Article Type: Letter

High-altitude adaptation in rhesus macaques

Zachary A. Szpiech†
Department of Biological Sciences
Auburn University
Auburn, AL 36842, USA
zas0020@auburn.edu

*Taylor E. Novak**
Department of Biological Sciences
Auburn University
Auburn, AL 36842, USA
tep0007@auburn.edu

*Nicholas P. Bailey**
Department of Biological Sciences
Auburn University
Auburn, AL 36842, USA
npb0015@auburn.edu

Laurie S. Stevison†
Department of Biological Sciences
Auburn University
Auburn, AL 36842, USA
lss0021@auburn.edu

†*Correspondence: zas0020@auburn.edu, lss0021@auburn.edu*
**Equal Contribution*

Running Title: ***High-altitude adaptation in rhesus macaques***

Keywords: Adaptation, high-altitude, EGLN1, selection, macaque, macaca mulatta

Abstract Word Count (300 max): 216
Total Word Count (5000 max): 3813

Abstract

When natural populations split and migrate to different environments, they may experience different selection pressures that can lead to local adaptation. For aerobic life, the low atmospheric oxygen content of high altitude living presents a special challenge and a strong selection pressure. Searching for evidence of adaptation to high altitude, we compare the whole genomes of 23 wild rhesus macaques captured at high altitude (mean altitude > 4000m above sea level) to 22 wild rhesus macaques captured at low altitude (mean altitude < 500m above sea level). To capture the genomic patterns of a positive selective sweep, we develop XP-nSL, a haplotype-based genomic scan for differential local adaptation with power to detect ongoing and recently completed hard and soft sweeps. We find evidence of local adaptation in the high-altitude population at or near 303 known genes and several unannotated regions. We find the strongest signal for adaptation at EGLN1, a classic target for convergent evolution in several species living in low oxygen environments. Furthermore, many of the 303 genes are involved in processes related to hypoxia, regulation of ROS, DNA damage repair, synaptic signaling, and metabolism. These results suggest that, beyond adapting via a beneficial mutation in one single gene, adaptation to high altitude in rhesus macaques is polygenic and spread across numerous important biological systems.

Impact Summary

Extreme environments pose a challenge to life on multiple fronts. Very high-altitude environments are one such example, with low atmospheric oxygen, increased ultraviolet light exposure, harsh temperatures, and reduced nutrition availability. In spite

of these challenges, many plants and animals, including humans, have genetically adapted to cope with these hardships. Here we study a population of rhesus macaques living at high altitude and compare their genomic patterns with those of a population living much closer to sea level, searching for evidence of genetic changes that are indicative of adaptation to their environment.

When positive selection is ongoing or a beneficial mutation has recently fixed in a population, genetic diversity is reduced in the vicinity of the adaptive allele, and we expect to observe long homozygous haplotypes at high frequency. Here we develop a statistic that summarizes these expected patterns and compares between two populations in order to search for evidence of adaptation that may have occurred in one but not the other. We implement this statistic in a popular and easy-to-use software package.

We find evidence for adaptation at a critical gene that helps control physiological response to low-oxygen, one that has been the target of repeated convergent evolution across many species. We also find evidence for positive selection across a range of traits, including metabolic and neurological. This work helps to explain the evolutionary history of the rhesus macaque and furthers our understanding about the ways organisms genetically adapt to high altitude environments.

Introduction

Selective sweeps produce regions of reduced genetic diversity in the vicinity of an adaptive mutation. These patterns manifest as long extended regions of homozygous haplotypes segregating at high frequency (Przeworski 2002; Sabeti *et al.* 2002; Kim and

94 Nielsen 2004; Garud *et al.* 2015). In the event of a *de novo* mutation that is adaptive in a
95 population, we expect the haplotype it resides on to rapidly rise in frequency in the
96 population (called a 'hard' sweep). On the other hand, if an ancestrally segregating
97 neutral or mildly deleterious allele turned out to be adaptive in a new environment, it
98 would likely reside on two or more haplotypes, which would rapidly rise in frequency in
99 the population (called a 'soft' sweep) (Hermisson and Pennings 2005; Pennings and
100 Hermisson 2006). As both of these processes happen on a time scale faster than
101 mutation or recombination can act to break up the sweeping haplotypes, we expect to
102 observe long and low diversity haplotypes at high frequency in the vicinity of an adaptive
103 mutation. However, if this mutation either does not exist or is not adaptive in a sister
104 population, we would not expect a sweep to occur and thus we would not expect to
105 observe similar haplotype patterns.

106 To capture these haplotype patterns and contrast them between a pair of
107 populations, we develop XP-nSL, a haplotype-based statistic with good power to detect
108 partial, fixed, and recently completed hard and soft sweeps by comparing a pair of
109 populations. XP-nSL is an extension nSL (Ferrer-Admetlla *et al.* 2014) and does not
110 require a genetic recombination map for computation. The lack of dependence on a
111 recombination map is important, as other statistics for identifying positive selection are
112 biased towards low-recombination regions (O'reilly *et al.* 2008), but the approach taken
113 by nSL has been shown to be more robust (Ferrer-Admetlla *et al.* 2014). Both nSL and
114 XP-nSL summarize haplotype diversity by computing the mean number of variant sites
115 in a region that are identical-by-state across all pairs of haplotypes. Whereas nSL
116 contrasts between haplotype pools carrying an ancestral or a derived allele in a single

117 population, XP-nSL contrasts between haplotype pools in two different populations,
118 allowing it to detect differential local adaptation.

119 An extreme example of adaptation to a local environment is the transition to high
120 altitude living. Organisms living at high altitude are confronted with many challenges,
121 including a low-oxygen atmosphere and increased ultraviolet light exposure, and these
122 harsh environments inevitably exert strong selection pressure. Indeed, adaptation to
123 high altitude living has been studied extensively across many organisms from plants,
124 including monocots (Gonzalo-Turpin and Hazard 2009; Ahmad *et al.* 2016) and dicots
125 (Kim and Donohue 2013; Liu *et al.* 2014; Munne-Bosch *et al.* 2016; Guo *et al.* 2018), to
126 numerous animals including amphibians (Yang *et al.* 2016), canids (Li *et al.* 2014; Wang
127 *et al.* 2014; Wang *et al.* 2020), humans (Bigham *et al.* 2009; Bigham *et al.* 2010; Xu *et*
128 *al.* 2010; Yi *et al.* 2010; Peng *et al.* 2011; Huerta-Sanchez *et al.* 2013; Huerta-Sanchez
129 *et al.* 2014; Jeong *et al.* 2014), yaks (Qiu *et al.* 2012), avians (Cai *et al.* 2013; Qu *et al.*
130 2013; Wang *et al.* 2015; Graham and Mccracken 2019), boars (Li *et al.* 2013), mice
131 (Storz *et al.* 2007; Cheviron *et al.* 2012; Schweizer *et al.* 2019; Storz *et al.* 2019; Velotta
132 *et al.* 2020), moles (Campbell *et al.* 2010), antelope (Ge *et al.* 2013), and horses
133 (Hendrickson 2013). Liu *et al.* (2018) recently sequenced and published the whole
134 genomes of 79 wild-born Chinese rhesus macaques collected from multiple sites in
135 China. Among these animals, 23 were sampled from far western Sichuan province in a
136 region with mean altitude > 4000 m above sea level (Liu *et al.* 2018), providing an
137 opportunity to study the genetics of local adaption to high altitude in rhesus macaques.

138 Rhesus macaques are the second most widely distributed primate, with a range
139 extending from Afghanistan to Vietnam and from a latitude of 15 to 38 degrees north

140 (Fooden 2000). Early ancestors of the macaque migrated out of Africa to the Eurasian
141 continent approximately 7 mya—the earliest catarrhine fossils on the continent are
142 macaque-like (Stewart and Disotell 1998). Modern rhesus macaques trace a recent
143 origin to Southeast Asia, with a major migratory split occurring approximately 162 kya
144 separating the ancestors of modern Indian and Chinese rhesus macaques (Hernandez
145 *et al.* 2007). Macaques have proven to be quite evolutionarily successful, demonstrating
146 ecological flexibility and adaptability via developmental plasticity and behavioral changes
147 (Richard *et al.* 1989; Madrid *et al.* 2018). Other studies have looked at how the rhesus
148 macaque radiation has led to population-level adaptation to climate and food availability
149 (Liu *et al.* 2018).

150 Here we test and evaluate our XP-nSL statistic and apply it to study the genomic
151 consequences of high altitude living in the rhesus macaque. We use it to compare the
152 haplotype patterns of the 23 animals from the high-altitude population with another 22
153 that Liu *et al.* (2018) sampled in lower-lying regions in eastern China with a mean
154 altitude < 500 m above sea level.

155 Methods

156 Data Preparation

157 Liu *et al.* (2018) generated whole genome sequencing data for 79 Chinese
158 macaques and called all biallelic polymorphic sites according to GATK best practices
159 using rheMac8, identifying 52,534,348 passing polymorphic autosomal sites. We then
160 filter all loci with > 10% missing data leaving 35,639,395 biallelic sites. Next, the
161 program SHAPEIT v4.1.2 (Delaneau *et al.* 2019) was used to phase haplotypes in the

162 full data set with a genetic map that was available for rheMac8 (Bcm-Hgsc 2020).
163 SHAPEIT performs imputation during phasing for any missing genotypes. Our analyses
164 here focus on 45 of the 79 samples from Liu *et al.* (2018), representing 23 from high-
165 altitude regions of China (*M. m. lasiotis*) and 22 from low-altitude regions of China (*M. m.*
166 *littoralis*), based on capture location information. See Table S1 for individual IDs used.
167 Liu *et al.* (2018) also inferred joint demographic histories for their five populations,
168 and we extract the demographic parameters for our two of interest. This demographic
169 history is recapitulated in Fig. 1 with detailed parameters given in Table 1, which are
170 then used for simulations to test XP-nSL.

171 **A Statistic for Detecting Local Adaptation**

172 We develop a cross-population haplotype-based statistic, XP-nSL, to scan for
173 regions of the genome implicated in differential local adaptation between two
174 populations by extending nSL (Ferrer-Admetlla *et al.* 2014), each of which is defined
175 below.

176 Consider the sets $A(k)$ and $D(k)$, representing the set of haplotypes at site k
177 carrying the ancestral or derived allele, respectively, and let $n_A(k) = |A(k)|$ and
178 $n_D(k) = |D(k)|$. Next $L_{ij}(k)$ is defined as the number of consecutive sites at which
179 haplotype i and j are identical-by-state (IBS) in the interval containing site k. Then nSL
180 at site k is $nS_L(k) = \log\frac{SL_A(k)}{SL_D(k)}$, where $SL_A(k) = \binom{n_A(k)}{2}\sum_{i<j\in A(k)} L_{ij}(k)$ and $SL_D(k) =$
181 $\binom{n_D(k)}{2}\sum_{i<j\in D(k)} L_{ij}(k)$. $SL_A(k)$ and $SL_D(k)$ represent the mean $L_{ij}(k)$ over all pairs of
182 haplotypes carrying either the ancestral or derived allele at locus k, respectively. nSL
183 scores are then normalized genome-wide in frequency bins either with respect to the

184 empirical background or neutral simulations with a matching demographic history. nSL is
185 implemented in `nsl` (Ferrer-Admetlla *et al.* 2014) and `selscan` v1.1.0+ (Szpiech and
186 Hernandez 2014).

187 XP-nSL is defined similarly, except instead of comparing pools of haplotypes
188 containing ancestral or derived alleles, it compares pools of haplotypes between two
189 different populations. Let $P_1(k)$ and $P_2(k)$, represent the set of haplotypes at site k in
190 population 1 and population 2, respectively, and let $n_{P_1}(k) = |P_1(k)|$ and $n_{P_2}(k) =$
191 $|P_2(k)|$. Then XP-nSL at site k is $XPnS_L(k) = \log\frac{SL_{P_1}(k)}{SL_{P_2}(k)}$, where $SL_{P_1}(k) =$
192 $\binom{n_{P_1}(k)}{2}\sum_{i<j\in P_1(k)} L_{ij}(k)$ and $SL_{P_2}(k) = \binom{n_{P_2}(k)}{2}\sum_{i<j\in P_2(k)} L_{ij}(k)$. $SL_{P_1}(k)$ and $SL_{P_2}(k)$
193 represent the mean $L_{ij}(k)$ over all pairs of haplotypes in population 1 or population 2 at
194 locus k, respectively. XP-nSL scores are then normalized genome-wide either with
195 respect to the empirical background or neutral simulations with a matching demographic
196 history. XP-nSL is implemented in `selscan` v1.3.0+ (Szpiech and Hernandez 2014). It
197 is worth noting that XP-nSL is analogous to XP-EHH (Sabeti *et al.* 2007) as nSL is
198 analogous to iHS (Voight *et al.* 2006).

199 The goal of these statistics is to capture a signal of extended regions of low
200 diversity on sweeping haplotypes (emblematic of an ongoing or recently completed
201 selective sweep) within a population (nSL) or on sweeping haplotypes in one population
202 versus another (XP-nSL). When XP-nSL scores are positive this suggests evidence for
203 a hard or soft sweep in population 1, and when XP-nSL scores are negative this
204 suggests evidence for a hard or soft sweep in population 2.

205 **Simulations**

In order to test the ability of XP-nSL to detect ongoing and recently completed hard and soft sweeps, coalescent simulations were performed conditional on an allele frequency trajectory with the program `discoal` (Kern and Schrider 2016). `discoal` simulates an allele frequency trajectory for a single non-neutral allele backwards in time and then simulates a neutral coalescent process conditional on this trajectory. This takes advantage of the speed and efficiency of the coalescent while still being able simulate genetic diversity patterns in the vicinity of a non-neutral locus.

All simulations were run with a two-population divergence demographic history (Fig. 1 and Table 1), as inferred by Liu *et al.* (2018), and given by the following discoal command line arguments `-p 2 46 44 -en 0 1 0.230410476572876 -en 0.071964666275442 1 0.278345739259351 -en 0.086558359329153 1 1.282885999320505 -ed 0.146214905642895 1 0 -en 0.146214905642895 0 4.089940389782871`. Here 46 haplotypes were sampled from population 0, which has the demographic history of the high-altitude population, and 44 haplotypes were sampled from population 1, which has the demographic history of the low-altitude population. A mutation rate of $\mu = 2.5 \times 10^{-8}$ (Fan *et al.* 2018) was used, along with a recombination rate of $r = 5.126 \times 10^{-9}$, which was computed as the genome-wide mean rate from the rheMac8 recombination map (Bcm-Hgsc 2020). A 500 kb region was simulated, thus giving a scaled mutation rate and scaled recombination parameters for discoal as `-t 809.425 -r 165.967`. 5,349 replicates of neutral sequence was simulated under this model, representing approximately the entire macaque genome minus 500 kb.

228 For non-neutral simulations, sweep scenarios were simulated with a positive
229 additive selection coefficient $s \in \{0.01, 0.02, 0.05\}$ (corresponding to `discoal` flag `-a`
230 selected from $\{323.77, 647.54, 1618.85\}$). Soft sweeps are simulated as a mutation that
231 arose neutral and turned beneficial at a particular establishment frequency $e \in$
232 $\{0.01, 0.02, 0.03, 0.04, 0.05, 0.10, 0.20\}$ (`discoal` flag `-f`). Hard sweeps are simulated from
233 a *de novo* mutation that was never neutral, i.e. $e = 0$. Finally, sweeps were conditioned
234 on having either reached a certain frequency in the population at the time of sampling,
235 $f \in \{0.7, 0.8, 0.9, 1.0\}$ (`discoal` flag `-c`), or that the adaptive mutation reached fixation
236 some number of generations prior to sampling, $g \in \{50, 100, 200\}$ (`discoal` flag `-ws`
237 selected from $\{7.7215 \times 10^{-4}, 1.5443 \times 10^{-3}, 3.0886 \times 10^{-3}\}$). For the sake of being
238 conservative in our estimation of power (Voight *et al.* 2006; Sabeti *et al.* 2007), it was
239 assumed the actual adaptive mutation remains unsampled (`discoal` flag `-h`) even
240 though whole genome sequencing data are being analyzed. For each combination of
241 parameter values, 500 replicates were simulated.

242 **Detecting Local Adaptation in Real Data**

243 From the phased data set, animals captured at high altitude (n = 23) and animals
244 captured at low altitude (n = 22) were subset (see Table S1). Using `selscan` v1.3.0
245 (Szpiech and Hernandez 2014) to compute raw XP-nSL scores across the genome,
246 scores were then normalized using the genome-wide empirical background. The low-
247 altitude population was used as the reference population, so positive XP-nSL scores
248 correspond to long homozygous haplotypes and a possible sweep in the high-altitude

population compared to the low-altitude population, and vice versa for negative XP-nSL scores.

In order to identify regions implicated as potentially adaptive, we search for clusters of extreme scores along a chromosome. Using `selscan`'s companion program `norm`, the genome is divided into non-overlapping 100kb regions and both the maximum XP-nSL score and the fraction of XP-nSL scores > 2 are computed. `norm` then creates 10 quantile bins for windows and identifies the top 1% of windows with the highest fraction of extreme scores. Each window is then annotated with the ensembl rheMac8 gene list, and a maximum XP-nSL score is assigned to a given gene based on the max-score in the 100 kb window with which it overlaps. If a gene overlaps more than one 100 kb window, it is assigned the top max-score from among the windows.

Results

Power Analysis of XP-nSL

First, we evaluate the performance of XP-nSL based on simulations. After computing XP-nSL for all sites in all simulations, scores were normalized by subtracting the mean and dividing by standard deviation of the neutral simulations, giving the neutral scores an approximately N(0,1) distribution (Fig. S1).

We consider the maximum score in a 100 kb interval as way to identify regions under positive selection similar to (Voight *et al.* 2006). To get the null distribution of max-scores, the maximum score is computed in the central 100 kb of all neutral simulations. The distribution of max-scores in neutral simulations had a median of 2.093 with 95% of the mass between 0.804 and 3.492, which is represented in Fig. 2 as a solid horizontal

271 black line (median) and two dashed horizontal black lines (95% interval). Next, the
272 maximum score is computed in the central 100 kb of all non-neutral simulations, and the
273 median and 95% intervals are plotted for each parameter combination. Fig. 2 shows
274 good separation between the neutral distribution of max-scores and the distribution of
275 max-scores for a range of non-neutral parameters, suggesting that our statistic can
276 distinguish between neutral and non-neutral scenarios. Note that soft sweeps that start
277 at 0.1 or 0.2 frequency see the least separation from the neutral distribution. To evaluate
278 the power of the max-score statistic, the 99th percentile of the max-score distribution is
279 computed in the neutral distribution ($neutral_{99} = 3.792$) and power is calculated as the
280 mass of each non-neutral max score distribution $> neutral_{99}$. This fixes the false
281 positive rate at 1%. Results are plotted in Fig. 3A, which shows good power to detect
282 both incomplete and completed sweeps. For soft sweeps that start at a frequency $\leq$
283 0.03, power is $> 75\%$ when the sweep is near or just past fixation; the ability to detect
284 soft sweeps falls off for sweeps that start > 0.03 frequency.

285 Next, we consider that due to linkage disequilibrium consecutive scores will be
286 correlated, and we should therefore expect clusters of extreme scores in true non-
287 neutral regions. We thus consider a window-based approach to identify selected regions
288 similar to (Voight *et al.* 2006). Taking the central 100 kb region of our simulations, the
289 fraction of XP-nSL scores > 2 is computed, representing approximately the highest 2%
290 of all neutral scores. Since each 100 kb window has different numbers of scores within
291 it, 10 quantile bins are created for windows with similar numbers of scores. Then the top
292 1% of windows in each bin of windows with the highest fraction of extreme scores are

293 taken as putatively selected regions. This approach also has the effect of ensuring the
294 false positive rate is at most 1%. The proportion of times the central 100 kb of each non-
295 neutral simulation is computed in order to get power. The results are plotted in Fig. 3B,
296 which shows improved power over the max-score approach across a wider range of
297 parameters. Indeed, using the window-based method, power to detect soft sweeps
298 improves substantially across the parameter space, with $> 75\%$ power to detect soft
299 sweeps at or near fixation that started at frequency ≤ 0.05.

300 **Identifying Evidence for Adaptation to High Altitude in Rhesus Macaques**

301 Next we analyze the pair of rhesus macaque populations using XP-nSL,
302 searching for evidence of local adaptation in the high-altitude population. Using the low-
303 altitude population as the reference population, normalized $XPnS_L > 0$ corresponds to
304 evidence for adaptation in the high-altitude population. Dividing the genome into 100 kb
305 windows, the maximum XP-nSL score of that region is assigned to each gene in it (see
306 Methods).

307 Using the per-gene max-scores, PANTHER (Mi *et al.* 2019) gene ontology
308 categories were tested for enrichment of high scores, where significance suggests an
309 enrichment of signals of positive selection among genes involved. Significant terms
310 related to regulation of ion transport and synaptic signaling (Table 2), each of which are
311 affected by hypoxic conditions (Karle *et al.* 2004; Corcoran and O'connor 2013).

312 From the genomic regions that were identified to contain a high proportion of
313 extreme positive scores (see Methods), 303 annotated genes were found. These
314 regions, their characteristics, and the genes contained therein are given in Table S2. A

315 PANTHER (Mi *et al.* 2019) gene ontology overrepresentation test indicates a 9.04-fold
316 enrichment of genes associated with monooxygenase activity ($FDR = 4.47 \times 10^{-2}$).
317 The monooxygenases in the selected regions include FMO2, FMO5, CYP2C8,
318 CYP2C9, CYP2C93, and ENSMMUT00000011129. These genes are important for the
319 metabolism of oxygen and the generation of reactive oxygen species (ROS) (Krueger
320 and Williams 2005). Under the physiological stress of a low-oxygen environment, ROS
321 levels increase and cause oxidative damage, and, in humans, long-term adaptation to
322 high altitudes includes adaptation to oxidative damage (Janocha *et al.* 2017). Indeed,
323 AOX1 is also identified in our top regions, mutations in which have been shown to affect
324 ROS levels in humans (Foti *et al.* 2017).
325 The genome-wide top ten 100 kb windows based on the percentage of extreme
326 XP-nSL scores are summarized in Table 3, and these windows overlap several genes,
327 including EGLN1. The EGLN1 locus is single strongest selection signal identified in the
328 entire genome (Fig. 4). This locus has the third highest cluster of extreme scores (Table
329 3), contains the highest XP-nSL score in the entire genome (chr1:207,698,003, $XPnS_L =$
330 6.54809), and contains six of the top ten genome wide XP-nSL scores (colored dark red
331 in Fig. 4). EGLN1 encodes Hypoxia-Inducible Factor Prolyl Hydroxylase 2 and acts as
332 an oxygen sensor by catalyzing the formation of HIF1α. The EGLN1 locus is a classic
333 target for adaptation to low-oxygen levels, having repeatedly been the target of
334 adaptation in numerous organisms living at high altitude around the world (Bigham *et al.*
335 2009; Bigham *et al.* 2010; Jeong *et al.* 2014; Graham and Mccracken 2019). In addition
336 to EGLN1, other genes related to lung function, oxygen use, and angiogenesis had

337 evidence for differential local adaptation between low- and high-altitude populations:
338 TRPM7, PAXIP1, RBPJ, and ENSMMUT00000040566 (Table S2). TRPM7 has been
339 found to be downregulated in a hypoxia-induced rat model and associated with
340 pulmonary hypertension (PAH) (Xing *et al.* 2019). PAXPIP1 may be involved in vascular
341 development and has also been associated with PAH (Jandl *et al.* 2019).
342 ENSMMUT00000040566 is a MAPK6 ortholog, which interacts with EGLN3 (Rodriguez
343 *et al.* 2016), and both it and RBPJ involved in angiogenesis (Ramasamy *et al.* 2014).
344 Due to the reduced oxygen levels at high altitudes, we expect genes involved in
345 metabolism and respiration may be under positive selection. Indeed, MDH1 encodes a
346 critical enzyme in the citrate cycle (Tanaka *et al.* 1996) and is found in the top ten
347 genome-wide regions (Table 3). A paralog of MDH1, MDH1B, has been previously
348 identified as a target of selection in humans living at high altitude (Yi *et al.* 2010).
349 ACADM and COX15 are also found in putatively adaptive regions (Table S2). Mutations
350 in and differing expression levels of ACADM are related to oxidative stress and
351 mitochondrial dysfunction in human disease (Xu *et al.* 2018). COX15 is involved in
352 oxidative phosphorylation (Alston *et al.* 2017), and cytochrome c oxidase (COX) genes
353 have previously been identified as under selection primates relative to other mammals
354 (Osada and Akashi 2012).
355 STXBP5L also appears in the top ten regions (Table 3) and is involved in
356 vesicular trafficking and neurotransmitter release (Kumar *et al.* 2015). As primate brains
357 use large amounts of oxygen and energy to function (Osada and Akashi 2012)
358 signatures of selection on neurological genes may be expected across populations living
359 at altitudes with differing oxygen levels. In addition to STXBP5L, several genes related

360 to neural development and synaptic formation (JAG2, TRPM7, DOCK7, NSG2, AUTS2)
361 were identified (Table S2). JAG2 is involved in the Notch signaling pathway that is
362 crucial for neuronal differentiation in the mammalian brain (Cardenas *et al.* 2018).
363 TRPM7, in addition to its association with PAH, plays a role in hypoxic neuronal cell
364 death (Aarts *et al.* 2003).

365 DNA damage, including double strand breaks and pyrimidine dimerization, can
366 manifest as a result of oxidative stress (Ye *et al.* 2016) or increased exposure to UV
367 radiation (Zhang *et al.* 2000; Greinert *et al.* 2012), both of which increase at high
368 altitudes. Two DNA damage repair genes are in our set of genes identified in positively
369 selected regions. The ring finger protein RNF138 is in our list of top ten genomic
370 windows (Table 3) and has been shown to promote DNA double-strand break repair
371 (Ismail *et al.* 2015). Furthermore, the DNA polymerase POLH also appears in a
372 putatively adaptive region (Table S2) and is known to be able to efficiently bypass
373 pyrimidine dimer lesions (Zhang *et al.* 2000).

374 Previous studies in rhesus macaques have shown there may be drastic
375 differences in diets between high-and low-altitude populations (Zhao 2018). Indeed,
376 genes related to lipid and fat metabolism (DOCK7, ST6GALNAC5, ANGPTL3, and
377 ACACA) were also found in putative adaptive regions. Across human populations, these
378 genes are all responsible for varying blood levels of fatty acids (Guo *et al.* 2016; Dewey
379 *et al.* 2017; Hebbar *et al.* 2018). Furthermore, ACACA has been shown to vary fatty acid
380 blood concentrations and be differentially expressed in highland versus lowland swine
381 populations (Shang *et al.* 2019).

382 ## Discussion

383 When populations split and migrate, they may adapt in different ways in response
384 to their local environments. Genetic adaptations that arise and sweep through the
385 population leave a characteristic genomic pattern of long haplotypes of low diversity and
386 high frequency. We develop a two-population haplotype-based statistic, XP-nSL, to
387 capture these patterns. With good power to detect hard and soft sweeps that occur in
388 one population but not another, XP-nSL can identify positively selected regions of the
389 genome likely the result of local adaptation. We apply this statistic to genomes sampled
390 from a pair of wild-born populations of Chinese rhesus macaques, inferred to have
391 diverged approximately 9500 generations ago, one of which lives at high altitude in far
392 western Sichuan province, and the other that lives close to sea level.

393 Life at high altitude presents extreme biological challenges, including low
394 atmospheric oxygen, increased UV exposure, harsh winters, and reduced nutrition
395 availability, which create strong selection pressure. Organisms that survive and persist
396 are likely to be carrying genetic mutations that confer an advantage for living in such
397 harsh environments. Common targets for adaptation to such an environment include
398 genes related to hypoxia, regulation of ROS, DNA damage repair, and metabolism.
399 Indeed, in the high-altitude macaque population, we identify a strong signal of positive
400 selection at the EGLN1 locus (Fig. 4), a classic target for adaptation to low-oxygen
401 environments, in addition to 302 other genes, many of which are related to the myriad
402 environmental selection pressures expected in high altitude environments. These results
403 suggest that, rather than a single adaptive mutation at a single locus, adaptation to this

404 extreme environment by rhesus macaques is polygenic and acts through multiple
405 biological systems.

406 Acknowledgments

407 The authors would like to thank members of the Stevison Lab for helpful discussions
408 and Lawrence Uricchio for helpful comments on early versions of the manuscript. This
409 work was supported by research start-up funds from the Department of Biological
410 Sciences at Auburn University (LSS).

411 Author Contributions

412 ZAS and LSS conceived of the study. ZAS performed all simulations and genomic
413 analyses and implemented novel statistics. ZAS, TEN, and NPB characterized gene
414 functions and ontologies with contributions LSS. ZAS wrote the manuscript with
415 contributions from NPB, TEN, and LSS. All authors read and approved of the
416 manuscript.

417 Data Accessibility

418 Macaque whole genome VCFs are available at http://dx.doi.org/10.5524/100484. Selection scan
419 data available at Dryad XXXX [Upon publication].

420

421

422

423

424

Tables

Table 1. Demographic parameters used for simulations. T values are given in number of generations before present. N values represent diploid effective population size.

Parameter	Value
T_1	4660
T_2	5605
T_3	9468
N_1	16188
N_2	3730
N_3	4506
N_4	20768
N_5	66210

Table 2. Gene ontology enrichment analysis results based on maximum XP-nSL scores per gene. Significant GO terms are enriched for high XP-nSL scores.

PANTHER GO-Slim Biological Process	Gene Ontology ID	p-value	FDR
anterograde trans-synaptic signaling	GO:0098916	4.43×10^{-5}	1.52×10^{-2}
chemical synaptic transmission	GO:0007268	4.43×10^{-5}	1.83×10^{-2}
regulation of transport	GO:0051049	3.60×10^{-5}	1.85×10^{-2}
trans-synaptic signaling	GO:0099537	6.32×10^{-5}	1.86×10^{-2}
synaptic signaling	GO:0099536	1.03×10^{-4}	2.66×10^{-2}
ion transport	GO:0006811	1.33×10^{-4}	2.75×10^{-2}
transmembrane transport	GO:0055085	1.57×10^{-4}	2.94×10^{-2}
regulation of ion transport	GO:0043269	1.30×10^{-4}	2.97×10^{-2}
metal ion transport	GO:0030001	2.21×10^{-4}	3.04×10^{-2}
regulation of localization	GO:0032879	2.13×10^{-4}	3.14×10^{-2}
regulation of ion transmembrane transport	GO:0034765	2.66×10^{-4}	3.23×10^{-2}
inorganic ion transmembrane transport	GO:0098660	3.01×10^{-4}	3.45×10^{-2}
sodium ion transport	GO:0006814	3.36×10^{-4}	3.65×10^{-2}
ion transmembrane transport	GO:0034220	4.50×10^{-4}	4.42×10^{-2}
regulation of transmembrane transport	GO:0034762	4.40×10^{-4}	4.54×10^{-2}
cation transport	GO:0006812	4.87×10^{-4}	4.57×10^{-2}

Table 3. The top ten 100 kb genomic regions as ranked by percentage of scores greater than 2. Concatenated region represents the genomic region merged with adjacent top 1% regions. Max score represents the max XP-nSL score in the concatenated region. Genes gives all genes overlapping the concatenated region.

Genomic Region	% > 2	Flanking Regions	Concatenated Region	Max Score	Genes
*chr2:42300001-42400000	81.62%	3	chr2:42200001-42600000	5.42043	STXBP5L
chr3:99100001-99200000	78.18%	3	chr3:99100001-99500000	5.54282	-
chr1:207600001-207700000	***77.39%***	***1***	***chr1:207600001-207800000***	***6.54809†***	***EGLN1, TSNAX***
chr10:69700001-69800000	76.60%	4	chr10:69700001-70200000	4.89388	PITPNB, ENSMMUT00000079195.1, TTC28
chr13:64300001-64400000	74.01%	1	chr13:64300001-64500000	4.17728	WDPCP, MDH1, ENSMMUT00000070474.1
chr7:96400001-96500000	73.86%	1	chr7:96400001-96600000	4.58886	FAM177A1, PPP2R3C, ENSMMUT00000039911.2, ENSMMUT00000027928.3
chr10:70400001-70500000	70.24%	0	chr10:70400001-70500000	3.69192	TTC28
*chr2:42400001-42500000	65.86%	3	chr2:42200001-42600000	5.42043	STXBP5L
chr7:120400001-120500000	62.54%	0	chr7:120400001-120500000	5.86114	KIAA0586, ENSMMUT00000059161.1
chr18:20600001-20700000	62.42%	0	chr18:20600001-20700000	5.37386	RNF138

***These regions are contained in the same concatenated region.**
†Top genome-wide score. This region contains 6 of the top 10 genome-wide scores.

Table S1. Population classification by individual ID.
See TableS1.xlsx

Table S2. List of genomic windows with the top 1% highest fraction of extreme XP-nSL scores.
See TableS2.xlsx

Figures

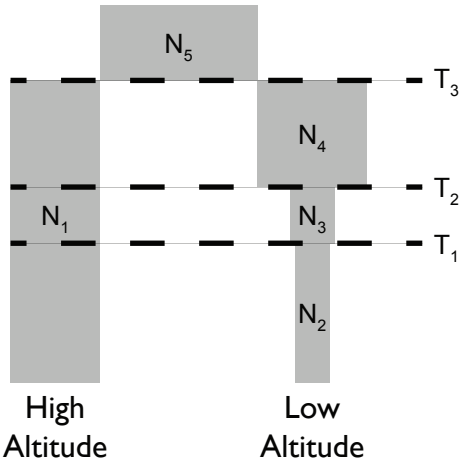

Figure 1. A representation of the demographic history for our high- and low-altitude populations as inferred by Liu *et al.* (2018).

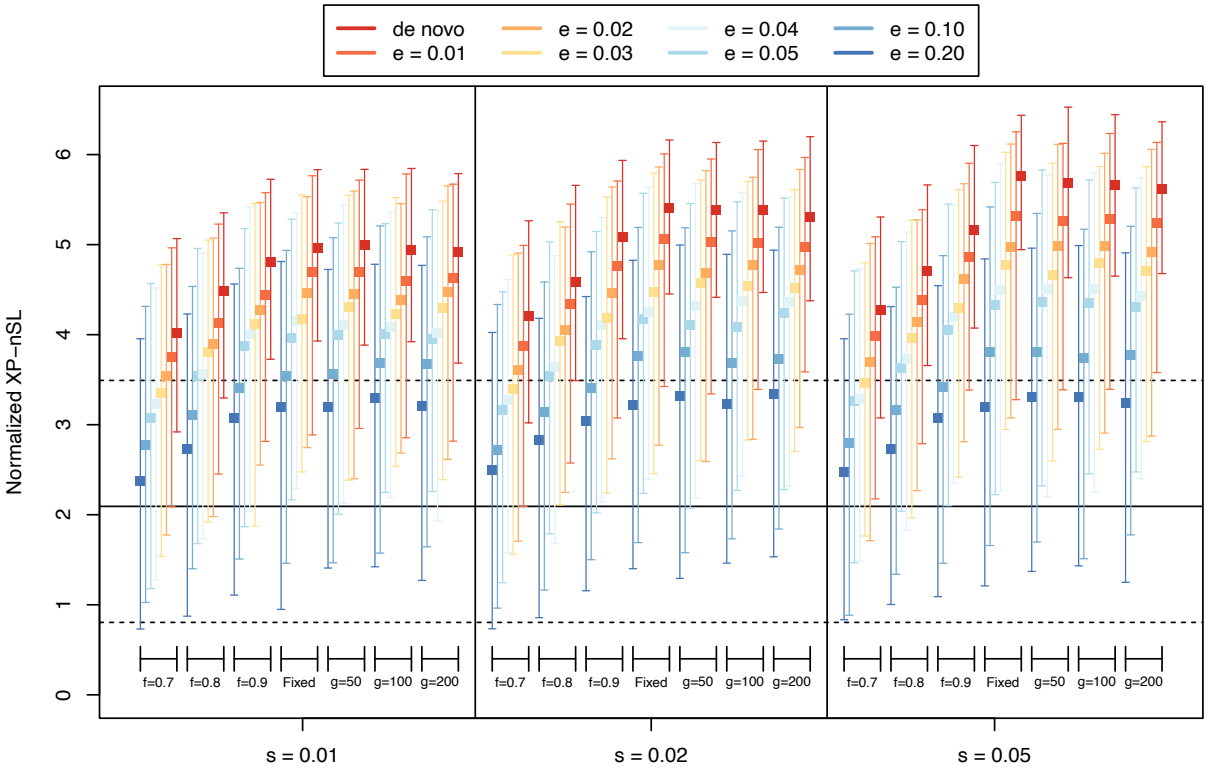

Figure 2. The distribution of maximum XP-nSL scores from simulations across various parameters, represented by medians and intervals containing 95% of the mass of the distribution. Neutral simulations are represented by the black solid horizonal line (median) and the black dashed horizonal line (95% interval). Non-neutral simulations represented by a colored box (median) and error bars (95% interval). The parameters are e (frequency at which selection begins, e > 0 indicates soft sweep), f (frequency of selected mutation at sampling), g (number of generations since fixation), and s (selection coefficient).

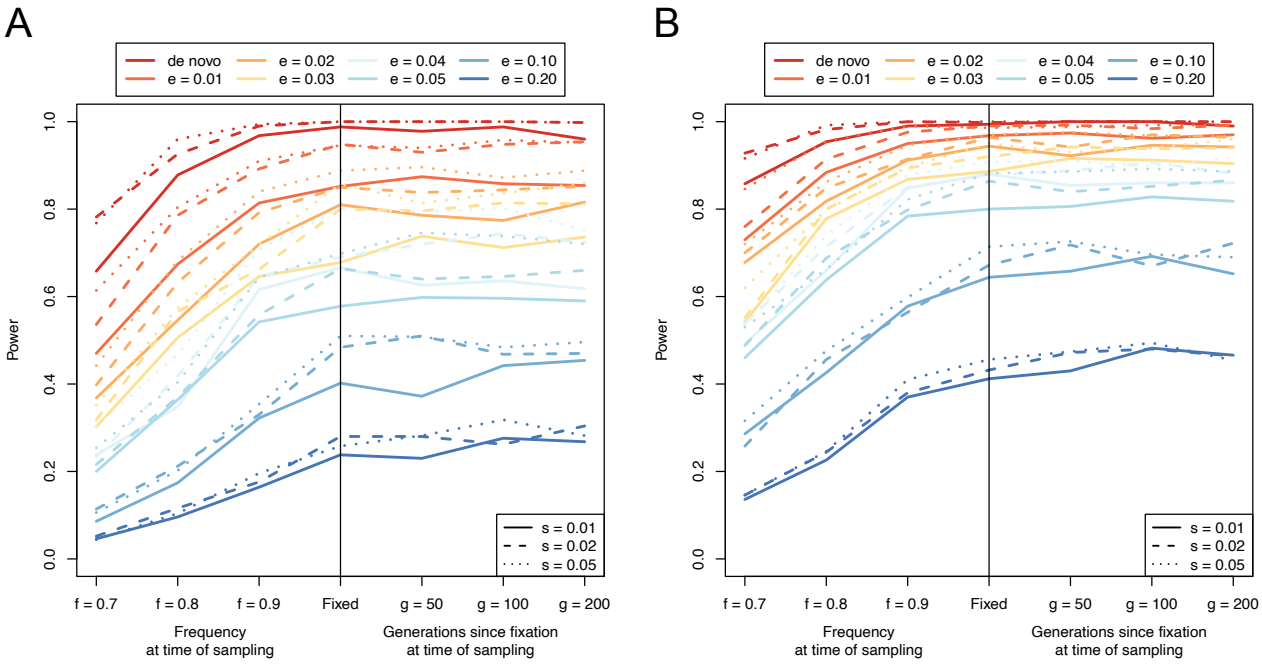

Figure 3. Power curves for (A) the max-score approach and (B) the window-based approach to identifying sweeps. The parameters are e (frequency at which selection begins, e > 0 indicates soft sweep), f (frequency of selected mutation at sampling), g (number of generations since fixation), and s (selection coefficient).

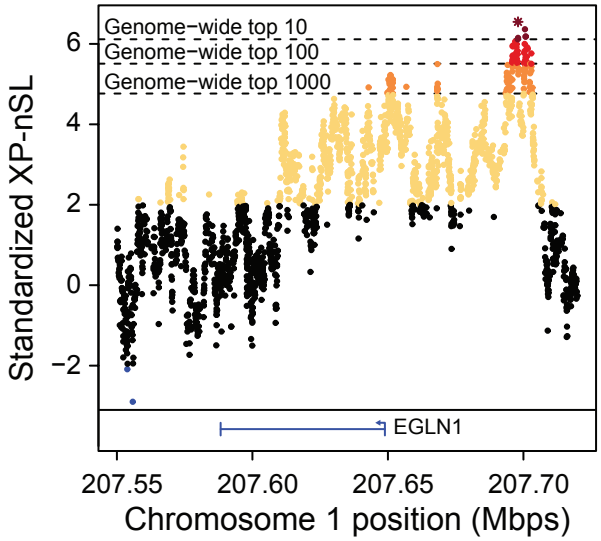

Figure 4. XP-nSL scores in the vicinity of the EGLN1 locus. This locus contains the genome wide top score (star) and six of the top ten genome wide scores (dark red).

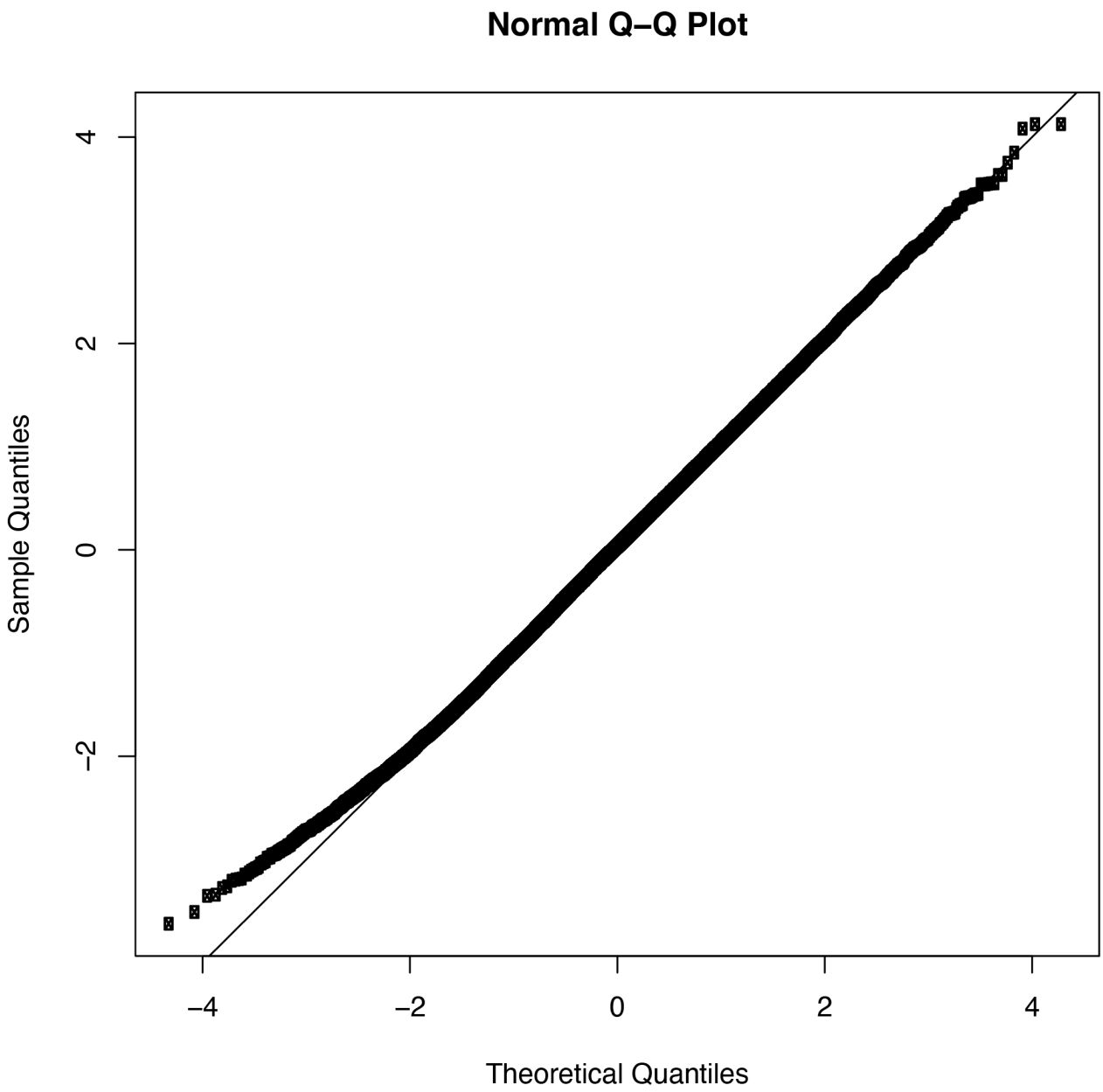

Figure S1. A normal quantile-quantile plot of neutral XP-nSL scores showing generally good adherence to a standard normal distribution. Due to autocorrelation along the genome, only every 1000th score is plotted.

522 ## References

523

524 Aarts, M., Iihara, K., Wei, W.L., Xiong, Z.G., Arundine, M., Cerwinski, W. *et al.* (2003). A key role
525 for TRPM7 channels in anoxic neuronal death. *Cell* 115**:** 863-877.
526 Ahmad, K.S., Hameed, M., Fatima, S., Ashraf, M., Ahmad, F., Naseer, M. *et al.* (2016). Morpho-
527 anatomical and physiological adaptations to high altitude in some Aveneae grasses from
528 Neelum Valley, Western Himalayan Kashmir. *Acta Physiologiae Plantarum* 38**:** 93.
529 Alston, C.L., Rocha, M.C., Lax, N.Z., Turnbull, D.M. & Taylor, R.W. (2017). The genetics and
530 pathology of mitochondrial disease. *J Pathol* 241**:** 236-250.
531 BCM-HGSC. (2020) Baylor College of Medicine Human Genome Sequencing Center Rhemac8
532 Recombination Map. Available at:
533 [ftp://ftp.hgsc.bcm.edu/ucscHub/rhesusSNVs/rheMac8/all.rate.bw].
534 Bigham, A., Bauchet, M., Pinto, D., Mao, X., Akey, J.M., Mei, R. *et al.* (2010). Identifying
535 signatures of natural selection in Tibetan and Andean populations using dense genome
536 scan data. *PLoS Genet* 6**:** e1001116.
537 Bigham, A.W., Mao, X., Mei, R., Brutsaert, T., Wilson, M.J., Julian, C.G. *et al.* (2009). Identifying
538 positive selection candidate loci for high-altitude adaptation in Andean populations.
539 *Hum Genomics* 4**:** 79-90.
540 Cai, Q., Qian, X., Lang, Y., Luo, Y., Xu, J., Pan, S. *et al.* (2013). Genome sequence of ground tit
541 Pseudopodoces humilis and its adaptation to high altitude. *Genome Biol* 14**:** R29.
542 Campbell, K.L., Storz, J.F., Signore, A.V., Moriyama, H., Catania, K.C., Payson, A.P. *et al.* (2010).
543 Molecular basis of a novel adaptation to hypoxic-hypercapnia in a strictly fossorial mole.
544 *BMC Evol Biol* 10**:** 214.
545 Cardenas, A., Villalba, A., de Juan Romero, C., Pico, E., Kyrousi, C., Tzika, A.C. *et al.* (2018).
546 Evolution of Cortical Neurogenesis in Amniotes Controlled by Robo Signaling Levels. *Cell*
547 174**:** 590-606 e521.
548 Cheviron, Z.A., Bachman, G.C., Connaty, A.D., McClelland, G.B. & Storz, J.F. (2012). Regulatory
549 changes contribute to the adaptive enhancement of thermogenic capacity in high-
550 altitude deer mice. *Proceedings of the National Academy of Sciences* 109**:** 8635-8640.
551 Corcoran, A. & O'Connor, J.J. (2013). Hypoxia-inducible factor signalling mechanisms in the
552 central nervous system. *Acta Physiol (Oxf)* 208**:** 298-310.
553 Delaneau, O., Zagury, J.F., Robinson, M.R., Marchini, J.L. & Dermitzakis, E.T. (2019). Accurate,
554 scalable and integrative haplotype estimation. *Nat Commun* 10**:** 5436.
555 Dewey, F.E., Gusarova, V., Dunbar, R.L., O'Dushlaine, C., Schurmann, C., Gottesman, O. *et al.*
556 (2017). Genetic and Pharmacologic Inactivation of ANGPTL3 and Cardiovascular Disease.
557 *N Engl J Med* 377**:** 211-221.
558 Fan, Z., Zhou, A., Osada, N., Yu, J., Jiang, J., Li, P. *et al.* (2018). Ancient hybridization and
559 admixture in macaques (genus Macaca) inferred from whole genome sequences. *Mol*
560 *Phylogenet Evol* 127**:** 376-386.
561 Ferrer-Admetlla, A., Liang, M., Korneliussen, T. & Nielsen, R. (2014). On detecting incomplete
562 soft or hard selective sweeps using haplotype structure. *Mol Biol Evol* 31**:** 1275-1291.

563 Fooden, J. (2000). *Systematic review of the rhesus macaque, Macaca mulatta (Zimmermann,*
564 *1780)*Field Museum of Natural History, Chicago, Ill. :.
565 Foti, A., Dorendorf, F. & Leimkuhler, S. (2017). A single nucleotide polymorphism causes
566 enhanced radical oxygen species production by human aldehyde oxidase. *PLoS One* 12**:**
567 e0182061.
568 Garud, N.R., Messer, P.W., Buzbas, E.O. & Petrov, D.A. (2015). Recent selective sweeps in North
569 American Drosophila melanogaster show signatures of soft sweeps. *PLoS Genet* 11**:**
570 e1005004.
571 Ge, R.-L., Cai, Q., Shen, Y.-Y., San, A., Ma, L., Zhang, Y. *et al.* (2013). Draft genome sequence of
572 the Tibetan antelope. *Nature Communications* 4**:** 1858.
573 Gonzalo-Turpin, H. & Hazard, L. (2009). Local adaptation occurs along altitudinal gradient
574 despite the existence of gene flow in the alpine plant species Festuca eskia. *Journal of*
575 *Ecology* 97**:** 742-751.
576 Graham, A.M. & McCracken, K.G. (2019). Convergent evolution on the hypoxia-inducible factor
577 (HIF) pathway genes EGLN1 and EPAS1 in high-altitude ducks. *Heredity (Edinb)* 122**:** 819-
578 832.
579 Greinert, R., Volkmer, B., Henning, S., Breitbart, E.W., Greulich, K.O., Cardoso, M.C. *et al.* (2012).
580 UVA-induced DNA double-strand breaks result from the repair of clustered oxidative
581 DNA damages. *Nucleic Acids Res* 40**:** 10263-10273.
582 Guo, T., Yin, R.X., Huang, F., Yao, L.M., Lin, W.X. & Pan, S.L. (2016). Association between the
583 DOCK7, PCSK9 and GALNT2 Gene Polymorphisms and Serum Lipid levels. *Sci Rep* 6**:**
584 19079.
585 Guo, X., Hu, Q., Hao, G., Wang, X., Zhang, D., Ma, T. *et al.* (2018). The genomes of two Eutrema
586 species provide insight into plant adaptation to high altitudes. *DNA Research* 25**:** 307-
587 315.
588 Hebbar, P., Nizam, R., Melhem, M., Alkayal, F., Elkum, N., John, S.E. *et al.* (2018). Genome-wide
589 association study identifies novel recessive genetic variants for high TGs in an Arab
590 population. *J Lipid Res* 59**:** 1951-1966.
591 Hendrickson, S.L. (2013). A genome wide study of genetic adaptation to high altitude in feral
592 Andean Horses of the paramo. *BMC Evol Biol* 13**:** 273.
593 Hermisson, J. & Pennings, P.S. (2005). Soft sweeps: molecular population genetics of adaptation
594 from standing genetic variation. *Genetics* 169**:** 2335-2352.
595 Hernandez, R.D., Hubisz, M.J., Wheeler, D.A., Smith, D.G., Ferguson, B., Rogers, J. *et al.* (2007).
596 Demographic histories and patterns of linkage disequilibrium in Chinese and Indian
597 rhesus macaques. *Science* 316**:** 240-243.
598 Huerta-Sanchez, E., Degiorgio, M., Pagani, L., Tarekegn, A., Ekong, R., Antao, T. *et al.* (2013).
599 Genetic signatures reveal high-altitude adaptation in a set of ethiopian populations. *Mol*
600 *Biol Evol* 30**:** 1877-1888.
601 Huerta-Sanchez, E., Jin, X., Asan, Bianba, Z., Peter, B.M., Vinckenbosch, N. *et al.* (2014). Altitude
602 adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature* 512**:** 194-
603 197.

604 Ismail, I.H., Gagne, J.P., Genois, M.M., Strickfaden, H., McDonald, D., Xu, Z. *et al.* (2015). The
605 RNF138 E3 ligase displaces Ku to promote DNA end resection and regulate DNA repair
606 pathway choice. *Nat Cell Biol* 17**:** 1446-1457.
607 Jandl, K., Thekkekara Puthenparampil, H., Marsh, L.M., Hoffmann, J., Wilhelm, J., Veith, C. *et al.*
608 (2019). Long non-coding RNAs influence the transcriptome in pulmonary arterial
609 hypertension: the role of PAXIP1-AS1. *J Pathol* 247**:** 357-370.
610 Janocha, A.J., Comhair, S.A.A., Basnyat, B., Neupane, M., Gebremedhin, A., Khan, A. *et al.*
611 (2017). Antioxidant defense and oxidative damage vary widely among high-altitude
612 residents. *Am J Hum Biol* 29.
613 Jeong, C., Alkorta-Aranburu, G., Basnyat, B., Neupane, M., Witonsky, D.B., Pritchard, J.K. *et al.*
614 (2014). Admixture facilitates genetic adaptations to high altitude in Tibet. *Nat Commun*
615 5**:** 3281.
616 Karle, C., Gehrig, T., Wodopia, R., Hoschele, S., Kreye, V.A., Katus, H.A. *et al.* (2004). Hypoxia-
617 induced inhibition of whole cell membrane currents and ion transport of A549 cells. *Am J*
618 *Physiol Lung Cell Mol Physiol* 286**:** L1154-1160.
619 Kern, A.D. & Schrider, D.R. (2016). Discoal: flexible coalescent simulations with selection.
620 *Bioinformatics* 32**:** 3839-3841.
621 Kim, E. & Donohue, K. (2013). Local adaptation and plasticity of Erysimum capitatum to altitude:
622 its implications for responses to climate change. *Journal of Ecology* 101**:** 796-805.
623 Kim, Y. & Nielsen, R. (2004). Linkage Disequilibrium as a Signature of Selective Sweeps. *Genetics*
624 167**:** 1513-1524.
625 Krueger, S.K. & Williams, D.E. (2005). Mammalian flavin-containing monooxygenases:
626 structure/function, genetic polymorphisms and role in drug metabolism. *Pharmacol Ther*
627 106**:** 357-387.
628 Kumar, R., Corbett, M.A., Smith, N.J., Jolly, L.A., Tan, C., Keating, D.J. *et al.* (2015). Homozygous
629 mutation of STXBP5L explains an autosomal recessive infantile-onset neurodegenerative
630 disorder. *Hum Mol Genet* 24**:** 2000-2010.
631 Li, M., Tian, S., Jin, L., Zhou, G., Li, Y., Zhang, Y. *et al.* (2013). Genomic analyses identify distinct
632 patterns of selection in domesticated pigs and Tibetan wild boars. *Nat Genet* 45**:** 1431-
633 1438.
634 Li, Y., Wu, D.D., Boyko, A.R., Wang, G.D., Wu, S.F., Irwin, D.M. *et al.* (2014). Population variation
635 revealed high-altitude adaptation of Tibetan mastiffs. *Mol Biol Evol* 31**:** 1200-1205.
636 Liu, J.-Q., Duan, Y.-W., Hao, G., Ge, X.-J. & Sun, H. (2014). Evolutionary history and underlying
637 adaptation of alpine plants on the Qinghai–Tibet Plateau. *Journal of Systematics and*
638 *Evolution* 52**:** 241-249.
639 Liu, Z., Tan, X., Orozco-terWengel, P., Zhou, X., Zhang, L., Tian, S. *et al.* (2018). Population
640 genomics of wild Chinese rhesus macaques reveals a dynamic demographic history and
641 local adaptation, with implications for biomedical research. *Gigascience* 7.
642 Madrid, J.E., Mandalaywala, T.M., Coyne, S.P., Ahloy-Dallaire, J., Garner, J.P., Barr, C.S. *et al.*
643 (2018). Adaptive developmental plasticity in rhesus macaques: the serotonin transporter
644 gene interacts with maternal care to affect juvenile social behaviour. *Proc Biol Sci* 285.

645 Mi, H., Muruganujan, A., Ebert, D., Huang, X. & Thomas, P.D. (2019). PANTHER version 14: more
646 genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools.
647 *Nucleic Acids Res* 47**:** D419-D426.
648 Munne-Bosch, S., Cotado, A., Morales, M., Fleta-Soriano, E., Villellas, J. & Garcia, M.B. (2016).
649 Adaptation of the Long-Lived Monocarpic Perennial Saxifraga longifolia to High Altitude.
650 *Plant Physiol* 172**:** 765-775.
651 O'Reilly, P.F., Birney, E. & Balding, D.J. (2008). Confounding between recombination and
652 selection, and the Ped/Pop method for detecting selection. *Genome Res* 18**:** 1304-1313.
653 Osada, N. & Akashi, H. (2012). Mitochondrial-nuclear interactions and accelerated
654 compensatory evolution: evidence from the primate cytochrome C oxidase complex. *Mol*
655 *Biol Evol* 29**:** 337-346.
656 Peng, Y., Yang, Z., Zhang, H., Cui, C., Qi, X., Luo, X. *et al.* (2011). Genetic variations in Tibetan
657 populations and high-altitude adaptation at the Himalayas. *Mol Biol Evol* 28**:** 1075-1081.
658 Pennings, P.S. & Hermisson, J. (2006). Soft sweeps II--molecular population genetics of
659 adaptation from recurrent mutation or migration. *Mol Biol Evol* 23**:** 1076-1084.
660 Przeworski, M. (2002). The Signature of Positive Selection at Randomly Chosen Loci. *Genetics*
661 160**:** 1179-1189.
662 Qiu, Q., Zhang, G., Ma, T., Qian, W., Wang, J., Ye, Z. *et al.* (2012). The yak genome and
663 adaptation to life at high altitude. *Nat Genet* 44**:** 946-949.
664 Qu, Y., Zhao, H., Han, N., Zhou, G., Song, G., Gao, B. *et al.* (2013). Ground tit genome reveals
665 avian adaptation to living at high altitudes in the Tibetan plateau. *Nat Commun* 4**:** 2071.
666 Ramasamy, S.K., Kusumbe, A.P., Wang, L. & Adams, R.H. (2014). Endothelial Notch activity
667 promotes angiogenesis and osteogenesis in bone. *Nature* 507**:** 376-380.
668 Richard, A.F., Goldstein, S.J. & Dewar, R.E. (1989). Weed macaques: The evolutionary
669 implications of macaque feeding ecology. *International Journal of Primatology* 10**:** 569.
670 Rodriguez, J., Pilkington, R., Garcia Munoz, A., Nguyen, L.K., Rauch, N., Kennedy, S. *et al.* (2016).
671 Substrate-Trapped Interactors of PHD3 and FIH Cluster in Distinct Signaling Pathways.
672 *Cell Rep* 14**:** 2745-2760.
673 Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z., Richter, D.J., Schaffner, S.F. *et al.* (2002).
674 Detecting recent positive selection in the human genome from haplotype structure.
675 *Nature* 419**:** 832-837.
676 Sabeti, P.C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C. *et al.* (2007). Genome-
677 wide detection and characterization of positive selection in human populations. *Nature*
678 449**:** 913-918.
679 Schweizer, R.M., Velotta, J.P., Ivy, C.M., Jones, M.R., Muir, S.M., Bradburd, G.S. *et al.* (2019).
680 Physiological and genomic evidence that selection on the transcription factor Epas1 has
681 altered cardiovascular function in high-altitude deer mice. *PLoS Genet* 15**:** e1008420.
682 Shang, P., Li, W., Liu, G., Zhang, J., Li, M., Wu, L. *et al.* (2019). Identification of lncRNAs and
683 Genes Responsible for Fatness and Fatty Acid Composition Traits between the Tibetan
684 and Yorkshire Pigs. *Int J Genomics* 2019**:** 5070975.
685 Stewart, C.B. & Disotell, T.R. (1998). Primate evolution - in and out of Africa. *Curr Biol* 8**:** R582-
686 588.

687 Storz, J.F., Cheviron, Z.A., McClelland, G.B. & Scott, G.R. (2019). Evolution of physiological
688 performance capacities and environmental adaptation: insights from high-elevation deer
689 mice (Peromyscus maniculatus). *J Mammal* 100**:** 910-922.
690 Storz, J.F., Sabatino, S.J., Hoffmann, F.G., Gering, E.J., Moriyama, H., Ferrand, N. *et al.* (2007).
691 The molecular basis of high-altitude adaptation in deer mice. *PLoS Genet* 3**:** e45.
692 Szpiech, Z.A. & Hernandez, R.D. (2014). selscan: an efficient multithreaded program to perform
693 EHH-based scans for positive selection. *Mol Biol Evol* 31**:** 2824-2827.
694 Tanaka, T., Inazawa, J. & Nakamura, Y. (1996). Molecular cloning and mapping of a human cDNA
695 for cytosolic malate dehydrogenase (MDH1). *Genomics* 32**:** 128-130.
696 Velotta, J.P., Robertson, C.E., Schweizer, R.M., McClelland, G.B. & Cheviron, Z.A. (2020).
697 Adaptive shifts in gene regulation underlie a developmental delay in thermogenesis in
698 high-altitude deer mice. *Molecular Biology and Evolution*.
699 Voight, B.F., Kudaravalli, S., Wen, X. & Pritchard, J.K. (2006). A map of recent positive selection
700 in the human genome. *Plos Biol* 4**:** e72.
701 Wang, G.D., Fan, R.X., Zhai, W., Liu, F., Wang, L., Zhong, L. *et al.* (2014). Genetic convergence in
702 the adaptation of dogs and humans to the high-altitude environment of the tibetan
703 plateau. *Genome Biol Evol* 6**:** 2122-2128.
704 Wang, M.-S., Li, Y., Peng, M.-S., Zhong, L., Wang, Z.-J., Li, Q.-Y. *et al.* (2015). Genomic Analyses
705 Reveal Potential Independent Adaptation to High Altitude in Tibetan Chickens. *Molecular*
706 *Biology and Evolution* 32**:** 1880-1889.
707 Wang, M.-S., Wang, S., Li, Y., Jhala, Y., Thakur, M., Otecko, N.O. *et al.* (2020). Ancient
708 hybridization with an unknown population facilitated high altitude adaptation of canids.
709 *Molecular Biology and Evolution*.
710 Xing, J., Wang, M., Hong, J., Gao, Y., Liu, Y., Gu, H. *et al.* (2019). TRPM7 channel inhibition
711 exacerbates pulmonary arterial hypertension through MEK/ERK pathway. *Aging (Albany*
712 *NY)* 11**:** 4050-4065.
713 Xu, S., Li, S., Yang, Y., Tan, J., Lou, H., Jin, W. *et al.* (2010). A Genome-Wide Search for Signals of
714 High-Altitude Adaptation in Tibetans. *Molecular Biology and Evolution* 28**:** 1003-1011.
715 Xu, Z., Jin, X., Cai, W., Zhou, M., Shao, P., Yang, Z. *et al.* (2018). Proteomics Analysis Reveals
716 Abnormal Electron Transport and Excessive Oxidative Stress Cause Mitochondrial
717 Dysfunction in Placental Tissues of Early-Onset Preeclampsia. *Proteomics Clin Appl* 12**:**
718 e1700165.
719 Yang, X., Wang, Y., Zhang, Y., Lee, W.H. & Zhang, Y. (2016). Rich diversity and potency of skin
720 antioxidant peptides revealed a novel molecular basis for high-altitude adaptation of
721 amphibians. *Sci Rep* 6**:** 19866.
722 Ye, B., Hou, N., Xiao, L., Xu, Y., Xu, H. & Li, F. (2016). Dynamic monitoring of oxidative DNA
723 double-strand break and repair in cardiomyocytes. *Cardiovasc Pathol* 25**:** 93-100.
724 Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z.X., Pool, J.E. *et al.* (2010). Sequencing of 50
725 human exomes reveals adaptation to high altitude. *Science* 329**:** 75-78.
726 Zhang, Y., Yuan, F., Wu, X., Rechkoblit, O., Taylor, J.S., Geacintov, N.E. *et al.* (2000). Error-prone
727 lesion bypass by human DNA polymerase eta. *Nucleic Acids Res* 28**:** 4717-4724.
728