Broad Concordance in the Spatial Distribution of Adaptive and Neutral Genetic Variation across an Elevational Gradient in Deer Mice

Rena M. Schweizer,*[,1] Matthew R. Jones,[1,2] Gideon S. Bradburd,[3] Jay F. Storz,[4] Nathan R. Senner,[1,5] Cole Wolf,[1] and Zachary A. Cheviron[1]

[1]Division of Biological Sciences, University of Montana, Missoula, MT, USA
[2]Southwest Biological Science Center, U.S. Geological Survey, Flagstaff, AZ, USA
[3]Ecology, Evolution, and Behavior Program, Department of Integrative Biology, Michigan State University, East Lansing, MI, USA
[4]School of Biological Sciences, University of Nebraska, Lincoln, NE, USA
[5]Department of Biological Sciences, University of South Carolina, Columbia, SC, USA

*Corresponding author: E-mail: rena.schweizer@umontana.edu.
Associate editor: Victoria Sork

Abstract

When species are continuously distributed across environmental gradients, the relative strength of selection and gene flow shape spatial patterns of genetic variation, potentially leading to variable levels of differentiation across loci. Determining whether adaptive genetic variation tends to be structured differently than neutral variation along environmental gradients is an open and important question in evolutionary genetics. We performed exome-wide population genomic analysis on deer mice sampled along an elevational gradient of nearly 4,000 m of vertical relief. Using a combination of selection scans, genotype–environment associations, and geographic cline analyses, we found that a large proportion of the exome has experienced a history of altitude-related selection. Elevational clines for nearly 30% of these putatively adaptive loci were shifted significantly up- or downslope of clines for loci that did not bear similar signatures of selection. Many of these selection targets can be plausibly linked to known phenotypic differences between highland and lowland deer mice, although the vast majority of these candidates have not been reported in other studies of highland taxa. Together, these results suggest new hypotheses about the genetic basis of physiological adaptation to high altitude, and the spatial distribution of adaptive genetic variation along environmental gradients.

Key words: **clinal selection, population structure, high-altitude adaptation, local adaptation, gene–environment association.**

Article

Introduction

For species that are continuously distributed across environmental gradients, the spatial scale of local adaptation is determined by the interplay between divergent selection and gene flow (Slatkin 1987; Lenormand 2002; Polechová and Barton 2015; Riesch et al. 2018; Bachmann et al. 2020). One way to gain insight into the spatial scale of local adaptation is through geographic cline analyses of allele frequencies across an environmental transect (Nagylaki 1975; Barton 1979, 1983; Stankowski et al. 2016; Storfer et al. 2018; Bradburd and Ralph 2019). When applied at a genomic scale, geographic cline analyses can be used to generate hypotheses about the environmental drivers of allele frequency variation and to identify processes that shape the distribution of adaptive genetic variation among interconnected populations (reviewed in Storfer et al. 2018). Locus-specific spatial patterns of genetic variation are determined by the combined effects of selection, gene flow, recombination rate, and the distribution of selection coefficients on nearby loci (Barton 1979, 1983; Lenormand 2002; Polechová and Barton 2015; Bachmann et al. 2020). These processes can result in clines in the frequencies of alleles at selected loci that are offset from one another, and from those of neutral loci, reflecting a combination of background population structure and selection at linked sites (Lenormand 2002; Yeaman and Whitlock 2011). However, because patterns of allele frequency variation are jointly determined by multiple demographic processes, as well as the spatial scale of selective pressures and the intensity of selection on specific loci, it is not always the case that neutral and adaptive clines are offset. For example, in zones of ecological transition, demographic processes and environmental selective pressures may align such that allele frequencies for both neutral and adaptive loci have similar geographic patterns (Endler 1977; Moore 1977). Under these scenarios,

clines for adaptive and neutral loci may often be concordant. Given these expectations, an open and important question in evolutionary genetics is whether adaptive genetic variation tends to be structured differently than neutral variation along environmental gradients.

Elevational gradients are particularly well-suited to address this question. High-elevation environments are characterized by extreme cold and low partial pressures of oxygen. Along elevational gradients, these environmental selection pressures covary in intensity and can often be combined with other covarying stressors (e.g., aridity). As a result, highland animals have evolved physiological modifications to cope with the environmental challenges of alpine environments. Many of these adaptations influence multiple steps of the oxygen transport cascade—the series of physiological processes that transport oxygen from the environment to respiring cells—to ensure matching of O_2 supply and demand (fig. 1; e.g., Storz et al. 2010; Storz and Scott 2019). In North America, deer mice (*Peromyscus maniculatus*) have an elevational range of almost 4,500 m and populations at different elevations experience a wide range of oxygen availability, temperature, precipitation, and snowpack. Deer mice native to high elevations in the Rocky Mountains have evolved a suite of physiological changes that contribute to adaptive enhancements of whole-animal aerobic performance in hypoxia (fig. 1; Cheviron et al. 2012, 2013, 2014; Lui et al. 2015; Ivy and Scott 2017, 2018; Lau et al. 2017; Dawson et al. 2018; Scott et al. 2015; Mahalingam et al. 2017, 2020; Nikel et al. 2018; Tate et al. 2017; Storz et al. 2019), a trait that influences survival in this species (Hayes and O'Connor 1999; Wilde et al. 2019).

Although the physiological mechanisms of high-elevation adaptation have been well-characterized in deer mice, our understanding of the genetic basis of these adaptations is far from complete (Storz et al. 2019; Storz 2021). Transcriptomic analyses have demonstrated that some phenotypic adaptations are associated with differential regulation of core metabolic and cell signaling pathways (Cheviron et al. 2012, 2014; Scott et al. 2015; Velotta et al. 2016, 2020), but direct connections between genotypic and phenotypic variation have been restricted to studies of a few candidate genes (e.g., Storz et al. 2009, 2010; Natarajan et al. 2015; Schweizer et al. 2019; Wearing et al. 2020; Song et al. 2021). These studies have revealed sharp clines in allele frequency that are centered at relatively low elevations in the geographic transition between the Great Plains of the central United States and the Front Range of the Rocky Mountains in Colorado (Storz et al. 2012; Schweizer et al. 2019). Two key examples include *Epas1* (endothelial PAS domain-containing protein 1), a gene involved in the transcriptional response to hypoxia, and tandemly linked gene duplicates that encode the β-subunits of adult hemoglobin (β-globin genes *Hbb-T1* and *Hbb-T2*). Across the Rocky Mountains, clines for both *Epas1* allele and β-globin haplotype frequencies are centered at ~1,400 m above sea level (a.s.l.) and span approximately 600 m of vertical relief (Storz et al. 2012; Schweizer et al. 2019). Clines for both of these putatively adaptive loci are also statistically indistinguishable from those representing background population structure (Schweizer et al. 2019). Whether these loci are representative of other putatively adaptive loci across the genome is not known.

Here, we examined elevational patterns of exome-wide variation to: 1) test whether genes that are associated with known, putatively adaptive, phenotypic differences between highland and lowland deer mice also bear signatures of positive selection in highland mice, and 2) identify additional loci that may have contributed to high-altitude adaptation. We complemented this analysis with a post hoc survey of similar population genomic studies of high-elevation adaptation in other endothermic species to determine whether targets of selection in deer mice are common in other taxa. Finally, once these putatively selected loci were identified, we then characterized geographic clines of allele frequency variation along an elevational gradient of almost 4,000 m of elevational relief. This analysis enabled us to examine the extent to which patterns of clinal variation at putatively adaptive loci differ from observed patterns across the remainder of the genome. Together, our results shed light on the spatial distribution of adaptive genetic variation across elevational gradients and suggest new hypotheses about the genetic basis of physiological adaptation to hypoxia.

Results

Sequencing Results

We sampled deer mice at low-elevation sites in the Great Plains, as well as along a transect spanning ~4,000 m of vertical relief in the southern Rocky Mountains in Colorado (fig. 2; table 1). In addition, we included samples from a low-elevation site in Merced, CA, as a reference population for some of our selection tests (supplementary table S1, Supplementary Material online). We captured ~77.4 Mb of sequence from each individual using a custom-designed capture array that targeted two sequence classes from the deer mouse genome, including: 1) the nuclear exome, and 2) 5,000 randomly selected nongenic segments of 500 bp each, to be used as a neutral control to model population structure (see Materials and Methods).

After filtering, we proceeded with a set of 256 individuals genotyped at 5,546,642 high-quality biallelic variable sites, with a mean depth of coverage of $22.24 \pm 7.90\times$ (supplementary fig. S1, Supplementary Material online) and mean missing data of $9.07 \pm 12.11\%$ (supplementary table S1, Supplementary Material online). We also generated a data set for analyses within the Rocky Mountain and Great Plains populations (table 1) that consisted of 168 unrelated individuals genotyped at 6,029,294 sites of which 267,264 were nongenic LD-pruned sites to be used for analyses of population structure. Exploratory PCA plots did not show evidence of any consistent effects of missing data (supplementary figs. S2 and S3, Supplementary Material online).

Population Genetic Structure of Rocky Mountain and Great Plains Mice

To determine how populations should be grouped for downstream analyses, we assessed fine-scale patterns of population

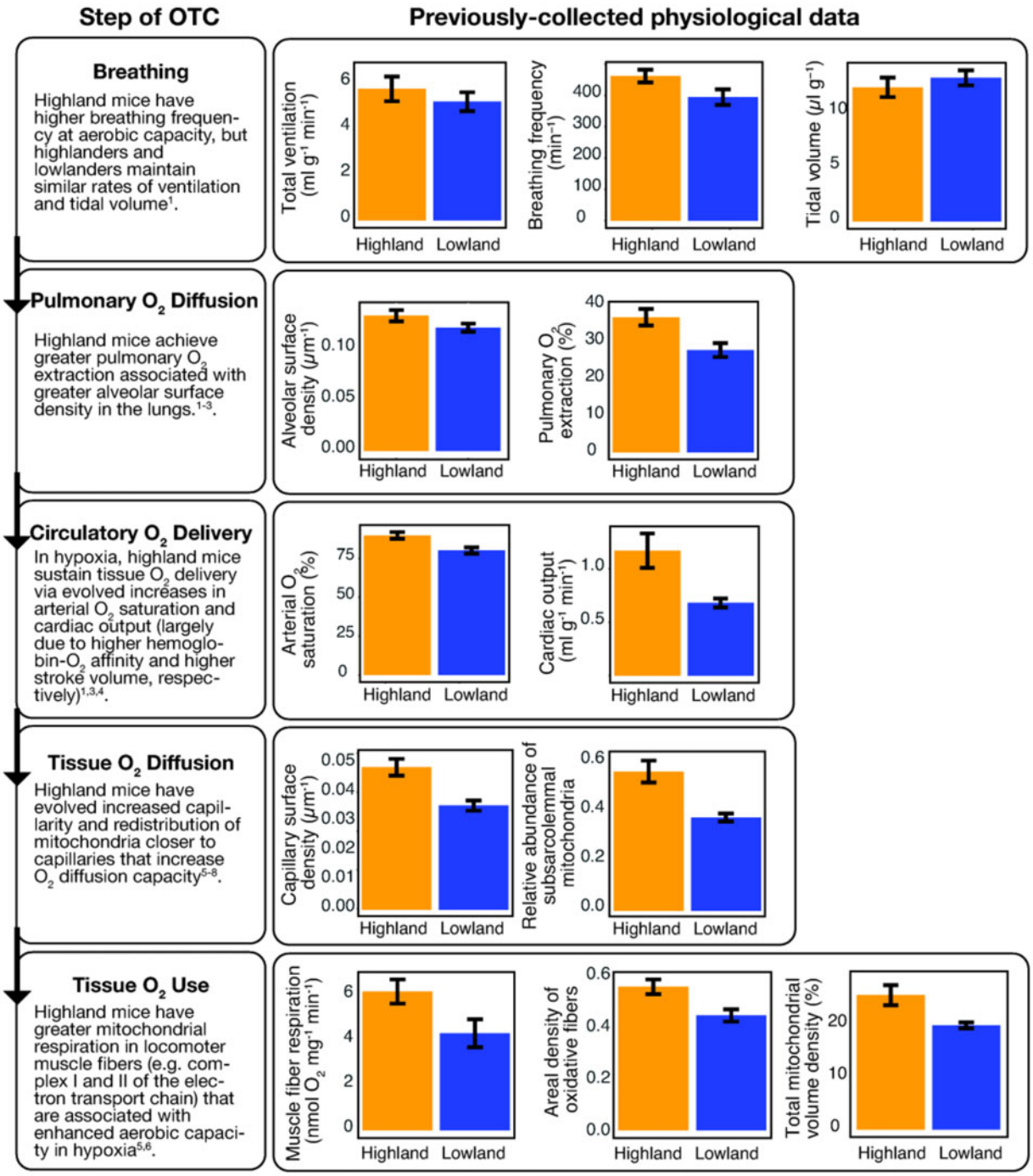

FIG. 1. Deer mice native to high elevations in the Rocky Mountains have evolved a suite of physiological changes that contribute to adaptive enhancements of whole-animal aerobic performance in hypoxia. For each step of the oxygen transport cascade (OTC), physiological differences between highland and lowland deer mice are summarized on the left and previously-collected representative physiological data are on the right. References: 1) Tate et al. (2020); 2) West et al. (2021); 3) Tate et al. (2017); 4) Ivy et al. (2020); 5) Scott et al. (2015); 6) Mahalingam et al. (2017); 7) Lui et al. (2015); 8) Nikel et al. (2018).

structure across the Great Plains-Rocky Mountains transition. We used a statistical framework that tests for discrete patterns of population structure against a backdrop of continuous geographic differentiation, implemented in the software conStruct (Bradburd et al. 2018). Analyses using conStruct showed strong and consistent support for two clusters—one comprised largely of Rocky Mountain populations (dashed box in fig. 2A: Pike Low, Pike Medium, Pike High, Boreas Pike, Lost Park, Colorado Trail, Mount Evans, Summit Lake, Echo Lake, Chicago Creek, Spring Gulch, Niwot Peak, St Vrain, Lefthand Canyon, Mesa, Big Thompson, CO), the other centered on the Great Plains samples (Pawnee, CO; Bonny Reservoir, CO; Ft. Larned, KS; Lincoln, NE)—with a clear pattern of isolation by distance within each cluster (supplementary figs. S4 and S5, Supplementary Material online). Given that the Rocky Mountain samples were well-described by a

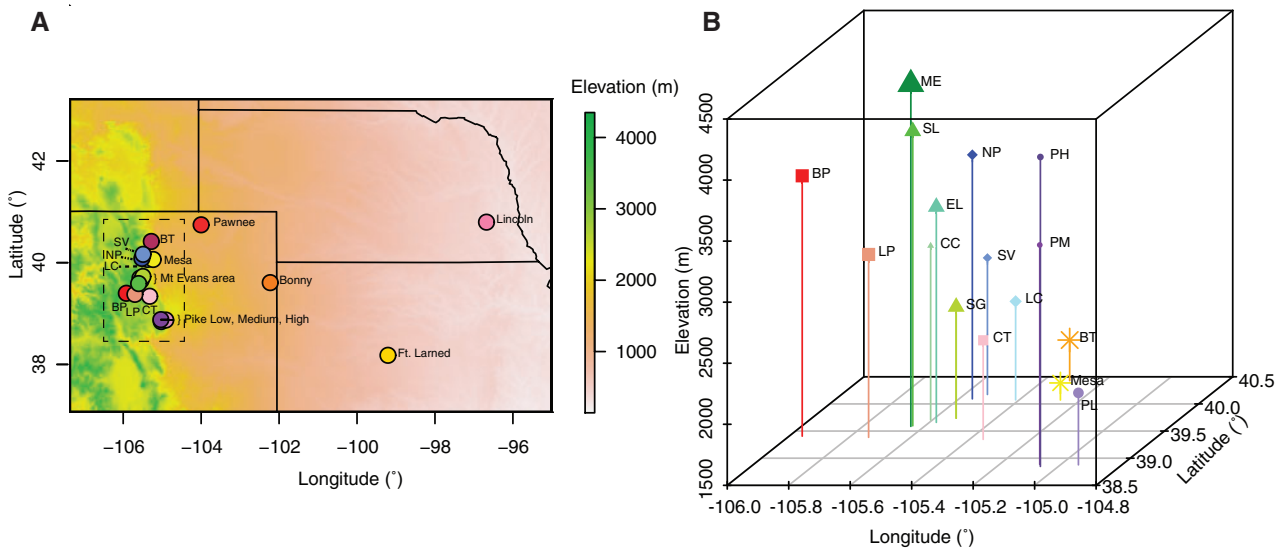

FIG. 2. (*A*) Sampling localities for deer mice along the transition between the Great Plains and the Front Range of the Southern Rocky Mountains. Dashed box represents those populations grouped within the "Rocky Mountains" for analyses. (*B*) Within the Front Range of the Southern Rocky Mountains, we sampled four elevational transects: 1) the Pikes Peak Transect: Pike Low, Pike Medium, Pike High (purples, circles); 2) the Boreas Pass Transect: Pike, Lost Park, Colorado Trail (reds, squares); 3) the Mount Evans Transect: Mount Evans, Summit Lake, Echo Lake, Chicago Creek, Spring Gulch (greens, triangles); 4) the Niwot Peak Transect: Niwot Peak, Saint Vrain, Lefthand Canyon (blues, diamonds). Shape sizes in (*B*) are proportional to number of individuals (see table 1 for sample sizes). SV, Saint Vrain; BT, Big Thompson; NP, Niwot Peak; LC, Lefthand Canyon; BP, Boreas Pike; LP, Lost Park; CT, Colorado Trail.

Table 1. Locations and Sampling Efforts for 256 Deer Mouse Samples.

Locality	Total No. of Indiv.	No. of Unrel. Indiv.	Mean Elevation (m)	Latitude (°)	Longitude (°)
Rocky Moutains					
Big Thompson	**11**	**8**	**1,832**	**40.4300**	**−105.3149**
Boreas Pike National Forest	**10**	**6**	**3,632**	**39.4071**	**−105.9577**
Chicago Creek	**5**	**4**	**2,926**	**39.7053**	**−105.6062**
Colorado Trail	**8**	**6**	**2,311**	**39.3480**	**−105.3558**
Echo Lake	**9**	**8**	**3,264**	**39.6611**	**−105.5779**
Lefthand Canyon	**10**	**6**	**2,306**	**40.0757**	**−105.4120**
Lost Park	**10**	**8**	**2,997**	**39.3856**	**−105.7370**
Mesa Reservoir	**10**	**5**	**1,639**	**40.0739**	**−105.2659**
Mount Evans	**48**	**32**	**4,302**	**39.5872**	**−105.6444**
Niwot Peak	**8**	**5**	**3,498**	**40.0948**	**−105.5571**
Pike High	**5**	**3**	**4,034**	**38.8493**	**−105.0589**
Pike Low	**9**	**6**	**2,089**	**38.8789**	**−104.9417**
Pike Middle	**4**	**4**	**3,296**	**38.8872**	**−105.0694**
Saint Vrain	**8**	**5**	**2,618**	**40.1754**	**−105.5259**
Spring Gulch	**10**	**5**	**2,411**	**39.7381**	**−105.5304**
Summit Lake	**10**	**5**	**3,912**	**39.5986**	**−105.6406**
Great Plains					
Bonny Reservoir	**10**	**9**	**1,158**	**39.6173**	**−102.2507**
Fort Larned	**10**	**7**	**620**	**38.1831**	**−99.2181**
Lincoln	**37**	**27**	**358**	**40.8106**	**−96.6803**
Pawnee	**9**	**9**	**1,594**	**40.7580**	**−104.0309**
Merced	**15**	**n/a**	**239**	**37.5986**	**−120.3415**
Total	**256**	**168**			

single, spatial cluster, we treated populations from across the elevational sampling (fig. 2*B*) as a single transect for downstream analyses.

Exome-Wide Signatures of Selection in Mt Evans Mice

To identify genes under positive selection in high-elevation mice, we used the population branch statistic (PBS; Yi et al. 2010) to measure the branch length of Mt Evans relative to both Lincoln—the lowest elevation site along our elevational transect—and Merced populations by a transformation of pairwise F_{ST} values (see Materials and Methods). Loci with elevated PBS values are indictive of loci under selection in Mt Evans. Overall, we observed relatively low differentiation across most of the 105,571 overlapping 5 kb

(supplementary fig. S6, Supplementary Material online) windows, which is consistent with previous analyses of broad-scale population structure (Storz and Kelly 2008; Storz et al. 2012; Natarajan et al. 2015; Schweizer et al. 2019). Using data simulated under an inferred demographic model (Schweizer et al. 2019), we determined the upper 0.1% of the simulated distribution of PBS (PBS $\geq$ 0.112) and used that as a threshold to identify candidate windows. A total of 4,118 windows exceeded the top 0.1% of the simulated PBS distribution, and these windows represented 1,993 unique genes (fig. 3A).

Genotype—Environment Analysis Narrows Down Top Candidate Genes

To complement the PBS analysis, we used redundancy analysis (RDA) to test for genotypic associations with elevation in the full set of samples from our Great Plains-Rocky Mountains transect. Our RDA analysis of the 1,109,794 sites (a subset with low missingness that were pruned for linkage, see Materials and Methods) revealed two significant RDA axes related to elevation (RDA1) and precipitation (RDA2) (supplementary fig. S7, Supplementary Material online). We identified a total of 15,713 unique candidate SNPs that loaded $\pm$3 standard

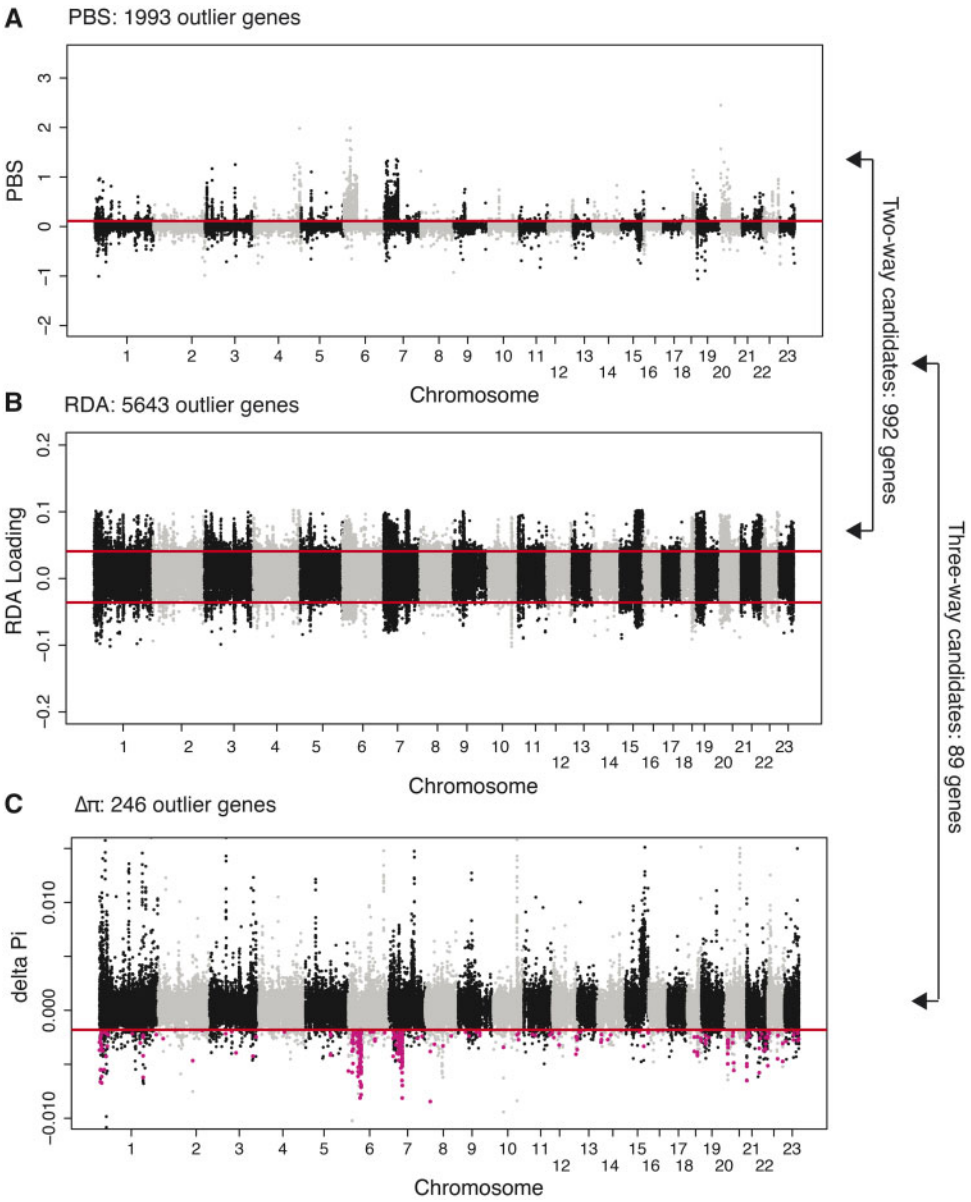

FIG. 3. Distribution of (*A*) PBS, (*B*) RDA, and (*C*) $\Delta\pi$ scores for 105,571 5-kb windows (PBS, $\Delta\pi$) and 1,109,794 SNPs (RDA) across the exome. (*A*) For PBS, points above the red line indicate windows with a PBS score above the 99.9th percentile of the demographically corrected null distribution. (*B*) For RDA, points above and below the upper and lower red lines, respectively, indicate SNPs with an RDA value that is $\pm$3 standard deviations of the mean. (*C*) For delta pi, points below the red line indicate windows with a $\Delta\pi$ in the 99.9th percentile of the null distribution for ME versus LN, and pink dots indicate windows that are outliers for ME relative to both LN and CA. See text for details.

deviations from the mean on each axis (12,637 on RDA1, 2,230 on RDA2, and 846 on RDA1 and RDA2). Of the total RDA outliers, 89.4% (14,055 SNPs representing 5,643 genes; fig. 3B) were more strongly correlated with elevation than precipitation. We chose to focus on the elevation outliers in subsequent analyses, but we acknowledge that there are potentially interesting links with precipitation.

A total of 992 unique genes overlapped between the SNP-based RDA elevation outliers and the window-based PBS outliers (top 35 in table 2; all 992 in supplementary table S2, Supplementary Material online). These genes represent a set of loci (henceforth, "two-way" candidate loci) that are associated with elevation and exhibit evidence of selection in the Mt. Evans population. Among these genes, there was a significant functional enrichment of eight Biological Process, four Cellular Component, and four Molecular Function categories (supplementary table S3, Supplementary Material online). The most significantly enriched category was "anatomical structure development" (GO: 0048856; *P*-value: 6.42E−06). Other significant categories of interest included "ion binding" (GO: 0043167, *P*-value: 0.0109), "regulation of gene expression" (GO:0010468; *P*-value: 0.024), "muscle cell migration" (GO:0014812; *P*-value: 0.0293), and "G-protein coupled purinergic nucleotide receptor signaling pathway" (GO:0035589; *P*-value: 0.0307; supplementary table S3, Supplementary Material online).

Evidence of Recent Selective Sweeps

To identify loci that bear signatures of recent selective sweeps at high elevations, we also identified windows with significantly reduced nucleotide diversity ($\Delta\pi$) in Mt Evans mice relative to the two lowland populations (see Materials and Methods). We used these data for two purposes. First, we tested whether the 992 two-way outliers exhibited significantly higher $\Delta\pi$ values than a random sample of nonoutlier loci, as would be expected if they had experienced a recent selective sweep at high elevation. This analysis revealed that the mean $\Delta\pi$ value for two-way outliers was indeed higher than that of random loci (Welch two sample *t*-test; $t = -11.283$, $P < 2.2 \times 10^{-16}$). Second, we performed a scan of $\Delta\pi$ outliers to identify additional loci that may have experienced a recent selective sweep at high elevation. As we did with PBS, we simulated nucleotide diversity for 100,000 5 kb windows under our demographic model to generate a null distribution of $\Delta\pi$ values. Our scan identified 314 windows with a $\Delta\pi$ value that was greater than the top 0.1% of the demographically corrected distribution for both highland/lowland comparisons (fig. 3C). Of the 246 unique genes in these windows, 89 overlapped with the 992 two-way candidate loci (supplementary table S4, Supplementary Material online). Gene enrichment analyses of this reduced candidate set (henceforth, "three-way" candidate loci) showed an enrichment of GO categories such as "negative regulation of catecholamine secretion" (GO:0033604) and "G-protein coupled nucleotide receptor activity" (GO:0001608) (supplementary table S5, Supplementary Material online). There was also significant enrichment within the reactome pathway (R-HSA-417957) "P2Y receptors."

Clinal Patterns of Variation for Candidate Loci

Given that we observed high levels of gene flow and low levels of differentiation among Rocky Mountain populations, we analyzed all populations, including those from the Great Plains, as a single elevational transect to assess clinal patterns of variation. For each of the 992 two-way candidate loci, we used the software HZAR (Derryberry et al. 2014) to fit a sigmoidal *tanh* cline model to the relationship between allele frequency and sampling elevation (see Materials and Methods). We estimated the cline center (*c*) and width (*w*) for 992 SNPs within our two-way candidate subset (supplementary table S6, Supplementary Material online). The variables *c* and *w* characterize the geographic location along the transect where the allele frequency turnover is greatest and the geographic region corresponding to the inverse of the maximum cline slope, respectively. We compared these best-fit clines with a cline generated using PC1 of our nongenic SNPs (supplementary fig. S8, Supplementary Material online; see Materials and Methods), and identified a subset of two-way candidate loci with cline centers ($n = 297$; 29.9%) and widths ($n = 240$; 24.2%) outside of the 95% confidence interval (CI) of the nongenic PC1 cline (fig. 4). Genes with clines centers that were offset upslope (i.e., occurring at higher elevation; $n = 158$ or 15.9%) of the nongenic PC1 cline showed an enrichment of functions related to catecholamine secretion and multiple categories related to odor perception ("sensory perception of smell," "odorant binding," and "olfactory receptor activity"; supplementary table S7, Supplementary Material online), whereas genes with cline centers that were offset downslope ($n = 297$ or 29.9%) showed enrichment of reactome pathways related to transport and cell junction organization (supplementary table S7, Supplementary Material online). There were no significantly enriched GO categories (supplementary table S8, Supplementary Material online) among the genes with cline widths that were narrower than the nongenic PC1 cline, whereas genes with cline widths greater than the nongenic PC1 cline only showed enrichment of broad categories such as "cell periphery" and "membrane part" (supplementary table S8, Supplementary Material online). The majority of our two-way candidate loci, however, were characterized by cline centers ($n = 695$ or 70.1%) or widths ($n = 752$ or 75.8%) that were indistinguishable from the PC1 cline representing background population structure.

Post Hoc Review of Candidate Genes in High-Elevation Vertebrates

To place the results of our selection scan into a broader context and to assess the degree of overlap in the genomic targets of selection across other similar studies, we performed a post hoc review of 14 studies in nine different species of terrestrial vertebrates (supplementary table S9, Supplementary Material online). We limited our survey to those that used population genomic data sets and allele-frequency-based tests of selection to allow for more direct comparison with our results. This survey identified a total of 3,983 unique genes that have been identified as selection candidates in other studies. Of our 992 two-way outlier

Table 2. Top 35 Candidate Genes That Are Outliers for Both PBS and RDA (two-way candidates).

Gene Symbol (*P. maniculatus*)	Gene Name (*P. maniculatus*)	Gene Symbol (*M. musculus*)	Window Location	PBS	SNP Location	RDA
Stx16	Syntaxin 16	*Stx16*	6501107.1:962501 − 967500	1.981	6501107.1:967816	0.795
Aff2	AF4/FMR2 family member 2	*Aff2*	6501704.1:907501 − 912500	1.788	6501704.1:910760	0.790
Hps3	HPS3 biogenesis of lysosomal organelles complex 2 subunit 1	*Hps3*	6501722.1:610001 − 615000	1.573	6501722.1:590083	0.737
Gabrq	Gamma-aminobutyric acid type A receptor theta subunit	*Gabrq*	6501587.1:1297501 − 1302500	1.530	6501587.1:1300279	0.657
Cpa3	Carboxypeptidase A3	*Cpa3*	6501722.1:272501 − 277500	1.381	6501722.1:283951	0.486
Cwf19l2	CWF19-like 2 cell cycle control (S. pombe)	*Cwf19l2*	6501598.1:270001 − 275000	1.324	6501598.1:334183	0.777
Usp28[a]	Ubiquitin specific peptidase 28	*Usp28*	6501163.1:165001 − 170000	1.320	6501163.1:208277	0.823
LOC102906935	Interstitial collagenase A-like	*Mmp1a*	6501245.1:2847501 − 2852500	1.299	6501245.1:2841407	0.596
Znfx1	Zinc finger protein OZF	*Znfx1*	6501708.1:275001 − 280000	1.275	6501708.1:284143	0.625
Dync2h1	Dynein cytoplasmic 2 heavy chain 1	*Dync2h1*	6501245.1:2427501 − 2432500	1.268	6501245.1:2399127	0.808
Exoc6b	Exocyst complex component 6B	*Exoc6b*	6501567.1:1457501 − 1462500	1.251	6501567.1:1061086	0.687
Gria4	Glutamate ionotropic receptor AMPA type subunit 4	*Gria4*	6501245.1:37501 − 42500	1.219	6501245.1:40269	0.837
Kmt2a	Lysine methyltransferase 2A	*Kmt2a*	6501694.1:475001 − 480000	1.186	6501694.1:481688	0.832
Ptprz1	Protein tyrosine phosphatase type IVA member 1	*Ptprz1*	6501405.1:145001 − 150000	1.169	6501405.1:188227	0.646
Rims2	Regulating synaptic membrane exocytosis 2	*Rims2*	6501417.1:1140001 − 1145000	1.152	6501417.1:1177690	0.456
Cpsf6	Cleavage and polyadenylation specific factor 6	*Cpsf6*	6501066.1:10150001 − 10155000	1.136	6501066.1:10172904	0.852
LOC102906541	Pyrethroid hydrolase Ces2e-like	*Ces2b*	6501344.1:235001 − 240000	1.102	6501344.1:244202	0.686
Ctsz	Cathepsin Z	*Ctsz*	6501107.1:1315001 − 1320000	1.095	6501107.1:1306134	0.417
Hspa4l	Heat shock protein family A (Hsp70) member 4 like	*Hspa4l*	6501366.1:1650001 − 1655000	1.091	6501366.1:1682358	0.736
Edn3	Endothelin 3	*Edn3*	6501107.1:1652501 − 1657500	1.071	6501107.1:1660223	0.656
Bud13	BUD13 homolog	*Bud13*	6501163.1:3185001 − 3190000	1.058	6501163.1:3186863	0.803
Ncam1[a]	Neural cell adhesion molecule 1	*Ncam1*	6501057.1:4725001 − 4730000	1.058	6501057.1:4733662	0.781
Cpb1	Carboxypeptidase B1	*Cpb1*	6501722.1:225001 − 230000	1.052	6501722.1:221496	0.736
Cadm1	Cell adhesion molecule 1	*Cadm1*	6501163.1:1602501 − 1607500	1.041	6501163.1:1552848	0.831
Nnat	Neuronatin	*Nnat*	6501257.1:617501 − 622500	1.037	6501257.1:620180	0.343
Supt20h	SPT20 homolog SAGA complex component	*Supt20*	6501125.1:3700001 − 3705000	1.035	6501125.1:3670022	0.530
Med12l[a]	Mediator complex subunit 12 like	*Med12l*	6501814.1:835001 − 840000	1.032	6501814.1:872041	0.661
LOC102923525	Probable G-protein coupled receptor 83	*Gpr165*	6501190.1:5030001 − 5035000	1.012	6501190.1:5031532	0.496

(continued)

Plch1[a]	Phospholipase C eta 1	*Plch1*	6501060.1:2525001 − 2530000	1.007	6501060.1:2440272	0.704
Fxr1	FMR1 autosomal homolog 1	*Fxr1*	6501046.1:2650001 − 2655000	0.994	6501046.1:2633664	0.512
Angpt1	Angiopoietin 1	*Angpt1*	6501143.1:3235001 − 3240000	0.991	6501143.1:3032740	0.735
Mme[a]	Membrane metallo-endopeptidase	*Mme*	6501060.1:2105001 − 2110000	0.972	6501060.1:2148909	0.693
Peg3	Paternally expressed 3	*Peg3*	6501712.1:890001 − 895000	0.967	6501712.1:889216	0.781
LOC102914701	Arylacetamide deacetylase-like 2	*Gm8298*	6501814.1:1227501 − 1232500	0.961	6501814.1:1255421	0.468
Nlgn1	Neuroligin 1	*Nlgn1*	6501046.1:11272501 − 11277500	0.960	6501046.1:11281042	0.700

[a]Gene is also a significant three-way outlier (see supplementary table S2, Supplementary Material online). Locations are provided as NW contig ID and position.

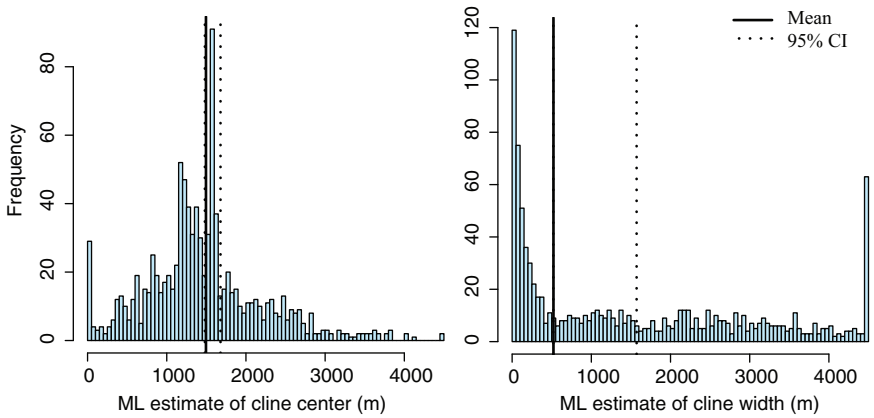

FIG. 4. Histograms of cline centers and cline widths for 992 SNPs from the two-way candidate loci. Solid and dashed lines represent the mean and upper and lower 95% CIs, respectively, of the nongenic PC1 cline.

candidate genes, only 163 (16.4%) have been identified in other selection scans of high-altitude populations. Of those 163 genes, 139 (85.3%) were only identified in one other study, 22 were identified in two additional studies, and 2 were identified in three or more additional studies (*Epas1* and *Kcnma1*).

Discussion

In this study, we took a population genomic approach to assess the concordance of spatial patterns of genetic variation between adaptive and neutral loci. We combined selection scans with genotype−environment associations and geographic cline analysis along an elevational gradient of nearly 4,000 m vertical relief to identify loci that bear the signature of natural selection at high elevations. We then compared spatial patterns of allele frequency variation at putatively adaptive loci with those at neutral loci. The combined results of these analyses reveal three primary insights. First, multiple lines of evidence indicate that altitude-related selection has shaped patterns of genetic variation across a large portion of the genome. As detailed below, we identified hundreds of loci that bear signatures of selection in highland population samples, many of which can be plausibly linked to known physiological differences between high- and low-elevation deer mice. Second, the vast majority of the selection candidates we identified have not been reported in similar studies of other highland taxa, suggesting potentially novel candidate genes and physiological pathways for adaptation to high elevation. Finally, our geographic cline analysis revealed that most loci under selection show clinal patterns of allele frequency variation that are concordant with background population structure, as only a small subset of putatively adaptive loci are characterized by cline centers shifted significantly up- or downslope from the genome-wide average.

Genomic Signatures of Selection on Genes That Span Multiple Physiological Systems

Recent studies have uncovered many of the physiological traits involved in high-elevation adaptation in deer mice, in addition to some of their gene regulatory underpinnings (reviewed in Storz et al. 2019; Storz and Cheviron 2021). This suite of physiological adaptations spans multiple steps in the transport pathways for oxygen and metabolic substrates (fig. 1). Consistent with this pattern of multitrait adaptation, we found hundreds of loci bearing signatures of positive selection at high elevations, many of which are likely involved in functions that relate to processes that influence oxygen homeostasis and aerobic metabolism (fig. 1). For brevity, we highlight just a few of the most promising candidate genes that may relate to known physiological differences between highland and lowland deer mice below. A list of the top 35 two-way candidates is presented in table 2, and the full list of selection candidates is provided in the supplemental materials (supplementary tables S2, S4, and S6, Supplementary Material online).

A key aspect of circulatory oxygen and metabolic fuel delivery is the regulation of blood flow via dynamic modifications of local blood pressure. Several recent studies have documented adaptive modifications of blood pressure regulation under hypoxia in highland deer mice. For example, highland mice exhibit reduced pulmonary hypertension under hypoxia compared with their lowland counterparts (Velotta et al. 2018); this lack of a global vasoconstrictive response likely contributes to their ability to achieve higher pulmonary O_2 extraction (Tate et al. 2017) (fig. 1). Angiotensins, and their precursor angiotensinogen, are key regulators of blood pressure and fluid homeostasis (Wu et al. 2011). One of our most compelling two-way outliers is an angiotensin receptor (*Agtr1b*, angiotensin II receptor type 1) that mediates cardiovascular effects of angiotensin, including vasoconstriction (Bonnardeaux et al. 1994). Similarly, several other outliers are P2Y receptors—purinergic G protein-coupled receptors that also play a role in vasodilation (Burnstock and Ralevic 2013; Sluyter 2015). It is conceivable that allelic variation at these loci could contribute to the modifications of adaptive pulmonary function in highland deer mice.

Highland deer mice also show a greater capacity for tissue oxygen extraction (Tate et al. 2020), in part because of evolved differences in skeletal muscle capillarity, fiber composition, and mitochondrial density and distribution (Lui et al. 2015; Scott et al. 2015; Mahalingam et al. 2017). One potential candidate gene that could contribute to these differences in muscle phenotype is *Itga7* (integrin subunit alpha 7; two-way outlier), which may have a role in the formation of muscle fibers (Mayer et al. 1997). Other candidate genes seem to be related to these phenotypic differences as well. One example is *Angpt1* (angiopoietin; two-way outlier), which plays an important role in vascular development and angiogenesis, as well as blood vessel maturation. Due to their effects on relevant muscle phenotypes, variants at these loci may contribute to known differences in tissue diffusion capacity between highlanders and lowlanders (Lui et al. 2015; Scott et al. 2015; Mahalingam et al. 2017; Nikel et al. 2018; Mahalingam et al. 2020).

Finally, highland deer mice exhibit greater whole-animal aerobic performance under hypoxia that is associated with a number of measured phenotypes, including differential regulation of core metabolic pathways that contribute to lipid and carbohydrate metabolism and oxidative phosphorylation (Cheviron et al. 2012, 2013; 2014; Lau et al. 2017). Several genes that participate in these processes also bear signatures of selection in the highland population. One example is *Ndufc1* (NADH: Ubiquinone Oxidoreductase Subunit C1), which encodes a subunit in the first enzyme complex of the electron transport chain in mitochondria. Previous work on deer mice has demonstrated an increased mitochondrial respiratory capacity in muscle, and other evolved changes in mitochondrial physiology of high-altitude populations (Mahalingam et al. 2017; fig. 1); variation at *Ndufc1* and other genes in related pathways may contribute to these differences in mitochondrial function (supplementary tables S2, S4, and S6, Supplementary Material online).

Importantly, although we have highlighted a few key genes related to several steps of the oxygen homeostasis and aerobic metabolism pathways, we have not demonstrated associations between putatively adaptive phenotypes and allelic variation at these loci. Many of the other candidate genes we have identified, but did not highlight, could also be involved in these complex physiological processes (supplementary tables S2, S4, and S6, Supplementary Material online), although it is also possible that our candidate gene lists contain false positives. Future work should aim to experimentally document phenotypic effects of mutations in some of the most compelling candidates, and to test for their effects on fitness (Barrett and Hoekstra 2011). Additionally, further efforts should focus on formal tests of polygenic adaptation by determining specific alleles associated with phenotypes of interest (e.g., through a genome-wide association study), and demonstrating that those alleles have population frequency differences that consistently increase or decrease (Berg and Coop 2014; Jeong et al. 2018).

Most Genomic Targets of Selection Are Unique to Deer Mice

Recent surveys in humans and domesticated animals have documented overlap in the genomic targets of selection among independent high-elevation populations, suggesting that adaptation to these environments may often involve repeated selection on a common set of genes (Witt and Huerta-Sanchez 2019; Storz and Cheviron 2021). Although a number of recent studies have documented selection on obvious candidate genes, such as *Epas1*, in independent lineages of wild vertebrates, these examples are often cherry-picked from a list of outliers specifically because they have been highlighted in other studies which may therefore give a biased view of the degree of convergence in selection targets at high elevation.

In our study, only two genes—*Epas1* and *KCNMA1* (potassium calcium-activated channel subfamily M alpha 1)—that were detected as outliers in highland deer mice have been identified in more than three additional highland populations of other species. *Epas1* encodes the oxygen sensitive subunit of hypoxia-inducible factor, a transcription factor that coordinates the transcriptional response to hypoxia, and was an outlier in 9 of the 14 studies representing 7 different species (Simonson et al. 2010; Yi et al. 2010; Li et al. 2014; Zhang et al. 2014; Song et al. 2016; Liu et al. 2019). *KCNMA1* encodes the alpha-pore of calcium-sensitive potassium channels that influences vascular tone and blood flow by regulating K+ efflux in vascular smooth muscle cells (Brayden and Nelson 1992; Knaus et al. 1995), and was a target of selection in four different species in addition to deer mice (Zhang et al. 2014; Qu et al. 2015; Song et al. 2016; Jeong et al. 2018). These two examples aside, the general lack of overlap suggests a diversity of different mechanisms underlying high-elevation adaptation. However, we cannot rule out the possibility that the lack of overlap is, to some extent, due to the presence of false positives in our or other studies. Nonetheless, the unique selection targets identified here may provide novel insight into physiological

mechanisms to surmount the challenges of high-elevation environments.

Spatial Patterns in Adaptive Genetic Variation Are Largely Consistent With Population Structure

Across environmental gradients, the relative strength of selection and gene flow shape spatial patterns of genetic variation (Slatkin 1987; Lenormand 2002). When many loci are subject to spatially varying selection, variation in local recombination rates may lead to variable levels of gene flow across loci (Endler 1977). This process can result in allele frequency clines for selected loci that are discordant with background population structure, though this need not always be the case (Yeaman and Whitlock 2011; Lenormand 2002). Our analysis revealed that the majority of putatively selected loci had cline centers (70.1%) and cline widths (75.8%) that fell within the 95% CI of genome-wide estimates from nongenic regions. This result suggests that, for the most part, loci that have experienced a history of altitude-related selection do not exhibit patterns of spatial structure that are distinct from background population structure. Not all cline shapes were concordant with population structure, however, suggesting that locus-specific levels of gene flow may structure allelic variation at such loci differently than background neutral genetic variation. This finding supports previous results from simulation studies suggesting that spatially varying selection can structure groups of locally adapted alleles over large geographic distances, even among populations that are connected by high rates of gene flow (Yeaman and Whitlock 2011).

Often, the loci with cline shapes that deviated from background population structure were not enriched for specific functions, suggesting that this class of loci participate in a broad range of biological functions. The one exception is for genes with clines that are centered upslope of the nongenic PC1 cline. This list of genes showed an enrichment of functions related to catecholamine secretion and odor perception. Catecholamine synthesis and secretion by the adrenal gland is responsive to environmental hypoxia, and catecholamines can affect many physiological processes that impinge on oxygen delivery and consumption, including heart rate and vasoconstrictive responses (Brown et al. 2009). High- and low-elevation deer mice differ in their catecholamine response to hypoxia (Scott et al. 2019), and it is possible that these loci may contribute to this physiological difference. For example, previous work has shown that *Epas1*, a two-way outlier, is a target of selection in deer mice and is associated with variation in heart rate under hypoxia (Schweizer et al. 2019). Allelic variation in *Epas1* causes a differential regulation of the catecholamine biosynthesis pathway (Schweizer et al. 2019), and the specific mutation under selection at high elevation disrupts interaction with a transcriptional coactivator, thus providing a possible mechanistic explanation (Song et al. 2021). We identified several additional promising candidate genes, such as *P2RY1*, *DRD2*, and *P2RY12*, that are related to catecholamine regulation under hypoxia (fig. 5). Studies of the phenotypic effects of allelic variation at these loci would be a fruitful area for future work.

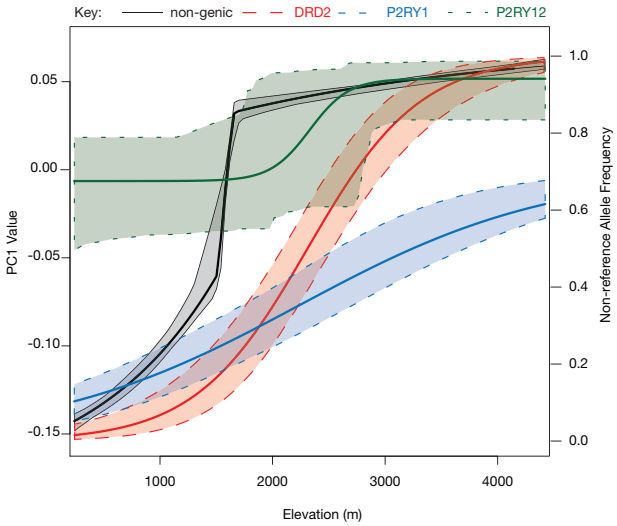

Fig. 5. Maximum-likelihood allele frequency clines for catecholamine genes located upslope of the nongenic PC1 cline. For the nongenic cline, the *y*-axis shows the PC1 values, whereas for the DRD2, P2RY1, and P2RY12 clines, the *y*-axis represents the frequency of the non-reference allele. Shaded regions show the 95% CIs.

No other grouping of loci with discordant clines exhibited clear gene ontology enrichment. Loci with clines that were centered downslope, as well as those with widths that were significantly narrower or greater than the PC1 cline, were not enriched for terms that were obviously related to abiotic selective pressures along elevational gradients. We note that functional enrichments such as those presented above are post hoc tests that generate hypotheses to be addressed in future work, rather than formal tests of previously specified hypotheses.

Conclusions

Our results demonstrate that hundreds of genes have experienced a history of spatially varying selection at high elevation in deer mice, and many of these loci participate in physiological processes that underlie known phenotypic differences between highland and lowland populations. The vast majority of selection targets we identified have not been reported in similar studies of other highland terrestrial vertebrates. Although convergence in phenotypic traits is relatively common in high-altitude vertebrates (reviewed in Storz and Scott 2019), the general lack of overlap among selection targets between deer mice and other highland species suggests that the genetic underpinnings of this phenotypic convergence may be more idiosyncratic. Finally, our results also show that, at least for deer mice along this elevational gradient, adaptive and neutral genetic variation tend to be structured similarly across the landscape. If this is a general outcome, it may have important implications not only for our understanding of the process of local adaptation, but also for more directed applications in conservation genetics, such as assisted migration and genetic rescue.

Materials and Methods

Sampling Scheme

Tissue samples of deer mice were collected from a variety of sources. In the field, we live-trapped deer mice using baited Sherman traps. We also used liver samples from euthanized mice, and blood or tail clip samples from mice that were part of a mark-recapture study (e.g., Wilde et al. 2019), in which individuals were released after capture. To augment our sample sizes for some sites, we used tissue samples from previously collected museum specimens that are cataloged in the mammal collections of the Denver Museum of Nature and Science, the University of Arizona Museum of Natural History Museum, the Museum of Vertebrate Zoology, and the Museum of Southwestern Biology (see Natarajan et al. 2015).

Exome Design, Capture, and High-Throughput Sequencing

To identify all annotated exons, we downloaded the *Peromyscus maniculatus bairdii* general feature file (GFF) v101 from NCBI (ftp://ftp.ncbi.nih.gov/genomes/Peromyscus_maniculatus_bairdii/GFF/), and extracted all features annotated as an exon. The final set of unique, non-pseudogenized exonic regions consisted of 218,065 exons in 25,246 genes. We designed 500 bp nongenic regions to be located at least 10 kb from any annotated gene, outside repetitive DNA regions, containing GC content within one standard deviation of the mean GC content, and located on contigs larger than 20 kb (Wall et al. 2008; Schweizer et al. 2016). Five thousand autosomal regions were randomly chosen that satisfied these criteria. In total, a custom Roche NimbleGen SeqCap EZ Library captured 226,973 regions (77,559,614 bp).

We used 256 mouse specimens for the analysis of exomic variation (table 1), 100 of which were used in a previous study (NCBI Short Read Archive PRJNA528923) (Schweizer et al. 2019). We extracted DNA from tissue samples of 156 deer mice (supplementary table S1, Supplementary Material online) using a Qiagen DNeasy kit and sheared DNA to ~300 bp using a Covaris E220 Focused Ultrasonicator. A genomic library for each individual was prepared using 200 ng of sheared DNA with a NEBNext Ultra II kit and unique index following the manufacturer's protocols (New England Biolabs). Batches of 24 indexed libraries were pooled, then target enriched and PCR amplified according to the NimbleGen Seq Cap EZ protocol (Roche). Quality control for each capture pool included a check of its size distribution on an Agilent TapeStation, as well as a check for enrichment of targeted regions and lack of enrichment of nontargeted regions using custom primers and the Luna qPCR master mix (NEB). Each capture pool of 24 individuals was sequenced with 100 bp paired-end sequencing on an Illumina HiSeq 4000. All 256 individuals were processed concurrently.

Data preprocessing and variant discovery on all samples followed the recommendations of the Broad Institute GATK v3.7-0-gcfedb67 Best Practices pipeline (https://software.broadinstitute.org/gatk/best-practices/workflow; last accessed June 2, 2021) and followed our previously published methods (Schweizer et al. 2019). Briefly, we trimmed sequence reads of adapter sequences and bases with quality below 20 using *fastq_illumina_filter 0.1* (http://cancan.cshl.edu/labmembers/gordon/fastq_illumina_filter/; last accessed November 1 2018) and *trim_galore 0.3.1* (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/; last accessed June 2, 2021). Forward and reverse reads were aligned and mapped to the *P. maniculatus baiardii* genome using *bwa mem* (Li and Durbin 2010), then duplicates were removed using *samtools rmdup* (Li et al. 2009). After two rounds of GATK *Base Quality Score Recalibration*, we genotyped each sample using *GATK HaplotypeCaller* with the "–emitRefConfidence" flag, then called variants using *GATK GenotypeGVCFs*. GVCFs were combined and filtered to remove SNPs with a quality of depth <2.0, an FS >60, mapping quality <40, mapping quality rank sum <-12.5, and read position rank sum <-8.0. This process identified 106,883,914 variable sites in at least 1 of our 256 individuals. However, three sampled individuals were dropped from further analysis due to high levels of missing data (>50% sites with missing genotype calls). After assessing the quality of filtered reads using the *vcftools* package (Danecek et al. 2011), we further filtered variants so that a site was called in at least 75% of individuals, was biallelic, and had a minimum depth of 5 and genotype quality of 20.

Population Genetic Structure of Rocky Mountain and Great Plains Mice

To focus our efforts on the Rocky Mountains-Great Plains transect, we removed the geographically disparate Merced samples and used an LD-pruned set of nongenic SNPs (using the "–indep-pairwise 50 5 0.5" flag in PLINK; Purcell et al. 2007). These nongenic SNPs are best suited for assessing neutral population processes. We also identified a subset of unrelated individuals using PRIMUS (Staples et al. 2013) and a maximum identity-by-descent of 0.1875, as recommended in Anderson et al. (2010).

The method implemented in the software conStruct is well-suited to our sampling design and the geographic distribution of deer mice, where there is a high likelihood of isolation-by-distance and the sampling is discontinuous (Bradburd et al. 2018). We ran conStruct on two data sets: All unrelated samples ($N = 168$), and just those east of the Rocky Mountains (i.e., excluding Bonny Reservoir, Ft. Larned, and Lincoln; $N = 117$). We evaluated both the spatial and nonspatial models with the number of discrete populations (K) varying between 1 and 5. For each model, we ran 2 replicate analyses, each for 5,000 iterations. The performance of the MCMC was assessed by comparison between replicate runs and visual inspection of marginal parameter estimate trace plots. We compared models using the "layer comparisons" approach outlined in Bradburd et al. (2018).

Exome Scan for Selection

To calculate PBS, for each population pair we used vcftools to calculate pairwise F_{ST} in 5-kb windows with a step size of 2.5 kb (following Jones et al. 2018) and specified vcftools to only use sequenced sites for those calculations. We then

transformed F_{ST} and calculated PBS for each window following Yi et al. (2010). We calculated PBS for 105,571 overlapping 5 kb windows from the exomes of 48 Mt Evans, 37 Lincoln, and 15 Merced mice, then, as a demographic control, simulated the distribution of F_{ST} under our previously characterized demographic history of these same populations (Schweizer et al. 2019). Using point estimates from our previously characterized demographic history of the Mt Evans, Lincoln, and Merced mice (Schweizer et al. 2019), we used msms (Ewing and Hermisson 2010) to simulate the distributions of F_{ST} for 17,000 5-kb windows. We calculated PBS values for all simulated windows, then used the 99.9% quantile of the simulated distribution to set the significance threshold for the empirical data using the *ecdf()* function in R. For each 5 kb window, we identified which genes overlapped that window using the *bedtools* intersect function and a bed file of targeted exonic regions.

Multivariate Genotype−Environment Associations

RDA characterizes a response matrix (here, genotypes) in relation to an explanatory matrix (here, environmental data) using multivariate linear regressions, followed by a PCA to produce canonical axes (Van Den Wollenberg 1977; Legendre et al. 2011). We implemented RDA in our set of mice sampled across the entire transect ($N = 165$ unrelated individuals) following the recommendations of (Forester et al. 2018). Three of the 168 individuals were not included in RDA because high missingness might bias the genotype imputation done for that analysis. RDA shows low false positive and high true positive rates under a variety of selection and demographic scenarios (Forester et al. 2018). Briefly, we obtained population-level environmental data for precipitation and temperature from BIOCLIM (Hijmans et al. 2005) using the *getData* function within the "raster" package (https://rspatial.org/raster/; last accessed June 2, 2021). We also obtained estimates of snowpack (measured as daily mean snow water equivalent from 1915 to 2011) from Livneh data provided by the NOAA/OAR/ESRL PSL, Boulder, Colorado, USA at their web site at https://psl.noaa.gov/ (last accessed June 2, 2021) (Livneh et al. 2013). We initially chose 11 environmental variables summarizing precipitation, temperature, and snowpack, then removed those variables with a Pearson correlation greater than $|0.70|$, as recommended by Dormann et al. (2013), while prioritizing the retention of elevation as a variable. Given that many environmental variables are highly correlated with elevation (supplementary fig. S9, Supplementary Material online), we subsequently chose a subset of two variables (elevation and annual precipitation) for further analysis. With this approach we aimed to identify the variable (or correlated variables) that caused the observed spatial patterning of allele frequencies. Due to computational limitations within RDA, we subsampled our genotype data by further pruning for linkage ($r^2 \leq 0.75$) and missingness ($\leq$1%) in PLINK, resulting in a set of 1,109,794 sites. Given that RDA requires no missing data and removing sites with any missing data would have resulted in the loss of 483,261 (43.5%) sites, we imputed missing genotypes for our data set ($n = 483{,}261$ or 0.26% of all data) using the most common genotype across all individuals (Forester et al. 2018). After running RDA, we identified significant constrained axes with a *P*-value of <0.05 after 999 permutations of the genotypic data. We identified candidate SNPs as those with a loading greater than 3 standard deviations of the mean and characterized each candidate SNP by the environmental variable with which it had the highest correlation.

Detection of Selective Sweeps

To identify loci that bear signatures of selective sweeps in the Mt Evans population, we identified 5 kb windows with a reduced nucleotide diversity (π) both in Mt Evans relative to Lincoln (by calculating delta pi ($\Delta\pi$), or $\pi_{\text{Mt Evans}} - \pi_{\text{Lincoln}}$), and in Mt Evans relative to Merced (by calculating $\Delta\pi$ as $\pi_{\text{Mt Evans}} - \pi_{\text{Merced}}$). With this approach we attempted to mirror the polarized calculations of the PBS. As with PBS, we simulated nucleotide diversity for 100,000 5 kb windows under our demographic model. Then, we used a custom Python script to calculate empirical nucleotide diversity in overlapping 5 kb windows and a demographically controlled threshold for significance of 99.9% (see Supplemental Materials).

GO Enrichment

We performed functional enrichment analysis to test for significantly enriched gene sets and functional categories of genes within our two-way and three-way outlier sets that may reflect a history of altitude-related selection. For each *P. maniculatus* gene, we identified the orthologous *Mus musculus* gene, then used gProfiler (Reimand et al. 2007, 2011, 2016) to analyze enrichment for GO, biological pathways such as Kyoto Encyclopedia of Genes and Genomes (Kanehisa and Goto 2000), and Reactome (Jassal et al. 2019), regulatory motifs in DNA, protein databases, and human phenotype ontology (Reimand et al. 2016). We used strong hierarchical filtering (returning only the most general term per parent term) to identify enriched gene functional categories below a false discovery rate corrected significance of $P < 0.05$.

Clinal Patterns of Variation for Candidate Genes

HZAR fits genetic data to equilibrium cline models using an MCMC algorithm and estimates parameters such as the cline center (c) and width (w); c and w characterize the geographic location along the transect where the allele frequency turnover is greatest and the geographic region corresponding to the inverse of the maximum cline slope, respectively. The values of these parameters can be estimated within HZAR by 15 models that vary in the number of estimated parameters (e.g., exponential decay on either side of the cline center, minimum and/or maximum allele frequencies). For each of our two-way candidate loci, we identified the highest ranked SNP (first by occurrence in an outlier PBS window and then ranked by maximum RDA correlation) with a $\geq$75% call rate amongst our 165 unrelated individuals from the Rocky Mountain and Great Plains populations, then calculated that SNP's allele sample frequency for each population sampled across our elevational transect. For each of those SNPs, we modeled the cline shape parameters using HZAR and

determined the cline center and cline width. We used sampling elevation in meters as a proxy for geographic distance and used a burn-in of 10,000 iterations.

To set a neutral background expectation for clinal patterns of allele frequency variation, we performed a PCA on the set of 282,617 LD-pruned nongenic SNPs within PLINK. There is a clear pattern of geographic structure on both the first (PC1; east/west) and second principal component axes (PC2; north/south). Therefore, we used PC1 values for each individual to fit clines in HZAR, as has recently been done elsewhere (Hague et al. 2020). We used similar parameters as when generating the allele frequency clines, with appropriate modifications for trait data (e.g., avoiding models that set scaling to a fixed minimum and maximum of 0 and 1, respectively). Because principal components, especially those generated on geographically structured data, can be shaped by mathematical artifacts that can make interpretation unintuitive (Novembre and Stephens 2008), we visually confirmed that the cline fit to PC1 is representative of the clines fit to 20 SNPs with the highest loadings on PC1 (supplementary fig. S10, Supplementary Material online). We identified statistically discordant clines as those SNPs whose cline center or cline width CIs do not overlap with the CIs of the neutral PC1 cline.

Degree of Candidate Gene Overlap with Other High-Elevation Studies

To determine the degree of overlap in the genomic targets of selection across other similar studies, we performed a post hoc review of 14 studies in nine different species of terrestrial vertebrates (supplementary table S9, Supplementary Material online). Prior to spring 2020 (our last search date), there were to our knowledge 56 studies of high-altitude adaptation; we subsequently eliminated studies that were not population genomic comparisons, did not publish a complete list of outlier genes, did not use a comparable allele-frequency based test of selection, did not sample a highland population at a high enough elevation to be comparable, and/or focused on an ectotherm.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

We are grateful to Brenna Forester for advice on running RDA, to Graham Scott for access to deer mouse physiology data and helpful feedback on our manuscript, to members of the Cheviron laboratory and three anonymous reviewers for constructive comments on earlier versions of the manuscript, and to Trey Sasser for assistance in the field. This work was supported by the National Science Foundation (PRFB 1612859 to R.M.S., IOS-1755411 to Z.A.C., OIA 1736249 to Z.A.C. and J.F.S.) and the National Institute of Health (R35GM137919-01 to G.S.B. and NIH HL087216 to J.F.S.). This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH (S10RR029668, S10RR027303). Scientific collections were kindly provided as follows: ROMO-SCI-2017-0034 from Rocky Mountain National Park, collecting permit CLC772 from the U.S. Forest Service, and license no. 17TR2168b from Colorado Parks and Wildlife.

Data Availability

Raw sequence reads in demultiplexed fastq format for 156 deer mice are available on NCBI SRA (PRJNA719846), with data from 100 deer mice already available on NCBI SRA (PRJNA528923). Additional meta data and scripts are provided within the Supplementary Material online.

References

Barton NH. 1983. Multilocus clines. *Evolution*. 37(3):454–471.
Barton NH. 1979. Gene flow past a cline. *Heredity*. 43(3):333–339.
Berg JJ, Coop G. 2014. A population genetic signal of polygenic adaptation. PLoS Genet. 10:e1004412.
Bonnardeaux A, Davies E, Jeunemaitre X, Féry I, Charru A, Clauser E, Tiret L, Cambien F, Corvol P, Soubrier F. 1994. Angiotensin II type 1 receptor gene polymorphisms in human essential hypertension. *Hypertension* 24(1):63–69.
Bradburd GS, Coop GM, Ralph PL. 2018. Inferring continuous and discrete population genetic structure across space. *Genetics* 210(1):33–52.
Bradburd GS, Ralph PL. 2019. Spatial population genetics: it's about time. *Annu Rev Ecol Evol Syst*. 50(1):427–449.
Brayden JE, Nelson MT. 1992. Regulation of arterial tone by activation of calcium-dependent potassium channels. *Science* 256(5056):532–535.
Brown ST, Kelly KF, Daniel JM, Nurse CA. 2009. Hypoxia inducible factor (HIF)-2α is required for the development of the catecholaminergic phenotype of sympathoadrenal cells. *J Neurochem*. 110(2):622–630.
Burnstock G, Ralevic V. 2013. Purinergic signaling and blood vessels in health and disease. Pharmacol Rev. 66:102–192.
Cheviron ZA, Bachman GC, Storz JF. 2013. Contributions of phenotypic plasticity to differences in thermogenic performance between highland and lowland deer mice. *J Exp Biol*. 216(Pt 7):1160–1166.
Cheviron ZA, Bachman GC, Connaty AD, McClelland GB, Storz JF. 2012. Regulatory changes contribute to the adaptive enhancement of thermogenic capacity in high-altitude deer mice. *Proc Natl Acad Sci U S A*. 109(22):8635–8640.
Cheviron ZA, Connaty AD, McClelland GB, Storz JF. 2014. Functional genomics of adaptation to hypoxic cold-stress in high-altitude deer mice: transcriptomic plasticity and thermogenic performance. *Evolution* 68(1):48–62.
Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al.; 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. *Bioinformatics* 27(15):2156–2158.
Dawson NJ, Lyons SA, Henry DA, Scott GR. 2018. Effects of chronic hypoxia on diaphragm function in deer mice native to high altitude. *Acta Physiol* 223:e13030–16.
Derryberry EP, Derryberry GE, Maley JM, Brumfield RT. 2014. hzar: hybrid zone analysis using an R software package. *Mol Ecol Resour*. 14(3):652–663.
Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, Marquéz JRG, Gruber B, Lafourcade B, Leitão PJ, et al. 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36(1):27–46.
Endler JA. 1977. Geographic variation, speciation, and clines. *Monogr Popul Biol*. 10:1–246.
Ewing G, Hermisson J. 2010. MSMS: a coalescent simulation program including recombination, demographic structure and selection at a single locus. *Bioinformatics* 26(16):2064–2065.
Forester BR, LASKY JR, Wagner HH, Urban DL. 2018. Comparing methods for detecting multilocus adaptation with multivariate genotype-environment associations. *Mol Ecol*. 27:1–31.

Hague MTJ, Stokes AN, Feldman CR, Brodie ED, Brodie ED. 2020. The geographic mosaic of arms race coevolution is closely matched to prey population structure. *Evol Lett*. 4(4):317–332.

Hayes J, O'Connor C. 1999. Natural selection on thermogenic capacity of high-altitude deer mice. *Evolution* 53(4):1280–1287.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol*. 25(15):1965–1978.

Ivy CM, Scott GR. 2017. Control of breathing and ventilatory acclimatization to hypoxia in deer mice native to high altitudes. *Acta Physiol*. 112:123–17.

Ivy CM, Scott GR. 2018. Evolved changes in breathing and CO2 sensitivity in deer mice native to high altitudes. *American Journal of Physiology- Regulatory, Integrative and Comparative Physiology* 315:R1027–R1037.

Ivy CM, Greaves MA, Sangster ED, Robertson CE, Natarajan C, Storz JF, McClelland GB, Scott GR. 2020. Ontogenesis of evolved changes in respiratory physiology in deer mice native to high altitude. *J Exp Biol*. 223:jeb219360.

Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A, Sidiropoulos K, Cook J, Gillespie M, Haw R, et al. 2019. The reactome pathway knowledgebase. *Nucleic Acids Res*. 47:D596.

Jeong C, Witonsky DB, Basnyat B, Neupane M, Beall CM, Childs G, Craig SR, Novembre J, Di Rienzo A. 2018. Detecting past and ongoing natural selection among ethnically Tibetan women at high altitude in Nepal. PLoS Genet. 14:e1007650–29.

Jones MR, Mills LS, Alves PC, Callahan CM, Alves JM, Lafferty DJR, Jiggins FM, Jensen JD, Melo-Ferreira J, Good JM. 2018. Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. *Science* 360(6395):1355–1358.

Kanehisa M, Goto S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*. 28(1):27–30.

Knaus HG, Eberhart A, Koch RO, Munujos P, Schmalhofer WA, Warmke JW, Kaczorowski GJ, Garcia ML. 1995. Characterization of tissue-expressed alpha subunits of the high conductance Ca(2+)-activated K+ channel. *J Biol Chem*. 270(38):22434–22439.

Lau DS, Connaty AD, Mahalingam S, Wall N, Cheviron ZA, Storz JF, Scott GR, McClelland GB. 2017. Acclimation to hypoxia increases carbohydrate use during exercise in high-altitude deer mice. *Am J Physiol Regul Integr Comp Physiol*. 312(3):R400–R411.

Legendre P, Oksanen J, Braak ter CJF. 2011. Testing the significance of canonical axes in redundancy analysis. *Methods Ecol Evol*. 2(3):269–277.

Lenormand T. 2002. Gene flow and the limits to natural selection. *TREE* 17(4):183–189.

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

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16):2078–2079.

Li Y, Wu DD, Boyko AR, Wang GD, Wu SF, Irwin DM, Zhang Y-P. 2014. Population variation revealed high-altitude adaptation of Tibetan Mastiffs. *Mol Biol Evol*. 31(5):1200–1205.

Liu X, Zhang Y, Li Y, Pan J, Wang D, Chen W, Zheng Z, He X, Zhao Q, Pu Y, et al. 2019. *EPAS1* gain-of-function mutation contributes to high-altitude adaptation in Tibetan horses. *Mol Biol Evol*. 36(11):2591–2603.

Livneh B, Rosenberg EA, Lin C, Nijssen B, Mishra V, Andreadis K, Maurer E, Lettenmaier D. 2013. A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States: update and extensions. *J Climate* 26(23):9384–9392.

Lui MA, Mahalingam S, Patel P, Connaty AD, Ivy CM, Cheviron ZA, Storz JF, McClelland GB, Scott GR. 2015. High-altitude ancestry and hypoxia acclimation have distinct effects on exercise capacity and muscle phenotype in deer mice. *Am J Physiol Regul Integr Comp Physiol*. 308(9):R779–R791.

Mahalingam S, McClelland GB, Scott GR. 2017. Evolved changes in the intracellular distribution and physiology of muscle mitochondria in high-altitude native deer mice. *J Physiol*. 595(14):4785–4801.

Mahalingam S, Cheviron ZA, Storz JF, McClelland GB, Scott GR. 2020. Chronic cold exposure induces mitochondrial plasticity in deer mice native to high altitudes. *J Physiol* 598:5411–5426.

Mayer U, Saher G, Fässler R, Bornemann A, Echtermeyer F, Mark von der H, Miosge N, Pöschl E, Mark von der K. 1997. Absence of integrin alpha 7 causes a novel form of muscular dystrophy. *Nat Genet*. 17(3):318–323.

Moore WS. 1977. An evaluation of narrow hybrid zones in vertebrates. *Quart Rev Biol*. 52(3):263–277.

Nagylaki T. 1975. Conditions for the existence of clines. *Genetics* 80(3):595–615.

Natarajan C, Hoffmann FG, Lanier HC, Wolf CJ, Cheviron ZA, Spangler ML, Weber RE, Fago A, Storz JF. 2015. Intraspecific polymorphism, interspecific divergence, and the origins of function-altering mutations in deer mouse hemoglobin. *Mol Biol Evol*. 32(4):978–997.

Nikel KE, Shanishchara NK, Ivy CM, Dawson NJ, Scott GR. 2018. Effects of hypoxia at different life stages on locomotory muscle phenotype in deer mice native to high altitudes. *Comparative Biochemistry and Physiology, Part B* 224:98–104.

Novembre J, Stephens M. 2008. Interpreting principal component analyses of spatial population genetic variation. *Nat Genet*. 40(5):646–649.

Polechová J, Barton NH. 2015. Limits to adaptation along environmental gradients. *Proc Natl Acad Sci U S A*. 112(20):6401–6406.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 81(3):559–575.

Qu Y, Tian S, Han N, Zhao H, Gao B, Fu J, Cheng Y, Song G, Ericson PGP, Zhang YE, et al. 2015. Genetic responses to seasonal variation in altitudinal stress: whole-genome resequencing of great tit in eastern Himalayas. *Scientific Reports* 5:14256.

Reimand J, Arak T, Adler P, Kolberg L, Reisberg S, Peterson H, Vilo J. 2016. g: profiler—a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Res*. 44(W1):W83–W89.

Reimand J, Arak T, Vilo J. 2011. g: profiler–a web server for functional interpretation of gene lists (2011 update). *Nucleic Acids Res*. 39(Web Server issue):W307–W315.

Reimand J, Kull M, Peterson H, Hansen J, Vilo J. 2007. g: profiler–a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res*. 35(Web Server issue):W193–W200.

Riesch R, Plath M, Bierbach D. 2018. Ecology and evolution along environmental gradients. *Curr Zool*. 64(2):193–196.

Schweizer RM, Robinson J, Harrigan RJ, Silva P, Galverni M, Musiani M, Green RE, Novembre J, Wayne RK. 2016. Targeted capture and resequencing of 1040 genes reveal environmentally driven functional variation in grey wolves. *Mol Ecol*. 25(1):357–379.

Schweizer RM, Velotta JP, Ivy CM, Jones MR, Muir SM, Bradburd GS, Storz JF, Scott GR, Cheviron ZA. 2019. Physiological and genomic evidence that selection on the transcription factor *Epas1* has altered cardiovascular function in high-altitude deer mice. *PLoS Genet*. 15(11):e1008420.

Scott AL, Pranckevicius NA, Nurse CA, Scott GR. 2019. Regulation of catecholamine release from the adrenal medulla is altered in deer mice (*Peromyscus maniculatus*) native to high altitudes. *Am J Physiol Regul Integr Comp Physiol*. 317(3):R407–R417.

Scott GR, Elogio TS, Lui MA, Storz JF, Cheviron ZA. 2015. Adaptive modifications of muscle phenotype in high-altitude deer mice are associated with evolved changes in gene regulation. *Mol Biol Evol*. 32(8):1962–1976.

Simonson TS, Yang Y, Huff CD, Yun H, Qin G, Witherspoon DJ, Bai Z, Lorenzo FR, Xing J, Jorde LB, et al. 2010. Genetic evidence for high-altitude adaptation in Tibet. *Science* 329(5987):72–75.

Slatkin M. 1987. Gene flow and the geographic structure of natural populations. *Science* 236(4803):787–792.

Sluyter R. 2015. P2X and P2Y receptor signaling in red blood cells. *Front Mol Biosci*. 2:60–287.
Song D, Bigham AW, Lee FS. 2021. High-altitude deer mouse hypoxia-inducible factor-2α shows defective interaction with CREB-binding protein. *J Biol Chem*. 296:100461.
Song S, Yao N, Yang M, Liu X, Dong K, Zhao Q, Pu Y, He X, Guan W, Yang N, et al. 2016. Exome sequencing reveals genetic differentiation due to high-altitude adaptation in the Tibetan cashmere goat (*Capra hircus*). *BMC Genomics* 17(1):12.
Stankowski S, Sobel JM, Streisfeld MA. 2016. Geographic cline analysis as a tool for studying genome-wide variation: a case study of pollinator-mediated divergence in a monkeyflower. Mol Ecol. 26:107–122.
Staples J, Nickerson DA, Below JE. 2013. Utilizing graph theory to select the largest set of unrelated individuals for genetic analysis. *Genet Epidemiol*. 37(2):136–141.
Storfer A, Patton A, Fraik AK. 2018. Navigating the interface between landscape genetics and landscape genomics. *Front Genet*. 9:68–14.
Storz JF, Scott GR. 2019. Life Ascending: Mechanism and Process in Physiological Adaptation to High-Altitude Hypoxia. *Annu Rev Ecol Evol Syst*. 50:503–526.
Storz JF, Cheviron ZA, McClelland GB, Scott GR. 2019. Evolution of physiological performance capacities and environmental adaptation: insights from high-elevation deer mice (*Peromyscus maniculatus*). *J Mammal*. 100(3):910–922.
Storz JF, Kelly JK. 2008. Effects of spatially varying selection on nucleotide diversity and linkage disequilibrium: insights from deer mouse globin genes. *Genetics* 180(1):367–379.
Storz JF, Natarajan C, Cheviron ZA, Hoffmann FG, Kelly JK. 2012. Altitudinal variation at duplicated-globin genes in deer mice: effects of selection, recombination, and gene conversion. *Genetics* 190(1):203–216.
Storz JF, Runck AM, Moriyama H, Weber RE, Fago A. 2010. Genetic differences in hemoglobin function between highland and lowland deer mice. *J Exp Biol*. 213(Pt 15):2565–2574.
Storz JF, Runck AM, Sabatino SJ, Kelly JK, Ferrand N, Moriyama H, Weber RE, Fago A. 2009. Evolutionary and functional insights into the mechanism underlying high-altitude adaptation of deer mouse hemoglobin. *Proc Natl Acad Sci U S A*. 106(34):14450–14455.
Storz JF. 2021. High-altitude adaptation: mechanistic insights from integrated genomics and physiology. *Mol Biol Evol*. doi: 10.1093/molbev/msab064.
Storz JF, Cheviron ZA. 2021. Physiological Genomics of Adaptation to High-Altitude Hypoxia. *Annu. Rev. Anim. Biosci.* 9:149–171.
Tate KB, Ivy CM, Velotta JP, Storz JF, McClelland GB, Cheviron ZA, Scott GR. 2017. Circulatory mechanisms underlying adaptive increases in thermogenic capacity in high-altitude deer mice. *J Exp Biol*. 220:3616–3620.
Tate KB, Wearing OH, Ivy CM, Cheviron ZA, Storz JF, McClelland GB, Scott GR. 2020. Coordinated changes across the O_2 transport pathway underlie adaptive increases in thermogenic capacity in high-altitude deer mice. *Proc Biol Sci*. 287(1927):20192750–20192759.
Van Den Wollenberg AL. 1977. Redundancy analysis an alternative for canonical correlation analysis. *Psychometrika* 42(2):207–219.
Velotta JP, Ivy CM, Wolf CJ, Scott GR, Cheviron ZA. 2018. Maladaptive phenotypic plasticity in cardiac muscle growth is suppressed in high-altitude deer mice. *Evolution* 72(12):2712–2716.
Velotta JP, Jones J, Wolf CJ, Cheviron ZA. 2016. Transcriptomic plasticity in brown adipose tissue contributes to an enhanced capacity for nonshivering thermogenesis in deer mice. *Mol Ecol*. 25(12):2870–2886.
Velotta JP, Robertson CE, Schweizer RM, McClelland GB, Cheviron ZA. 2020. Adaptive shifts in gene regulation underlie a developmental delay in thermogenesis in high-altitude deer mice. *Mol Biol Evol*. 37(8):2309–2321.
Wall JD, Cox MP, Mendez FL, Woerner A, Severson T, Hammer MF. 2008. A novel DNA sequence database for analyzing human demographic history. *Genome Res*. 18(8):1354–1361.
Wearing OH, Ivy CM, Guttierrez-Pinto N, Velotta JP, Campbell-Staton SC, Natarajan C, Cheviron ZA, Storz JF, Scott GR. 2020. The adaptive benefit of increases in hemoglobin-O2 affinity is contingent on tissue O2 diffusing capacity in high-altitude deer mice. biorxiv.org
West CM, Ivy CM, Husnudinov R, Scott GR. 2021. Evolution and developmental plasticity of lung structure in high-altitude deer mice. J Comp Physiol B, *Biochem Syst Environ Physiol*:1–12.
Wilde LR, Wolf CJ, Porter SM, Stager M, Cheviron ZA, Senner NR. 2019. Botfly infections impair the aerobic performance and survival of montane populations of deer mice, *Peromyscus maniculatus rufinus*. *Funct Ecol*. 33(4):608–618.
Witt KE, Huerta-Sanchez E. 2019. Convergent evolution in human and domesticate adaptation to high-altitude environments. *Philosophical Transactions of the Royal Society B: Biological Sciences* 374:20180235.
Wu C, Lu H, Cassis LA, Daugherty A. 2011. Molecular and pathophysiological features of angiotensinogen: a mini review. *N Am J Med Sci (Boston)* 4(4):183–190.
Yeaman S, Whitlock MC. 2011. The genetic architecture of adaptation under migration-selection balance. *Evolution* 65(7):1897–1911.
Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZXP, Pool JE, Xu X, Jiang H, Vinckenbosch N, Korneliussen TS, et al. 2010. Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* 329:75–78.
Zhang W, Fan Z, Han E, Hou R, Zhang L, Galaverni M, Huang J, Liu H, Silva P, Li P, et al. 2014. Hypoxia adaptations in the Grey Wolf (*Canis lupus chanco*) from Qinghai-Tibet Plateau. PLoS Genet. 10:e1004466.