HIGHLIGHTED TOPIC | Hypoxia 2015

Archaic inheritance: supporting high-altitude life in Tibet

Emilia Huerta-Sánchez1 and Fergal P. Casey2

1Molecular Cell Biology, University of California, Merced, California; and 2Merced, California

Submitted 20 April 2015; accepted in final form 14 August 2015

Huerta-Sánchez E, Casey FP. Archaic inheritance: supporting high-altitude life in Tibet. J Appl Physiol 119: 1129–1134, 2015. First published August 20, 2015; doi:10.1152/japplphysiol.00322.2015.—The Tibetan Plateau, often called the roof of the world, sits at an average altitude exceeding 4,500 m. Because of its extreme altitude, the Plateau is one of the harshest human-inhabited environments in the world. This, however, did not impede human colonization, and the Tibetan people have made the Tibetan Plateau their home for many generations. Many studies have quantified their markedly different physiological response to altitude and proposed that Tibetans were genetically adapted. Recently, advances in sequencing technologies led to the discovery of a set of candidate genes which harbor mutations that are likely beneficial at high altitudes in Tibetans. Since then, other studies have further characterized this impressive adaptation. Here, in this minireview, we discuss the progress made since the discovery of the genes involved in Tibetans’ adaptation to high altitude with a particular emphasis on describing the series of studies that led us to conclude that archaic human DNA likely contributed to this impressive adaptation.

THE COLONIZATION of the Tibetan Plateau, characterized by low oxygen concentrations, extreme cold, and a harsh arid environment, is undoubtedly one of the most challenging of human settlements. For decades, many studies proposed that Tibetans were genetically adapted to live at high altitude, and it was only in 2010 that seven studies uncovered the genetic basis underlying this adaptation. In those studies (3, 4, 28, 39, 42, 46, 48), two genes were highlighted as having the strongest signatures of positive selection: EGLN1 and EPAS1. Not only did these genes exhibit strong signatures of positive selection, but particular single nucleotide polymorphisms (SNPs) in these genes were statistically associated with the distinct physiology of Tibetans which set them apart from other high-altitude populations (3, 39, 48). More recently, a study discovered a 3.4 kb deletion located 80 kb downstream of EPAS1 that also appears to be at high frequency and specific to high-altitude Tibetans (18), and therefore may have a role in the adaptation. For detailed reviews of the genetic selection scans and the unique physiological adaptations in Tibetans, see Refs. 2, 19, 29, 37, and 38. For example, Tibetans are not prone to pulmonary hypertension due to vasoconstriction compared with Andean highlanders. Neither do they exhibit an increase in hemoglobin concentration (associated with a higher risk of chronic mountain sickness), unlike other high-altitude populations, a trait which is statistically associated with EPAS1 variants (3, 39, 48). Detailed in vitro measurements on peripheral blood lymphocytes and in vivo measurements of lung function of Tibetans living at sea level have shown that the overall effect of their particular EPAS1 variants appears to be a blunted hypoxic response (30). Importantly, the occurrence of intrauterine growth restriction and pregnancy-related pre-eclampsia, as well as infant mortality, is several times lower in Tibetans than unadapted Han Chinese living on the Tibetan Plateau (23, 24, 26). The lower reproductive success associated with living at high altitude is likely to be a very strong selective pressure, and both Tibetan and Andean high-altitude populations possess genetic adaptations that improve reproductive success.

Coincidentally, in 2010, the first draft of the Neanderthal genome was published (10), which provided evidence that modern humans admixed with Neanderthals. Interestingly, a few months after this publication, a distinct archaic human genome, termed the Denisovan, was published, and it was discovered that Denisovans had also contributed DNA to modern human populations (15). This suggests that introgression of archaic human DNA into modern human populations may have accelerated human evolution, and may have been the source for one of the clearest examples of positive selection discovered to date. Therefore, access to archaic genome sequences has provided an additional source of information on human high-altitude adaptation.
Here we discuss the accumulating evidence for genetic adaptation to high altitude since the original discoveries of EPAS1 and EGLN1, and we emphasize the role of adaptive introgression in EPAS1. Increasingly, genetic evidence is revealing that adaptive introgression is a widespread phenomenon in humans and was a key evolutionary force in other human adaptations, such as to UV exposure and pathogen defense (1a, 21, 33, 36, 41).

**How Do We Know that Ancestors of Modern Humans Mated with Archaic Humans?**

Archeological evidence and the fossil record place the Neanderthals in Eurasia from about 250,000 years ago to around 30,000 years ago (13, 14). When we consider that anatomically modern humans left Africa around 60,000 years ago (9, 11, 12), it seems likely that modern humans encountered other hominins like the Neanderthals, and cohabited the same geographical range for at least a few thousand years (13). However, whether or not humans and Neanderthals interbred remained a mystery and a controversial topic. Finally, after the sequencing of the first Neanderthal genome, scientists were able to compare with modern human genomes and found traces of Neanderthal DNA in modern humans.

To test the hypothesis of introgression of Neanderthals into modern humans, Green et al. (10) compared non-Africans to sub-Saharan Africans and concluded that Eurasians carried more Neanderthal alleles than sub-Saharan Africans. Their results suggested that Neanderthal ancestry in Eurasians was about 1-4%, and has since been confirmed by further sequencing of higher quality Neanderthal DNA such as the Altai Neanderthal (31).

**Who Are the Denisovans?**

Unlike the Neanderthals, for which we have an extensive fossil record, the Denisovans are a mysterious population whose only bones found thus far are a distal phalanx and two molars (17, 22, 34). In 2010 the first low-coverage Denisovan genome was sequenced from that phalanx (34). Later, in 2012, better DNA library technology led to the publication of a high-coverage Denisovan genome (22). This genome was called “the Denisovan genome” because the bone was found at the Denisovan cave (at an altitude of 700 m) in the Altai Mountains in southern Siberia. Interestingly, this cave has been an active excavation site for the last 30 years, and it also housed the Neanderthal bones (31) from which the high-coverage Altai Neanderthal genome was sequenced (published in 2014).

![Diagram](http://jap.physiology.org/)

**Fig. 1. D statistic.** The evolutionary relationship between two modern human populations (H1 and H2), an archaic human population (Archaic) and an outgroup (Chimp). “A” is the ancestral allele and “B” is a mutation that occurs in the archaic human lineage after they split from modern human lineage (H1 and H2). The horizontal arrow represents admixture from an archaic population into a modern human population, and the time of that gene flow is denoted t_{gf}. The D statistic measures the enrichment of ABBA or BABA sites. If admixture from the archaic human into H2 occurred, on average, H2 would share more mutations with the archaic individual (one ABBA configuration, panel 1). If admixture from the archaic human into H1 occurred, on average, H1 would share more mutations with the archaic individual (more BABA configurations, panel 2). D is the difference between the numbers of ABBA sites and the BABA sites normalized by their sum. D is computed by using sites where the archaic human carries a derived allele (allele that is different from the chimp allele). Under a model of no introgression, H1 and H2 should be closer to each other, and the expected difference of ABBA and BABA sites should be 0. However, if introgression occurred, a deviation from zero will be observed.

\[
D(H_1, H_2, \text{Archaic, Chimp}) = \frac{\sum_{i=1}^{t} I_{ABBA}(i) - I_{BABA}(i)}{\sum_{i=t} I_{ABBA}(i) + I_{BABA}(i)}
\]
Even though we lack a Denisovan fossil record, its genome provides us with some information about this enigmatic population. For example, we know the Denisovan genome is more similar to the Neanderthal than to any modern human population, indicating that these two groups likely descended from the same ancestral population (22, 34). Estimates of heterozygosity, the portion of sites that differ between the paternal and maternal genomes, are extremely low (about 20% of that seen in Africans) which suggests that the Denisovan population size was small (22). Using $D$ statistics (see Fig. 1 and How Do We Know that Ancestors of Modern Humans Mated with Archaic Humans?), Reich et al. (34) found evidence of Denisovan admixture into some modern human populations. Notably, the Melanesians were the only population found to harbor considerable (4–6%) Denisovan ancestry (34). However, those first estimates were based on comparisons with only a few modern human genomes. Another study that included fewer DNA sites but many more individuals found that Southeast Asians also carried a small proportion of Denisovan DNA (40). Recently, improved methods that include more complex features of the data (not just the information to compute $D$ statistics), such as the correlation between nearby sites, has identified the DNA segments in modern humans that derive from both Neanderthals and Denisovans (31, 33). These new results confirm that East Asians do carry small amounts of Denisovan ancestry of around 0.2% (31, 33).

Denisovan-like alleles at the EPAS1 gene in Tibetans

Since the genome sequences of the Denisovan and Neanderthals became available, not only do we know that interbreeding with modern humans took place, but that sometimes modern humans received an evolutionary advantage from doing so. This process is called adaptive introgression, and to date a few studies have reviewed examples of this process in modern humans (33). Surprisingly, a key example of adaptive introgression comes from one of the most celebrated cases of human adaptation: adaptation to high altitude in Tibetans.

Pinpointing the genes involved in Tibetans’ adaptation to high altitude was the first step in unraveling the molecular basis of this adaptation. The discovery of the $EPAS1$ gene as a functionally relevant target of selection attracted attention in the field, but many questions remained unanswered. Were there other mutations at high frequency in Tibetans and absent in other populations? Which was the beneficial mutation(s)? Did those mutations modulate the expression of $EPAS1$? How the Tibetans acquired these mutations was of considerable interest. Did these mutations occur de novo in Tibetans after they arrived at the Tibetan Plateau, or were they preexisting? When did positive selection begin, and what was the age of the beneficial mutation? If Tibetans acquired the beneficial mutation de novo [selection de novo (SDN)], then the mutation likely occurred after the population split between the Tibetans and the Han (see Fig. 2A) and it would be immediately beneficial. In this case, the age of the allele would be the same as the timing of selection. However, Yi et al. (48) found that the mutation with the highest frequency difference was also present in the Han and absent in the Danes, suggesting that the mutation likely occurred before the Tibetans split from the Han (Fig. 2B). In this case, since the mutation probably existed in the ancestral Han-Tibetan population and selection would have only taken place after the population split, a model describing selection acting on a preexisting mutation [also known as selection on standing variation (SSV)] appeared to be more suitable, and under this model the age of the mutation is older than the timing of selection.

To begin addressing these questions, Huerta-Sánchez et al. (15) resequenced a region of roughly 130 kb that contains the whole sequence of the EPAS1 gene, and revealed a segment of 32 kb bases exhibiting large genetic differentiation ($\sim$0.80) between the Tibetans and the Han population in contrast to the genome-wide estimates ($\sim$0.02). These markedly large frequency differences had previously been observed by Peng et al. (28). These mutations are located in intronic regions, suggesting that they may affect the transcriptional regulation of the gene, either by alternative splicing or by modulating expression levels. Specifically, in this region, 32 mutations had frequency differences of at least 75%, one of which had been identified as the mutation with the largest frequency difference between Tibetan and Han in an earlier study (48). Closer inspection of the haplotypes suggested that some of these highly differentiated mutations were absent not only in the Han individuals but also in virtually all of the 1000 Genomes Project sequences composed of 13 populations (1). The absence of many of these mutations in the 1000 Genomes Project populations suggested another scenario: that Tibetans had acquired the haplotype from a distant population. Indeed, simu...
lations under the two positive selection scenarios of SSV or SDN could not explain the genetic variation observed at $EPAS1$ (15). A model where Tibetans received this haplotype from another population (through admixture) seemed a more plausible explanation (see Fig. 2C), and, remarkably, a comparison to the Neanderthal and Denisovan genomes showed many Tibetan SNPs were distinctly Denisovan derived (15). Figure 3 illustrates the similarity between the Tibetan selected haplotype, the two Neanderthal and two Denisovan haplotypes, as well as the closest modern human haplotype. To find the closest modern human haplotype to the Tibetan haplotype, we combined the 2,184 haplotypes from the 1000 Genomes Project (1) data with the 80 Tibetan haplotypes. In this plot, we prefiltered sequences to remove sites that are rare (<5%) or common (>95%) in both Tibetans and the 1000 Genomes Project. Clearly, Tibetans are closest to the Denisovan haplotype with only nine differences between them at these sites. In contrast, the Tibetans have 33 differences to Neanderthal, so the archaic component is clearly more likely to be of Denisovan origin. The closest modern human haplotype to Tibetans is from a southern Han Chinese (CHS) individual with 13 differences between them. In addition, calculations of $D$ statistics at this 32-kb region comparing Tibetans to Han, Europeans or Yorubans are statistically significant, so we can reject the null hypothesis of no introgression (15). Other statistics used to identify introgressed segments that depend on the linkage disequilibrium between SNPs (30a, 41) are also statistically significant in this region. Since the Tibetan haplotype does not perfectly match the Denisovan haplotype, the possibility remains that the archaic haplotype merely hitchhiked on an advantageous de novo mutation. Functional analyses will hopefully resolve which mutation is the causal one.

Another possibility to consider is that the Tibetans and Denisovans may share haplotypes by chance. For example, the haplotype could have been present in the ancestral Tibetan-Denisovan population and it could have survived in both Tibetans and Denisovans. However, over time, recombination would fragment the haplotype, and subsequently the expected haplotype length would be small. Assuming a divergence time between Denisovans and Tibetans of 200,000 and a recombination rate of 2.3e-8, we found the expected length of a shared haplotype to be significantly shorter than 32 kb (15). Therefore we can exclude this possibility of similarity by chance.

Remaining Questions

Even though we have identified the possible source of the beneficial mutations that made it possible for Tibetans to thrive at high altitude, many questions remain to be answered. We still do not know the biological mechanism that enables Tibetans to deal with the low-oxygen conditions at high altitude. The absence of coding mutations under selection suggests that the adaptation might instead modulate the expression of the gene rather than change its protein sequence or structure. We believe that the $EPAS1$ haplotype carried by Tibetans may cause a partial loss of function of the gene. The putatively selected SNPs all occur within the $EPAS1$ introns, one very close to a splice site of exon 6, so it is possible that splicing is affected and there is an associated shift in isoform abundances.

Loss of function is supported by the recessive nature of the genotype, and many previous studies have concluded that gain-of-function mutations in $EPAS1$ (leading to an elevated response in hypoxia-inducible genes) are strongly associated with erythrocytosis and pulmonary hypertension in both non-human (8, 25, 43) and human observational studies (7). Conversely, heterozygous loss-of-function $EPAS1$ mouse models are protected against pulmonary hypertension when subjected to a low-oxygen environment for several weeks, compared with wild type (5). The precise mechanism by which the protective effect is conferred might be related to vasoconstrictive and vascular remodeling factors (such as endothelin-1) that are known to be mediated by $EPAS1$ response to hypoxia (5).

Other open questions are the extent of Denisovan admixture in Tibetans genome-wide, and whether other loci under adaptive pressures (e.g., $EGLN1$) harbor Denisovan-like haplotypes. In addition, the previous estimates of the time of selection and selection strength are based on a model where the mutations occur de novo in the Tibetans (28), which, in light of the new evidence for adaptive introgression, should be revised. That may imply that while the age of the mutation is old, the time of the selection event may be quite recent.

Fig. 3. Comparison with Denisovan and Neanderthal haplotypes. Each column represents a single nucleotide polymorphism (SNP). The first two rows are the two Denisovan haplotypes (in green), followed by the Tibetan (TIB) most common haplotype (in pink), the closest modern human haplotype in the 1000 Genomes Project data to the Tibetan haplotype (in yellow, Han Chinese (CHS)), and the two Neanderthal (NEAN) haplotypes (in blue). If an allele is different from the chimp allele, we say it is derived (shown as black). If the allele matches the chimp, we say it is ancestral (shown in gray). Note that the Neanderthal carries two heterozygous sites, and for illustration purposes we have assumed a phase for the Neanderthal haplotypes.
Another key parameter is the Tibetan-Han population split time. Genetic estimates range from the very recent (~2,800 yr) to the very ancient (~30,000 yr). Yi et al. (48) used the frequency distribution of synonymous mutations and estimated the Tibetan-Han divergence to be around 2,800 years (48). On the other hand, analysis of common Tibetan mitochondrial haplogroups put most divergence estimates within 11,000–22,000 years (32, Table 1). Y chromosome data yielded similar estimates: 8,000–30,000 years (32, Table 1). However, it is worth noting that estimates from single loci (like the mitochondria or Y chromosome) tend to overestimate population split times if the ancestral population is large (35). Estimates from genome-wide SNP data varied from 6,500–12,000 years depending on the sampling of SNPs included in the analysis (32). Recently, using multivariate statistical methods such as principal components analysis, Jeong et al. (16) found evidence that Tibetans are a mixture of two ancestral populations (related to the East Asians and the Sherpas) and confirmed that SNPs nearest to EPAS1 have the strongest differentiated signal within Sherpas, indicative of selection. Another study of Deedu Mongolians living in the Qinghai-Tibetan Plateau identified positive selection on EPAS1 (45). Therefore, it appears that the same locus is under selection in these three high-altitude populations, and may have shared the allele by admixture with each other or inherited the allele from a common ancestor. A complete demographic scenario would include both recent and ancient events, and would provide estimates of key demographic parameters such as the time of admixture. Once we have a complete demographic model, then we can extend the model to incorporate positive selection and quantify the strength and time of selection. A combination of approaches and the availability of whole-genome sequence data from multiple individuals will provide a clearer picture of Tibetan demographic history and confirm the sequence of events that led to this remarkable example of adaptation.

Conclusions

In light of the new evidence that archaic humans have donated genetic material to modern humans, and modern humans may have gained a valuable selective advantage from the exchange (33), we need to consider adaptive introgression as one of the main evolutionary forces behind human adaptation. One of the clearest examples of positive selection found to date in the EPAS1 gene in Tibetans shows very strong evidence for adaptive introgression.

GRANTS

E. Huerta-Sánchez was supported by University of California Merced startup funds.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

AUTHOR CONTRIBUTIONS

E.H.-S. and F.P.C. prepared figures; E.H.-S. drafted manuscript; E.H.-S. and F.P.C. edited and revised manuscript; E.H.-S. and F.P.C. approved final version of manuscript.

REFERENCES


