals in Vertebrates* 3, 321.

Rebar, C. E. (1995). Ability of *Dipodomys merriami* and *Chaetodipus intermedios* to Locate Resource Distributions. *Journal of Mammalogy*, 76(2), 437-447.

Roach, N. 2018. *Dipodomys stephensi*. The IUCN Red List of Threatened Species 2018: e.T6682A22228640. https://dx.doi.org/10.2305/IUCN.UK.2018-2.RLTS.T6682A22228640.en. Downloaded on 21 May 2021.

Robertson, C. E., & Wilsterman, K. (2020). Developmental and reproductive physiology of small mammals at high altitude: challenges and evolutionary innovations. *Journal of Experimental Biology*, 223(24), jeb215350.

Rosenblum, E. B., Hoekstra, H. E., & Nachman, M. W. (2004). Adaptive reptile color variation and the evolution of the Mc1r gene. Evolution, 58(8), 1794-1808.

Savolainen, O., Lascoux, M., & Merilä, J. (2013). Ecological genomics of local adaptation. *Nature Reviews Genetics*, 14(11), 807-820.

Sayers, E. (2010). A General Introduction to the E-utilities. *Entrez Programming Utilities Help [Internet]. Bethesda (MD): National Center for Biotechnology Information (US)*.

Schweizer, R. M., Jones, M. R., Bradburd, G. S., Storz, J. F., Senner, N. R., Wolf, C., & Cheviron, Z. A. (2021). Broad concordance in the spatial distribution of adaptive and neutral genetic variation across an elevational gradient in deer mice. *Molecular biology and evolution*, 38(10), 4286-4300.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., ... & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology*, 7(1), 539.

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. *Molecular biology and evolution*, 36(10), 2227-2237.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19), 3210-3212.

Shurtliff, Q. R. (2013). Mammalian hybrid zones: a review. *Mammal Review*, 43(1), 1-21.

Sjodin, B. M., & Russello, M. A. (2022). Comparative genomics reveals putative evidence for high-elevation adaptation in the American pika (*Ochotona princeps*). *G3*, 12(11), jkac241.

Smit, A. F. A., Hubley, R., & Green, P. (2015). RepeatMasker Open-4.0. 2013–2015.

Stock, A. D. (1974). Chromosome evolution in the genus Dipodomys and its taxonomic and phylogenetic implications. *Journal of Mammalogy*, 505-526.

Storz, J. F. (2016). Hemoglobin–oxygen affinity in high-altitude vertebrates: is there evidence for an adaptive trend? *Journal of Experimental Biology*, 219(20), 3190-3203.

Storz, J. F., Sabatino, S. J., Hoffmann, F. G., Gering, E. J., Moriyama, H., Ferrand, N., ... & Nachman, M. W. (2007). The molecular basis of high-altitude adaptation in deer mice. *PLoS genetics*, 3(3), e45.

Suarez-Gonzalez, A., Hefer, C. A., Lexer, C., Cronk, Q. C., & Douglas, C. J. (2018). Scale and direction of adaptive introgression between black cottonwood (Populus trichocarpa) and balsam poplar (P. balsamifera). *Molecular ecology*, 27(7), 1667-1680.

Subramanian, B., Gao, S., Lercher, M. J., Hu, S., & Chen, W. H. (2019). Evolview v3: a webserver for visualization, annotation, and management of phylogenetic trees. *Nucleic acids research,* 47(W1), W270-W275

Sullivan, R. M., & Best, T. L. (1997). Systematics and Morphologic Variation in Two Chromosomal Forms of the Agile Kangaroo Rat (Dipodomys agius). *Journal of mammalogy*, 78(3), 775-797.

Suyama, M., Torrents, D., & Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic acids research*, 34(suppl_2), W609-W612.

Swanson, B. J. (2001). *An empirical evaluation of effective population size estimators* (Doctoral dissertation, Purdue University).

Swanson, M. T., Oliveros, C. H., & Esselstyn, J. A. (2019). A phylogenomic rodent tree reveals the repeated evolution of masseter architectures. *Proceedings of the Royal Society B*, 286(1902), 20190672.

Swofford, D. L. (2002). Phylogenetic analysis using parsimony (* and other methods).

Tao, R., Xiong, X., Liangpunsakul, S., & Dong, X. C. (2015). Sestrin 3 protein enhances hepatic insulin sensitivity by direct activation of the mTORC2-Akt signaling. *Diabetes*, 64(4), 1211-1223.

Theron, E., Hawkins, K., Bermingham, E., Ricklefs, R. E., & Mundy, N. I. (2001). The molecular basis of an avian plumage polymorphism in the wild: a melanocortin-1-receptor point mutation is perfectly associated with the melanic plumage morph of the bananaquit, Coereba flaveola. *Current Biology*, 11(8), 550-557.

Thorpe, R. S., Barlow, A., Malhotra, A., & Surget-Groba, Y. (2015). Widespread parallel population adaptation to climate variation across a radiation: implications for adaptation to climate change. *Molecular Ecology*, 24(5), 1019-1030.

Tian, R., Wang, Z., Niu, X., Zhou, K., Xu, S., & Yang, G. (2016). Evolutionary genetics of hypoxia tolerance in cetaceans during diving. *Genome biology and evolution*, 8(3), 827-839.

Tiffin, P., & Ross-Ibarra, J. (2014). Advances and limits of using population genetics to understand local adaptation. *Trends in ecology & evolution*, 29(12), 673-680.

Tishkoff, S. A., Reed, F. A., Ranciaro, A., Voight, B. F., Babbitt, C. C., Silverman, J. S., ... & Deloukas, P. (2007). Convergent adaptation of human lactase persistence in Africa and Europe. *Nature genetics*, 39(1), 31-40.

Tomasco, I. H., Boullosa, N., Hoffmann, F. G., & Lessa, E. P. (2017). Molecular adaptive convergence in the α-globin gene in subterranean octodontid rodents. *Gene*, 628, 275-280.

Uchimura, A., Higuchi, M., Minakuchi, Y., Ohno, M., Toyoda, A., Fujiyama, A., ... & Yagi, T. (2015). Germline mutation rates and the long-term phenotypic effects of mutation accumulation in wild-type laboratory mice and mutator mice. *Genome research*, 25(8), 1125-1134.

Vincze, E., Papp, S., Preiszner, B., Seress, G., Bókony, V., & Liker, A. (2016). Habituation to human disturbance is faster in urban than rural house sparrows. *Behavioral Ecology*, 27(5), 1304-1313.

van't Hof, A. E., Edmonds, N., Dalíková, M., Marec, F., & Saccheri, I. J. (2011). Industrial melanism in British peppered moths has a singular and recent mutational origin. Science, 332(6032), 958-960.

Wang, P., Ha, A. Y., Kidd, K. K., Koehle, M. S., & Rupert, J. L. (2010). A variant of the endothelial nitric oxide synthase gene (NOS3) associated with AMS susceptibility is less common in the Quechua, a high altitude Native population. *High altitude medicine & biology*, 11(1), 27-30.

Webster, D. B. (1962). A function of the enlarged middle-ear cavities of the kangaroo rat, Dipodomys. *Physiological Zoology*, 35(3), 248-255.

Webster, D. B., & Webster, M. (1971). Adaptive value of hearing and vision in kangaroo rat predator avoidance. Brain, Behavior and Evolution, 4(4), 310-322.

Webster, D. B., & Webster, M. (1984). The specialized auditory system of kangaroo rats. In *Contributions to sensory physiology* (Vol. 8, pp. 161-196). Elsevier.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, 1358-1370.

Wilsterman, K., Moore, E. C., Schweizer, R. M., Cunningham, K., Good, J. M., & Cheviron, Z. A. (2023). Adaptive structural and functional evolution of the placenta protects fetal growth in high-elevation deer mice. *Proceedings of the National Academy of Sciences*, 120(25), e2218049120.

Witt, K. E., & Huerta-Sánchez, E. (2019). Convergent evolution in human and domesticate adaptation to high-altitude environments. *Philosophical Transactions of the Royal Society B*, 374(1777), 20180235.

Wong, E. L., Hiscock, S. J., & Filatov, D. A. (2022). The role of interspecific hybridisation in adaptation and speciation: Insights from studies in Senecio. *Frontiers in Plant Science*, 13, 907363.

Wright, S. (1949). The genetical structure of populations. *Annals of eugenics*, 15(1), 323-354.

Wu, H., Guang, X., Al-Fageeh, M. B., Cao, J., Pan, S., Zhou, H., ... & Wang, J. (2014). Camelid genomes reveal evolution and adaptation to desert environments. *Nature communications*, 5(1), 5188.

Xiang, K., Ouzhuluobu, Peng, Y., Yang, Z., Zhang, X., Cui, C., ... & Su, B. (2013). Identification of a Tibetan-specific mutation in the hypoxic gene EGLN1 and its contribution to high-altitude adaptation. *Molecular biology and evolution*, 30(8), 1889-1898.

Yamamoto, H., Williams, E. G., Mouchiroud, L., Cantó, C., Fan, W., Downes, M., ... & Auwerx, J. (2011). NCoR1 is a conserved physiological modulator of muscle mass and oxidative function. *Cell*, 147(4), 827-839.

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and evolution*, 24(8), 1586-1591.

Ye, Y., Liu, H., Zhang, F., & Hu, F. (2019). mTOR signaling in Brown and Beige adipocytes: implications for thermogenesis and obesity. *Nutrition & metabolism*, 16, 1-14.

Yuan, H. X., Russell, R. C., & Guan, K. L. (2013). Regulation of PIK3C3/VPS34 complexes by MTOR in nutrient stress-induced autophagy. *Autophagy*, 9(12), 1983-1995.

Zerilli, M., Zito, G., Martorana, A., Pitrone, M., Cabibi, D., Cappello, F., ... & Rodolico, V. (2010). BRAFV600E mutation influences hypoxia-inducible factor-1α expression levels in papillary thyroid cancer. *Modern Pathology*, 23(8), 1052-1060.

Zhang, J., Kobert, K., Flouri, T., & Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, 30(5), 614-620.