A Genome Scan for Adaptation to High Altitude in Wild Rhesus Macaques

Zachary A. Szpiech, Taylor E. Novak*, Nicholas P. Bailey*, and Laurie S. Stevison

Department of Biological Sciences, Auburn University, Auburn, AL, USA

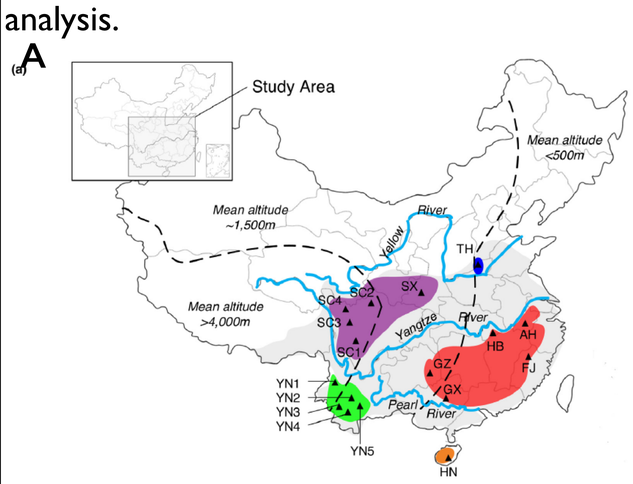
Abstract

When natural populations split and migrate to different environments, they may experience different selection pressures that can lead to localized 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 analyze the whole genomes of 23 wild rhesus macaques from a population living at high altitude (>4000m above sea level) alongside 22 wild rhesus macaques from a population living at lower altitude (<500m above sea level), which split approximately 100 kya. We extend the nSL statistic and develop XP-nSL, a haplotype-based genomic scan for local adaptation. We compare haplotype patterns between the high and low altitude populations. We find that a set of 197 genes related to hypoxia response and oxygen homeostasis is enriched for signals of local adaption in the high-altitude population. We also find a strong signal at the start of EGLN1 (a classic target for convergent adaptation to high altitude) suggesting a possibly regulatory adaptation. Other signals were found overlapping genes associated with neuronal death during hypoxia, total lung capacity in humans, and a critical enzyme in the citric acid cycle.

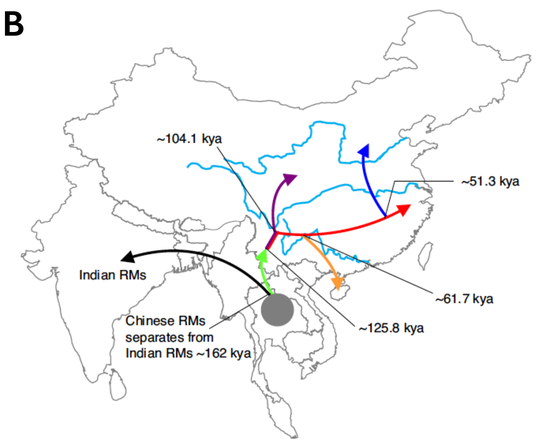
Samples

We take a subset of whole genome sequences of wild-caught Chinese rhesus macaques (*M. m. mulatta*) from among twenty-three sampling locations. High-altitude animals were caught in sampling locations SC1, SC2, SC3, and SC4 (Figure A), and the twenty-two low-altitude animals were caught in sampling locations AH, HB, FJ, and GX (Figure A). These two populations of macaques have an inferred population divergence of ~104kya (Figure B). After filtering rare alleles (any site with MAF < 0.05 in one or both populations), we are left with 5,149,637 autosomal SNPs for analysis.

A



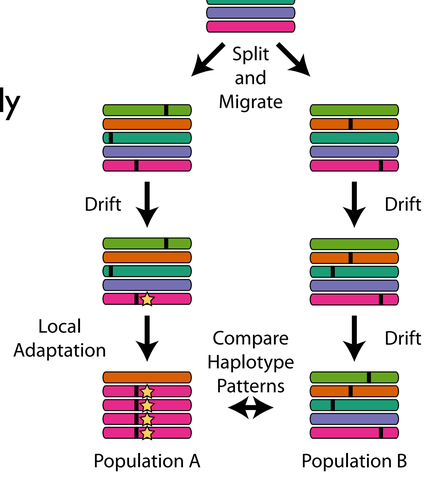
B

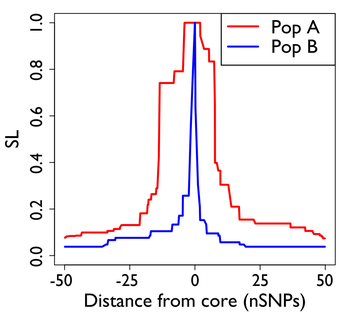


Local Adaptation

In the event of an adaptive mutation sweeping on one or more haplotype backgrounds in one population, a characteristic pattern of long haplotypes and high homozygosity can be found in the vicinity of the locus. This pattern is summarized in each population separately by computing haplotype homozygosity in successively wider windows around the locus and integrating the resulting curve.

We adapt the nSL statistic as a two-population statistic, implemented as XP-nSL in selscan v1.3.0. XP-nSL is then computed as the normalized log-ratio of the integrated SL curves in each population.



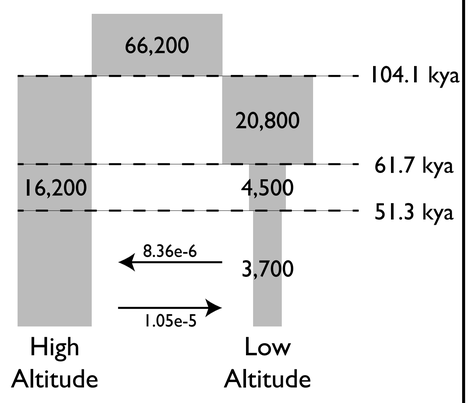


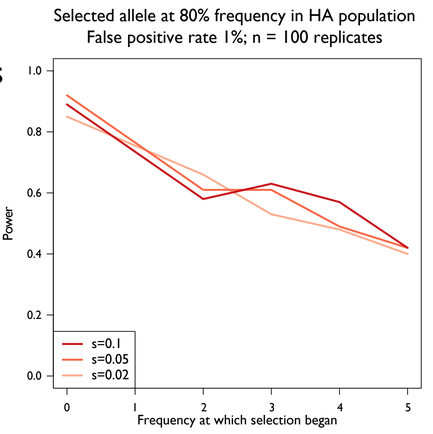
$$XPnSL = \frac{ln\left(\frac{\sum_n SL_A}{\sum_n SL_B}\right) - E\left[ln\left(\frac{\sum_n SL_A}{\sum_n SL_B}\right)\right]}{SD\left[ln\left(\frac{\sum_n SL_A}{\sum_n SL_B}\right)\right]}$$

In this formulation positive values correspond to a possible selective sweep in Population A, and negative values correspond to a possible selective sweep in Population B. We then further search for regions that are enriched for large numbers of consecutive positive or negative scores.

Simulations

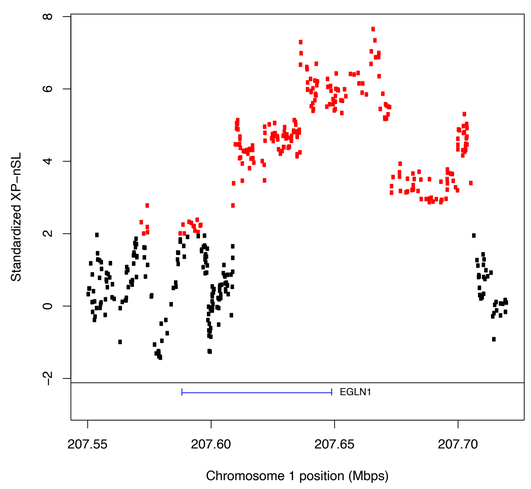
We simulate hard and soft sweep scenarios in a two population divergence model with migration. Hard sweeps were simulated as a de novo mutation with a selection coefficient set to s = 0.02, 0.05, or 0.1, conditional on establishment in the population. These sweeps are characterized by a single haplotype rising quickly to high frequency. Soft sweeps are simulated as selection on standing variation, where a random (previously neutral) variant at a given allele frequency (2%, 3%, 4%, or 5%) is chosen and reassigned a selection cofficient s = 0.02, 0.05, or 0.1. We also allow for migration between the two populations, and, importantly, we enforce that all mutations in the Low Altitude population are neutral regardless of their effect in the High Altitude population. We find that the method has good power to detect local adaptation across a range of parameters and even in the presence of gene flow.

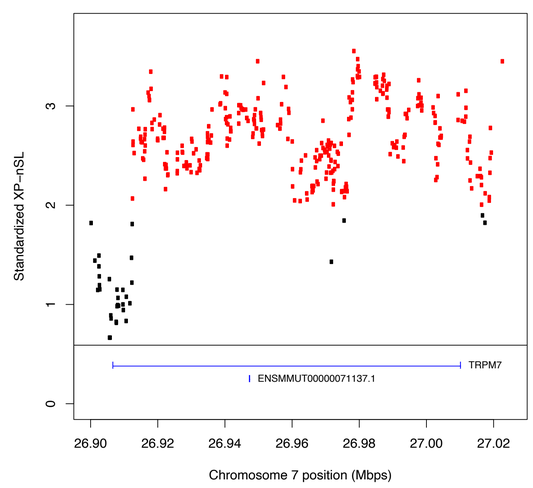




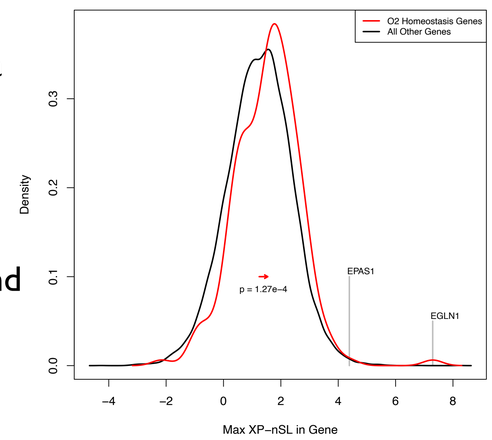
Results

We identify 476 genes putatively under selection in the high-altitude population. We find that EGLN1, a classic target for high-altitude adaptation, is implicated. We also find a strong signal in the vicinity of the gene TRPM7, which regulates calcium homeostasis and plays an important role in hypoxic brain injury and neuronal cell death.





Next we consider a set of 197 genes *a priori* known to be related to hypoxia response and oxygen homeostasis, and compare the distrubution of max XP-nSL scores in this set to the distribution of max scores in the rest of the genes in the genome. We find a statistically significant difference in means, with O2 homeostasis genes having higher scores on average p = 1.27E-4.



Discussion

Our preliminary results suggest several possible adaptations in rhesus macaques to cope with a low-oxygen environment. In addition to finding a classic target of high-altitude adaptation (EGLN1), a strong signal in the vicinity of TRPM7 is intriguing. The protein encoded by this gene, when suppressed *in vivo*, has been shown to provide neuroprotection against brain injury during hypoxia.