The Genomic Basis of Desert Adaptation in Rodents

By

Noëlle K. J. Bittner

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Integrative Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Michael W. Nachman, Chair
Professor Erica B. Rosenblum
Professor Rasmus Nielsen

Summer 2020

ABSTRACT

The Genomic Basis of Desert Adaptation in Rodents

By

Noëlle Kristen Jeffery Bittner

Doctor of Philosophy in Integrative Biology

University of California, Berkeley

Professor Michael W. Nachman, Chair

Understanding the genomic architecture of complex adaptive traits in natural populations has long been a central goal in evolutionary biology. Of particular interest is the extent of the "genetic toolbox," that is: how many evolutionarily viable solutions are there to a common evolutionary problem? To get at this question, I looked at convergent adaptation to desert environments across the order Rodentia. In mammals, the problem of living in an extremely arid environment with limited access to free water has been solved in a variety of different ways. For granivorous desert rodents, many have developed ultra-efficient osmoregulatory systems, including highly modified kidneys, to handle the problem of salt and water homeostasis. Here, I examined the genomic basis of adaptation across multiple evolutionary timescales.

In chapter one, I studied the recent invasion of a North American desert by the house mouse, *Mus musculus*. I compared this desert population with a non-desert population and examined organismal phenotypes and gene expression profiles in the lab for experimentally dehydrated mice and for control mice. I discovered significant differences in the response to water stress in these two populations. Mice derived from a desert population showed a less extreme response to water stress, suggesting they may have adapted to live in deserts. Non-desert mice showed shifts towards desert-like gene expression after water stress, consistent with adaptive plasticity. Further, I identified candidate genes underlying desert adaptation in this population.

In chapter two, I leveraged three phylogenetically distinct species pairs, one a desert specialist and the other a mesic representative, from three families of rodents over ~70 million years of divergence and found convergent patterns of evolution using sequences of expressed genes (RNA-seq data). First, I looked for convergent shifts in gene expression associated with desert living in each species pair and identified genes displaying this pattern. Next I used multiple methods to test for aspects of convergence at the DNA sequence level and found that few genes showed both convergence in DNA expression and convergence in protein structure. I found that a greater proportion of genes tested show convergent patterns in gene expression suggesting that changes in the regulation of genes

may have a greater impact than changes in protein structure during adaptation to desert environments.

In chapter three, I sequenced and assembled *de novo* the genome of the rock pocket mouse, *Chaetodipus intermedius* (Heteromyidae) using MaSuRCA. This species, found in the Sonoran and Chihuahuan deserts of North America, is a desert specialist and a model for understanding the genetics of coat color variation in wild populations. This genome is a resource to be used to further understand the genomics of adaptation in this unique lineage.

DEDICATION

To the place where I first fell in love with the desert, the Sonoran desert, and its endless forms most beautiful and most wonderful.

ACKNOWLEDGEMENTS

This work was generously funded by The Louise Kellogg Fund and David and Marvalee Wake Fund through the Museum of Vertebrate Zoology, the UC Berkeley Graduate Division, and the NSF in the form of a Doctoral Dissertation Improvement Grant. This work was also supported by NIH grants to Michael Nachman.

First, I would like to thank my advisor, Michael Nachman, for his support and wisdom. I have learned so much from him as a teacher, mentor, mammalogist, and fellow human, and for that I will be forever grateful. I would like to thank my committee members, Rasmus Nielsen and Bree Rosenblum, for their thoughtful comments, guidance and support.

I am extremely grateful for the mentorship and wisdom I've received from so many including Jim Patton, Jim McGuire, Rauri Bowie, David Wake, Eileen Lacey, Alan Shabel, Rosemary Gillespie, Noah Whiteman, and for the tireless enthusiasm of Mohamed Noor. Special thanks to my undergraduate mentor, Stephen Tilley, for showing me the ropes of academia (and salamandering) and to my high school biology teacher Judy Kemlitz.

I am endlessly thankful to have had so many brilliant lab mates throughout my long tenure in the Nachman lab: Pennie Rabago, Jeremy Jonas, Dana Lin, Michael Sheehan, Andreas Chavez, Kathleen Ferris, Christopher Emerling, Gideon Bradburd, Andrew Moeller, Jessica Castillo Vardaro, Libby Beckman, Erin Voss, David Manahan, Kennedy Agwamba, Sylvia Durkin, and honorary member Rachel Thayer. Extra special thanks for the friendship of Megan Phifer-Rixey who taught me (among other things) how to catch my first *Chaetodipus*, Polly Campell who also taught me (among other things) how to mouse wrangle, Sarah Banker a great adventure buddy, and Mallory Ballinger for all the sanity checks. I would never have gotten to the finish line without the endless encouragement, optimism, and creativity of Taichi Suzuki and the genius, love, and late night support of Katya Mack.

My graduate career would have been much less full without my fellow graduate students including Zach Hanna, Phil Georgakakos, Dave Armitage, Luke Bloch, Alexander Stubbs, Andy Gloss, and Brianna McTeague among others too numerous to list. I have learned so much from the people and specimens of the Museum of Vertebrate Zoology and consider myself so lucky to have had the opportunity to be a part of it. This research could not have been completed without the expertise and encouragement of Lydia Smith and Ke Bi, and for that they are very much appreciated.

Lastly, there are not enough words to express my gratitude for my family: to my parents, Carol and Tom, who have always encouraged my love of creatures and the natural world and who have gone above and beyond to help me get wherever I've wanted to go; to my sister, Fudgie, who challenges me and makes me better; to my Uncle Kenny, who always supported my search for knowledge; to the whole of the Bittner and Jeffery families; and to my wonderful chosen family who keep me going and who bring my life so much joy.

Chapter 1

Plasticity in gene expression facilitates invasion of the desert environment in house mice

ABSTRACT

Understanding how organisms adapt to new environments is a key problem in evolution, yet it remains unclear whether phenotypic plasticity generally facilitates or hinders this process. Here we studied evolved and plastic responses to water-stress in lab-born descendants of wild house mice (*Mus musculus domesticus*) collected from desert and non-desert environments and measured organismal phenotypes and gene expression under normal and water-stressed conditions. After many generations in the lab, desert mice consumed significantly less water than mice from other localities, indicating that this difference has a genetic basis. Under water-stress, desert mice lost less weight than non-desert mice, and exhibited differences in blood chemistry related to osmoregulatory function. Gene expression in the kidney revealed evolved differences between mice from different environments as well as plastic responses between hydrated and dehydrated mice. Desert mice showed reduced expression plasticity under water-stress compared to non-desert mice. Importantly, non-desert mice generally showed shifts towards desert-like expression under water-stress, consistent with adaptive plasticity. Finally, patterns of gene expression identified several candidate genes for adaptation to the desert, including *Aqp1* and *Apoe*. These findings provide evidence for local adaptation after a recent invasion and suggest that adaptive plasticity may have facilitated colonization of the desert environment.

INTRODUCTION

Understanding the origin and genetic architecture of complex traits associated with local adaptation is a central goal of evolutionary biology. One ongoing debate concerns the extent to which phenotypic plasticity may facilitate or constrain adaptation to new environments (Baldwin 1896; Price et al. 2003; Ghalambor et al. 2007; Levis and Pfennig 2016). For example, adaptive plasticity, defined as an environmentally induced phenotypic change that brings individuals closer to the local optimum, may enable organisms to invade new environments. Subsequent genetically encoded changes in the same direction as the plastic changes may then accrue, bringing individuals even closer to the optimum, as seen for coloration in lizards living on dark substrates (Corl et al. 2018). Conversely, plastic changes may be non-adaptive, moving individuals farther from the local optimum. In such cases, selection is expected to favor genetic changes underlying phenotypes that go in the opposite direction of the plastic change and thereby bring the individual closer to the optimum. This pattern of non-adaptive plasticity is seen for gene expression changes in

guppies reared in the absence of predators (Ghalambor et al. 2015). Which of these two outcomes is most likely remains unclear and may depend both on the phenotype in question and the environmental heterogeneity to which populations have been exposed (e.g. Huang and Agrawal 2016).

House mice (*Mus musculus domesticus*) provide an opportunity to study plastic and evolved changes in the context of adaptation to novel environments. House mice are native to western Europe but were recently introduced to the Americas with European colonization, approximately 400-600 generations ago (Phifer-Rixey and Nachman 2015). In this short time, they have colonized a wide variety of different environments. In eastern North America, house mice show strong evidence of local adaptation for several complex phenotypes such as body size, activity, and nest-building behavior (e.g. Lynch 1992; Mack, et al. 2018; Phifer-Rixey et al. 2018).

In the Sonoran Desert of North America, house mice must contend with low to seasonally-absent water as well as extreme heat. Although house mice are human commensals, they frequently live in sheds, grain storage areas, barns, and other habitats where they are not well shielded from the environment. They can also live in situations where they are not associated with humans (Sage 1981). House mouse urine concentration, a metric often associated with specialization to xeric environments, is very high (Haines et al. 1973) and falls within the range of many known desert specialists (Beuchat 1990). Previous experiments of wild mice brought into the lab have found that mice can survive beyond 14 days without access to free water on a diet of dried seeds (Koford 1968). In other experiments with varying relative humidity, mice survived up to 41 days without free water (Haines and Schmidt-Nielsen 1967).

Previous studies have examined the behavioral, morphological, and physiological adaptations that allow desert mammals to persist under xeric conditions (reviewed in Schmidt-Nielsen 1964; Donald and Pannabecker 2015). Recent work has also begun to identify some of the genes associated with these phenotypes (Marra et al. 2014; Wu et al. 2014; MacManes 2017; Giorello et al. 2018). Far less is known about the role of phenotypic plasticity in the context of desert adaptation. Since all mice, including those living in more mesic environments, occasionally go through periods of water stress, selection may have favored plastic responses that enable mice to survive periods of water shortage (i.e. adaptive plasticity). Here we are interested in whether the evolved differences between mice from more mesic environments and mice from more xeric environments mirror the plastic responses seen within populations, both for organismal level phenotypes and for gene expression in the kidney. Changes in gene expression in the kidney may also help identify candidate genes for adaptation to xeric conditions.

To assess the contribution of plastic and evolved changes to a xeric environment, we studied lab-born descendants of wild mice from two populations, one from Edmonton, Canada and one from Tucson, Arizona. While neither population experiences high precipitation, the annual precipitation in Edmonton is 57% more than in Tucson, and these two locations differ dramatically in average temperature. We address four primary questions. (1) Do progeny of wild mice from these two populations exhibit phenotypic

differences when reared in a common laboratory setting, indicating that the phenotypes have a genetic basis? (2) Do these same phenotypes exhibit plastic (non-genetic) changes when mice are exposed to water limitation? (3) Are plastic changes generally in the same or opposite direction as the evolved changes? (4) Do gene co-expression networks identify sets of genes and corresponding phenotypes that underlie adaptation to xeric conditions? We find that in a common environment, Tucson mice drink less water than Edmonton mice and differ in blood chemistry and gene expression in the kidney. These same traits exhibit significant plasticity when mice are fully hydrated compared to mice under water stress, and evolved differences are generally in the same direction as plastic differences, both for gene expression and for organism-level phenotypes. Finally, co-expression networks identify groups of genes that likely underlie adaptation to xeric conditions. These findings suggest an important role for adaptive plasticity in the colonization of the desert environment and stand in contrast to some recent studies documenting non-adaptive plasticity in gene expression (e.g. Ghalambor et al. 2015).

MATERIALS AND METHODS

Mice

To assess whether mice from the Sonoran Desert differ in water consumption compared to mice from other habitats, we used wild-derived mouse lines developed in our lab from a range of localities in the Americas. Wild house mice were caught from five populations in different habitats and used to create inbred lines through sib-sib mating over multiple generations. The five localities were Tucson, AZ, USA (TUC); Edmonton, Alberta, Canada (EDM); Gainesville, FL, USA (GAI); Saratoga Springs, NY, USA (SAR); and Manaus, Amazonas, Brazil (MAN). In nearly all cases, lines were established from unrelated individuals. Lines were maintained in the laboratory for 6-19 generations on a diet of standard mouse chow. All mice were handled in accordance with a UC Berkeley Animal Care and Use protocol (protocol AUP-2016-03-8548-1).

Measuring water consumption

To quantify differences among populations, we measured water consumption over 72 hours in 163 adult males representing 45 different inbred lines (Table S1). In total, 40 individuals were measured from Tucson, 24 from Gainesville, 48 from Edmonton, 23 from Saratoga Springs, and 28 from Manaus. Mice were between 90 and 200 days of age and were housed singly in cages at 23°C with a 10 hour dark and 14 hour light cycle on standard Teklad Global rodent chow (18% protein, 6% fat). Body weight and amount of water consumed after 72 hours were recorded, and relative water consumption (RWC) was calculated (grams of water consumed per gram of mouse). Relative water consumption was used as a metric due to the population level variation in body weight.

Measuring phenotypic plasticity

To study evolved and plastic responses to xeric conditions, we chose one wild-derived inbred line each from Tucson and Edmonton [Tucson: TUSA4xA8 (TUCC/Nach), Edmonton: TAS111x165 (EDMA/Nach)]. These lines were chosen because they showed large differences in water consumption as well as little variance within lines. Male littermates from these lines were assigned at random to either control or water restriction treatments

and housed individually post-weaning with water *ad libitum*. After 90 days of age, mice were weighed and phenotyped for relative water consumption. Following phenotyping (average age = 99 days), mice assigned to the water restriction treatment (Edmonton n=11, Tucson n=7) were restricted from all water consumption for 72 hours. Mice assigned to the control treatment (Edmonton n=10, Tucson n=6) were maintained with water *ad libitum*. All mice were weighed every 24 hours and monitored for markers of drastically declining health. All animal care was conducted in accordance with procedures approved by the UC Berkeley Animal Care and Use Committee (protocol AUP-2017-05-9940). After sacrifice with isofluorane, left kidney, liver, and caecum were immediately removed and stored in RNAlater according to manufacturer's instructions. The right kidney was weighed and stored in 10% neutral buffered formalin for morphological analysis at the UC Davis Comparative Pathology Laboratory. Five hydrated mice per population (ten mice total) were phenotyped for renal cortex thickness and papilla thickness. Blood was extracted from the heart and body cavity using a syringe and centrifuged in BD SST Microtainer tubes to extract serum. Levels of blood urea nitrogen (BUN), total protein, creatinine, chloride, potassium, and sodium levels were analyzed at the UC Davis Comparative Pathology Laboratory for 20 mice (five individuals per population per treatment). These serum solutes were measured to quantify kidney health and glomerular function in treated and control samples.

mRNA library preparation and sequencing

RNA was extracted from half of a kidney preserved in RNAlater from twenty mice total, (five mice per population per treatment), using the MoBio Laboratories Powerlyzer Ultraclean Tissue & Cells RNA Isolation Kit. RNA libraries were prepared using the KAPA Hyper Prep Kit and then pooled and sequenced across two lanes of 100bp PE Illumina HiSeq4000 at the Vincent J. Coates Genomics Sequencing Center at UC Berkeley.

mRNA read mapping and quantification of gene expression

Reads were trimmed for quality and adaptor contamination with Trimmomatic v0.36 (Bolger et al. 2014) and mapped to the *Mus musculus* reference genome (GRCm38/mm10) using STAR v2.6.0c (Dobin et al. 2013). Reads overlapping exons were counted using the program HTSeq 0.6.1 (Anders et al. 2015) to estimate per-gene mRNA abundance. We removed genes with a mean fewer than ten reads across samples from additional analyses. The R package DESeq2 (Love et al. 2014) was used to test for differential expression between (1) treatments within each population, and (2) populations within each treatment. Genes were retained as significant at a false-discovery rate of 5%.

Gene co-expression analyses

We used standard protocols (Langfelder and Horvath 2008) to perform a weighted gene co-expression network analysis (WGCNA) on expression residuals for the 20 individuals to identify expression modules. We tested for associations between eigengenes (the first principle component of a module) and each of the nine phenotypes described (RWC, kidney weight, proportion of weight maintained, serum BUN, serum creatinine, serum total protein, serum chloride, serum potassium, and serum sodium) as well as population of origin and treatment group. We were able to assign genes membership to expression modules as well as position relative to the center of the module. Genes that are more

central (i.e. those that have more connections with other genes in a module) are good targets for putative candidate genes related to phenotypes of interest, population of origin, or treatment group. Additionally, to identify genes that show differential co-expression between Tucson and Edmonton we used the program DGCA (Differential Gene Correlation Analysis) (McKenzie et al. 2016), which calculates the average change in correlation between the two lines across all gene pairs.

Enrichment analyses
GO category enrichment on gene sets of interests were performed with GOrilla (Eden et al. 2009) by testing the foreground set against the background set of all genes expressed in the kidney. Phenotype enrichment tests were performed with modPhea (Weng and Liao 2017) by comparing the foreground set against the background set of all genes expressed in the kidney.

RESULTS

Relative water consumption is lowest in mice from the Sonoran desert
To determine whether mice from Tucson, Arizona exhibit lower water consumption compared to mice from other populations, we took advantage of a set of wild-derived inbred lines of mice from five localities across the Americas (Figure 1a). We assayed relative water consumption for 163 male mice representing 45 wild derived inbred lines over a 72-hour period from five founder populations (Tucson (TUC), Edmonton (EDM), Saratoga Springs (SAR), Gainesville (GAI), and Manaus (MAN). Mice from Tucson drank significantly less water than mice from any other population except Saratoga Springs (Median RWC: TUC: 0.34, SAR: 0.40, MAN: 0.42, GAI: 43, EDM: 0.52)(Figure 1b). The greatest difference was seen between mice from Tucson and mice from Edmonton (Mann-Whitney U, $p < 0.00001$)(Figure 1b). For this reason, we chose to focus on comparisons between lines from these two populations in all subsequent analyses.

Evolved and plastic phenotypic differences associated with xeric conditions
Water consumption is a complex trait with many factors contributing to the ultimate phenotype. To further characterize this phenotypic variation, we compared the inbred line from Tucson with the lowest average water consumption (TUCC/Nach) to the inbred line from Edmonton with the greatest average water consumption (EDMA/Nach).

First, we compared mice from these two lines under standard (hereafter, “hydrated”) conditions. In addition to the difference in water consumption seen between mice from these lines (Figure S1), we found that hydrated mice from Tucson and Edmonton showed several phenotypic differences related to fluid consumption and homeostasis in a standard laboratory environment. Edmonton mice had heavier kidneys relative to their body weight than Tucson mice ($p = 0.00015$)(Figure S2), but did not show significant differences in two aspects of gross morphology: renal cortex thickness nor the ratio of papilla to cortex thickness (Figure S3). The relative thickness of the papillae in the medulla is correlated with urine concentrating ability; thicker medullas are often associated with animals inhabiting more arid climates (Al-kahtani et al. 2004). Anecdotally, mice from Tucson appeared to produce far less urine than mice from Edmonton, consistent with many desert

rodents, but this was not measured in this study due to difficulty in obtaining urine from Tucson mice. In blood chemistry comparisons between hydrated Tucson and Edmonton mice, Tucson mice had higher chloride (median mmol/L: Tucson: 114.9, Edmonton: 112.5, p=0.03, Figure 2a) and creatinine (median mg/dL: Tucson: 0.17, Edmonton: 0.11, p=0.02, Figure 2b) levels. Chloride, an electrolyte, is a marker of dehydration and thus expected to be at higher concentrations in the blood of dehydrated animals (MacManes 2017). Creatinine is a waste product of normal muscle metabolism that is removed from the blood to be excreted by the kidney mainly through glomerular filtration. Blood creatinine levels increase as glomerular filtration, and thus kidney function, decreases and therefore is often used as a measure of kidney health (Kassirer 1971). The increased levels of serum chloride and creatinine, commonly warnings for declining osmoregulatory function, in healthy Tucson mice suggests that their baseline kidney function differs from that of Edmonton mice and they may be able to function normally despite higher blood osmolarity.

Next, we asked how mice from Tucson and mice from Edmonton differed in their response to water stress. We took male full-siblings from the same litter as mice from our hydrated comparison and withheld water from these mice for 72-hours. Hereafter, we refer to the water-restricted group as "dehydrated." We found that Tucson mice lost significantly less weight than Edmonton mice over the course of 72 hours without access to water (median percent weight maintained: Tucson: 0.82, Edmonton: 0.78, p=0.027, Figure 2c) suggesting that Tucson mice are more buffered against water stress. Comparing blood chemistry measures after mice were subjected to water stress, we found that Tucson mice measured higher in BUN (median mg/dL: Tucson: 64.45, Edmonton: 43.80, p=0.03, Figure 2d), chloride (median mmol/L: Tucson: 129.60, Edmonton: 111.95, p=0.03, Figure 2a), and potassium (median mmol/L: Tucson: 18.06, Edmonton: 7.78, p=0.03, Figure S4a) than Edmonton mice. BUN is a waste product from the liver during the metabolism of protein and increases as glomerular filtration rate and blood volume decreases (Baum et al. 1975). Similarly, high levels of serum potassium often reflect a decrease in filtration of the solute from the blood and thus decreased kidney function (Schwartz 1955) although this measure is particularly sensitive to lysed blood cells during collection and could reflect the challenge of collecting blood from dehydrated animals. Regardless of treatment, Tucson mice maintained higher serum chloride levels than Edmonton mice. In contrast, potassium levels were higher in Tucson mice only in the dehydrated treatment. High levels of both of these electrolytes are consistent with dehydration. Dehydrated mice from Tucson showed greater indicators for dehydration and kidney dysfunction than Edmonton mice but lost less weight when water stressed. This may reflect a greater evolved capacity to respond to the stress of dehydration. The fact that phenotypic differences in both hydrated and dehydrated animals persist in a common environment indicates that they may have a genetic basis.

While differences between the Tucson and Edmonton lines represent evolved differences, differences between hydrated and dehydrated mice are evidence of plastic responses to water stress. Many of the traits that differed between lines also exhibited phenotypic plasticity in comparisons between hydrated and dehydrated mice within lines. Dehydrated Tucson mice had higher levels of serum BUN compared with hydrated Tucson mice (median mg/dL: Hydrated: 25.20, Dehydrated: 64.45, p=0.008, Figure 2d). BUN differed

both between stressed and control mice from Tucson and was higher than in Edmonton mice when water stressed. Dehydrated Edmonton mice had higher levels of serum creatinine (median mg/dL: Hydrated: 0.11, Dehydrated: 0.14, p=0.03, Figure 2b) and sodium (mean mmol/L: Hydrated: 153, Dehydrated: 161, p=0.01, Figure S4b) compared with hydrated Edmonton mice, reaching the levels for both solutes that were seen in Tucson hydrated and dehydrated animals. The only solute that responded to water stress in both lines was total protein (median g/dL: Edmonton: Hydrated: 5.29, Dehydrated: 6.89, p=0.01; Tucson: Hydrated: 5.68, Dehydrated: 7.29, p=0.008). Total protein measures the concentration of both albumin and gobulin in the blood which increases with dehydration (Senay and Christensen 1965). These results indicate that while both lines react physiologically to the stress of dehydration, they may do this through different mechanisms.

Evolved and plastic transcriptional responses to xeric conditions

Changes in gene expression provide a flexible mechanism for rapidly responding to changes in the local environment, and can also underlie evolutionary divergence. Kidneys are the primary osmoregulatory organ and are essential for homeostasis and solute excretion. To identify candidate genes underlying adaptation to low water environments over short evolutionary timescales as well as plastic responses to water restriction, we sequenced mRNA from kidneys of ten Tucson and ten Edmonton mice, five from the dehydrated and five from the hydrated treatment. Differences in expression between Tucson and Edmonton mice in a common environment represent evolved differences, while differences between dehydrated and hydrated treatments represent a plastic response to water restriction.

We sequenced a total of ~1.3 billion reads for an average of 26,406,068 uniquely mapped reads per sample which were used to quantify mRNA expression levels and differential expression between samples. Gene expression was measured in a total of 54,233 genes as defined by Ensembl GRCm38 (mm10) with 19,105 genes expressed over a mean of ten reads per sample. Sampling the 1000 genes with the greatest variance, we found that Tucson and Edmonton individuals clustered separately, indicating that more of the gene expression variation was partitioned between line-of-origin than between treatment group (Figure 3a). Edmonton mice clustered into two distinct groups based on treatment (dehydrated vs. hydrated individuals), but dehydrated and hydrated Tucson mice did not form distinct clusters (Figure 3a).

To identify differential gene expression, we used DESeq2 (Love et al. 2014) to perform pairwise comparisons between: 1) Tucson hydrated *vs.* Edmonton hydrated, and 2) Tucson dehydrated *vs.* Edmonton dehydrated, 3) Tucson dehydrated *vs*. Tucson hydrated, and 4) Edmonton dehydrated *vs.* Edmonton hydrated individuals. Overall, we found more genes were differentially expressed between the two lines than between the treatment groups (FDR of 5%; see methods). A total of 3,935 genes were differentially expressed between hydrated Edmonton and Tucson mice while a total of 3,419 genes were differentially expressed between dehydrated Tucson and Edmonton individuals, a 51% overlap with differences seen between hydrated Tucson and Edmonton mice.

Comparing the hydrated and dehydrated groups within each population (Tucson dehydrated *vs*. Tucson hydrated, and Edmonton dehydrated *vs*. Edmonton hydrated), we found that twice as many genes were differentially expressed in the Edmonton (1354) than in the Tucson comparisons (677 genes) (Chi-square test with Yates Correction, $p < 0.0001$, Figure 3b), with a 225 gene overlap. This 225 gene overlap represents the shared transcriptional response to dehydration with respect to these two lines and is enriched for phenotypes including dehydration ($q=9x10^{-3}$) and decreased vasodilation ($q=3.8x10^{-2}$) and GO terms involved in regulation of blood pressure ($q = 4.99x10^{-2}$). Genes differentially expressed between hydrated and dehydrated Edmonton mice were also enriched for GO terms relevant to water stress, such as renal system processes ($q = 5.08 \times 10^{-2}$) and regulation of body fluids ($q = 1.30x10^{-2}$). Within genes solely differentially expressed between the Tucson groups, we saw enrichment for homeostasis related GO terms (see File S1), but not for any kidney-specific categories. In addition to having a greater number of differentially expressed genes in Edmonton comparisons, we also found that the average magnitude of expression differences between hydrated and dehydrated treatments ($|\log_2$ fold change$|$) was significantly greater for Edmonton mice than for Tucson mice (Mann-Whitney *U*, $p < 2.2 \times 10^{-16}$) (Figure 3c). Together, these results are consistent with Edmonton mice being farther from their physiological optimum when water stressed than are Tucson mice, consistent with the hypothesis that Tucson mice are locally adapted to a water limited environment.

Dehydrated Edmonton mice show shifts towards Tucson-like expression

Next, we were interested in asking whether the plastic changes in response to water stress in the non-xeric mice (i.e. Edmonton) go in the same or opposite direction as the evolved differences between mice from xeric and non-xeric habitats. Specifically, we were interested in whether Edmonton mice placed under water stress would show transcriptional responses that make them more similar to the base-line condition of Tucson mice. Therefore, we focused on differentially expressed genes between the Edmonton hydrated and dehydrated groups and asked whether the Edmonton dehydrated group showed shifts in expression in the direction of the Tucson hydrated condition (Figure 4a). The majority of these genes (85%, 416 genes) showed changes in the same direction, meaning that the putatively adaptive and plastic responses were in the same direction (+/+ and -/-) (Binomial exact test, $P<0.0001$). Only 15% (74 genes) show changes in opposite directions (+/- and -/+) (Figure 4b).

The group of 416 genes with evolved and plastic changes in the same direction was enriched for GO terms involved in homeostasis and ion transport and for mutant renal/urinary system phenotypes ($q = 0.035$). For example, one of these genes, *Aquaporin 1* (*Aqp1*), showed differences in expression between hydrated and dehydrated Edmonton mice ($q = 0.0016$) and between the hydrated conditions of both lines ($q = 0.00019$)(Figure 5a). In Tucson mice, there was no significant effect of treatment on expression level, but in Edmonton mice, expression increased in response to water stress recapitulating the constitutive expression level of the mice from Tucson. Aquaporins are a family of membrane proteins which form channels used to transport water and small solutes across cell membranes. *Aqp1* is expressed in the descending loop of Henle, and channels formed from this protein are the main route through which water is reabsorbed in this region

(Chou et al. 1999). It is known to affect urine concentrating ability, response to dehydration, and renal water transport in lab lines of house mice (Ma et al. 1998; Sohara et al. 2005) and has been identified in a several studies related to desert adaptation in rodents (reviewed in Pannabecker, 2015). In our analyses, expression of *Aqp1* was associated with variation in six of the nine measured phenotypes (Creatinine, p=0.020; total protein, p=0.012; potassium, p=0.0094; weight loss, p=0.016; kidney weight, p=0.0077; RWC, p=0.022).

Of the 416 genes where dehydrated Edmonton mice showed shifts towards Tucson-like expression, the majority (87%) were not differentially expressed between hydrated and dehydrated Tucson mice. For the 54 of these genes that were differentially expressed in the Tucson comparison, in all cases the plastic response was observed to be changing in the same direction as in Edmonton mice.

Co-expressed sets of genes are associated with phenotypic variation in Tucson and Edmonton mice

To identify transcriptional networks associated with phenotypic variation in Tucson and Edmonton mice, we performed a weighted gene co-expression network analysis (WGCNA). This analysis uncovers genes with highly correlated expression profiles and groups them into modules reflecting hypotheses about connectivity. We identified 54 co-expression modules, with at least one module significantly associated with each of the nine phenotypes (Figure S5). Of these, the "salmon" module (Figure S6a) was of particular interest because it was significantly associated with all nine phenotypes (BUN, p=0.02; Creatinine, p=0.006; total protein, $p = 1 \times 10^{-4}$; Chloride, p = 0.002; Potassium, p = 0.001; Sodium level, p = 0.01; weight loss, p = 0.007; kidney weight, p = 0.03; RWC, p= 0.006) as well as population of origin (p = 0.006) and treatment (p = 0.001). Genes in this module were enriched for several metabolic processes, including glyceraldehyde-3-phosphate metabolic process ($q = 6.85 \times 10^{-9}$), ribose phosphate metabolic process ($q = 1.45 \times 10^{-8}$), and carbohydrate derivative metabolic process ($q = 1.3 \times 10^{-7}$). Visualizing the most connected genes in the salmon module, we identified *Tmtc1*, *Apoe*, and *Sult1a1* as the most centrally located hub genes (Figure S6a). All three of these genes were identified in our previous analysis as genes for which dehydrated Edmonton mice showed shifts towards Tucson-like expression (Figure 4). *Apoe* is of particular interest because of its documented role in kidney function. It is thought to play an important role in renal damage protection (Wen et al. 2002; Bonomini et al. 2011), and laboratory mutants show a number physiological and morphological changes similar to kidney disease phenotypes (Bonomini et al. 2011). A*poe* expression responded in the same direction and in a similar magnitude in both lines under water stress. Interestingly, under water stress, the expression level in Edmonton recapitulated the constitutive expression level of hydrated Tucson mice (Figure 5b). *Apoe* expression was also correlated with eight of the nine measured phenotypes (Creatinine, p=0.034; total protein, p=0.00077; chloride, p=0.0055; potassium, p=0.0041; sodium, p=0.034; weight loss, p=0.014; kidney weight, p=0.025; RWC, p=0.0036).

In addition to shifts in the expression of entire co-expression modules, populations may differ as a consequence of altered co-expression between pairs of genes, called differential co-expression. In order to identify genes that show differential co-expression between

Tucson and Edmonton, we calculated the average change in correlation between the two lines across all gene pairs using the program DGCA (Differential Gene Correlation Analysis) (McKenzie et al. 2016). We identified 182 genes that showed significantly different co-expression between Tucson and Edmonton individuals (q <0.10, see methods). These genes were enriched for several GO categories including glycolysis: generation of precursor metabolites and energy (q = 0.0013), cellular amino acid metabolic process (q = 0.00046), and lipid metabolic process (q = 0.0014). This gene set was also enriched for mutant phenotypes related to renal/urinary system (q = 0.00001), abnormal urine homeostasis (q = 0.005), and abnormal ion homeostasis (q = 0.01). One of the genes with significant changes in co-expression was *Aqp1*, which is involved in renal water transport. We also found that several solute carriers (*Slc22a19, Slc11a2, Slc36a1, Slc47a1, Slc12a1, Slco3a1*) showed evidence of differential co-expression, including one (*Slc47a1*) that has been shown to be under positive selection in the desert-adapted cactus mouse, *Peromyscus eremicus* (Kordonowy et al. 2017). *Slc47a1* mouse mutants are also associated with increased BUN, increased circulating creatinine level, and kidney degeneration (Tsuda et al. 2009). Altogether, these results suggest that Tucson and Edmonton mice show shifts in the expression of co-expression modules as well as changes in the co-expression associations between sets of genes.

DISCUSSION

Here we have described phenotypic and transcriptional divergence between descendants of mice from a desert environment and descendants of mice from a more mesic environment when reared under identical laboratory conditions. We also described plastic responses in these mice under conditions of water stress. First, we showed that inbred lines derived from the Sonoran desert consume less water than do mice from other populations in the Americas. Next, comparisons between a single line from Tucson and a single line from Edmonton revealed many phenotypic differences in a common environment, both at the organismal level and at the gene expression level in the kidney. The fact that these differences were present after multiple generations in the lab indicates that they are genetically based. Nonetheless, these same traits reveal considerable plasticity in comparisons between control mice and mice under conditions of water stress. Notably, we found that Tucson mice showed attenuated responses to water stress. After a 72 hour period without water, Tucson mice lost less weight and showed fewer expression differences in the kidney. Surprisingly, the blood chemistry of Tucson mice was consistent with higher levels of dehydration and reduced kidney function, both under standard and water-restricted conditions. However, phenotypes that appear maladaptive in one genomic or environmental context may be adaptive in another (e.g., Riddle et al., 2018). Altogether, these results are consistent with genetic changes to Tucson mice following their invasion of the desert environment.

Phenotypic plasticity may allow animals to persist in harsh new environments if the plastic responses bring individuals closer to the local optimum (reviewed by Ghalambor et al., 2007). Adaptive plasticity can be followed by genetic changes as populations become established, in a process called “genetic assimilation” or the “Baldwin Effect” (Waddington 1942, 1952, 1953; Simpson 1953; Robinson and Dukas 1999; Price et al. 2003). We found

that the evolved differences in kidney gene expression between Tucson and Edmonton mice were generally in the same direction as plastic changes in Edmonton individuals under water stress. Consequently, under water stress, gene expression in Edmonton mice becomes more similar to that of Tucson mice under hydrated conditions. This result is consistent with the idea that plastic responses to short-term water stress are an example of adaptive plasticity. Thus, plastic responses to water stress may have helped facilitate the colonization and subsequent adaptation of house mice to the desert environment.

Our finding that plastic responses generally go in the same direction as evolved responses stands in contrast to several recent studies. For example, an allele of the *Epas1* gene confers adaptation to high altitude in Tibetans by attenuating the maladaptive plastic response of increased hemoglobin concentration (Beall et al. 2010; Simonson et al. 2010; Yi et al. 2010; Jeong et al. 2018). Similarly, most gene expression changes in the brains of guppies reared in the absence of predators go in the opposite direction of those seen in populations that have evolved without predators (Ghalambor et al. 2015). In both cases, the selective agent (hypoxia in humans or absence of predators in fish) may have not been present in the recent history of the populations exhibiting non-adaptive plasticity. In contrast, we speculate that occasional periods of water stress are probably common in many populations of mice, including in places, like Edmonton, that are not in deserts. Under such situations, selection is expected to favor an adaptive plastic response.

While adaptive plasticity may facilitate the colonization of new environments, it can also slow or impede adaptive evolution if, by moving individuals closer to the optimum, genetic variation is shielded from natural selection (Ghalambor et al. 2007; Price et al. 2003). However, when plastic responses to a new environment are incomplete, directional selection may favor a more extreme phenotype and lead to subsequent genetic changes (Price et al. 2003). While all house mice, even those in mesic environments, likely undergo short periods of water stress, water stress is likely to be more severe in desert environments. Phenotypic comparisons between Edmonton and Tucson mice suggest that plastic responses to water stress in Edmonton mice may be suboptimal. Tucson mice drink less water on average and lose less weight in response to dehydration, indicating that these animals are buffered against water stress in a way that Edmonton mice are not. The blood chemistry comparisons reported here also suggest there are differences between Tucson and Edmonton kidney function and homeostasis. Consequently, we suggest that while phenotypic plasticity likely helped house mice colonize the Sonoran desert, the xeric environment still imposed sufficient selective pressure for subsequent genetic changes.

Finally, the comparison of gene expression changes both between treatments and between lines has identified a few genes that may be important in the adaptive response. Expression changes pinpoint a number of interesting candidates, including *Aqp1* and *Apoe*. Future studies aimed at identifying *cis*-regulatory variation at these genes might help to pinpoint causative mutations underlying adaptation to desert environments.

ACKNOWLEDGEMENTS

I thank Katya Mack and Michael Nachman for their substantial contributions to this paper and the design and execution of this study. I also thank members of the Nachman Lab for valuable comments and discussions. I thank Taichi Suzuki, Megan Phifer-Rixey, Felipe Martins, Dana Lin, Michael Sheehan, and Mallory Ballinger for help with animal husbandry or for collecting the wild mice used to establish the lines for this study. I thank Lydia Smith of UC Berkeley and Eugene Dunn of UC Davis for their technical expertise. This work was facilitated by an Extreme Science and Engineering Discovery Environment (XSEDE) allocation to M.W.N. XSEDE is supported by National Science Foundation (NSF) grant number ACI-1548562. This work was supported by an NIH grant to MWN (R01 GM127468) and a NSF Doctoral Dissertation Improvement Grant to NKJB (1601827).

DATA ACCESSIBILITY STATEMENT

Illumina sequencing data from this study has been submitted to the NCBI Sequence Read Archive under the Bioproject PRJNA614581.

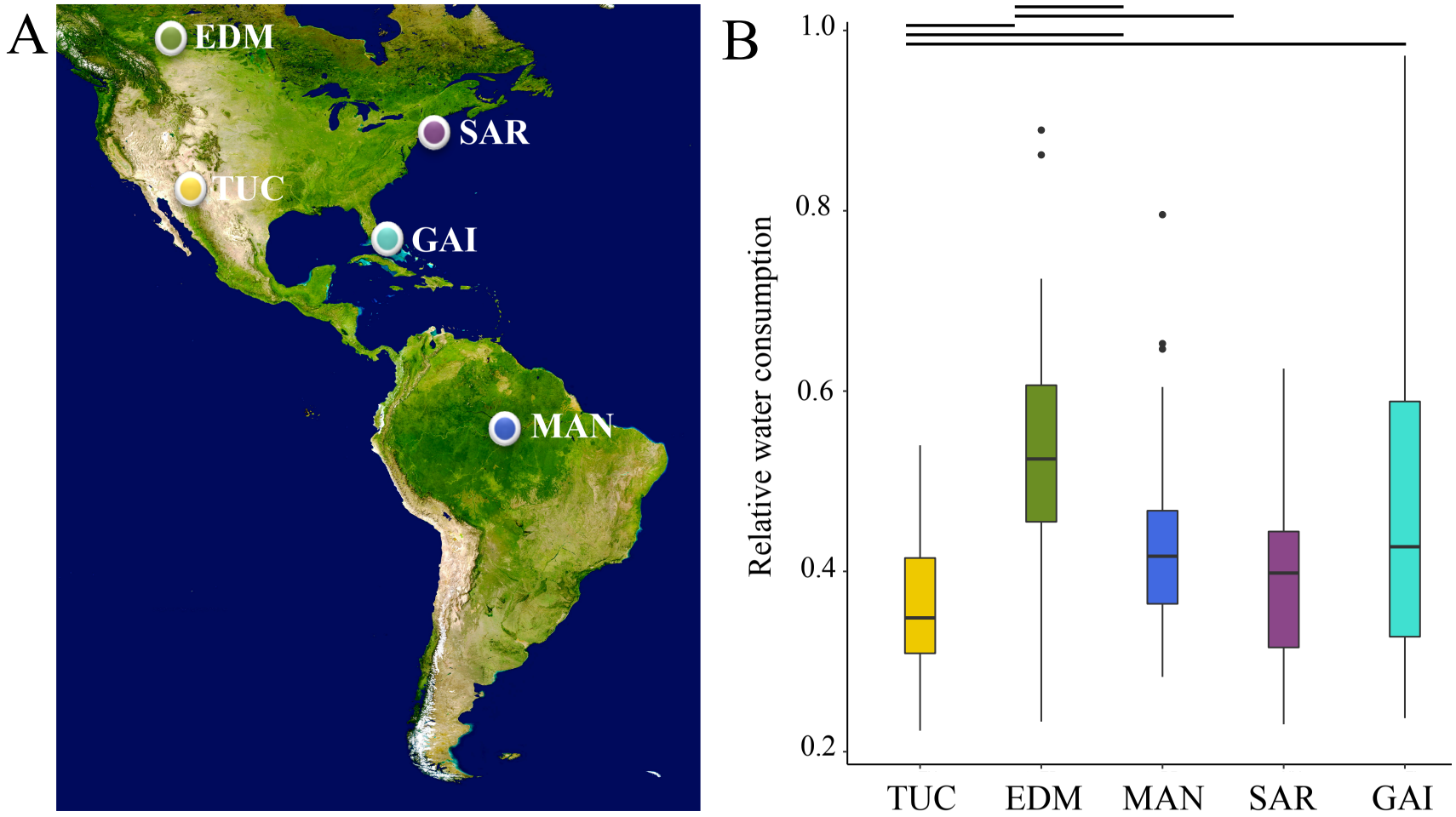

Figure 1. Relative water consumption in lab-born descendants of wild mice from different environments. (A) Sampling localities of wild-caught mice used to establish inbred lines in this study (map obtained from Google satelites.pro): Edmonton, Canada (EDM), Tucson, AZ (TUC), Gainesville, FL (GAI), Saratoga Springs, NY (SAR), and Manaus, Brazil (MAN). (B) Relative water consumption (g water consumed/ g mouse) in descendants of mice from different localities. Lines indicate comparisons that are significant ($p < 0.05$; Mann-Whitney U tests). Vertical lines denote 1.5 * the interquartile range.

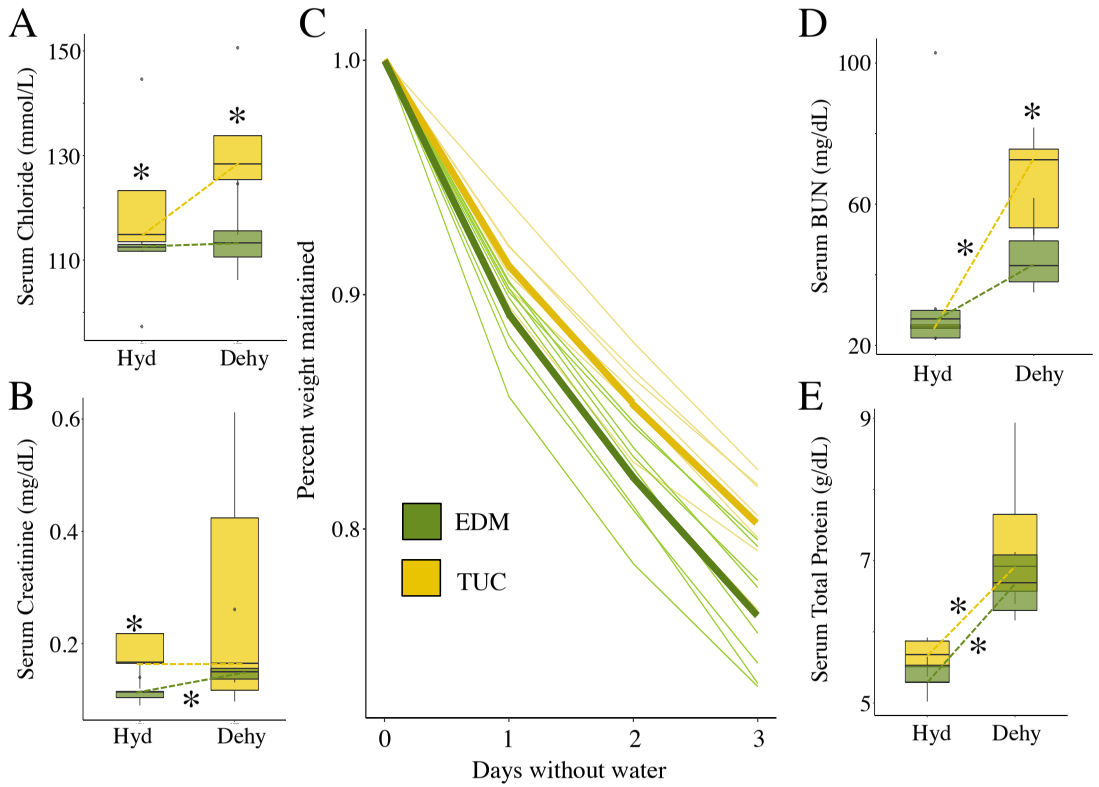

Figure 2. Evolved and plastic phenotypic variation among hydrated and dehydrated mice from desert and non-desert environments. Figures A,B, D, and E show reaction norms between hydrated and dehydrated states of both lines (Tucson, Edmonton) for each phenotype. Dotted lines connecting the medians represent the pattern of phenotypic expression as a response to the treatment (hydrated, dehydrated). Asterisks denote significant comparisons ($p < 0.05$) either between lines (above boxes) or between treatments (next to dotted lines). (A) Tucson mice show higher levels of serum chloride than Edmonton mice, both when hydrated and when dehydrated (p=0.03 for both). (B) Significant differences between hydrated Edmonton and Tucson mice (p=0.02) and between hydrated and dehydrated Edmonton mice (p=0.03) in creatinine. (C) Tucson mice show significantly less weight loss after 72 hours of water restriction (p=0.027). (D) Significant differences in serum BUN between hydrated and dehydrated Tucson mice (p = 0.008) and dehydrated Tucson and Edmonton mice (p = 0.03). (E) Significant differences in total protein between hydrated and dehydrated Tucson mice (p=0.008) and between hydrated and dehydrated Edmonton mice (p=0.01). Vertical lines denote 1.5 * the interquartile range.

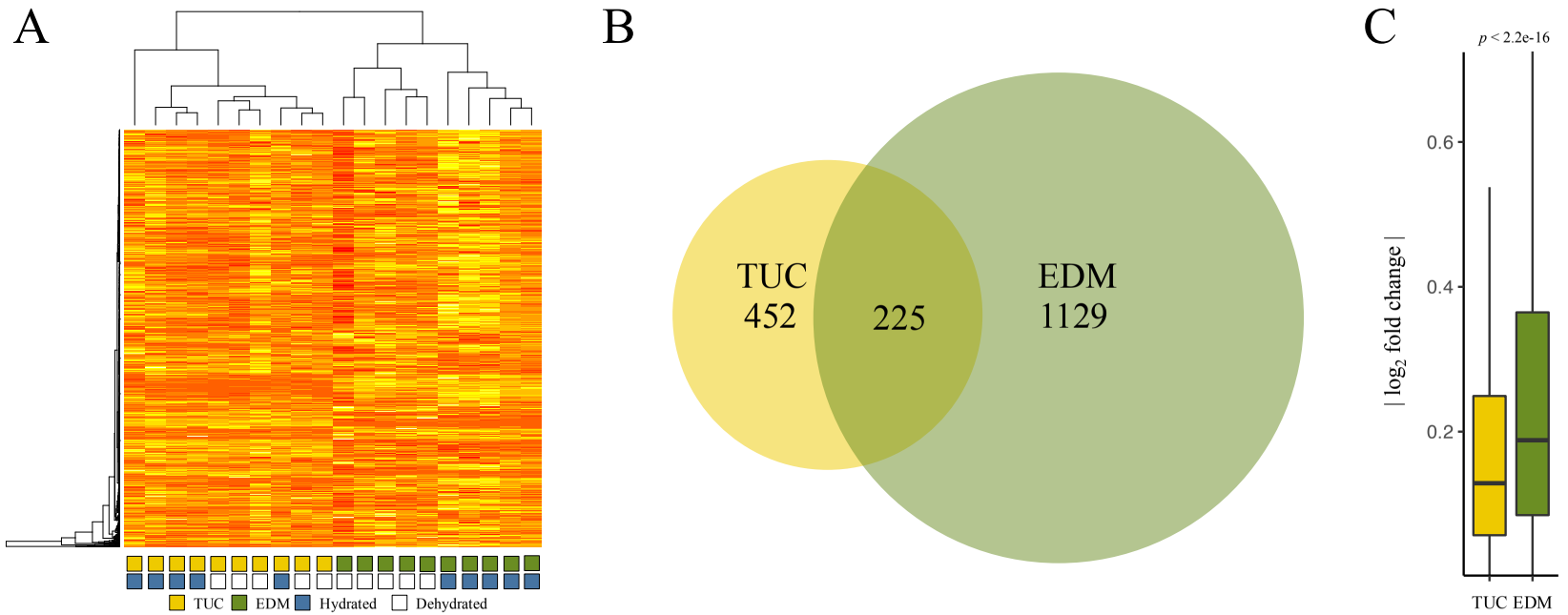

Figure 3. Evolved and plastic gene expression variation among hydrated and dehydrated mice from desert and non-desert environments. (A) Heat map depicting relationships among samples for the top 1000 genes with greatest variance in expression. Expression patterns form two major groups, corresponding to line of origin (Tucson versus Edmonton). Edmonton samples also form clusters based on treatment (hydrated versus dehydrated) while Tucson samples do not. (B) Numbers of differentially expressed genes between dehydrated and hydrated samples in Tucson and Edmonton. Edmonton mice exhibit twice as many genes with differential expression between dehydrated and hydrated conditions compared to Tucson mice. The 225 genes at the intersection represent the shared transcriptional response to water stress. (C) Magnitude of fold changes in each population between dehydrated and hydrated samples. Vertical lines denote 1.5 * the interquartile range.

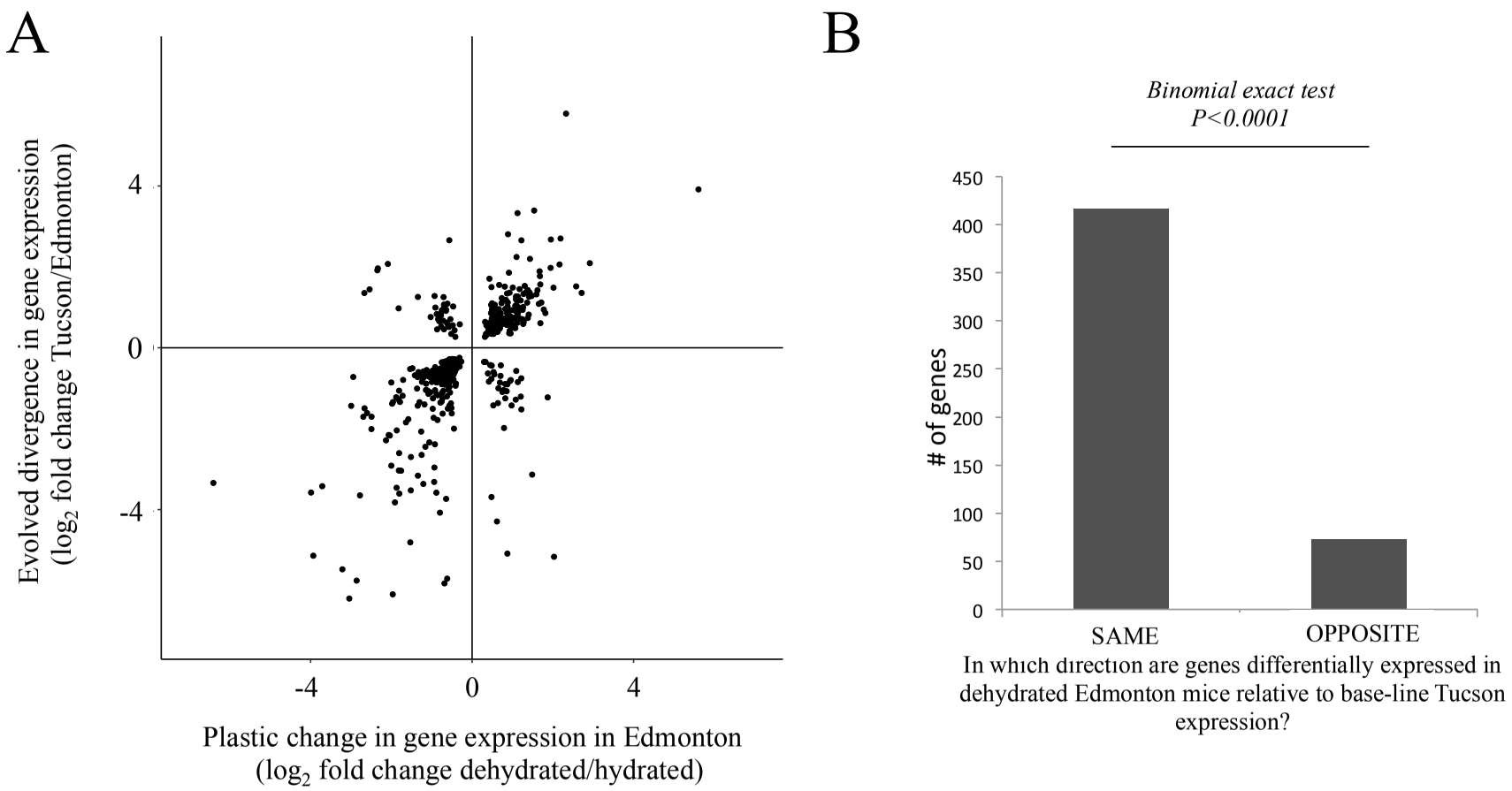

Figure 4. Plastic responses to dehydration in non-desert (Edmonton) mice are mostly in the same direction as evolved differences between non-desert (Edmonton) and desert (Tucson) hydrated mice. (A) Scatterplot comparing the evolved and plastic changes in gene expression. Each point represents a gene and the log2 fold change between Edmonton hydrated vs. dehydrated on the x-axis (plastic response) and Tucson hydrated vs. Edmonton hydrated on the y-axis (evolved divergence). (B) Number of genes in which the evolved and plastic transcriptional responses go in the same direction, and number of genes in which the evolved and plastic transcriptional responses go in the opposite direction.

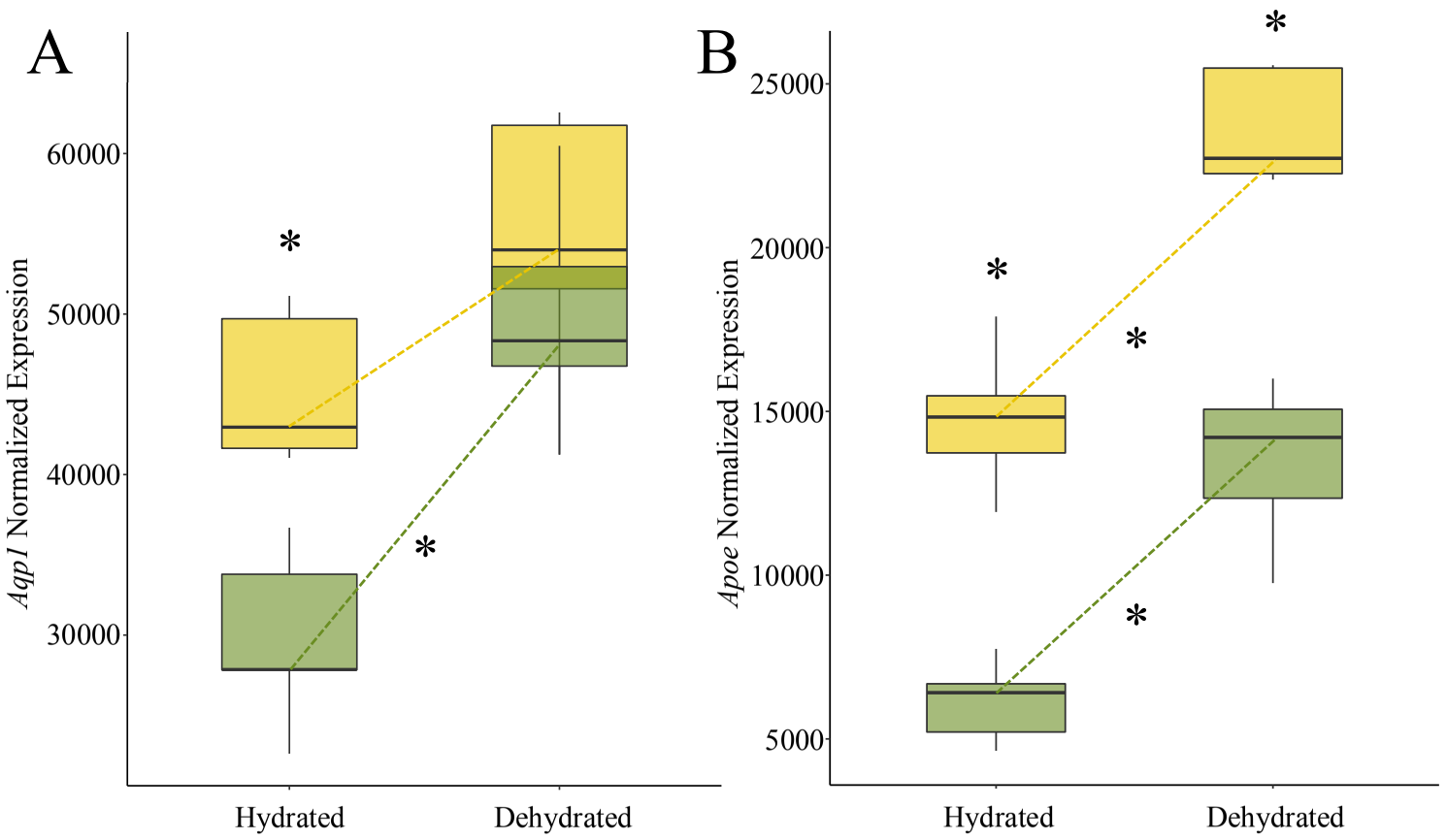

Figure 5. Expression variation at two genes that may underlie adaptation to a desert environment. Reaction norms showing normalized gene expression in hydrated and dehydrated mice from Tucson (gold) and Edmonton (green). Dotted lines connecting the medians represent the pattern of phenotypic expression as a response to the treatment (hydrated, dehydrated). Asterisks denote significant comparisons ($p < 0.05$) either between lines (Tucson, Edmonton; above boxes) or between treatments (hydrated, dehydrated; next to dotted lines). Asterisks denote significant comparisons ($p < 0.05$) either between lines (Tucson, Edmonton) or between treatments (hydrated, dehydrated). (A) *Aqp1*. (B) *Apoe*. For both genes, the dehydrated Edmonton mice recapitulate the baseline expression level seen in hydrated Tucson mice. For *Aqp1*, the plastic response is attenuated in Tucson mice (A), while for *Apoe*, it is not (B). Vertical lines denote 1.5 * the interquartile range.

SUPPLEMENT

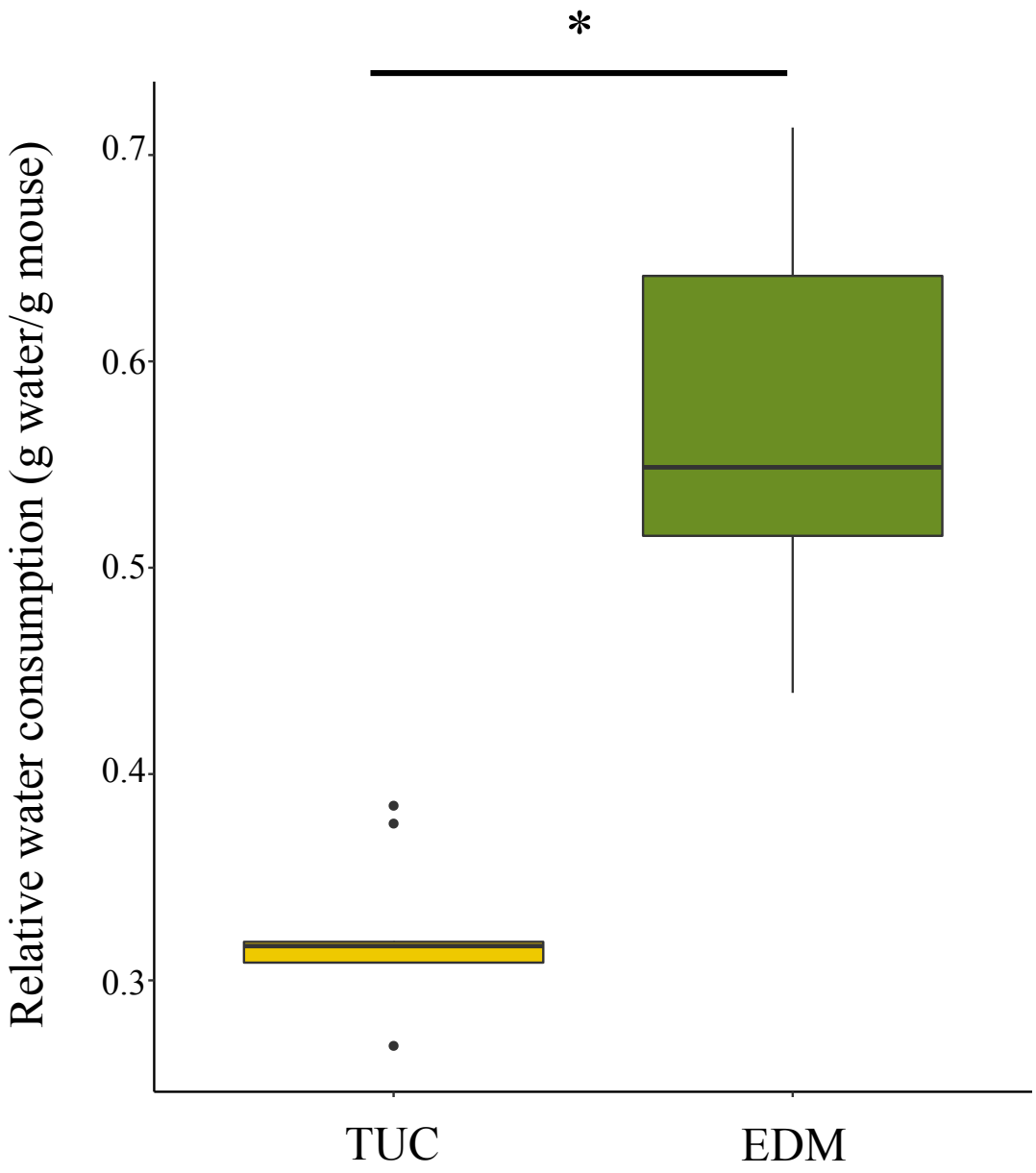

Figure S1. Relative water consumption in hydrated control mice from different environments. Edmonton mice consume significantly more water adjusted for body mass than do Tucson mice (*p = 0.0026). Vertical lines denote 1.5 * the interquartile range.

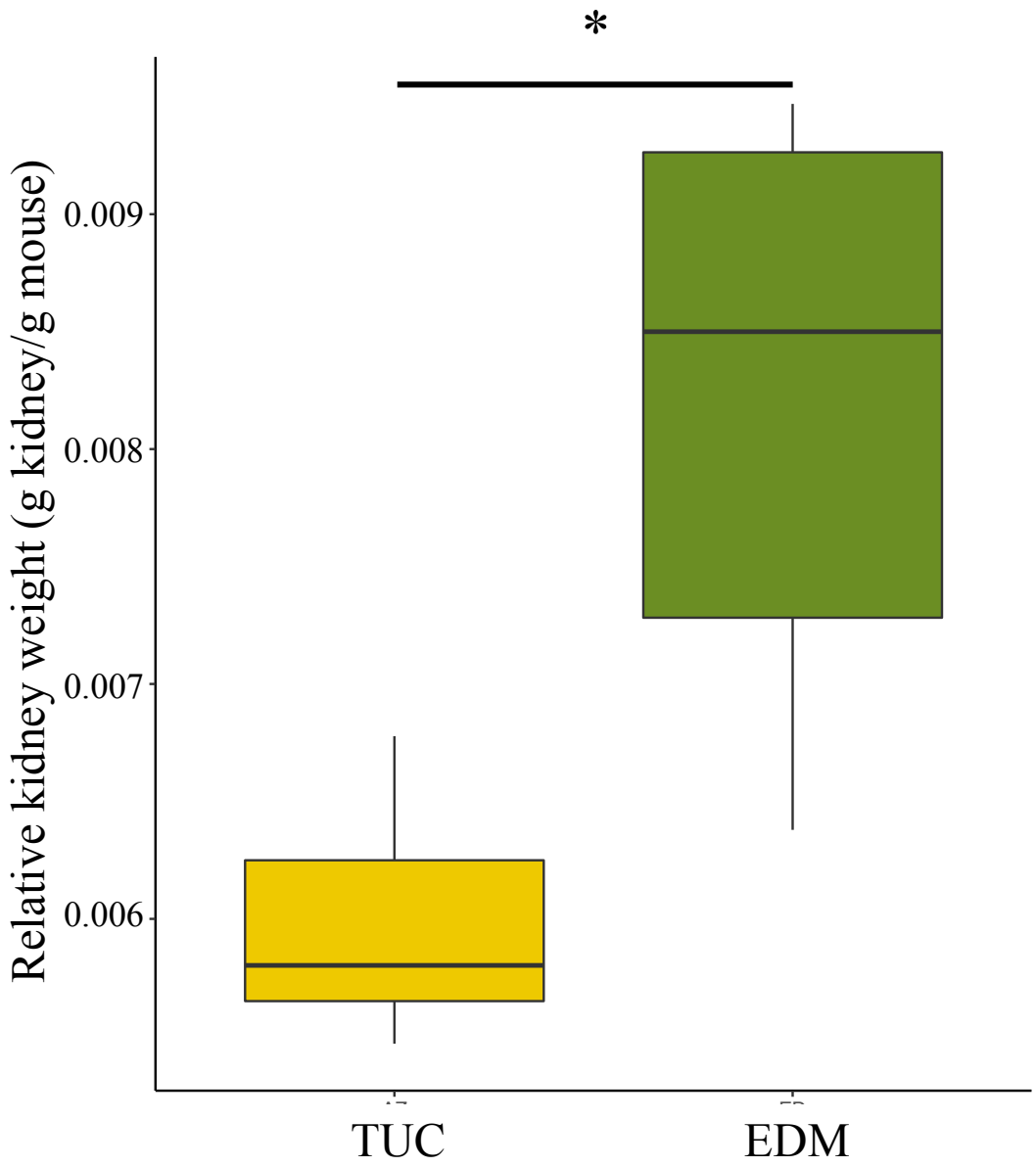

Figure S2. Relative kidney weight in hydrated control mice from different environments. Edmonton mice have heavier kidneys relative to their body weight than Tucson mice (*p = 0.00015). Vertical lines denote 1.5 * the interquartile range.

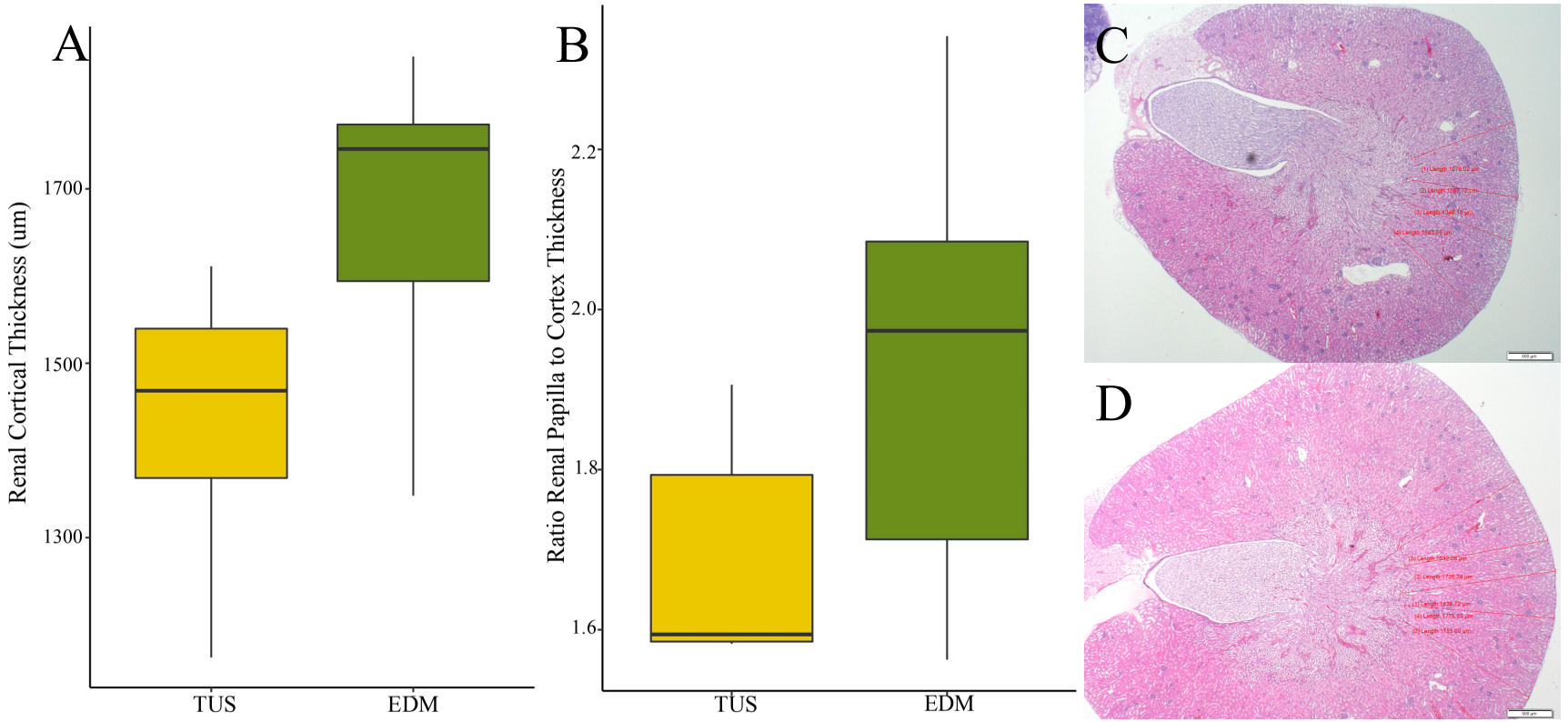

Figure S3. Kidney histology from hydrated control mice from different environments. Tucson and Edmonton mice do not show significant differences in (a) renal cortex thickness (p = 0.15) nor (b) the ratio of renal papilla thickness to cortical thickness (p = 0.31). Vertical lines denote 1.5 * the interquartile range. (c) Example cross-section with measurements of a Tucson kidney. (d) Example cross-section with measurements of an Edmonton kidney.

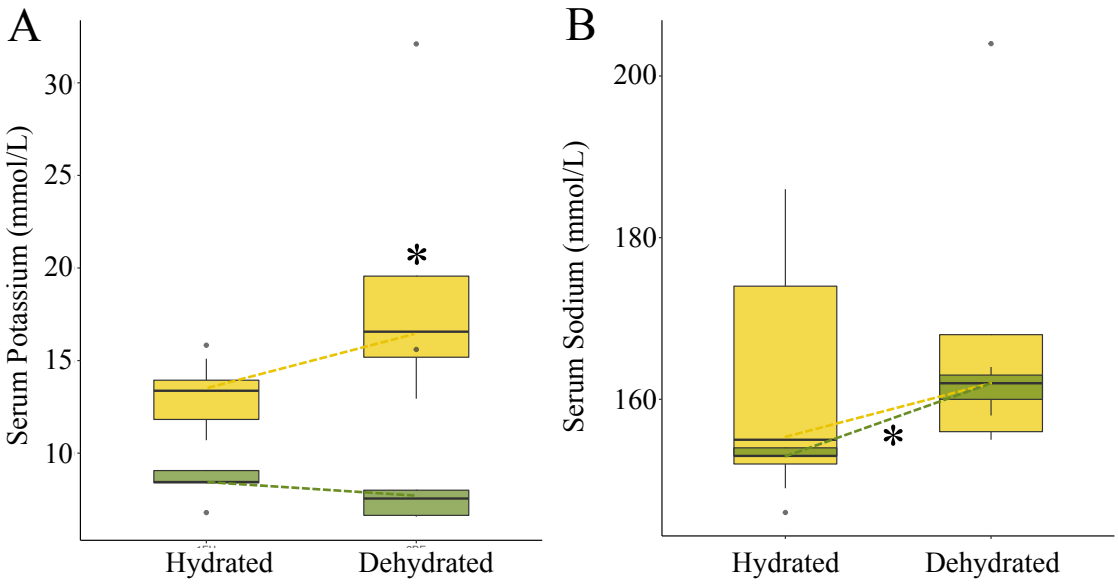

Figure S4. Evolved and plastic phenotypic variation among hydrated and dehydrated mice from desert and non-desert environments. Asterisks denote significant comparisons ($p < 0.05$) either between lines (Tucson, Edmonton) or between treatments (hydrated, dehydrated). (a) Serum potassium concentration is significantly different between dehydrated Tucson mice and dehydrated Edmonton mice (median mmol/L: Tucson: 7.78, Edmonton: 18.06, p=0.32). (b) Serum sodium concentration is significantly different between hydrated and dehydrated Edmonton mice (median mmol/L: Hydrated: 153, Dehydrated: 161, p=0.01). Vertical lines denote 1.5 * the interquartile range.

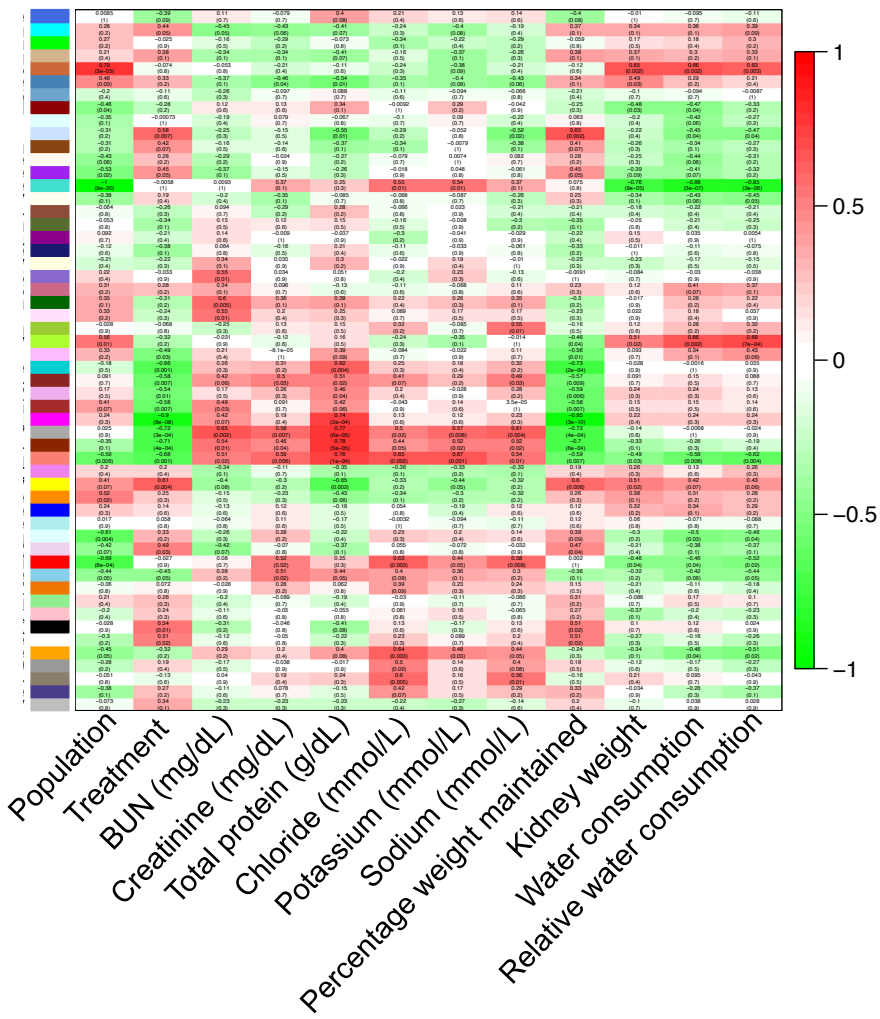

Figure S5. Results from a weighted gene co-expression analysis (WGCNA). Modules (rows represented by colors in the first column) represent groups of genes with highly correlated expression values. Each module is represented by the significance of its association with each phenotype, population, or treatment in each column. The “salmon” module is significantly associated with all nine parameters of interest.

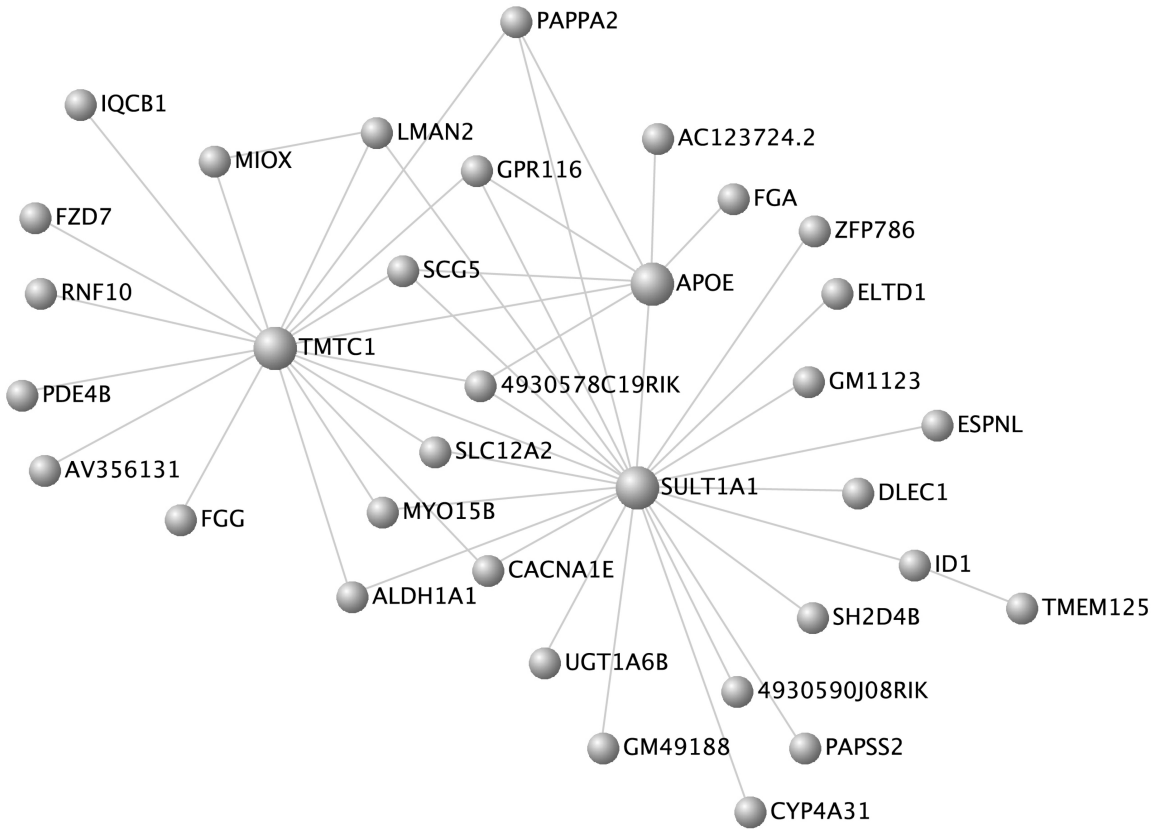

Figure S6. Network of genes identified by WGCNA in the "salmon" module. Genes with the greatest number of connections are of particular interest including *Apoe* which is associated with kidney-related phenotypes.

Table S1. Wild derived inbred lines phenotyped for water consumption. Lines with a second name followed by "/Nach" represent universal strain names. All other names are in-lab notations. Some strains are assigned both names. Strains denoted with * were used in experimental portion of the study.

Tucson, AZ, USA (TUC)	Gainsville, FL, USA (GAI)	Edmonton, Canada (EDM)	Saratoga Springs, NY, USA (SAR)	Manaus, Brazil (MAN)
TUS(30x34)C	DL(51x50)C	TAS(69X186)A	MJS(9x11)A	FMM(208x209)A
TUS(12x10)B	DL(60x78)C	TAS(130X152)B	MJS(34x38)B	FMM(218X238)B
TUS(32X34)B	DL(62X61)A	TAS(92X87)C	MJS(66x67)B	FMM(224x225)B
TUS(30X34)A; TUCA/Nach	DL(64x74)B	TAS(125X120)A	MJS(76x56)C	FMM(228x238)B
TUS(9x25)B; TUCB/Nach	DL(76x87)A	TAS(89x90)B	MJS(83x80)A	FMM(250x247)A
TUS(A4xA8)C; TUCC/Nach*	DL(81X82)A	TAS(111X165)B; EDMA/Nach*	MJS(84x89)A	FMM(255x193)B
	DL(107x83)B	TAS(96X154)C; EDMB/Nach	MJS(92x91)B	FMM(260x220)A
	DL(69X77)C; GAIB/Nach	TAS(93X101)B; EDMC/Nach	MJS(19x13)B; SARA/Nach	FMM(261x239)B
	DL(95x54)B; GAIC/Nach		MJS(82x81)C; SARB/Nach	FMM(263X249)A
				FMM(193x255)A; MANA/Nach
				FMM(222x254)A; MANB/Nach
				FMM(221x215)A; MANC/Nach
				FMM(254x215)B; MAND/Nach

Chapter 2

Convergent patterns of gene expression and protein evolution associated with adaptation to desert environments in rodents

ABSTRACT

Desert specialization has arisen multiple times across rodents and is often associated with a suite of convergent phenotypes, including modification of the kidneys to mitigate water loss. However, the extent to which phenotypic convergence in desert rodents is mirrored on the molecular level is unknown. Here, we sequenced kidney mRNA and assembled transcriptomes of three pairs of rodent species to search for convergence in gene expression and amino acid sequence associated with adaptation to deserts. We conducted phylogenetically-independent comparisons between a desert specialist and a non-desert relative in three families representing ~70 million years of evolution. Overall patterns of gene expression faithfully recapitulated the phylogeny of these six taxa. However, we found that 8.6% of all genes tested showed convergent patterns of expression evolution between desert and non-desert taxa, a proportion that is much higher than expected by chance. In addition to these convergent changes, we observed many species-pair specific changes in gene expression indicating that different instances of adaptation to deserts include a combination of unique and shared changes. Patterns of protein evolution revealed a small number of genes showing evidence of positive selection, and most of these did not show convergent changes in gene expression. Convergence in gene expression was more frequent than convergence in amino acid sequence, suggesting that changes in gene regulation play a primary role in desert adaptation in rodents.

INTRODUCTION

The repeatability of adaptive evolution at the molecular level remains an open question. In situations where the mutational target is small and constraints exist due to epistasis or pleiotropy, the molecular paths available to adaptation may be highly limited (Weinreich et al. 2006; Karageorgi et al. 2019). Indeed, there are a number of excellent examples of convergent molecular evolution underlying simple traits (e.g. Stewart and Wilson 1987; Mundy 2005; Zhen et al. 2012). For highly polygenic traits, however, convergence may be less expected simply because the mutational target is large and multiple paths may be available on which selection can act. Nonetheless, several studies have found evidence for convergence at the molecular level even for complex traits (e.g. Marcovitz et al. 2019; Sackton et al. 2019).

Convergent phenotypic evolution may be due to changes in gene regulation or to changes in protein structure, or both, yet these processes are rarely studied together in the context of complex adaptive traits (but see Hao et al. 2019). There is evidence that gene expression divergence and amino acid sequence divergence are correlated between paralogs following gene duplications (Gu et al. 2002; Makova and Li 2003), and more generally that rates of gene expression and rates of protein evolution are coupled in some lineages (e.g. Nuzhdin et al. 2004; Lemos et al. 2005). These observations raise the possibility that changes in both gene expression and protein sequence may contribute to the repeated evolution of complex adaptive traits.

Adaptation to desert environments in rodents provides an opportunity to study repeated evolution in both gene expression and protein sequence. Desert ecosystems present the challenge of extreme aridity and low or seasonally absent water, yet multiple lineages of rodents have independently evolved the ability to survive in these unusually harsh environments (reviewed in Degen 1997) Rodents have solved these challenges in different ways, including dietary specialization on plants that are relatively high in water content (Schmidt-Nielsen 1979) or modifications to reduce evaporative water loss (Schmidt-Nielsen and Schmidt-Nielsen 1952; Schmidt-Nielsen 1964; Degen 1997). However, a common feature of most desert rodents is a modified kidney capable of producing highly concentrated urine (MacMillen and Lee 1967; Beuchat 1990; Al-kahtani et al. 2004; Donald and Pannabecker 2015). Final excreted urine concentration depends on the development and maintenance of a corticomedullary osmotic gradient within the kidney. Studies have shown that many aspects of kidney morphology and physiology have been modified in different lineages to produce hyper-concentrated urine (Bankir and de Rouffignac 1985; Donald and Pannabecker 2015).

To study the molecular basis of convergent adaptation to deserts in rodents, we compared kidney gene expression and protein sequence divergence between a desert and a non-desert species in each of three pairs of phylogenetically independent comparisons representing transitions to desert living in three different rodent families (Heteromyidae, Dipodidae, and Muridae). Desert species were chosen based on their high urine concentration, a proxy for increased osmoregulatory capacity (Figure 1). Within Muridae, we compared the Australian Spinifex Hopping Mouse, *Notomys alexis*, the mammal with the highest known urine concentration and well studied for its modifications to desert life (MacMillen and Lee 1967; Macmillen and Lee 1969; Baudinette 1972), to the house mouse (*Mus musculus*), a widespread generalist. Within Dipodidae, we compared the desert-dwelling Lesser Egyptian Jerboa, *Jaculus jaculus,* previously studied for its kidney modifications associated with granivorous desert living (Schmidt-Nielsen and Schmidt-Nielsen 1952; Khalil and Tawfic 1963), to the riparian Western Jumping Mouse, *Zapus princeps*, a common North American species found in mesic environments. Within Heteromyidae, we compared the Rock Pocket Mouse, *Chaetodipus intermedius* (Bradley et al. 1975; Altschuler et al. 1979), native to the North American Sonoran desert , to the Desmarest's spiny pocket mouse, *Heteromys desmarestianus,* a neotropical species found in mesic areas that cannot survive without free water (Fleming 1977). All three of these desert-adapted species can survive indefinitely without free water.

We sequenced kidney mRNA from these pairs of taxa, assembled and annotated *de novo* transcriptomes, analyzed rates of evolution in single copy orthologs to identify genes putatively under selection, and examined amino acid substitutions for evidence of convergence. We also performed mRNA-sequencing on multiple individuals within each species to study gene expression divergence between desert and non-desert species. Overall patterns of gene expression recapitulated the phylogeny of these six species. However, we also discovered a significantly greater number of convergent changes in gene expression between desert and non-desert species than expected by chance. In contrast, convergent changes in amino acid sequence occurred at a much smaller proportion of genes, suggesting that most convergence at the molecular level occurs in gene regulation rather than in protein structure. Finally, a small subset of genes showed convergent expression changes and evidence of positive selection on amino acid sequence; these genes are strong candidates underlying adaptation to deserts in rodents.

MATERIALS AND METHODS

Sample collection

We included five adult male mice for each species, with the exception of *H. desmarestianus*, for which only four samples could be obtained. *C. intermedius*, *Z. princeps*, and *M. musculus* were caught by N. Bittner using Sherman live traps set over-night following the guidelines of the American Society of Mammalogists (Sikes and Gannon 2011) and an ACUC protocol approved by UC Berkeley (AUP-2016-03-8536). Animals were given apple after capture to avoid dehydration for the short period of time they were in traps. Mice were euthanized by cervical dislocation, and kidney and liver were removed and preserved in RNAlater. *C. intermedius* were trapped near Tucson, AZ, USA, *Z. princeps* were trapped at Sagehen Creek Field Station near Truckee, CA, USA, and *M. musculus* were trapped near Berkeley, CA, USA. *H. desmarestianus* were collected in Chiapas, Mexico by Beatriz Jimenez, and *N. alexis* were collected by Kevin Rowe in Northern Territory, Australia. Mice collected by N. Bittner were prepared as museum specimens (skins and skulls) and deposited in the collections of the UC Berkeley Museum of Vertebrate Zoology (MVZ). Animals collected by K. Rowe were prepared as museum specimens and deposited at Museums Victoria. The collecting localities, collector's numbers, and museum catalog numbers for each specimen are provided for all wild-caught animals in Table S1. Samples from *Jaculus jaculus* were provided by Kim Cooper at UC San Diego from an outbred lab colony. Despite the fact that the *Jaculus* were from a laboratory colony while all other animals were wild caught, patterns of gene expression among all individuals reflected the phylogeny of these taxa, suggesting that the laboratory environment for *Jaculus* did not obscure overall expression patterns (see Results).

mRNA library preparation and sequencing

To target loci underlying adaptation to xeric environments, we focused on genes expressed in the kidney. RNA was extracted from each individual from kidney preserved in RNAlater using the MoBio Laboratories Powerlyzer Ultraclean Tissue & Cells RNA Isolation Kit. Remaining DNA was removed with DNAse-1 followed by a Zymo RNA Clean and Concentrator column clean-up. Due to the poor quality of some samples (RIN scores below 5), a ribosomal RNA depletion step was performed with a KAPA Riboerase Kit before

libraries were prepared with the KAPA HyperPrep Kit. Libraries were pooled and sequenced across two lanes of 150 bp PE NovaSeq (one lane of S1 and one of SP) at the Vincent J. Coates Genomics Sequencing Center at UC Berkeley. One library from each species (except *Mus musculus*; see below) was sequenced at greater depth to ensure transcriptome assembly; these were sequenced to a target of 100M read pairs while the remaining 24 libraries, intended for expression analysis, were sequenced to a target of 20M read pairs (see File S1).

Transcriptome assembly

For each of the five 100M-read-pair libraries, reads were examined for quality metrics with FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc) and then corrected by removing erroneous k-mers using rCorrector (Song and Florea 2015). Adapters and poor quality sequence were trimmed using Trim Galore! (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore). Since FastQC revealed a large quantity of duplicates within the sequenced libraries, which is likely in part due to rRNA contamination, we chose to remove all reads that mapped to known rodent rRNA from NCBI using bowtie2 (Langmead and Salzberg 2012) We ran Trinity v2.1.1 (Grabherr et al. 2011) to generate a transcriptome assembly for each species. Because transcriptome-depth (i.e. 100M read pairs) sequencing was not done for *Mus musculus*, reads from all five individuals (approximately equal to the sequencing depth for transcriptome individuals) were combined to assemble the transcriptome of a local individual as above (Table S2). To remove redundant transcripts from the Trinity assembly, transcripts with equal to or greater than 95% sequence identity were clustered with cd-hit-est (settings: -c 0.95 -n 8) (Li and Godzik 2006) to create representative transcripts before use in downstream analysis (Table S2). This is done to collapse transcript isoforms as well as to remove transcripts created by assembly errors (chimeras, duplicates, misassembled transcripts and the like). Transrate (Smith-Unna et al. 2016) was used to calculate assembly statistics. To assess assembly completeness, we used Benchmarking Universal Single-Copy Orthologs (BUSCO) (Seppey et al. 2019) to look for the 6,192 orthologs found in the Euarchontoglires odb9 database and thus expected to exist in the taxa studied here.

Transcriptome annotation and ortholog detection

To identify coding regions within our assembled transcripts for downstream analyses, we utilized TransDecoder v. 5.5.0 (http://transdecoder.sourceforge.net). We identified the longest open reading frame (ORF) and searched for matches to both the Pfam protein domain database (Bateman et al. 2004) and mouse specific SwissProt database (Bairoch and Apweiler 2000) to retain ORFs based on homology. Since high quality gene annotations were available for *M. musculus*, we used the curated RefSeq protein database for this species. Orthologous gene groups across all six taxa were identified using OrthoFinder v 2.3.3 (setting: -S diamond) (Emms and Kelly 2015). To minimize the number of alternate isoforms used in the analysis, we only used the longest ORF identified per gene.

mRNA read mapping

Raw reads from all libraries were examined for quality with FastQC. Adapters and poor quality sequence were trimmed using Trimmomatic v0.36 (Bolger et al. 2014). The five libraries that were generated for transcriptome assembly were subsampled to the average

read number of the libraries generated for expression (27,787,405 reads). Reads were mapped to transcriptomes generated for each species with Salmon v 0.14.1 (Patro et al. 2017). To compare across genera, transcripts were annotated using BLASTn to the Refseq cdna database for *Mus musculus*. Read counts were summed across transcripts for each annotated gene.

Quantification of gene expression and identification of convergent differential expression
DESeq2 (Love et al. 2014) was used to normalize for differences in library size and to call differential expression between species within each family and across all samples. As transcripts between species can differ in length, a length correction was applied. Reads were subsequently transformed with a variance stabilizing transformation for principal component analysis.

We used DESeq2 to identify convergent changes in gene expression between desert and non-desert species across all three families using an approach similar to that used by Parker et al. (2019). In particular, we fit a generalized linear model for gene expression as a function of habitat (desert vs. non-desert), family (species-pair), and their interaction. Genes were classified as convergently differentially expressed in cases where there was a significant effect of habitat (desert vs. non-desert, FDR<0.01) but no interaction effect of species-pair by habitat (FDR > 0.05). This analysis was restricted to genes with greater than an average of 20 reads per sample for each species, resulting in a total of 8,174 genes. P-values were adjusted for multiple testing using a Benjamini & Hochberg (Benjamini and Hochberg 1995) correction. Differential expression within each species pair was identified using pairwise contrasts. For pairwise contrasts, genes with a mean of fewer than 10 reads per sample were removed from the analysis.

Permutation tests were used to assess whether more genes showed convergent shifts by habitat type than expected by chance, as described in Parker et al. (2019). For each gene, read counts were randomly assigned to habitat within each species pair. All biological replicates (i.e. all five individuals) in each species were assigned to the same habitat. This process was used to create 10,000 permuted datasets. The number of convergently differentially expressed genes in these datasets were compared to that of the observed dataset

Estimating rates of molecular evolution and identification of genes under positive selection
Using the single copy ortholog groups generated by OrthoFinder for all six species, we aligned these using MAFFT through Guidance2 (Sela et al. 2015), which provided alignment quality scores for all 1,855 genes and removed those for which the alignment quality score was poor (mean column score <0.80). Alignments for which quality scores were poor were removed from subsequent analyses, resulting in a set of 1,474 genes with aligned protein coding alignments for subsequent analyses.

We used a maximum likelihood approach in a phylogenetic context by implementing the codeml package in PAML (Yang 1997) to identify genes in desert lineages with evidence of selection. We performed three analyses using the 1,474 single copy orthologs with high quality alignments present in all species. First, we defined the three desert species together

as “foreground” lineages and compared these to the three non-desert species as “background” lineages using a foreground-background branch analysis implemented in PAML (Yang 1998, 2007). This analysis estimates ω or d_N/d_S (the rate of nonsynonymous substitutions per nonsynonymous site divided by the rate of synonymous substitutions per synonymous site) and compares branches of interest (e.g., “foreground” branches) to the other “background” branches. Elevated rates of d_N/d_S compared with a null model are considered evidence for selection. This analysis was intended to identify genes underlying desert adaptation common to all three species. Second, we performed three separate foreground-background branch analyses, in which each desert species by itself was compared to the other five species. This analysis was intended to identify species-specific adaptations. Third, we performed a branch site model which allows for ω to vary both across sites in a gene and across branches on the tree.

Testing for convergent shifts in evolutionary rates

Tests for convergent shifts in evolutionary rates and convergent amino acids require an outgroup and benefit from additional taxa; therefore we included several additional non-desert species in these analyses. Sequence data for orthologs were downloaded for the American beaver (*Castor canadensis*, C.can_genome_v1.0), the guinea pig (*Cavia porcellus*, Cavpor3.0), the brown rat (*Rattus norvegicus*, Rnor_6.0) and the prairie vole (*Microtus ochrogaster,* MichOch1.0) from Ensembl (Figure S4). The longest annotated transcript for each gene was used for alignments. Alignments were performed as described above for each single copy ortholog (1,351; 868 retained after alignment cutoff). Branch lengths were estimated using the baseml program of the PAML package under the general reversible process (REV) model for nucleotide substitution rates.

To identify convergent shifts in relative evolutionary rates (RER) among desert taxa, we used the package RERconverge (Kowalczyk et al. 2019). RERconverge calculates relative branch lengths by normalizing branches for a focal gene to the distribution of branch lengths across all genes, enabling the identification of convergent changes in evolutionary rates among focal taxa while also accounting for differences in rates across taxa and phylogenetic divergence.

Identification of convergent amino acid substitutions

In the strictest definition, a convergent amino acid substitution at a particular site is represented by an identical amino acid that is shared by all species of a convergent phenotype to the exclusion of species that do not share this phenotype as described by Zhang and Kumar (1997). However, this definition may be overly restrictive in situations where alternate amino acids may have similar biochemical properties and serve the same function. Several more permissive, and maybe more biologically relevant, methods have been proposed that take into account convergent amino acid profiles rather than specific identical substitutions. Here we use the “Profile Change with One Change” (PCOC) model (Rey et al. 2018). This method accounts for shifts in the profile of amino acids along branches leading to convergent phenotypes. Amino acid profiles are the stationary distribution of amino acid frequencies, empirically built from alignments, and assumed to reflect fitness (where more frequent amino acids have higher fitness). Thus, this method identifies convergent changes as those where shifts occur in the amino acids preferred at a

given site. Simulations suggest that this method has higher specificity and sensitivity than other methods (Rey et al. 2018). To implement PCOC, protein alignments were generated for the longest annotated transcript of available single copy orthologs for the same set of ten species described above (1,351 genes) (Figure S4). For each gene, branch lengths were recalibrated utilizing the aaml option of the codeml package in PAML under a poisson distribution (model 0). We used these as input for PCOC with "scenario" set to include branches leading to *C. intermedius*, *J. jaculus*, and *N. alexis* to identify convergent amino acid changes in these lineages.

Enrichment analyses
For gene sets of interest, GO category enrichment tests were performed with GOrilla (Eden et al. 2009) to test a foreground gene set of interest against a background set of all other genes included in the analysis. Phenotype enrichment tests were performed with modPhea (Weng and Liao 2017) using the same framework.

RESULTS

Sequencing, assembly, and annotation
We generated on average ~123 million reads per sample for the assembly of *de-novo* kidney transcriptomes in each species. For *Mus musculus*, five smaller libraries were concatenated for assembly. After read correction, quality filtering, and adapter trimming, each library had an average of ~103 million reads which were used for the assembly. Each assembly contained 965,227 transcripts on average. We reduced the number of redundant transcripts in the assembly to improve accuracy of downstream analyses by clustering similar transcripts together using CD-HIT-EST. This decreased the number of transcripts by ~20% per sample to an average of 793,887 transcripts (Table S2). We used BUSCO to check assembly completeness to determine how many of the 6,192 orthologs found in the Euarchontoglires odb9 were present in our assembled transcriptomes. The six assemblies ranged in completeness from 80 - 87% (Figure S1). This level of completeness reflects a single tissue (kidney) taken at one developmental time point. After ORF prediction, we annotated each transcript to known *M. musculus* proteins. We were able to assign transcripts to 395,029 putative ortholog groups.

Global gene expression reflects phylogenetic relationships and habitat type
To identify patterns of differential gene expression, we sequenced kidney mRNA from additional individuals in each of the six species for an average of ~27 million reads per individual. We retrieved 13,305 genes in *C. intermedius*, 11,749 genes in *H. desmarestianus*, 14,891 genes in *J. jaculus*, 14,380 genes in *Z. princeps*, 18,622 genes in *N. alexis* and 19,913 genes *M. musculus* for which we were able to quantify expression levels.

Gene expression profiles largely recapitulated the known phylogenetic relationships of these six species (Figure 2A). Individuals within each species mostly form well-defined clusters (with the exception of a single *Heteromys* individual), and the different genera within each family share expression profiles that are more similar to each other than they are to genera in different families. Further, Muridae and Dipodidae are more similar to each other in expression profiles than either is to Heteromyidae, reflecting the known

evolutionary relationships of these families. Thus, the overall expression patterns reflect evolutionary history more than habitat type. These patterns are also seen in a principal component analysis (PCA) based on expression level co-variance (Figure 2B), where PC1 (accounting for 33% of the variance) largely reflects phylogeny.

Despite the overall phylogenetic pattern of gene expression, consistent differences in expression were seen between desert and non-desert species within each family. In particular, PC4 captures this variation, separating desert from non-desert taxa (explaining 11% of the variation) (Figure 2C).

Convergent differential expression in desert rodent kidneys
First, we quantified differential expression (DE) between desert and non-desert species within each family. In pairwise contrasts between desert and non-desert species in Heteromyidae, Dipodidae, and Muridae, we identified >4,000 genes in each comparison with evidence of significant DE (Table S3, FDR<0.01). Individual pairwise comparisons between desert and non-desert species found uniquely in each of the three families (to the exclusion of the two others) were associated with several GO categories, including cellular metabolic processes and nitrogen metabolic processes (Table S4). We identified a total of 654 genes that showed significant differential expression in all three species pairs (Figure S2), with 145 of these genes showing shifts in the same direction in each case.

To identify convergent shifts in gene expression associated with desert-living, we also modeled gene expression as a function of species pair (i.e., family), habitat, and their interaction. Convergent changes were identified as those for which there was a significant effect of habitat (FDR<0.01), but no interaction between species pair and habitat (FDR>0.05) (see Methods; Parker et al. 2019). We identified 702 genes with shared shifts in desert rodents relative to the mesic comparison (Figure 3A). This set includes all of the 145 genes identified above in pairwise tests. Thus, 8.6% (702/8,174) of genes showed convergent shifts in expression in desert rodents compared to their non-desert relatives.

Shared shifts in gene expression can be a consequence of selection in response to shared environmental pressures or stochastic processes. To ask if the observed number of genes with convergent differential expression was more than expected by chance, we performed a permutation test in which we took each gene and randomly switched habitat assignment within species pairs, while always maintaining the same label for all biological replicates within a species, to create 10,000 permuted data sets (Figure 3B, see Methods). Permuted datasets never identified more convergent genes than our observed set of convergent genes, suggesting an enrichment of convergent differential expression associated with habitat type.

Fold changes between individual desert-mesic pairs were often modest in one or more contrasts between species pairs (Figure S3); only 208 of genes with shared expression shifts showed an average of greater >0.5 $\log_2$ fold change difference between each desert-mesic species pair. The number of genes showing higher expression in desert rodents compared to non-desert relatives (335 genes, shown in blue in Figure 3A) was slightly fewer than the number of genes showing lower expression in desert rodents compared to

non-desert relatives (367 genes, shown in red in Figure 3A). Additionally, across all genes, fold changes between individual desert-mesic species were found to be significantly correlated in 2 of the 3 comparisons of species pairs (Spearman's rank correlation rho, *C. intermedius/H. desmarestianus* vs*. J. jaculus/Z. princeps, p*=0.0062*, rho*=0.03; *C. intermedius/H. desmarestianus* vs. *N. alexis/M. musculus*, *p*< 2.2e-16, *rho*=0.10; *N. alexis/M. musculus* vs*. J. jaculus/Z. Princeps, p*=0.12, *rho*=-0.017).

To identify genes and pathways of interest, we divided the set of convergently expressed genes into those that are upregulated with respect to the desert taxa in all comparisons and those that are downregulated with respect to to the desert taxa in all comparisons and performed phenotype and GO term enrichment tests on these (see methods). Genes convergently upregulated across desert rodents were enriched for several GO terms related to gene regulation, including regulation of RNA metabolic process (q=2.55 x 10^{-5}), regulation of gene expression (q=1.34E-5), and regulation of RNA biosynthetic process (3.87 x 10^{5}). Genes downregulated in desert rodents were enriched for GO terms related to metabolic processes, including metabolic process (q=1.56 x 10^{3}), organic substance metabolic process (q=3.93 x 10^{-3}), and cellular metabolic process (3.54 x 10^{-3}). Genes with evidence for convergent differential expression included genes with mouse mutant phenotypes related to kidney development and physiology or homeostasis (Table S5). For example, Aquaporin 11 (*Aqp11*) is expressed at a lower level in all desert species compared to non-desert species in all three comparisons (Figure 3C). This gene is part of a family of genes encoding membrane-integrated channels responsible for water transfer across membranes throughout the body. Aquaporins have been repeatedly implicated in studies of desert adaptation across rodents (Marra et al. 2012, 2014; Pannabecker 2015; Giorello et al. 2018). Mouse knockouts have demonstrated that *Aqp11* is necessary for proximal tubular function and the formation of healthy kidneys (Morishita et al. 2005; Tchekneva et al. 2008). In addition, *Aqp11* plays a role in salivary gland development (Larsen et al. 2010). This set also includes genes associated with human phenotypes related to kidney and renal diseases (Table S6); for example, mutations in the gene *col4a5,* which is downregulated in desert species, have been associated with Alport syndrome, a disease characterized by kidney inflammation (Köhler et al. 2019).

Genes under selection in desert lineages

Next, we tested for evidence of selection on protein coding sequences using well aligned one-to-one orthologs found in all desert-mesic species pairs (1474 genes). We searched for genes showing signatures of selection using a model that compares the rate of nonsynonymous substitutions with the rate of synonymous substitutions in a phylogenetic context, ω or d_N/d_S (Yang, 1998, 2007). We performed three analyses using the 1,474 single copy orthologs with high quality alignments present in all species. We tested whether there was evidence of selection only on desert species when compared to non-desert species. We uncovered 39 genes (39/1474= 2.6%) for which ω was significantly higher in the three "foreground" desert lineages compared with the three "background" non-desert lineages (Table S7 ; FDR<0.1). This group is enriched for phenotypes related to multiple aspects of the immune response as well as to hearing/vestibular/ear phenotypes and other aspects of osteology (Table S8). Immune genes are some of the fastest evolving genes in the genome and are disproportionately found to be under selection in many studies (Hurst and Smith

1999; Schlenke and Begun 2003; Nielsen et al. 2005). One gene of particular interest, unrelated to immunity, is FAT atypical cadherin 4 (*FAT4*) (q = 0.018). *FAT4* has been implicated in human kidney diseases (Alders et al. 2014) and is involved in normal kidney development through modulating the RET signaling pathway in mouse models (Mao et al. 2015; Zhang et al. 2019). *Fat4* homozygous knockout mice have smaller kidneys with the presence of cysts in renal tubules when compared with wild type mice and they die within a few hours of birth (Saburi et al. 2008). These phenotypes in laboratory mice make this an interesting candidate gene for future studies in desert rodents. We found three genes that showed evidence of positive selection (when the three desert species were treated together as foreground lineages) and also showed convergent shifts in gene expression (Rows 1-3 in Table S9), however they are not known to be associated with phenotypes of interest. This amount of overlap is no more than expected by chance (hypergeometric test, p=0.64). Further, these genes with convergent shifts in gene expression do not show evidence of either increased or decreased dN/dS (p = 0.68) compared with other genes.

We then tested whether ω was significantly higher in each of the three "foreground" desert lineages individually compared with the five remaining taxa. We identified 23 genes in *C. intermedius*, 19 in *J. jaculus*, and 18 in *N. alexis* where ω was significantly elevated (at FDR <0.1)(Table S10). These genes are candidates for lineage-specific adaptations. In *C. intermedius*, enriched phenotypes were related to immunity and morphological traits including kidney size, while in *J. jaculus* and *N. alexis*, enriched phenotypes terms were related to behavioral and electrophysiological traits (Table S11). In the *Chaetodipus* comparison, *Dusp4* is of some interest as it has been associated with aberrant circulating solute levels in mouse models. Deletion of this gene has been associated with increased excreted protein and altered kidney structure in diabetic mice (Denhez et al. 2019). It is also convergently differentially expressed. Overall, the amount of overlap (hypergeometric test, p > 0.06 in all comparisons) between any of these lists and differentially expressed genes between lineage pairs is no more than expected by chance (Table S9).

In the third analysis, we employed a branch-site model to identify genes in which specific codons may be under positive selection. In this approach, genes for which specific codons have a ω > 1 in the "foreground" branch (defined to include all three desert species) compared with the "background" branch are identified. Seven genes were identified (Table S12) with codons under selection in all three desert lineages, including *Coro2b*, a gene implicated in abnormal renal glomerulus morphology (Schwarz et al. 2019) and urine protein level (Rogg et al. 2017) and *Bloc1s4*, which is implicated in abnormal renal physiology (Gwynn et al. 2000). Again, there was no significant overlap with the genes identified in the differential expression analysis (p=0.47; Table S9).

Limited evidence for shifts in relative evolutionary rates associated with desert-living
Next, we performed evolutionary rate tests on orthologs to identify genes that showed shifts in evolutionary rates associated with desert-living using RERconverge (Kowalczyk et al. 2019). RERconverge compares the rate of change in focal branches (desert species) and non-focal branches (mesic species) to identify genes for which focal branches have convergent shifts in relative evolutionary rates. While we identified a small number of genes with evidence for convergent shifts in evolutionary rates (p<0.05, 25 genes), none of

these were significant after a false-discovery rate correction (FDR<0.1). Genes with evidence for convergent differential expression did not show higher relative evolutionary rates compared to genes without evidence for expression convergence (Mann-Whitney U, p=0.57).

Amino acids exhibiting convergence

Finally, we looked for evidence of convergence at the amino acid level using the PCOC framework (Rey et al. 2018). This was implemented on 1,351 genes for ten species (Figure S4). We identified 88 sites within 74 genes with evidence of convergence in amino acid sequence (posterior probability >0.99) in all desert taxa. Thus the proportion of genes showing convergence in amino acid composition (74/1351=5.5%) is smaller than the proportion of genes showing convergence in expression (702/8174=8.6%; see above) with no significant overlap (Table S9). These genes were not enriched for any GO terms. Of the genes with phenotype annotations relating to kidney morphology and osmoregulation (Table S13), there are multiple interesting candidates including plasmalemma vesicle associated protein (*Plvap*). *Plvap* is involved water and solute movement in organs through capillary endothelia, and mouse mutants develop smaller kidneys (Herrnberger et al. 2012).

DISCUSSION

The molecular basis of convergent evolution has been well studied for a number of simple traits, but has been less studied for complex traits. Even fewer studies have compared convergence in both gene expression and protein evolution for complex traits. Here, we studied convergence in gene expression and amino acid sequence in three species of desert rodents and their non-desert relatives. We chose species with high measures of urine concentration, a proxy for increased osmoregulatory capacity, from across the rodent tree representing ~70 million years of evolution.

Despite the long evolutionary timeframe, we identified a large number of genes (702/8174=8.6%) that showed convergent changes in gene expression. This number is far more than expected by chance. However, we caution that this number does not reflect the number of causative changes (i.e. mutational events in evolution), since many of these convergent changes in expression might reflect downstream consequences of a smaller number of changes at upstream regulators that govern networks of co-regulated genes. Nonetheless, the large number of convergent changes in expression suggests that a measureable amount of desert adaptation is mediated by a large set of shared changes in gene regulation, whether at the level of individual genes or through sets of co-regulated genes.

In addition to these shared changes in gene expression, we identified a large number of species-specific changes in gene expression in each species pair.

While this paper represents fairly long evolutionary timescales, we previously showed that a desert population of *Mus musculus* may have adapted to survive in desert conditions in as few as 200-300 years with associated changes in kidney gene expression (Bittner et al.,

chapter 1). The previous study identified 3,935 differentially expressed genes of which 99 are found in this study to be convergent across all three desert lineages. The lack of significant overlap (hypergeometric test, p=0.99) suggests that over long evolutionary timescales, adaptive responses to xeric conditions may be quite different from the evolved changes in gene expression on short evolutionary timescales.

In contrast to the large number of convergent changes in gene expression, we observed only a modest number of convergent changes in amino acid sequence (74/1351=5.5%). An even smaller number of genes showed evidence of positive selection (39/1474= 2.6%), although these proportions are not directly comparable since the methods used to detect convergence and positive selection are quite different. Nonetheless, our analyses suggest that the phenotypic convergence seen in urine concentration is reflected at the molecular level more in patterns of gene regulation than in patterns of protein evolution.

Although expression evolution and amino acid sequence evolution has been found to be correlated in some cases, we did not find significant overlap in the number of genes showing convergent gene expression and convergent amino acid sequence evolution. The small amount of overlap might reflect differences in the selection pressures on these two classes of changes. For example, *cis*-regulatory changes in gene expression are often controlled in a tissue-specific and developmental-stage-specific manner, and as such are expected to be less pleiotropic and thus less constrained in evolution (e.g. Wray 2007). Protein-coding changes, on the other hand, affect all tissues and developmental stages in which the protein is expressed and thus may be more pleiotropic and consequently more constrained. However, the small amount of overlap might also reflect both statistical and methodological limitations of our study. The analytic methods used to detect convergent expression changes and convergent amino acid changes are quite distinct and likely have different false-negative and false-positive rates. In addition, we only studied kidney gene expression in one sex (male) and at one developmental stage (adult). Gene expression varies considerably during kidney development (Schwab et al. 2003) and early expression is undoubtedly important in morphological differences between desert and non-desert kidneys. Finally, kidneys have a heterogenous cellular composition, and changes in cellular composition between species are likely to affect measures of gene expression in bulk preparations.

Despite these caveats, we identified a small number of intriguing candidate genes underlying desert adaptation, including some that showed both convergent gene expression and convergent amino acid sequence evolution. The target available to selection in a trait as complex as desert adaptation is likely large and constrained along each lineage to a different degree by other aspects of the organism's morphology and physiology. Despite this, an interesting outcome of our analysis is that the genes and pathways identified here are similar to those identified in other studies of rodent and mammalian desert adaptation (Marra et al. 2012, 2014; Wu et al. 2014; MacManes 2017; Giorello et al. 2018; Tigano et al. 2020). It is clear that gene families such as aquaporins, which are responsible for facilitating water transport across membranes, and solute carriers, may play a role in mitigating water loss across multiple systems and therefore underlie convergent evolution at the genetic level to desert environments.

ACKNOWLEDGEMENTS

I thank Katya Mack and Michael Nachman for their substantial contributions to this paper and the design and execution of this study. I also thank members of the Nachman Lab for valuable comments and discussions. I thank Dr. Kim Cooper, Dr. Kevin Rowe, and Dr. Beatriz Otero Jiménez for generously sharing their samples as well as Dr. Andy Gloss, Dr. Shea Lambert, Dr. Aaron Ragsdale, Dr. Taichi Suzuki, Ned McAllister and the Oakland Zoo, and Kim Hemmer and Golden Gate Fields for their help with field collections. I thank Dr. Jim Patton, Dr. Andy Gloss, Lydia Smith, and Jeremy Richardson for their technical expertise. This work was supported by a NSF Doctoral Dissertation Improvement Grant to NKJB (1601827), grants from the Wilhelm L.F. Martens fund and David and Marvalee fund through the Museum of Vertebrate Zoology, and an NIH grant to MWN (R01 GM127468).

DATA ACCESSIBILITY STATEMENT

Illumina sequencing data from this is stored in the NCBI Sequence Read Archive (SUB7183219). Samples collected by NKJB are accessioned into the Museum of Vertebrate Zoology collection.

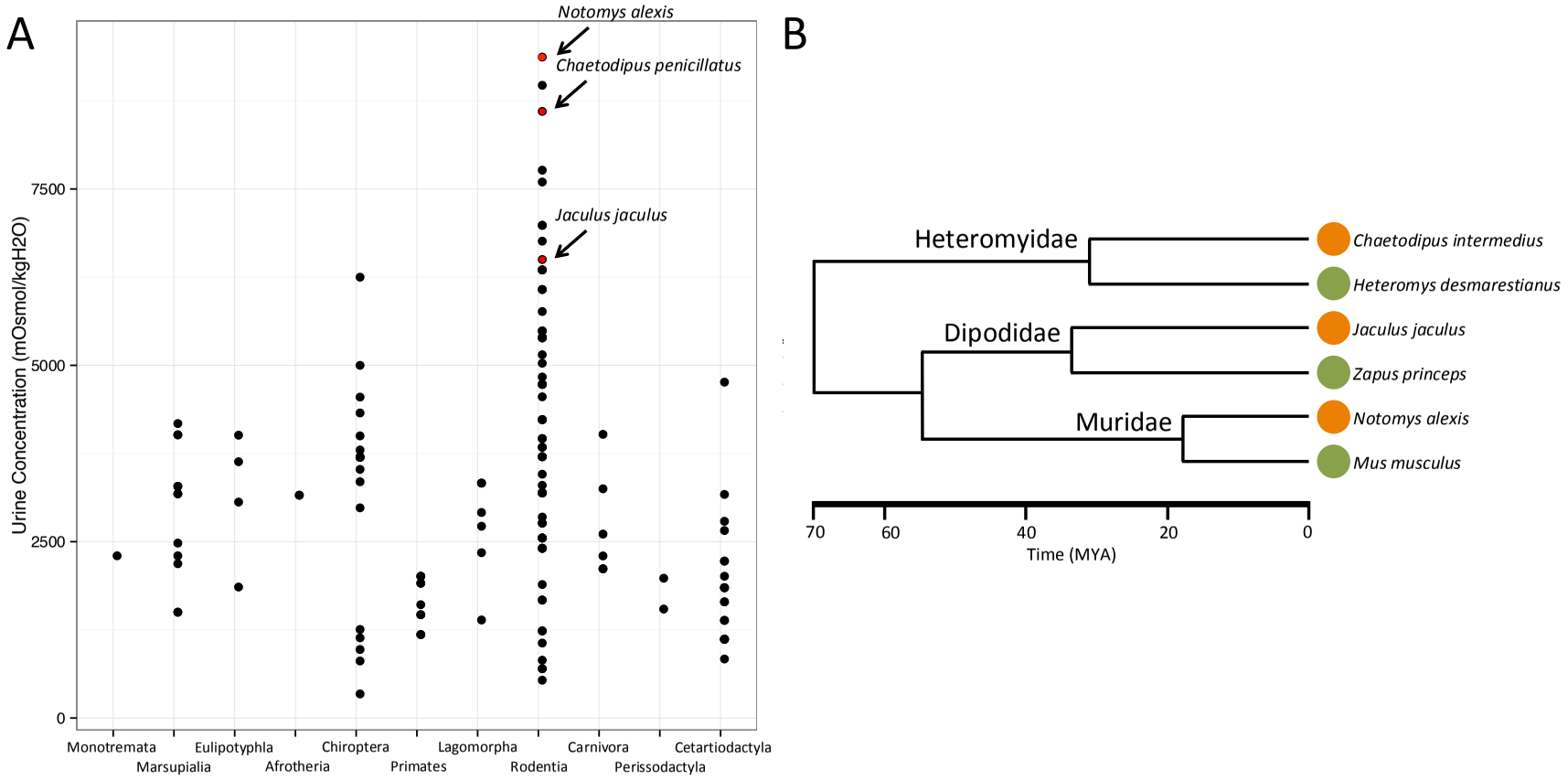

Figure 1. A. Estimates of urine concentration from across Mammalia accumulated by Beuchat 1990. Notably, Rodentia has representatives with the highest urine concentrations recorded in mammals. The three desert specialists in this study have among the highest urine concentrations measured in rodents. Note, while *C. intermedius* has not been measured for urine concentration, *C. penicillatus* is its sister taxon and found in the same environment. B. Phylogenetic relationships of target species coded by habitat type (desert in orange, non-desert in green). Divergence time estimates from TimeTree.org.

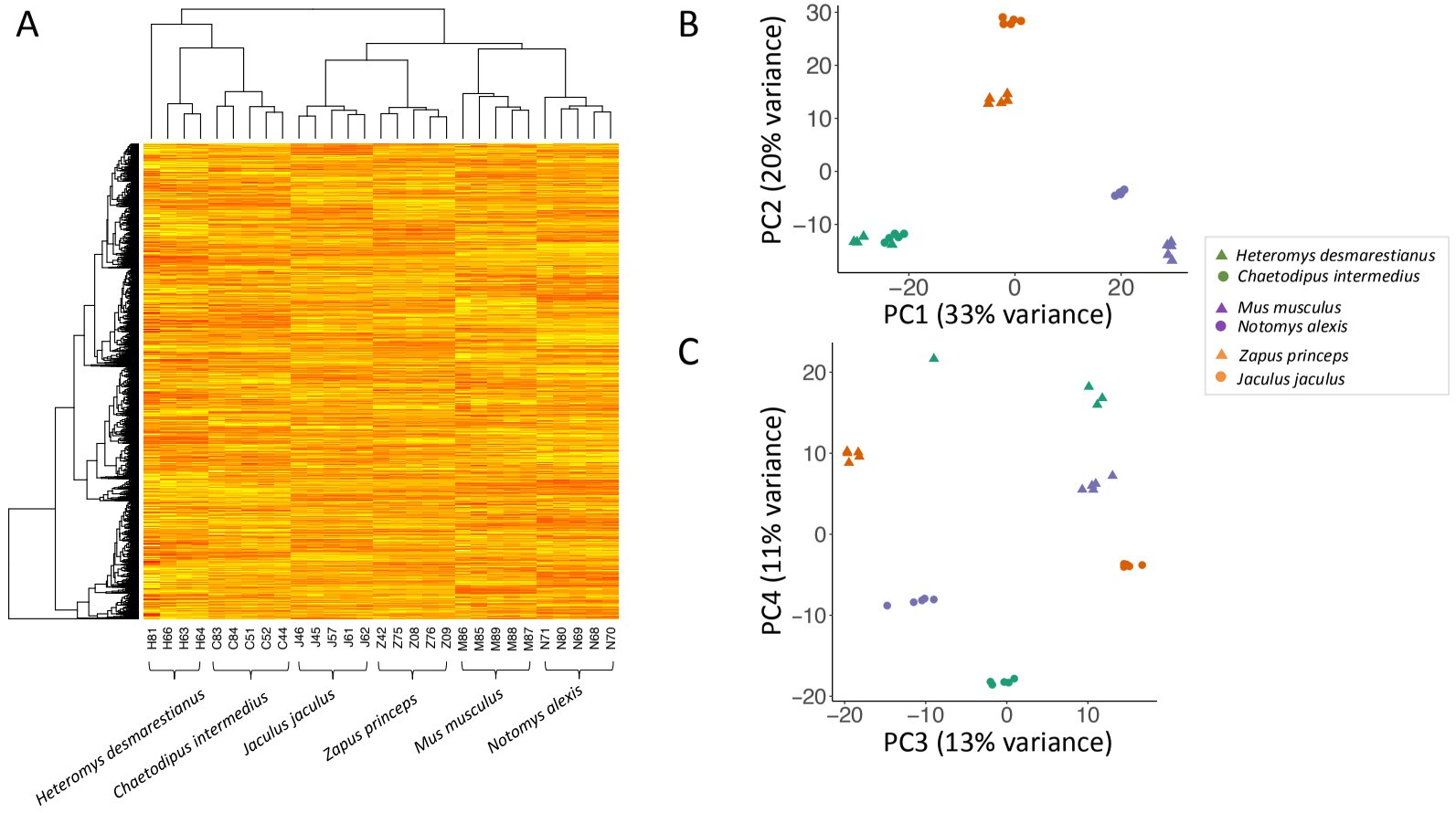

Figure 2. Expression level variation differentiates species and habitat type. A) Heat map showing relationships among samples based on gene expression clustering. With the exception of one sample (H81), expression patterns reflect phylogenetic relationships (see Figure 1). B) Principal components (PC1 and PC2) for the expression data. PC1 explains 33% of the variance and reflects the phylogenetic relationships of the species. PC2 explains 20% of the variance. C) Principal components (PC3 and PC4) for the expression data. PC4 explains 11% of the variance and differentiates samples by habitat type.

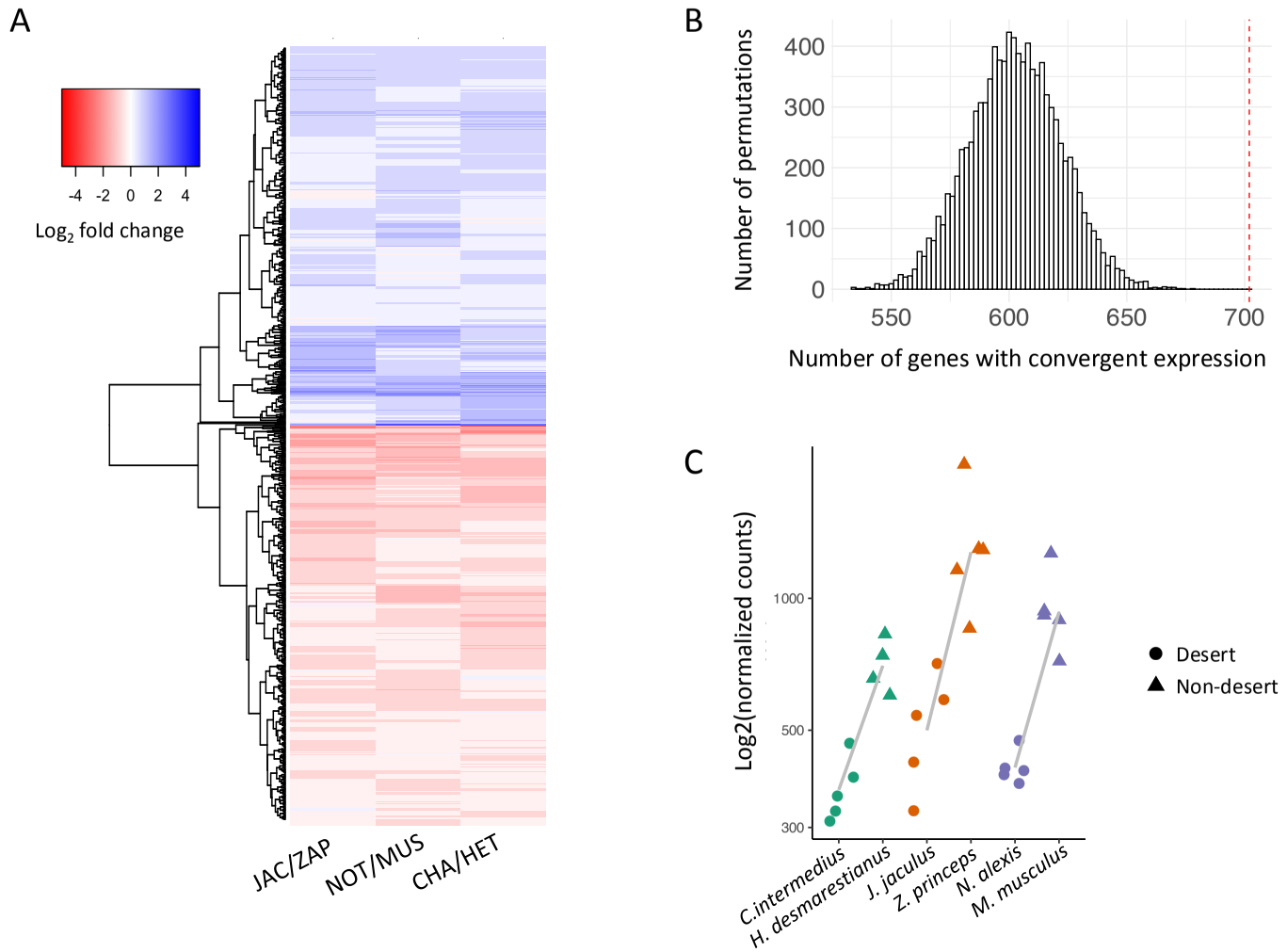

Figure 3. A) Heatmap of the 704 genes with evidence of convergent gene expression patterns. Each row is a gene. Each of the three columns shows the mean expression value among all desert individuals compared to the mean expression value of all non-desert individuals for each family. B) Number of genes expected by chance to show convergent expression after 10,000 permutations. Observed number of genes (blue line) is greater than the distribution expected by chance ($p < 0.0001$). C) Expression values for *Aqp11*, a gene showing convergent gene expression. In all comparisons, desert species show lower expression levels compared to their non-desert relative.

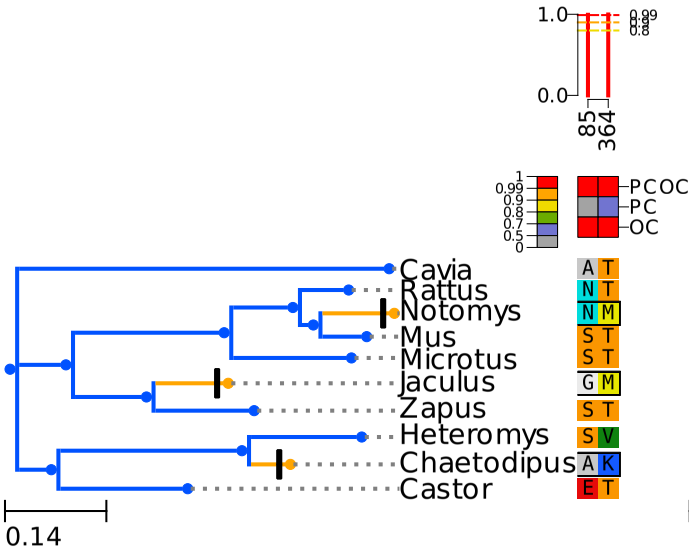

Figure 4. Convergent amino acid substitutions were identified across desert species using the PCOC model with *Plvap* represented here. PCOC, OC, and PC posterior probabilities are color coded above amino acids and below alignment site. Orange branches represent desert lineages.

Table 1. Transcriptome assembly statistics

	Number of contigs	Contig N50	Number of annotated transcripts
C. intermedius	829275	1277	141461
H. desmarestianus	560083	893	79843
J. jaculus	952270	782	106708
Z. princeps	813290	977	113307
N. alexis	814518	1324	137036
M. musculus	861908	1499	N/A

SUPPLEMENT

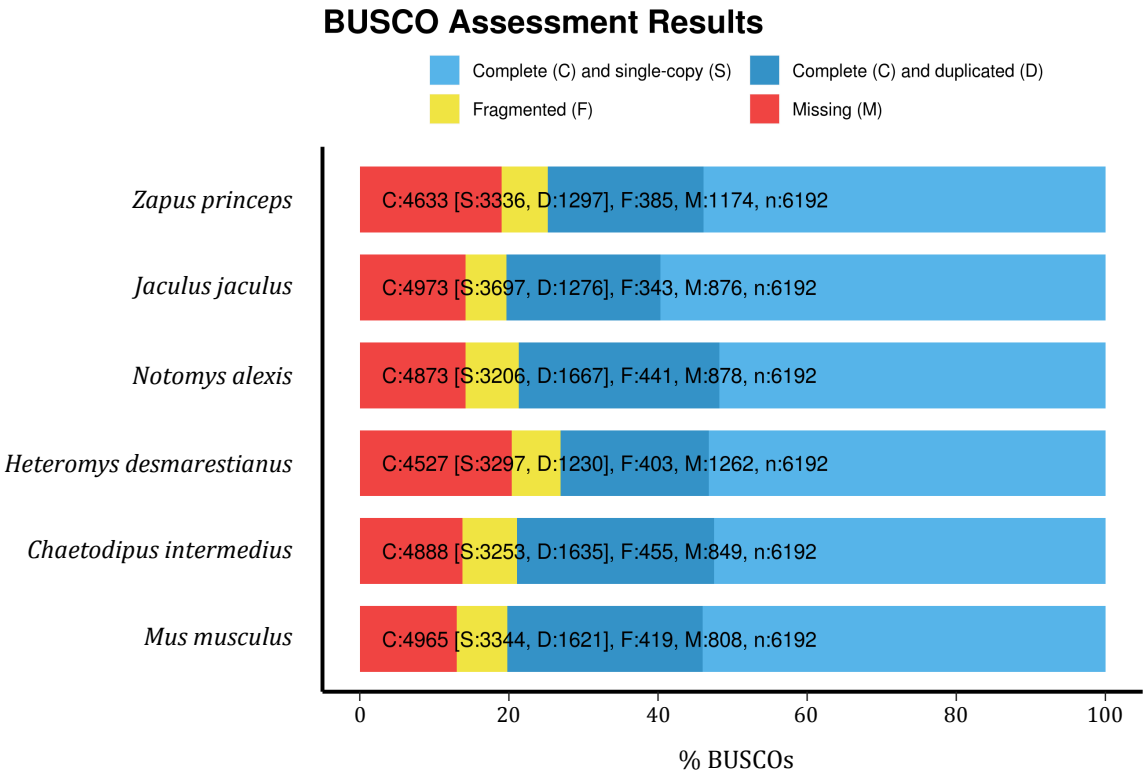

Figure S1. Benchmarking Using Single Copy Orthologs (BUSCO) score for each of the transcriptome assemblies

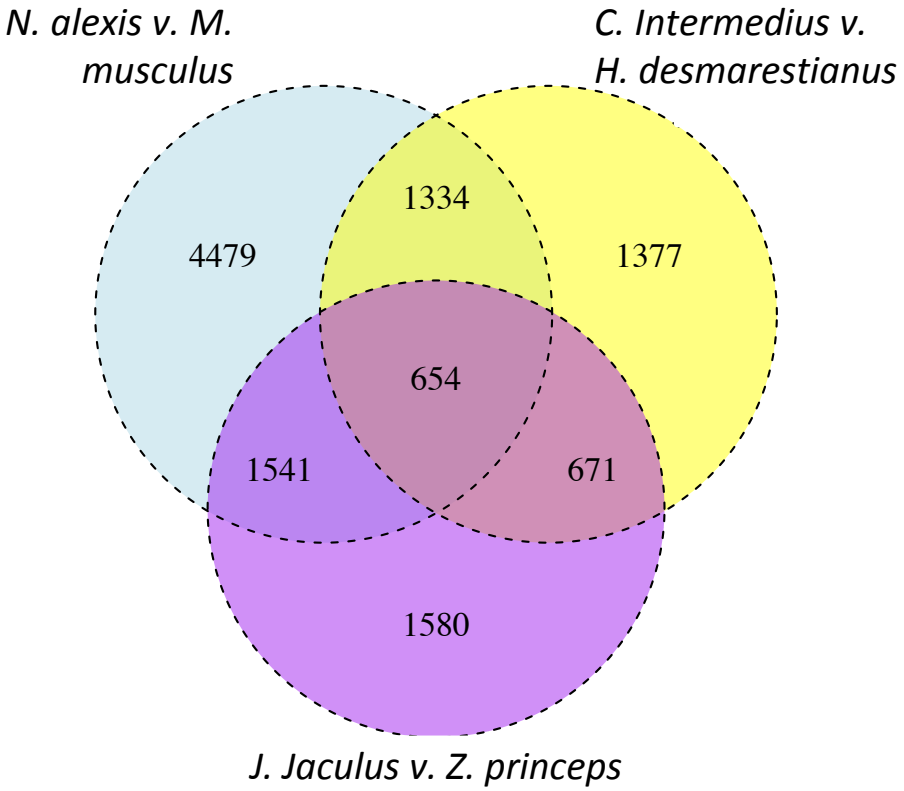

Figure S2. Genes that are differentially expressed between each desert and non-desert comparisons within each family and the overlaps among families.

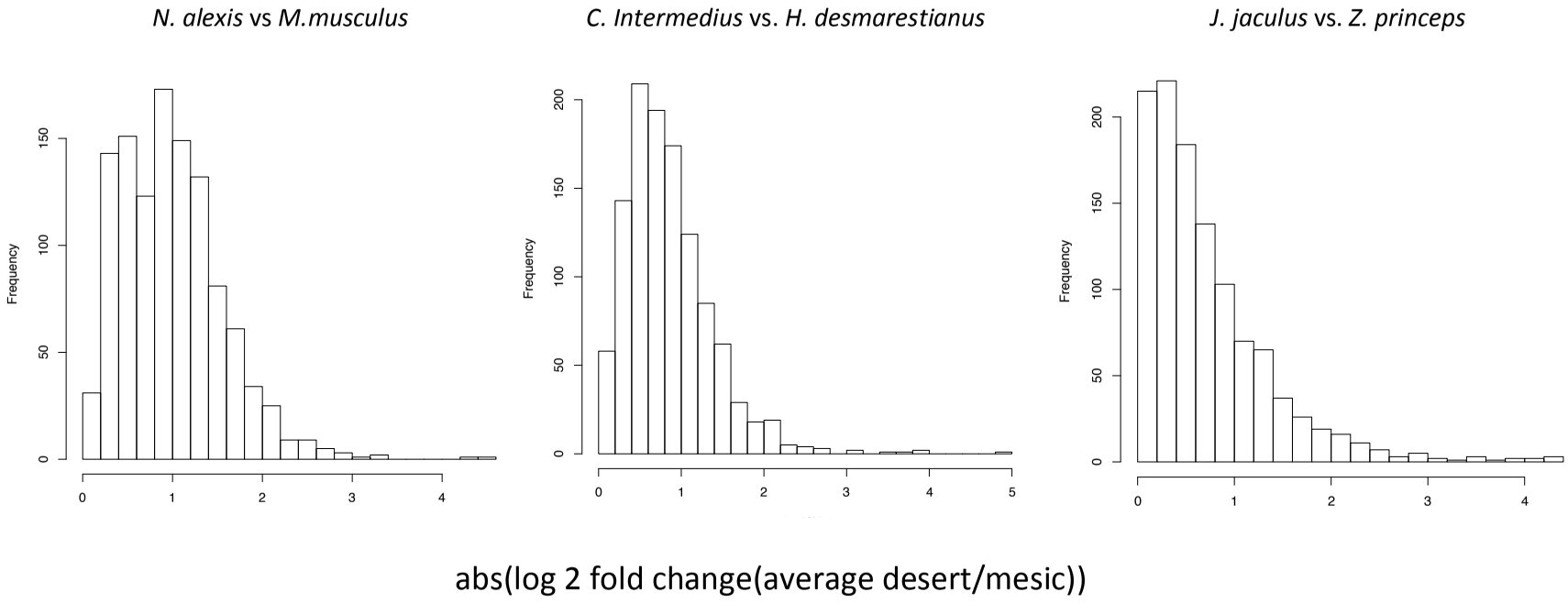

Figure S3. Magnitude of expression differences between each desert-mesic species pair

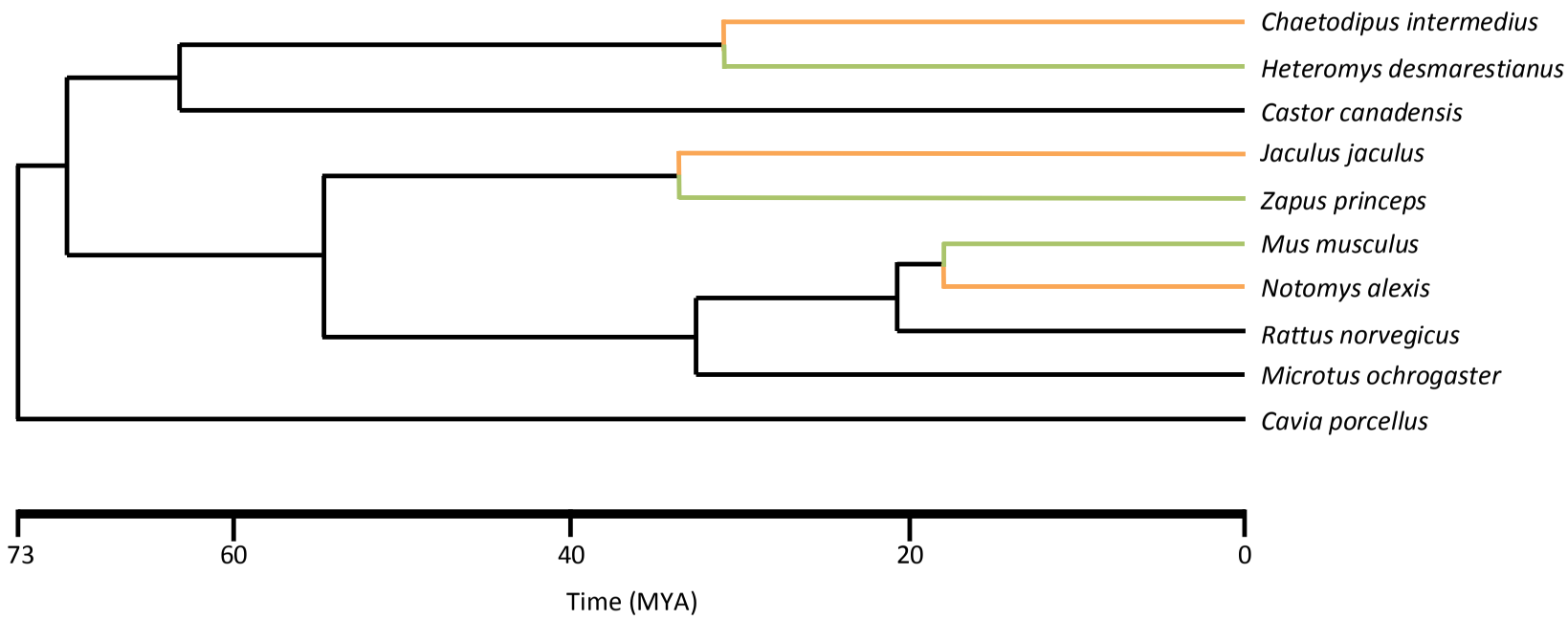

Figure S4. Tree including all samples including both ones generated for this project and externally downloaded. Branches in orange represent desert species sequenced for this project. Branches in green represent non-desert comparisons sequenced for this project.

Table S1. Localities for samples sequenced for this study. * denotes samples used for transcriptome assembly

Family	Species	Collection ID	Accessioned	Collector	Locality
Heteromyidae	*C. intermedius*	NJB20	Museum of Vertebrate Zoology	Noëlle Bittner	Avra Valley Road, Pima Co., AZ, USA
Heteromyidae	*C. intermedius*	NJB24	Museum of Vertebrate Zoology	Noëlle Bittner	Avra Valley Road, Pima Co., AZ, USA
Heteromyidae	*C. intermedius*	NJB25	Museum of Vertebrate Zoology	Noëlle Bittner	Avra Valley Road, Pima Co., AZ, USA
Heteromyidae	*C. intermedius*	NJB28	Museum of Vertebrate Zoology	Noëlle Bittner	Avra Valley Road, Pima Co., AZ, USA
Heteromyidae	*C. intermedius*	NJB29*	Museum of Vertebrate Zoology	Noëlle Bittner	Avra Valley Road, Pima Co., AZ, USA
Heteromyidae	*H. desmarestianus*	279	Museum of Vertebrate Zoology	Beatriz Jimenez	Finca Irlanda Research Station, Chiapas, Mexico
Heteromyidae	*H. desmarestianus*	1003	Museum of Vertebrate Zoology	Beatriz Jimenez	Finca Irlanda Research Station, Chiapas, Mexico
Heteromyidae	*H. desmarestianus*	1004	Museum of Vertebrate Zoology	Beatriz Jimenez	Finca Irlanda Research Station, Chiapas, Mexico
Heteromyidae	*H. desmarestianus*	1005*	Museum of Vertebrate Zoology	Beatriz Jimenez	Finca Irlanda Research Station, Chiapas, Mexico
Dipodidae	*J. jaculus*	JJ108	Cooper Lab, UCSD	Cooper Lab	Breeding colony, Cooper Lab, UCSD
Dipodidae	*J. jaculus*	JJ0200	Cooper Lab, UCSD	Cooper Lab	Breeding colony, Cooper Lab, UCSD
Dipodidae	*J. jaculus*	JJ0206	Cooper Lab, UCSD	Cooper Lab	Breeding colony, Cooper Lab, UCSD
Dipodidae	*J. jaculus*	JJ223*	Cooper Lab, UCSD	Cooper Lab	Breeding colony, Cooper Lab, UCSD
Dipodidae	*J. jaculus*	JJ0192B	Cooper Lab, UCSD	Cooper Lab	Breeding colony, Cooper Lab, UCSD
Dipodidae	*Z. princeps*	NJB032	Museum of Vertebrate Zoology	Noëlle Bittner	Sagehen Creek Field Station, Nevada Co., CA, USA
Dipodidae	*Z. princeps*	NJB033*	Museum of Vertebrate Zoology	Noëlle Bittner	Sagehen Creek Field Station, Nevada Co., CA, USA
Dipodidae	*Z. princeps*	NJB034	Museum of Vertebrate Zoology	Noëlle Bittner	Sagehen Creek Field Station, Nevada Co., CA, USA
Dipodidae	*Z. princeps*	NJB36	Museum of Vertebrate Zoology	Noëlle Bittner	Sagehen Creek Field Station, Nevada Co., CA, USA
Dipodidae	*Z. princeps*	NJB37	Museum of Vertebrate Zoology	Noëlle Bittner	Sagehen Creek Field Station, Nevada Co., CA, USA
Muridae	*N. alexis*	KCR1220	Museums Victoria	Kevin Rowe	New Crown Station, Nothern Territory, Australia
Muridae	*N. alexis*	KMCR374*	Museums Victoria	Kevin Rowe	New Crown Station, Nothern Territory, Australia
Muridae	*N. alexis*	KCR1280	Museums Victoria	Kevin Rowe	Repeater Station 8396, Northern Territory, Australia
Muridae	*N. alexis*	KCR1284	Museums Victoria	Kevin Rowe	Lyndavale Rd, Northern Territory, Australia
Muridae	*N. alexis*	KCR1285	Museums Victoria	Kevin Rowe	Lyndavale Rd, Northern Territory, Australia
Muridae	*M. musculus*	NJB42*	Museum of Vertebrate Zoology	Noëlle Bittner	Urban Ore, Oakland, Alameda Co, CA, USA
Muridae	*M. musculus*	NJB43*	Museum of Vertebrate Zoology	Noëlle Bittner	Kismet Farms Barn, Martinez, Contra Costa Co, CA, USA
Muridae	*M. musculus*	NJB44*	Museum of Vertebrate Zoology	Noëlle Bittner	Golden Gate Fields, Berkeley, Alameda Co, CA, USA
Muridae	*M. musculus*	NJB45*	Museum of Vertebrate Zoology	Noëlle Bittner	Golden Gate Fields, Berkeley, Alameda Co, CA, USA
Muridae	*M. musculus*	NJB48*	Museum of Vertebrate Zoology	Noëlle Bittner	Golden Gate Fields, Berkeley, Alameda Co, CA, USA

Table S2. Transcriptome assembly statistics

Species	RIN score	No. raw reads	No. reads post trimming	No. transcripts	No. transcripts post cd-hit
Zapus princeps	8.2	103215575	85452424	986946	813290
Jaculus jaculus	8.2	117005698	105962476	1110470	952270
Notomys alexis	4.5	118017751	85297596	1017904	814518
Heteromys desmarestianus	7.3	128217251	107205068	668408	560083
Chaetodipus intermedius	9.1	149131536	133095027	1042408	829275
Mus musculus	8-9.2	149873239	136440739	1099754	861908

Table S3. Number of differentially expressed genes in all pairwise comparisons

	Genes analyzed	Differentially expressed genes	Unique DE genes
Heteromyidae	10,460	4,036	1,377
Dipodidae	12,919	4,446	1,580
Muridae	17,039	8,008	4,479

Table S4. GO categories enriched in pairwise species comparisons

	GO categories	P	Q
Heteromyidae	peptide biosynthetic process	2.50E-04	1.00E+00
	translation	4.34E-04	1.00E+00
	aspartate family amino acid metabolic process	7.72E-04	1.00E+00
Dipodidae	postsynaptic specialization, intracellular component	8.62E-06	1.61E-02
	postsynaptic density, intracellular component	1.47E-05	1.37E-02
	detection of biotic stimulus	2.43E-05	3.52E-01
Muridae	cellular metabolic process	1.68E-06	2.53E-02
	metabolic process	3.01E-06	2.27E-02
	nitrogen compound metabolic process	4.77E-06	2.39E-02
	intra-Golgi vesicle-mediated transport	6.44E-06	2.42E-02
	organic substance metabolic process	9.75E-06	2.94E-02

Table S5. Genes with evidence for convergent differential expression with phenotypes related to kidney development and physiology and homeostasis

Gene Name	ENS	Relevant Phenotypes
Col4a5	ENSMUSG00000031274	abnormal kidney morphology; pale kidney; increased urine protein level; renal tubule atrophy; dilated renal tubules; abnormal renal glomerulus morphology; abnormal renal glomerulus basement membrane morphology
Plekhm1	ENSMUSG00000034247	polycystic kidney
Morc2a	ENSMUSG00000034543	small kidney
Robo2	ENSMUSG00000052516	abnormal kidney morphology; kidney cysts
Klf7	ENSMUSG00000025959	increased kidney weight
Atp7a	ENSMUSG00000033792	decreased kidney weight; increased kidney copper level; abnormal renal tubule morphology; decreased renal glomerulus number
Gas6	ENSMUSG00000031451	decreased kidney cell proliferation; abnormal renal glomerulus morphology
Dlg5	ENSMUSG00000021782	abnormal kidney collecting duct morphology; dilated kidney collecting duct; kidney cysts; abnormal renal tubule morphology
Aqp11	ENSMUSG00000042797	abnormal kidney cortex morphology; enlarged kidney; kidney cortex cysts; kidney failure; pale kidney; decreased urine osmolality; abnormal renal tubule epithelium morphology; abnormal renal tubule morphology
Cep290	ENSMUSG00000019971	abnormal kidney collecting duct morphology; absent kidney epithelial cell primary cilium; kidney cysts; polycystic kidney; abnormal renal tubule epithelial cell primary cilium morphology
Por	ENSMUSG00000005514	decreased kidney weight
Gata2	ENSMUSG00000015053	abnormal kidney morphology; abnormal renal/urinary system morphology
Nek1	ENSMUSG00000031644	kidney cortex cysts; kidney cysts; polycystic kidney; abnormal renal tubule morphology
Ptgis	ENSMUSG00000017969	abnormal kidney morphology; kidney atrophy; kidney cortex atrophy; kidney cysts; kidney vascular congestion; abnormal renal tubule morphology; dilated renal glomerular capsule; renal fibrosis; renal necrosis
Mplkip	ENSMUSG00000012429	abnormal kidney morphology
Zdhhc5	ENSMUSG00000034075	polycystic kidney
Fstl1	ENSMUSG00000022816	small kidney; kidney papillary atrophy; abnormal kidney pelvis morphology
Etv5	ENSMUSG00000013089	abnormal kidney morphology

Table S5 (cont.). Genes with evidence for convergent differential expression with phenotypes related to kidney development and physiology and homeostasis

Gene Name	ENS	Relevant Phenotypes
Erap1	ENSMUSG00000021583	enlarged kidney
Arntl	ENSMUSG00000055116	decreased kidney weight
Abca7	ENSMUSG00000035722	small kidney
Cmpk1	ENSMUSG00000028719	polycystic kidney
Chtop	ENSMUSG00000001017	pelvic kidney
Pax2	ENSMUSG00000004231	small kidney; kidney medulla cysts; kidney failure; increased kidney apoptosis; absent kidney; abnormal kidney lobule morphology; abnormal kidney development
Frem2	ENSMUSG00000037016	absent kidney
Frem1	ENSMUSG00000059049	abnormal kidney morphology; absent kidney; single kidney; abnormal kidney development
Lamc1	ENSMUSG00000026478	single kidney; kidney failure; kidney cysts; absent kidney; abnormal kidney development; abnormal renal glomerular capsule morphology; abnormal renal glomerulus morphology
Sgsh	ENSMUSG00000005043	abnormal kidney morphology
Bcl2	ENSMUSG00000057329	small kidney; polycystic kidney; pale kidney; kidney medulla cysts; kidney failure; kidney degeneration; kidney cysts; kidney cortex cysts; increased kidney cell proliferation; increased kidney apoptosis; enlarged kidney; delayed kidney development; decreased kidney weight; abnormal kidney morphology; abnormal kidney cortex morphology; abnormal kidney blood vessel morphology; abnormal kidney medulla morphology
Zbed4	ENSMUSG00000034333	abnormal kidney morphology; enlarged kidney
Sgk1	ENSMUSG00000019970	increased urine potassium level; increased urine sodium level; abnormal renal sodium ion transport; decreased renal glomerular filtration rate
Lrrc8a	ENSMUSG00000007476	abnormal renal tubule morphology
Tusc2	ENSMUSG00000010054	abnormal renal glomerulus morphology; renal cast
Bhlhe40	ENSMUSG00000030103	abnormal renal glomerulus morphology
Tie1	ENSMUSG00000033191	abnormal renal glomerulus morphology
Itpr3	ENSMUSG00000042644	increased fluid intake
Prkab1	ENSMUSG00000029513	increased fluid intake

Table S6: Human disease phenotypes from Human Phenotype Ontology database associated with convergent differential expression

Gene name	Human Disease
Col4a5	Alport Syndrome
Nrip1	Congenital anomalies of kidney and urinary tract 3
Frem1	Renal agenesis
Pax2	Papillorenal syndrome; Renal coloboma syndrome; Renal hypoplasia bilateral
Gata3	Hypoparathyroidism; Sensorineural deafness and Renal Disease
Plce1	Nephrotic syndrome type 3

Table S7. Genes found to be under selection in all three desert taxa

Gene stable ID	Gene name
ENSMUSG00000000282	Mnt
ENSMUSG00000000439	Mkrn2
ENSMUSG00000004032	Gstm5
ENSMUSG00000004356	Utp20
ENSMUSG00000006005	Tpr
ENSMUSG00000006315	Tmem147
ENSMUSG00000012076	Brms1l
ENSMUSG00000017057	Il13ra1
ENSMUSG00000019188	H13
ENSMUSG00000019792	Trmt11
ENSMUSG00000021686	Ap3b1
ENSMUSG00000021694	Ercc8
ENSMUSG00000022707	Gbe1
ENSMUSG00000023938	Aars2
ENSMUSG00000024816	Frmd8
ENSMUSG00000025017	Pik3ap1
ENSMUSG00000026096	Osgepl1
ENSMUSG00000028409	Smu1
ENSMUSG00000028414	Fktn
ENSMUSG00000031530	Dusp4
ENSMUSG00000031595	Pdgfrl
ENSMUSG00000032377	Plscr4
ENSMUSG00000032512	Wdr48
ENSMUSG00000032593	Amigo3
ENSMUSG00000033114	Slc35d2
ENSMUSG00000034361	Cpne2
ENSMUSG00000035284	Vps13c
ENSMUSG00000039318	Rab3gap2
ENSMUSG00000039512	Uhrf1bp1
ENSMUSG00000044976	Wdr72
ENSMUSG00000045538	Ddx28
ENSMUSG00000046743	Fat4
ENSMUSG00000046756	Mrps7
ENSMUSG00000047371	Zfp768
ENSMUSG00000054889	Dsp
ENSMUSG00000062908	Acadm
ENSMUSG00000067995	Gtf2f2
ENSMUSG00000069539	Scyl2
ENSMUSG00000079469	Pigb

Table S8. Enriched phenotypes for genes found to be under selection in all three desert taxa

Phenotype ID	FDR-corrected *p* value	Description
MP:0005377	0.025	hearing/vestibular/ear phenotype
MP:0008499	0.02	increased IgG1 level
MP:0005269	0.039	abnormal occipital bone morphology
MP:0003049	0.039	abnormal lumbar vertebrae morphology
MP:0003047	0.049	abnormal thoracic vertebrae morphology
MP:0005011	0.049	increased eosinophil cell number
MP:0002460	0.049	decreased immunoglobulin level
MP:0008074	0.016	increased CD4-positive, alpha beta T ce ...
MP:0011088	0.049	neonatal lethality, incomplete penetran ...
MP:0005061	0.049	abnormal eosinophil morphology
MP:0002602	0.049	abnormal eosinophil cell number
MP:0002461	0.049	increased immunoglobulin level
MP:0004399	0.049	abnormal cochlear outer hair cell morph ...
MP:0005013	0.049	increased lymphocyte cell number
MP:0009546	0.049	absent gastric milk in neonates
MP:0012764	0.032	increased alpha-beta T cell number
MP:0008073	0.032	abnormal CD4-positive, alpha beta T cel ...
MP:0013804	0.032	decreased IgG2 level

Table S9: Overlap between sequence based tests and significantly differentially expressed genes. Rows 1-31 refer to PAML analyses, rows 32-36 refer to PCOC analyses

Analysis	Gene stable ID	Gene name
All desert taxa	ENSMUSG00000026096	Osgepl1
All desert taxa	ENSMUSG00000032512	Wdr48
All desert taxa	ENSMUSG00000034361	Cpne2
Notomys vs. all other taxa	ENSMUSG00000033114	Slc35d2
Notomys vs. all other taxa	ENSMUSG00000030204	Ddx47
Notomys vs. all other taxa	ENSMUSG00000039318	Rab3gap2
Notomys vs. all other taxa	ENSMUSG00000006378	Gcat
Notomys vs. all other taxa	ENSMUSG00000045538	Ddx28
Notomys vs. all other taxa	ENSMUSG00000043881	Kbtbd7
Notomys vs. all other taxa	ENSMUSG00000021694	Ercc8
Notomys vs. all other taxa	ENSMUSG00000021149	Gtpbp4
Notomys vs. all other taxa	ENSMUSG00000034297	Med13
Notomys vs. all other taxa	ENSMUSG00000046756	Mrps7
Chaetodipus vs. all other taxa	ENSMUSG00000028409	Smu1
Chaetodipus vs. all other taxa	ENSMUSG00000031622	Sin3b
Chaetodipus vs. all other taxa	ENSMUSG00000022336	Eif3e
Chaetodipus vs. all other taxa	ENSMUSG00000046743	Fat4
Chaetodipus vs. all other taxa	ENSMUSG00000031530	Dusp4
Chaetodipus vs. all other taxa	ENSMUSG00000020840	Blmh
Chaetodipus vs. all other taxa	ENSMUSG00000020780	Srp68
Chaetodipus vs. all other taxa	ENSMUSG00000052595	A1cf
Chaetodipus vs. all other taxa	ENSMUSG00000024120	Lrpprc
Jaculus vs. all other taxa	ENSMUSG00000033114	Slc35d2
Jaculus vs. all other taxa	ENSMUSG00000047371	Zfp768
Jaculus vs. all other taxa	ENSMUSG00000043881	Kbtbd7
Jaculus vs. all other taxa	ENSMUSG00000006378	Gcat
Jaculus vs. all other taxa	ENSMUSG00000021694	Ercc8
Jaculus vs. all other taxa	ENSMUSG00000021149	Gtpbp4
Jaculus vs. all other taxa	ENSMUSG00000030204	Ddx28
Jaculus vs. all other taxa	ENSMUSG00000045690	Wdr89
Jaculus vs. all other taxa	ENSMUSG00000000282	Mnt
All desert taxa	ENSMUSG00000003458	Ncstn
All desert taxa	ENSMUSG00000015396	Cd83
All desert taxa	ENSMUSG00000025464	Paox
All desert taxa	ENSMUSG00000027809	Etfdh
All desert taxa	ENSMUSG00000040697	Dnajc16

Table S10. Genes found to be under selection in individual lineages

Chaetodipus intermeidus		*Jaculus jaculus*		*Notomys alexis*	
Gene stable ID	Gene name	Gene stable ID	Gene name	Gene stable ID	Gene name
ENSMUSG00000004032	Gstm5	ENSMUSG00000000282	Mnt	ENSMUSG00000000282	Mnt
ENSMUSG00000017057	Il13ra1	ENSMUSG00000006005	Tpr	ENSMUSG00000006005	Tpr
ENSMUSG00000019188	H13	ENSMUSG00000006378	Gcat	ENSMUSG00000006378	Gcat
ENSMUSG00000019792	Trmt11	ENSMUSG00000021149	Gtpbp4	ENSMUSG00000021149	Gtpbp4
ENSMUSG00000020780	Srp68	ENSMUSG00000021694	Ercc8	ENSMUSG00000021694	Ercc8
ENSMUSG00000020840	Blmh	ENSMUSG00000030204	Ddx47	ENSMUSG00000030204	Ddx47
ENSMUSG00000022336	Eif3e	ENSMUSG00000033114	Slc35d2	ENSMUSG00000033114	Slc35d2
ENSMUSG00000022351	Sqle	ENSMUSG00000034297	Med13	ENSMUSG00000034297	Med13
ENSMUSG00000024120	Lrpprc	ENSMUSG00000035284	Vps13c	ENSMUSG00000035284	Vps13c
ENSMUSG00000027222	Pex16	ENSMUSG00000039318	Rab3gap2	ENSMUSG00000039318	Rab3gap2
ENSMUSG00000028409	Smu1	ENSMUSG00000068566	Myadm	ENSMUSG00000043881	Kbtbd7
ENSMUSG00000031530	Dusp4	ENSMUSG00000043881	Kbtbd7	ENSMUSG00000044442	N6amt1
ENSMUSG00000031622	Sin3b	ENSMUSG00000044442	N6amt1	ENSMUSG00000045538	Ddx28
ENSMUSG00000032377	Plscr4	ENSMUSG00000045538	Ddx28	ENSMUSG00000045690	Wdr89
ENSMUSG00000032512	Wdr48	ENSMUSG00000045690	Wdr89	ENSMUSG00000046756	Mrps7
ENSMUSG00000038859	Baiap2l1	ENSMUSG00000046756	Mrps7	ENSMUSG00000047371	Zfp768
ENSMUSG00000041895	Wipi1	ENSMUSG00000047371	Zfp768	ENSMUSG00000049504	Proser1
ENSMUSG00000041945	Mfsd9	ENSMUSG00000067995	Gtf2f2	ENSMUSG00000067995	Gtf2f2
ENSMUSG00000046743	Fat4	ENSMUSG00000022724	Riox2		
ENSMUSG00000052595	A1cf				
ENSMUSG00000063800	Prpf38a				
ENSMUSG00000068566	Myadm				
ENSMUSG00000079469	Pigb				

Table S11. Enriched phenotypes for genes found to be under positive selection in lineage-specific PAML analyses.

Chaetodipus intermedius		*Jaculus jaculus*		*Notomys alexis*	
Description	FDR-corrected *p* value	Description	FDR-corrected *p* value	Description	FDR corrected *p* value
abnormal CD4-positive, alpha beta T cel ...	0.044	abnormal behavioral response to light	0.012	abnormal behavioral response to light	0.012
abnormal IgG1 level	0.04	abnormal impulse conducting system cond ...	0.018	abnormal impulse conducting system cond ...	0.018
abnormal interatrial septum morphology	0.035	abnormal heart electrocardiography wave ...	0.021	abnormal heart electrocardiography wave ...	0.021
abnormal lumbar vertebrae morphology	0.008	hyperactivity	0.011	hyperactivity	0.011
abnormal myelopoiesis	0.016				
abnormal presacral vertebrae morphology	0.007				
abnormal rib morphology	0.046				
abnormal tail morphology	0.007				
abnormal thoracic vertebrae morphology	0.011				
abnormal vertebrae morphology	0.031				
altered susceptibility to parasitic inf ...	0.031				
decreased rib number	0.018				
increased alpha-beta T cell number	0.044				
increased CD4-positive, alpha beta T ce ...	0.03				
increased IgG level	0.045				
increased IgG1 level	0.004				
small kidney	0.046				

Table S12. Genes with codons found to be under selection in all desert taxa

Gene stable ID	Gene name
ENSMUSG00000006005	Tpr
ENSMUSG00000026885	Ttll11
ENSMUSG00000038102	Trappc11
ENSMUSG00000041729	Coro2b
ENSMUSG00000043207	Zmpste24
ENSMUSG00000044442	N6amt1
ENSMUSG00000060708	Bloc1s4

Table S13. Amino acids with convergent substitutions identified by PCOC with phenotypes related to kidney morphology and osmoregulation

Gene stable ID	Gene name	Site	PCOC	PC	OC	Phenotypes
ENSMUSG00000002980	Bcam	63	0.9954	0.9491	0.9816	abnormal colon morphology, abnormal glomerular capillary morphology, abnormal jejunum morphology, abnormal renal glomerulus morphology
ENSMUSG00000002980	Bcam	328	0.9995	0.4349	0.9995	
ENSMUSG00000007038	Neu1	49	0.9958	0.6253	0.9988	abnormal urine homeostasis, accumulation of giant lysosomes in kidney/renal tubule cells, decreased circulating total protein level, distended urinary bladder, hydronephrosis, ischuria
ENSMUSG00000003549	Ercc1	91	0.9992	0.3407	0.9998	abnormal kidney physiology, dilated renal tubules, increased urine protein level
ENSMUSG00000028164	Manba	551	0.9980	0.6539	0.9967	abnormal renal tubule morphology
ENSMUSG00000028673	Fuca1	265	0.9977	0.3877	0.9992	abnormal urine homeostasis, enlarged urinary bladder,
ENSMUSG00000028673	Fuca1	430	0.9935	0.6063	0.9952	
ENSMUSG00000028164	Manba	753	0.9976	0.7057	0.9987	abnormal renal tubule morphology
ENSMUSG00000029007	Agtrap	119	0.9975	0.1705	0.9997	abnormal blood homeostasis, decreased urine pH, hypervolemia, polyuria
ENSMUSG00000002812	Flii	1051	0.9977	0.8954	0.9962	abnormal kidney morphology
ENSMUSG00000034845	Plvap	85	0.9936	0.4339	0.9976	abnormal circulating protein level, abnormal kidney capillary morphology, decreased circulating calcium level, decreased circulating total protein level,
ENSMUSG00000034845	Plvap	364	0.9971	0.5978	0.9990	
ENSMUSG00000005981	Trap1	28	0.9918	0.8342	0.9947	
ENSMUSG00000005981	Trap1	140	0.9943	0.3027	0.9976	decreased circulating creatinine level
ENSMUSG00000005981	Trap1	694	0.9959	0.1593	0.9991	
ENSMUSG00000010047	Hyal2	245	0.9934	0.7443	0.9937	abnormal blood homeostasis, increased kidney iron level
ENSMUSG00000032602	Slc25a20	57	0.9902	0.3574	0.9975	double ureter, increased circulating sodium level

Chapter 3

A *de novo* assembly of the genome of the rock pocket mouse (*Chaetodipus intermedius*)

ABSTRACT

The rock pocket mouse (*Chaetodipus intermedius*) is a specialized desert rodent that has been well studied for the variation in pelage color it exhibits throughout its range. Despite being an early example of a species in which a single gene underlying an adaptive phenotype in natusre was identified, an assembled genome has until now been unavailable. Here, we provide a low-coverage draft genome from a wild-caught female *C. intermedius* from Pima County, Arizona, USA assembled using MaSuRCA from short read and mate pair libraries sequenced on an Illumina platform. We generated a 15.3x coverage 1.98 Gb assembly with a scaffold N50 of 83 Kb and L50 of 5,263. The assembly had representation from 66.5% of conserved single-copy orthologs found throughout euarchontoglires. This genome provides a resource for further study on the genetics of this species.

INTRODUCTION

As the number of publically available and taxonomically diverse genome assemblies increases, so too does our ability to explore the underlying genomic architecture of a host of adaptive phenotypes in non-model organisms. The rock pocket mouse, *Chaetodipus intermedius* (Merriam 1889), formerly *Perognathus intermedius*, is a small nocturnal, desert rodent in the family Heteromyidae. It is distributed across rocky areas of the southwestern United States and northern Mexico, principally in the Sonoran and Chichuahuan deserts. Aspects of its physiology (Bradley et al. 1975), diet (Reichman 1975), reproduction (Reichman and Van De Graaff 1973), behavior (Rebar 1995), habitat (Hoover et al. 1977), morphology (Dice and Blossom 1937; Weckerly and Best 1985; Dayan and Simberloff 1994), and taxonomy have been studied.

While the number of well assembled rodent genomes continues to increase, the best studied of these, the house mouse, *Mus musculus*, shared a common ancestor with *C. intermedius* approximately 70 million years ago. We have much to learn about other more specialized lineages of rodents. In recent years, genomes have been made available on the NCBI genome browser for three species in two genera of the Heteromyidae: two species of *Dipodomys* and one species of *Perognathus*, but a *Chaetodipus* genome has not yet been made public. As yet, there are no published genomes for the remaining three genera of this family: *Heteromys, Liomys* or *Microdipodops* (Figure 1).

Like many members of the Heteromyidae, rock pocket mice are striking in their ability to thrive in extreme desert conditions. This family is of particular interest in studies of desert adaptation because four of the six genera inhabit arid ecosystems in North America (Alexander and Riddle 2005). To mitigate the effect of evaporative water loss despite high temperatures and to maintain salt and water homeostasis despite low-to-seasonally absent free water, many members of this family have developed modifications associated with desert living (Schmidt-Nielsen 1964). For example, many species in this family do not drink water, relying instead on the small volume of metabolic water they produce from their dry granivorous diet and their ultra-efficient kidneys to produce urine more than three times as concentrated as that of the lab rat to excrete waste (reviewed in Altschuler et al. 1979; Beuchat 1990; Degen 1997; Donald and Pannabecker 2015). These observations have motivated several studies on the genomic basis of adaptation to deserts in this family in comparison to other families of desert rodents (Marra et al. 2012, 2014; MacManes 2017; Giorello et al. 2018; Tigano et al. 2019, Bittner et al., chapter 2). The sequence of the rock pocket mouse genome will allow further comparisons of the genomic architecture of desert adaptation and the genetic basis of lineage specific traits in *C. intermedius*.

A second topic of considerable interest with respect to the rock pocket mouse is the marked coat color polymorphism exhibited throughout its range associated with variation in substrate color. Mice with dark (melanic) pelage are found on darker substrates while mice inhabiting lighter rocky outcroppings are lighter in color (Benson 1933; Dice and Blossom 1937). This variation presumably confers an adaptive advantage against avian predators (Hoekstra et al. 2005). Mutations in a single gene, *Mc1r*, have been shown to be associated with color variation in one population (Nachman et al. 2003; Hoekstra et al. 2004), however other dark populations have not been found to share these mutations (Hoekstra and Nachman 2003) suggesting that multiple independent origins of adaptive melanism have arisen in this species.

The mechanism and genetics of the mammalian pigmentation pathway are well studied with around 80 genes that have been identified that alter coat color in the house mouse *Mus musculus* (reviewed in Jackson 1994; Barsh 1996) . This information has been leveraged to map the genetic basis of color variation in many other mammals (Ritland et al. 2001; Steiner et al. 2009; Jones et al. 2020). Interestingly, despite the many genes in this pathway, two genes, *Mc1r* and *Agouti* represent the majority of variants found to underlie pelage color variation in the wild. While some of this undoubtedly reflects bias due to candidate gene approaches, it sets up an important question about the repeatability and mechanism of adaptive evolution. To what extent similar or identical mutations underlie adaptive color variation in nature is not well known. In addition, the relative contribution and effect sizes of protein coding and regulatory variants are largely unknown. Therefore, while the genetic basis of coat color variation is understood in one population, this genome can form the foundation for future studies.

Further, this is an opportunity to study local adaptation in patchy landscapes. Populations of rock pocket mice are heterogeneously distributed in association with rocky outcroppings in the desert while melanic populations are restricted to lava formations. These lava formations range in age from 1000 years old to millions of years old, in size

from a few km^2 to greater than 1500 km^2 and are at varying distances from each other. This sets up a natural experiment to study the relative contribution of *de novo* mutations and migrant alleles to melanic populations in this landscape across space and time. A first application would be to further test Ralph and Coop's (2015) work on the spatial scale of selection where they modeled a critical distance between populations above which a population is likely to evolve a *de novo* mutation instead of capturing a beneficial migrant allele. This distance is proportional to the spatial scale of selection, which is influenced by the dispersal distance and the selective cost of the allele between populations (in a mismatched environment) and is dependent on the mutation rate. Additional questions about the age of adaptive alleles , the population size, and the influence of migration of non adaptive alleles in small populations could be empirically tested if the genetic basis of melanism were more completely understood. The rock pocket mouse genome provides the opportunity to use low coverage whole genomes to rapidly and efficiently identify variants underlying color variation in other populations and study convergence at the molecular level.

Here, we provide a draft *de novo* genome assembly of *C. intermedius* from a wild female collected from Avra Valley near Tucson, Arizona, USA. This population represents the ancestral wild type "light" pelage coloration and lives on rocky non-lava substrates (Hoekstra et al. 2005). We used MaSuRCA (Zimin et al. 2013) to assemble Illumina data from small insert and mate pair libraries with moderate insert sizes. This resource will allow for further studies on the evolution and natural history of this species.

MATERIALS AND METHODS

Sample collection

One adult female, MPR42 (accessioned as UAZ 27660), was live trapped in Avra Valley, Pima County, Arizona, USA (32.40179, -111.14754) using Sherman traps set over-night in accordance with the guidelines set out by the Journal of Mammalogy (Sikes and Gannon 2011) and a protocol approved by the University of Arizona Institutional Animal Care and Use Committee (IACUC). The individual was euthanized by cervical dislocation, and tissues (liver, kidney, spleen) were removed and flash frozen in liquid nitrogen. A voucher specimen (skin and skull) was prepared and deposited in the mammal collections of the Department of Ecology and Evolutionary Biology at the University of Arizona Museum of Mammalogy.

Library preparation and sequencing

DNA from the same individual was extracted and libraries were built and sequenced twice. The first round of library preparation was executed by the University of Arizona Genomics Core (UAGC). There, they prepared paired end libraries with 180bp and 500bp insert sizes and a mate pair library of 3-5kb insert size. These were each sequenced across two lanes (six lanes total) as 100bp reads from an Illumina HiSeq.

For the second round of sequencing, DNA was extracted from frozen liver using a Qiagen Puregene kit and sent to the UC Davis Genome Center for library preparation and sequencing. There, they generated a small insert library designed with a target insert of

260bp and three mate-pair libraries targeted at 3kb, 5kb, and 10kb. These libraries were sequenced on four lanes of an Illumina HiSeq3000 to produce 150bp PE reads: two lanes for the small insert library, one lane for the 3kb mate pair library, and one lane for the combination of the two larger mate pair libraries.

Genome assembly
The best assembly resulted from MaSuRCA v2.31.10 (Zimin et al. 2013) using only the second round of sequencing data (see above). This assembler works best with minimal pre-processing therefore we limited ours to adapter and read length trimming with fastp v0.20.0 (Chen et al. 2018) for the small insert library and NxTrim v0.4.3 (O'Connell et al. 2015) for mate pair libraries. We also removed reads shorter than 70bp using fastp. To better estimate sequenced insert sizes, we mapped each library to the genome of *Dipodomys ordii* (~30 MY diverged) using Minimap2 v2.17 (Li 2018). We used *D. ordii* because it is the best-assembled genome from the Heteromyidae. For this we downloaded the Dord_2.0 genome from the NCBI genome database. This assembly is 2.2Mb in size with ~65,000 scaffolds and an assembly N50 of 48,087. It was assembled with ALLPATHS-LG (Gnerre et al. 2011) from a combination of shotgun Sanger sequencing (2.5x) and Illumina NGS sequencing (181x). We used the mean and standard deviation estimated by mapping each library to configure the assembler. Using the output from NxTrim, we utilized all of the read pairs identified as paired end for contig assembly and the read pairs identified as mate pair for the scaffolding assembly. To estimate genome size from the data, we used jellyfish v.2.3.0 (Marçais and Kingsford 2011). MaSuRCA was then utilized for the assembly.

Following this, assembly completeness was benchmarked using BUSCO v.4 (Seppey et al. 2019) to check for the presence of 6,192 orthologs found in the Euarchontoglire and 3,354 orthologs found in the vertebrata odb10 databases. These are genes conserved in all species in each of these clades and therefore expected to be in *C. intermedius* . The greater the number of single copy orthologs from each of these databases found in the assembly, the more complete the assembly is assumed to be. The genome was assembled and memory-intensive downstream analyses were completed on a private lab server with 32 cores and 756 Gb of RAM. With this setup, the assembly took just over four weeks.

Additional assembly efforts
Earlier efforts to assemble these data resulted in worse quality or incomplete assemblies that failed to finish (see Discussion). The sequencing strategy was initially designed to utilize ALLPATHS-LG for assembly. Using only the first round of sequencing, this failed due to insufficient memory but was not revisited with all of the available sequence data on a more contemporary machine.

Using the combination of first and second round sequencing, assemblies were attempted with SOAPdenovo2 (Luo et al. 2012). Before assembling, the short insert reads were trimmed and adapters removed using Trimmomatic (Bolger et al. 2014) and the mate pair adapters were removed with NxTrim and Trimmomatic. Next, the reads were error corrected using a k-mer based approach with SOAPec part of the SOAPdenovo2 toolkit. The assembler was then run with the 127 k-mer option. Following this, gap closing was attempted with GapCloser from the SOAPdenovo2 toolkit but was not completed.

Genome annotation

Following the assembly process, we used RepeatMasker 4.1.0 (Smit et al. 2013) with the HMMER 3.2.1 (Eddy et al. 2018) and the "rodentia" option to repeat mask the genome. We annotated this masked genome using MAKER 2.31.10 (Cantarel et al. 2008) trained with *Dipodomys ordii* and *Mus musculus* proteomes available from the Uniprot database (The UniProt Consortium 2007) and an assembled *C. intermedius* kidney transcriptome for transcribed sequence evidence. This transcriptome was assembled from an RNA-seq library generated from an individual sampled from the Avra Valley population and assembled using Trinity v 2.1.1 (Grabherr et al. 2011) as described in chapter 2.

RESULTS

For the second round of sequencing, over four lanes, we generated $2.57x10^9$ total reads: $6.39x10^8$ pairs of small insert reads, $3.25x10^8$ pairs of 3kb mate pair reads, $1.78x10^8$ pairs of 5kb mate pair reads and $1.44x10^8$ pairs of 10kb mate pair reads (Table S1) for an estimated total of 64x small insert coverage and 64x of mate pair libraries of a 3.0 Gb genome. After mapping to the *D. ordii* genome, insert sizes were confirmed for the libraries and were within 10% of the target size. Sequence duplication levels were moderate for the short insert libraries (17.6%) but were much higher for the mate pair libraries (3kb: 76%, 5kb: 85.1%, 10kb: 87%) reported by FastQC. Further, the majority of reads in these libraries were determined to be paired end, or of unknown orientation, rather than mate pair reads. For example, only 66% of the the 3kb insert mate pair library before de-duplicating were mate pair reads that could be used for scaffolding. For the remining library: 14% were paired end reads that were of use for the contig building section of the assembly, and 18.6% were reads of unknown orientation and which are not usable in the scaffolding process. These library problems likely contributed to the weakly scaffolded assembly.

After assembly, the genome remained highly fragmented with 119,988 total scaffolds representing ~1.98Gb. The scaffold N50 was 83.0 Kb with the longest scaffold 2.0 Mb (Table 1). The assembled coverage was 15.3x of a computationally estimated genome size of 2.1 Gb. Using a k-mer based approach, the unique genome length without repeats was estimated to be 1.45Gb. We expect final assembly coverage to be less than raw read coverage in mammals due to large highly repetitive regions that do not assemble. Read pairs with overlapping inserts (as is designed in the small insert library) are merged during assembly and contribute to lower depth. Additionally, low quality and short reads were thrown out before assembly.

To estimate the completeness of the assembled genome, we used BUSCO to identify the presence and completeness of single copy orthologs conserved across euarchontoglires (and thus rodents) as well as across vertebrates. The genome had representation from 66.5% of euarchontoglire genes and 79.1% of vertebrate genes (Table 2). The euarchontoglire set is nearly four times the number of genes as the vertebrata set, which may partially account for the lower BUSCO score despite being a more specific group. The genome has been annotated to some extent and these efforts can be continued following a

more contiguous genome assembly. This assembly would benefit from additional sequencing efforts using long-range sequencing technologies to increase contiguity. Further, attempting different assembly strategies may provide a more contiguous genome.

Genome assemblies attempted with SOAPdenovo2 resulted in even more highly fragmented assemblies with 295,022 scaffolds. They were less complete with 51.3% of BUSCOs represented. Of these, only 27.1% were complete BUSCOs. This was likely due to methodological problems with the libraries, including the first round mate pair library. The inclusion of the first round of sequencing data likely contributed to this low quality assembly as many of the libraries contained smaller than target insert sizes including the mate pair library.

For the first round of sequencing, which was not included in the final assembly, we generated $1.58x10^8$ pairs of 180bp insert reads $3.77x10^8$ pairs of 500bp insert reads, and $1.18x10^8$ pairs of 3-5kb insert reads (Table S1).

DISCUSSION

The rock pocket mouse is an excellent model for understanding the genomic basis of adaptation in desert rodents because of its extreme phenotypic modifications associated with desert living. Further, nearly a century of work has formed the basis of our understanding of coat color variation in the wild, to be leveraged for studies on the origin of novel alleles in locally adapted populations and the genetic basis of convergent traits.

While the genome is not sufficiently annotated to determine conclusively the basis of the fragmented scaffolding, there are clues as to its cause. The mate pair library preparation and sequencing provided fewer usable mate pairs than expected with large portions of the libraries without enough information to be used in the assembly. In addition, these libraries had large proportions of PCR duplicates. These duplicates provide no additional information to the assembler and thus do not contribute to the assembly building. However, the small insert library was high quality and sufficient to get representation of over 65% of genes conserved across all euarchontoglires in this assembly. Future steps such as attempting a reference guided assembly to another Heteromyid rodent might result in a more complete assembly. Additionally, while the developer of MaSuRCA cautions against performing additional gap closing steps which may introduce chimeric scaffolds, this step could be attempted using reads from the first sequencing round.

Here, we provide the genome sequence for *Chaetodipus intermedius*, the fourth public genome for the family Heteromyidae to be used for future studies on the evolution, ecology, and natural history this species and for evolutionary questions more broadly.

ACKNOWLEDGEMENTS

I thank Megan Phifer-Rixey and Michael Nachman for their substantial contributions to this paper and the design and execution of this study. I thank members of the Nachman Lab for valuable comments and discussions, to Rohit Kolora, Ke Bi and Stefan Prost for their technical expertise, and to the UAGC for its sequencing efforts. This work was supported by grants from the Wilhelm L.F. Martens fund and David and Marvalee fund through the Museum of Vertebrate Zoology to NKJB and by support from UC Berkeley and the University of Arizona to MWN.

DATA ACCESSIBILITY STATEMENT

Illumina sequencing data from this is stored in the NCBI Sequence Read Archive (SUB7899990).

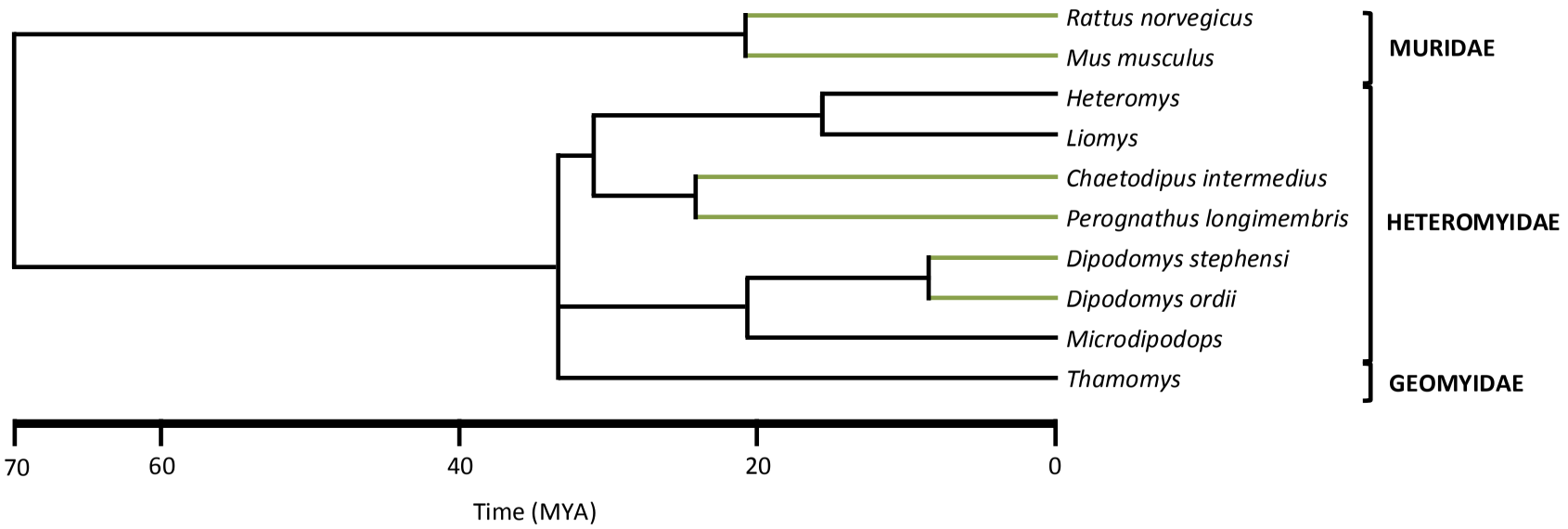

Figure 1. Phylogenetic tree displaying relationships of genera within the Heteromyidae, its sister family, the Geomyidae, and their relationships to some of the most well studied model rodents, the lab mouse and lab rat. Green lineages depict publically available genomes including the one made available by this study. Time estimates from TimeTree.org.

Table 1. Sequencing statistics

Sequencing Round	Library type	Number of lanes sequenced	Number of raw reads	Total bp	Read depth
2	260bp small insert	1	644988584	96748287600	32.2
2	260bp small insert	1	633713350	95057002500	31.7
2	3kb insert mate pair	1	650352788	97552918200	32.5
2	5kb insert mate pair	0.5	356396002	53459400300	17.8
2	10kb insert mate pair	0.5	288703704	43305555600	14.4
1	180bp small insert	2	317645698	31764569800	10.6
1	500bp small insert	2	755407218	75540721800	23.2
1	3-5kb insert mate pair	2	237496542	23749654200	7.9

Table 2. MaSurCA assembly statistics

Statistic	Scaffold	Contig
Number	119988	149886
Largest	2016071	1383913
N50	83028	49770
L50	5263	8489
GC	41.98%	
Total Length	1.98 Gb	

Table 3. BUSCO assembly completeness statistics

	Euarchontoglires	Vertebrata
Complete BUSCOs	6172 (48.6%)	1699 (50.6%)
Complete and single copy BUSCOs	6055 (47.7%)	1641 (48.9%)
Complete and duplicated BUSCOs	117 (0.9%)	58 (1.7%)
Fragmented BUSCOs	2270 (17.9%)	956 (28.5%)
Missing BUSCOs	4250 (33.5%)	699 (20.9%)
Total BUSCO groups searched	12692	3354

REFERENCES

Alders, M., L. Al-Gazali, I. Cordeiro, B. Dallapiccola, L. Garavelli, B. Tuysuz, F. Salehi, M. A. Haagmans, O. R. Mook, C. B. Majoie, M. M. Mannens, and R. C. Hennekam. 2014. Hennekam syndrome can be caused by FAT4 mutations and be allelic to Van Maldergem syndrome. Human Genetics 133:1161–1167.

Alexander, L. F., and B. R. Riddle. 2005. Phylogenetics of the New World Rodent Family Heteromyidae. Journal of Mammalogy 86:366–379.

Al-kahtani, M. A., C. Zuleta, E. Caviedes-Vidal, and Jr. Garland Theodore. 2004. Kidney Mass and Relative Medullary Thickness of Rodents in Relation to Habitat, Body Size, and Phylogeny. Physiological and Biochemical Zoology 77:346–365.

Altschuler, E. M., R. B. Nagle, E. J. Braun, S. L. Lindstedt, and P. H. Krutzsch. 1979. Morphological study of the desert heteromyid kidney with emphasis on the genus perognathus. The Anatomical Record 194:461–468.

Anders, S., P. T. Pyl, and W. Huber. 2015. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31:166–169.

Bairoch, A., and R. Apweiler. 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Research 28:45–48.

Baldwin, J. M. 1896. A New Factor in Evolution. The American Naturalist 30:441–451.

Bankir, L., and C. de Rouffignac. 1985. Urinary concentrating ability: insights from comparative anatomy. American Journal of Physiology-Regulatory, Integrative and Comparative Physiology 249:R643–R666.

Barsh, G. S. 1996. The genetics of pigmentation: from fancy genes to complex traits. Trends in Genetics 12:299–305.

Bateman, A., L. Coin, R. Durbin, R. D. Finn, V. Hollich, S. Griffiths-Jones, A. Khanna, M. Marshall, S. Moxon, E. L. L. Sonnhammer, D. J. Studholme, C. Yeats, and S. R. Eddy. 2004. The Pfam protein families database. Nucleic Acids Res 32:D138–D141.

Baudinette, R. V. 1972. The impact of social aggregation on the respiratory physiology of Australian hopping mice. Comparative Biochemistry and Physiology. A, Comparative Physiology 41:35–38.

Baum, N., C. C. Dichoso, and C. E. Carlton. 1975. Blood urea nitrogen and serum creatinine: Physiology and interpretations. Urology 5:583–588.

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

Benjamini, Y., and Y. Hochberg. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B (Methodological) 57:289–300.

Benson, S. B. 1933. Concealing coloration among some desert rodents of the southwestern United States. University of California Press. Berkeley, CA.

Beuchat, C. A. 1990. Body size, medullary thickness, and urine concentrating ability in mammals. American Journal of Physiology-Regulatory, Integrative and Comparative Physiology 258:R298–R308.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.
Bonomini, F., L. F. Rodella, M. Moghadasian, C. Lonati, R. Coleman, and R. Rezzani. 2011. Role of apolipoprotein E in renal damage protection. Histochemistry and Cell Biology 135:571–579.
Bradley, W. G., M. K. Yousef, and I. M. Scott. 1975. Physiological studies on the rock pocket mouse, Perognathus intermedius. Comparative Biochemistry and Physiology Part A: Physiology 50:331–337.
Cantarel, B. L., I. Korf, S. M. C. Robb, G. Parra, E. Ross, B. Moore, C. Holt, A. S. Alvarado, and M. Yandell. 2008. MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Research 18:188–196.
Chen, S., Y. Zhou, Y. Chen, and J. Gu. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:i884–i890.
Chou, C.-L., M. A. Knepper, A. N. van Hoek, D. Brown, B. Yang, T. Ma, and A. S. Verkman. 1999. Reduced water permeability and altered ultrastructure in thin descending limb of Henle in aquaporin-1 null mice. The Journal of Clinical Investigation 103:491–496.
Corl, A., K. Bi, C. Luke, A. S. Challa, A. J. Stern, B. Sinervo, and R. Nielsen. 2018. The Genetic Basis of Adaptation following Plastic Changes in Coloration in a Novel Environment. Current Biology 28:2970-2977.e7.
Dayan, T., and D. Simberloff. 1994. Morphological Relationships Among Coexisting Heteromyids: An Incisive Dental Character. The American Naturalist 143:462–477.
Degen, A. A. 1997. Water Requirements and Water Balance. Pp. 93–162 *in* Ecophysiology of Small Desert Mammals. Springer Berlin Heidelberg, Berlin, Heidelberg.
Denhez, B., M. Rousseau, D.-A. Dancosst, F. Lizotte, A. Guay, M. Auger-Messier, A. M. Côté, and P. Geraldes. 2019. Diabetes-Induced DUSP4 Reduction Promotes Podocyte Dysfunction and Progression of Diabetic Nephropathy. Diabetes 68:1026–1039.
Dice, L. R., and P. M. Blossom. 1937. Studies of mammalian ecology in southwestern North America with special attention to the colors of desert mammals. Carnegie Institution of Washington, Washington, D.C.
Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T. R. Gingeras. 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21.
Donald, J., and T. L. Pannabecker. 2015. Osmoregulation in Desert-Adapted Mammals. Pp. 191–211 *in* K. A. Hyndman and T. L. Pannabecker, eds. Sodium and Water Homeostasis: Comparative, Evolutionary and Genetic Models. Springer, New York, NY.
Eddy, S. R., and HMMR development team. 2018. HMMER 3.2.1.
Eden, E., R. Navon, I. Steinfeld, D. Lipson, and Z. Yakhini. 2009. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics 10:48.
Emms, D. M., and S. Kelly. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biology 16:157.
Fleming, T. H. 1977. Response of Two Species of Tropical Heteromyid Rodents to Reduced Food and Water Availability. Journal of Mammalogy 58:102–106.

Ghalambor, C. K., K. L. Hoke, E. W. Ruell, E. K. Fischer, D. N. Reznick, and K. A. Hughes. 2015. Non-adaptive plasticity potentiates rapid adaptive evolution of gene expression in nature. Nature 525:372–375.
Ghalambor, C. K., J. K. McKay, S. P. Carroll, and D. N. Reznick. 2007. Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. Functional Ecology 21:394–407.
Giorello, F. M., M. Feijoo, G. D'Elía, D. E. Naya, L. Valdez, J. C. Opazo, and E. P. Lessa. 2018. An association between differential expression and genetic divergence in the Patagonian olive mouse (Abrothrix olivacea). Molecular Ecology 27:3274–3286.
Gnerre, S., I. MacCallum, D. Przybylski, F. J. Ribeiro, J. N. Burton, B. J. Walker, T. Sharpe, G. Hall, T. P. Shea, S. Sykes, A. M. Berlin, D. Aird, M. Costello, R. Daza, L. Williams, R. Nicol, A. Gnirke, C. Nusbaum, E. S. Lander, and D. B. Jaffe. 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proceedings of the National Academy of Sciences 108:1513–1518.
Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, and A. Regev. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology 29:644–652.
Gu, Z., D. Nicolae, H. H.-S. Lu, and W.-H. Li. 2002. Rapid divergence in expression between duplicate genes inferred from microarray data. Trends in Genetics 18:609–613.
Gwynn, B., S. L. Ciciotte, S. J. Hunter, L. L. Washburn, R. S. Smith, S. G. Andersen, R. T. Swank, E. C. Dell'Angelica, J. S. Bonifacino, E. M. Eicher, and L. L. Peters. 2000. Defects in the cappuccino (cno) gene on mouse chromosome 5 and human 4p cause Hermansky-Pudlak syndrome by an AP-3-independent mechanism. Blood 96:4227–4235.
Haines, H., C. Ciskowski, and V. Harms. 1973. Acclimation to Chronic Water Restriction in the Wild House Mouse Mus musculus. Physiological Zoology 46:110–128.
Haines, H., and K. Schmidt-Nielsen. 1967. Water Deprivation in Wild House Mice. Physiological Zoology 40:424–431.
Hao, Y., Y. Xiong, Y. Cheng, G. Song, C. Jia, Y. Qu, and F. Lei. 2019. Comparative transcriptomics of 3 high-altitude passerine birds and their low-altitude relatives. Proceedings of the National Academy of Sciences 116:11851–11856.
Herrnberger, L., R. Seitz, S. Kuespert, M. R. Bösl, R. Fuchshofer, and E. R. Tamm. 2012. Lack of endothelial diaphragms in fenestrae and caveolae of mutant Plvap-deficient mice. Histochemistry and Cell Biology 138:709–724.
Hoekstra, H. E., K. E. Drumm, and M. W. Nachman. 2004. Ecological genetics of adaptive color polymorphism in pocket mice: geographic variation in selected and neutral genes. Evolution 58:1329–1341.
Hoekstra, H. E., J. G. Krenz, and M. W. Nachman. 2005. Local adaptation in the rock pocket mouse (Chaetodipus intermedius): natural selection and phylogenetic history of populations. Heredity 94:217–228.
Hoekstra, H. E., and M. W. Nachman. 2003. Different genes underlie adaptive melanism in different populations of rock pocket mice. Molecular Ecology 12:1185–1194.
Hoover, K. D., W. G. Whitford, and P. Flavill. 1977. Factors Influencing the Distributions of Two Species of Perognathus. Ecology 58:877–884.

Huang, Y., and A. F. Agrawal. 2016. Experimental Evolution of Gene Expression and Plasticity in Alternative Selective Regimes. PLOS Genetics 12:e1006336.
Hurst, L. D., and N. G. Smith. 1999. Do essential genes evolve slowly? Curr. Biol. 9:747–750.
Jackson, I. J. 1994. Molecular and Developmental Genetics of Mouse Coat Color. Annual Review of Genetics 28:189–217.
Jeong, C., D. B. Witonsky, B. Basnyat, M. Neupane, C. M. Beall, G. Childs, S. R. Craig, J. Novembre, and A. D. Rienzo. 2018. Detecting past and ongoing natural selection among ethnically Tibetan women at high altitude in Nepal. PLOS Genetics 14:e1007650.
Jones, M. R., L. S. Mills, J. D. Jensen, and J. M. Good. 2020. Convergent evolution of seasonal camouflage in response to reduced snow cover across the snowshoe hare range. Evolution 2020.
Karageorgi, M., S. C. Groen, F. Sumbul, J. N. Pelaez, K. I. Verster, J. M. Aguilar, A. P. Hastings, S. L. Bernstein, T. Matsunaga, M. Astourian, G. Guerra, F. Rico, S. Dobler, A. A. Agrawal, and N. K. Whiteman. 2019. Genome editing retraces the evolution of toxin resistance in the monarch butterfly. Nature 574:409–412.
Kassirer, J. P. 1971. Clinical Evaluation of Kidney Function: Glomerular Function. New England Journal of Medicine 285:385–389.
Khalil, F., and J. Tawfic. 1963. Some observations on the kidney of the desert J. jaculus and G. gerbillus and their possible bearing on the water economy of these animals. Journal of Experimental Zoology 154:259–271.
Koford, C. B. 1968. Peruvian Desert Mice: Water Independence, Competition, and Breeding Cycle near the Equator. Science 160:552–553.
Köhler, S., L. Carmody, N. Vasilevsky, J. O. B. Jacobsen, D. Danis, J.-P. Gourdine, M. Gargano, N. L. Harris, N. Matentzoglu, J. A. McMurry, D. Osumi-Sutherland, V. Cipriani, J. P. Balhoff, T. Conlin, H. Blau, G. Baynam, R. Palmer, D. Gratian, H. Dawkins, M. Segal, A. C. Jansen, A. Muaz, W. H. Chang, J. Bergerson, S. J. F. Laulederkind, Z. Yüksel, S. Beltran, A. F. Freeman, P. I. Sergouniotis, D. Durkin, A. L. Storm, M. Hanauer, M. Brudno, S. M. Bello, M. Sincan, K. Rageth, M. T. Wheeler, R. Oegema, H. Lourghi, M. G. Della Rocca, R. Thompson, F. Castellanos, J. Priest, C. Cunningham-Rundles, A. Hegde, R. C. Lovering, C. Hajek, A. Olry, L. Notarangelo, M. Similuk, X. A. Zhang, D. Gómez-Andrés, H. Lochmüller, H. Dollfus, S. Rosenzweig, S. Marwaha, A. Rath, K. Sullivan, C. Smith, J. D. Milner, D. Leroux, C. F. Boerkoel, A. Klion, M. C. Carter, T. Groza, D. Smedley, M. A. Haendel, C. Mungall, and P. N. Robinson. 2019. Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. Nucleic Acids Research 47:D1018–D1027.
Kordonowy, L., K. D. Lombardo, H. L. Green, M. D. Dawson, E. A. Bolton, S. LaCourse, and M. D. MacManes. 2017. Physiological and biochemical changes associated with acute experimental dehydration in the desert adapted mouse, Peromyscus eremicus. Physiological Reports 5:6.
Kowalczyk, A., W. K. Meyer, R. Partha, W. Mao, N. L. Clark, and M. Chikina. 2019. RERconverge: an R package for associating evolutionary rates with convergent traits. Bioinformatics 35:4815–4817.
Langfelder, P., and S. Horvath. 2008. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9:559.

Langmead, B., and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9:357–359.
Larsen, H. S., A.-K. Ruus, O. Schreurs, and H. K. Galtung. 2010. Aquaporin 11 in the developing mouse submandibular gland. European Journal of Oral Sciences 118:9–13.
Lemos, B., B. R. Bettencourt, C. D. Meiklejohn, and D. L. Hartl. 2005. Evolution of Proteins and Gene Expression Levels are Coupled in Drosophila and are Independently Associated with mRNA Abundance, Protein Length, and Number of Protein-Protein Interactions. Molecular Biology and Evolution 22:1345–1354.
Levis, N. A., and D. W. Pfennig. 2016. Evaluating 'Plasticity-First' Evolution in Nature: Key Criteria and Empirical Approaches. Trends in Ecology & Evolution 31:563–574.
Li, H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094–3100.
Li, W., and A. Godzik. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22:1658–1659.
Love, M. I., W. Huber, and S. Anders. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15:550.
Luo, R., B. Liu, Y. Xie, Z. Li, W. Huang, J. Yuan, G. He, Y. Chen, Q. Pan, Y. Liu, J. Tang, G. Wu, H. Zhang, Y. Shi, Y. Liu, C. Yu, B. Wang, Y. Lu, C. Han, D. W. Cheung, S.-M. Yiu, S. Peng, Z. Xiaoqian, G. Liu, X. Liao, Y. Li, H. Yang, J. Wang, T.-W. Lam, and J. Wang. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience 1:18.
Lynch, C. B. 1992. Clinal Variation in Cold Adaptation in Mus domesticus: Verification of Predictions from Laboratory Populations. The American Naturalist 139:1219–1236.
Ma, T., B. Yang, A. Gillespie, E. J. Carlson, C. J. Epstein, and A. S. Verkman. 1998. Severely Impaired Urinary Concentrating Ability in Transgenic Mice Lacking Aquaporin-1 Water Channels. Journal of Biological Chemistry 273:4296–4299.
Mack, K. L., M. A. Ballinger, M. Phifer-Rixey, and M. W. Nachman. 2018. Gene regulation underlies environmental adaptation in house mice. Genome Research 28:1636–1645.
MacManes, M. D. 2017. Severe acute dehydration in a desert rodent elicits a transcriptional response that effectively prevents kidney injury. American Journal of Physiology-Renal Physiology 313:F262–F272.
MacMillen, R. E., and A. K. Lee. 1967. Australian desert mice: independence of exogenous water. Science 158:383–385.
Macmillen, R. E., and A. K. Lee. 1969. Water metabolism of Australian hopping mice. Comparative Biochemistry and Physiology 28:493–514.
Makova, K. D., and W.-H. Li. 2003. Divergence in the Spatial Pattern of Gene Expression Between Human Duplicate Genes. Genome Research 13:1638–1645.
Mao, Y., P. Francis-West, and K. D. Irvine. 2015. Fat4/Dchs1 signaling between stromal and cap mesenchyme cells influences nephrogenesis and ureteric bud branching. Development 142:2574–2585.
Marçais, G., and C. Kingsford. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 27:764–770.
Marcovitz, A., Y. Turakhia, H. I. Chen, M. Gloudemans, B. A. Braun, H. Wang, and G. Bejerano. 2019. A functional enrichment test for molecular convergent evolution finds a clear

protein-coding signal in echolocating bats and whales. Proceedings of the National Academy of Sciences 116:21094–21103.
Marra, N. J., S. H. Eo, M. C. Hale, P. M. Waser, and J. A. DeWoody. 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 & Proteomics 7:328–339.
Marra, N. J., A. Romero, and J. A. DeWoody. 2014. Natural selection and the genetic basis of osmoregulation in heteromyid rodents as revealed by RNA-seq. Molecular Ecology 23:2699–2711.
McKenzie, A. T., I. Katsyv, W.-M. Song, M. Wang, and B. Zhang. 2016. DGCA: A comprehensive R package for Differential Gene Correlation Analysis. BMC Systems Biology 10:106.
Merriam, C. H. 1889. Preliminary revision of the North American Pocket Mice. North American Fauna 1:1–36.
Morishita, Y., T. Matsuzaki, M. Hara-chikuma, A. Andoo, M. Shimono, A. Matsuki, K. Kobayashi, M. Ikeda, T. Yamamoto, A. Verkman, E. Kusano, S. Ookawara, K. Takata, S. Sasaki, and K. Ishibashi. 2005. Disruption of Aquaporin-11 Produces Polycystic Kidneys following Vacuolization of the Proximal Tubule. Molecular and Cellular Biology 25:7770–7779.
Mundy, N. I. 2005. A window on the genetics of evolution: MC1R and plumage colouration in birds. Proceedings of the Royal Society B: Biological Sciences 272:1633–1640.
Nachman, M. W., H. E. Hoekstra, and S. L. D'Agostino. 2003. The genetic basis of adaptive melanism in pocket mice. Proceedings of the National Academy of Sciences 100:5268–5273.
Nielsen, R., C. Bustamante, A. G. Clark, S. Glanowski, T. B. Sackton, M. J. Hubisz, A. Fledel-Alon, D. M. Tanenbaum, D. Civello, T. J. White, J. J Sninsky, M. D. Adams, and M. Cargill. 2005. A scan for positively selected genes in the genomes of humans and chimpanzees. PLoS Biol. 3:e170.
Nuzhdin, S. V., M. L. Wayne, K. L. Harmon, and L. M. McIntyre. 2004. Common Pattern of Evolution of Gene Expression Level and Protein Sequence in Drosophila. Molecular Biology and Evolution 21:1308–1317.
O'Connell, J., O. Schulz-Trieglaff, E. Carlson, M. M. Hims, N. A. Gormley, and A. J. Cox. 2015. NxTrim: optimized trimming of Illumina mate pair reads. Bioinformatics 31:2035–2037.
Pannabecker, T. L. 2015. Aquaporins in Desert Rodent Physiology. The Biological Bulletin 229:120–128.
Parker, D. J., J. Bast, K. Jalvingh, Z. Dumas, M. Robinson-Rechavi, and T. Schwander. 2019. Repeated Evolution of Asexuality Involves Convergent Gene Expression Changes. Molecular Biology and Evolution 36:350–364.
Patro, R., G. Duggal, M. I. Love, R. A. Irizarry, and C. Kingsford. 2017. Salmon provides fast and bias-aware quantification of transcript expression. Nature Methods 14:417–419.
Phifer-Rixey, M., K. Bi, K. G. Ferris, M. J. Sheehan, D. Lin, K. L. Mack, S. M. Keeble, T. A. Suzuki, J. M. Good, and M. W. Nachman. 2018. The genomic basis of environmental adaptation in house mice. PLOS Genetics 14:e1007672.

Phifer-Rixey, M., and M. W. Nachman. 2015. Insights into mammalian biology from the wild house mouse Mus musculus. eLife 4:e05959.
Price, T. D., A. Qvarnström, and D. E. Irwin. 2003. The role of phenotypic plasticity in driving genetic evolution. Proceedings of the Royal Society of London. Series B: Biological Sciences 270:1433–1440.
Ralph, P. L., and G. Coop. 2015. Convergent Evolution During Local Adaptation to Patchy Landscapes. PLOS Genetics 11:e1005630.
Rebar, C. E. 1995. Ability of Dipodomys merriami and Chaetodipus intermedius to Locate Resource Distributions. Journal of Mammalogy 76:437–447.
Reichman, O. J. 1975. Relation of Desert Rodent Diets to Available Resources. Journal of Mammalogy 56:731–751.
Reichman, O. J., and K. M. Van De Graaff. 1973. Seasonal Activity and Reproductive Patterns of Five Species of Sonoran Desert Rodents. The American Midland Naturalist 90:118–126.
Rey, C., L. Guéguen, M. Sémon, and B. Boussau. 2018. Accurate Detection of Convergent Amino-Acid Evolution with PCOC. Molecular Biology and Evolution 35:2296–2306.
Riddle, M. R., A. C. Aspiras, K. Gaudenz, R. Peuß, J. Y. Sung, B. Martineau, M. Peavey, A. C. Box, J. A. Tabin, S. McGaugh, R. Borowsky, C. J. Tabin, and N. Rohner. 2018. Insulin resistance in cavefish as an adaptation to a nutrient-limited environment. Nature 555:647–651.
Ritland, K., C. Newton, and H. D. Marshall. 2001. Inheritance and population structure of the white-phased "Kermode" black bear. Current Biology 11:1468–1472.
Robinson, B. W., and R. Dukas. 1999. The Influence of Phenotypic Modifications on Evolution: The Baldwin Effect and Modern Perspectives. Oikos 85:582–589.
Rogg, M., M. Yasuda-Yamahara, A. Abed, P. Dinse, M. Helmstädter, A. C. Conzelmann, J. Frimmel, D. Sellung, M. L. Biniossek, O. Kretz, F. Grahammer, O. Schilling, T. B. Huber, and C. Schell. 2017. The WD40-domain containing protein CORO2B is specifically enriched in glomerular podocytes and regulates the ventral actin cytoskeleton. Scientific Reports 7:15910.
Saburi, S., I. Hester, E. Fischer, M. Pontoglio, V. Eremina, M. Gessler, S. E. Quaggin, R. Harrison, R. Mount, and H. McNeill. 2008. Loss of Fat4 disrupts PCP signaling and oriented cell division and leads to cystic kidney disease. Nature Genetics 40:1010–1015.
Sackton, T. B., P. Grayson, A. Cloutier, Z. Hu, J. S. Liu, N. E. Wheeler, P. P. Gardner, J. A. Clarke, A. J. Baker, M. Clamp, and S. V. Edwards. 2019. Convergent regulatory evolution and loss of flight in paleognathous birds. Science 364:74–78.
Sage, R. D. 1981. Wild Mice. Pp. 39–90 *in* H. L. Foster, D. Small J., and J. G. Fox, eds. The Mouse in Biomedical Research: History, Genetics, and Wild Mice. Academic Press: New York.
Schlenke, T. A., and D. J. Begun. 2003. Natural selection drives Drosophila immune system evolution. Genetics 164:1471–1480.
Schmidt-Nielsen, K. 1964. Desert Animals: Physiological Problems of Heat and Water. Dover Publications, New York.
Schmidt-Nielsen, K., and B. Schmidt-Nielsen. 1952. Water Metabolism of Desert Mammals. Physiological Reviews 32:135–166.

Schwab, K., L. T. Patterson, B. J. Aronow, R. Luckas, H.-C. Liang, and S. S. Potter. 2003. A catalogue of gene expression in the developing kidney. Genetic Disorders - Development 64:1588–1604.
Schwartz, W. B. 1955. Potassium and the Kidney. New England Journal of Medicine 253:601–608.
Schwarz, A., K. Möller-Hackbarth, L. Ebarasi, D. U. Jess, S. Zambrano, H. Blom, A. Wernerson, M. Lal, and J. Patrakka. 2019. Coro2b, a podocyte protein downregulated in human diabetic nephropathy, is involved in the development of protamine sulphate-induced foot process effacement. Scientific Reports 9:1–11.
Sela, I., H. Ashkenazy, K. Katoh, and T. Pupko. 2015. GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. Nucleic Acids Research 43:W7–W14.
Senay, L. C., and M. L. Christensen. 1965. Changes in blood plasma during progressive dehydration. Journal of Applied Physiology 20:1136–1140.
Seppey, M., M. Manni, and E. M. Zdobnov. 2019. BUSCO: Assessing Genome Assembly and Annotation Completeness. Pp. 227–245 *in* M. Kollmar, ed. Gene Prediction: Methods and Protocols. Springer, New York, NY.
Sikes, R. S., and W. L. Gannon. 2011. Guidelines of the American Society of Mammalogists for the use of wild mammals in research. Journal of Mammalogy 92:235–253.
Simonson, T. S., Y. Yang, C. D. Huff, H. Yun, G. Qin, D. J. Witherspoon, Z. Bai, F. R. Lorenzo, J. Xing, L. B. Jorde, J. T. Prchal, and R. Ge. 2010. Genetic Evidence for High-Altitude Adaptation in Tibet. Science 329:72–75.
Simpson, G. G. 1953. The Baldwin Effect. Evolution 7:110–117.
Smit, A., R. Hubley, and P. Green. 2013. RepeatMasker Open-4.0.
Smith-Unna, R., C. Boursnell, R. Patro, J. M. Hibberd, and S. Kelly. 2016. TransRate: reference-free quality assessment of de novo transcriptome assemblies. Genome Res. 26:1134–1144.
Sohara, E., T. Rai, J. Miyazaki, A. S. Verkman, S. Sasaki, and S. Uchida. 2005. Defective water and glycerol transport in the proximal tubules of AQP7 knockout mice. American Journal of Physiology-Renal Physiology 289:F1195–F1200.
Song, L., and L. Florea. 2015. Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads. GigaScience 4:48.
Steiner, C. C., H. Römpler, L. M. Boettger, T. Schöneberg, and H. E. Hoekstra. 2009. The Genetic Basis of Phenotypic Convergence in Beach Mice: Similar Pigment Patterns but Different Genes. Molecular Biology and Evolution 26:35–45.
Stewart, C.-B., and A. C. Wilson. 1987. Sequence Convergence and Functional Adaptation of Stomach Lysozymes from Foregut Fermenters. Cold Spring Harb Symp Quant Biol 52:891–899.
Tchekneva, E. E., Z. Khuchua, L. S. Davis, V. Kadkina, S. R. Dunn, S. Bachman, K. Ishibashi, E. M. Rinchik, R. C. Harris, M. M. Dikov, and M. D. Breyer. 2008. Single Amino Acid Substitution in Aquaporin 11 Causes Renal Failure. Journal of the American Society of Nephrology 19:1955–1964.
The UniProt Consortium. 2007. The Universal Protein Resource (UniProt). Nucleic Acids Research 36:D190–D195.

Tigano, A., J. P. Colella, and M. D. MacManes. 2020. Comparative and population genomics approaches reveal the basis of adaptation to deserts in a small rodent. Molecular Ecology 29:1300–1314.
Tsuda, H., Y. Isaka, S. Takahara, and M. Horio. 2009. Discrepancy between serum levels of low molecular weight proteins in acute kidney injury model rats with bilateral ureteral obstruction and bilateral nephrectomy. Clinical and Experimental Nephrology 13:567–570.
Waddington, C. H. 1942. Canalization of Development and the Inheritance of Acquired Characters. Nature 150:563–565.
Waddington, C. H. 1953. Genetic Assimilation of an Acquired Character. Evolution 7:118–126.
Waddington, C. H. 1952. Selection of the Genetic Basis for an Acquired Character. Nature 170:71–71.
Weckerly, F. W., and T. L. Best. 1985. Morphologic Variation among Rock Pocket Mice (Chaetodipus intermedius) from New Mexico Lava Fields. The Southwestern Naturalist 30:491–501.
Weinreich, D. M., N. F. Delaney, M. A. DePristo, and D. L. Hartl. 2006. Darwinian Evolution Can Follow Only Very Few Mutational Paths to Fitter Proteins. Science 312:111–114.
Wen, M., S. Segerer, M. Dantas, P. A. Brown, K. L. Hudkins, T. Goodpaster, E. Kirk, R. C. LeBoeuf, and C. E. Alpers. 2002. Renal Injury in Apolipoprotein E–Deficient Mice. Laboratory Investigation 82:999–1006.
Weng, M.-P., and B.-Y. Liao. 2017. modPhEA: model organism Phenotype Enrichment Analysis of eukaryotic gene sets. Bioinformatics 33:3505–3507.
Wray, G. A. 2007. The evolutionary significance of cis -regulatory mutations. Nature Reviews Genetics 8:206–216.
Wu, H., X. Guang, M. B. Al-Fageeh, J. Cao, S. Pan, H. Zhou, L. Zhang, M. H. Abutarboush, Y. Xing, Z. Xie, A. S. Alshanqeeti, Y. Zhang, Q. Yao, B. M. Al-Shomrani, D. Zhang, J. Li, M. M. Manee, Z. Yang, L. Yang, Y. Liu, J. Zhang, M. A. Altammami, S. Wang, L. Yu, W. Zhang, S. Liu, L. Ba, C. Liu, X. Yang, F. Meng, S. Wang, L. Li, E. Li, X. Li, K. Wu, S. Zhang, J. Wang, Y. Yin, H. Yang, A. M. Al-Swailem, and J. Wang. 2014. Camelid genomes reveal evolution and adaptation to desert environments. Nature Communications 5:1–10.
Yang, Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Molecular Biology and Evolution 15:568–573.
Yang, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution 24:1586–1591.
Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Computer applications in the biosciences : CABIOS 13:555–556.
Yi, X., Y. Liang, E. Huerta-Sanchez, X. Jin, Z. X. P. Cuo, J. E. Pool, X. Xu, H. Jiang, N. Vinckenbosch, T. S. Korneliussen, H. Zheng, T. Liu, W. He, K. Li, R. Luo, X. Nie, H. Wu, M. Zhao, H. Cao, J. Zou, Y. Shan, S. Li, Q. Yang, Asan, P. Ni, G. Tian, J. Xu, X. Liu, T. Jiang, R. Wu, G. Zhou, M. Tang, J. Qin, T. Wang, S. Feng, G. Li, Huasang, J. Luosang, W. Wang, F. Chen, Y. Wang, X. Zheng, Z. Li, Z. Bianba, G. Yang, X. Wang, S. Tang, G. Gao, Y. Chen, Z. Luo, L. Gusang, Z. Cao, Q. Zhang, W. Ouyang, X. Ren, H. Liang, H. Zheng, Y. Huang, J. Li, L. Bolund, K. Kristiansen, Y. Li, Y. Zhang, X. Zhang, R. Li, S. Li, H. Yang, R. Nielsen, J.

Wang, and J. Wang. 2010. Sequencing of 50 Human Exomes Reveals Adaptation to High Altitude. Science 329:75–78.
Zhang, H., M. Bagherie-Lachidan, C. Badouel, L. Enderle, P. Peidis, R. Bremner, S. Kuure, S. Jain, and H. McNeill. 2019. FAT4 Fine-Tunes Kidney Development by Regulating RET Signaling. Developmental Cell 48:780-792.e4.
Zhang, J., and S. Kumar. 1997. Detection of convergent and parallel evolution at the amino acid sequence level. Molecular Biology and Evolution 14:527–536.
Zhen, Y., M. L. Aardema, E. M. Medina, M. Schumer, and P. Andolfatto. 2012. Parallel Molecular Evolution in an Herbivore Community. Science 337:1634–1637.
Zimin, A. V., G. Marçais, D. Puiu, M. Roberts, S. L. Salzberg, and J. A. Yorke. 2013. The MaSuRCA genome assembler. Bioinformatics 29:2669–2677.