Original article

Multiple evolutionary pathways to achieve thermal adaptation in small mammals

Jocelyn P. Colella[1,2,†]*, Anna Tigano[1,2], Olga Dudchenko[3,4,5], Arina D. Omer[3], Ruqayya Khan[3,4], Ivan D. Bochkov[3,4], Erez L. Aiden[3,4,5,6,7], Matthew D. MacManes[1,2]

[1]University of New Hampshire, Molecular, Cellular, and Biomedical Sciences Department, Durham, NH 03824, USA (JPC: Jocelyn.Colella@unh.edu; AT: Anna.Tigano@unh.edu; MDM: Matthew.MacManes@unh.edu)

[2]Hubbard Genome Center, University of New Hampshire, Durham, NH 03824, USA

[3]The Center for Genome Architecture, Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA

[4]Department of Computer Science, Department of Computational and Applied Mathematics, Rice University, Houston, TX 77030, USA

[5]Center for Theoretical and Biological Physics, Rice University, Houston, TX 77030, USA

[6]Shanghai Institute for Advanced Immunochemical Studies, ShanghaiTech University, Shanghai 201210, China

[7]School of Agriculture and Environment, University of Western Australia, Perth, WA 6009, Australia

[†]University of Kansas Biodiversity Institute and Ecology and Evolutionary Biology Department, Lawrence, KS 66045, USA

* Corresponding Author: Jocelyn.Colella@unh.edu

ABSTRACT

Rapid ecological radiations provide useful models for identifying instances of parallel evolution, which can highlight critical genomic architecture involved in shared adaptations. Thermoregulatory innovations have allowed deer mice of the genus *Peromyscus* to radiate throughout North America, exploiting extreme thermal environments from mountain tops to desert valleys, and positioning this taxon as a model for understanding thermal adaptation. We compare signatures of selective sweeps across population-level genomic resequencing data from two desert-adapted *Peromyscus* species (*P. crinitus* and *P. eremicus*) and a third, widely-distributed habitat generalist (*P. maniculatus*) to test for signatures of parallel evolution and identify shared genomic architecture involved in adaptation to hot deserts. We found limited evidence of parallel evolution. Instead, we identified divergent molecular mechanisms of adaptation to similar environments potentially tied to species-specific historical demography that may limit or enhance adaptive variation. We also identified numerous genes under selection in *P. crinitus* that are implicated in osmoregulation (*Trypsin, Prostasin*) and metabolic responses to desert life (*Kallikrein, eIF2-alpha kinase GCN2, APPL1/2*). Evidence of varied evolutionary routes to achieve the same phenotype suggest there may be many molecular trajectories for small mammals to accommodate anthropogenic climate change.

Key words: convergence, dehydration, desert, parallel evolution, *Peromyscus,* thermoregulation

INTRODUCTION

Increasing global temperatures and altered patterns of precipitation threaten biodiversity worldwide (Moritz et al. 2008; Cahill et al. 2013; Urban 2015). Phenotypic plasticity enables an immediate response to changing conditions but evolutionary change through adaptation will be critical for the long-term survival of most species (Hoffman and Sgro 2011; Cahill et al. 2013). Range shifts upward in elevation and latitude have been documented in a number of terrestrial species and interpreted as a response to warming (Chen et al. 2011; Tingley and Beissinger 2013; Freeman et al. 2018); however, responses vary even among closely-related species or populations (Hoffman and Willi 2008; Moritz et al. 2008). The physiological limits responsible for organismal range shifts are in part governed by genetics, which can facilitate adaptation to specific environmental conditions. Population genomic methods enable the identification of genes and molecular pathways involved in local adaptation by scanning the genome for signatures of selection (Bassham et al. 2018; Garcia-Elfring et al. 2019). For species that have independently adapted to similar environments, parallel or convergent evolution can be inferred if a greater number of genes or phenotypes share signatures of selection than would be expected under a purely stochastic model of evolution (e.g., drift). Convergence typically implies the evolution of similar adaptive responses independently among distantly related taxa in response to similar environmental or ecological conditions; in turn, parallel evolution is defined as the occurrence of similar adaptive changes in groups with common ancestry (Simpson 1961; Wood et al. 2005). Convergent evolution is often presumed to be driven by different underlying molecular mechanisms, whereas parallel evolution may be driven by similar mechanisms; however, this generalization may not reflect reality (Arendt and Reznick 2008). Evidence of parallel or convergent evolution can suggest a deterministic effect of selection and highlight conserved genomic architecture involved in shared adaptive phenotypes (Rundle et al. 2000; McDonald et al. 2009), while a lack of concerted evolution may identify novel evolutionary strategies to achieve the same phenotypic result.

As a model taxon (Dewey and Dawson 2001; Bedford and Hoekstra 2015) inhabiting varied environments throughout North America, deer mice in the genus *Peromyscus* are a frequent and productive subject of classical adaptation studies (e.g., physiological, Storz 2007; behavioral, Hu and Hoekstra 2017; genetic, Cheviron et al. 2012; Storz and Cheviron 2016; Tigano et al. 2020). Physiological similarity of deer mice to lab mice (*Mus musculus*) further broadens the implications of evolutionary and ecological investigations of *Peromyscus* by linking relevant results to biomedical sciences. The genus *Peromyscus* (N = 67 species; mammaldiversity.org) is hypothesized to be the product of a rapid ecological radiation across North America, evident in their varied ecological niches and rich species diversity (Glazier 1980; Riddle et al. 2000; Bradley et al. 2007; Platt et al. 2015; Lindsey 2020). Adaptive radiations are useful natural experiments for identifying patterns of parallel or convergent evolution, or the lack thereof. Short generation times and accelerated thermoregulatory evolution relative to other mammals, among other adaptive responses, appear to have enabled *Peroymscus* rodents to exploit extreme thermal environments, ranging from cold, high elevations (Pierce and Vogt 1993; Cheviron et al. 2012, 2014; Kaseloo et al. 2014; Garcia-Elfring et al. 2019) to arid, hot deserts (Riddle et al. 2000; MacManes 2017; Tigano et al. 2020). Thermoregulation and dehydration tolerance are complex physiological traits and there are several potential evolutionary routes to achieve the same phenotypic outcome. Within this framework, comparisons among divergent *Peromyscus* species adapted to similar environments may highlight shared adaptive polymorphisms or disparate evolutionary paths central to achieving the same phenotype (Cheviron et al. 2012; Ivy and Scott 2017; Hu and Hoekstra 2017; Storz et al. 2019). In cold environments, endotherms rely on aerobic thermogenesis to maintain constant internal body temperatures. Changes in both gene expression and the functional properties of proteins in high-altitude adapted deer mice suggest that changes in multiple hierarchical molecular pathways may be common in the evolution of complex physiological traits, such as thermoregulation (Wichman and Lynch 1991; Storz 2007; Cheviron et al. 2012; Storz and

Cheviron 2016; Garcia-Elfring et al. 2019). Nonetheless, investigations of thermoregulation among high-elevation species may be confounded by concurrent selection on hemoglobin oxygen-binding affinity as a consequence of a reduction in the partial pressure of oxygen as elevation increases substantially (Storz and Kelly 2008; Storz et al. 2010; Natarajan et al. 2015). In hot environments, endotherms are challenged with balancing heat dissipation, energy expenditure, and water retention (Anderson and Jetz 2005), resulting in a different suite of behavioral, physiological, and molecular adaptations that enable survival (Schwimmer and Haim 2009; Degen 2012; Kordonowy et al. 2016), but may be confounded by acute or chronic dehydration. Understanding the biochemical mechanisms that enable survival under extreme environmental stress can provide important insights into the nature of physiological adaptation.

Rapid thermoregulatory and ecological diversification among *Peromyscus* species (origin ~8 Mya, radiation ~5.71 Mya; Platt et al. 2015) positions these small rodents as models for anticipating species responses to accelerated warming (Cahill et al. 2013). Desert specialist phenotypes have evolved repeatedly during the course of the *Peromyscus* radiation, with each species and populations therein subject to distinct histories of demographic variation and gene flow. These idiosyncratic histories can have a direct impact on evolution, as effective population sizes are inextricably linked to the efficacy of selection and maintenance of genetic diversity in wild populations (Charlesworth 2009). Further, contemporary or historical gene flow may help or hinder adaptive evolution through homogenization or adaptive introgression, respectively (Coyne and Orr 2004; Morjan and Reiseberg 2004; Jones et al. 2018; Tigano and Friesen 2016). Native to the American West, the canyon mouse (*P. crinitus*, Fig 1.) is well adapted to desert life. In the lab, *P. crinitus* can survive in the absence of exogenous water, with urine concentration levels similar to that of desert-adapted kangaroo rats (*Dipodomys merriami*; Abbott 1971; MacMillen 1972; MacMillen and Christopher 1975; MacMillien 1983), but without equivalently specialized renal anatomy (Issaian et al. 2012). Canyon mice also exhibit a lower-than-expected body temperature relative to their size and can enter environmentally-mediated

torpor in response to drought, food limitation, or low external temperatures (McNab 1968; McNab and Morrison 1963; Morhardt and Hudson 1966; Johnson and Armstrong 1987), which facilitates survival in highly-variable and extreme desert environments. These phenotypes persist for multiple generations in the lab indicating they have a genomic basis (McNab and Morrison 1968). Cactus mice (*P. eremicus*) are related to *P. crinitus* and the two species are frequently sympatric. Cactus mice exhibit similar adaptations to desert environments, including urine concentration, reduced water requirements, and environmentally-induced torpor (Veal and Caire 1979). In contrast, the habitat generalist *P. maniculatus* (North American deer mouse) is geographically-widespread, native to both cool, high-elevations and hot southwestern deserts. Whole-genome assemblies are publicly available for both *P. eremicus* (Tigano et al. 2020) and *P. maniculatus* (Harvard University, Howard Hughes Medical Institute), which positions these species as ideal comparatives against *P. crinitus* to identify genes and regulatory regions associated with desert adaptation. Without fossil evidence of divergence and subsequent convergence between desert-adapted *Peromyscus* species, similar patterns of selection are interpreted as evidence of parallel evolution.

Here, we investigate genomic signatures of selection in desert-adapted *P. crinitus*. We contrast signatures of selective sweeps across three related *Peromyscus* species, two desert specialists (*P. crinitus* and *P. eremicus*) and one habitat generalist (*P. maniculatus*), to test for signatures of parallel evolution. We hypothesize that similar genes or functional pathways will be under selection in both desert-adapted species and not under selection in *P. maniculatus*, providing a signature of parallel evolution. Finally, we place selective sweep analyses into an evolutionary framework to interpret the varied evolutionary trajectories available to small mammals to respond to changing environmental conditions and to account for demographic and gene flow events.

METHODS

105 *De novo genome sequencing and assembly*

106 Wild mice were handled and sampled in accordance with the University of New Hampshire and
107 University of California Berkeley's Institutional Animal Care and Use Committee (130902 and
108 R224-0310, respectively) and California Department of Fish and Wildlife (SC-008135) and the
109 American Society of Mammalogists best practices (Sikes and Animal Care and Use Committee
110 of the American Society of Mammalogists 2016).

111 For the assembly of the *P. crinitus* genome, DNA was extracted from a liver subsample
112 from a *P. crinitus* individual collected in 2009 from the Philip L. Boyd Deep Canyon Desert
113 Research Center in Apple Valley, California. To generate a high-quality, chromosome-length
114 genome assembly for this individual we extracted high-molecular-weight genomic DNA using a
115 Qiagen genomic tip kit (Qiagen, Inc, Hilden, Germany). A 10X Genomics linked-reads library
116 was prepared according to manufacturer protocol at Mount Sinai and sequenced to a depth of
117 70X on a HiSeq 4000 (Novogene, Sacramento, California, USA). 10X Genomics reads were *de*
118 *novo* assembled into contigs using *Supernova* 2.1.1 (Weisenfeld et al. 2017). To arrange the
119 scaffolds thus obtained in chromosomes, a Hi-C library for *P. crinitus* was constructed and
120 sequenced from primary fibroblasts from the T.C. Hsu Cryo-Zoo at the University of Texas MD
121 Anderson Cancer Center. The Hi-C data were aligned to the supernova assembly using *Juicer*
122 (Durand et al. 2016). Hi-C genome assembly was performed using the *3D-DNA* pipeline
123 (Dudchenko et al. 2017) and the output was reviewed using *Juicebox Assembly Tools*
124 (Dudchenko et al. 2018). The Hi-C data are available on
125 www.dnazoo.org/assemblies/Peromyscus_crinitus visualized using *Juicebox.js*, a cloud-based
126 visualization system for Hi-C data (Robinson et al. 2018).

127 Benchmarking Universal Single-Copy Orthologs (*BUSCO* v3, using the Mammalia odb9
128 database; Simão et al. 2015) and *OrthoFinder2* (Emms and Kelly 2015) were used to assess
129 genome quality and completeness. Genome sizes were estimated for each species using
130 *abyss-fac* (Simpson et al. 2009) and the *assemblathon_stats.pl* script available at:

https://github.com/ucdavis-bioinformatics/assemblathon2-analysis/. *RepeatMasker* v.4.0 (Smit et al. 2015) was used to identify repetitive elements. The genome was annotated using the software package *MAKER* (3.01.02; Campbell et al. 2014). Control files, protein, and transcript data used for this process are available at https://github.com/macmanes-lab/pecr_genome/tree/master/annotation. We used *Mashmap* (-f one-to-one --pi 90 -s 300000; Jain et al. 2017, 2018) to assess and plot (generateDotPlot.pl) syntenic conservation between *P. crinitus* and *P. maniculatus* genomes. *Peromyscus crinitus* chromosomes were renamed and sorted using *seqtk* (github.com/lh3/seqtk) following the *P. maniculatus* chromosome naming.

For comparative genomic analyses, we generated low-coverage whole-genome resequencing data for nine *P. crinitus* and five *P. maniculatus* individuals collected from syntopic areas of southern California (Table S1). *Peromyscus crinitus* samples were collected from the University of California (UC) Philip L. Boyd Deep Canyon Desert Research Center (DCDRC) near Palm Desert, California, and *P. maniculatus* were collected further East from the UC Motte Rimrock Reserve and Elliot Chaparral Reserves. In addition to these, we used publicly available whole-genome resequencing data from 26 *P. eremicus* individuals, also collected from DCDRC and Motte Rimrock Reserve and prepared and sequenced in parallel (Tigano et al. 2020). All samples were collected in 2009, with the exception of eight *P. eremicus* samples which were collected in 2018. Animals were collected live in Sherman traps and a 25 mg ear-clip was taken from each individual and stored at -80ºC in 95% ethanol. Animals were sampled from arid areas with average monthly temperatures between 9-40ºC and mean annual rainfall of 15-18 cm. The Biotechnology Resource Center at Cornell University (Ithaca, NY, USA) prepared genomic libraries using the Illumina Nextera Library Preparation kit (e.g., skim-seq). Libraries were sequenced at Novogene (Sacramento, CA, USA) using 150 bp paired-end reads from one lane on the Illumina NovaSeq S4 platform. *fastp* v. 1 (Chen et al. 2018) was used to assess read quality and trim adapters. Sequences from all samples and all species were mapped with *BWA* (Li and Durbin 2010) to the *P. crinitus* reference genome to enable comparative analyses,

duplicates removed with *samblaster v. 0.1.24* (Faust and Hall 2014), and alignments indexed and sorted using *samtools v. 1.10* (Li et al. 2009).

Population Genomics

We used the software package *ANGSD v. 0.93* (Korneliussen et al. 2014) to call variants from low-coverage population genomic data with high confidence with the general options: *-SNP_pval 1e-6, -minMapQ 20, -minQ 20, -setMinDepth 20, -minInd 20, -minMaf 0.01*. *ANGSD* was run across all species and again within each species, where we required a minimum of half (*-minInd*) *P. crinitus* and *P. eremicus* samples and all *P. maniculatus* samples to meet independent quantity (*-minMapQ*) and quality (*-minQ*) thresholds and sample representation for each variable site in each species.

Differentiation among species was examined using a Principle Component Analysis (PCA, see Supporting Materials; *ngsTools,* Fumagalli et al. 2014) and multidimensional scaling (MDS) of principal components in *NGSadmix v. 33* (Skotte et al. 2013). MDS plots were generated in *R* v.3.6.1 (R Core Team 2017) based on the covariance matrix. Cook's D was used to identify MDS outliers, using the broken stick method to identify single samples with undue influence (Cook and Weisberg 1984; Williams 1987). *NGSadmix* was used to fit genomic data into K populations to parse species-level differences and provide a preliminary screen for genomic admixture under a maximum-likelihood model. Nonetheless, expanded sample sizes, including representatives from additional populations of each species are necessary to thoroughly investigate patterns of population structure and introgression. We tested K = 1 through K = (N - 1), where N is the number of total individuals examined. *NGSadmix* was run for all species combined and for each species independently. As an additional measure of differentiation, we estimated weighted and unweighted global F_{ST} values for each species pair using *realSFS* in *ANGSD*.

We used Pairwise sequential Markovian Coalescent (*PSMC v. 0.6.5-r67;* Li and Durbin 2011) to examine patterns of historical demography through time for each species. The original reads used to generate the high-quality, chromosome-length assemblies for each species (*P. crinitus* generated here; *P. eremicus*, SAMEA5799953, Tigano et al. 2020; *P. maniculatus*: GCA_003704035.1, Harvard University) were mapped to their assembly reference to identify heterozygous sites and indexed in *BWA*. *Samblaster* removed PCR duplicates and *picard* (http://broadinstitute.github.io/picard/) added a read group to the resulting bam file and generated a sequence dictionary (*CreateSequenceDictionary*) from the reference assembly. *Samtools* was used to sort, index, and variants called (*mpileup*) for each species, with *bcftools v1.10.2* (*call*, Li et al. 2009) and *VCFtools v 0.1.16* (*vcf2fq*, Danecek et al. 2011). *PSMC* distributions of effective population size (N_e) were estimated with 100 bootstrap replicates. *PSMC* results were visualized through *gnuplot v. 5.2* (Williams and Kelley 2010), using perl scripts available at github.com/lh3/psmc, and scaled by a generation time of 6 months (0.5 yr, Millar, 1989; Pergams and Lacy 2008) and a general mammalian mutation rate of 2.2×10^{-9} substitutions/site/year (Kumar and Subramanian 2002).

Tests for selection & convergence

To detect recent selective sweeps in low-coverage whole-genome data, we used *Sweepfinder2* (DeGiorgio et al. 2016; Nielsen et al. 2005). *Sweepfinder2* was run on both variant and invariant sites (Huber et al. 2016) for each species, excluding sex chromosomes. Sex-chromosomes were excluded for three reasons: (1) sex chromosome evolution is both rapid and complex relative to autosomes, (2) we had different sample sizes of each sex across species, and (3) desert adaptations, the focus of this study, are unlikely to be sex-specific. We repeated *Sweepfinder2* analyses on *P. eremicus*, initially analyzed by Tigano et al. (2020), using an improved annotation scheme based on *Peromyscus*-specific data rather than *Mus musculus* genes. Allele frequencies were estimated in *ANGSD*, converted to allele counts, and the site

frequency spectrum (SFS) was estimated from autosomes only in *Sweepfinder2*. Sweeps were estimated from the pre-computed SFS and the composite likelihood ratio (CLR) and alpha values, indicating the strength of selection, were calculated every 10,000 sites. Per Tigano et al. (2020), a 10 kb window size was selected as a trade-off between computational time and resolution. CLR values above the 99.9^{th} percentile of the empirical distribution for each species were considered to be evolving under a model of natural selection, hereafter referred to as significant sweep sites. Smaller sample sizes produce fewer bins in the SFS and a diminishing number of rare alleles which may impact both the overall SFS and local estimate surrounding testing sites; therefore, we explored the impact of sample sizes on *Sweepfinder2* results in the Supporting Information.

For each species, mean Tajima's D was calculated across the entire genome in non-overlapping windows of 10 kb and 1 kb in *ANGSD*. Nucleotide diversity (π) was also calculated in 10 kb and 1 kb windows and corrected based on the number of sites genotyped (variant and invariant) per window. Tajima's D and π are expected to be significantly reduced in regions surrounding selective sweeps (Smith and Haigh 1974; Kim and Stephan 2002), therefore we used a Mann-Whitney test ($p < 0.05$, after a Bonferroni correction for multiple tests) to measure significant deviations from the global mean in 1 kb and 10 kb flanking regions surrounding significant sweep sites and 27 candidate genes identified in previous studies (MacManes 2017; Table S2). Candidate loci include aquaporins (N = 12), sodium-calcium exchangers (*SLC8a1*), and *Cyp4* genes belonging to the Cytochrome P450 gene family (N = 14). We used custom python scripts to functionally annotated (I) the closest gene to each significant sweep site, (II) the nearest upstream and downstream gene, regardless of strand (sense/antisense), and (III) the nearest upstream and downstream gene on each strand. Dataset I follows the general assumption that proximity between a significant sweep site and a protein-coding gene suggests interaction. Dataset II represents an extension of that model by encompassing the most proximal gene in each direction. Because *Sweepfinder2* is performed on the consensus

sequence and our data is unphased, we do not have information indicating on which strand a significant sweep site occurs. Therefore, dataset III encompasses strand-uncertainty by including the two nearest genes to a significant sweep site on both strands. It should be noted that the genes identified in smaller datasets (I, II) are nested within the larger datasets (II, III) and by definition, the larger datasets include more noise which may dilute a signature of parallel evolution, but may better capture the true signal of selection. Hence, it is important to critically examine numerous hierarchical gene subsets. We tested genes from each dataset for functional and pathway enrichment in Gene Ontology (GO) categories using *Panther v. 15.0* (Mi et al. 2017) and extracted GO terms for each enriched functional group. We used *Mus musculus* as a reference and a Bonferroni correction for multiple tests ($p < 0.05$) to correct for false discoveries. Enriched GO terms were summarized and visualized in *REVIGO* (Reduce and Visualize Gene Ontology, Supek et al. 2011) implemented at: http://revigo.irb.hr/index.jsp?error=expired. As a test for convergence, the overlap in the gene names and enriched GO terms associated with significant selective sweeps was assessed for each dataset. Overlap was visualized in the R package *VennDiagram* (Chen and Boutros 2011). To test for convergence, we used a Fisher's Exact Test ($p < 0.05$) in the *GeneOverlap* package (Shen 2016) in R to assess whether gene or enriched GO term overlap between species was greater than expected based on the total number of genes in the genome.

To compare patterns of gene family expansion and contraction potentially involved in adaptation within the genus *Peromyscus*, we analyzed 14 additional genomes, including ten *Peromyscus* species and four near outgroup rodent species: *Microtus ochrogaster, Neotoma lepida, Sigmodon hispidus*, and *Mus musculus* (Table S3). To prevent bias driven by variable assembly qualities, samples with < 70% complete mammalian BUSCOs were excluded from downstream analyses, resulting in the final analysis of ten species. Groups of orthologous sequences (orthogroups) were identified in *Orthofinder2*. Invariant orthogroups and groups that varied by more than 25 genes across taxa (custom python script: ortho2cafe.py) were excluded.

Our rooted species tree, estimated in *Orthofinder2*, was used to calculate a single birth-death parameter (lambda) and estimate changes in gene family size using *CAFE v.4.2.1* (Han et al. 2013). Results were summarized using the python script *cafetutorial_report_analysis.py* available from the Hahn Lab: hahnlab.github.io/CAFÉ/manual.html.

RESULTS

Chromosome-length genome assembly for P. crinitus

Linked reads combined with Hi-C scaffolding produced a high-quality, chromosome-length genome assembly for *P. crinitus*. Our assembly has a contig N50 of 137,026 bp and scaffold N50 of 97,468,232 bp, with 24 chromosome-length scaffolds. The anchored sequences in the three *Peromyscus* genome assemblies were as follows: *P. crinitus* genome ~2.27 Gb, *P. eremicus* ~2.5 Gb, and *P. maniculatus* ~2.39 Gb (Table 1). Our assembly has high contiguity and completeness and low redundancy, as demonstrated by the presence of 89.3% complete BUSCOs, 0.9% of duplicates and 9.0% missing, excluding unplaced scaffolds. As anticipated (Smalec et al. 2019), we found no significant variation in chromosome number or major interchromosomal rearrangements between *P. crinitus* and *P. maniculatus* (Figure S4). We annotated 17,265 total protein coding genes in the *P. crinitus* genome. Similar to other *Peromyscus* species, LINES1 (long interspersed nuclear elements) and LTR (long terminal repeats) elements comprised 22.7% of the repeats in the *P. crinitus* genome, with SINEs (short interspersed nuclear elements) representing an additional 9.6% (Table S5). Although similar to other *Peromyscus* species, *P. crinitus* has the greatest total repeat content (>37%; see Tigano et al. 2020 Supplementary Table 2).

Population Genomics

MDS analysis parsed the three species into well-separated clusters and identified no outliers or evidence of admixture (Fig. S6). Analysis of genetic structure identified all three species as a

single group (K = 1) with the highest likelihood. A three-population model neatly parsed the three species, as expected (Fig. S7, Table S8). We found evidence of potential admixture in *P. crinitus* with at least three individuals containing 11-27% ancestry from *P. eremicus* and additional material from *P. maniculatus* (4-16%), although variable samples sizes may impact assignment certainty and expanded sequencing of additional species and populations will be required to identify the specific sources of introgressed material. Four *P. eremicus* individuals had < 90% assignment probability to the *P. eremicus* species cluster, with a maximum of 15% assignment to a different species cluster. Identification of admixture in both species is not biased by differences in coverage, as low (2X), medium (8X), and high coverage (17X) samples were found to be admixed at a < 90% assignment threshold. No *P. maniculatus* individuals were identified as admixed. *PSMC* estimates of historical demography show greater variance and a higher overall N_e for *P. crinitus* relative to *P. eremicus* (Fig. 2). Demographic estimates for *P. maniculatus* are included as a reference but should be interpreted with caution as they are based on a captive-bred individual and may not accurately reflect the demography of wild populations.

Average Tajima's D (1 kb windows) was negative for all species and ranged from -0.69 to -1.61. *Peromyscus crinitus* had the lowest Tajima's D value and *P. maniculatus* the highest (Fig. S9). Global pairwise F_{ST} was moderate between all species ranging from 0.20-0.27 (unweighted: 0.12-0.17). Mean global π (1 kb windows) was 0.005 (±0.005) for *P. crinitus,* 0.007 (± 0.007) for *P. eremicus*, and 0.012 (± 0.010) for *P. maniculatus* (Fig. S10).

Selection & convergence

Within *P. crinitus* we identified a total of 209 significant sweep sites (Table S11), with 104 sites localized on chromosome 9 and 16 regions experiencing major selective sweeps (Fig. 3). Despite the large size of chromosome 1, we found no significant sweep sites on this chromosome for *P. crinitus*. We found 239 total significant sweep sites for *P. eremicus* (Table

S12), with 56 sites concentrated on chromosome 1. Finally, we identified a total of 213 significant sweep sites for *P. maniculatus* (Table S13), with 103 sites located on chromosome 4.

A significant selective sweep (CLR > 99.9%) affects the area surrounding the site tested - including both protein coding and non-coding regions - and cannot identify the specific nucleotides under selection. Under the assumption that proximity suggests interaction between a functional region of the genome and a sweep site, we hierarchically examined protein coding genes most proximal to each sweep site. On average the distance from a sweep site to the nearest coding gene was 45 kbp in *P. crinitus* (range: 31 - 439,122 bp), and much greater for both *P. maniculatus* (average: 152 kbp; range: 190 - 513,562 bp) and *P. eremicus* (average: 117 kbp; range: 38 - 1,423,027 bp), which may be partially explained by differences in assembly quality. For both *P. eremicus* and *P. maniculatus*, only two significant sweep sites were identified within protein-coding genes (*P. eremicus:* Meiosis-specific with OB domain-containing protein [gene name: *MEIOB*], Harmonin [*Ush1c*]; *P. maniculatus:* Dehydrogenase/reductase SDR family member 7B [*DRS7B*] and Zinc finger protein 217 [*ZN217*]). In contrast, for *P. crinitus* 12 significant sweep sites fell within 19 distinct candidate loci, many of which code for numerous alternatively spliced transcripts (Table 2). Among *P. crinitus* significant sweep sites localized within coding sequences, we identified 19 enriched GO terms (3 Biological Process [BP], 9 Molecular Function [MF], 7 Cellular Component [CC]), with functionality ranging from 'proteolysis' to 'hydrolase' activity (Fig. 4). Functional examination of those genes identifies solute regulation as a key function, with genes pertaining to calcium (*Trypsin-2* [*PRSS2*]) and zinc (*Kallikrein-4* [*KLK4*]) binding and sodium regulation (*Prostasin* [*PRSS8*]) identified as under selection. Examination of the two nearest genes to each sweep site (dataset II: one upstream, one downstream gene, regardless of strand) in *P. crinitus* identified 121 unique genes and 26 enriched GO (gene ontology) terms (8 Biological Processes [BP], 10 Molecular Functions [MF], 8 Cellular Component [CC]), with functionality pertaining to metabolism (e.g., 'protein metabolic process', 'organonitrogen compound metabolic process', 'peptide metabolic process') and

ribosomes (Fig. 4; Tables S11-13). For *P. eremicus,* we identified 202 unique genes and 14 enriched GO terms (0 BP, 1 MF, 13 CC) associated with selective sweeps, with functionality again centered around ribosomes. For dataset II, two genes and seven enriched GO terms were shared between the two desert-adapted species and we found no relationship between genes identified under selection between the two species. Functional enrichment of *P. eremicus* and *P. maniculatus* across all datasets was limited to ribosomes (e.g., 'structural constituent of ribosome', 'cytosolic ribosome', 'ribosomal subunit'; Fig. 4; Tables S14-16). In contrast, functionality of enriched GO terms for *P. crinitus* centered on metabolic processes, including protein breakdown, hydrolysis, and cellular functionality (e.g., 'organelle', 'intracellular', 'cytoplasm'; Fig. 4; Table S16), in addition to ribosomes. See Supporting Information for detailed results for datasets I and III. *Peromyscus eremicus* and *P. maniculatus* shared significant overlap ($p < 0.05$) in enriched GO terms across all hierarchical data subsets (I, II, III; Fig. 5). Significant overlap of enriched GO terms was also detected between *P. crinitus* and both other *Peromyscus* species for datasets II and III only, with zero overlap for dataset I (Fig. 5). Significant overlap in the genes located near significant sweep sites among desert-adapted *Peromyscus* was only detected in dataset III. Overall, *P. crinitus* consistently had greater diversity of functional enrichment relative to the other two species and the GO terms and genes involved in ribosomal functionality were frequently shared among all species.

Species tree estimates (Fig. S17) were consistent with previous phylogenetic investigations (Bradley et al. 2007). *Peromyscus crinitus* and *P. eremicus* are related in our species tree, however a number of intermediate taxa remain unsampled (e.g., *P. merriami*, *P. californicus*). Among the species examined here, the desert adapted *crinitus-eremicus* clade is related to a clade comprised of *P. leucopus, P. polionotus*, and *P. maniculatus,* with the *nasutus-attwateri* clade most basal within *Peromyscus.* For the *Peromyscus* genus, we found 19,925 gene families that had experienced contractions, 502 expansions, and 12 families that

were evolving rapidly. However, we found no gene families experiencing significant expansions, contractions, or rapid evolution within *Peromyscus* or below the genus level.

Counter expectations, Tajima's D and π surrounding significant selective sweep sites were significantly higher than the global mean estimates for each species (Table S18). Only in *P. maniculatus* did we detect a significant reduction in π surrounding significant sweep sites. Mean Tajima's D surrounding *a priori* candidate loci were also significantly more positive in all three species.

DISCUSSION

Continued and accelerating environmental change increases the exigency of accurately anticipating species responses to anthropogenic climate change. Adaptive evolutionary responses vary among species and populations, even when subjected to seemingly synonymous environmental selective pressures (Bi et al. 2015; Garcia-Elfring et al. 2019), but evidence of parallel or convergent evolution can highlight critical genomic architecture involved in key adaptations. We broadly define convergence as the evolution of similar adaptive responses among distantly related taxa in response to similar environmental or ecological conditions and parallel evolution as the occurrence of similar adaptive changes in groups that share recent common ancestry (Simpson 1961; Wood et al. 2005). We analyzed genome-wide patterns of selective sweeps among three species of deer mice within the North American genus *Peromyscus* to identify critical genomic architecture or alternative evolutionary pathways to achieve desert adaptation. We hypothesized that desert specialists *P. crinitus* and *P. eremicus* would share similar patterns of selective sweeps related to surviving high-temperature, low-water environments indicative of parallel or convergent evolution. The species examined here share a common ancestor and the two desert adapted taxa share an even more recent ancestor (Fig. S17); however, there are number of unsampled species separating *P. crinitus* from *P. eremicus* (e.g., *P. merriami*, *P. californicus*) and these two taxa are relatively

389 divergent (average Fst of 0.25; divergence > 1 Mya, Platt et al. 2015). Nonetheless, without
390 evidence of ancestral divergence followed by subsequent convergence (molecular or
391 phenotypic) we cannot disentangle convergence from parallel evolution within the context of this
392 investigation. Instead, we infer parallel evolution as the most parsimonious explanation for
393 coordinated selection among desert-adapted *Peromyscus* species. As such, both selection on
394 novel beneficial mutations (typically deemed convergence) and the coincident retention of
395 advantageous ancestral polymorphisms are interpreted as evidence in support of parallel
396 evolution, as we do not have ancestral state information to distinguish the former.

397 Overall, we rejected our hypothesis that desert adapted species share parallel
398 signatures of selective sweeps reflecting adaptation to similar environments. Instead, we
399 identified divergent molecular mechanisms of adaptation to desert environments, with *P. crinitus*
400 potentially responding primarily through genomic changes to protein coding genes and *P.*
401 *eremicus* through transcriptional regulation of gene expression. This result potentially
402 contradicts the generalization that convergent evolution is driven by different molecular
403 mechanisms and parallel evolution, by similar mechanisms (Arendt and Reznick 2008).
404 Molecular flexibility of thermoregulatory responses may have catalyzed the radiation of
405 *Peromyscus* in North America by enabling rapid exploitation of novel thermal environments.
406 Finally, the application of an evolutionary lens to the interpretation of genomic patterns of
407 selective sweeps, particularly one that integrates historical demography and gene flow, may
408 serve to inform which evolutionary mechanism (genomic vs. transcriptomic) will be most
409 efficacious to achieve similar adaptive phenotypes in other small mammals.
410
411 *Limited evidence of parallel evolution*
412 Identification of similar genes or functions under selection in different species adapted to similar
413 environments suggests a deterministic effect of natural selection (Rosenblum et al. 2014) and
414 provides evidence in support of parallel evolution. In contrast, we found limited evidence of

parallel evolution among desert-adapted *Peromyscus*. Few to no enriched GO terms overlapped between desert-specialists (Fig. 5). Instead, GO terms relating to ribosomes (e.g., 'ribosome', 'ribosomal subunit', 'cystolic ribosome', etc.) broadly overlapped between all three *Peromyscus* species examined, with the most significant overlap occurring between desert adapted *P. eremicus* and generalist *P. maniculatus*. Although *P. maniculatus* are not desert specialists, the individuals sequenced here were collected in arid regions of southern California; therefore, the shared signature of selection on ribosomes across all three species may reflect broadly shared adaptations to hot and dry conditions or relate to thermoregulatory plasticity among *Peromyscus* rodents. We found few genes experiencing significant selective sweeps shared among *Peromyscus* species, with only one significant relationship: 10 genes, among hundreds, were shared between the two desert specialists (dataset III; Fig. 5) and were consistent with a signature of parallel evolution. Nonetheless, many of the overlapping genes are directly related to broad ribosomal functionality (e.g., *RL36, RS26, RL15, RS2*) also shared with *P. maniculatus*. Further, selective sweeps are only one way to detect signatures of parallel evolution and this hypothesis remains to be explored in greater detail using additional methods (Booker et al. 2017, Weigand and Leese 2018).

Although we found no significantly expanded or contracted gene families within the genus *Peromyscus*, previous investigations of the entire Myodonta clade within Rodentia identified multiple expanded or contracted gene families associated with ribosomes in *P. eremicus* (Tigano et al. 2020). Ribosomes play a critical role in protein synthesis and degradation. Cellular damage accumulates quickly in desert environments as a consequence of increased thermal- and osmotic-stress (Lamitina et al. 2006; Burg et al. 2007). In response, changes in gene expression modulate osmoregulation by removing and replacing damaged proteins to prevent cell death (Lamitina et al. 2006). While ribosomes appear to be a target of parallel evolution in desert-adapted *Peromyscus*, this genomic signature is not unique. Instead, selection on ribosomal functionality may be convergent across many species adapted to distinct

441 thermal environments (metazoans; Porcelli et al. 2015). Ribosomes are evolutionarily linked to
442 the mitochondrial genomes of animals (Barreto and Burton 2012; Bar-Yaacov et al. 2012) and
443 accelerated mitochondrial evolution in animals has led to compensatory, rapid evolution of
444 ribosomal proteins (Osada and Akashi 2012; Barreto and Burton 2013; Bar-Yaacov et al. 2012).
445 Rapid mitochondrial diversification within *Peromyscus* (Riddle et al. 2000; Bradley et al. 2007;
446 Platt et al. 2015), coincident with the ecological radiation of this genus (Lindsey 2020), suggests
447 that equivalent, rapid selection on ribosomal proteins may be a key evolutionary innovation that
448 enabled Peromyscine rodents to successfully and quickly adapt to varied thermal environments.
449 Comparisons among additional desert- and non-desert-adapted *Peromyscus* species will be
450 necessary to test this hypothesis within an evolutionary framework. Although not unique to
451 desert-species, rapid ribosomal evolution may provide a common evolutionary mechanism to
452 respond to anthropogenic climate change.

453 Evaporative cooling through sweating, panting, or salivating increases water loss and
454 challenges osmoregulatory homeostasis while maintaining thermoregulation in a hot and dry
455 climate (McKinley et al. 2018). Thermal stress exacerbates dehydration by increasing
456 evaporative water loss and if untreated, can result in cognitive dysfunction, motor impairment,
457 and eventually death. In consequence, osmoregulatory mechanisms are often under selection in
458 extreme thermal environments (MacManes and Eisen 2014; Marra et al. 2014). Consistently,
459 40% of the 10 genes targeted by selective sweeps and shared between desert adapted
460 *Peromyscus* are involved in ion balance (Table 3). Proteins Trypsin-2 (*TRY2*) and Trimeric
461 intracellular cation channel type B (*TM38B*) are associated with selective sweeps in both desert-
462 adapted species and involved in calcium ion (Ca^{2+}) binding and release, respectively. DNA-
463 directed RNA polymerase III (*RPC1*) has also experienced a significant selective sweep in both
464 desert species and influences magnesium (Mg^{2+}) binding. Calcium and magnesium cations are
465 among those essential for osmoregulation (also, Na^{+}, K^{+}, Cl^{-}, HCO_3^{-}; Stockham and Scott 2008)
466 and parallel selection on these genes is consistent with the hypothesis that solute-carrier

proteins are essential to maintaining homeostasis in desert-specialized rodents (Marra et al. 2014; Kordonowy and MacManes 2017). Additional genes implicated in osmoregulation were identified as experiencing selective sweeps only in *P. crinitus* (Table 2; Table S16). In addition to those shared with *P. eremicus,* Prostatin (*PRSS8*), only under selection in *P. crinitus,* is critically responsible for increasing the activity of epithelial sodium (Na^+) channels, which mediate Na^+ reabsorption through the kidneys (Narikiyo et al. 2002). Two more genes associated with Ca^{2+} regulation (Anionic Trypsin-2 [*PRSS2*] and Trypsin [*TRYP*]) and other genes regulating zinc (*KLK4*) and iron (*NCOA4*) were also identified as targets of selective sweeps exclusively in *P. crinitus.*

Metabolic tuning: proteins-for-water or lipids-for-torpor?

Hot deserts experience dramatic fluctuations in both food and water availability that challenge species survival (Noy-Meir 1973; Silanikove 1994). Mammals accommodate high temperatures by increasing their body temperature, up to a point, and cold temperatures by aerobic thermogenesis or metabolic suppression and the initiation of torpor or hibernation (Levesque et al. 2016). When resources are scarce, metabolism relies exclusively on endogenous nutrients; carbohydrates (e.g., sugars, glucose) are consumed immediately, then lipids, and eventually, proteins. Protein oxidation has a low energy return relative to lipid catabolism (Bar and Volkoff 2012), but yields five times more metabolic water (Jenni and Jenni-Eiermann 1998; Gerson and Guglielmo 2011a, b; McCue et al. 2017). In a low-water, desert environment an early shift to protein catabolism during periods of resource limitation may represent an important source of water for desert species (e.g., protein-for-water hypothesis; Mosin 1984; Jenni and Jenni-Eiermann, 1998; Gerson and Guglielmo, 2011a, b), as demonstrated for migrating birds. Consistent with this hypothesis, we identified numerous candidate genes that experienced selective sweeps in *P. crinitus* and that are involved in the detection of metabolic-stress and shifts in metabolic fuel consumption. For example, the gene eIF-2-alpha kinase GCN2 (*E2AK4*),

493 which is responsible for sensing metabolic stress in mammals and required for adaptation to
494 amino acid starvation, experienced the strongest selective sweep on chromosome 4 in *P.*
495 *crinitus* and (Fig. 3; Harding et al. 2003; Baker et al. 2012; Taniuchi et al. 2016). Numerous
496 candidate genes involved in oxidation (Oxidoreductase NAD-binding domain-containing protein
497 1 [*Oxnad1*]), fat catabolism (Kallikrein-6 [*KLK6*]), protein processing (Kallikrein-13 [*KLK13*]), and
498 proteolysis (Kallikrein [*KLK4*, *KLK13*], Trypsin [*PRSS2, TRYP, TRY2*], Chymotrypsin-like
499 elastase family member 2A [*CELA2A*]) were significantly enriched in *P. crinitus*, with proteolysis
500 as the most enriched functional group (Fig. 4; Table S16), potentially supporting the protein-for-
501 water hypothesis.

502 For desert species, including both desert specialists examined here (Morhardt and
503 Hudson 1966; MacMillen 1983), heat- and drought-induced torpor enable long duration, low
504 energy survival. Lipid acquisition and storage are critical to the initiation and maintenance of
505 torpor (Buck et al. 2002; Melvin and Andrews 2009). Significant weight loss in experimentally-
506 dehydrated *P. eremicus* and enhanced thermogenic performance of high-altitude-adapted deer
507 mice have been associated with enhanced lipid metabolism (Cheviron et al. 2012; Kordonowy et
508 al. 2016). At high-latitudes, increased lipid oxidation enables aerobic thermogenesis, but in hot
509 deserts lipids may represent a valuable energy source in a food-scarce environment (e.g.,
510 lipids-for-torpor hypothesis).

511 Metabolic processes were enriched in *P. crinitus*, but not *P. eremicus.* Two additional
512 candidate genes DCC-interacting protein 13-alpha and -beta (*APPL1*, *APPL2*), experienced a
513 significant selective sweep in *P. crinitus* and are important in glucose regulation, insulin
514 response, and fatty acid oxidation, reinforcing the hypothesis that enhanced lipid oxidation may
515 be critical to thermoregulatory responses. Laboratory manipulations of *APPL1* demonstrate
516 protection against high-fat diet-induced cardiomyopathy in rodents (Park et al. 2013) and *APPL2*
517 is responsible for dietary regulation, cold-induced thermogenesis, and cold acclimation
518 (uniprot.org). Together, these genes play a role in both obesity and dietary regulation. Both

519 *APPL* genes are associated with obesity and non-alcoholic fatty liver disease and their sweep
520 signature in *P. crinitus* may have relevant connections to biomedical research that remain to be
521 explored (Jiang et al. 2012; Barbieri et al. 2013). Physiological tests will be essential to
522 determine whether desert-adapted deer mice prioritize proteins or fats during periods of
523 resource limitation (e.g., lipids-for-torpor) or extreme dehydration (e.g., protein-for-water
524 hypothesis).

525 Molecular rewiring of metabolic processes in response to environmental conditions has
526 been documented in a number of species (e.g., mammals, Cheviron et al. 2012, Velotta et al.
527 2020; birds, Xie et al. 2018; fruit flies, Mallard et al. 2018). Nonetheless, expression changes
528 can also impact metabolism (Cheviron et al. 2012; Storz and Cheviron 2016). The capacity for
529 rapid adaptation to distinct thermal environments through multiple evolutionary mechanisms,
530 combined with thermoregulatory behavioral fine-tuning (e.g., nocturnality, aestivation, food
531 caching, burrowing, dietary shifts), suggests there may be many more evolutionary responses
532 available for small mammals to accommodate changing climates than previously realized.
533 Therein, metabolism and metabolic plasticity represents a fundamental phenotype for
534 anticipating species survival under altered climate scenarios.

535

536 *Different evolutionary trajectories, same result*

537 There are multiple evolutionary pathways to achieve environmental adaptation, most notably
538 through genomic changes in protein coding genes or transcriptional regulation of gene
539 expression. Although patterns of gene expression remain to be explored for *P. crinitus*, our
540 comparative genomic investigation suggests alternative evolutionary strategies for each desert-
541 adapted *Peromyscus* species shaped by their demographic histories: *P. crinitus* primarily
542 through changes in protein-coding genes and *P. eremicus* primarily through transcriptional
543 regulation.

Diverse functional enrichment of the *P. crinitus* genome (Fig. 4), spanning metabolic and osmoregulatory functions in addition to the general functional enrichment of ribosomes, points to polygenic adaptation to desert life. Here, we have identified a number of candidate loci worthy of detailed examination across populations and in a laboratory setting to understand the varied influence of different loci in thermoregulation and dehydration tolerance. In contrast, evidence of many significant sweep sites in the *P. eremicus* genome, generally located more distant from protein coding genes, and with functional enrichment restricted exclusively to ribosomes, suggests local adaptation in this species may have been driven more by selection on regulatory or non-coding regions of the genome that disproportionately impact gene expression. Indeed, transcriptomic investigations have identified significant expression changes implicated in osmoregulation in *P. eremicus* (MacManes and Eisen 2014; Kordonowy and MacManes 2017) and in thermoregulation in other *Peromyscus* species and rodents (Cheviron et al. 2012; Marra et al. 2014; Storz and Cheviron 2016). Transcriptional regulation is a particularly useful mechanism for environmental acclimation, as these changes are more transient relative to genomic changes and can enhance phenotypic flexibility (Garrett and Rosenthal 2012; Rieder et al. 2015; Liscovitch-Brauer et al. 2017). For example, transcription factors (TFs), which are a common target of selection located outside of protein-coding genes, can coordinate the expression of many genes, which allows multiple phenotypic changes to occur simultaneously (Wagner and Lynch 2008). Reduced genomic variation is expected near selective sweeps and can encompass tens to thousands of adjacent nucleotides depending on recombination and the strength of selection (Fay and Wu 2000; Carlson et al. 2005), yet counter expectation, Tajima's D and nucleotide diversity for regions flanking putative selective sweeps were significantly higher than the global average for most comparisons (Table S18).

Placing the results of selective sweep analyses within an evolutionary framework is critical to interpreting varied evolutionary responses to similar environments. The expansion of North American deserts following the conclusion of the last glacial maximum (LGM, ~11 Kya;

570 Pavlik et al. 2008) constrains the evolutionary and adaptive timescales of contemporary desert
571 species. Assuming simultaneous colonization of southwestern deserts, the stable but low
572 effective population size of *P. eremicus* suggests two hypotheses: (I) selection has removed
573 variation in this species over time (Murray et al. 2017) or (II) *P. eremicus* historically harbored
574 less genetic variation for selection to act on, despite equivalent contemporary diversity relative
575 to *P. crinitus* (Table S18). As a consequence of demography, the evolution of *P. eremicus* is
576 likely to have been more impacted by genetic drift (Fig. 2; Allendorf 1986; Masel 2011), so we
577 expect genomic evolution to be slower in *P. eremicus* relative to *P. crinitus*, which historically
578 has a larger effective population size and broader pool of variation for selection to act upon.
579 Within this context, environmental adaptation could be achieved more efficiently for *P. eremicus*
580 through transcriptomic plasticity or changes in regulatory elements (Allendorf 1986; Neme and
581 Tautz 2016; Mallard et al. 2018). In contrast, the larger historical effective population size of *P.*
582 *crinitus* is more conducive to the maintenance of genetic diversity and rapid evolution of protein
583 coding sequences through mutational stochasticity, reduced impact of genetic drift, and
584 potentially, gene flow. *Peromyscus crinitus* experienced a historical demographic bottleneck
585 prior to the formation of North American deserts; Nevertheless, the recovered effective
586 population size of *P. crinitus* is much larger than *P. eremicus* and consistent with low levels of
587 detected admixture (Fig. S7, Table S8). Also consistent with a history of admixture, *P. crinitus*
588 has the least negative Tajima's D, while low nucleotide diversity may indicate that admixture
589 may not be recent. Repeated growth and contraction of rivers in the American Southwest during
590 Pleistocene glacial-interglacial cycles (0.7-0.01 Mya; Muhs et al. 2003; Van Dam and Matzke
591 2016) would have provided iterative opportunities for connectivity and introgression between
592 incompletely-isolated *Peromyscus* species. Historical hybridization between *P. crinitus* and one
593 or more other Peromyscine species, likely unsampled here, may have accelerated adaptation in
594 *P. crinitus* through the rapid influx of novel mutation combinations through adaptive
595 introgression, a hypothesis that warrants further investigation through expanded sampling and

596 tests of adaptive introgression. Low-coverage whole-genome resequencing is optimal for
597 comparative and population genomic investigations (O'Rawe et al. 2015; da Fonseca et al.
598 2016), but detailed analyses of historical introgression are limited and we look forward to testing
599 this hypothesis with expanded population sampling and increased sequencing depth. Overall,
600 incorporating an evolutionary perspective into the interpretation of selection patterns has
601 important implications for anticipating species responses to anthropogenic climate change, as
602 historical demography and gene flow, in addition to selection, are responsible for shaping
603 contemporary diversity.
604
605 *Conclusion*
606 Contrasting patterns of selective sweeps and evolutionary histories between different species
607 experiencing similar environmental pressures can provide powerful insights into the adaptive
608 potential of species. We used comparative and population genomic analyses of three
609 *Peromyscus* species to identify candidate loci that may underlie adaptations to desert
610 environments in North America that serve to inform future investigations focused on predicting
611 potential for adaptation and identifying the causes of warming-related population declines (Cahill
612 et al. 2013). The identification of numerous targets of selection within *P. crinitus* highlights
613 multiple molecular mechanisms (metabolic switching, osmoregulatory tuning) associated with
614 physiological responses to deserts that warrant further investigation. Even in species recently
615 adapted to similar environments we identified divergent evolutionary trajectories, with one
616 species accommodating desert conditions primarily through genomic changes within protein
617 coding genes and the other, through transcriptional regulation mediated by historical
618 demographic processes. Our approach demonstrates the importance of placing genomic
619 selection analyses into an evolutionary framework to anticipate evolutionary responses to
620 change.
621

AUTHOR CONTRIBUTIONS

JPC performed analyses and wrote the first version of the paper. AT collected whole genome resequencing data, designed the bioinformatic pipeline. MDM conceptualized the project, performed *de novo* genome assembly and annotation, and funded data generation. OD, ADO, RK, IB, and ELA generated and assembled Hi-C data for *P. crinitus* as part of the DNA Zoo consortium effort. All authors reviewed and edited the manuscript.

ACKNOWLEDGEMENTS

We thank A. S. Westbrook for computational support; the Premise computing cluster at the University of New Hampshire, where all analyses were conducted; the Biotechnology Resource Center at Cornell University for preparation of whole-genome resequencing libraries; Christopher Tracy for access to the Boyd Deep Canyon Reserve; Jim Patton for desert field expertise; Sen Pathak, Asha Multani, Richard Behringer for providing the fibroblast samples from the T.C. Hsu Cryo-Zoo at the University of Texas M.D. Anderson Cancer Center; DNA Zoo for generating Hi-C data; Pawsey Supercomputing Centre with funding from the Australian Government and the Government of Western Australia for computational support of the DNA Zoo effort and the Museum of Southwestern Biology at the University of New Mexico for loaned tissue materials. This work was funded by the National Institute of Health National Institute of General Medical Sciences to MDM (1R35GM128843).

DATA ACCESSIBILITY STATEMENT

The draft assembly data are housed on the European Nucleotide Archive (ENA) under project ID PRJEB33592. The Hi-C data is available on SRA (SRX7041777, SRX7041776, SRX7041773) under the DNA Zoo project accession PRJNA512907. The *P. crinitus* genome assembly is available at https://www.dnazoo.org/assemblies/Peromyscus_crinitus. Whole-genome resequencing data for *P. crinitus* are available on ENA under project ID PRJEB35488.

Custom python scripts and other bash scripts used in analysis are available at:

https://github.com/jpcolella/Peromyscus_crinitus.

LITERATURE CITED

Abbott, K. (1971). Water economy of the canyon mouse, *Peromyscus crinitus stephensi*. *Comparative Biochemistry and Physiology, 38A*, 37–52.

Allendorf, F. (1986). Genetic drift and the loss of alleles versus heterozygosity. *Zoo Biology, 5*(2), 181–190.

Anderson, K., & Jetz, W. (2005). The broad-scale ecology of energy expenditure of endotherms. *Ecology Letters, 8*(3), 310–318.

Arendt, J., & Reznick, D. (2008). Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends in Ecology & Evolution, 23*(1), 25–32.

Baker, B., Nargund, A., Sun, T., & Haynes, C. (2012). Protective coupling of mitochondrial function and protein synthesis via the eIF2α kinase GCN-2. *PLoS Genetics, 8*(6), e1002760.

Bar, N., & Volkoff, H. (2012). Adaptation of the physiological, endocrine and digestive system functions to prolonged food deprivation in fish. In Comparative physiology of fasting, starvation, and food limitation (pp. 69–90). New York: Springer.

Bar-Yaacov, D., Blumberg, A., & Mishmar, D. (2012). Mitochondrial-nuclear co-evolution and its effects on OXPHOS activity and regulation. *Biochimica et Biophsyica Acta (BBA)-Gene Regulatory Mechanisms, 1819*(9–10), 1107–1111.

Barbieri, M., Esposito, A., Angellotti, E., Rizzo, M., Marfella, R., & Paolisso, G. (2013). Association of genetic variation in adaptor protein APPL1/APPL2 loci with non-alcoholic fatty liver disease. *PLOS ONE, 8*(8), e71391.

Barreto, F., & Burton, R. (2013). Evidence for compensatory evolution of ribosomal proteins in response to rapid divergence of mitochondrial rRNA. *Molecular Biology and Evolution, 30*(2), 310–314.

Bassham, S., Catchen, J., Lescak, E., von Hippel, F., & Cresko, W. (2018). Repeated selection of alternatively adapted haplotypes creates sweeping genomic remodeling in stickleback. *Genetics, 209*(3), 921–939.

Bedford, M., & Hoekstra, H. (2015). The natural history of model organisms: *Peromyscus* mice as a model for studying natural variation. *Elife, 4*, e06813.

Bi, K., Linderoth, T., Singhal, S., Vanderpool, D., Patton, J., Nielsen, R., … Good, J. (2015). Temporal genomic contrasts reveal rapid evolutionary responses in an alpine mammal during recent climate change. *PLoS Genetics, 15*(5), e1008119.

Booker, T., Jackson, B., & Keightley, P. (2017). Detecting positive selection in the genome. *BMC Biology, 15*(98). doi: https://doi.org/10.1186/s12915-017-0434-y

Bradley, R., Durish, N., Rogers, D., Miller, J., Engstrom, M., & Kilpatrick, C. (2007). Toward a molecular phylogeny for *Peromyscus*: Evidence from mitochondrial cytochrome-b sequences. *Journal of Mammalogy, 88*(5), 1146–1159.

Buck, M., Squire, T., & Andrews, M. (2002). Coordinate expression of PDK4 gene: a means of regulating fuel selection in a hibernating mammal. *Physiological Genomics, 8*(1), 5–13.

Burg, M., Ferraris, J., & Dmitrieva, N. (2007). Cellular response to hyperosmotic stresses. *Physiological Reviews, 87*(4), 1441–1474.

Cahill, A., Aiello-Lammens, M., Fisher-Reid, M., Hua, X., Karanewsky, C., Yeong Ryu, H., … Wiens, J. (2013). How does climate change cause extinction? *Proceedings of the Royal Society B: Biological Sciences, 280*(1750), 20121890.

Campbell, M., Holt, C., Moore, B., & Yandell, M. (2014). Genome annotation and curation using MAKER and MAKER-P. *Current Protocols in Bioinformatics, 48*(1), 4–11.

Carlson, C., Thomas, D., Eberle, M., Swanson, J., Livingston, R., Rieder, M., & Nickerson, D. (2005). Genomic regions exhibiting positive selection identified from dense genotype data. *Genome Research, 15*(11), 1553–1565.

Charlesworth, B. (2009). Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics, 10*(3), 195–205.

Chen, H., & Boutros, P. (2011). VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics, 12*(1), 35

Chen, I., Hill, J., Ohlemueller, R., Roy, D., & Thomas, C. (2011). Rapid range shifts of species associated with high levels of climate warming. *Science, 333*(6045), 1024–1026.

Chen, S., Zhou, Y., Chen, Y., & Gu, H. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics, 34*(17), i884–i890.

Cheviron, Z., Bachman, G., Connaty, A., McClelland, G., & Storz, J. (2012). Regulatory changes contribute to the adaptive enhancement of thermogenic capacity in high-altitude deer mice. *Proceedings of the National Academy of Sciences, 109*(22), 8635–8640.

Cheviron, Z., Connaty, A., McClelland, G., & Storz, J. (2014). Functional genomics of adaptation to hypoxic cold-stress in high-altitude deer mice: transcriptomic plasticity and thermogenic performance. *Evolution, 68*(1), 48–62.

Cook, R., & Weisberg, S. (1984). Residuals and influence in Regression. Wiley.

Coyne, J. A., & Orr, H. A. (2004). Speciation. Sunderland, MA: Sinauer Associates.

da Fonseca, R. R., Albrechtsen, A., Themudo, G., Ramos-Madrigal, J., Sibbesen, J., Maretty, L., … Pereira, R. (2016). Next-generation biology: sequencing and data analysis approaches for non-model organisms. *Marine Genomics, 30*, 3–13. doi: 10.1016/j.margen.2016.04.012

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … 1000 Genome Project Data Process Subgroup. (2011). The variant call format and VCFtools. *Bioinformatics, 27*(15), 2156–2158.

Degen, A. (2012). Ecophysiology of small desert mammals. Springer Science & Business Media.

DeGiorgio, M., Huber, C., Hubisz, M., Hellmann, I., & Nielsen, R. (2016). SweepFinder2: increased sensitivity, robustness and flexibility. *Bioinformatics, 32*(12), 1895–1897.

Dewey, M., & Dawson, W. (2001). Deer mice: "the *Drosophila* of North American mammalogy. *Genesis, 29*(3), 105–109.

Dudchenko, O., Batra, S., Omer, A., Nyquist, S., Hoeger, M., Durand, N., … Aiden, E. (2017). De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science, 356*(6333), 92–95.

Dudchenko, O., Shamim, M., Batra, S., Durand, N., Musial, N., Mostofa, R., … Aiden, E. (2018). The Juicebox Assembly Tools module facilitates de novo assembly of mammalian genomes with chromosome-length scaffolds for under $1000. *BioRxiv.* doi: https://doi.org/10.1101/254797

Durand, N., Shamim, M., Machol, I., Rao, S., Huntley, M., Lander, E., & Aiden, E. (2016). Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Systematics, 3*(1), 95–98.

Emms, D. M., & Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology, 16*(1), 157.

Faust, G., & Hall, I. (2014). SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics, 30*(17), 2503–2505.

Fay, J., & Wu, C.-I. (2000). Hitchhiking under positive Darwinian selection. *Genetics, 155*(3), 1405–1413.

Freeman, B., Lee-Yaw, J., Sunday, J., & Hargreaves, A. (2018). Expanding, shifting and shrinking: The impact of global warming on species' elevational distributions. *Global Ecology and Biogeography, 27*(11), 1268–1276.

747 Fumagalli, M., Vieira, F., Linderoth, T., & Nielsen, R. (2014). ngsTools: methods for population
748 genetics analyses from next-generation sequencing data. *Bioinformatics, 30*(10), 1486–1487.
749 Garcia-Elfring, A., Barrett, R., & Millien, V. (2019). Genomic signatures of selection along a
750 climatic gradient in the northern range margin of the white-footed mouse (*Peromyscus*
751 *leucopus*). *Journal of Heredity, 110*(6), 684–695.
752 Garrett, S., & Rosenthal, J. (2012). RNA editing underlies temperature adaptation in K+
753 channels from polar octopuses. *Science, 335*(6070), 848–851.
754 Gerson, A., & Guglielmo, C. (2011a). Flight at low ambient humidity increases protein
755 catabolism in migratory birds. *Science, 333*(6048), 1434–1436.
756 Gerson, A., & Guglielmo, C. (2011b). House sparrows (*Passer domesticus*) increase protein
757 catabolism in response to water restriction. *American Journal of Physiology, 300*(4), R925–
758 R930.
759 Glazier, D. (1980). Ecological shifts and the evolution of geographically restricted species of
760 North American *Peromyscus* (mice). *Journal of Biogeography*, *7*(1), 63–83.
761 Han, M., Thomas, G., Lugo-Martinez, J., & Hahn, M. (2013). Estimating gene gain and loss
762 rates in the presence of error in genome assembly and annotation using CAFE 3*. Molecular*
763 *Biology and Evolution, 30*(8), 1987-1997.
764 Harding, H., Zhang, Y., Zeng, H., Novoa, I., Lu, P., Calfon, M., … Ron, D. (2003). An integrated
765 stress response regulates amino acid metabolism and resistance to oxidative stress.
766 *Molecular Cell, 11*(3), 619–633.
767 Hoffmann, A., & Sgro, C. (2011). Climate change and evolutionary adaptation. *Nature,*
768 *470*(7335), 479–485.
769 Hoffmann, A., & Willi, Y. (2008). Detecting genetic responses to environmental change. Nature
770 Reviews: *Genetics, 9*(6)*,* 421–432.
771 Hu, C., & Hoekstra, H. (2017). *Peromyscus* burrowing: A model system for behavioral evolution.
772 *Seminars in Cell and Developmental Biology, 61*, 107–114.
773 Huber, C., DeGiorgio, M., Hellmann, I., & Nielsen, R. (2016). Detecting recent selective sweeps
774 while controlling for mutation rate and background selection. *Molecular Ecology, 25*(1), 142–
775 156.
776 Issaian, T., Urity, V., Dantzler, W., & Pannabecker, T. (2012). Architecture of vasa recta in the
777 renal inner medulla of the desert rodent *Dipodomys merriami*: potential impact on the urine
778 concentrating mechanism. *American Journal of Physiology - Regulatory, Integrative and*
779 *Comparative Physiology, 303*(7), R748–R756.
780 Ivy, C., & Scott, G. (2017). Control of breathing and ventilatory acclimatization to hypoxia in
781 deer mice native to high altitudes. *Acta Physiologica, 221*(4), 266–282.
782 Jain, C., Koren, S., Dilthey, A., Phillippy, A. M., Aluru, S. (2018) A fast adaptive algorithm for
783 computing whole-genome homology maps. *Bioinformatics, 34*(17), i748-i756.
784 Jain, C., Dilthey A, Koren, S, Aluru, S., Phillippy, A. M. (2017) A fast approximate algorithm for
785 mapping long reads to large reference databases. In: Sahinalp S. (eds) *Research in*
786 *Computational Molecular Biology*. RECOMB 2017. Lecture Notes in Computer Science,
787 10229. Springer, Cham.
788 Jenni, L., & Jenni-Eiermann, S. (1998). Fuel supply and metabolic constraints in migrating birds.
789 *Journal of Avian Biology, 29*(4), 521–528.
790 Jiang, S., Fang, Q., Yu, W., Zhang, R., Hu, C., Dong, K., … Jia, W. (2012). Genetic variations in
791 APPL2 are associated with overweight and obesity in a Chinese population with normal
792 glucose tolerance. *BMC Medical Genetics, 13*(1), 22.
793 Johnson, D., & Armstrong, D. (1987). *Peromyscus crinitus*. *Mammalian Species, 287*, 1–8.
794 Jones, M., Mills, L., Alves, P., Callahan, C., Alves, J., Lafferty, D., … Good, J. (2018). Adaptive
795 introgression underlies polymorphic seasonal camouflage in snowshoe hares. *Science,*
796 *360*(6395), 1355–1358.

Kaseloo, P., Crowell, M., & Heideman, P. (2014). Heritable variation in reaction norms of metabolism and activity across temperatures in a wild-derived population of white-footed mice (*Peromyscus leucopus*). *Journal of Comparative Physiology B, 184*(4), 525–534.

Kim, Y., & Stephan, W. (2002). Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics, 160*(2), 765–777.

Kordonowy, L., & MacManes, M. (2017). Characterizing the reproductive transcriptomic correlates of acute dehydration in males in the desert-adapted rodent, *Peromyscus eremicus*. *BMC Genomics, 18*(1), 473.

Kordonowy, L., Lombardo, K., Green, H., Dawson, M., Bolton, E., LaCourse, S., & MacManes, M. (2016). Physiological and biochemical changes associated with acute experimental dehydration in the desert adapted mouse, *Peromyscus eremicus*. *Physiological Reports, 5*(6), e13218.

Korneliussen, T., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. *Bioinformatics, 15*(1), 356.

Kumar, S., & Subramanian, S. (2002). Mutation rates in mammalian genomes. Proceedings of the National Academy of Sciences, 99(2), 803–808.

Lamitina, T., Haung, C., & Strange, K. (2006). Genome-wide RNAi screening identifies protein damage as a regulator of osmoprotective gene expression. *Proceedings of the National Academy of Sciences of the United States of America, 103*(32), 12173–12178.

Levesque, D., Nowack, J., & Stawski, C. (2016). Modelling mammalian energetics: the heterothermy problem. *Climate Change Responses, 3*(1), 7.

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics, 26*(5), 589–595.

Li, Heng, & Durbin, R. (2011). Inference of human population history from whole genome sequence of a single individual. *Nature, 475*(7357), 493–496.

Li, H., Handsaker, B., Wysoker, A., Fennel, T., Ruan, J., Homer, N., … 1000 Genome Project Data Process Subgroup. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics, 25*(16), 2078–2079.

Lindsey, L. (2020). Utilizing genomic applications to examine patterns of diversification in deermice (Rodentia: Cricetidae: *Peromyscus*). Texas Tech Dissertation.

Liscovitch-Brauer, N., Alon, S., Porath, H., Elstein, B., Unger, R., Ziv, T., … Eisenberg, E. (2017). Trade-off between transcriptome plasticity and genome evolution in Cephalopods. *Cell, 169*(2), 191–202.

MacManes, M., & Eisen, M. (2014). Characterization of the transcriptome, nucleotide sequence polymorphism, and natural selection in the desert adapted mouse *Peromyscus eremicus. PeerJ, 2*, e642.

MacManes, M. (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.

MacMillen, R., & Christopher, E. (1975). The water relations of two populations of non-captive desert rodents. In Environmental physiology of desert organisms ((N. F. Hadley, ed.), pp. 117–137). Stroudsburg, Pennsylvania: Dowden, Hutchinson, and Ross.

MacMillen, R. (1972). Water economy of nocturnal desert rodents. *Symposia of the Zoological Society of London, 31*, 147–174.

MacMillen, R. (1983). Water regulation in *Peromyscus*. *Journal of Mammalogy, 64*(1), 38–47.

Mallard, F., Nolte, V., Tobler, R., Kapun, M., & Schlötterer, C. (2018). A simple genetic basis of adaptation to a novel thermal environment results in complex metabolic rewiring in *Drosophila*. *Genome Biology, 19*(119), 1-15. doi: 10.1186/s13059-018-1503-4

Marra, N., Romero, A., & DeWoody, J. (2014). Natural selection and the genetic basis of osmoregulation in heteromyid rodents as revealed by RNA-seq. *Molecular Ecology, 23*(11), 2699–2711.

Masel, J. (2011). Genetic drift. *Current Biology, 21*(20), R837–R838.

McCue, M., Sandoval, J., Beltran, J., & Gerson, A. (2017). Dehydration causes increased reliance on protein oxidation in mice: a test of the protein-for-water hypothesis in a mammal. *Physiological and Biochemical Zoology, 90*(3), 359–369.

McDonald, M., Gehrig, S., Meintjes, P., Zhang, X.-X., & Rainey, P. (2009). Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. IV. Genetic constraints guide evolutionary trajectories in a parallel adaptive radiation. *Genetics, 183*(3), 1041–1053.

McKinley, M., Martelli, D., Pennington, G., Trevaks, D., & McAllen, R. (2018). Integrating competing demands of osmoregulatory and thermoregulatory homeostasis. *Physiology, 33*(3), 170–181.

McNab, B. (1963). The influence of fat deposits on the basal rate of metabolism in desert homoiotherms. *Comparative Biochemistry and Physiology, 26*, 337–343.

McNab, B., & Morrison, P. (1963). Body temperature and metabolism in subspecies of Peromyscus from arid and mesic environments. *Ecological Monographs, 33*(1), 63–82.

Melvin, R., & Andrews, M. (2009). Torpor induction in mammals: Recent discoveries fueling new ideas. *Trends in Endocrinology and Metabolism, 20*(10), 490–498.

Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., & Thomas, P. (2017). PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Research, 45*(D1), F183-D189.

Millar, J. (1989). Reproduction and development. In Advances in the Study of *Peromyscus* (Rodentia) (eds Kirkland Gl Jr, Layne JN, pp. 169–232). Lubbock, Texas: University of Texas Press.

Morhardt, J., & Hudson, J. (1966). Daily torpor induced in white-footed mice (*Peromyscus* spp.) by starvation. *Nature, 212*(5066), 1046–1047.

Moritz, C., Patton, J., Conroy, C., Parra, J., White, G., & Beissinger, S. (2008). Impact of a century of climate change on small-mammal communities in Yosemite National Park, USA. *Science, 322*(5899), 261–264.

Morjan, C., & Rieseberg, L. (2004). How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. *Molecular Ecology, 13*(6), 1341–1356.

Mosin, A. (1984). On the energy fuel in voles during their starvation. *Comparative Biochemistry and Physiology, 77*(3), 563–565.

Muhs, D., Reynolds, R., Been, J., & Skipp, G. (2003). Eolian sand transport pathways in the southwestern United States: importance of the Colorado river and local sources. *Quaternary International, 104*(1), 3–18.

Murray, G. G. R., Soares, A. E. R., Novak, B. J., Schaefer, N. K., Cahill, J. A., Baker, A. J., … Shapiro, B. (2017) Natural selection shaped the rise and fall of passenger pigeon genomic diversity. *Science 17*(358), 951-954.

Narikiyo, T., Kitamura, K., Adachi, M., Miyoshi, T., Iwashita, K., Shiraishi, N., … Tomita, K. (2002). Regulation of prostasin by aldosterone in the kidney. *The Journal of Clinical Investigation, 109*(3), 401–408.

Natarajan, C., Hoffmann, F., Lanier, H., Wolf, C., Cheviron, Z., Spangler, M., … Storz, J. (2015). Intraspecific polymorphism, interspecific divergence, and the origins of function-altering mutations in deer mouse hemoglobin. *Molecular Biology and Evolution, 32*(4), 978–997.

Neme, R., & Tautz, D. (2016). Fast turnover of genome transcription across evolutionary time exposes entire non-coding DNA to de novo gene emergence. *ELife, 5*, e09977.

Nielsen, R., Williamson, S., Kim, Y., Hubisz, M., Clark, A., & Bustamante, C. (2005). Genomic scans for selective sweeps using SNP data. *Genome Research, 15*(11), 1566–1575.

Noy-Meir, I. (1973). Desert ecosystems: Environment and producers. *Annual Review of Ecology and Systematics, 4*(1), 25–51.

O'Rawe, J., Ferson, S., & Lyon, G. (2015). Accounting for uncertainty in DNA sequencing data. *Trends in Genetics, 31*(2), 61–66.

Osada, N., & Akashi, H. (2012). Mitochondrial-nuclear interactions accelerated compensatory evolution: evidence from the primate cytochrome C oxidase complex. *Molecular Biology and Evolution, 29*(1), 337.

Park, M., Wu, D., Park, T., Choi, C., Li, R.-K., Cheng, K., … Sweeney, G. (2013). APPL1 transgenic mice are protected from high-fat diet-induced cardiac dysfunction. *American Journal of Physiology: Endocrinology and Metabolism, 305*(7), E795–E804.

Pavlik, B. M. (2008). The California deserts: An ecological rediscovery. Berkeley: University of California Press.

Pergams, O., & Lacy, R. (2008). Rapid morphological and genetic change in Chicago-area *Peromyscus. Molecular Ecology, 17*(1), 450–463.

Pierce, S., & Vogt, F. (1993). Winter acclimatization in *Peromyscus maniculatus gracilis P leucopus noveboracensis,* and *P. l. leucopus*. *Journal of Mammalogy, 74*(3), 665–677.

Platt, II, R., Amman, B., Keith, M., Thompson, C., & Bradley, R. (2015). What is *Peromyscus*? Evidence from nuclear and mitochondrial DNA sequences suggests the need for a new classification. *Journal of Mammalogy, 96*(4), 708–719.

Porcelli, D., Butlin, R., Gaston, K., Joly, D., & Snook, R. (2015). The environmental genomics of metazoan thermal adaptation. *Heredity, 114*(5), 502–514.

R Core Team. (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Riddle, B., Hafner, D., & Alexander, L. (2000). Phylogeography and systematics of the Peromyscus eremicus species group and the historical biogeography of North American warm regional deserts. *Molecular Phylogenetics and Evolution, 17*(2), 145–160.

Rieder, L., Savva, Y., Reyna, M., Chang, Y.-J., Dorsky, J., Rezaei, A., & Reenan, R. (2015). Dynamic response of RNA editing to temperature in *Drosophila*. *BMC Biology, 13*(1), 1.

Robinson, J., Turner, D., Durand, N., Thorvaldsdottir, H., Mesirov, J., & Aiden, E. (2018). Juicebox.js provides a cloud-based visualization system for Hi-C data. *Cell Systems, 6*(2), 256–258.

Rosenblum, E., Parent, C., & Brandt, E. (2014). The molecular basis of phenotypic convergence. *Annual Review of Ecology and Systematics, 45*, 203–226.

Rundle, H., Nagel, L., Boughman, J., & Schluter, D. (2000). Natural selection and parallel speciation in sympatric sticklebacks. *Science, 287*(5451), 306–308.

Schwimmer, H., & Haim, A. (2009). Physiological adaptations of small mammals to desert ecosystems. *Integrative Zoology, 4*(4), 357–366.

Shen, L. (2016). GeneOverlap: R package for testing and visualizing gene overlaps. New York City, New York: Ichan School of Medicine at Mount Sinai.

Sikes, R. S., & The Animal Care and Use Committee of the American Society of Mammalogists. (2016). 2016 Guidelines of the American Society of Mammalogists for the use of wild mammals in research and education. *Journal of Mammalogy, 97*(3), 663–688.

Silanikove, N. (1994). The struggle to maintain hydration and osmoregulation in animals experiencing severe dehydration and rapid rehydration: the story of ruminants. *Experimental Physiology, 79*(3), 281–300.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics, 31*(19), 32-10–3212.

Simpson, G. (1961). Principles of Animal Taxonomy. New York: Columbia University Press.

Simpson, J., Wong, K., Jackman, S., Schein, J., Jones, S., & Birol, I. (2009). ABySS: a parallel assembler for short read sequence data. *Genome Research, 19*(6), 1117–1123.

Skotte, L., Korneliussen, T., & Albrechtsen, A. (2013). Estimating individual admixture proportions from next generation sequencing data. *Genetics, 195*(3), 693–702.

949 Smalec, B., Heider, T., Flynn, B., & O'Neill, R. (2019). A centromere satellite concomitant with
950 extensive karyotypic diversity across the *Peromyscus* genus defies predictions of molecular
951 drive. Chromosome Research: An International Journal on the Molecular, Supramolecular
952 and Evolutionary Aspects of Chromosome Biology. doi: 10.1007/s10577-019-09605-1
953 Smit, A., Hubley, R., & Green, P. (2013). RepeatMasker Open-4.0. Available at:
954 http://www.repeatmasker.org/
955 Smith, J., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetical Research,*
956 *23*(1), 23–35.
957 Steppan, S., Adkins, R., & Anderson, J. (2004). Phylogeny and divergence-date estimates of
958 rapid radiations in Muroid rodents based on multiple nuclear genes. *Systematic Biology,*
959 *53*(4), 533–553.
960 Stockham, S., & Scott, M. (2008). Fundamental of veterinary clinical pathology. Ames, IA:
961 Wiley-Blackwell.
962 Storz, J., & Cheviron, Z. (2016). Functional genomic insights into regulatory mechanisms of
963 high-altitude adaptation. *Advances in Experimental Medicine and Biology, 903*, 113–128. doi:
964 10.1007/978-1-4988-7678-9_8.
965 Storz, J., & Kelly, J. (2008). Effects of spatially varying selection on nucleotide diversity and
966 linkage disequilibrium: insights from deer mouse globin genes. Genetics, 180, 367–379.
967 Storz, J. (2007). Hemoglobin function and physiological adaptation to hypoxia in high-altitude
968 mammals. *Journal of Mammalogy, 88*(1), 24–31.
969 Storz, J., Runck, A., Moriyama, H., Weber, R., & Fago, A. (2010). Genetic differences in
970 hemoglobin function between highland and lowland deer mice. *Journal of Experimental*
971 *Biology, 213*(15), 2565–2574.
972 Storz, J., Cheviron, Z., McClelland, G., & Scott, G. (2019). Evolution of physiological
973 performance capacities and environmental adaptation: insights from high-elevation deer mice
974 (*Peromyscus maniculatus*). *Journal of Mammalogy, 100*(3), 910–922.
975 Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. (2011). REVIGO summarizes and visualizes
976 long lists of gene ontology terms. *PloS One, 6*(7), e21800.
977 Taniuchi, S., Miyake, M., Tsugawa, K., Oyadormari, M., & Oyadomari, S. (2016). Integrated
978 stress response of vertebrates is regulated by four eIF2α kinases. *Scientific Reports, 6*,
979 32886.
980 Tigano, A., Colella, J., & MacManes, M. (2020). Comparative and population genomics
981 approaches reveal the basis of adaptation to deserts in a small rodent. *Molecular Ecology,*
982 *29*(7), 1300–1314.
983 Tigano, A. & Friesen, V. L. (2016). Genomics of local adaptation with gene flow. *Molecular*
984 *Ecology, 25*, 2144-2164.
985 Tingley, M., & Beissinger, S. (2013). Cryptic loss of montane avian richness and high
986 community turnover over 100 years. *Ecology, 94*(3), 598–609.
987 Urban, M. (2015). Accelerating extinction risk from climate change. *Science, 348*(6234), 571–
988 573.
989 Van Dam, M., & Matzke, N. (2016). Evaluating the influence of connectivity and distance on
990 biogeographical patterns in the south-western deserts of North America. *Journal of*
991 *Biogeography, 43*(8), 1514–1532.
992 Veal, R., & Caire, W. (1979). *Peromyscus eremicus*. *Mammalian Species, 118*(8), 1–6.
993 Velotta, J., Robertson, C., Schweizer, R., McClelland, G., & Cheviron, Z. (2020). Adaptive shifts
994 in gene regulation underlie a developmental delay in thermogenesis in high-altitude deer
995 mice. *Molecular Biology and Evolution*, msaa086. doi:
996 https://doi.org/10.1093/molbev/msaa086
997 Wagner, G., & Lynch, V. (2008). The gene regulatory logic of transcription factor evolution.
998 *Trends in Ecology & Evolution, 23*(7), 277–385.

Weigand, H., & Leese, F. (2018). Detecting signatures of positive selection in non-model species using genomic data. *Zoological Journal of the Linnean Society, 184*(2), 528–583.

Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M., & Jaffe, D. B. (2017). Direct determination of diploid genome sequences. *Genome Research, 27*(5), 757–767.

Wichman, H., & Lynch, C. (1991). Genetic variation for seasonal adaptation in *Peromyscus leucopus*: nonreciprocal breakdown in a population cross. *Journal of Heredity, 82*(3), 197–204.

Williams, T., & Kelley, C. (2010). Gnuplot 4.4: an interactive plotting program (Version 4.4). Retrieved from http://gnuplot.sourceforge.net/

Williams, D. (1987). Generalized linear model diagnostics using the deviance and single case deletions*. Applied Statistics, 36*(2), 181–191.

Wood, T., Burke, J., & Rieseberg, L. (2005). Parallel genotypic adaptation: when evolution repeats itself. *Genetica, 123*(1–2), 157–170.

Xie, S., Yang, X., Wang, D., Zhu, F., Yang, N., Hou, Z., & Ning, Z. (2018). Thyroid transcriptome analysis reveals different adaptive responses to cold environmental conditions between two chicken breeds. PLOS ONE, 13(1), 2018.

TABLES

Table 1. Assembly stats, genome size, and global Tajima's D and pi (1 kb windows) for each *Peromyscus* species.

Species	N	Scaffold N50	Contig N50	Size (Gb)[a]	Size (Gb)[b]	Taj. D	π
P. crinitus	9	94,816,992	204,461	2.27	2.28	-0.69	0.005
P eremicus	26	119,957,392	76,024	2.45	2.54	-1.27	0.007
P. maniculatus	5	115,033,041	42,400	2.33	2.44	-1.62	0.012

[a] abyss-fac estimate
[b] assemblathon estimate

Table 2. Significant sweep sites localized within protein-coding genes for each *Peromyscus* species. Species (Spp.), chromosome (Chr.) sweep position (Pos.), gene name, protein, general function (based on UniProt database: uniprot.org), and direction (Dir.) of gene transcription. Abbreviation definitions: thermogen./thermogenesis, neg./negative, pos./positive, reg./regulation, IGF/insulin-like growth factor, ISR/Integrated Stress Response. *no gene name alternative available.

Spp.	Chr.	Pos.	Gene	Protein	General function	Dir.
P. maniculatus	4	145409180	*ZNF217*	Zinc finger protein 217	DNA-binding transcription factor, transcription regulation, zinc binding	-
	20	36260251	*DHRS7B*	Dehydrogenase/reductase SDR family member 7B	Oxidoreductase activity	-
P. eremicus	1	42451454	*Ush1c*	Harmonin	Mechanotransduction in cochlear hair cells	-
	8	67956	*MEIOB*	Meiosis-specific with OB domain-containing protein	Meiosis	+
P. crinitus	3	52113514	*PRSS2*	Trypsin-2	Calcium ion binding	+
			KLK13	Kallikrein-13	Protein processing, proteolysis, reg. of IGF	+
			PRSS8	Prostasin	Sodium balance	+
			KLK4	Kallikrein-4	Zinc ion binding, proteolysis	+
			PRTN3	Myeloblastin	Degrades collagen (I, III, IV), elastin, fibronectin, laminin, vitronectin; blood coagulation, immune response	+
			KLK14	Kallikrein-14	Varied (epidermis morphogenesis)	+
			TRYP_PIG*	Trypsin	Calcium ion binding, proteolysis	+
			KLK6	Kallikrein-6	Varied (collagen catabolism, tissue regen)	+
			CELA2A	Chymotrypsin-like elastase family member 2A	Cleavage and elastin hydrolase, proteolysis	+
	4	57673659	*EIF2AK4*	eIF-2-alpha kinase GCN2	Metabolic stress sensing protein kinase, role in ISR required for adaptation to amino acid starvation, protein synthesis repression	-
	6	66203934	*Nes*	Nestin	Brain, eye development (neg. reg. catalytic activity)	-
	9	23303147	*DENND64*	DENND64	Endocytic recycling pathway component	+
		23323150	*DENND64*	DENND64	Endocytic recycling pathway component	+
		43305800	*Nynrin*	NYNRIN	Nucleic acid binding	+
		22243007	*Parg*	Poly(ADP-ribose) glycohydrolase	Prevent detrimental accumulation of poly(ADP-ribose) upon prolonged replicative stress	-
			Parg	Poly(ADP-ribose) glycohydrolase	Prevent detrimental accumulation of poly(ADP-ribose) upon prolonged replicative stress	-
		22283012	*NCOA4*	Nuclear receptor coactivator 4	Androgen receptor (iron ion homeostasis)	-
		22303015	*Oxnad1*	Oxidoreductase NAD-binding domain-containing protein 1	Oxidoreductase activity	-
	18	450151	*APPL2*	DCC-interacting protein 13-beta	Varied (cold acclimation, diet induced thermogen., glucose homeostasis, neg. reg. of insulin response/fatty acid oxidation/glucose import, pos. reg. of cold-induced thermogen.)	-
			APPL1	DCC-interacting protein 13-alpha	Varied. (insulin receptor signaling pathway, pos. reg. of glucose	-

				import)	
		APPL2	DCC-interacting protein 13-beta	Varied (cold acclimation, diet induced thermogen., glucose homeostasis, neg. reg. of insulin response/fatty acid oxidation/ glucose import, pos. reg. of cold-induced thermogen.)	-
	460153	*APPL1*	DCC-interacting protein 13-alpha	Varied (insulin receptor signaling pathway, pos. reg. of glucose import)	-
23	28065942	*Tctn1*	Tectonic-1	Neural development	+
		Tctn1	Tectonic-1	Neural development	+

Table 3. Functional annotation of proximal gene names and enriched GO (gene ontology) terms associated with significant selective sweeps and shared between desert-adapted *P. crinitus* and *P. eremicus*. * indicates genes or GO terms also shared with *P. maniculatus*

Data set	Dataset	Gene Name / GO term	Function	Protein/Class
Gene Names	I	none	-	-
	II	*BTD*	Hydrolase, biotin transport/metabolism	Biotinidase
		RL36	Ribosomal protein, translation	60S ribosomal protein L36
	III	*BTD*	Hydrolase, biotin transport/metabolism	Biotinidase
		ENV	Zn binding, virion attachment	Envelope glycoprotein
		H3X	DNA binidng, protein heterodimerization	Putative histone H3.X
		RL15	RNA binding, ribosome constituent, translation	60S ribosomal protein L15
		RL36	Ribosomal protein, translation	60S ribosomal protein L36
		RPC1	Zn/Mg binding, immune response	DNA-directed RNA polymerase III subunit RPC1
		*RS2**	Ribosomal protein, enzyme binding	40S ribosomal protein S2
		RS26	mRNA binding, ribosome, translation	40S ribosomal protein S26
		TM38B	Rapid Ca2+ release, K+ channel, ossification	Trimeric intracellular cation channel type B
		TRY2	Ca2+ binding, collagen catabolism, proteolysis, cell growth	Trypsin-2
Enriched GO terms	I	none	-	-
	II	GO:0005622	intracellular	cellular component
		GO:0005840*	ribosome	cellular component
		GO:0022626*	cystolic ribosome	cellular component
		GO:0043226	organelle	cellular component
		GO:0043229	intracellular organelle	cellular component
		GO:0044391*	ribosomal subunit	cellular component
		GO:1990904*	ribonucleoprotein complex	cellular component
	III	GO:0003735*	structural constituent of ribosome	molecular function
		GO:0022626*	cystolic ribosome	cellular component
		GO:0022625*	cystolic large ribosomal subunit	cellular component
		GO:0044391*	ribosomal subunit	cellular component
		GO:0005840*	ribosome	cellular component
		GO:0015934	large ribosomal subunit	cellular component
		GO:1990904*	ribonucleoprotein complex	cellular component

FIGURES & FIGURE LEGENDS

Figure 1. Geographic ranges of the three *Peromyscus* species examined in this study with major southwestern North American deserts denoted by diagonal hashing. *P. crinitus* range is in red, *P. eremicus* in blue, and *P. maniculatus* in yellow. Areas of sympatry denoted by color overlap: dark purple = yellow + red + blue and green = yellow + blue.

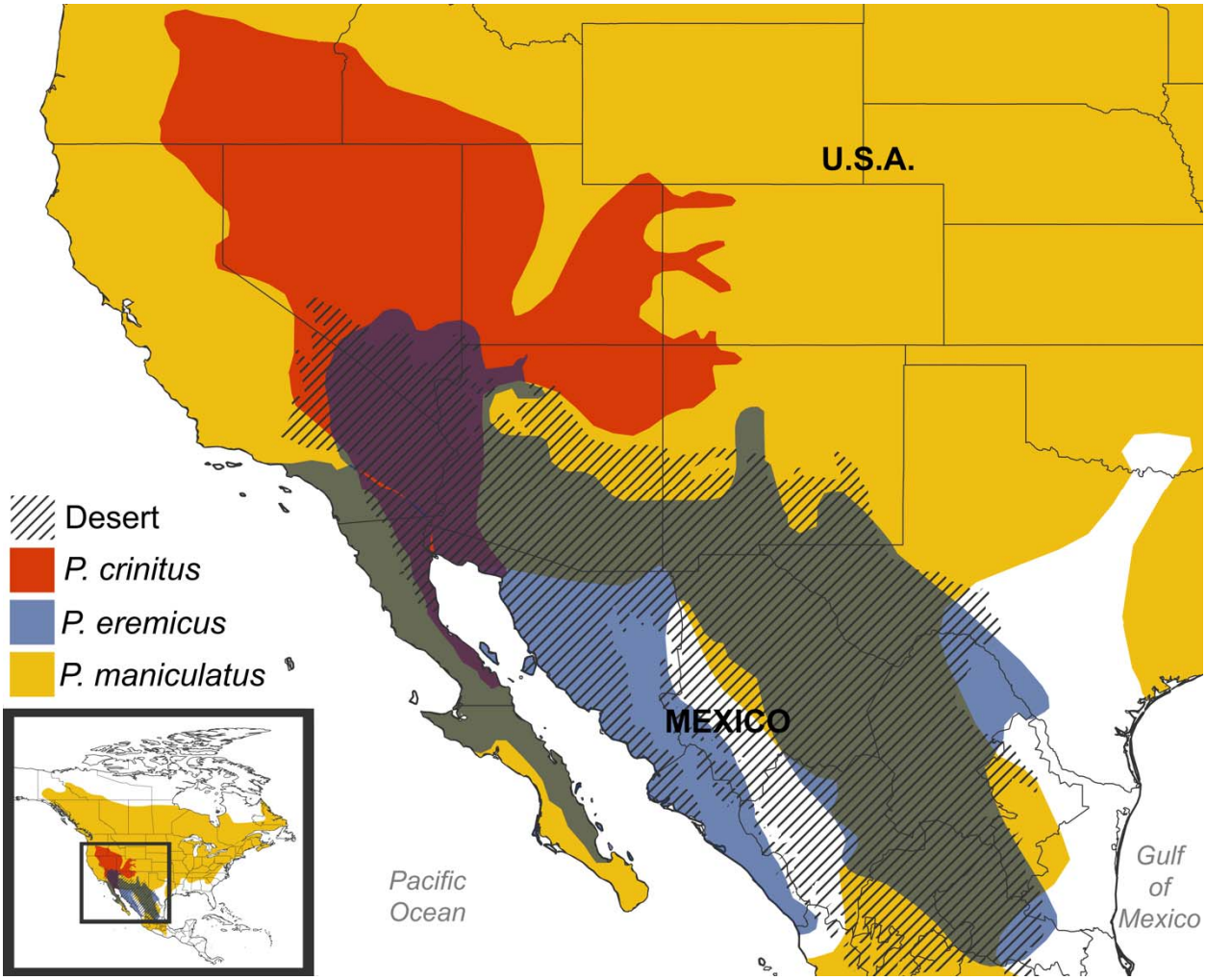

Figure 2. Distributions of effective population size (N_e) through time for *P. crinitus* (red), *P. eremicus* (blue), and *P. maniculatus* (yellow) based on a generation time of 6 months (0.5 years) and a general mammalian mutation rate of 2.2×10^{-9} substitutions/site/year. Note that the *P. maniculatus* genome was sequenced from a captive individual and therefore does not reflect natural populations trends of this species.

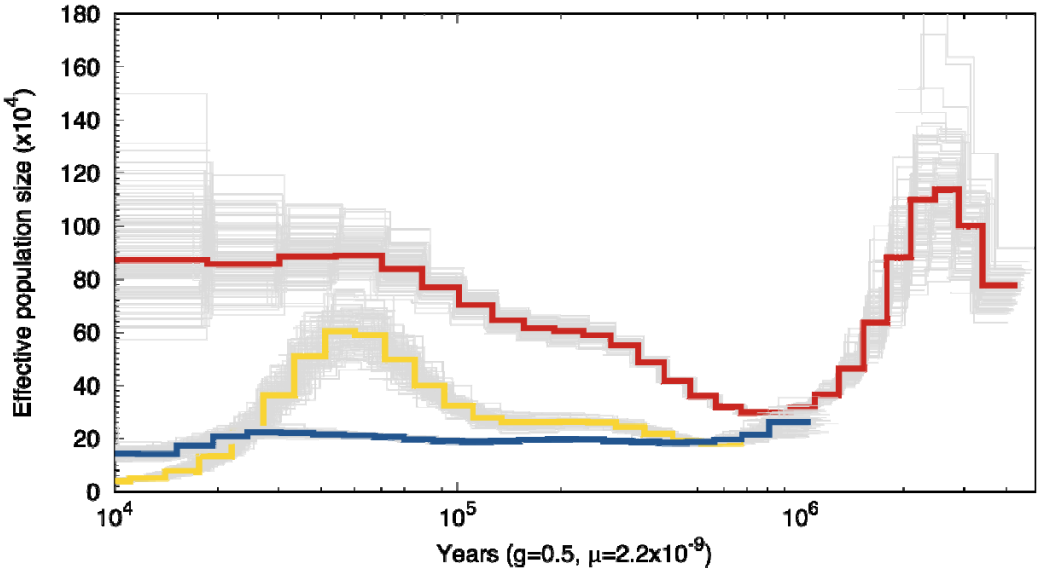

Figure 3. *REVIGO* plots of enriched functional groups for *P. crinitus* (top row) and *P. eremicus* (bottom row) based on functional annotation of the two nearest protein-coding genes to each site (dataset II) identified as the subject of a selective sweep. Darker colors indicate greater significance. MP = metabolic process, MB = membrane-bound.

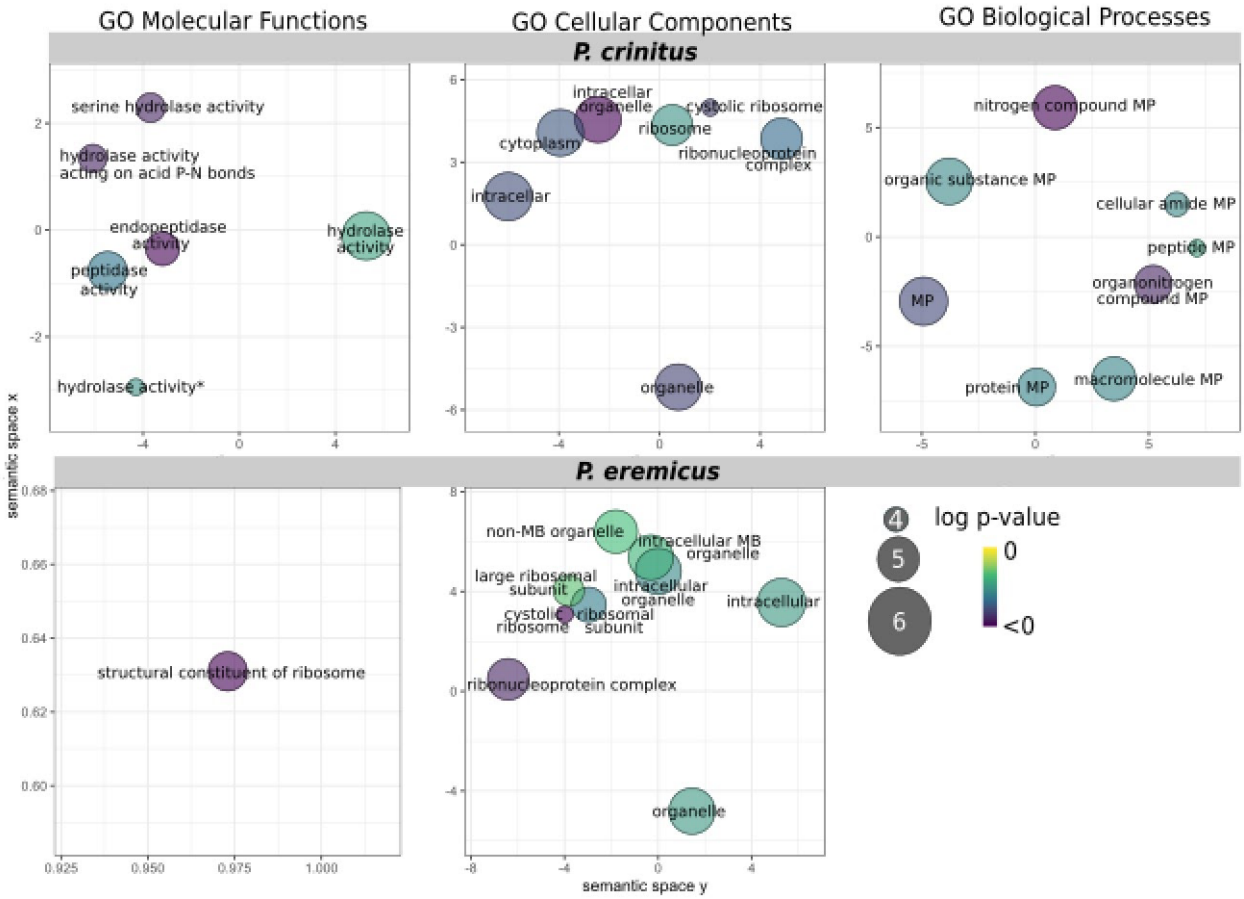

Figure 4. Composite likelihood ratio (CLR) scores for *P. crinitus* based on *Sweepfinder2* results. Values above the horizontal red line surpass the 99.9th percentile. The top five or fewer unique genes are labeled for each chromosome.

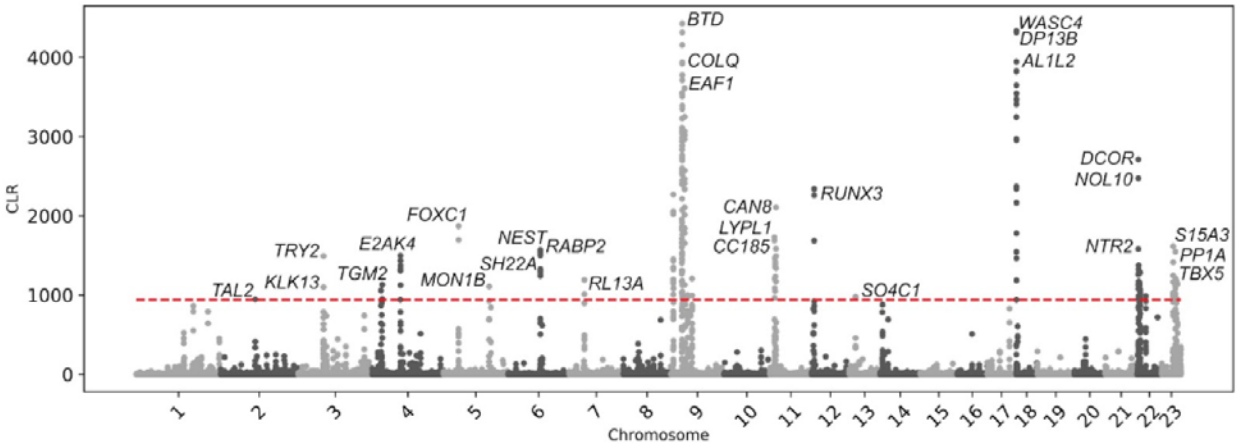

Figure 5. Overlap in proximal gene names (top row) and enriched GO terms (bottom row) for datasets I (left column), II (center), and III (right). *indicates significant overlap between species