OXFORD GENETICS

GENETICS, 2022, 220(2), iyab226
https://doi.org/10.1093/genetics/iyab226
Advance Access Publication Date: 13 December 2021
Investigation
Highlighted Article

The genomic basis of high-elevation adaptation in wild house mice (*Mus musculus domesticus*) from South America

Elizabeth J. Beckman [1,*,†] Felipe Martins [1,†] Taichi A. Suzuki,[1,2] Ke Bi,[1] Sara Keeble [3] Jeffrey M. Good [3,4] Andreas S. Chavez [1,5] Mallory A. Ballinger,[1] Kennedy Agwamba,[1] and Michael W. Nachman [1,*]

[1]Museum of Vertebrate Zoology and Department of Integrative Biology, University of California, Berkeley, Berkeley, CA 94720, USA,
[2]Department of Microbiome Science, Max Planck Institute for Developmental Biology, Tübingen 72076, Germany,
[3]Division of Biological Sciences, University of Montana, Missoula, MT 59812, USA,
[4]Wildlife Biology Program, University of Montana, Missoula, MT 59812, USA, and
[5]Department of Evolution, Ecology, and Organismal Biology and the Translational Data Analytics Institute, The Ohio State University, Columbus, OH 43210, USA

*Corresponding author: Museum of Vertebrate Zoology, University of California, Berkeley, 3101 Valley Life Sciences Building, Berkeley, CA 94720, USA.
Email: libby.beckman@berkeley.edu (E.J.B.); mnachman@berkeley.edu (M.W.N.)
Reference number for public data: Sequence Read Archive BioProject ID PRJNA766703.
†These authors contributed equally to this work.

Abstract

Understanding the genetic basis of environmental adaptation in natural populations is a central goal in evolutionary biology. The conditions at high elevation, particularly the low oxygen available in the ambient air, impose a significant and chronic environmental challenge to metabolically active animals with lowland ancestry. To understand the process of adaptation to these novel conditions and to assess the repeatability of evolution over short timescales, we examined the signature of selection from complete exome sequences of house mice (*Mus musculus domesticus*) sampled across two elevational transects in the Andes of South America. Using phylogenetic analysis, we show that house mice colonized high elevations independently in Ecuador and Bolivia. Overall, we found distinct responses to selection in each transect and largely nonoverlapping sets of candidate genes, consistent with the complex nature of traits that underlie adaptation to low oxygen availability (hypoxia) in other species. Nonetheless, we also identified a small subset of the genome that appears to be under parallel selection at the gene and SNP levels. In particular, three genes (*Col22a1*, *Fgf14*, and *srGAP1*) bore strong signatures of selection in both transects. Finally, we observed several patterns that were common to both transects, including an excess of derived alleles at high elevation, and a number of hypoxia-associated genes exhibiting a threshold effect, with a large allele frequency change only at the highest elevations. This threshold effect suggests that selection pressures may increase disproportionately at high elevations in mammals, consistent with observations of some high-elevation diseases in humans.

Keywords: high elevation; adaptive evolution; parallelism; Andes; *Mus musculus domesticus*

Introduction

A key problem in evolutionary genetics is to understand how organisms adapt to novel environments. High elevation environments present a suite of challenges to metabolically active animals, including low oxygen availability in ambient air, increased UV radiation exposure, and colder temperatures. In particular, the decrease of the partial pressure of oxygen (P_{O2}) with increasing elevation is a strong environmental stressor (Storz *et al.* 2010; Storz and Scott 2019). Animals with lowland ancestry that colonize high elevations cannot avoid these difficult conditions. Many species that live at high elevations have evolved genetic adaptations (Beall *et al.* 2010; Simonson *et al.* 2010; Yi *et al.* 2010; Alkorta-Aranburu *et al.* 2012; Qiu *et al.* 2012; Ge *et al.* 2013; Lorenzo *et al.* 2014; Liu *et al.* 2019; Ma *et al.* 2019; Richardson *et al.* 2019; Schweizer *et al.* 2019; Signore and Storz 2020) and exhibit unique physiological traits affecting the oxygen transport system (reviewed in Storz and Scott 2019), the nervous system (Hendrickson 2013; Ai *et al.* 2014; Velotta *et al.* 2020), skin color (Ai *et al.* 2014), body size, and metabolism (Gering *et al.* 2009; Cheviron *et al.* 2012; Cheviron *et al.* 2014). Thus, analysis of the molecular signature of adaptation in recent high elevation colonists can provide insight into the process of adaptation to a novel environment.

There is a considerable body of literature on high elevation adaptation, notably but not exclusively in humans and domesticated animals (reviewed in Storz and Scott 2019; Witt and Huerta-Sánchez 2019; Storz 2021), and several broad patterns have emerged. First, physiological and genetic adaptations to high elevation can occur over short timescales (~200–1600 generations; Beall 2006, 2007; Witt and Huerta-Sánchez 2019; Wu *et al.* 2020). Second, the phenotypes exhibited by recent high-elevation colonists are often complex and may involve both genetic and plastic responses (Monge and León-Velarde 1991; Cheviron *et al.* 2012; Childebayeva *et al.* 2021). Third, genetic studies have revealed that genes that regulate or are regulated by the hypoxia

Received: September 29, 2021. **Accepted:** December 04, 2021
© The Author(s) 2021. Published by Oxford University Press on behalf of Genetics Society of America. All rights reserved.
For permissions, please email: journals.permissions@oup.com

Downloaded from https://academic.oup.com/genetics/article/220/2/iyab226/6460342 by University of California Berkeley user on 07 February 2022

inducible factor (HIF) pathway often play a large role in high-elevation adaptation (Scheinfeldt *et al.* 2012; Foll *et al.* 2014; Eichstaedt *et al.* 2014). For example, *Egln1*, which regulates the HIF pathway, and *Epas1*, which codes for the hypoxia inducible factor α2, bear a signature of selection in several different high elevation lineages (Bigham *et al.* 2010; Peng *et al.* 2011; Xing *et al.* 2013; Eichstaedt *et al.* 2017; Schweizer *et al.* 2019; Szpiech *et al.* 2020; Wu *et al.* 2020; reviewed in Witt and Huerta-Sánchez 2019; Pamenter *et al.* 2020). At face value, this suggests that changes at these genes may reflect convergent evolution driven by associations with a common phenotype that is under selection at high elevation in different lineages. However, Storz (2021) points out that underlying causal mechanisms may be radically different even when genic targets are shared. For example, *Epas1* is under selection in both humans and deer mice at high elevations, yet the phenotypic associations with *Epas1* variants are different in these species (Schweizer *et al.* 2019; Storz 2021). Fourth, many genes that are under selection associated with high elevation are unique and population-specific, even in closely related lineages that independently colonized high elevation (Bigham *et al.* 2010; Alkorta-Aranburu *et al.* 2012; Scheinfeldt *et al.* 2012; Yang *et al.* 2017; Jacovas *et al.* 2018). Finally, studies on humans and deer mice indicate that the genetic basis of high elevation adaptation is highly polygenic (Jeong *et al.* 2018; Velotta *et al.* 2020), consistent with the idea that many systems interact to optimize oxygen transport and use within the body (Monge and León-Velarde 1991).

There have been comparatively fewer studies of natural populations of nondomestic mammals, with several notable exceptions, such as work on deer mice (Cheviron *et al.* 2014; Storz *et al.* 2019), artiodactyls (Qiu *et al.* 2012; Guang-Xin *et al.* 2019; Signore and Storz 2020) and Old World monkeys (Yu *et al.* 2016; Szpiech *et al.* 2020). There are also few studies that characterize the genomic basis of adaptation to high elevation in multiple closely related lineages with independent colonization histories. With the important exception of humans on the Qinghai-Tibetan Plateau, the Ethiopian highlands and the Andean Altiplano, and a few domestic taxa, current assessment of convergent adaptation to high elevation in mammals come from lineages that are separated by several million years (Storz 2016; Witt and Huerta-Sánchez 2019; Wu *et al.* 2020). Finally, it is unclear if there is a threshold elevation above which selection pressures in mammals increase disproportionately. Though P_{O2} decreases roughly linearly with increasing elevation, there are several human diseases including high altitude pulmonary edema, acute mountain sickness, and Sudden Infant Death Syndrome where the disease incidence in individuals with lowland ancestry increases significantly above ∼2000 m (Hackett and Roach 2001; Gabry *et al.* 2003; West 2004; Katz *et al.* 2015). Pulmonary artery pressure and oxygen saturation in blood show an exponential relationship with elevation in lowland humans and change rapidly above 3000 m (Penaloza and Arias-Stella 2007). These results suggest a nonlinear relationship between selection pressure and elevation, but few studies have the intraspecific sampling across different elevations to test this hypothesis.

Here, we examine the molecular signature of selection associated with high elevation environments in two transects of wild house mice, *Mus musculus domesticus*, on the slopes of the Andes in South America. Originating on the Indian subcontinent, *M. musculus* is a human commensal that is native to Europe and Asia and mostly found at low elevations. *M. m. domesticus* expanded its range in historic times through its association with humans, arriving in the Americas with European colonists. The exact timing of the establishment of stable populations in the Americas is unclear, but likely occurred several hundred years ago, corresponding to 400–600 generations or more (Phifer-Rixey *et al.* 2018). As house mice colonized new regions, they have adapted to many diverse environments (Lynch 1992; Phifer-Rixey *et al.* 2018; Ferris *et al.* 2021). On the Andean Altiplano in South America, house mice have been found up to ∼4000 m (Storz *et al.* 2007). House mice demonstrate a strong acute response to hypoxia (Dzal and Milsom 2019) and characteristic physiological changes when reared under chronic hypoxic conditions (Jochmans-Lemoine *et al.* 2015). Previous work on house mice in the Andes compared one high elevation population from Bolivia and one low elevation population from Peru and focused on alpha and beta hemoglobin genes. That study found no evidence for adaptation to high elevation involving these genes (Storz *et al.* 2007), leaving open the larger question of genome-wide signatures of selection.

To assess genome-wide signatures of selection associated with high elevation, we sampled 96 wild-caught house mice from 10 populations across two elevational transects from sea level to over 3000 m in the Andes. One transect was in Ecuador and the other transect was mainly in Bolivia. These two transects are separated by more than 2000 km. We sequenced the complete exomes of all individuals and addressed four primary questions. (1) What are the evolutionary relationships of house mice in the Andes? (2) What genes underlie adaptation to high elevation in house mice? (3) Does adaptation to high elevation in the two regions show convergence at the SNP, gene or pathway levels? (4) Do some genes exhibit a "threshold effect," where changes in allele frequency in response to selection are only observed above a threshold elevation? We found that mice colonized high elevations independently in Ecuador and Bolivia, and that in each case many genes throughout the genome have responded to selection. Consistent with the multifactorial and highly polygenic nature of high-elevation adaptation, we found little evidence for broad convergence in underlying genes; however, we identified a small set of genes that appear to be under parallel selection at the gene and SNP levels at high elevation. Furthermore, we identified a subset of hypoxia-associated genes that exhibited a threshold effect, showing a dramatic change in allele frequency only at the highest elevations.

Materials and methods

Sampling

We sampled wild *M. musculus domesticus* in two regions of South America (Figure 1; Supplementary File S1) as described by Suzuki *et al.* (2019). On the west slope of the Andes in Ecuador, hereafter the northern transect, we collected 50 mice from five elevations with nearly even numbers of males and females, corresponding approximately to: ∼50 m ($n = 11$), ∼400 m ($n = 10$), ∼1600 m ($n = 9$), ∼2500 m ($n = 10$), and ∼2900 m ($n = 10$). See Figure 2 for elevation ranges for each population and Supplementary File S1 for complete details. On the east slope of the Andes in Bolivia and Brazil, hereafter the southern transect, we collected 46 mice from five elevations: ∼90 m ($n = 10$), ∼300 m ($n = 6$), ∼2600 m ($n = 12$), ∼3300 m ($n = 8$), and ∼3800 m ($n = 10$). These sample sizes are sufficient to provide reasonable estimates of nucleotide diversity for each population (Pluzhnikov and Donnelly 1996). To minimize relatedness among individuals and avoid uneven sampling of ephemeral, local demes, each mouse was collected at least 500 m from every other mouse (Pocock *et al.* 2005; Harr *et al.* 2016). All mice were handled in accordance with a protocol approved by

Downloaded from https://academic.oup.com/genetics/article/220/2/iyab226/6460342 by University of California Berkeley user on 07 February 2022

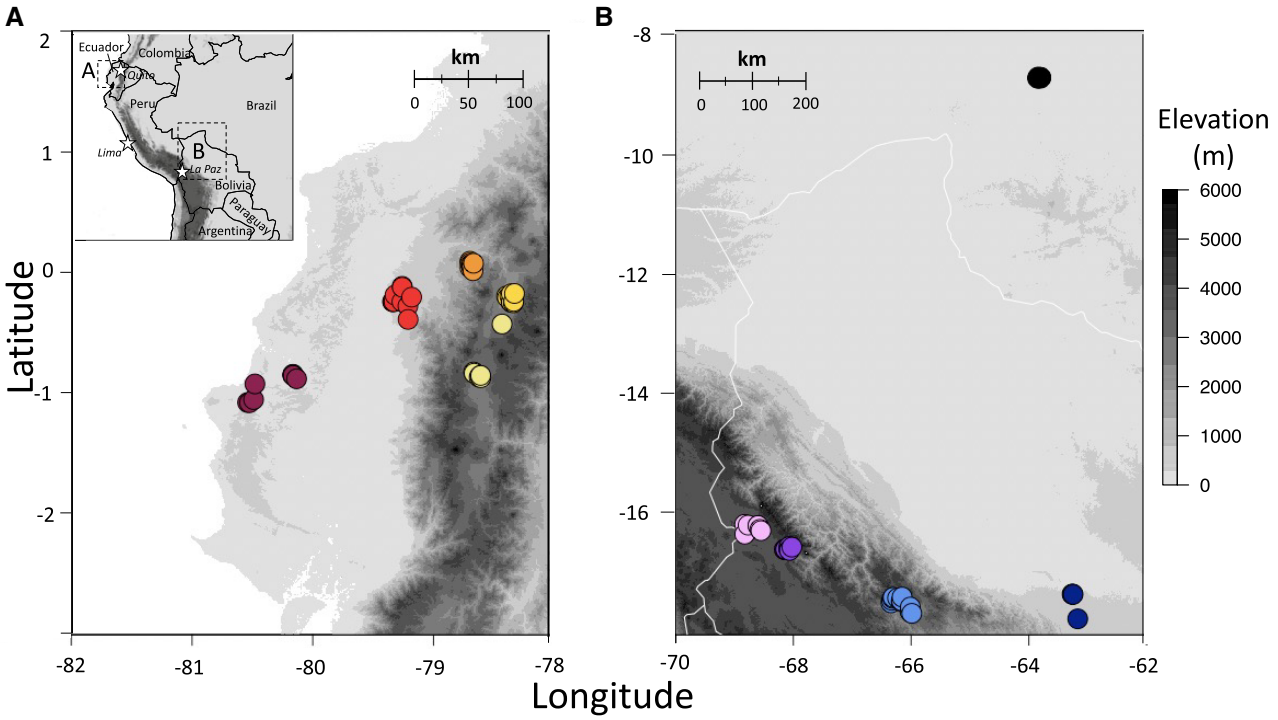

Figure 1 Sampling map of *Mus musculus domesticus*. Circles indicate sampling localities for elevational transects in (A). Northern transect ($n = 50$) and (B) Southern transect ($n = 46$). Inset shows regions of western South America depicted in (A) and (B) with major cities indicated by white stars. See Figure 2 for color legend indicating elevational range.

the University of California, Berkeley Institutional Animal Care and Use Committee. We preserved tissue in 100% ethanol or liquid nitrogen, and prepared study specimens which were deposited at the Museum of Vertebrate Zoology (MVZ) at UC Berkeley (MVZ catalog numbers are given in Supplementary File S1).

Sequencing

We extracted DNA from frozen liver tissue with the Qiagen DNeasy Blood and Tissue kit (Qiagen) for all individuals except samples from Santa Cruz, Bolivia, which were handled separately as described below. We followed Meyer and Kircher (2010) with the modifications recommended in Bi *et al.* (2013) to prepare individually barcoded libraries for each sample. We pooled 16–19 individuals in equimolar proportions and enriched the libraries for exonic regions using the SeqCap EZ Developer Library: Mouse Exome kit (Roche Inc.). Individuals from the northern transect ($n = 50$) were sequenced on three lanes of Illumina HiSeq 2000 (100 paired-end) with 16–17 individuals per lane. Forty individuals from the southern transect were sequenced on two lanes of Illumina HiSeq 4000 (150 paired-end).

For the six individuals from Santa Cruz, Bolivia, we extracted DNA from frozen tissue using the MasterPure™ Complete DNA and RNA Purification Kit (Lucigen). We prepared individually barcoded libraries using the Kapa HyperPrep kit (Roche Inc), and pooled the libraries with *M. m. domesticus* samples from another project for equimolar pools of 18–19 individuals. We enriched the pools for exonic regions as above. We sequenced 56 *M. m. domesticus* samples, including six for this project, on one lane of Illumina NovaSeq S4 (150 paired-end). Sequencing for all samples was done at the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley.

We used FastP (Chen *et al.* 2018) to clean raw sequence reads and aligned them to the *M. musculus* reference genome (GRCm38) with BWA (v 0.7.13, Li 2013) using the BWA-MEM algorithm. We indexed aligned reads in Samtools (v 1.9, Li *et al.* 2009). We used the Genome Analysis Toolkit (v 4.0.11.0, Van der Auwera *et al.* 2013) to merge data across lanes, remove duplicates and recalibrate base quality. We retained properly paired reads that uniquely matched to the reference for further analysis. We performed these tasks using the Extreme Science and Engineering Discovery Environment (XSEDE) Comet (Towns *et al.* 2014).

We used Angsd (v0.922, Korneliussen *et al.* 2014) to generate genotype likelihoods and to call variants. We employed several filters to produce high confidence variant sites given the modest coverage of 15–20× per locus per individual. At the transect level, we excluded sites with exceptionally high average coverage (over three times the average total depth) or low coverage (under five reads average per individual) (after Bi *et al.* 2013). At the population level, we removed sites that were significantly out of Hardy-Weinberg equilibrium (P-value < 0.001), and we required at least 3× coverage in eight or more individuals per population per locus, except for the Santa Cruz population where we required all six individuals. For the X chromosome, we ran males and females separately in Angsd, defined heterozygous genotypes in males as missing, and combined genotype data from the sexes with custom scripts. Last, we eliminated tandem repeats or pseudogenes in the *M. musculus* genome with bedtools (v2.27.1, Quinlan and Hall 2010).

Population structure analyses

We assessed population structure among South American house mice using autosomal SNPs in several ways. First, we performed principal component analyses on genetic variation using PCAngsd (v 0.99, Meisner and Albrechtsen 2018) for all 96 samples and for each transect separately. Second, we estimated a maximum likelihood phylogenetic tree with RAxML (v 8.2.12,

Downloaded from https://academic.oup.com/genetics/article/220/2/iyab226/6460342 by University of California Berkeley user on 07 February 2022

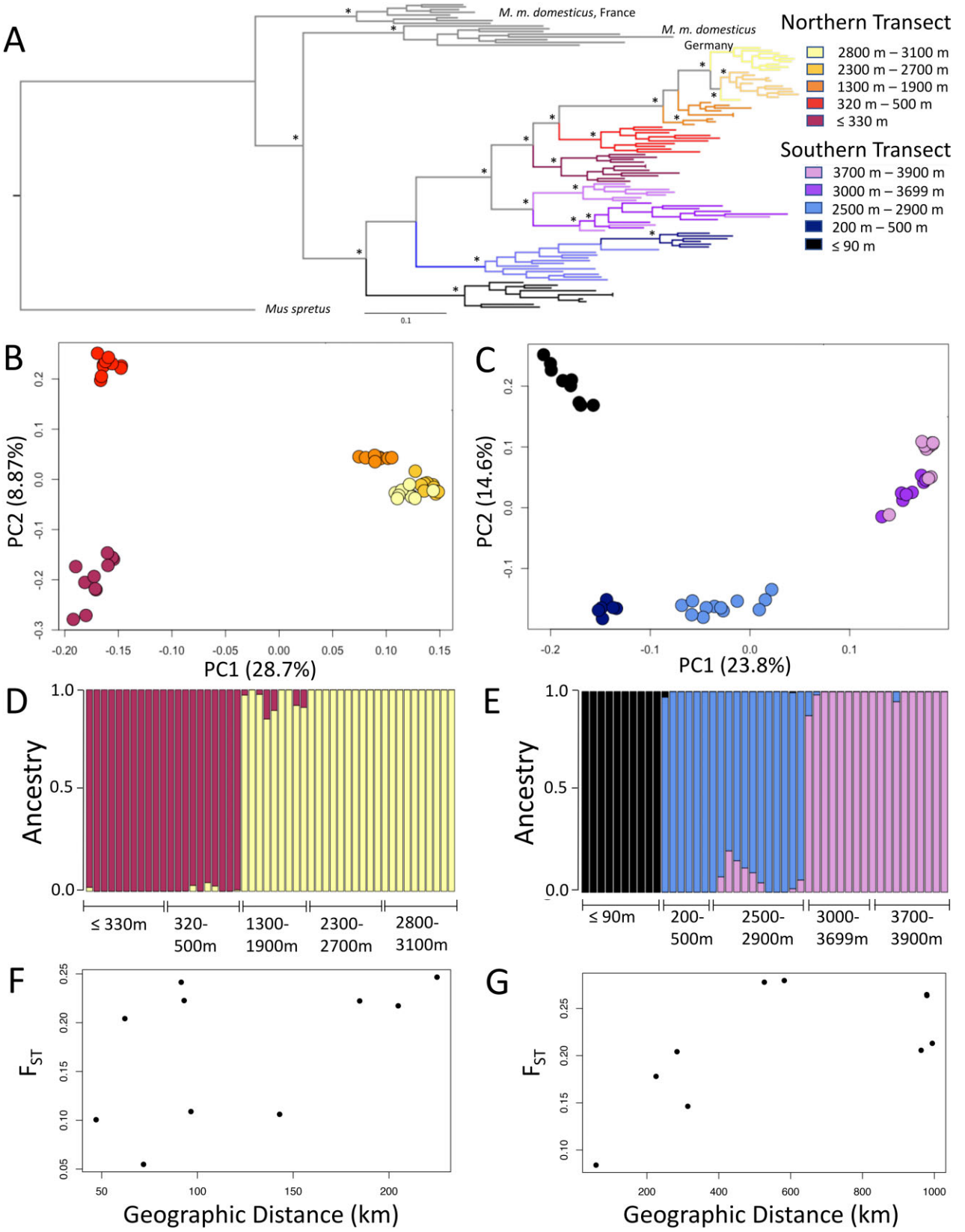

Figure 2 Population structure of *M. m. domesticus* in northern and southern transects. (A) Maximum likelihood phylogenetic tree of 96 individuals estimated with 104,044 SNPs. Bootstrap support (bs) is shown at key nodes; * indicates bs ≥ 97. Legend for elevation at right. (B, C) Principal component analyses of northern and southern transects, respectively. See (A) for legend. (D, E) Admixture analysis of northern (D) and southern (E) transects. (F, G) Genetic differentiation (F_{ST}) by geographic distance in kilometers for northern (F) and southern (G) transects.

Downloaded from https://academic.oup.com/genetics/article/220/2/iyab226/6460342 by University of California Berkeley user on 07 February 2022

Stamatakis 2014) with the GTRCAT model and rapid bootstrapping using sites present in 80% of individuals in a population. For this analysis, we included our samples as well as *M. m. domesticus* from France ($n = 8$), *M. m. domesticus* from Germany ($n = 8$), and *Mus spretus* ($n = 1$) as an outgroup (Harr *et al.* 2016). Third, we used Admixture (v 1.3.0, Alexander and Lange 2011) to estimate ancestry proportions for all our samples and for each respective transect after we pruned autosomal biallelic SNPs for linkage disequilibrium with PLINK (v 1.90, Chang *et al.* 2015) using nonoverlapping 50 kb windows and an r^2 threshold of 0.5. For each transect, we varied the number of populations from 1 to 5 ($K = 1$, 2, 3, 4, 5), and considered the K with the lowest cross-validation error as the best fit to the data.

For the remaining analyses, we used the maximum number of unrelated individuals (Supplementary File S1), defined as individuals without a first degree relationship in Primus (v 1.9.0, Staples *et al.* 2013) and kinship under ~0.2 in ngsRelate (v 2, Hanghøj *et al.* 2019).

We characterized the genetic variation within and among populations in each transect by calculating nucleotide diversity and pairwise F_{ST} in Angsd. We estimated the proportion of segregating sites (Watterson's θ) and nucleotide diversity (π) per site for coding regions and introns in each population, and calculated the unweighted average pairwise F_{ST} (Reynolds *et al.* 1983). We also calculated the average π in 100 kb sliding windows (25 kb step) across the genome. For each transect, we tested for significant isolation by distance (IBD) with a Mantel test (R package ade4 v 1.7, Dray and Dufour 2007) in R (v 3.6.0, R Core Team 2019). Distances were calculated from straight lines between points based on latitude and longitude. We tested for IBD in the southern transect with and without the most distant population, Porto-Velho, Brazil, to see if this population had an outsized impact on the results.

Autosomal selection scans

We tested for selection associated with high elevation on autosomes using two methods, Latent Factor Mixed Models (LFMM) (Frichot *et al.* 2013; Caye *et al.* 2019) and the modified Population Branch Statistic, PBS*n1* (Crawford *et al.* 2017). The goal of Latent Factor Mixed Models (LFMM) is to identify SNPs that show significant correlations with a specified environmental variable once population history is taken into account; the method incorporates population structure and unidentified demographic processes as latent factors (Caye *et al.* 2019). The modified Population Branch Statistic, PBS*n1* (Crawford *et al.* 2017), allowed us to identify high elevation population-specific allele frequency change. For these analyses, we required any variant to be present with a minor allele frequency of at least 5% across the transect and included only unrelated individuals. For both methods, we incorporated all SNPs that passed the filters specified above, rather than pruning SNPs that were in linkage disequilibrium. This was done to include all SNPs that could be of biological interest.

We performed LFMM using the package lfmm (v 0.0, Caye *et al.* 2019) in R (R Core Team 2019) with elevation as the environmental variable. We performed LFMM analyses using populations from all elevations for each transect alone and for the two transects combined. To account for demographic history among populations, we evaluated the effect of using a different number of latent factors ($K = 2$–5 for each transect, $K = 2$–10 for the combined data), chose the K that best summarized the population structure results following Caye *et al.* (2019), and estimated the genomic inflation factor over 50 independent runs. To correct for multiple testing, we used an optimized false discovery rate (FDR) approach (Benjamini and Hochberg 1995) and defined significance for the LFMM analysis at two thresholds: $q = 0.05$ and $q = 0.01$. To approximate the size of the region impacted by selection, we chose the SNP with the highest *P*-value in an LFMM candidate gene and searched upstream and downstream from the focal SNP in 2-kb adjacent windows for SNPs with a signature of selection as in Phifer-Rixey *et al.* (2018), skipping windows with no information.

We calculated PBS*n1*, the Population Branch Statistic (PBS, Yi *et al.* 2010) modified to avoid artificially high values when differentiation is high among populations (Crawford *et al.* 2017). We used PBS*n1* rather than PBS because of the high F_{ST} values observed among populations (see Results). We conducted PBS*n1* analyses for the northern and southern transects separately. For each analysis, we defined the highest-elevation population in the transect of interest as the target, and used the lowest populations in each transect as the other two populations. After estimating F_{ST} per site as described above, we calculated PBS*n1* using custom Python (v 3.7.7) scripts. We identified candidate SNPs as the top 5%, top 1%, and top 0.1% of all SNPs. We approximated the size of the region impacted by selection for the top 1% of PBS*n1* by searching outward in 2 kb increments starting at the highest PBS*n1* value in a candidate gene as above (Phifer-Rixey *et al.* 2018).

The power and false positive rate for both LFMM and PBS have been studied under various demographic and selective models in the original publications (Yi *et al.* 2010; Frichot *et al.* 2013; Caye *et al.* 2019). In general, both of these methods perform well, although the details depend on the exact demographic or selective scenarios. A full demographic model for house mice in the Andes is beyond the scope of this study; however, we conducted neutral simulations under two different models to assess the false positive rate for PBS*n1*. Both models incorporated divergence without gene flow or selection. One model was based on constant population sizes and the other model incorporated bottlenecks. Parameter values for these models were chosen to match the levels of heterozygosity and differentiation observed in the actual data (see details in Supplementary File S2).

We also examined the difference in nucleotide diversity between the highest and lowest populations in each transect in 100 kb nonoverlapping windows. For each population, we estimated per site nucleotide diversity in Angsd and averaged the nucleotide diversity for each 100 kb window. We reasoned that, given a selective sweep associated with high elevation, the nucleotide diversity of a genomic region at high elevation should be lower than the nucleotide diversity of the same region at low elevation. Thus, $\pi_{low} - \pi_{high} = \pi_{diff}$ may be positive and of greatest magnitude in regions that have undergone selection with respect to high elevation. We identified the top 1% of π_{diff} across windows that contained 500 or more sites with adequate coverage.

Each selection scan (LFMM or PBS*n1*) performed on house mice from the northern or southern transect produced a set of candidate SNPs and candidate genes. We used the variant effect predictor (VEP) for *M. musculus* (v 101, McLaren *et al.* 2016) to characterize candidate SNPs as: (1) nonsynonymous mutations, (2) in 3′ or 5′ untranslated regions (UTRs), (3) synonymous mutations and (4) other. We used bootstrapping to ask if the proportion of each category (1–4) in the candidate SNPs was the same when compared to all the SNPs from each transect. Next, defining the allelic state of *M. spretus* as ancestral, we assigned the alleles of candidate SNPs as ancestral or derived in VCFtools (v 0.1.16, Danecek *et al.* 2011). We tested if the candidate SNPs in a transect were equally likely to be ancestral or derived at high elevation using a Fisher's exact test. Last, we asked if candidate gene sets were enriched for Gene

Downloaded from https://academic.oup.com/genetics/article/220/2/iyab226/6460342 by University of California Berkeley user on 07 February 2022

Ontology (GO) terms and functional pathways using gprofiler2 (Kolberg *et al.* 2020) in R (R Core Team 2019).

To investigate whether targets of selection were the same in both transects, we used permutation tests to ask if the putative targets of selection were more similar than expected by chance using permTest function in regioneR (v 1.22.0, Gel *et al.* 2016) in R (R Core Team 2019). We tested for gene overlap between transects at several different significance thresholds including $q = 0.05$ and $q = 0.01$ for LFMM and the top 5%, top 1%, top 0.5%, and top 0.1% of PBS*n1* values. Using the same significance levels, we tested if there was more overlap between candidate SNPs in each transect than expected. For shared candidate genes based on the top 1% of PBS*n1* values, we also asked if there was more overlap in SNPs than expected. Last, we characterized the major allele at high elevation in shared candidate genes from both transects.

To assess the change in allele frequency in candidate SNPs across each transect, we used VCFtools (Danecek *et al.* 2011) and R (R Core Team, 2019) to plot allele frequency by elevation. To identify regions of exceptional allele frequency change at high elevation, we defined a large allele frequency change (LAFC) as a change greater than the average allele frequency change plus two standard deviations between adjacent populations. We examined allele frequency change for the following population pairs: (1) ~2500 and ~2900 m (north), (2) ~2600 and ~3300 m (south), and (3) ~3300 and ~3800 m (south). We conducted these calculations for SNPs across the genome as a baseline, and for PBS*n1* candidates alone. To highlight regions of major allele change, we merged LAFC SNPs within 100 kb of one another.

X chromosome selection scans

We conducted separate selection scans using PBS*n1* and LFMM on the X chromosome with minor adjustments to the methods described above. Specifically, we separately analyzed genotype data using LFMM as above; however, we defined the genotype of males as homozygous [0, 1] while females could be homozygous or heterozygous [0, 0.5, or 1]. To calculate the PBS*n1* statistic for the sites on the X chromosome, we calculated F_{ST} (Weir and Cockerham 1984) in VCFtools (Danecek *et al.* 2011) after we defined males as haploid. We conducted the selection scans using our original site filters, as well as on a dataset that excluded all missing data. We calculated the difference in nucleotide diversity for high and low populations in each transect, $\pi_{low} - \pi_{high} = \pi_{diff}$, as above.

To investigate the structure of genetic variation on the X chromosome, we phased the chromosome using fastPHASE (v 1.4, Scheet and Stephens 2006). We estimated maximum likelihood trees in RAxML (Stamatakis 2014) using the Lewis ascertainment correction, evolutionary model GTRCAT and rapid bootstrapping (-f a -N 100 -x 12345) for nonoverlapping windows of 50 SNPs. We used TWISST (Martin and Van Belleghem 2017) to assess relative tree topologies across the chromosome.

Results

We sequenced the exomes of 50 *M. m. domesticus* from the northern transect and 46 individuals from the southern transect at moderate coverage (15–20x). The number of unrelated individuals was 48 in the northern transect and 45 in the southern transect (Supplementary File S1). We identified 287,897 variant sites in the northern transect and 472,455 variant sites in the southern transect; coverage was comparable at exonic and intronic sites. We used these high-confidence SNPs to (1) assess the history of high elevation colonization of house mice in South America, (2) identify the genic targets of high elevation adaptation, (3) test if the genetic basis of high elevation adaptation is shared among high elevation populations, and (4) evaluate evidence of an elevation-based threshold effect. The results below are based on autosomal variation unless otherwise specified.

Independent origin of high elevation populations

A maximum likelihood phylogenetic tree estimated with 104,044 SNPs provides strong evidence that house mice colonized high elevations independently in the north and the south (Figure 2A). All mice from the northern transect formed a monophyletic group (bootstrap support = 100). Within the northern transect, mice were grouped largely by population. Moreover, populations in the north were arranged by elevation, with lower elevation populations at the base of the northern clade and higher-elevation populations nested within the northern clade (Figure 2A). This pattern is consistent with the colonization of high-elevation populations in Ecuador from nearby low-elevation populations. Mice from the southern transect did not form a monophyletic group. As in the north, mice were grouped largely by population. The low elevation populations from the southern transect diverged near the base of the South American clade, and the high-elevation populations in the south formed the sister group to all mice from Ecuador. The topology of the tree in Figure 2A thus demonstrates that high-elevation populations in the two transects are not each other's closest relatives, indicating that house mice colonized the high Andes independently in the north and the south.

We detected population structure in both transects using several methods (Figure 2 and Supplementary Figure S2). In the northern transect, the first principal component (PC1) explained 28.7% of the genetic variance, and it separated individuals collected below and above 500 m (Figure 2B). Admixture analysis (Figure 2D) and pairwise F_{ST} (Supplementary Table S1) also supported a genetic break between populations above and below 500 m. We found evidence of isolation by distance across the northern transect for the autosomes (Figure 2F, mantel test, P-value = 0.016) and the X chromosome (mantel test, P-value = 0.024); autosomal F_{ST} was also correlated with elevation ($R^2 = 0.62$). Levels of genetic variation, estimated from the proportion of segregating sites (Watterson's θ) and nucleotide diversity (π) for both coding and intronic regions, generally decreased from low to high elevations (Table 1). The decrease in nucleotide variation with increasing elevation is consistent with smaller founding populations colonizing higher elevations. Moreover, the genetic variation found at the highest elevation was largely a subset of the variation present at the lowest elevation (Supplementary Figure S3). In this comparison, 88% of the SNPs found in the highest population were also variable in the lowest population.

In the southern transect, the PC1 explained 23.8% of the genetic variation, and PC1 largely separated populations by elevation (Figure 2C). Ancestry assignments in Admixture with $K = 3$–5 closely followed a priori population designations based on elevation (Figure 2E and Supplementary Figure S4). We found significant isolation-by-distance among populations in the southern transect (Figure 2G and Supplementary Table S2; mantel test, $P = 0.006$). There was still significant isolation-by-distance when the most distant population (Porto Velho, Brazil) was removed (Supplementary Figure S5; mantel test, $P = 0.04$). Similarly, F_{ST} was also correlated with elevation ($R^2 = 0.49$, Supplementary Figure S5). The IBD pattern was recapitulated on the X chromosome ($P = 0.01$). In contrast to the northern transect, we found similar levels of nucleotide variability in all five populations of the southern transect (Table 1), and genetic

Downloaded from https://academic.oup.com/genetics/article/220/2/iyab226/6460342 by University of California Berkeley user on 07 February 2022

Table 1 Proportion of segregating sites (θ) (%) and nucleotide diversity (π) (%) for *M. m. domesticus*

	Intronic		Coding	
	θ	π	θ	π
Northern transect, ≤330 m	0.10145	0.1308	0.06310	0.07878
Northern transect, 320–500 m	0.11025	0.1353	0.06794	0.08207
Northern transect, 1300–1900 m	0.06527	0.0779	0.04049	0.04781
Northern transect, 2300–2700 m	0.05662	0.0688	0.03506	0.04153
Northern transect, 2800–3100 m	0.06438	0.0806	0.03924	0.04813
Southern transect, ≤90 m	0.11072	0.1406	0.06985	0.08647
Southern transect, 200–500 m	0.10555	0.1071	0.06598	0.06649
Southern transect, 2500–2900 m	0.10393	0.1428	0.06657	0.08966
Southern transect, 3000–3699 m	0.10495	0.1237	0.06605	0.07629
Southern transect, 3700 – 3900 m	0.09968	0.1207	0.06257	0.07443

variation at the highest elevation was not a simple subset of the variation at the lowest elevation (Supplementary Figure S3). This result, in combination with the topology of the maximum likelihood tree, does not fit a simple scenario where the high-elevation populations in Bolivia were founded by mice from the lowland populations that we sampled.

Genome-wide signals of selection across the northern transect

We looked for signals of selection in the northern transect with LFMM using 196,251 SNPs with MAF ≥ 0.05 in 48 unrelated individuals (Supplementary File S1). We chose $K = 2$ as the number of latent factors with the best fit to the data. We also looked for signals of selection with PBS*n1* using 126,473 SNPs with an MAF ≥ 0.05 across the three populations (see Materials and Methods). From neutral simulations, we observed an FDR for PBS*n1* of < 10e-7 when population sizes remained constant. In simulations with bottlenecks, we observed an FDR of 0.00164. These simulations suggest that most of the loci identified as significant using PBS*n1* are true positives.

Significant candidate SNPs were distributed across the genome using both LFMM and PBS*n1* (Figure 3). With LFMM, we identified 1865 SNPs in 625 unique gene regions with a signature of selection at a moderate threshold ($q = 0.05$). With a moderate PBS*n1* threshold (top 1% of PBS*n1* values), we found 1219 SNPs in 594 unique gene regions with a signature of selection. Significant peaks were also present on all autosomes using both methods at more stringent thresholds ($q = 0.01$ for LFMM; top 0.1% for PBS*n1*). There was significant overlap between LFMM and PBS*n1* candidate SNPs at both moderate and stringent thresholds (permutation tests, P-value ≤ 0.001 for both) as well as significant overlap for gene regions at both moderate and stringent thresholds (permutation tests, P-value ≤ 0.001 for both).

With both LFMM and PBS*n1*, we found that the average window size over which signals of selection were detected was small (~16 kb, distribution shown in Supplementary Figure S6) and comparable to the scale over which linkage disequilibrium decays in natural populations of *M. m. domesticus* (Laurie *et al.* 2007). This suggests that selection is generally not creating long haplotypes as might be expected under hard selective sweeps, but rather that selection is acting primarily on standing genetic variation.

The proportions of nonsynonymous SNPs and SNPs in UTRs were not significantly different between candidate SNPs and all SNPs either for LFMM or PBS*n1* (Supplementary Table S3). Surprisingly, however, the proportion of synonymous candidate SNPs was significantly greater than expected using both LFMM and PBS*n1* (permutation test, $P = 0.01$ for all subsets).

Since the ancestral range of *M. musculus* was at low elevation, we predicted that adaptation to high elevation might preferentially involve new, derived alleles. We tested this prediction by polarizing alleles with respect to *Mus spretus* and comparing the frequency of derived alleles across elevations. *M. spretus* diverged from *M. musculus* ~2 million years ago, shares little ancestral polymorphism with house mice and thus serves as a good outgroup (Harr *et al.* 2016). The derived allele was at high frequency at high elevation for candidate SNPs more often than expected by chance in LFMM and PBS*n1* results (Supplementary Table S4; Fisher's exact test, $P < 0.001$ for all subsets); approximately 73–76% of all candidate SNPs in LFMM and PBS*n1* analyses had a derived allele at highest frequency at high elevation. This excess contrasted sharply with the pattern seen in the remainder of the genome where the derived allele was at higher frequency at high elevation for approximately half the sites (48.5%).

We also found that estimates of π_{diff}, the difference in π between individuals collected below 300 m (π_{low}) and individuals collected above 2800 m (π_{high}) in 100-kb windows, were not normally distributed (Kolmogorov–Smirnov test, P-value = 2.2×10^{-16}) (Supplementary Figure S7). The distribution of π_{diff} was skewed to the right (kurtosis = 14.4, skewness = 0.85), meaning more regions showed elevated π at low elevation and/or depressed π at high elevation than expected in a normal distribution, consistent with selection acting at high elevations.

Genes that were identified at the moderate significance thresholds in the LFMM analysis ($q = 0.05$) and PBS*n1* values (top 1%) as well as in the top 1% of π_{diff} were strong candidates for high elevation adaptation. Twelve genes in the northern transect were significant in all three analyses (Supplementary Figure S8). Two of these genes (*Elmo1* and *Col22a1*) displayed large genomic windows impacted by selection for both LFMM and PBS*n1* (top 1% of all windows). Four additional genes displayed a large genomic window for PBS*n1* alone (Supplementary Figure S8B).

Genome-wide signals of selection across the southern transect

The LFMM analysis for the southern transect included 351,786 SNPs with an MAF ≥ 0.05 among 45 unrelated individuals (Supplementary File S1). We selected $K = 5$ as the number of latent factors as a conservative approach to correcting for the population structure we observed; however, P-values from LFMM analyses with $K = 3$–5 were highly correlated (correlation coefficient ≥ 0.99). We calculated PBS*n1* for 176,993 SNPs with an MAF ≥ 0.05 across all three populations.

Signatures of selection across the southern transect revealed many similarities and some differences compared to the patterns we observed in the northern transect. As with the northern transect, significant candidate SNPs were distributed across the genome using both LFMM and PBS*n1* at both significance thresholds (Figure 3). With LFMM, we identified 1027 SNPs in 342 unique gene regions with a signature of selection at a moderate threshold ($q = 0.05$). With a moderate PBS*n1* threshold (top 1% of PBS*n1* values), we found 1760 SNPs in 717 unique gene regions with a signature of selection. LFMM and PBS*n1* candidate gene

Downloaded from https://academic.oup.com/genetics/article/220/2/iyab226/6460342 by University of California Berkeley user on 07 February 2022

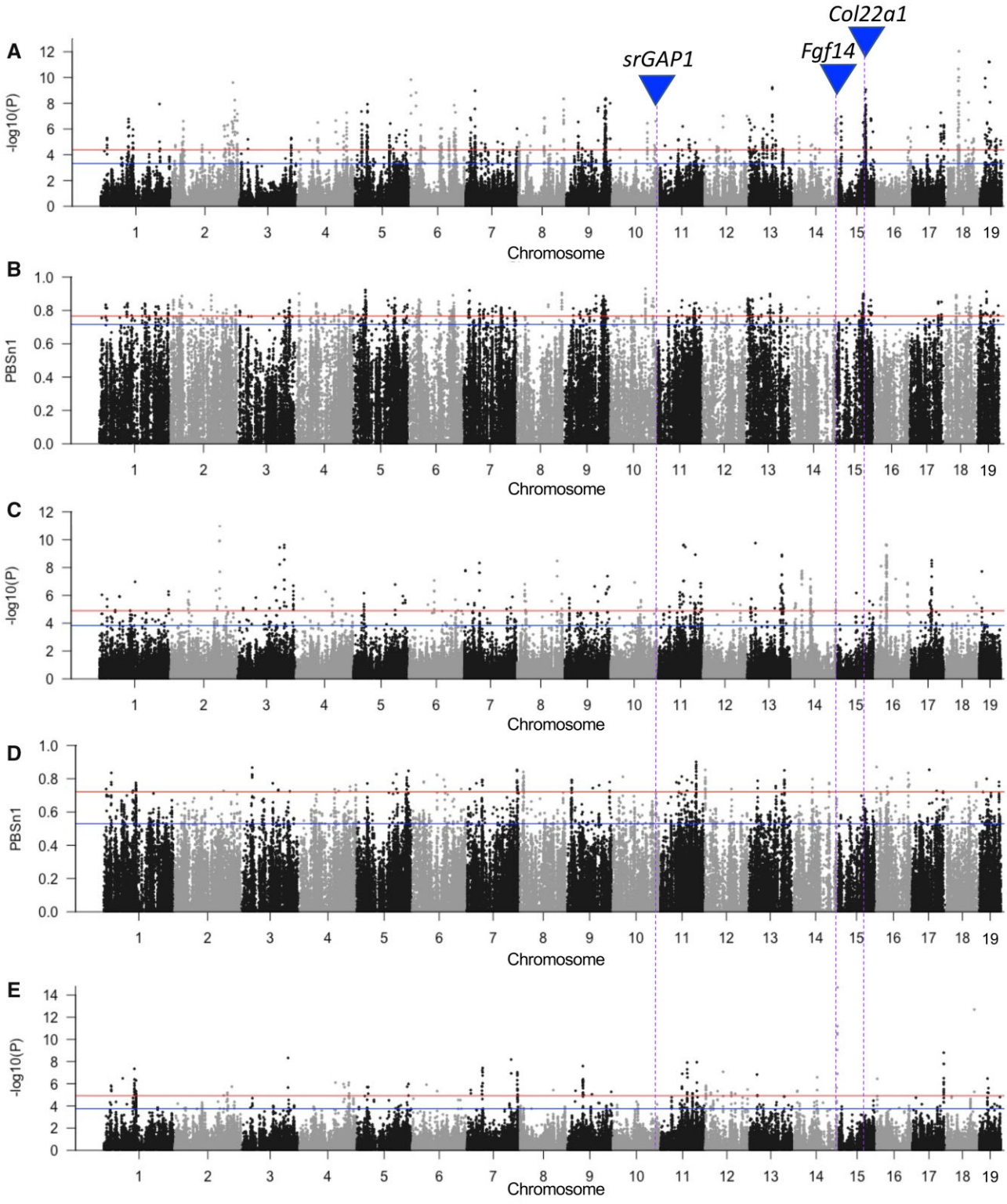

Figure 3 Autosomal selection scans for *M. m. domesticus* across elevation. (A) Latent Factor Mixed Modeling (LFMM) results with elevation in northern transect ($n = 48$; 196,251 SNPs). (B) PBS*n1* results for northern transect, ($n = 48$; 126,473 SNPs). (C) LFMM results with elevation in southern transect ($n = 45$; 351,786 SNPs). (D) PBS*n1* results for southern transect ($n = 45$; 176,993 SNPs). (E) LFMM results with elevation for all individuals ($n = 93$; 131,701 SNPs). LFMM plots include $q = 0.05$ (blue) and $q = 0.01$ (red) lines. PBS*n1* thresholds of top 1% (blue) and 0.1% (red) shown in (B) and (D) Genes discussed in detail in text are indicated with blue triangles in A; the purple dashed line below each triangle indicates the gene position in panels (B–E).

regions overlapped significantly more than expected by chance (permutation test, P-value ≤ 0.001), but SNPs did not (permutation test, P-value $= 0.42$). As in the northern transect, the genomic distance over which signals of selection extended, on average, was short (~16 kb using either LFMM or PBS*n1* at the moderate threshold; Supplementary Figure S6). In contrast to the northern transect, candidate SNPs using LFMM or PBS*n1* in the southern transect showed significantly more nonsynonymous

Downloaded from https://academic.oup.com/genetics/article/220/2/iyab226/6460342 by University of California Berkeley user on 07 February 2022

and 3′ or 5′ UTR changes than expected based on the entire SNP dataset (Supplementary Table S3). Similar to the northern transect, the top 1% and top 0.1% PBS*n1* candidates had more derived alleles at high frequency at high elevation in the southern transect (Fisher's exact test, $P < 0.001$), but the LFMM candidate SNPs did not (Supplementary Table S4). For PBS*n1* candidates, between 67% and 74% of sites had a derived allele at high frequency at high elevation. Again, in noncandidate regions, the derived allele was at higher frequency at high elevation for close to half of all sites (50.8%). The difference in nucleotide diversity (π_{diff}) between the populations collected under 90 m and above 3700 m did not fit the expectations of a normal distribution (Kolmogorov-Smirnov test, *P*-value $= 2.2 \times 10^{-16}$); as in the north, there were more outliers on the right side of the curve than expected (kurtosis = 12.29, skewness = 0.448) (Supplementary Figure S7).

Across the southern transect, there were three genes (*Cntnap5b*, *2210408I21Rik*, and *Atp12a*) which were significant outliers based on LFMM and PBS*n1* analyses at moderate thresholds as well as in the top 1% of π_{diff} values (Supplementary Figure S9). The gene *Atp12a* also bore the signature of selection across a large region of the genome based on LFMM and PBS*n1* values (top 1% of windows).

X chromosome selection scans

We analyzed the X chromosome separately in the northern and southern transects. For the LFMM analyses, we kept the same number of latent factors as in the autosomal analyses ($K = 2$ for the north; $K = 5$ for the south), and included 1530 SNPs from the northern transect and 3914 SNPs from the southern transect. For the PBS*n1* analyses, we included 1159 SNPs from the northern transect and 1364 SNPs from the southern transect.

Across the northern transect, we found 13 significant SNPs in 13 unique gene regions in the LFMM analysis at the $q = 0.05$ threshold. Eleven of these 13 SNPS were located in a single 7.8 Mb region on the X chromosome (94253117–102069835) (Figure 4). When we phased the variant sites across the X chromosome, we found that the 7.8 Mb region corresponds to a large haplotype that is nearly fixed for different alleles in comparison between the two low-elevation populations and the three high-elevation populations in the northern transect. Populations above 1300 m and below 500 m had low genetic variation across this 7.8 Mb region; average π in the long haplotype was 37% that of the remainder of the X chromosome. Candidate alleles at high elevation were equally likely to be derived or ancestral, and this region does not correspond to a known chromosomal rearrangement in *M. musculus*. The functions of the genes in this region are not related to known hypoxia response pathways. *PBSn1* analyses identified 10 SNPs in nine genes at the 1% level across the X chromosome, including five SNPs within this long haplotype and five SNPs outside this region. Across the X chromosome, PBS*n1* alleles at high frequency at high elevation were more likely to be derived than expected by chance (Supplementary Table S5, fisher's exact test, $P < 0.01$). LFMM and PBS*n1* candidate gene regions overlapped significantly more than expected by chance (permutation test, *P*-value ≤ 0.001), due to three shared genes within this long haplotype. In contrast, LFMM and PBS*n1* candidate SNPs did not overlap more than expected by chance.

Across the southern transect, we identified 100 SNPs in 97 gene regions using LFMM and 14 SNPs in 14 gene regions using PBS*n1* with a moderate threshold. Significant gene regions and SNPs were not shared between LFMM and PBS*n1* analyses more than expected by chance. Alleles at high elevation were more likely to be derived among LFMM candidate SNPs (Fisher's exact test, $P < 0.01$), but not among PBS*n1* candidate SNPs (Supplementary Table S5).

Unique and shared responses to selection

Next, we compared the signatures of selection in the northern and southern transects. Several genome-wide patterns were broadly consistent between the two transects, including a larger than expected number of derived alleles at high elevation, a skewed distribution of π_{diff} values, and the observation of small genomic regions bearing the signature of selection across all autosomes. Despite these broad similarities, there were many differences between the two transects in the response to selection. We explored these differences at the pathway, gene, and SNP levels.

At the pathway level, no candidate gene set from a single selection scan was enriched for any GO terms or pathways. Furthermore, candidate genes that were significant in both LFMM and PBS*n1* analyses in each transect were not enriched for any GO terms or pathways. Last, candidate genes that were significant in both transects based on LFMM alone, PBS*n1* alone, or both methods of analysis were not enriched for any GO terms or pathways. Thus, the selection candidates we identified were not enriched for functional pathways associated with hypoxia, and there was no evidence of a broad-scale shared response between the independent high elevation populations at the pathway level.

At the level of genes, the degree of overlap between the transects for the autosomes depended on the method of analysis (Figure 5; Supplementary Table S6). Using LFMM, we found that the targets of selection were primarily transect-specific. At the $q = 0.05$ and $q = 0.01$ thresholds, we observed 17 and 1 gene regions, respectively, that were shared between the transects; this was not more than expected by chance (Supplementary Table S6; permutation test, *P*-value > 0.05). In contrast, using *PBSn1*, we found greater overlap of gene regions than expected by chance between transects across a range of thresholds (Supplementary Table S6). In the top 5%, top 1% and top 0.5% of PBS*n1* values, we found significant overlap between transects (permutation test, *P*-value ≤ 0.001 in all comparisons). In the top 0.1% of PBS*n1* values, we did not observe more overlap than expected by chance, although the power of this comparison is probably reduced due to the smaller number of candidate gene regions (Supplementary Table S6). We also conducted an LFMM analysis using all individuals from both transects (93 samples; 131,701 SNPs); 156 out of 210 unique gene regions were shared between this analysis and any transect-specific analyses (Figures 3 and 5).

Despite the largely distinct response to selection in each transect, several genes were identified in both transects by one or more analytic methods, and these are strong candidates for loci underlying adaptation to high elevation. We found 43 gene regions (containing 46 genes) that were outliers at a moderate threshold for at least three of the four transect-specific selection scans (Supplementary Table S7), and six gene regions (containing seven genes) that were significant in all four selection scans (Table 2). We used the MGI database (Bult *et al.* 2019), based largely on mutant alleles in inbred lines, to characterize the function of these six genes and found that they primarily affect the nervous system (four genes) and the cardiovascular system (three genes). Cellular functions associated with more than one candidate gene included signal transduction (2), cell migration (2), and cell adhesion (2). One of these genes, *Col22a1*, is involved in extracellular matrix organization and angiogenesis, and was also in the top 1% of π_{diff} values in the northern transect. Three genes (*Sh3pxd2b*, *Fgf14*, and *Ston1-Gtf2a1l*) were also significant at

Downloaded from https://academic.oup.com/genetics/article/220/2/iyab226/6460342 by University of California Berkeley user on 07 February 2022

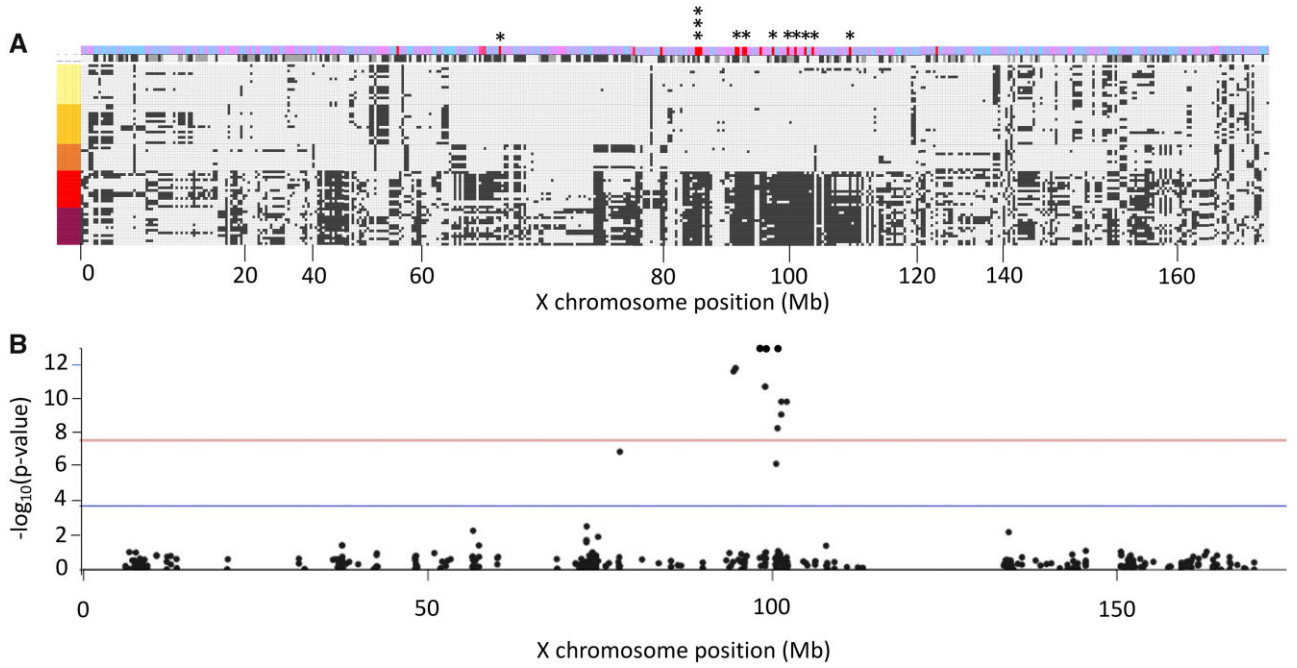

Figure 4 Summary of genetic variation on X chromosome of northern *M. m. domesticus*. (A) Phased haplotypes of variable sites only ($n = 50$; 68 chromosomes; 482 SNPs) showing the major (light) and minor (dark) alleles per SNP. From the top of (A), SNPs indicated by * are significant at $q = 0.05$ in latent factor mixed models (LFMM) analysis. In the next row, the LFMM *P*-value for each SNP is colored from the most significant (red) to the least significant (blue). Next, a row indicates which allele (light or dark) is ancestral; ancestry of SNPs with medium gray could not be assigned. Below this, individual haplotypes across the X chromosome are shown as narrow rows. The elevation of each chromosome is shown on the left and corresponds to < 330 m (maroon), 320–500 m (red), 1300–1900 m (dark orange), 2300–2700 m (light orange), or 2800–3100 m (yellow). (B) LFMM analysis ($n = 48$; 482 SNPs) with $q = 0.05$ threshold in blue and $q = 0.01$ in red.

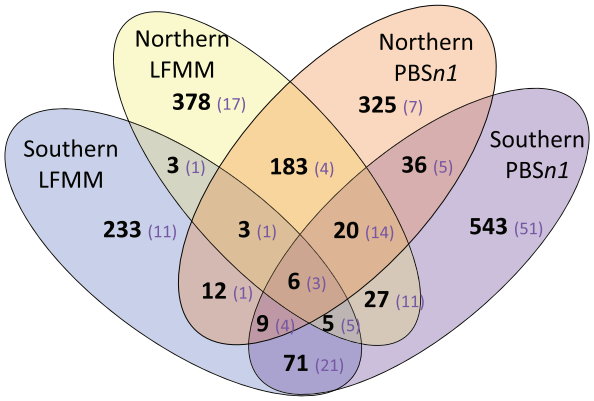

Figure 5 Autosomal unique gene region overlap of selection scan candidates in northern and southern *M. m. domesticus*. Significance thresholds for each transect are $q = 0.05$ for LFMM and top 1% of PBS*n1* values. Numbers in parentheses are unique gene regions of set that are also found in the combined transect analysis with $q = 0.05$.

the $q = 0.05$ threshold in the LFMM analysis based on the combined data from both transects (Figure 5). Finally, we identified genomic regions that showed signatures of selection in both transects using either PBS*n1* or LFMM. This comparison identified 121 regions (Figure 5) containing 127 genes. We then asked whether any of these 127 genes showed evidence of selection in previous studies of other high-elevation taxa. Seventeen such genes were identified (Table 3), including three of the six from Table 2 (i.e., those showing signatures of selection using *both* LFMM and PBSn1 in *both* transects). There are thus multiple lines of evidence implicating these three genes (*Col22a1*, *Fgf14*, and *srGAP1*) in high-elevation adaptation.

At the SNP level, there was not more overlap than expected by chance between the transects for candidates identified by either LFMM or PBS*n1* at any significance threshold (Supplementary Table S8). In fact, we observed significantly *less* overlap of candidate SNPs than expected by chance between the transects for both LFMM ($q = 0.05$ threshold) and PBS*n1* (top 5%, 1%, and 0.5% thresholds) (Z-scores ≤ -2, corresponding to *P*-values ≤ 0.02 for each of these four analyses; Supplementary Table S8), suggesting that selection has acted on different alleles in each transect. We also explored whether SNPs were shared among the set of 71 overlapping genes observed in the top 1% of PBS*n1* values. When restricted to these 71 genes alone, we found that PBS*n1* candidate SNPs overlapped between the transects more than expected by chance (permutation test, *P*-value ≤ 0.001), suggesting there may be common alleles under selection in both transects in a subset of the genome. Furthermore, candidate SNPs that were significant in both transects consistently demonstrated the same major allele at high elevation in the north and south.

Signatures of selection on the X chromosome generally differed between the northern and southern transects. The amount of overlap between the transects was not greater than expected by chance either for gene regions (Supplementary Table S9) or for SNPs (Supplementary Table S10). In contrast to the autosomes, there were no genes on the X chromosome that were significant in three or more selection tests (Supplementary Figure S10). Finally, the large haplotype block present on the X chromosome in the northern transect was not seen in the southern transect.

Downloaded from https://academic.oup.com/genetics/article/220/2/iyab226/6460342 by University of California Berkeley user on 07 February 2022

Table 2 Genes identified in both selection scans[a] in *M. m. domesticus* in both transects

Gene	Position	Northern transect			Southern transect		All data	Tissue[b]	Functional summary
		LFMM	PBS*n1*	π_{diff}	LFMM	PBS*n1*	LFMM		
SRGAP1	10:121780991	X**	X	–	X	X	–	NERV	Signal transduction and cell migration
SH3PXD2B	11:32347820	X	X	–	X**	X	X	CARD MUSC NERV OCUL	Essential to organ development and osteoblast fate; null allele mutants have severe skeletal and cardiac defects
FGF14	14:123977907	X**	X	–	X	X	X	CARD NERV RESP	Growth factor activity that includes heparin binding activity; involved in cardiac conduction and muscle contraction
COL22A1	15:71795795	X**	X	X	X	X	–	NERV CARD	Extracellular matrix organization, Key to collagen formation and angiogenesis
MROH5	15:73786936	X	X	–	X	X	–	-	–
STON1	17:88597615	X**	X	–	X	X	X	GENI	Cell adhesion, cellular signaling, cell migration

[a] "X" indicates a significance threshold of $q \leq 0.05$ for latent factor mixed modeling (LFMM) analyses, the top 1% of PBS*n1* values, and the top 1% of the difference in π between low and high populations (π_{diff}).
[**] Indicates $q \leq 0.01$ for LFMM.
[b] Affected tissues are abbreviated as CARD: cardiovascular system, GENI: genitourinary system, MUSC: musculoskeletal system, NERV: Nervous system, OCUL: ocular, RESP: respiratory system.

Threshold effects

To explore whether any genes exhibit a threshold effect, we examined allele frequencies across elevations for candidate SNPs from the LFMM and PBS*n1* analyses (Figure 6). This comparison illustrates how the different analytic approaches identify distinct but overlapping candidate SNPs. As expected, we found that candidate SNPs from LFMM analyses in the northern (Figure 6A) and southern (Figure 6B) transects showed a linear relationship between elevation and allele frequency. In contrast, the PBS*n1* statistic does not rely upon information from intermediate elevations, and we found considerable variation in the pattern of allele frequency change when comparing allele frequency with elevation for candidate SNPs (Figure 6, C and D). Notably, there were a number of gene regions showing LAFCs at the highest elevation. We observed nine such LAFC gene regions, defined as 2 or more SNPs exhibiting LAFC within 100 kb, in the northern transect and 16 LAFC gene regions in the southern transect.

LAFC regions contained a number of genes associated with pathways involved in hypoxia resistance (Table 4). In the northern transect, three out of nine regions were involved in oxygen transport (*UBXN4*, *Slc4a1*), morphogenesis of epithelial cells including in blood vessels and lungs (*Esrp1*), and cardiac and erythrocyte morphology (*Slc4a1*). In the southern transect, eight gene regions out of 16 were related to hypoxia response. These included the critical gene, erythropoietin (*Epo*), a gene in the HIF pathway that plays a fundamental role in erythrocyte proliferation and hemoglobin biosynthesis. Across the 198 kb region containing *Epo*, there were 14 SNPs with exceptionally LAFC from 3300 to 3800 m. Another gene, encoding the apelin receptor (*Aplnr*), is involved in the regulation of blood vessel cell proliferation in angiogenesis, and the region includes 8 SNPs with LAFC over 1.8 kb. *Acta2* is a gene involved in the regulation of blood pressure and vascular smooth muscle contraction, and it exhibited 3 SNPs with LAFC within an 11 bp segment.

Overall, we found no overlap of either genes or SNPs in LAFC regions between the northern and southern transects. However, many of the transect-specific LAFC regions which demonstrate a threshold effect contain genes known to be relevant to high elevation adaptation.

Discussion

We used gene sequences to uncover the history of house mice in the Andes and identify genetic signatures of natural selection associated with high elevation. We found that house mice colonized high elevations independently in the north and in the south. Targets of selection were widely distributed across the genome, and candidate SNPs were enriched for derived alleles at high elevation. Notably, targets of selection were largely transect-specific; however, a small number of genes showed signatures of selection in both transects. Finally, we found that many of the genomic regions exhibiting an LAFC at or above 2900 m possessed a gene in the HIF pathway or contributed to the development of the cardiovascular system, angiogenesis, and oxygen transport.

House mice colonization in the high Andes

The phylogeny in Figure 2 indicates that house mice colonized high elevations independently in Ecuador and in Bolivia since high elevation mice in Ecuador and Bolivia are not each other's closest relatives. In Ecuador, the pattern of decreasing heterozygosity with increasing elevation, as well as the nesting of high-elevation populations within the phylogeny of low-elevation populations, suggests that house mice that reside at or above 1300 m were derived from lowland mice on or near the Pacific coast. The lowland population from which high elevation mice in Bolivia arose is uncertain, but our analyses suggest house mice are not moving large distances latitudinally across the Andean Altiplano. Additional sampling and demographic analyses could elucidate the connectivity among Andean populations in detail and the source population for mice in the Bolivian Andes.

Dispersal of house mice into the high Andes was probably mediated by human transport; thus, post-colonial trade routes in South America provide a context for interpreting the phylogeographic patterns we observed. The monophyletic clade of house mice in Ecuador mirrors the import and export trade connections between Spanish administrative power in Quito and the Pacific coast port at Guayaquil beginning in the early 1500 s (Conniff 1977; Super 1979), though trade dynamics between the coast

Downloaded from https://academic.oup.com/genetics/article/220/2/iyab226/6460342 by University of California Berkeley user on 07 February 2022

Table 3 Candidate genes[a] with evidence of selection from previous studies[b]

Gene	Functional summary	HEAS[b]	Citation
ANKMY2	Enzyme binding activity, Colocalizes with Egln1, regulator of the HIF pathway	Interaction with EGLN1	Pourhaghighi *et al.* (2020) Simonson *et al.* (2010) Yi *et al.* (2010)
CDH4	Cell adhesion; Neuron generation	Paralog under selection (1)	Bigham *et al.* (2009)
COL9A1	Extracellular matrix component; Involved in chondrocyte proliferation and cartilage development	Paralog under selection (3)	Schweizer *et al.* (2019) Arciero *et al.* (2018) Alkorta-Aranburu *et al.* (2012)
COL22A1	Role in extracellular matrix organization, collagen formation, and angiogenesis	Paralog under selection (3)	" "
COL25A1	Axonogenesis, involved in innervation	Paralog under selection (3)	" "
CYP2U1	Heme binding activity; involved in metabolism	Cytochrome P450	Pokreisz *et al.* (2006)
DNAH11	Cardiac septum morphogenesis, determination of left/right symmetry	Paralog under selection (1)	Li *et al.* (2014)
ELMO2	Actin filament organization and cell chemotaxis	VEGF pathway VEGFA pathway	Li and Eriksson (2001)
FGF14	Growth factor activity, heparin binding activity; Involved in cardiac conduction, muscle contraction	Paralog under selection (3)	Gorkhali *et al.* (2016) Edea *et al.* (2019) Wang *et al.* (2016) Dong *et al.* (2014) Yang *et al.* (2017) Jha *et al.* (2016)
HADH	Positive regulation of cold-induced thermogenesis; Direct interaction with ITGA2 in HIF pathway	Interaction with ITGA2	Tseng *et al.* (2018)
ITGBL1	Cell adhesion, integrin-mediated signaling pathway Colocalizes with MMP2, integral to endothelial cell angiogenesis	Interaction with MMP2	Pourhaghighi *et al.* (2020) Ben-Yosef *et al.* (2005)
MB	Response to hypoxia, brown fat cell differentiation, enucleate erythrocyte differentiation	Response to hypoxia	De Miranda *et al.* (2012)
RYR2	Heart morphogenesis, regulation of heart contraction	RYR2 (3)	Zhang *et al.* (2014) Wu *et al.* (2020) Wang *et al.* (2015)
SLC35B4	Regulation of gluconeogenesis	SLC35 isoform (2)	Huerta-Sánchez *et al.* (2013) Wang *et al.* (2015)
SRGAP1	Signal transduction and cell migration	Selection on SRGAP1 (3)	Wu *et al.* (2020)
STAT5B	Involved in organ development, cell surface receptor signaling pathway and lymphocyte activation	Selection on STAT5B	Simonson *et al.* (2010)
TGFBR3	Angiogenesis; blood vessel and heart development	TGF ß signaling	Chu *et al.* (2011)

[a] One-hundred twenty-one genomic regions that showed signatures of selection in both transects using either PBS*n1* or LFMM were identified. These 121 regions contained 127 genes, 17 of which were identified in previous studies as showing signatures of selection at high elevations as summarized here.
[b] Genes have reported high elevation associated selection (HEAS) documented for (1) the gene itself, (2) a paralog with similar function, (3) an associated pathway, or (4) a direct protein-protein interaction with any of the former. The number of wild populations in which a gene or paralog is under HEAS is indicated in parentheses.

and the Sierras varied dramatically over time (Hanratty 1991; Newsom 1995). The silver trade routes may also provide a clue to the sister relationship we observed between mice from Ecuador and mice from La Paz in Bolivia. One of the destinations for silver extracted from the rich mines of Potosí and Sucre was the city of Lima, on the Pacific Ocean. The city of La Paz initially played a limited role as a stopover site along this journey before developing into a major trade center with significant population growth in the 1800 s (Leonard 1948). Thus, the source of populations in Ecuador and the Bolivian Andes may have been house mice introduced at ports on the Pacific coast. An alternative scenario, that house mice arrived in La Paz via the trade route from Buenos Aires to Potosí (Cobb 1949), could be addressed with additional sampling across the Andes and lowland South America.

Genetic basis of high elevation adaptation in house mice

We used multiple methods to identify several hundred genes from each transect that were under selection at high elevation. While many of these genes may play a direct role in hypoxia adaptation, the six genes in Table 2 stand out for showing a signature of selection using two methods of analysis in both transects. Three of these genes (*Col22a1*, *Fgf14*, and *srGAP1*) are especially

Downloaded from https://academic.oup.com/genetics/article/220/2/iyab226/6460342 by University of California Berkeley user on 07 February 2022

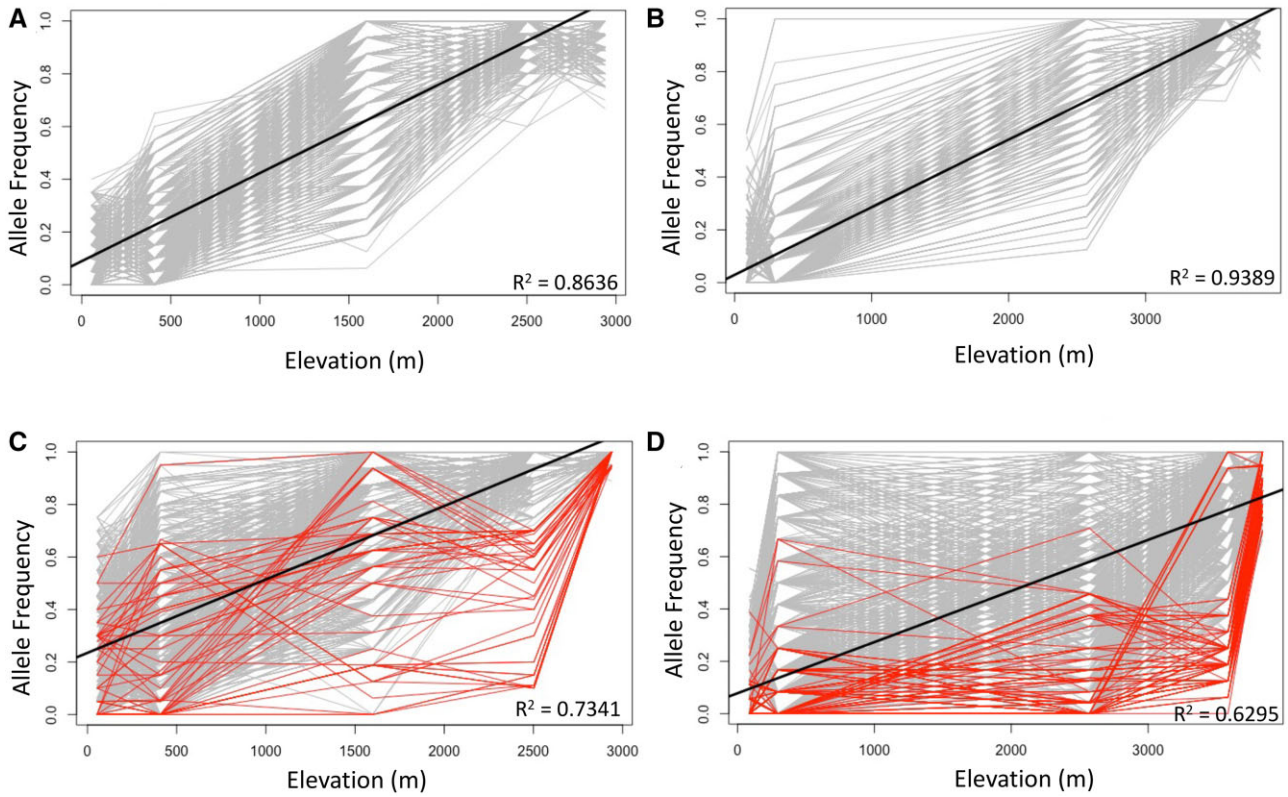

Figure 6 Population allele frequency of autosomal candidate SNPs by elevation. Individual SNPs (gray) and linear regression with all candidates (black) across elevation. (A) Northern transect LFMM candidates with $q = 0.05$. (B) Southern transect LFMM candidates with $q = 0.05$. (C) Northern transect, SNPs from top 1 % of PBS*n1* results. (D) Southern transect, SNPs from top 1 % of PBS*n1* results. SNPs with large allele frequency change (LAFC) from below to above ~2900 m in red.

good candidates for high-elevation adaptation because of information from previous studies (Table 3). Functional studies on these genes in *M. musculus domesticus* could provide insight into the initial stages of adaptation to high elevation.

Experiments on the minor collagen gene *Col22a1* (collagen, type XXII, alpha1) implicate it in several high-elevation associated phenotypes including (1) heart mass in laboratory mice (Shorter *et al.* 2018), (2) angiogenesis (Gaudet *et al.* 2011), and (3) vascular integrity and permeability (Ton *et al.* 2018). *Col22a1* demonstrates an early response to TGFß signaling stimulation (Watanabe *et al.* 2019) and modulates the expression of *Acta2*. Interestingly, *Acta2* and a second gene, *Aplnr*, are both members of the Apelin signaling pathway which affects blood pressure, angiogenesis, cardiac contractility, and energy metabolism (Wysocka *et al.* 2018). In the southern transect, *Acta2* and *Aplnr* were also under selection, and both gene regions included multiple SNPs that exhibited an LAFC at high elevation (Table 4). These three potentially interacting genes are located on separate chromosomes. Previously reported collagen paralogs under high-elevation associated selection include *Col1a2* in deer mice (Schweizer *et al.* 2019), *Col4a4* in Tibetan humans (Arciero *et al.* 2018), and *Col6a1* in Ethiopian humans (Alkorta-Aranburu *et al.* 2012).

Fgf14 (fibroblast growth factor 14) is the fourth member of the fibroblast growth factor (FGF) family to be implicated in adaptation to high elevation in mammals. *Fgf5* showed signatures of selection in domestic sheep (Edea *et al.* 2019), goats (Wang *et al.* 2016), and pigs (Dong *et al.* 2014) at high elevation. In addition, *Fgf7* was under selection in sheep (Gorkhali *et al.* 2016) and *Fgf10* bore evidence of selection in pigs (Li *et al.* 2013; Dong *et al.* 2014) on the Qinghai-Tibetan Plateau. These paralogs share core conserved amino acid residues (Wang *et al.* 2002). However, the functional roles of FGF proteins in high-elevation adaptation remain unclear and may vary substantially among paralogs. *Fgf14* modulates neuronal and cardiac voltage-gated Na^+ channels (Wang *et al.* 2002; Lou *et al.* 2005; Li *et al.* 2014). *Fgf14* knockout mice demonstrate changes in locomotory behavior (Wang *et al.* 2002; Strupp *et al.* 2020), social behavior (Hoxha *et al.* 2019), and spatial learning (Wozniak *et al.* 2007). We found two SNPs in the 3' UTR and one SNP in an intron of *Fgf14* that were significant in both transects; these SNPs may play a regulatory role or may be in LD with unsampled causative sites.

Finally, *srGAP1* (SLIT-ROBO Rho GTPase activating protein 1) mediates endothelial cell migration during sprouting angiogenesis (Blockus and Chédotal 2016; Genet *et al.* 2019). Wu *et al.* (2020) report *srGAP1* as under convergent, positive selection in domestic cattle, horses, and sheep living on the Qinghai-Tibetan Plateau. Interestingly, *srGAP1* also modulates *Rac1* activity (Yamazaki *et al.* 2013); *Rac1* is a component of the VEGF signaling pathway, a critical pathway that regulates angiogenesis. We found that *Rac1* showed signatures of selection and LAFC at high elevation in the southern transect (Table 4). Again, these interacting genes are on separate chromosomes.

Unique and shared responses to selection

By sampling two transects separated by over 2000 km, we were able to assess the repeatability of evolution over short evolutionary timescales since house mice in Ecuador and Bolivia independently colonized high elevations. The over-arching pattern was that most signals of selection were specific to each transect

Downloaded from https://academic.oup.com/genetics/article/220/2/iyab226/6460342 by University of California Berkeley user on 07 February 2022

Table 4 Candidate regions[a] of large allele frequency change above 2900 m

Transect	Position	Length (kb)	SNPs	Gene[b]	Functional summary[b]
North	1:128244465	0.2	2	**Ubxn4**	**Involved in oxygen transport, binds embryonic hemoglobin (Hba-x)**
North	4:11219755	148.4	9	**Esrp1** . . .	**Morphogenesis of branching epithelium including blood vessels and lung**
North	7:68738456	3.5	3	Arrdc4	Extracellular vesicle biogenesis
North	11:68557194	37.5	2	Mfsd6l	Integral component of plasma membrane
North	11:102350132	11.1	**3**	**Slc4a1**	**Oxygen transport, erythrocyte morphology, cardiac morphology**
North	15:65887656	0.3	2	Oc90	Bone mineralization
North	19:10901049	19.4	3	Prpf19 . . .	Cell proliferation, Neurogenesis
North	19:11251559	1.7	4	Ms4a1	Involved in the immune system including B cell number and physiology
North	19:11461388	18.8	5	Ms4a4b Ms4a6c	Integral components of plasma membrane
South	2:84678316	45.5	3	Tmx2 . . .	Exhibits disulfide oxidoreductase activity. Involved in brain development
South	2:85137930	1.8	8	**Aplnr**	**Apelin receptor activity**. **Role in circulatory system development, and regulation of blood vessel endothelial cell proliferation in angiogenesis.**
South	3:126364943	0.23	3	Arsj	Metal ion binding activity.
South	4:47039636	9.6	4	Anks6	Involved in animal organ development; determination of left/right symmetry
South	4:155583106	83.9	2	**Nadk**, Slc35e2, Mmp23	Nadk: involved in regulation of insulin secretion; **direct interaction with ITGA2b**, **downstream target of HIF pathway** Slc35e2: transmembrane transporter Mmp23: involved in collagen catabolic process;
South	4:155903388	155.8	3	. . .	—
South	5:137474042	198.4	14	**Epo** . . .	**HIF signaling pathway; erythrocyte differentiation; and hemoglobin biosynthetic process.**
South	5:139364147	31.7	13	**Gpr146**	Localizes to plasma membrane. Null allele mutants exhibit abnormal **hemoglobin content**, **red blood cell morphology** and **platelet cell number**
South	5:140629527	5.9	2	Ttyh3	Integral component of plasma membrane, involved in chloride transport
South	5:143481619	27.9	11	**Rac1**	GTP binding activity, involved in cell morphogenesis, cell differentiation, and neuron generation. Null allele mutants exhibit **cardiac hypertrophy, oxidative stress** and **abnormal platelet aggregation**. In **MAPK signaling pathway and VEGF signaling pathway**
South	6:49378460	11.4	2	Fam221a	—-
South	10:20023776	86.8	3	**Map3k5**	Involved in intracellular signal transduction, regulates MAPK cascade. Included in **thermogenesis** pathway. Null allele mutants have **altered response to myocardial infarction** and decreased **response of heart to induced stress.**
South	10:20273661	49.2	3	Bclaf1 . . .	Regulation of response to stress and apoptosis; null allele mutants exhibit impaired lung development, B and T cell abnormalities
South	11:99215323	171.4	14	Smarce1 Krt genes	Smarce1: Involved in neurogenesis. Included in **thermogenesis** pathway Krt*: keratin/keratin-like genes involved in intermediate filaments
South	18:65276770	30.2	4	**Alpk2**	**Involved in epicardium morphogenesis and heart development**
South	19:34243280	0.01	3	**Acta2**	Involved in **blood pressure regulation and vascular associated smooth muscle contraction. Apelin signaling pathway,** Relaxin signaling pathway.

[a] Only regions containing more than one qualifying SNP within 100 kb of one another are included. The number of qualifying SNPs in the region (SNPs column) and the length of the region in kb (Length column) are shown.
[b] Genes and key words associated with hypoxia response in bold. For brevity, the symbol "…" indicates additional, unlisted genes.

Downloaded from https://academic.oup.com/genetics/article/220/2/iyab226/6460342 by University of California Berkeley user on 07 February 2022

(Figure 5). For example, using LFMM, 625 genomic regions were identified in the northern transect, and only 17 of these were also identified in the southern transect. Using PBSn1, 594 genomic regions were identified in the northern transect, and only 71 of these were also identified in the southern transect. Similarly, there was no overlap of LAFC regions above 2900 m between the two transects.

Despite these largely transect-specific signatures of selection, there were a number of patterns that were shared between transects; these are consistent with selection acting primarily in the high-elevation populations. First, there was more overlap of PBS*n1* candidate genes than expected by chance at all but the most stringent thresholds (Supplementary Table S6). Moreover, we observed more shared candidate SNPs than expected when we only considered the overlap of the top 1% of PBS*n1* candidate genes. Second, we observed a skew in the distribution of π_{diff} values in both transects. Last, we observed a large excess of derived alleles at high frequency at high elevation for candidate SNPs in three of four selection scans (Supplementary Table S4) but not for SNPs in the remainder of the genome. Similarly, Wu *et al.* (2020) also reported more derived alleles in nonsynonymous candidate variants at high elevation than expected by chance in domestic sheep, dogs, and cattle on the Qinghai-Tibetan Plateau. If candidate SNPs are the direct targets of selection, derived alleles may be favored at high elevation. Consistent with this, nonsynonymous candidate SNPs showed an even more pronounced enrichment of derived alleles at high elevation. However, many of the candidate SNPs we identified are probably linked to targets of selection rather than being direct targets themselves. In the case of linked selection, the observed excess of derived alleles at high elevations remains somewhat puzzling. While gene flow can produce an excess of high-frequency derived alleles genome-wide (*e.g.*, Marchi and Excoffier 2020), the fact that we see this pattern only at candidate SNPs and not genome-wide raises the possibility that the pattern reflects a combination of linked selection and demographic processes such as gene flow or bottlenecks. Future simulations could be useful for exploring this idea.

Our observations of population-specific responses to selection in house mice mirror some patterns seen in humans. Humans independently colonized high elevations between 200 and 1600 generations ago on the Andean Altiplano, the Qinghai-Tibetan Plateau, and the Ethiopian highlands (Aldenderfer 2003, 2011; Pleurdeau 2006; Rademaker *et al.* 2014; Zhang *et al.* 2018). These three groups evolved distinct physiological responses to hypoxia that distinguish each population from their closest lowland relatives as well as from each other (Beall 2006, 2007). While many of the genes underlying high-elevation adaptation in these populations are in the HIF pathway, the majority of genes and individual SNPs under selection are population specific (Beall *et al.* 2010; Bigham *et al.* 2010; Simonson *et al.* 2010; Yi *et al.* 2010; Xu *et al.* 2011; Scheinfeldt *et al.* 2012; Huerta-Sánchez *et al.* 2013; Hu *et al.* 2017). With the exception of *Egln1* and *Epas1*, which are under selection in both Tibetans and Andeans (Bigham *et al.* 2010; Simonson *et al.* 2010; Yi *et al.* 2010; Foll *et al.* 2014, Eichstaedt *et al.* 2017), most studies report that (1) individual genes under selection in high elevation populations were not shared (Bigham *et al.* 2010; Alkorta-Aranburu *et al.* 2012; Scheinfeldt *et al.* 2012; Yang *et al.* 2017; Jacovas *et al.* 2018) and reflected the population-specific physiological adaptations to high elevation (Eichstaedt *et al.* 2014; Gnecchi-Ruscone *et al.* 2018), and (2) genetic variants that were associated with phenotypic variation in one population did not show the same association in other populations (Alkorta-Aranburu *et al.* 2012).

In contrast, there is evidence of a shared genomic basis to adaptation in some domestic animal breeds with independent high elevation origins (Wu *et al.* 2020). For example, an identical variant in a transcription factor binding site upstream of *Fgf7* is under selection in two sheep breeds in the Himalayas (Gorkhali *et al.* 2016). Another study reported identical variants with a signature of selection in breeds of high- and mid-elevation pigs (Dong *et al.* 2014). On the whole, parallel selection at the SNP level appears to be comparatively rare; more frequently, a small number of genes are reported to be under convergent high elevation associated selection in two independent, closely related lineages (Hendrickson 2013; Ai *et al.* 2014; Dong *et al.* 2014; Edea *et al.* 2019; Liu *et al.* 2019). Thus, there is some evidence for a shared genomic basis for high elevation adaptation in domestic animals, but the pattern is mixed and may be confounded by gene flow (Huerta-Sánchez *et al.* 2014, but see Lou *et al.* 2015; Miao *et al.* 2017; Signore *et al.* 2019).

Here, we found that on a broad scale the response to selection at high elevation was population-specific. Given the recent ancestry of these house mouse populations and the fact that most selection was probably acting on standing variation rather than on new mutations, the absence of a stronger shared response to selection is perhaps surprising. What might account for this? First, we note that adaptation to high elevation is exceedingly complex and may involve changes in many integrated physiological systems (Storz and Scott 2019). For highly polygenic suites of traits, responses in different subsets of genes may be more likely. Second, while both transects extend from sea level to high elevations, they also differ in many environmental variables that may impose distinct selection pressures. The high elevation habitat in Ecuador is more temperate, more humid, and less seasonal than the high elevation environment in Bolivia. Moreover, the highest elevation sampled in Ecuador was 2900 m, while the highest elevation sampled in Bolivia was 3800 m. Differing responses to selection may reflect some of these environmental differences. Last, it is possible that the same underlying *causative* SNPs are present in both transects, but occur on different haplotype backgrounds. This scenario is reasonable given the likelihood that selection has acted on standing variation where functional mutations may have had an opportunity to recombine and become associated with different haplotypes.

Evidence for a threshold effect

Our sampling strategy enabled us to ask about large changes in allele frequency at high elevations. We identified a small number of PBS*n1* candidate gene regions that demonstrated a steep transition between populations at and above 2900 m. We found that 33% of these LAFC regions in the northern transect are implicated in oxygen transport and cardiovascular development. *Esrp1*, which is involved in branching epithelium of blood vessels and lungs, exhibited one of the greatest allele frequency shifts in the entire genome. In the southern transect, a surprising 50% of LAFC regions contained genes involved in heart morphogenesis, erythropoiesis, vasculature development, or angiogenesis. We originally hypothesized a threshold effect for elevation based on the risk of certain human diseases increasing disproportionately at higher elevations. Interestingly, one of the LAFC genes we discovered was *Aplnr*, the apelin receptor. In humans, *Aplnr* variants affect the risk of high-altitude pulmonary edema (Mishra *et al.* 2015), a disease which increases significantly above 3000 m (West 2004). Another four LAFC genes (*Slc4a1*, *Nadk*, *Epo*, and *Rac1*) are implicated in pulmonary artery pressure in humans (Wojciak-Stothard *et al.* 2006; Karamanian *et al.* 2014; Patel *et al.* 2017; Zhao

Downloaded from https://academic.oup.com/genetics/article/220/2/iyab226/6460342 by University of California Berkeley user on 07 February 2022

et al. 2021) which increases exponentially above ~3000 m (Penaloza and Arias-Stella 2007).

Conclusion

We identified a number of shared genomic patterns in response to selection in two independent elevational transects in house mice. These included (1) a genome-wide distribution of putative targets of selection, (2) short-genomic distances over which signatures of selection extend, (3) an excess of derived, high-frequency alleles in high-elevation populations for candidate SNPs but not for other SNPs, (4) reduced nucleotide diversity in high elevation populations relative to low elevation populations, and (5) a small set of genes showing large changes in allele frequency only at the highest elevations. These observations are consistent with selection acting on standing genetic variation at many loci and primarily at high elevations. Despite these broad similarities, the identity of genes showing signatures of selection was largely distinct in each transect, similar to patterns seen in humans and other species. Finally, we identified a small subset of genes with strong evidence of selection in both transects. Of course, such patterns are only one step in understanding the genetic basis of adaptation, but the genes identified here provide a good starting point for evaluating the effects of specific mutations in functional studies. In this regard, the house mouse is a useful model, since it exhibits significant variation in the wild and is also amenable to manipulation in the lab.

Data availability

Sequences are available at the Sequence Read Archive, BioProject ID PRJNA766703. Supplementary File S1 includes full locality information, sex and Museum of Vertebrate Zoology vouchered specimen number for all sequenced individuals. Supplementary File S2 includes supplemental methods, including text, model figures, and complete code for demographic simulations. Supplementary material is available at figshare: https://doi.org/10.25386/genetics.16695040.

Acknowledgments

The authors thank Santiago Borneo, Alejandra Camacho, and Daniel Chavez from the Universidad Católica de Quito for facilitating work in Ecuador, and they thank Marcel Caminer and Cristian Gangotena for help in the field. They would like to thank Lydia Smith for her significant contributions in the laboratory. They also thank Olivier François and Megan Phifer-Rixey for their technical advice. Last, they thank Scott Edwards and Dave Begun as well as three anonymous reviewers for their helpful comments on the manuscript.

Funding

This work was supported by the National Institutes of Health (grants R01 GM074245 and R01 GM127468 to M.W.N., R01 HD073439, R01 HD094787, and R01 GM098536 to J.M.G.) and by grant BEX 8985/11-1 from Coordenacão de Aperfeiçoamento de Pessoal de nível Superior (to F.M.). A.S.C. was supported by an National Science Foundation postdoctoral fellowship (PRFB-1402539). This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by the National Science Foundation grant number ACI-1053575.

Conflicts of interest

The authors declare that there is no conflict of interest.

Literature cited

Ai H, Yang B, Li J, Xie X, Chen H, *et al.* 2014. Population history and genomic signatures for high-altitude adaptation in Tibetan pigs. BMC Genomics. 15:834.doi:10.1186/1471–2164-15–834.

Aldenderfer MS. 2003. Moving up in the world: archaeologists seek to understand how and when people came to occupy the Andean and Tibetan Plateaus. Am Sci. 91:542–549. https://www.jstor.org/stable/27858304.

Aldenderfer MS. 2011. Peopling the Tibetan Plateau: insights from archaeology. High Alt Med Biol. 12:141–147. doi:10.1089/ham.2010.1094.

Alexander DH, Lange K. 2011. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics. 12:246.doi:10.1186/1471–2105-12–246.

Alkorta-Aranburu G, Beall CM, Witonsky DB, Gebremedhin A, Pritchard JK, *et al.* 2012. The genetic architecture of adaptations to high altitude in Ethiopia. PLoS Genet. 8:e1003110.doi:10.1371/journal.pgen.1003110.

Arciero E, Kraaijenbrink Asan T, Haber M, Mezzavilla M, *et al.* 2018. Demographic history and genetic adaptation in the Himalayan region inferred from genome-wide SNP genotypes of 49 populations. Mol Biol Evol. 35:1916–1933. doi:10.1093/molbev/msy094.

Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, *et al.* 2013. From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. Curr Protoc Bioinformatics. 43:11.10.1–11.10.33. doi:10.1002/0471250953.bi1110s43.

Beall CM. 2006. Andean, Tibetan, and Ethiopian patterns of adaptation to high-altitude hypoxia. Integr Comp Biol. 46:18–24. doi:10.1093/icb/icj004.

Beall CM. 2007. Two routes to functional adaptation: Tibetan and Andean high-altitude natives. Proc Natl Acad Sci USA. 104:8655–8660. doi:10.1073/pnas.0701985104.

Beall CM, Cavalleri GL, Deng L, Elston RC, Gao Y, *et al.* 2010. Natural selection on EPAS1 (HIF2) associated with low hemoglobin concentration in Tibetan highlanders. Proc Natl Acad Sci USA. 107:11459–11464. doi:10.1073/pnas.1002443107.

Ben-Yosef Y, Miller A, Shapiro S, Lahat N. 2005. Hypoxia of endothelial cells leads to MMP-2-dependent survival and death. Am J Physiol Cell Physiol. 289:C1321–C1331. doi:10.1152/ajpcell.00079.2005.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B Stat Methodol. 57:289–300. doi:10.1111/j.2517–6161.1995.tb02031.x.

Bi K, Linderoth T, Vanderpool D, Good JM, Nielsen R, *et al.* 2013. Unlocking the vault: next-generation museum population genomics. Mol Ecol. 22:6018–6032. doi:10.1111/mec.12516.

Bigham A, Bauchet M, Pinto D, Mao X, Akey JM, *et al.* 2010. Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data. PLoS Genet. 6:e1001116.doi:10.1371/journal.pgen.1001116.

Bigham AW, Mao X, Mei R, Brutsaert T, Wilson MJ, *et al.* 2009. Identifying positive selection candidate loci for high-altitude adaptation in Andean populations. Hum Genomics. 4:79–90. doi:10.1186/1479–7364-4-2-79.

Blockus H, Chédotal A. 2016. Slit-robo signaling. Development. 143:3037–3044. doi:10.1242/dev.132829.

Bult CJ, Blake JA, Smith CL, Kadin JA, Richardson JE; Mouse Genome Database Group. 2019. Mouse Genome Database (MGD) 2019. Nucleic Acids Res. 47:D801–D806. doi:10.1093/nar/gky1056.

Downloaded from https://academic.oup.com/genetics/article/220/2/iyab226/6460342 by University of California Berkeley user on 07 February 2022

Caye K, Jumentier B, Lepeule J, François O. 2019. LFMM 2: fast and accurate inference of gene-environment associations in genome-wide studies. Mol Biol Evol. 36:852–860. doi:10.1093/molbev/msz008.

Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, *et al.* 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience. 4:7. doi:10.1186/s13742-015–0047-8.

Chen S, Zhou Y, Chen Y, Gu J. 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 34:i884–i890. doi:10.1093/bioinformatics/bty560.

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 USA. 109:8635–8640. doi:10.1073/pnas.1120523109.

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. 68:48–62. doi:10.1111/evo.12257.

Childebayeva A, Goodrich JM, Leon-Velarde F, Rivera-Chira M, Kiyamu M, *et al.* 2021. Genome-wide epigenetic signatures of adaptive developmental plasticity in the Andes. Genome Biol Evol. 13:evaa239. doi:10.1093/gbe/evaa239

Chu W, Li X, Li C, Wan L, Shi H, *et al.* 2011. TGFBR3, a potential negative regulator of TGF-β signaling, protects cardiac fibroblasts from hypoxia-induced apoptosis. J Cell Physiol. 226:2586–2594. doi:10.1002/jcp.22604.

Cobb GB. 1949. Supply and transportation for the Potosí mines, 1545–1640. Hispanic Am Historical Rev. 29:25–45. https://www.jstor.org/stable/2508292.

Conniff ML. 1977. Guayaquil through independence: urban development in a colonial system. The Americas 33:385–410. https://www.jstor.org/stable/980945.

Crawford JE, Amaru R, Song J, Julian CG, Racimo F, *et al.* 2017. Natural selection on genes related to cardiovascular health in high-altitude adapted Andeans. Am J Hum Genet. 101:752–767. : 10.1016/j.ajhg.2017.09.023.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, *et al.*; 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. Bioinformatics. 27:2156–2158. doi:10.1093/bioinformatics/btr330.

De Miranda MA, Jr, Schlater AE, Green TL, Kanatous SB. 2012. In the face of hypoxia: myoglobin increases in response to hypoxic conditions and lipid supplementation in cultured Weddell seal skeletal muscle cells. J Exp Biol. 215:806–813. doi:10.1242/jeb.060681.

Dong K, Yao N, Pu Y, He X, Zhao Q, *et al.* 2014. Genomic scan reveals loci under altitude adaptation in Tibetan and Dahe pigs. PLoS One. 9:e110520. doi:10.1371/journal.pone.0110520.

Dray S, Dufour A. 2007. The ade4 package: implementing the duality diagram for ecologists. J Stat Soft. 22:1–20. doi:10.18637/jss.v022.i04.

Dzal YA, Milsom WK. 2019. Hypoxia alters the thermogenic response to cold in adult homeothermic and heterothermic rodents. J Physiol. 597:4809–4829. doi:10.1113/JP277183.

Edea Z, Dadi H, Dessie T, Kim KS. 2019. Genomic signatures of high-altitude adaptation in Ethiopian sheep populations. Genes Genom. 41:973–981. doi:10.1007/s13258-019–00820-y.

Eichstaedt CA, Antão T, Pagani L, Cardona A, Kivisild T, *et al.* 2014. The Andean adaptive toolkit to counteract high altitude maladaptation: genome-wide and phenotypic analysis of the Collas. PLoS One. 9:e93314. e93314. doi:10.1371/journal.pone.0093314.

Eichstaedt CA, Pagani L, Antao T, Inchley CE, Cardona A, *et al.* 2017. Evidence of early-stage selection on EPAS1 and GPR126 genes in Andean high altitude populations. Sci Rep. 7:1–9. doi:10.1038/s41598-017–13382-4.

Ferris KG, Chavez AS, Suzuki TA, Beckman EJ, Bi K, *et al.* 2021. The genomics of rapid climatic adaptation and parallel evolution in North American house mice. PLoS Genet. 17:e1009495.doi:10.1371/journal.pgen.1009495

Foll M, Gaggiotti OE, Daub JT, Vatsiou A, Excoffier L. 2014. Widespread signals of convergent adaptation to high altitude in Asia and America. Am J Hum Genet. 95:394–407. doi:10.1016/j.ajhg.2014.09.002.

Frichot E, Schoville SD, Bouchard G, François O. 2013. Testing for associations between loci and environmental gradients using latent factor mixed models. Mol Biol Evol. 30:1687–1699. doi:10.1093/molbev/mst063.

Gabry AL, Ledoux X, Mozziconacci M, Martin C. 2003. High-altitude pulmonary edema at moderate altitude (< 2,400 m; 7,870 feet): a series of 52 patients. Chest. 123:49–53. doi:10.1378/chest.123.1.49.

Gaudet P, Livstone MS, Lewis SE, Thomas PD. 2011. Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium. Brief Bioinform. 12:449–462. doi:10.1093/bib/bbr042.

Ge RL, Cai Q, Shen YY, San A, Ma L, *et al.* 2013. Draft genome sequence of the Tibetan antelope. Nat Commun. 4:1858.doi:10.1038/ncomms2860.

Gel B, Diez-Villanueva A, Serra E, Buschbeck M, Peinado MA, *et al.* 2016. regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. Bioinformatics. 32:btv562.doi:10.1093/bioinformatics/btv562.

Genet G, Boyé K, Mathivet T, Ola R, Zhang F, *et al.* 2019. Endophilin-A2 dependent VEGFR2 endocytosis promotes sprouting angiogenesis. Nat Commun. 10:2350.doi:10.1038/s41467-019–10359-x.

Gering EJ, Opazo JC, Storz JF. 2009. Molecular evolution of cytochrome b in high- and low-altitude deer mice (genus *Peromyscus*). Heredity (Edinb). 102:226–235. doi:10.1038/hdy.2008.124.

Gnecchi-Ruscone GA, Abondio P, De Fanti S, Sarno S, Sherpa M, *et al.* 2018. Evidence of polygenic adaptation to high altitude from Tibetan and Sherpa genomes. Genome Biol Evol. 10:2919–2930. doi:10.1093/gbe/evy233.

Gorkhali NA, Dong K, Yang M, Song S, Kader A, *et al.* 2016. Genomic analysis identified a potential novel molecular mechanism for high- altitude adaptation in sheep at the Himalayas. Sci Rep. 6:29963. doi:10.1038/srep29963.

Guang-Xin, E, Basang W-D, Zhu Y-B. 2019. Whole-genome analysis identifying candidate genes of altitude adaptive ecological thresholds in yak populations. J Anim Breed Genet. 136:371–377. doi:10.1111/jbg.12403.

Hackett PH, Roach RC. 2001. High-altitude illness. N Engl J Med. 345:107–114. doi:10.1056/NEJM200107123450206.

Hanghøj K, Moltke I, Andersen PA, Manica A, Korneliussen TS. 2019. Fast and accurate relatedness estimation from high-throughput sequencing data in the presence of inbreeding. GigaScience. 8:giz034. doi:10.1093/gigascience/giz034.

Hanratty DM. 1991. Ecuador: A Country Study. Washington, D.C.: United States Govt Printing Office.

Harr B, Karakoc E, Neme R, Teschke M, Pfeifle C, *et al.* 2016. Genomic resources for wild populations of the house mouse, *Mus musculus* and its close relative *Mus spretus*. Sci Data. 3:160075. doi:10.1038/sdata.2016.75.

Hendrickson SL. 2013. A genome wide study of genetic adaptation to high altitude in feral Andean Horses of the páramo. BMC Evol Biol. 13:273.doi:10.1186/1471–2148-13–273.

Downloaded from https://academic.oup.com/genetics/article/220/2/iyab226/6460342 by University of California Berkeley user on 07 February 2022

Hoxha E, Marcinnò A, Montarolo F, Masante L, Balbo I, *et al.* 2019. Emerging roles of Fgf14 in behavioral control. Behav Brain Res. 356:257–265. doi:10.1016/j.bbr.2018.08.034.

Hu H, Petousi N, Glusman G, Yu Y, Bohlender R, *et al.* 2017. Evolutionary history of Tibetans inferred from whole-genome sequencing. PLoS Genet. 13:e1006675.doi:10.1371/journal.pgen.1006675.

Huerta-Sánchez E, DeGiorgio M, Pagani L, Tarekegn A, Ekong R, *et al.* 2013. Genetic signatures reveal high-altitude adaptation in a set of Ethiopian populations. Mol Biol Evol. 30:1877–1888.. doi: 10.1093/molbev/mst089.

Huerta-Sánchez E, Jin Asan X, Bianba Z, Peter BM, *et al.* 2014. Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature. 512:194–197. doi:10.1038/nature13408.

Jacovas VC, Couto-Silva CM, Nunes K, Lemes RB, de Oliveira MZ, *et al.* 2018. Selection scan reveals three new loci related to high altitude adaptation in Native Andeans. Sci Rep. 8:12733.doi: 10.1038/s41598-018–31100-6.

Jeong C, Witonsky DB, Basnyat B, Neupane M, Beall CM, *et al.* 2018. Detecting past and ongoing natural selection among ethnically Tibetan women at high altitude in Nepal. PLoS Genet. 14: e1007650.doi:10.1371/journal.pgen.1007650.

Jochmans-Lemoine A, Villalpando G, Gonzales M, Valverde I, Soria R, *et al.* 2015. Divergent physiological responses in laboratory rats and mice raised at high altitude. J Exp Biol. 218:1035–1043. doi: 10.1242/jeb.112862.

Jha AR, Zhou D, Brown CD, Kreitman M, Haddad GG, *et al.* 2016. Shared genetic signals of hypoxia adaptation in *Drosophila* and in high-altitude human populations. Mol Biol Evol. 33:501–517. doi: 10.1093/molbev/msv248.

Karamanian VA, Harhay M, Grant GR, Palevsky HI, Grizzle WE, *et al.* 2014. Erythropoietin upregulation in pulmonary arterial hypertension. Pulm Circ. 4:269–279. doi:10.1086/675990.

Katz D, Shore S, Bandle B, Niermeyer S, Bol KA, *et al.* 2015. Sudden infant death syndrome and residential altitude. Pediatrics. 135: e1442–e1449. doi:10.1542/peds.2014–2697.

Kolberg L, Raudvere U, Kuzmin I, Vilo J, Peterson H. 2020. gprofiler2—an R package for gene list functional enrichment analysis and namespace conversion toolset g: profiler. F1000Research. 9: 709.doi:10.12688/f1000research.24956.2.

Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: analysis of next generation sequencing data. BMC Bioinformatics. 15: 356.doi:10.1186/s12859-014–0356-4.

Laurie CC, Nickerson DA, Anderson AD, Weir BS, Livingston RJ, *et al.* 2007. Linkage disequilibrium in wild mice. PLoS Genet. 3: e144.doi:10.1371/journal.pgen.0030144.

Leonard OE. 1948. La Paz, Bolivia: its population and growth. Am Sociol Rev. 13:448–454.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v2 [q-bio.GN].

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, *et al.*; 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map (SAM) format and SAMtools. Bioinformatics. 25: 2078–2079. doi:10.1093/bioinformatics/btp352.

Li M, Tian S, Jin L, Zhou G, Li Y, *et al.* 2013. Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. Nat Genet. 45:1431–1438. doi:10.1038/ng.2811.

Li X, Eriksson U. 2001. Novel VEGF family members: VEGF-B, VEGF-C and VEGF-D. Int J Biochem Cell Biol. 33:421–426. doi: 10.1016/s1357-2725(01)00027–9.

Li Y, Wu D-D, Boyko AR, Wang G-D, Wu S-F, *et al.* 2014. Population variation revealed high-altitude adaptation of Tibetan mastiffs. Mol Biol Evol. 31:1200–1205. doi:10.1093/molbev/msu070.

Liu X, Zhang Y, Li Y, Pan J, Wang D, *et al.* 2019. EPAS1 gain-of-function mutation contributes to high-altitude adaptation in Tibetan horses. Mol Biol Evol. 36:2591–2603. doi: 10.1093/molbev/msz158.

Lorenzo FR, Huff C, Myllymaki M, Olenchock B, Swierczek S, *et al.* 2014. A genetic mechanism for Tibetan high-altitude adaptation. Nat Genet. 46:951–956. doi:10.1038/ng.3067.

Lou H, Lu Y, Lu D, Fu R, Wang X, *et al.* 2015. A 3.4-kb copy-number deletion near EPAS1 is significantly enriched in high-altitude Tibetans but absent from the Denisovan sequence. Am J Hum Genet. 97:54–66. doi:10.1016/j.ajhg.2015.05.005.

Lou JY, Laezza F, Gerber BR, Xiao M, Yamada K, *et al.* 2005. Fibroblast growth factor 14 is an intracellular modulator of voltage-gated sodium channels. J Physiol. 569:179–193. doi:10.1113/jphysiol.2005.097220.

Lynch CB. 1992. Clinal variation in cold adaptation in *Mus domesticus*: verification of predictions from laboratory populations. Am Nat. 139:1219–1236. doi:10.1086/285383.

Ma YF, Han XM, Huang CP, Zhong L, Adeola AC, *et al.* 2019. Population genomics analysis revealed origin and high-altitude adaptation of Tibetan pigs. Sci Rep. 9:11463. doi: 10.1038/s41598-019–47711-6.

Marchi N, Excoffier L. 2020. Gene flow as a simple cause for an excess of high-frequency-derived alleles. Evol Appl. 13:2254–2263. doi: 10.1111/eva.12998.

Martin SH, Van Belleghem SM. 2017. Exploring evolutionary relationships across the genome using topology weighting. Genetics. 206: 429–438. doi:10.1534/genetics.116.194720.

McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, *et al.* 2016. The Ensembl variant effect predictor. Genome Biol. 17:122. doi: 10.1186/s13059-016–0974-4.

Meisner J, Albrechtsen A. 2018. Inferring population structure and admixture proportions in low-depth NGS data. Genetics. 210: 719–731. doi:10.1534/genetics.118.301336.

Meyer M, Kircher M. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb Protoc. 2010:pdb.prot5448.

Miao B, Wang Z, Li Y. 2017. Genomic analysis reveals hypoxia adaptation in the Tibetan mastiff by introgression of the Gray wolf from the Tibetan Plateau. Mol Biol Evol. 34:734–743. doi: 10.1093/molbev/msw274.

Mishra A, Kohli S, Dua S, Thinlas T, Mohammad G, *et al.* 2015. Genetic differences and aberrant methylation in the apelin system predict the risk of high-altitude pulmonary edema. Proc Natl Acad Sci USA. 112:e3311. doi:10.1073/pnas.1509391112.

Monge C, León-Velarde F. 1991. Physiological adaptation to high altitude: oxygen transport in mammals and birds. Physiol Rev. 71: 1135–1172. doi:10.1152/physrev.1991.71.4.1135.

Newsom LA. 1995. Life and Death in Early Colonial Ecuador. Norman, OK: University of Oklahoma Press.

Pamenter ME, Hall JE, Tanabe Y, Simonson TS. 2020. Cross-species insights into genomic adaptations to hypoxia. Front Genet. 11: 743. doi:10.3389/fgene.2020.00743.

Patel RS, Masi S, Taddei S. 2017. Understanding the role of genetics in hypertension. Eur Heart J. 38:2309–2312. doi:10.1093/eurheartj/ehx273.

Penaloza D, Arias-Stella J. 2007. The heart and pulmonary circulation at high altitudes: healthy highlanders and chronic mountain

Downloaded from https://academic.oup.com/genetics/article/220/2/iyab226/6460342 by University of California Berkeley user on 07 February 2022

sickness. Circulation. 115:1132–1146. doi:10.1161/CIRCULATIONAHA.106.624544.

Peng Y, Yang Z, Zhang H, Cui C, Qi X, *et al.* 2011. Genetic variations in Tibetan populations and high-altitude adaptation at the Himalayas. Mol Biol Evol. 28:1075–1081. doi:10.1093/molbev/msq290.

Phifer-Rixey M, Bi K, Ferris KG, Sheehan MJ, Lin D, *et al.* 2018. The genomic basis of environmental adaptation in house mice. PLoS Genet. 14:e1007672.doi:10.1371/journal.pgen.1007672.

Pleurdeau D. 2006. Human technical behavior in the African middle stone age: the lithic assemblage of Porc-Epic Cave (Dire Dawa, Ethiopia). Afr Archaeol Rev. 22:177–197.

Pluzhnikov A, Donnelly P. 1996. Optimal sequencing strategies for surveying molecular genetic diversity. Genetics. 144:1247–1262.

Pocock M, Hauffe H, Searle J. 2005. Dispersal in house mice. Biol J Linn Soc. 84:565–583. doi:10.1111/j.1095–8312.2005.00455.x.

Pokreisz P, Fleming I, Kiss L, Barbosa-Sicard E, Fisslthaler B, *et al.* 2006. Cytochrome P450 epoxygenase gene function in hypoxic pulmonary vasoconstriction and pulmonary vascular remodeling. Hypertension. 47:762–770. doi:10.1161/01.hyp.0000208299.62535.58.

Pourhaghighi R, Ash PEA, Phanse S, Goebels F, Hu LZM, *et al.* 2020. BraInMap elucidates the macromolecular connectivity landscape of mammalian brain. Cell Syst. 10:333–350.e14. doi:10.1016/j.cels.2020.03.003.

Qiu Q, Zhang G, Ma T, Qian W, Wang J, *et al.* 2012. The yak genome and adaptation to life at high altitude. Nat Genet. 44:946–949. doi:10.1038/ng.2343.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 26:841–842. doi:10.1093/bioinformatics/btq033.

R Core Team. 2019. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/

Rademaker K, Hodgins G, Moore K, Zarrillo S, Miller C, *et al.* 2014. Paleoindian settlement of the high-altitude Peruvian Andes. Science. 346:466–469. doi:10.1126/science.1258260.

Reynolds J, Weir BS, Cockerham CC. 1983. Estimation of the coancestry coefficient: basis for a short-term genetic distance. Genetics. 105:767–779.

Richardson MF, Munyard K, Croft LJ, Allnutt TR, Jackling F, *et al.* 2019. Chromosome-level alpaca reference genome VicPac3.1 improves genomic insight into the biology of new world camelids. Front Genet. 10:58.doi:10.3389/fgene.2019.00586.

Scheet P, Stephens M. 2006. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. Am J Hum Genet. 78:629–644. doi:10.1086/502802.

Scheinfeldt LB, Soi S, Thompson S, Ranciaro A, Woldemeskel D, *et al.* 2012. Genetic adaptation to high altitude in the Ethiopian highlands. Genome Biol. 13:R1. doi:10.1186/gb-2012-13-1-r1.

Schweizer RM, Velotta JP, Ivy CM, Jones MR, Muir SM, *et al.* 2019. Physiological and genomic evidence that selection on the transcription factor *Epas1* has altered cardiovascular function in high-altitude deer mice. PLoS Genet. 15:e1008420. doi:10.1371/journal.pgen.1008420.

Shorter JR, Huang W, Beak JY, Hua K, Gatti DM, *et al.* 2018. Quantitative trait mapping in Diversity Outbred mice identifies two genomic regions associated with heart size. Mamm Genome. 29:80–89. doi:10.1007/s00335-017–9730-7.

Signore AV, Storz JF. 2020. Biochemical pedomorphosis and genetic assimilation in the hypoxia adaptation of Tibetan antelope. Sci Adv. 6:eabb5447. doi:10.1126/sciadv.abb5447.

Signore AV, Yang Y-Z, Yang Q-Y, Qin G, Moriyama H, *et al.* 2019. Adaptive changes in hemoglobin function in high-altitude Tibetan canids were derived via gene conversion and introgression. Mol Biol Evol. 36:2227–2237. doi:10.1093/molbev/msz097.

Simonson TS, Yang Y, Huff CD, Yun H, Qin G, *et al.* 2010. Genetic evidence for high-altitude adaptation in Tibet. Science. 329:72–75. doi:10.1126/science.1189406.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30:1312–1313. doi:10.1093/bioinformatics/btu033.

Staples J, Nickerson DA, Below JE. 2013. Utilizing graph theory to select the largest set of unrelated individuals for genetic analysis. Genet Epidemiol. 37:136–141. doi:10.1002/gepi.21684.

Storz JF. 2016. Causes of molecular convergence and parallelism in protein evolution. Nat Rev Genet. 17:239–250. doi:10.1038/nrg.2016.11.

Storz JF. 2021. High-altitude adaptation: mechanistic insights from integrated genomics and physiology. Mol Biol Evol. 38:2677–2691. doi:10.1093/molbev/msab064.

Storz JF, Baze M, Waite JL, Hoffmann FG, Opazo JC, *et al.* 2007. Complex signatures of selection and gene conversion in the duplicated globin genes of house mice. Genetics. 177:481–500. doi:10.1534/genetics.107.078550.

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. 100:910–922. doi:10.1093/jmammal/gyy173.

Storz JF, Scott GR. 2019. Life ascending: mechanism and process in physiological adaptation to high-altitude hypoxia. Annu Rev Ecol Evol Syst. 50:503–522. doi:10.1146/annurev-ecolsys-110218-025014.

Storz JF, Scott GR, Cheviron ZA. 2010. Phenotypic plasticity and genetic adaptation to high-altitude hypoxia in vertebrates. J Exp Biol. 213:4125–4136. doi:10.1242/jeb.048181.

Strupp M, Maul S, Konte B, Hartmann AM, Giegling I, *et al.* 2020. A variation in FGF14 is associated with downbeat nystagmus in a genome-wide association study. Cerebellum. 19:348–357. doi:10.1007/s12311-020–01113-x.

Super JC. 1979. Partnership and profit in the early Andean trade: the experiences of Quito merchants, 1580. J Lat Am Stud. 11:265–281.

Suzuki TA, Martins FM, Nachman MW. 2019. Altitudinal variation of the gut microbiota in wild house mice. Mol Ecol. 28:2378–2390. doi:10.1111/mec.14905.

Szpiech ZA, Novak TE, Bailey NP, Stevison LS. 2020. Application of a novel haplotype-based scan for local adaptation to study high-altitude adaptation in rhesus macaques. bioRxiv 2020.05.19.104380. doi:10.1101/2020.05.19.104380.

Ton QV, Leino D, Mowery SA, Bredemeier NO, Lafontant PJ, *et al.* 2018. Collagen COL22A1 maintains vascular stability and mutations in COL22A1 are potentially associated with intracranial aneurysms. Dis Models Mech. 11:dmm033654. doi:10.1242/dmm.033654.

Towns J, Cockerill T, Dahan M, Foster I, Gaither K, *et al.* 2014. XSEDE: accelerating scientific discovery. Comput Sci Eng. 16:62–74. doi:10.1109/MCSE.2014.80.

Tseng HY, Samarelli AV, Kammerer P, Scholze S, Ziegler T, *et al.* 2018. LCP1 preferentially binds clasped αMβ2 integrin and attenuates leukocyte adhesion under flow. J Cell Sci. 131:jcs218214. doi:10.1242/jcs.218214.

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. 37:2309–2321. doi:10.1093/molbev/msaa086.

Downloaded from https://academic.oup.com/genetics/article/220/2/iyab226/6460342 by University of California Berkeley user on 07 February 2022

Wang M-S, Li Y, Peng M-S, Zhong L, Wang Z-J, *et al.* 2015. Genomic analyses reveal potential independent adaptation to high altitude in Tibetan chickens. Mol Biol Evol. 32:1880–1889. doi: 10.1093/molbev/msv071.

Wang Q, Bardgett ME, Wong M, Wozniak DF, Lou J, *et al.* 2002. Ataxia and paroxysmal dyskinesia in mice lacking axonally transported FGF14. Neuron. 35:25–38. doi:10.1016/S0896-6273(02)00744-4

Wang X, Liu J, Zhou G, Guo J, Yan H, *et al.* 2016. Whole-genome sequencing of eight goat populations for the detection of selection signatures underlying production and adaptive traits. Sci Rep. 6: 38932.doi:10.1038/srep38932.

Watanabe T, Frost DAB, Mlakar L, Heywood J, Da Silveira WA, *et al.* 2019. A human skin model recapitulates systemic sclerosis dermal fibrosis and identifies COL22A1 as a TGFβ early response gene that mediates fibroblast to myofibroblast transition. Genes. 10:75. doi:10.3390/genes10020075.

Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. Evolution. 38:1358–1370. doi: 10.1111/j.1558–5646.1984.tb05657.x

West JB; American Physiological Society 2004. The physiologic basis of high-altitude diseases. Ann Intern Med. 141:789–800. doi: 10.7326/0003–4819-141-10-200411160-00010.

Witt KE, Huerta-Sanchez E. 2019. Convergent evolution in human and domesticate adaptation to high-altitude environments. Phil Trans R Soc B. 374:20180235.doi:10.1098/rstb.2018.0235.

Wojciak-Stothard B, Tsang LYF, Paleolog E, Hall SM, Haworth SG. 2006. Rac1 and RhoA as regulators of endothelial phenotype and barrier function in hypoxia-induced neonatal pulmonary hypertension. Am J Physiol Lung Cell Mol Physiol. 290:L1173–L1182. doi:10.1152/ajplung.00309.2005.

Wozniak DF, Xiao M, Xu L, Yamada KA, Ornitz DM. 2007. Impaired spatial learning and defective theta burst induced LTP in mice lacking fibroblast growth factor 14. Neurobiol Dis. 26:14–26. doi: 10.1016/j.nbd.2006.11.014.

Wu D-D, Yang C-P, Wang M-S, Dong K-Z, Yan D-W, *et al.* 2020. Convergent genomic signatures of high-altitude adaptation among domestic mammals. Natl Sci Rev. 7:952–963. doi: 10.1093/nsr/nwz213.

Wysocka MB, Pietraszek-Gremplewicz K, Nowak D. 2018. The role of apelin in cardiovascular diseases, obesity and cancer. Front Physiol. 9:557. doi:10.3389/fphys.2018.00557.

Xing J, Wuren T, Simonson TS, Watkins WS, Witherspoon DJ, *et al.* 2013. Genomic analysis of natural selection and phenotypic variation in high-altitude Mongolians. PLoS Genet. 9:e1003634. doi: 10.1371/journal.pgen.1003634.

Xu S, Li S, Yang Y, Tan J, Lou H, *et al.* 2011. A genome-wide search for signals of high-altitude adaptation in Tibetans. Mol Biol Evol. 28: 1003–1011. doi:10.1093/molbev/msq277.

Yamazaki D, Itoh T, Miki H, Takenawa T. 2013. SrGAP1 regulates lamellipodial dynamics and cell migratory behavior by modulating Rac1 activity. Mol Biol Cell. 24:3393–3405. doi: 10.1091/mbc.E13-04–0178.

Yang J, Jin Z-B, Chen J, Huang X-F, Li X-M, *et al.* 2017. Genetic signatures of high-altitude adaptation in Tibetans. Proc Natl Acad Sci USA. 114:4189–4194. doi:10.1073/pnas.1617042114.

Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo Z, *et al.* 2010. Sequencing of 50 human exomes reveals adaptation to high altitude. Science. 329:75–78. doi:10.1126/science.1190371.

Yu L, Wang G-D, Ruan J, Chen Y-B, Yang C-P, *et al.* 2016. Genomic analysis of snub-nosed monkeys (*Rhinopithecus*) identifies genes and processes related to high-altitude adaptation. Nat Genet. 48: 947–952. doi:10.1038/ng.3615.

Zhang W, Fan Z, Han E, Hou R, Zhang L, *et al.* 2014. Hypoxia adaptations in the Grey Wolf (*Canis lupus chanco*) from Qinghai-Tibet Plateau. PLoS Genet. 10:e1004466.doi:10.1371/journal.pgen.1004466.

Zhang XL, Ha BB, Wang SJ, Chen ZJ, Ge JY, *et al.* 2018. The earliest human occupation of the high-altitude Tibetan Plateau 40 thousand to 30 thousand years ago. Science. 362:1049–1051. doi: 10.1126/science.aat8824.

Zhao E, Xie H, Zhang Y. 2021. Identification of differentially expressed genes associated with idiopathic pulmonary arterial hypertension by integrated bioinformatics approaches. J Comput Biol. 28:79–88. doi:10.1089/cmb.2019.0433.

Communicating editor: S. Edwards

Downloaded from https://academic.oup.com/genetics/article/220/2/iyab226/6460342 by University of California Berkeley user on 07 February 2022