bioRxiv preprint doi: https://doi.org/10.1101/2021.09.10.459863; this version posted September 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

1
2
3
4
5
6
7
8
9 Convergent patterns of gene expression and protein evolution associated with
10 adaptation to desert environments in rodents
11
12
13
14
15
16
17
18
19
20 Noëlle K. J. Bittner[1], Katya L. Mack[1,2], Michael W. Nachman[1]
21
22
23
24
25
26
27
28
29
30
31 [1]Deparment of Integrative Biology and Museum of Vertebrate Zoology, University of
32 California Berkeley, Berkeley, CA 94720, USA
33 [2]Present address: Department of Biology, Stanford University, Stanford, CA 94305,
34 USA
35
36 Corresponding Author:
37 Michael Nachman
38 Department of Integrative Biology and Museum of Vertebrate Zoology
39 3101 Valley Life Sciences Building
40 University of California Berkeley, Berkeley, CA 94720
41 Phone: 510-642-1792
42 Email: mnachman@berkeley.edu
43

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.10.459863; this version posted September 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

44 **Abstract**

45 Desert specialization has arisen multiple times across rodents and is often
46 associated with a suite of convergent phenotypes, including modification of the
47 kidneys to mitigate water loss. However, the extent to which phenotypic
48 convergence in desert rodents is mirrored at the molecular level is unknown. Here,
49 we sequenced kidney mRNA and assembled transcriptomes for three pairs of rodent
50 species to search for convergence in gene expression and amino acid sequence
51 associated with adaptation to deserts. We conducted phylogenetically-independent
52 comparisons between a desert specialist and a non-desert relative in three families
53 representing ~70 million years of evolution. Overall, patterns of gene expression
54 faithfully recapitulated the phylogeny of these six taxa. However, we found that
55 8.6% of all genes showed convergent patterns of expression evolution between
56 desert and non-desert taxa, a proportion that is much higher than expected by
57 chance. In addition to these convergent changes, we observed many species-pair
58 specific changes in gene expression indicating that different instances of adaptation
59 to deserts include a combination of unique and shared changes. Patterns of protein
60 evolution revealed a small number of genes showing evidence of positive selection,
61 the majority of which did not show convergent changes in gene expression. Overall,
62 our results suggest convergent changes in gene regulation play a primary role in the
63 complex trait of desert adaptation in rodents.

64

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.10.459863; this version posted September 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

65

66 **Introduction**

67 The repeatability of adaptive evolution at the molecular level remains an open
68 question. In situations where the mutational target is small and constraints exist
69 due to epistasis or pleiotropy, the molecular paths available to adaptation may be
70 highly limited (Weinreich et al. 2006; Karageorgi et al. 2019). Indeed, there are a
71 number of excellent examples of convergent molecular evolution underlying simple
72 traits (e.g. Stewart and Wilson 1987; Mundy 2005; Zhen et al. 2012). For highly
73 polygenic traits, however, convergence may be less expected simply because the
74 mutational target is large and multiple paths may be available on which selection
75 can act. Nonetheless, several studies have found evidence for convergence at the
76 molecular level even for complex traits (e.g. Marcovitz et al. 2019; Sackton et al.
77 2019).

78

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

89

90 Adaptation to desert environments in rodents provides an opportunity to study
91 repeated evolution in both gene expression and protein sequence in a complex trait.
92 Desert ecosystems present the challenge of extreme aridity and low or seasonally
93 absent water, yet multiple lineages of rodents have independently evolved the
94 ability to survive in these unusually harsh environments (reviewed in Degen 1997).
95 Rodents have solved these challenges in myriad ways, including dietary

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.10.459863; this version posted September 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

96 specialization on plants that are relatively high in water content or modifications to
97 reduce evaporative water loss (Schmidt-Nielsen and Schmidt-Nielsen 1952;
98 Schmidt-Nielsen 1964; Degen 1997). However, a common feature of most desert
99 rodents is a modified kidney capable of producing highly concentrated urine
100 (MacMillen and Lee 1967; Beuchat 1990; Al-kahtani et al. 2004; Donald and
101 Pannabecker 2015). Final excreted urine concentration depends on the
102 development and maintenance of a corticomedullary osmotic gradient within the
103 kidney. Studies have shown that many aspects of kidney morphology and
104 physiology have been modified in different lineages to produce hyper-concentrated
105 urine (Bankir and de Rouffignac 1985; Donald and Pannabecker 2015).
106
107 The genetic basis of desert adaptation has been studied independently in a handful
108 of species and individual genes and pathways which may underlie this adaptive
109 phenotype have been identified (Rocha et al. 2021). Here, we leverage three
110 phylogenetically independent lineages of rodents that have all converged on a
111 common phenotype, ultra-high urine concentration associated with desert living, to
112 identify shared molecular changes associated with habitat type. We compared
113 kidney gene expression and protein sequence divergence between a desert and a
114 non-desert species in each of three pairs of phylogenetically independent
115 comparisons representing transitions to desert living in three different rodent
116 families (Heteromyidae, Dipodidae, and Muridae). Desert species were chosen based
117 on their high urine concentration, a proxy for increased osmoregulatory capacity
118 (Figure 1). Within Muridae, we compared the Australian Spinifex Hopping Mouse,
119 *Notomys alexis*, the mammal with the highest known urine concentration and well
120 studied for its modifications to desert life (MacMillen and Lee 1967; Macmillen and
121 Lee 1969; Baudinette 1972; Donald et al. 2012), to the house mouse (*Mus musculus*),
122 a widespread generalist. Within Dipodidae, we compared the desert-dwelling Lesser
123 Egyptian Jerboa, *Jaculus jaculus,* previously studied for its kidney modifications
124 associated with granivorous desert living (Schmidt-Nielsen and Schmidt-Nielsen
125 1952; Khalil and Tawfic 1963), to the Western Jumping Mouse, *Zapus princeps*, a
126 North American species found in riparian environments. Within Heteromyidae, we

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.10.459863; this version posted September 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

127 compared the Rock Pocket Mouse, *Chaetodipus intermedius* (Bradley et al. 1975;
128 Altschuler et al. 1979), native to the North American Sonoran desert , to the
129 Desmarest's spiny pocket mouse, *Heteromys desmarestianus,* a neotropical species
130 found in mesic areas that cannot survive without free water (Fleming 1977).
131
132 We sequenced kidney mRNA from these pairs of taxa and assembled and annotated
133 *de novo* transcriptomes for 3 desert-mesic species pairs spanning ~70 million years
134 of evolution. Assembled transcriptomes were used to analyze rates of evolution in
135 single copy orthologs to identify genes putatively under selection across desert
136 lineages. We also performed mRNA-sequencing on multiple individuals within each
137 species to study gene expression divergence between desert and non-desert species.
138 Global patterns of gene expression recapitulated the phylogeny of these six species.
139 However, we also discovered a significantly greater number of convergent changes
140 in gene expression than expected by chance between desert and non-desert species.
141 In contrast, convergent changes in amino acid sequence were identified at a smaller
142 proportion of genes. Overall, we identified genes with shared patterns of molecular
143 evolution associated with habitat type.
144

145 Materials and Methods

146 *Sample collection*

147 Five adult male mice for each species, with the exception of *H. desmarestianus*, for
148 which only four samples could be obtained, were included in this study. *C.*
149 *intermedius*, *Z. princeps*, and *M. musculus* were caught by N. Bittner using Sherman
150 live traps set over-night following the guidelines of the American Society of
151 Mammalogists (Sikes and Gannon 2011) and an ACUC protocol approved by UC
152 Berkeley (AUP-2016-03-8536). Animals were given apple after capture to avoid
153 dehydration for the short period of time they were in traps. Mice were euthanized
154 by cervical dislocation, and kidney and liver were removed and preserved in
155 RNAlater. *C. intermedius* were trapped near Tucson, AZ, USA, *Z. princeps* were
156 trapped at Sagehen Creek Field Station near Truckee, CA, USA, and *M. musculus* were

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.10.459863; this version posted September 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

157 trapped near Berkeley, CA, USA. *H. desmarestianus* were collected in Chiapas, Mexico
158 by Beatriz Jimenez, and *N. alexis* were collected by Kevin Rowe in Northern
159 Territory, Australia. Mice collected by N. Bittner were prepared as museum
160 specimens (skins and skulls) and deposited in the collections of the UC Berkeley
161 Museum of Vertebrate Zoology (MVZ). Animals collected by K. Rowe were prepared
162 as museum specimens and deposited at Museums Victoria. The collecting localities,
163 collector's numbers, and museum catalog numbers for each specimen are provided
164 for all wild-caught animals in Table S1. Samples from *Jaculus jaculus* were provided
165 by Kim Cooper at UC San Diego from an outbred lab colony. Despite the fact that the
166 *Jaculus* were from a laboratory colony while all other animals were wild caught,
167 patterns of gene expression among all individuals reflected the phylogeny of these
168 taxa, suggesting that the laboratory environment for *Jaculus* did not obscure overall
169 expression patterns (see Results).
170
171 *mRNA library preparation and sequencing*
172 To target loci underlying adaptation to xeric environments, we focused on genes
173 expressed in the kidney. RNA was extracted from kidney preserved in RNAlater
174 using the MoBio Laboratories Powerlyzer Ultraclean Tissue & Cells RNA Isolation
175 Kit. Remaining DNA was removed with DNAse-1 followed by a Zymo RNA Clean and
176 Concentrator column clean-up. Due to the poor quality of some samples (RIN scores
177 below 5), a ribosomal RNA depletion step was performed with a KAPA Riboerase Kit
178 before libraries were prepared with the KAPA HyperPrep Kit. Libraries were pooled
179 and sequenced across two lanes of 150 bp PE NovaSeq (one lane of S1 and one of
180 SP) at the Vincent J. Coates Genomics Sequencing Center at UC Berkeley. One library
181 from each species (except *Mus musculus*; see below) was sequenced at greater depth
182 for transcriptome assembly; these were sequenced to a target of 100M read pairs
183 while the remaining 24 libraries, intended for expression analysis, were sequenced
184 to a target of 20M read pairs (see File S1).
185
186 *Transcriptome assembly*

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.10.459863; this version posted September 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

187 For each of the five 100M-read-pair libraries, reads were examined for quality
188 metrics with FastQC (https://www.bioinformatics.babraham.ac.uk/
189 projects/fastqc/) and then corrected by removing erroneous k-mers using
190 rCorrector (Song and Florea 2015). Adapters and poor quality sequence were
191 trimmed using Trim Galore! (https://www.bioinformatics.babraham.ac.uk/
192 projects/trim_galore/). Since FastQC revealed a large quantity of duplicates within
193 the sequenced libraries, which is likely in part due to rRNA contamination, we chose
194 to remove all reads that mapped to known rodent rRNA from NCBI using bowtie2
195 (Langmead and Salzberg 2012). We ran Trinity v2.1.1 (Grabherr et al. 2011) to
196 generate a transcriptome assembly for each species. Because transcriptome-depth
197 (i.e. 100M read pairs) sequencing was not done for *Mus musculus*, reads from all five
198 individuals (approximately equal to the sequencing depth for transcriptome
199 individuals) were combined to assemble the transcriptome of a local individual as
200 above (Table S2). To remove redundant transcripts from the Trinity assembly,
201 transcripts with equal to or greater than 95% sequence identity were clustered with
202 cd-hit-est (settings: -c 0.95 -n 8) (Li and Godzik 2006) to create representative
203 transcripts before use in downstream analysis (Table S2). This was done to collapse
204 transcript isoforms as well as to remove transcripts created by assembly errors
205 (chimeras, duplicates, misassembled transcripts and the like). Transrate (Smith-
206 Unna et al. 2016) was used to calculate assembly statistics. To assess assembly
207 completeness, we used Benchmarking Universal Single-Copy Orthologs (BUSCO)
208 (Seppey et al. 2019) to look for the 6,192 orthologs found in the Euarchontoglires
209 odb9 database and thus expected to exist in the taxa studied here.
210
211 *Transcriptome annotation and ortholog detection*
212 To identify coding regions within our assembled transcripts for downstream
213 analyses, we utilized TransDecoder v. 5.5.0 (http://transdecoder.sourceforge.net).
214 We identified the longest open reading frame (ORF) and searched for matches to
215 both the Pfam protein domain database (Bateman et al. 2004) and mouse specific
216 SwissProt database (Bairoch and Apweiler 2000) to retain ORFs based on
217 homology. Since high quality gene annotations were available for *M. musculus*, we

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.10.459863; this version posted September 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

218 used the curated RefSeq protein database for this species. Orthologous gene groups
219 across all six taxa were identified using OrthoFinder v 2.3.3 (setting: -S diamond)
220 (Emms and Kelly 2015). To minimize the number of alternate isoforms used in the
221 analysis, we only used the longest ORF identified per gene.
222
223 *mRNA read mapping*
224 Raw reads from all libraries were examined for quality with FastQC. Adapters and
225 poor quality sequence were trimmed using Trimmomatic v0.36 (Bolger et al. 2014).
226 The five libraries that were generated for transcriptome assembly were subsampled
227 to the average read number of the libraries generated for expression (27,787,405
228 reads). Reads were mapped to transcriptomes generated for each species with
229 Salmon v 0.14.1 (Patro et al. 2017). To compare across genera, transcripts were
230 annotated using BLASTn to the Refseq cdna database for *Mus musculus*. Read counts
231 were summed across transcripts for each annotated gene.
232
233 *Quantification of gene expression and identification of convergent differential*
234 *expression*
235 DESeq2 (Love et al. 2014) was used to normalize for differences in library size and
236 to call differential expression between species within each family and across all
237 samples. As transcripts between species can differ in length, a length correction was
238 applied. Reads were subsequently transformed with a variance stabilizing
239 transformation for principal component analysis.
240
241 We used DESeq2 to identify convergent changes in gene expression between desert
242 and non-desert species across all three families using an approach similar to that
243 used by Parker et al. (2019). In particular, we fit a generalized linear model for gene
244 expression as a function of habitat (desert vs. non-desert), family (species-pair), and
245 their interaction. Genes were classified as convergently differentially expressed in
246 cases where there was a significant effect of habitat (desert vs. non-desert,
247 $FDR<0.01$) but no interaction effect of species-pair by habitat ($FDR > 0.05$). This
248 analysis was restricted to genes with greater than an average of 20 reads per sample

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.10.459863; this version posted September 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

249 for each species, resulting in a total of 8,174 genes. P-values were adjusted for
250 multiple testing using a Benjamini & Hochberg (Benjamini and Hochberg 1995)
251 correction. Differential expression within each species pair was identified using
252 pairwise contrasts. For pairwise contrasts, genes with a mean of fewer than 10
253 reads per sample were removed from the analysis.
254
255 Permutation tests were used to assess whether more genes showed convergent
256 shifts by habitat type than expected by chance, as described in Parker et al. (2019).
257 For each gene, read counts were randomly assigned to habitat within each species
258 pair. All biological replicates (i.e. all five individuals) in each species were assigned
259 to the same habitat. This process was used to create 10,000 permuted datasets. The
260 number of convergently differentially expressed genes in these datasets were
261 compared to that of the observed dataset
262
263 *Estimating rates of molecular evolution and identification of genes under positive*
264 *selection*
265 Using the single copy ortholog groups generated by OrthoFinder for all six species,
266 we aligned these using MAFFT through Guidance2 (Sela et al. 2015), which provided
267 alignment quality scores for all 1,855 genes and removed those for which the
268 alignment quality score was poor (mean column score <0.8). Alignments for which
269 quality scores were poor were removed from subsequent analyses, resulting in a set
270 of 1,474 genes with aligned protein coding alignments for subsequent analyses.
271
272 We used a maximum likelihood approach in a phylogenetic context by implementing
273 the codeml package in PAML (Yang 1997) to identify genes in desert lineages with
274 evidence of selection. We performed three analyses using the 1,474 single copy
275 orthologs with high quality alignments present in all species. First, we defined the
276 three desert species together as "foreground" lineages and compared these to the
277 three non-desert species as "background" lineages using a foreground-background
278 branch analysis implemented in PAML (Yang 1998; Yang 2007). This analysis
279 estimates ω or dN/dS (the rate of nonsynonymous substitutions per

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.10.459863; this version posted September 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

280 nonsynonymous site divided by the rate of synonymous substitutions per
281 synonymous site) and compares branches of interest (e.g., "foreground" branches)
282 to the other "background" branches. Elevated rates of dN/dS compared with a null
283 model are considered evidence for selection. This analysis was intended to identify
284 genes underlying desert adaptation common to all three species. Second, we
285 performed three separate foreground-background branch analyses, in which each
286 desert species by itself was compared to the other five species. This analysis was
287 intended to identify species-specific adaptations. Third, we performed a branch site
288 model which allows for ω to vary both across sites in a gene and across branches on
289 the tree.
290
291 *Enrichment analyses*
292 For gene sets of interest, GO category enrichment tests were performed with GOrilla
293 (Eden et al. 2009) to test a foreground gene set of interest against a background set of all
294 other genes included in the analysis. Phenotype enrichment tests were performed with
295 modPhea (Weng and Liao 2017) using the same framework.
296
297
298 **Results**
299
300 *Sequencing, assembly, and annotation*
301
302 We generated on average ~123 million reads per sample for the assembly of *de-*
303 *novo* kidney transcriptomes in each species. For *Mus musculus*, five smaller libraries
304 were concatenated for assembly. After read correction, quality filtering, and adapter
305 trimming, each library had an average of ~103 million reads which were used for
306 the assembly. Each assembly contained 965,227 transcripts on average. We reduced
307 the number of redundant transcripts in the assembly to improve accuracy of
308 downstream analyses by clustering similar transcripts together using CD-HIT-EST.
309 This decreased the number of transcripts by ~20% per sample to an average of
310 793,887 transcripts (Table S2). We used BUSCO to check assembly completeness to

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.10.459863; this version posted September 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

311 determine how many of the 6,192 orthologs found in the Euarchontoglires odb9
312 were present in our assembled transcriptomes. The six assemblies ranged in
313 completeness from 80 - 87% (Figure S1). This level of completeness reflects a single
314 tissue (kidney) taken at one developmental time point. After ORF prediction, we
315 annotated each transcript to known *M. musculus* proteins. We were able to assign
316 transcripts to 395,029 putative ortholog groups.
317
318 *Global gene expression reflects phylogenetic relationships and habitat type*
319 To identify patterns of differential gene expression, we sequenced kidney mRNA
320 from additional individuals in each of the six species for an average of ~27 million
321 reads per individual. We retrieved 13,305 genes in *C. intermedius*, 11,749 genes in *H.*
322 *desmarestianus*, 14,891 genes in *J. jaculus*, 14,380 genes in *Z. princeps*, 18,622 genes
323 in *N. alexis* and 19,913 genes in *M. musculus* for which we were able to quantify
324 expression levels. These genes were annotated using *M. musculus* transcripts so as
325 expected, the number of genes we were able to annotate in more divergent species
326 is more limited.
327
328 Gene expression profiles largely recapitulated the known phylogenetic relationships
329 of these six species (Figure 2A). Individuals within each species form well-defined
330 clusters (with the exception of a single *Heteromys* individual), and the different
331 genera within each family share expression profiles that are more similar to each
332 other than they are to genera in different families. Further, Muridae and Dipodidae
333 are more similar to each other in expression profiles than either is to Heteromyidae,
334 reflecting the known evolutionary relationships of these families. Thus, the overall
335 expression patterns reflect evolutionary history more than habitat type. These
336 patterns are also seen in a principal component analysis (PCA) based on expression
337 level co-variance (Figure 2B), where PC1 (accounting for 33% of the variance)
338 largely reflects phylogeny. Despite the overall phylogenetic pattern of gene
339 expression, consistent differences in expression were seen between desert and non-
340 desert species within each family. In particular, PC4 captures this variation,

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.10.459863; this version posted September 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

341 separating desert from non-desert taxa (explaining 11% of the variation) (Figure
342 2C).
343
344 *Convergent differential expression in desert rodent kidneys*
345 We quantified differential expression (DE) between desert and non-desert species
346 within each family. In pairwise contrasts between desert and non-desert species in
347 Heteromyidae, Dipodidae, and Muridae, we identified >4,000 genes in each
348 comparison with evidence of significant DE (Table S3, FDR<0.01). Individual
349 pairwise comparisons between desert and non-desert species found uniquely in
350 each of the three families (to the exclusion of the two others) were associated with
351 several GO categories, including cellular metabolic processes and nitrogen
352 metabolic processes (Table S4). We identified a total of 654 genes that showed
353 significant differential expression in all three species pairs (Figure S2), with 145 of
354 these genes showing shifts in the same direction in each comparison.
355
356 To identify convergent shifts in gene expression associated with desert-living, we
357 also modeled gene expression as a function of species pair (i.e., family), habitat, and
358 their interaction. Convergent changes were identified as those for which there was a
359 significant effect of habitat (FDR<0.01), but no interaction between species pair and
360 habitat (FDR>0.05) (see Methods; Parker et al. 2019). We identified 702 genes with
361 shared shifts in desert rodents relative to the mesic comparison (Figure 3A). This
362 set includes all of the 145 genes identified above in pairwise tests. Thus, 8.6%
363 (702/8,174) of genes showed convergent shifts in expression in desert rodents
364 compared to their non-desert relatives.
365
366 Shared shifts in gene expression can be a consequence of selection in response to
367 shared environmental pressures or stochastic processes. To ask if the observed
368 number of genes with convergent differential expression was more than expected
369 by chance, we performed a permutation test in which we took each gene and
370 randomly switched habitat assignment within species pairs, while always
371 maintaining the same label for all biological replicates within a species, to create

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.10.459863; this version posted September 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

372 10,000 permuted data sets (Figure 3B, see Methods). Permuted datasets never
373 identified more convergent genes than the observed set of convergent genes,
374 suggesting an enrichment of convergent differential expression associated with
375 habitat type.
376
377 Fold changes between individual desert-mesic pairs were often modest in one or
378 more contrasts between species pairs (Figure S3); only 208 genes with shared
379 expression shifts showed an average of greater >0.5 log2 fold change difference
380 between each desert-mesic species pair. The number of genes showing higher
381 expression in desert rodents compared to non-desert relatives (335 genes, shown in
382 blue in Figure 3A) was slightly fewer than the number of genes showing lower
383 expression in desert rodents compared to non-desert relatives (367 genes, shown in
384 red in Figure 3A). Additionally, across all genes, fold changes between individual
385 desert-mesic species were found to be significantly correlated in 2 of the 3
386 comparisons of species pairs (Spearman's rank correlation rho, *C. intermedius/H.*
387 *desmarestianus* vs. *J. jaculus/Z. princeps, p*=0.0062, *rho*=0.03; *C. intermedius/H.*
388 *desmarestianus* vs. *N. alexis/M. musculus*, *p*< 2.2e-16, *rho*=0.10; *N. alexis/M. musculus*
389 vs. *J. jaculus/Z. Princeps, p*=0.12, *rho*=-0.017).
390
391 To identify genes and pathways of interest, we divided the set of convergently
392 expressed genes into those that are upregulated with respect to the desert taxa in all
393 comparisons and those that are downregulated with respect to the desert taxa in all
394 comparisons and performed phenotype and GO term enrichment tests on these (see
395 methods). Genes convergently upregulated across desert rodents were enriched for
396 several GO terms related to gene regulation, including regulation of RNA metabolic
397 process (q=2.55 x 10^{-5}), regulation of gene expression (q=1.34E-5), and regulation of
398 RNA biosynthetic process (3.87 x 10^{5}). Genes downregulated in desert rodents
399 were enriched for GO terms related to metabolic processes, including metabolic
400 process (q=1.56 x 10^{3}), organic substance metabolic process (q=3.93 x 10^{-3}), and
401 cellular metabolic process (3.54 x 10^{-3}). Genes with evidence for convergent
402 differential expression included genes with mouse mutant phenotypes related to

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.10.459863; this version posted September 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

403 kidney development and physiology or homeostasis (Table S5). For example,
404 Aquaporin 11 (*Aqp11*) is expressed at a lower level in all desert species compared to
405 non-desert species in all three comparisons (Figure 3C). This gene is part of a family
406 of genes encoding membrane-integrated channels responsible for water transfer
407 across membranes throughout the body. Aquaporins have been repeatedly
408 implicated in studies of desert adaptation across rodents (Marra et al. 2012; Marra
409 et al. 2014; Pannabecker 2015; Giorello et al. 2018). Mouse knockouts have
410 demonstrated that *Aqp11* is necessary for proximal tubular function and the
411 formation of healthy kidneys (Morishita et al. 2005; Tchekneva et al. 2008). In
412 addition, *Aqp11* plays a role in salivary gland development (Larsen et al. 2010). This
413 set also includes genes associated with human phenotypes related to kidney and
414 renal diseases (Table S6); for example, mutations in the gene *col4a5,* which is
415 downregulated in desert species, have been associated with Alport syndrome, a
416 disease characterized by kidney inflammation (Köhler et al. 2019).
417
418 *Genes under selection in desert lineages*
419 Next, we tested for evidence of selection on protein coding sequences using well
420 aligned one-to-one orthologs found in all desert-mesic species pairs (1474 genes).
421 We searched for genes showing signatures of selection using a model that compares
422 the rate of nonsynonymous substitutions with the rate of synonymous substitutions
423 in a phylogenetic context, ω or dN/dS (Yang, 1998, 2007). We performed three
424 analyses using the 1,474 single copy orthologs with high quality alignments present
425 in all species. When testing for evidence of selection on desert species compared to
426 non-desert species, we uncovered 39 genes (39/1474= 2.6%) for which ω was
427 significantly higher in the three "foreground" desert lineages compared with the
428 three "background" non-desert lineages (Table S7, FDR<0.1). This group is enriched
429 for phenotypes related to multiple aspects of the immune response as well as to
430 hearing/vestibular/ear phenotypes and other aspects of osteology (Table S8).
431 Immune genes are some of the fastest evolving genes in the genome and are
432 disproportionately found to be under selection in many studies (Hurst and Smith
433 1999; Schlenke and Begun 2003; Nielsen et al. 2005). One gene of particular

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.10.459863; this version posted September 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

434 interest, unrelated to immunity, is FAT atypical cadherin 4 (*FAT4*) (q = 0.018). *FAT4*
435 has been implicated in human kidney diseases (Alders et al. 2014) and is involved in
436 normal kidney development through modulating the RET signaling pathway in
437 mouse models (Mao et al. 2015; Zhang et al. 2019). *FAT4* homozygous knockout
438 mice have smaller kidneys with the presence of cysts in renal tubules when
439 compared with wild type mice and they die within a few hours of birth (Saburi et al.
440 2008). These phenotypes in laboratory mice make this an interesting candidate
441 gene for future studies in desert rodents. We found three genes that showed
442 evidence of positive selection (when the three desert species were treated together
443 as foreground lineages) and also showed convergent shifts in gene expression
444 (Rows 1-3 in Table S9), however they are not known to be associated with
445 phenotypes of interest. This amount of overlap is no more than expected by chance
446 (hypergeometric test, p=0.64).
447
448 We then tested whether ω was significantly higher in each of the three "foreground"
449 desert lineages individually compared with the five remaining taxa. We identified 23
450 genes in *C. intermedius*, 19 in *J. jaculus*, and 18 in *N. alexis* where ω was significantly
451 elevated (at FDR <0.1)(Table S10). These genes are candidates for lineage-specific
452 adaptations. In *C. intermedius*, enriched phenotypes were related to immunity and
453 morphological traits including kidney size, while in *J. jaculus* and *N. alexis*, enriched
454 phenotype terms were related to behavioral and electrophysiological traits (Table
455 S11). In the *Chaetodipus* comparison, *Dusp4* is of some interest as it has been
456 associated with aberrant circulating solute levels in mouse models. Deletion of this
457 gene has been associated with increased excreted protein and altered kidney
458 structure in diabetic mice (Denhez et al. 2019). It is also convergently differentially
459 expressed. Overall, the amount of overlap (hypergeometric test, p > 0.06 in all
460 comparisons) between any of these lists and differentially expressed genes between
461 lineage pairs is no more than expected by chance (Table S9).
462
463 In the third analysis, we employed a branch-site model to identify genes in which
464 specific codons may be under positive selection. In this approach, genes for which

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.10.459863; this version posted September 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

465 specific codons have a ω > 1 in the "foreground" branch (defined to include all three
466 desert species) compared with the "background" branch are identified. Seven genes
467 were identified (Table S12) with codons under selection in all three desert lineages,
468 including *Coro2b*, a gene implicated in abnormal renal glomerulus morphology
469 (Schwarz et al. 2019) and urine protein level (Rogg et al. 2017) and *Bloc1s4*, which
470 is implicated in abnormal renal physiology (Gwynn et al. 2000). Again, there was no
471 significant overlap with the genes identified in the differential expression analysis
472 (p=0.47; Table S9).
473
474
475

476 **Discussion**

477 The molecular basis of convergent evolution has been well studied for a number of
478 simple traits, but has been less studied for complex traits. Even fewer studies have
479 compared convergence in both gene expression and protein evolution for complex
480 traits. Here, we studied convergence in gene expression and amino acid sequence in
481 three species of desert rodents and their non-desert relatives, from across the
482 rodent tree, representing ~70 million years of evolution.
483
484 Despite the long evolutionary timeframe and the fact that most expression evolution
485 tracked phylogeny (Figure 2), we identified a surprising number of genes
486 (702/8174=8.6%) that showed convergent shifts in gene expression (Figure 3). This
487 number is more than expected by chance, and this result is highly significant (Figure
488 3B). We note, however, that the number of genes showing convergent expression
489 does not reflect the number of causative changes (i.e. mutational events in
490 evolution), since many of these convergent changes in expression might reflect
491 downstream consequences of a smaller number of changes at upstream regulators
492 that govern networks of co-regulated genes. Nonetheless, the large number of
493 convergent changes in expression suggests that a measureable amount of desert
494 adaptation is mediated by a large set of shared changes in gene regulation, whether
495 at the level of individual genes or through sets of co-regulated genes.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.10.459863; this version posted September 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

496
497 In addition to these shared changes in gene expression, we identified a large
498 number of species-specific changes in gene expression in each species pair. Perhaps
499 not surprising given the long evolutionary timescales and complexity of
500 osmoregulatory function, much of the evolutionary response appears to be specific
501 to individual lineages.
502
503 In contrast to the fairly long evolutionary timescales in the present study, we
504 recently documented differences in kidney gene expression between desert and
505 non-desert populations of *Mus musculus* separated by only a few hundred
506 generations of evolution (Bittner et al. 2021). In that study, we identified 3,935
507 differentially expressed genes of which 99 were found to be convergent across all
508 three desert lineages in the present study. The lack of significant overlap
509 (hypergeometric test, p=0.99) suggests that over long evolutionary timescales,
510 adaptive responses to xeric conditions may be quite different from the evolved
511 changes in gene expression over short evolutionary timescales.
512
513 In contrast to the number of convergent changes in gene expression, we observed
514 few genes that showed evidence of positive selection on amino acid sequences
515 (39/1474= 2.6%) among desert species. These proportions are not directly
516 comparable since the methods used to detect convergence and positive selection are
517 quite different. Nonetheless, our analyses suggest that the phenotypic convergence
518 seen in urine concentration is reflected at the molecular level more in patterns of
519 gene regulation than in patterns of protein evolution.
520
521 Although expression evolution and amino acid sequence evolution have been found
522 to be correlated in some cases (Nuzhdin et al. 2004; Lemos et al. 2005), we did not
523 find significant overlap in the number of genes showing convergent gene expression
524 and convergent amino acid sequence evolution. The small amount of overlap might
525 reflect differences in the selection pressures on these two classes of changes. For
526 example, *cis*-regulatory changes in gene expression are often controlled in a tissue-

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.10.459863; this version posted September 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

527 specific and developmental-stage-specific manner, and as such are expected to be
528 less pleiotropic and thus less constrained in evolution (e.g. Wray 2007). Protein-
529 coding changes, on the other hand, affect all tissues and developmental stages in
530 which the protein is expressed and thus may be more pleiotropic and consequently
531 more constrained. The small amount of overlap might also reflect both statistical
532 and methodological limitations of our study. First, the analytic methods used to
533 detect convergent expression changes and convergent amino acid changes are quite
534 distinct and likely have different false-negative and false-positive rates. Second, we
535 studied gene expression in adults, yet gene expression varies considerably during
536 kidney development (Schwab et al. 2003) and early expression is undoubtedly
537 important in establishing morphological differences between desert and non-desert
538 kidneys. Third, kidneys have a heterogenous cellular composition, and changes in
539 cellular composition between species are likely to affect measures of gene
540 expression in bulk preparations. Future studies of gene expression in single cell
541 preparations at early developmental time points might uncover additional signals of
542 adaptation to desert conditions. It would also be worthwhile to study expression
543 changes in both males and females since the demands of water balance may be
544 especially acute in lactating females.
545
546 Despite these caveats, we identified a number of potential candidate genes
547 associated with desert adaptation, including some that showed both convergent
548 gene expression and convergent amino acid sequence evolution. The target
549 available to selection in a trait as complex as desert adaptation is likely large and
550 constrained along each lineage to a different degree by other aspects of the
551 organism's morphology and physiology. Nonetheless, an interesting outcome of our
552 analysis is that a number of the genes and pathways identified here are similar to
553 those identified in other studies of rodent and mammalian desert adaptation (Marra
554 et al. 2012; Marra et al. 2014; Wu et al. 2014; MacManes 2017; Giorello et al. 2018;
555 Tigano et al. 2020). It is clear that gene families such as aquaporins, which are
556 responsible for facilitating water transport across membranes, and solute carriers,
557 may play a role in mitigating water loss across multiple systems and therefore

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.10.459863; this version posted September 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

558 underlie convergent evolution at the genetic level to desert environments.
559 Altogether, our results demonstrate the power of studying convergent evolution at
560 multiple levels by integrating scans for convergent evolution on both the amino acid
561 and gene expression level to identify genes and pathways of interest. Our results
562 suggest that changes in gene regulation may play an essential role in how the
563 kidneys of species separated by 70 million years of evolution have convergently
564 evolved ultra-efficient osmoregulatory capacity to contend with the stressors of
565 harsh desert environments.
566
567

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.10.459863; this version posted September 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

568 **Acknowledgements**

569 We thank members of the Nachman Lab for valuable comments and discussions. We
570 thank Dr. Kim Cooper, Dr. Kevin Rowe, and Dr. Beatriz Otero Jiménez for generously
571 sharing their samples as well as Dr. Andy Gloss, Dr. Shea Lambert, Dr. Aaron
572 Ragsdale, Dr. Taichi Suzuki, Ned McAllister and the Oakland Zoo, and Kim Hemmer
573 and Golden Gate Fields for their help with field collections. We thank Dr. Jim Patton,
574 Dr. Andy Gloss, Lydia Smith, and Jeremy Richardson for their technical expertise.
575 This work was supported by a NSF Doctoral Dissertation Improvement Grant to
576 NKJB (1601827), grants from the Wilhelm L.F. Martens Fund and the David and
577 Marvalee Wake Fund of the Museum of Vertebrate Zoology, and an NIH grant to
578 MWN (R01 GM127468).

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.10.459863; this version posted September 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

References

Alders M, Al-Gazali L, Cordeiro I, Dallapiccola B, Garavelli L, Tuysuz B, Salehi F, Haagmans MA, Mook OR, Majoie CB, et al. 2014. Hennekam syndrome can be caused by FAT4 mutations and be allelic to Van Maldergem syndrome. Hum Genet. 133(9):1161–1167. doi:10.1007/s00439-014-1456-y.

Al-kahtani MA, Zuleta C, Caviedes-Vidal E, Garland Jr Theodore. 2004. Kidney Mass and Relative Medullary Thickness of Rodents in Relation to Habitat, Body Size, and Phylogeny. Physiological and Biochemical Zoology. 77(3):346–365. doi:10.1086/420941.

Altschuler EM, Nagle RB, Braun EJ, Lindstedt SL, Krutzsch PH. 1979. Morphological study of the desert heteromyid kidney with emphasis on the genus perognathus. The Anatomical Record. 194(3):461–468. doi:10.1002/ar.1091940311.

Bairoch A, Apweiler R. 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Res. 28(1):45–48. doi:10.1093/nar/28.1.45.

Bankir L, de Rouffignac C. 1985. Urinary concentrating ability: insights from comparative anatomy. American Journal of Physiology-Regulatory, Integrative and Comparative Physiology. 249(6):R643–R666. doi:10.1152/ajpregu.1985.249.6.R643.

Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer ELL, et al. 2004. The Pfam protein families database. Nucleic Acids Res. 32(suppl_1):D138–D141. doi:10.1093/nar/gkh121.

Baudinette RV. 1972. The impact of social aggregation on the respiratory physiology of Australian hopping mice. Comp Biochem Physiol A Comp Physiol. 41(1):35–38. doi:10.1016/0300-9629(72)90030-8.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.10.459863; this version posted September 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological). 57(1):289–300.

Beuchat CA. 1990. Body size, medullary thickness, and urine concentrating ability in mammals. American Journal of Physiology-Regulatory, Integrative and Comparative Physiology. 258(2):R298–R308. doi:10.1152/ajpregu.1990.258.2.R298.

Bittner NKJ, Mack KL, Nachman MW. 2021. Gene expression plasticity and desert adaptation in house mice. Evolution. n/a(n/a). doi:https://doi.org/10.1111/evo.14172. https://onlinelibrary.wiley.com/doi/abs/10.1111/evo.14172.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30(15):2114–2120. doi:10.1093/bioinformatics/btu170.

Bradley WG, Yousef MK, Scott IM. 1975. Physiological studies on the rock pocket mouse, Perognathus intermedius. Comparative Biochemistry and Physiology Part A: Physiology. 50(2):331–337. doi:10.1016/0300-9629(75)90022-5.

Degen AA. 1997. Water Requirements and Water Balance. In: Ecophysiology of Small Desert Mammals. Berlin, Heidelberg: Springer Berlin Heidelberg. (Cloudsley-Thompson JL, editor.). p. 93–162. http://link.springer.com/10.1007/978-3-642-60351-8_6.

Denhez B, Rousseau M, Dancosst D-A, Lizotte F, Guay A, Auger-Messier M, Côté AM, Geraldes P. 2019. Diabetes-Induced DUSP4 Reduction Promotes Podocyte Dysfunction and Progression of Diabetic Nephropathy. Diabetes. 68(5):1026–1039. doi:10.2337/db18-0837.

Donald J, Pannabecker TL. 2015. Osmoregulation in Desert-Adapted Mammals. In: Hyndman KA, Pannabecker TL, editors. Sodium and Water Homeostasis: Comparative, Evolutionary and Genetic Models. New York, NY: Springer. (Physiology

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.10.459863; this version posted September 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

in Health and Disease). p. 191–211. https://doi.org/10.1007/978-1-4939-3213-9_10.

Donald JA, Hamid NKA, Horvath P, McLeod J. 2012. Appetite regulation in desert-adapted Spinifex hopping mice during water deprivation. The FASEB Journal. 26:1069.4-1069.4.

Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. 2009. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics. 10(1):48. doi:10.1186/1471-2105-10-48.

Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biology. 16(1):157. doi:10.1186/s13059-015-0721-2.

Fleming TH. 1977. Response of Two Species of Tropical Heteromyid Rodents to Reduced Food and Water Availability. J Mammal. 58(1):102–106. doi:10.2307/1379741.

Giorello FM, Feijoo M, D'Elía G, Naya DE, Valdez L, Opazo JC, Lessa EP. 2018. An association between differential expression and genetic divergence in the Patagonian olive mouse (Abrothrix olivacea). Molecular Ecology. 27(16):3274–3286. doi:10.1111/mec.14778.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology. 29(7):644–652. doi:10.1038/nbt.1883.

Gu Z, Nicolae D, Lu HH-S, Li W-H. 2002. Rapid divergence in expression between duplicate genes inferred from microarray data. Trends in Genetics. 18(12):609–613. doi:10.1016/S0168-9525(02)02837-8.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.10.459863; this version posted September 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Gwynn B, Ciciotte SL, Hunter SJ, Washburn LL, Smith RS, Andersen SG, Swank RT, Dell'Angelica EC, Bonifacino JS, Eicher EM, et al. 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(13):4227–4235.

Hao Y, Xiong Y, Cheng Y, Song G, Jia C, Qu Y, Lei F. 2019. Comparative transcriptomics of 3 high-altitude passerine birds and their low-altitude relatives. PNAS. 116(24):11851–11856. doi:10.1073/pnas.1819657116.

Hurst LD, Smith NG. 1999. Do essential genes evolve slowly? Curr Biol. 9(14):747–750. doi:10.1016/s0960-9822(99)80334-0.

Karageorgi M, Groen SC, Sumbul F, Pelaez JN, Verster KI, Aguilar JM, Hastings AP, Bernstein SL, Matsunaga T, Astourian M, et al. 2019. Genome editing retraces the evolution of toxin resistance in the monarch butterfly. Nature. 574(7778):409–412. doi:10.1038/s41586-019-1610-8.

Khalil F, Tawfic J. 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(3):259–271. doi:10.1002/jez.1401540302.

Köhler S, Carmody L, Vasilevsky N, Jacobsen JOB, Danis D, Gourdine J-P, Gargano M, Harris NL, Matentzoglu N, McMurry JA, et al. 2019. Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. Nucleic Acids Res. 47(D1):D1018–D1027. doi:10.1093/nar/gky1105.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Clin Adv Hematol Oncol. 9(4):357–359. doi:10.1038/nmeth.1923.

Larsen HS, Ruus A-K, Schreurs O, Galtung HK. 2010. Aquaporin 11 in the developing mouse submandibular gland. European Journal of Oral Sciences. 118(1):9–13. doi:10.1111/j.1600-0722.2009.00708.x.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.10.459863; this version posted September 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Lemos B, Bettencourt BR, Meiklejohn CD, Hartl DL. 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. Mol Biol Evol. 22(5):1345–1354. doi:10.1093/molbev/msi122.

Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 22(13):1658–1659. doi:10.1093/bioinformatics/btl158.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology. 15(12):550. doi:10.1186/s13059-014-0550-8.

MacManes MD. 2017. Severe acute dehydration in a desert rodent elicits a transcriptional response that effectively prevents kidney injury. American Journal of Physiology-Renal Physiology. 313(2):F262–F272. doi:10.1152/ajprenal.00067.2017.

MacMillen RE, Lee AK. 1967. Australian desert mice: independence of exogenous water. Science. 158(3799):383–385. doi:10.1126/science.158.3799.383.

Macmillen RE, Lee AK. 1969. Water metabolism of Australian hopping mice. Comparative Biochemistry and Physiology. 28:493–514.

Makova KD, Li W-H. 2003. Divergence in the Spatial Pattern of Gene Expression Between Human Duplicate Genes. Genome Res. 13(7):1638–1645. doi:10.1101/gr.1133803.

Mao Y, Francis-West P, Irvine KD. 2015. Fat4/Dchs1 signaling between stromal and cap mesenchyme cells influences nephrogenesis and ureteric bud branching. Development. 142(15):2574–2585. doi:10.1242/dev.122630.

Marcovitz A, Turakhia Y, Chen HI, Gloudemans M, Braun BA, Wang H, Bejerano G. 2019. A functional enrichment test for molecular convergent evolution finds a clear

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.10.459863; this version posted September 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

protein-coding signal in echolocating bats and whales. PNAS. 116(42):21094–21103. doi:10.1073/pnas.1818532116.

Marra NJ, Eo SH, Hale MC, Waser PM, DeWoody JA. 2012. A priori and a posteriori approaches for finding genes of evolutionary interest in non-model species: osmoregulatory genes in the kidney transcriptome of the desert rodent Dipodomys spectabilis (banner-tailed kangaroo rat). Comp Biochem Physiol Part D Genomics Proteomics. 7(4):328–339. doi:10.1016/j.cbd.2012.07.001.

Marra NJ, Romero A, DeWoody JA. 2014. Natural selection and the genetic basis of osmoregulation in heteromyid rodents as revealed by RNA-seq. Molecular Ecology. 23(11):2699–2711. doi:10.1111/mec.12764.

Morishita Y, Matsuzaki T, Hara-chikuma M, Andoo A, Shimono M, Matsuki A, Kobayashi K, Ikeda M, Yamamoto T, Verkman A, et al. 2005. Disruption of Aquaporin-11 Produces Polycystic Kidneys following Vacuolization of the Proximal Tubule. Molecular and Cellular Biology. 25(17):7770–7779. doi:10.1128/MCB.25.17.7770-7779.2005.

Mundy NI. 2005. A window on the genetics of evolution: MC1R and plumage colouration in birds. Proceedings of the Royal Society B: Biological Sciences. 272(1573):1633–1640. doi:10.1098/rspb.2005.3107.

Nielsen R, Bustamante C, Clark AG, Glanowski S, Sackton TB, Hubisz MJ, Fledel-Alon A, Tanenbaum DM, Civello D, White TJ, et al. 2005. A scan for positively selected genes in the genomes of humans and chimpanzees. PLoS Biol. 3(6):e170. doi:10.1371/journal.pbio.0030170.

Nuzhdin SV, Wayne ML, Harmon KL, McIntyre LM. 2004. Common Pattern of Evolution of Gene Expression Level and Protein Sequence in Drosophila. Mol Biol Evol. 21(7):1308–1317. doi:10.1093/molbev/msh128.

Pannabecker TL. 2015. Aquaporins in Desert Rodent Physiology. The Biological Bulletin. 229(1):120–128. doi:10.1086/BBLv229n1p120.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.10.459863; this version posted September 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Parker DJ, Bast J, Jalvingh K, Dumas Z, Robinson-Rechavi M, Schwander T. 2019. Repeated Evolution of Asexuality Involves Convergent Gene Expression Changes. Mol Biol Evol. 36(2):350–364. doi:10.1093/molbev/msy217.

Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. Nature Methods. 14(4):417–419. doi:10.1038/nmeth.4197.

Rocha JL, Godinho R, Brito JC, Nielsen R. 2021 Apr 13. Life in Deserts: The Genetic Basis of Mammalian Desert Adaptation. Trends in Ecology & Evolution. doi:10.1016/j.tree.2021.03.007. https://www.sciencedirect.com/science/article/pii/S0169534721000744.

Rogg M, Yasuda-Yamahara M, Abed A, Dinse P, Helmstädter M, Conzelmann AC, Frimmel J, Sellung D, Biniossek ML, Kretz O, et al. 2017. The WD40-domain containing protein CORO2B is specifically enriched in glomerular podocytes and regulates the ventral actin cytoskeleton. Sci Rep. 7(1):15910. doi:10.1038/s41598-017-15844-1.

Saburi S, Hester I, Fischer E, Pontoglio M, Eremina V, Gessler M, Quaggin SE, Harrison R, Mount R, McNeill H. 2008. Loss of Fat4 disrupts PCP signaling and oriented cell division and leads to cystic kidney disease. Nat Genet. 40(8):1010–1015. doi:10.1038/ng.179.

Sackton TB, Grayson P, Cloutier A, Hu Z, Liu JS, Wheeler NE, Gardner PP, Clarke JA, Baker AJ, Clamp M, et al. 2019. Convergent regulatory evolution and loss of flight in paleognathous birds. Science. 364(6435):74–78. doi:10.1126/science.aat7244.

Schlenke TA, Begun DJ. 2003. Natural selection drives Drosophila immune system evolution. Genetics. 164(4):1471–1480.

Schmidt-Nielsen K. 1964. Desert Animals: Physiological Problems of Heat and Water. New York: Dover Publications.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.10.459863; this version posted September 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Schmidt-Nielsen K, Schmidt-Nielsen B. 1952. Water Metabolism of Desert Mammals. Physiological Reviews. 32(2):135–166. doi:10.1152/physrev.1952.32.2.135.

Schwab K, Patterson LT, Aronow BJ, Luckas R, Liang H-C, Potter SS. 2003. A catalogue of gene expression in the developing kidney. Kidney Int. 64(5):1588–1604. doi:10.1046/j.1523-1755.2003.00276.x.

Schwarz A, Möller-Hackbarth K, Ebarasi L, Jess DU, Zambrano S, Blom H, Wernerson A, Lal M, Patrakka J. 2019. Coro2b, a podocyte protein downregulated in human diabetic nephropathy, is involved in the development of protamine sulphate-induced foot process effacement. Sci Rep. 9(1):1–11. doi:10.1038/s41598-019-45303-y.

Sela I, Ashkenazy H, Katoh K, Pupko T. 2015. GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. Nucleic Acids Res. 43(W1):W7–W14. doi:10.1093/nar/gkv318.

Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: Assessing Genome Assembly and Annotation Completeness. In: Kollmar M, editor. Gene Prediction: Methods and Protocols. New York, NY: Springer. (Methods in Molecular Biology). p. 227–245. https://doi.org/10.1007/978-1-4939-9173-0_14.

Sikes RS, Gannon WL. 2011. Guidelines of the American Society of Mammalogists for the use of wild mammals in research. Journal of Mammalogy. 92(1):235–253. doi:10.1644/10-MAMM-F-355.1.

Smith-Unna R, Boursnell C, Patro R, Hibberd JM, Kelly S. 2016. TransRate: reference-free quality assessment of de novo transcriptome assemblies. Genome Res. 26(8):1134–1144. doi:10.1101/gr.196469.115.

Song L, Florea L. 2015. Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads. GigaScience. 4(1):48. doi:10.1186/s13742-015-0089-y.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.10.459863; this version posted September 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Stewart C-B, Wilson AC. 1987. Sequence Convergence and Functional Adaptation of Stomach Lysozymes from Foregut Fermenters. Cold Spring Harb Symp Quant Biol. 52:891–899. doi:10.1101/SQB.1987.052.01.097.

Tchekneva EE, Khuchua Z, Davis LS, Kadkina V, Dunn SR, Bachman S, Ishibashi K, Rinchik EM, Harris RC, Dikov MM, et al. 2008. Single Amino Acid Substitution in Aquaporin 11 Causes Renal Failure. JASN. 19(10):1955–1964. doi:10.1681/ASN.2008030296.

Tigano A, Colella JP, MacManes MD. 2020. Comparative and population genomics approaches reveal the basis of adaptation to deserts in a small rodent. Molecular Ecology. 29(7):1300–1314. doi:10.1111/mec.15401.

Weinreich DM, Delaney NF, DePristo MA, Hartl DL. 2006. Darwinian Evolution Can Follow Only Very Few Mutational Paths to Fitter Proteins. Science. 312(5770):111–114. doi:10.1126/science.1123539.

Weng M-P, Liao B-Y. 2017. modPhEA: model organism Phenotype Enrichment Analysis of eukaryotic gene sets. Bioinformatics. 33(21):3505–3507. doi:10.1093/bioinformatics/btx426.

Wray GA. 2007. The evolutionary significance of cis -regulatory mutations. Nature Reviews Genetics. 8(3):206–216. doi:10.1038/nrg2063.

Wu H, Guang X, Al-Fageeh MB, Cao J, Pan S, Zhou H, Zhang L, Abutarboush MH, Xing Y, Xie Z, et al. 2014. Camelid genomes reveal evolution and adaptation to desert environments. Nature Communications. 5(1):1–10. doi:10.1038/ncomms6188.

Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Computer applications in the biosciences : CABIOS. 13(5):555–556. doi:10.1093/bioinformatics/13.5.555.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.10.459863; this version posted September 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Yang Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol Biol Evol. 15(5):568–573. doi:10.1093/oxfordjournals.molbev.a025957.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24(8):1586–1591. doi:10.1093/molbev/msm088.

Zhang H, Bagherie-Lachidan M, Badouel C, Enderle L, Peidis P, Bremner R, Kuure S, Jain S, McNeill H. 2019. FAT4 Fine-Tunes Kidney Development by Regulating RET Signaling. Developmental Cell. 48(6):780-792.e4. doi:10.1016/j.devcel.2019.02.004.

Zhen Y, Aardema ML, Medina EM, Schumer M, Andolfatto P. 2012. Parallel Molecular Evolution in an Herbivore Community. Science. 337(6102):1634–1637. doi:10.1126/science.1226630.

Data Accessibility Statement

Illumina sequencing data from this study is available is available through the NCBI Sequence Read Archive under accession PRJNA656179. Samples collected by NKJB are accessioned into the Museum of Vertebrate Zoology collection.

Author Contributions

This study was designed by all authors. NKJB conducted the experiments, and NKJB and KLM analyzed the data. The paper was written by NKJB and edited by MWN and KLM.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.10.459863; this version posted September 11, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure Legends

Figure 1. A. Estimates of urine concentration from across Mammalia from 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 is 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 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 (red 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. Grey dotted lines are drawn between the mean of normalized expression for each desert-mesic pair. Points are jittered for clarity.

Figure 1

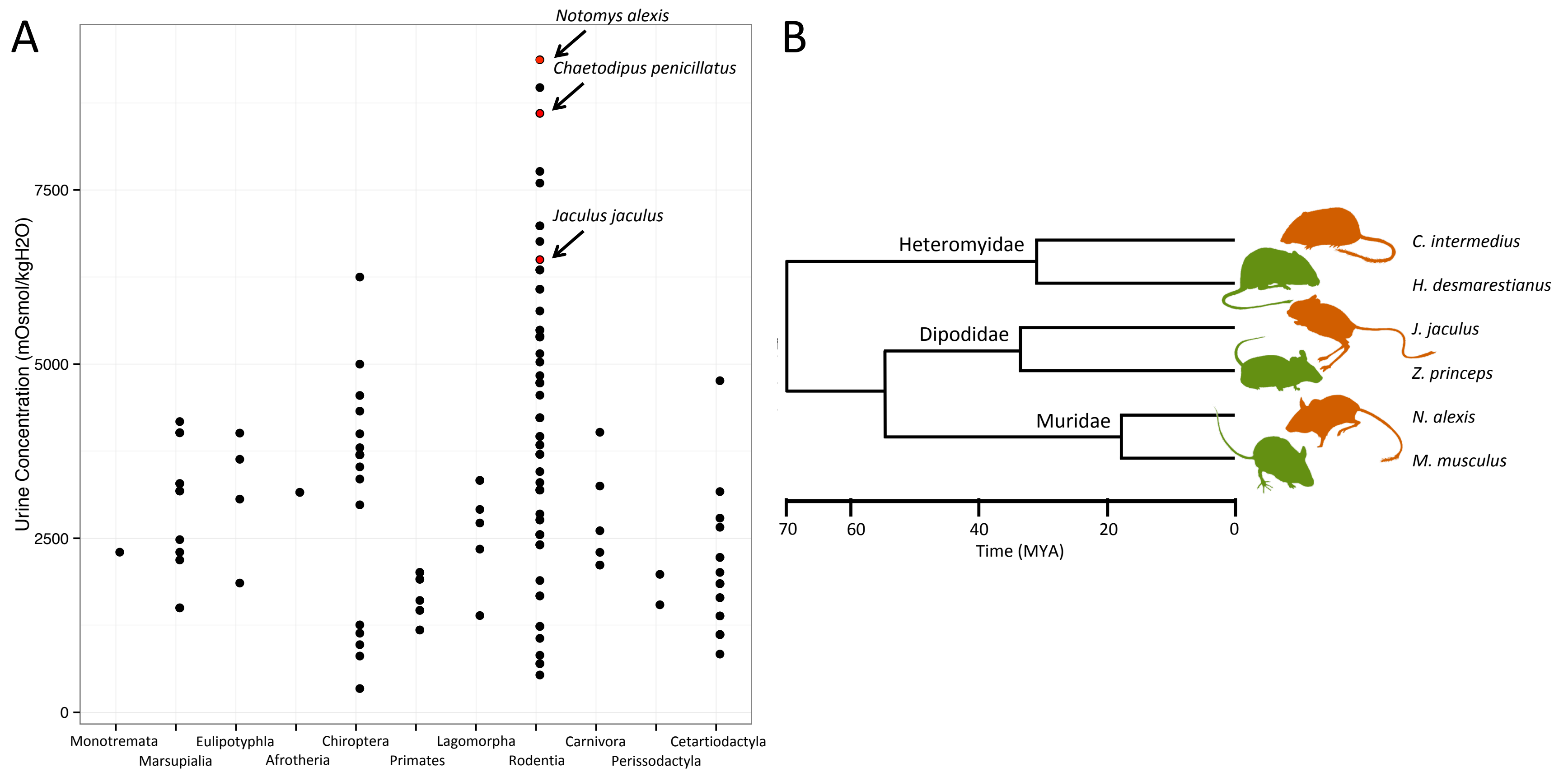


Figure 2

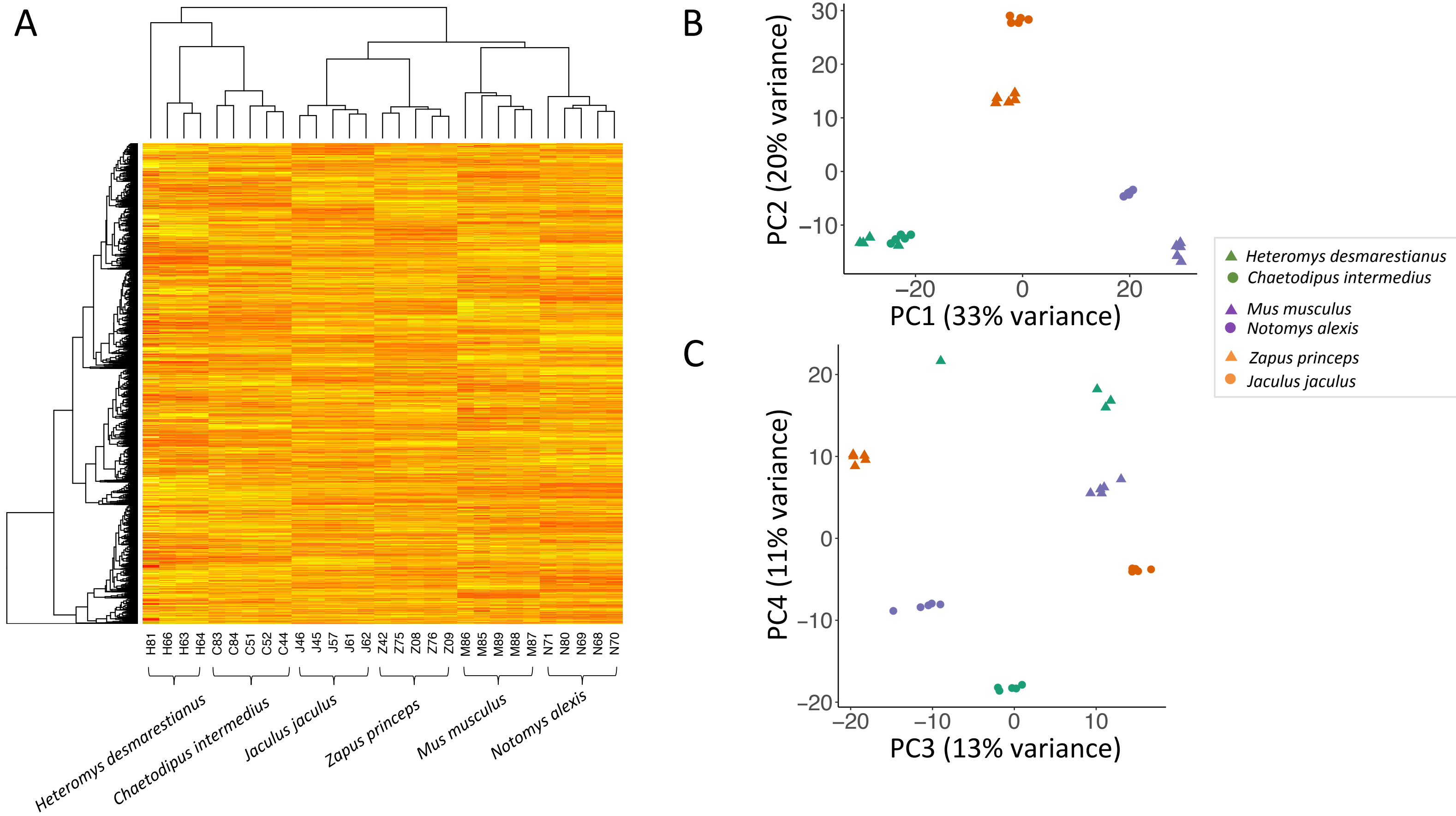
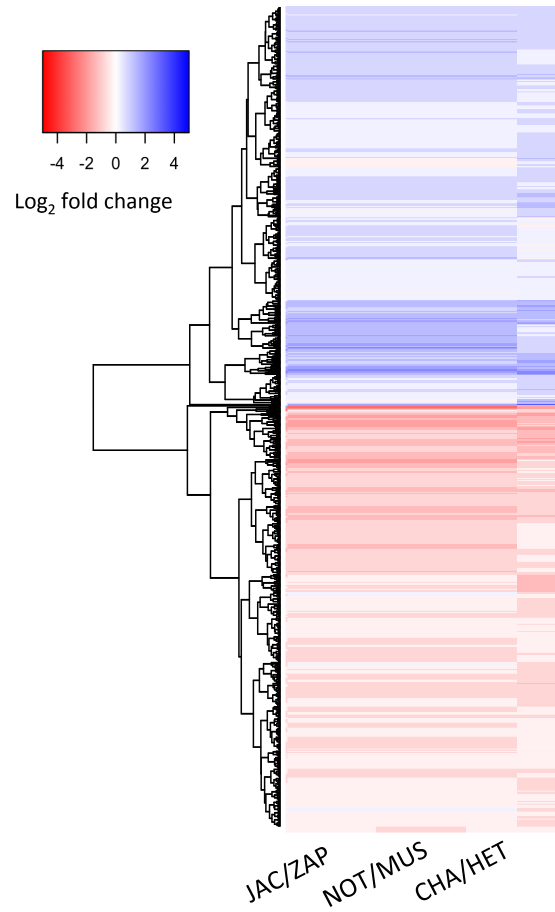
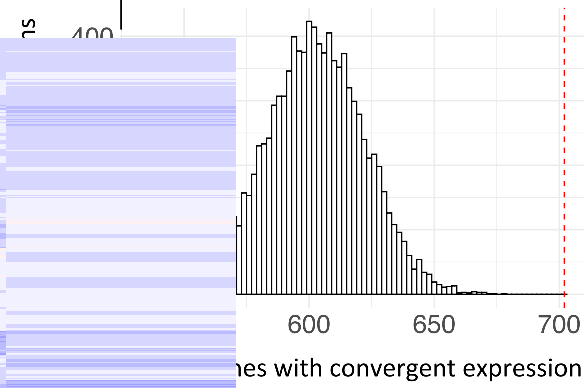


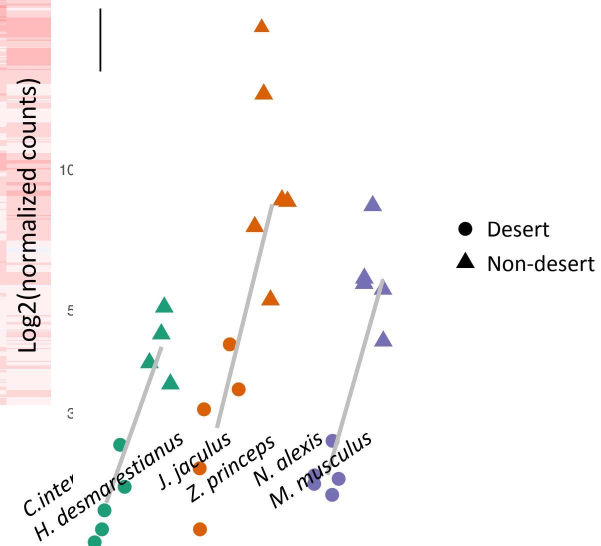
Figure 3

A



B





Supplemental material

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

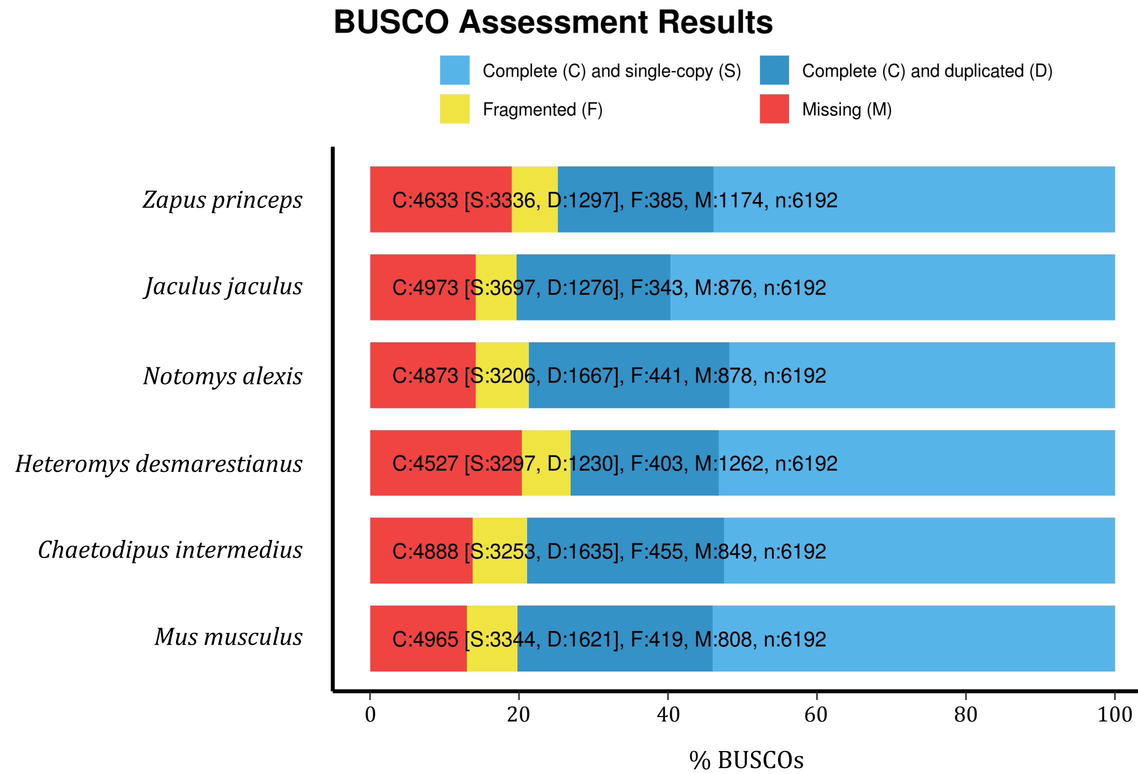


Figure S2. Genes that are differentially expressed between each desert and non-desert comparisons within each family and the overlaps among families.

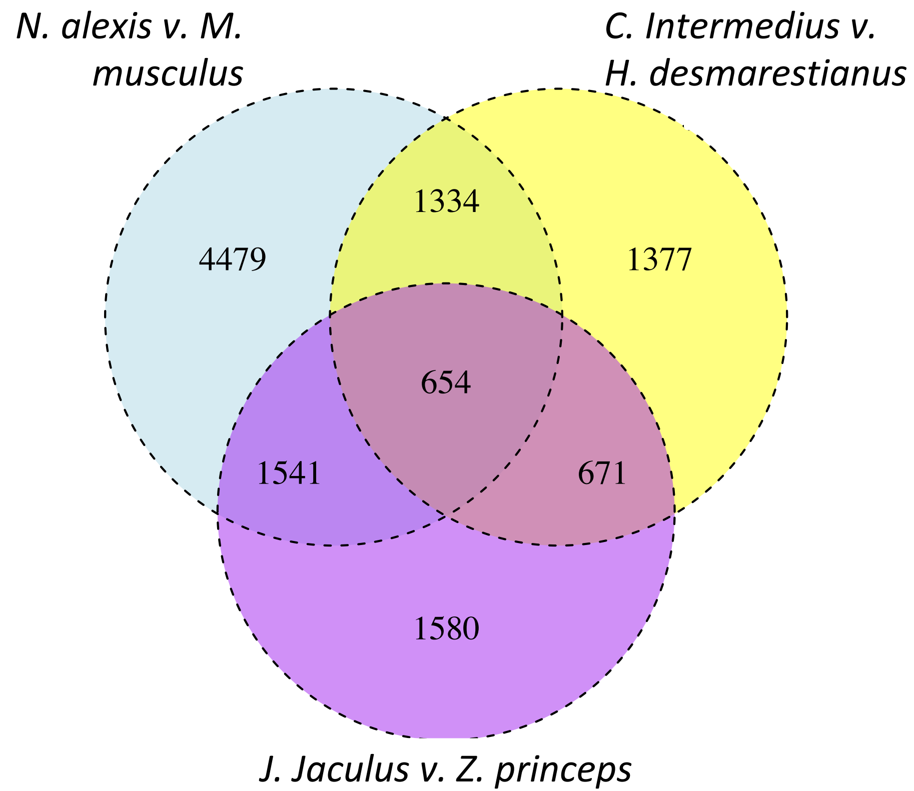


Figure S3. Magnitude of expression differences between each desert-mesic species pair