



Tracing the Selection on the Human *ADH1B* Gene

Hui Li, Sheng Gu, Kenneth K. Kidd
Department of Genetics, Yale School of Medicine, New Haven, CT06520

Alcohol dehydrogenase (ADH) is a widely studied enzyme as is the gene family encoding the forms of this enzyme. Previous studies have shown that the *ADH1B**47His allele is associated with a decrease in the risk of alcoholism and the core region with this allele has undergone positive selection in some populations. A literature review identified studies reporting allele frequencies of this polymorphism for 131 population samples (for a total of 168 when combined with our new data on 37 populations). The derived *ADH1B**47His allele reaches high frequencies only in West and East Asia, but has a low frequency in the region between East and West Asia, suggesting that the derived allele increased in frequency independently in the two regions. We tested seven single nucleotide polymorphisms (SNPs) and two short tandem repeat polymorphisms (STRPs) in the *ADH1B* region in the world sample to form the haplotypes. Seven haplogroups were defined with different SNP allele patterns. H5, H6, and H7 are haplogroups with the derived *ADH1B**47His allele. H5 is restricted to West Asia and H6 to the East Asia and Pacific region. H7 has in addition to H6 the derived allele of rs3811801 in the regulatory region, and is restricted to East Asia. We analyzed 24 population samples from East Asia covering six ethnic families and find H7 is enriched in the Hmong, Han Chinese, and Altaic families. We typed 23 more SNPs in about 170kb flanking region of *ADH1B*. The extended haplotype homozygosity (EHH) and relative EHH (REHH) tests for the *ADH1B* core were consistent with selection for the haplogroups with derived SNP alleles in the Hmong and Altaic. Other populations showed only a weak signal at best. The selection distribution is significantly correlated with the frequency of the derived allele at the regulatory polymorphism rs3811801, not the derived amino-acid altering allele *ADH1B**47His. Thus, the real focus of selection may be the regulatory region of H7. The *ADH1B* downstream STRP provides relative age estimates for the SNP-based haplogroups that are in general agreement with the Out of Africa pattern and a recent expansion of H7 in East Asia. The refined pattern of variation for *ADH1B* haplogroups will help design studies to understand the selective force(s) that may have operated.

1. Special Geographic distribution of *ADH1B**47His allele

ADH1B Arg47His is the first polymorphism studied in the ADH region. This polymorphism results in the different structures of the enzyme, and therefore, was noticed and easy to examine as a protein. In the past 30 years, many populations have been studied for this polymorphism. We collected data from the literature as well as from new population samples. The frequency of *ADH1B**47His varies enormously among the geographic regions[3]. High frequency of this SNP was only found in Southwestern and Eastern Asia (Fig.1)

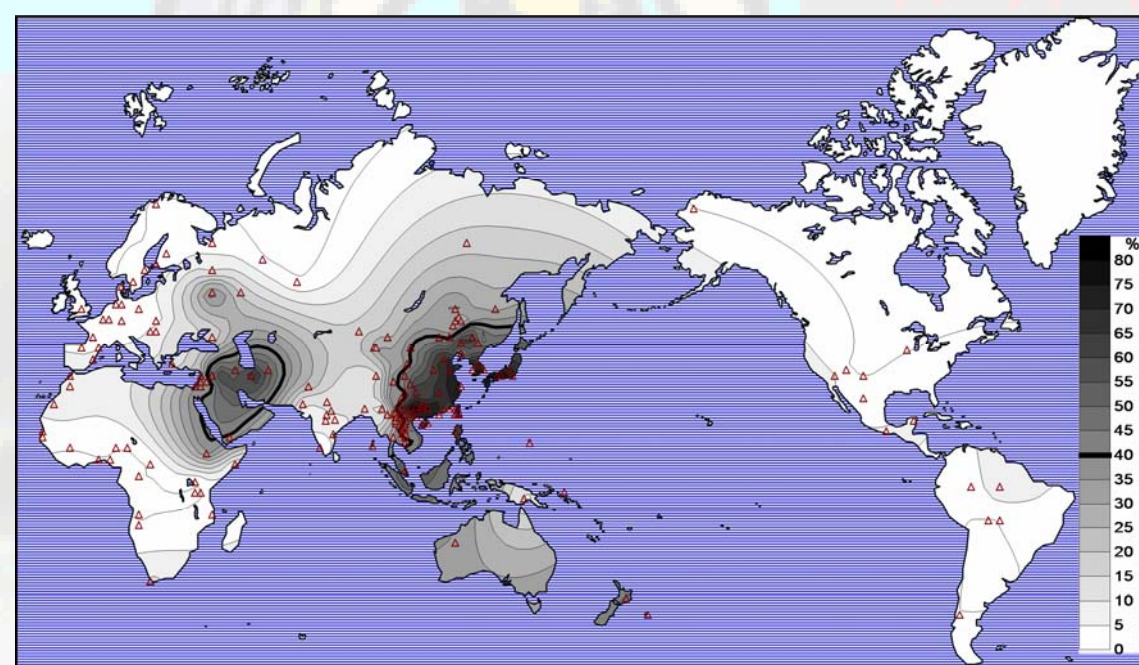


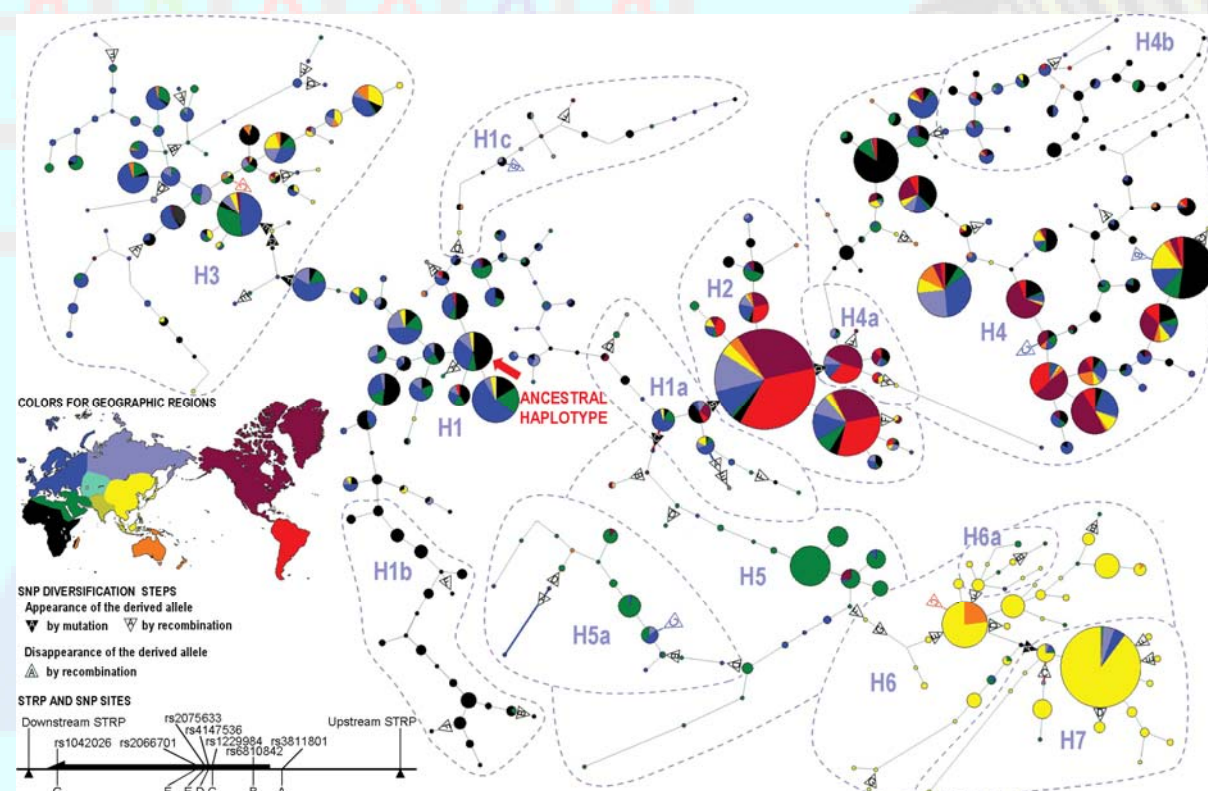
Fig.1 Distribution of *ADH1B47His allele**
Triangles are population samples included in the analysis. Bold lines outline the high frequency areas of the *ADH1B**47His allele.

2. Different haplogroups in Eastern and Western Asia

To examine whether the high frequencies of *ADH1B**47His allele in Southwestern and Eastern Asia arose independently[1], we analyzed six more SNPs and two STRPs in the *ADH1B* region. The haplotypes of the nine polymorphisms were estimated by PHASE, and the network of the haplotypes was constructed by NETWORK program (Fig.2). In the network, seven major haplogroups were determined according to the SNPs. Three of the haplogroups with the *ADH1B**47His allele, H5, H6, and H7, are the most geographic specific. H5 is mostly found in Western Asia, while H6 and H7 are mostly in Eastern Asian. Therefore, the *ADH1B**47His allele is in the different haplogroups in the two geographic regions, and the high frequency must have occurred independently. The two most frequent haplotypes, the central haplotypes of H2 and H7 may have reaches high frequencies for different reasons: the Amerindians went through a bottleneck into America; while positive selection appears to have occurred in Eastern Asia.

Fig.2 Network of *ADH1B* haplotypes

Seven haplogroups were defined based on the SNPs in the order from A to G.
H1:CGGGACA
H2:CTGGACA
H3:CGGGGTG
H4:CTGTACA
H5:CGAGACA
H6:CGAGGTG
H7:TGAGGTG



3. Ethnic associated positive selection in Eastern Asia

We previously found evidence for positive selection for H7 in some East Asian populations[2]. Using the same methods, Long Range Haplotype Homozygosity, the new populations we have studied cover all of the ethnic/linguistic families in Eastern Asia (Fig.3), including Altaic (Uigur, Khazak, Yakut, Mongol, Korean, Japanese), Sino-Tibetan (Tibetan, Qiang, Dee, Wu Chinese, Minnan Chinese, Cantonese, Hakka, Phunoi), Hmong-Mien (Black Hmong and White Hmong), Daic (Kam, Hlai, Laka, Zhuang, Lao, Saek), Austro-Asiatic (Ava, Khmu, Bo, Katu, Laven, Khmer), and Austronesian (Atayal, Amis, Tsat, Micronesian). All of the haplogroups in the Eastern Asian populations were examined and only H7 showed high EHH and REHH values, which are signals for potential positive selection. However, the strength of the evidence for selection varies among populations. Highest REHH values were found in Altaic and Hmong-Mien populations, especially Mongol (MVF). Four Chinese populations also show fairly high REHH. Therefore, the REHH value is highly associated with the ethnic classification. One can speculate that

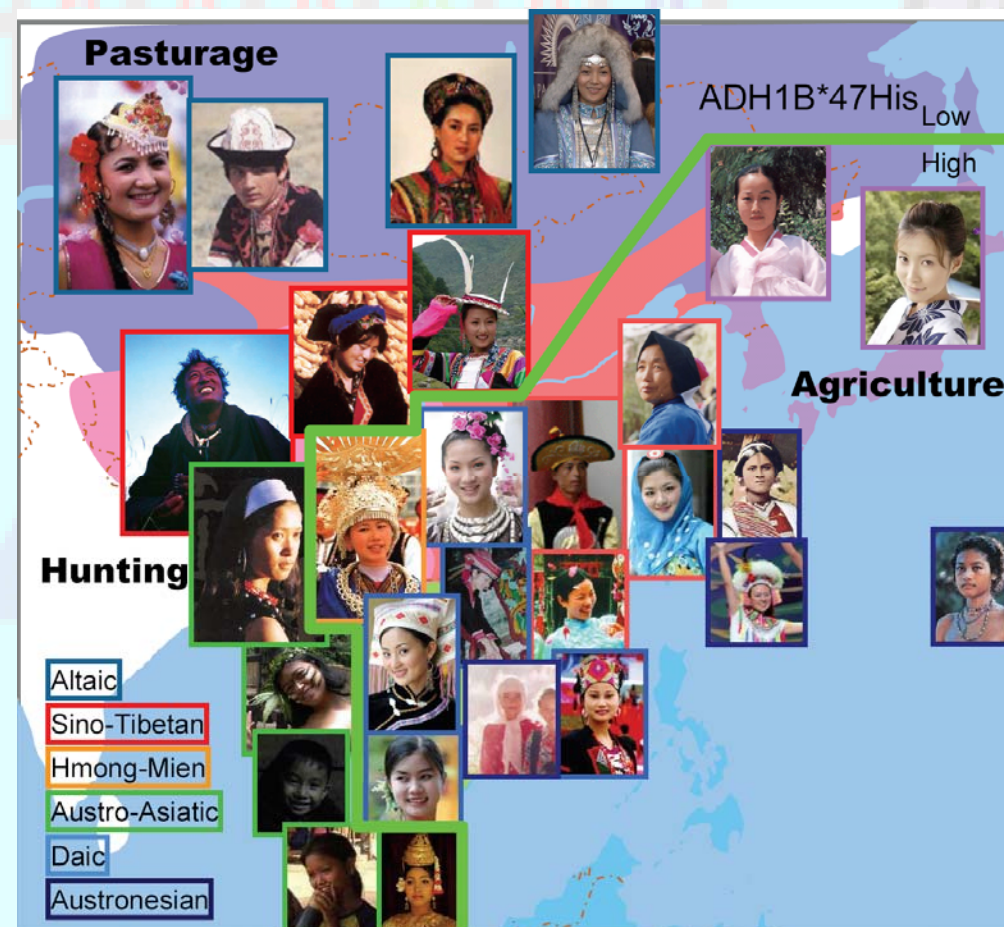


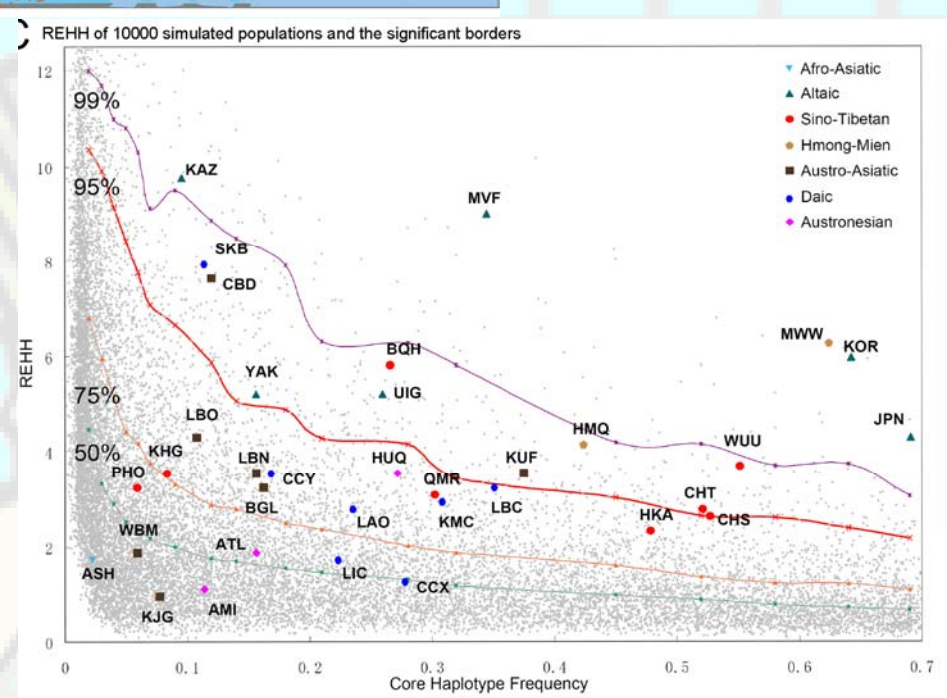
Fig.3 Populations studied in Eastern Asia

Populations are classified into linguistic families marked by different colors of the frames. A green bold line divided the region into east and west parts. To the east of the line is the area with a high frequency of *ADH1B**47His. The eastern populations are all agriculture groups, indicating a potential association between life styles and *ADH1B* gene diversity.

positive selection might have happened in the ancient Altaic population in the northeast region of the Eastern Asian.

Fig.4 REHH values of H7 in the Eastern Asian population and the simulated REHH values

Observed REHH values of the populations are shown in the colored dots, while simulated values are the gray dots. Four significance levels were established by the simulations. The most significant values were found in Mongol (MVF), White Hmong (MWW), Korean (KOR), and Japanese (JPN).



4. Ages of the haplogroups and the recent selection

The variances of STRPs can be used to determine the ages of both the populations and the haplogroups. The *ADH1B* downstream STRP contains sufficient variance, and was used to estimate the population ages and haplogroup ages. Figure 5 displays the distributions and variances of the STRP in different regions of the world. No surprise that highest variance was found in sub-Saharan Africa, representing the African origin of modern humans. A continuous reduction of variance was observed from Europe via Siberia to America, which is a major route of early migration suggested by other studies. The major repeat number of the STRP in Eastern Asia is most different from the other regions, indicating the particularity of *ADH1B* in Eastern Asia. The two peaks in the distribution of the STRP in Oceania suggest the dual origin of the Oceanians. Combined with the geographic distributions of the haplogroups in Figure 2, the probable migration routes of the haplogroups are given in Figure 6.

Fig.5 Distributions and variances of *ADH1B* downstream STRP

X-axis is the repeat numbers of the STRP. Y-axis is the individual sample number.

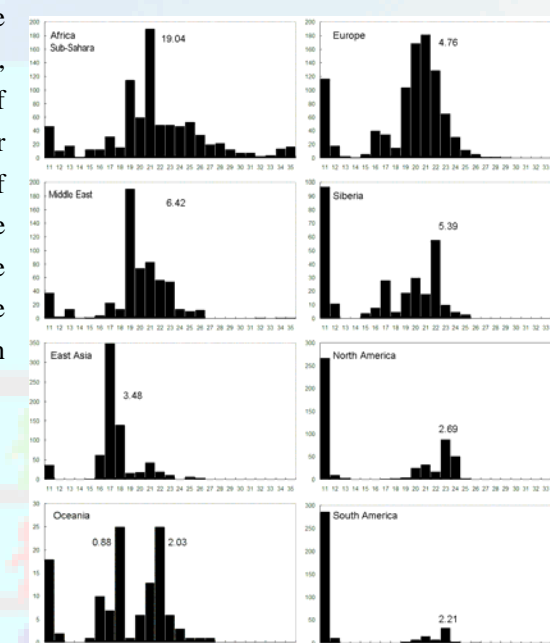
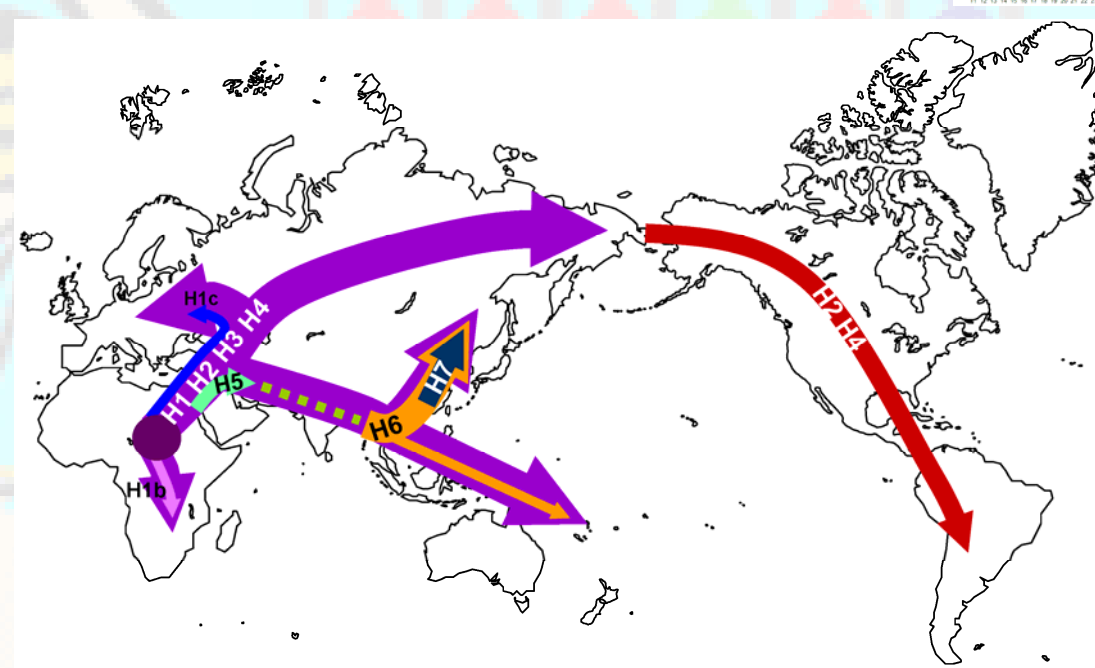


Fig.6 Migration routes of the *ADH1B* haplogroups

H1, H2, H3, and H4 form a background in the Old World. H2 and H4 went through a bottleneck into the New World. H5 occurred in Western Asia. H6 and H7 occurred in Eastern Asia.

Giving that the age of modern humans is around 200,000 years, we estimated



the mutation rate of the *ADH1B* downstream STRP and the ages of each haplogroups. H5 is around 53 thousand years old, which is close to the age for Modern Human out of Africa. H6 is around 15 thousand years old, approximately the age of the earliest archaeological cultures found in Eastern Asia. H7 is the youngest, with an age of 2.5 thousand years, indicating a really recent increase in frequency of this haplogroup, i.e., selection.

Conclusion

We demonstrated the geographic associated distribution and recent selection of the human *ADH1B* gene. The *ADH1B* gene must have played an important role in the history of modern humans and its changes in life style.

Main references:

- Osier MV, Pakstis AJ, Soodyall H, Comas D, Goldman D, Odunsi A, Okonofua F, Parnas J, Schulz LO, Bertranpetit J, Bonne-Tamir B, Lu RB, Kidd JR, Kidd KK (2002) A global perspective on genetic variation at the ADH genes reveals unusual patterns of linkage disequilibrium and diversity. *Am J Hum Genet* 71:84-99.
- Han Y, Gu S, Oota H, Osier MV, Pakstis AJ, Speed WC, Kidd JR, Kidd KK (2007) Evidence of Positive Selection on a Class I ADH Locus. *Am J Hum Genet* 80:441-456.
- Li H, Mukherjee N, Soundararajan U, Tarnok Z, Barta C, Khaliq S, Mohyuddin A, Kajuna SLB, Mehdi SQ, Kidd JR, Kidd KK (2007) Geographically Separate Increases in the Frequency of the Derived *ADH1B**47His Allele in Eastern and Western Asia. *Am J Hum Genet* 81: 842-846.