Genome **Biology**

RESEARCH HIGHLIGHT

Living the high life: high-altitude adaptation

Laura B Scheinfeldt[1] and Sarah A Tishkoff[1,2]*

Abstract

Genome-wide scans demonstrate that genetic variants associated with high-altitude adaptation in Tibetans and Andeans arose independently as a result of convergent adaptation.

There is widespread interest in the identification of human genes subject to positive selection, in part because they may elucidate the biological basis of human adaptation to novel environments and, therefore, may lead to the identification of genes and variants that play a role in disease susceptibility. To this end, a number of computational approaches have been developed to perform scans for positive selection. Unlike demographic processes such as migration, population expansion and bottlenecks, which affect whole-genome patterns of variation, positive selection shapes variation in a locus-specific manner.

Furthermore, populations in divergent environments with distinct selective pressures may be subject to local adaptation. Genetic variants that are targets of positive selection in locally adapted populations are expected to show higher levels of population differentiation (that is, differences between populations) and, in some cases, extended regions of allelic association or linkage disequilibrium. Thus, genome-wide scans for selection often identify outliers in the empirical distribution of summary statistics that characterize population differentiation, linkage disequilibrium or some combination of the two, and these outliers are enriched for loci that have been subject to positive selection (reviewed in [1]).

Physiological adaptations to high-altitude

One of the classic examples of adaptation to a novel environment is adaptation to high-altitude. At high-altitude, differences in barometric pressure result in insufficient oxygen in the air, thereby causing hypoxia

(that is, reduced oxygen levels in the blood). People at high-altitude for short periods of time, who are not adapted to that environment, are at increased risk for acute altitude sickness involving pulmonary and cerebral edema, and after extended periods of time at high-altitude, for chronic altitude sickness involving pulmonary hypertension and related complications. Moreover, pregnant women living in high-altitude environments are at increased risk for intrauterine growth restriction and pre-eclampsia, which can lead to more serious complications, including perinatal death [2]. Therefore, individuals possessing physiological traits that offset the effects of hypoxia have a higher fitness in high-altitude environments. For example, Tibetan women with high oxygen saturation of hemoglobin have more than twice the number of surviving children as those with low oxygen saturation of hemoglobin [3], indicating a very strong selective force on the physiological response to high altitude.

Studies of hypoxia-related physiological traits among high-altitude populations suggest that adaptive phenotypes have arisen independently in these populations. Populations in Asia, Africa and America each have sets of physiological traits that are distinct from each other. For example, hemoglobin concentrations are elevated in high-altitude populations living in the Andes relative to Tibetan and Ethiopian high-altitude populations, and oxygen saturation of arterial hemoglobin is reduced in high-altitude Tibetan and Andean populations relative to high-altitude Ethiopian populations [4]. Moreover, estimates of heritability, or the proportion of phenotypic variation that is attributed to genotypic variation (h^2), for these traits vary among high-altitude populations. Hemoglobin concentration is estimated to have a strong heritability in both Tibetan and Andean populations ($h^2 = 0.65$ and 0.89, respectively), but oxygen saturation of hemoglobin is only significantly heritable ($h^2 = 0.40$) in Tibetans [4].

The genetic basis of high-altitude adaptation

A growing body of work is focused on the genetic basis of high-altitude adaptation. Moore *et al.* [2] reported one of the first genome-wide scans for selection in high-altitude populations. The authors looked for a signal of positive selection in the genomes of high-altitude Andeans using

*Correspondence: tishkoff@mail.med.upenn.edu
[1]Department of Genetics, University of Pennsylvania, Philadelphia, PA 19104, USA
Full list of author information is available at the end of the article

a set of more than 11,000 genome-wide single nucleotide polymorphisms (SNPs). Specifically, the authors calculated two summary statistics that identify the subset of variants that are highly differentiated between high- and low-altitude populations and may be involved in regionally restricted adaptation to high altitude. From this list of structured SNPs, they identified a subset of SNPs located in genes involved in the HIF (hypoxia inducible factor) pathway, which plays a role in the response to hypoxia by aiding the survival of cells during oxygen depletion as well as the restoration of normal oxygen levels [5]. Using this approach, the authors identified four HIF pathway genes that are candidate targets of positive selection in the high-altitude Andean population (*NOS2A* (nitric oxide synthase 2), *ADRA1b* (alpha-1B-adrenergic receptor), *EDN1* (endothelin 1) and *PHD3* (HIF-prolyl hydroxylase 3)).

More recently, Bigham *et al.* [6] extended the analysis of Moore *et al.* [2] by using an expanded set of approximately 500,000 SNPs in a panel of high-altitude Native American populations (Andeans, *N* = 50) and low-altitude Native Americans (*N* = 55). The authors utilized four summary statistics to generate lists of candidate genes for high-altitude adaptation, including methods for identifying regions of the genome that are highly structured between high- and low-altitude populations and methods for identifying regions with unusual patterns of linkage disequilibrium in Andean highlanders. The authors further refined their analysis by excluding candidates that were shared between high- and low-altitude populations to enrich for candidates specifically related to high-altitude adaptation. Several candidate genes were identified by at least two summary statistics, including a subset of genes that are in the HIF pathway (*VEGF* (vascular endothelial growth factor), *TNC* (tenascin C), *CDH1* (cadherin 1), *EDNRA* (endothelin receptor A), *PRKAA1* (protein kinase, AMP-activated, alpha 1 catalytic subunit), *NOS2A* (nitric oxide synthase 2), *ELF2* (E74-like factor 2), and *PIK3CA* (phosphoinositide-3-kinase, catalytic, alpha polypeptide)), but several additional candidates were identified with a single summary statistic, including the enzyme EGLN1 (which catalyzes the post-transcriptional formation of 4-hydroxyproline). EGLN1 downregulates HIF targets, including EPO (erythropoietin), which is involved in red blood cell production.

Two recent studies [7,8] have identified candidate regions that may play a role in adaptation to high altitude in highland Tibetan populations. Simonson *et al.* [7] first used tests of neutrality based on extended haplotype homozygosity (iHS), which is useful for identifying intermediate frequency variants within a population, and XP-EHH (cross-population extended haplotype homozygosity test), which is useful for identifying fixed or nearly fixed variants that differ between populations, to identify candidate regions that may have been targets of selection in the Tibetans (*N* = 31) (and excluded regions that were also candidate regions in closely related low-altitude Asians). The authors then narrowed their results to genes involved in candidate pathways for high-altitude adaptation, including genes involved with the HIF pathway and genes in Gene Ontology categories related to high-altitude physiology. The results of this analysis include six candidate loci that were identified with the XP-EHH test (*EPAS1* (endothelial PAS domain protein 1), *CYP2E1* (a cytochrome P450 enzyme), *EDNRA*, *ANGPTL4* (angiopoietin-like 4), *CAMK2D* (calcium/calmodulin-dependent protein kinase II delta) and *EGLN1*), five loci that were identified with the iHS test (*EGLN1*, *HMOX2* (heme oxygenase 2), *CYP17A1* (a cytochrome P450 enzyme), *PPARA* (peroxisome proliferator-activated receptor alpha) and *PTEN* (a phosphatase)), and one locus that was identified with both tests (*EGLN1*). Only the five genes identified with the iHS test were useful for phenotypic association studies because this test is designed to identify regions that are variable within a population. Three phenotypes were included in the analysis: hemoglobin concentrations, hematocrit values and oxygen saturation. The authors used haplotypes, or sets of associated SNPs, with extended haplotype homozygosity (that is, that appear to be a target of selection), to test for association between genotypic and phenotypic variation, and report significant additive associations between hemoglobin concentrations and two haplotypes. These haplotypes overlap two candidate genes, *EGLN1* and *PPARA*. As noted earlier, *EGLN1* downregulates HIF targets. *PPARA* is a transcription factor that is not in the HIF pathway but is classified as one of the genes in the Gene Ontology biological process 'response to hypoxia' [9]. This analysis shows that combining information from candidate pathways and genome-wide scans for selection increases the power to detect genotype-phenotype associations, especially those with effects that are too modest to reach genome-wide significance.

Similarly, Beall *et al.* [8] identified a signal of positive selection in a sample of high-altitude Tibetans at the *EPAS1* locus. More specifically, the authors identified an extended haplotype (that overlaps *EPAS1*) present at high frequency (46%) in the Tibetans (*N* = 35) and low frequency (2%) in a closely related sample of Han Chinese (*N* = 84). Furthermore, additional studies were carried out that demonstrate a significant association between *EPAS1* SNPs and hemoglobin concentrations in two separate samples of high-altitude Tibetans (*N* = 70, *N* = 91). While the function of *EPAS1* is not completely understood, it is a HIF pathway gene that is involved in embryonic development; therefore, it is not clear whether selection has acted on hemoglobin levels in adults, perinatal physiology or some combination of physiological

traits [8]. Nevertheless, these new studies [7,8], which combine scans for selection with candidate gene phenotype analyses, are the first to successfully identify a significant association between a hypoxia-related physiological trait and variation at particular candidate genes for selection. The next obvious step would be to resequence all three genes (*EGLN1*, *PPARA* and *EPAS1*) in the Tibetans to identify specific functional variants that affect hemoglobin concentrations, including potential regulatory variants. The authors also note that future work involving functional analyses is necessary to better elucidate the roles of these genes in high-altitude adaptation [7,8].

Even more recently, Bigham *et al.* [10] carried out an extensive scan for adaptation in high-altitude Tibetan ($N = 49$) and Andean ($N = 49$) populations with a dense genotyping array including over 900,000 SNPs. The authors performed several scans for selection in each high-altitude population and identified 38 and 14 candidate regions in the Andeans and Tibetans, respectively. Interestingly, there is no overlap in candidate regions identified in the two populations. When the authors restrict their analysis to candidate genes and pathways involved in high-altitude physiology, *PRKAA1* and *NOS2A* are the best-supported candidates in the Andeans and *EPAS1* is the best candidate in the Tibetans. There is one HIF pathway gene that is a candidate in both high-altitude populations, *EGLN1*; however, the authors identified population-specific *EGLN1* haplotypes. These combined results indicate that the genetic variants associated with high-altitude adaptation in Tibetans and Andeans arose independently due to convergent adaptation.

The studies described above, which use population genomic approaches, have furthered our understanding of the genetic basis of adaptation to high altitude. It is interesting that one of the genes that is a target of selection and associated with hemoglobin concentrations in Tibetans, *EGLN1* [7], was also identified as a candidate for involvement in high-altitude adaptation in Andean populations [6,10]. *EPAS1*, on the other hand, was only identified in the Tibetan populations, and *PPARA* was not considered with the HIF pathway genes in the Andean analysis, so it remains to be seen whether this gene is involved in high-altitude adaptation in populations other than the Tibetans. Future work that includes the characterization of sequence variation in coding regions will identify any protein altering variants involved in high-altitude adapted physiology. In addition, analyses of gene-expression variation will contribute to the identification of regulatory variants (that is, expression quantitative trait loci, eQTLs) that play a role in adaptation to high altitude. Both these approaches will be facilitated by the increased accessibility of next-generation sequencing and functional assays.

Note added in proof
Since this research highlight was written and in proof, an additional publication by Yi *et al.* (2010, *Science*, 329:75-78) reported a study of sequence variation in exomes and flanking regions in high-altitude Tibetans and identified a candidate intronic SNP in *EPAS1* that is significantly associated with hemoglobin levels and with erythrocyte count.

Author details
[1]Department of Genetics, University of Pennsylvania, Philadelphia, PA 19104, USA. [2]Department of Biology, University of Pennsylvania, Philadelphia, PA 19104, USA.

Published: 10 September 2010

References
1. Akey JM: **Constructing genomic maps of positive selection in humans: where do we go from here?** *Genome Res* 2009, **19**:711-722.
2. Moore LG, Shriver M, Bemis L, Hickler B, Wilson M, Brutsaert T, Parra E, Vargas E: **Maternal adaptation to high-altitude pregnancy: an experiment of nature - a review.** *Placenta* 2004, **25 Suppl A**:S60-S71.
3. Beall CM, Song K, Elston RC, Goldstein MC: **Higher offspring survival among Tibetan women with high oxygen saturation genotypes residing at 4,000 m.** *Proc Natl Acad Sci USA* 2004, **101**:14300-14304.
4. Beall CM: **Andean, Tibetan, and Ethiopian patterns of adaptation to high-altitude hypoxia.** *Integr Comp Biol* 2006, **46**:18-24.
5. Guillemin K, Krasnow MA: **The hypoxic response: huffing and HIFing.** *Cell* 1997, **89**:9-12.
6. Bigham AW, Mao X, Mei R, Brutsaert T, Wilson MJ, Julian CG, Parra EJ, Akey JM, Moore LG, Shriver MD: **Identifying positive selection candidate loci for high-altitude adaptation in Andean populations.** *Hum Genomics* 2009, **4**:79-90.
7. Simonson TS, Yang Y, Huff CD, Yun H, Qin G, Witherspoon DJ, Bai Z, Lorenzo FR, Xing J, Jorde LB, Prchal, JT, Ge R: **Genetic evidence for high-altitude adaptation in Tibet.** *Science* 2010, **329**:72-75.
8. Beall CM, Cavalleri GL, Deng L, Elston RC, Gao Y, Knight J, Li C, Li JC, Liang Y, McCormack M, Montgomery HE, Pan H, Robbins PA, Shianna KV, Tam SC, Tsering N, Veeramah KR, Wang W, Wangdui P, Weale ME, Xu Y, Xu Z, Yang L, Zaman MJ, Zeng C, Zhang L, Zhang X, Zhaxi P, Zheng YT: **Natural selection on EPAS1 (HIF2α) associated with low hemoglobin concentration in Tibetan highlanders.** *Proc Natl Acad Sci USA* 2010, **107**:11459-11464.
9. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.** *Nat Genet* 2000, **25**:25-29.
10. Bigham A, Pinto D, Mao X, Akey JM, Mei R, Scherer SW, Wilson MJ, Herráez DL, Tom B, Brutsaert EJ, Moore LG, Shriver MD: **Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data.** *PLoS Genet* 2010, **6**:e1001116.