Quaternary International 426 (2016) 187–194

Contents lists available at ScienceDirect

Quaternary International

journal homepage: www.elsevier.com/locate/quaint

Molecular adaption of alcohol metabolism to agriculture in East Asia

Ling-Xiang Wang, Shaoqing Wen, Chuan-Chao Wang, Boyan Zhou, Hui Li*

MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai, China

ARTICLE INFO

Article history:
Available online 7 April 2016

Keywords:
Alcohol metabolism
Positive selection
Agriculture
East Asia
Neolithic age

ABSTRACT

Human Alcohol metabolic gene families ADH and ALDH mainly comprise ten genes, expressing in different organs and tissues. Two of the variants, ADH1B*47His and ALDH2*504Lys reach very high frequency in East Asia, while are almost absent in the rest of the world, which is believed to be results of positive selection related to the development of agriculture in Neolithic time. In addition to the *ADH* alcohol metabolic pathway, a microsomal ethanol oxidizing system involving the gene *CYP2E1* has also been identified. The study on the micro-evolution of alcohol metabolic genes will help us understand how human adapted to the artificial environment of agriculture.

We collected 1211 samples from 44 worldwide populations including 19 representative populations in East Asia, combined with 2504 samples of 26 populations from 1000 Genome project. We scanned all the 23 missense mutations or pathogenic SNPs in *ADH* and *CYP2E1* pathways. Then we examined the selection signature on the genetic polymorphism in East Asia and estimated the allele ages. Our analyses revealed that the long-term farming ethnic groups in East Asia, such as Han, differed from the nomadic populations in the pattern of alcohol-related genetic polymorphism. This divergence was mainly attributed to the 6 closely related functional SNPs rs1229984 (*ADH1B*), rs671 (*ALDH2*), rs8187929 (*ALDH1A1*), rs2228093 (*ALDH1B1*), rs3813867 (*CYP2E1*), and rs2031920 (*CYP2E1*). The derived core haplotypes of the new detected SNPs showed moderate to strong selection signals and the estimated allele ages coincided with the Neolithic time. The driving force tended to be the emergence and expansion of agriculture in East Asia.

1. Introduction

The domestication of plants and animals since the Neolithic age has triggered a rapid increase in human population, coupled with vast changes in cultures and ecology, creating new opportunities for adaptation (Bocquet-Appel, 2011). Demographic expansion provided the fundamental for new adaptive evolution. Cultural and ecological transition changed the selective pressures. Some of the most radical new selective pressures have been associated with the transition to agriculture (Armelagos and Harper, 2005). For example, genes related to disease resistance are among the inferred functional classes most likely to show evidence of recent positive selection (Wang et al., 2006). Virulent epidemic diseases, including smallpox, malaria, yellow fever, typhus, and cholera, became important causes of mortality after the origin and spread of agriculture (McNeill, 2010). Likewise, subsistence and dietary changes have led to selection on genes such as lactase (Bersaglieri et al., 2004).

In East Asia, the emergence of agriculture about 10,000 years ago and its subsequent longtime flourishing (Gong et al., 2007; Yang et al., 2012; Gross and Zhao, 2014) made a dramatic change in people's diet, behavior, and culture, which might leave selection signals in the genomes of modern Asians. The alcohol metabolism is among the most significant selection targets, as fermented food and beverages produced by rice can date back to the Neolithic period in China (McGovern et al., 2004).

Current study shows that alcohol metabolism is a complex system with multiple genes involved, and is related to interaction of genetic and environmental factors. There is also a high individual variability in ethanol metabolism, with alcohol elimination rates varying as much as three to four-fold among individuals (Li et al., 2001). Such an individual variability is mainly due to genetic variations in the main ethanol and acetaldehyde metabolizing enzymes. Particularly, there are multiple molecular forms of the alcohol dehydrogenase (ADH1A, ADH1B, ADH1C, ADH4, ADH5, ADH6, ADH7) and aldehyde dehydrogenase (ALDH1A1, ALDH1B1, ALDH2), expressing in different organs and tissues. Two of the most studied functional variants, ADH1B*47His (rs1229984) and

* Corresponding author.
E-mail address: huili.fudan@gmail.com (H. Li).

http://dx.doi.org/10.1016/j.quaint.2016.03.008
1040-6182/

ALDH2*504Lys (rs671) together are responsible for the Asian flushing reaction due to high acetaldehyde levels (Harada et al., 1999; Osier et al., 1999, 2002; Oota et al., 2004; Lu et al., 2005; Li et al., 2008, 2009). They both reach high frequencies in East Asia, while are almost absent in the rest of the world, which is believed to be results of positive selection and is closely related to the development of Neolithic agriculture in China (Peng et al., 2010; Li et al., 2011). Further, the obvious ethnicity-related distributions of *ADH1B* diversities suggest the existence of some culture-related selective forces acting on the *ADH1B* region (Li et al., 2008). Additionally, a strong signature of positive selection was detected for the *ADH* gene cluster in a genome-wide analyses (Voight et al., 2006).

In addition to the *ADH* pathway, a microsomal ethanol oxidizing system involving mostly the gene *CYP2E1* on chromosome 10 has also been identified (Lieber and DeCarli, 1968; Lieber, 1999). All these genes involving in the alcohol metabolism have important functions, high diversities, and complex inter-gene reactions. Previous studies mostly focused on one or two certain genes, which made the whole evolution pattern of alcohol metabolism in East Asia unclear. Here, we scanned all the missense or pathogenic SNPs reported to be functional in *ADH* and *CYP2E1* pathways in a worldwide sample. We also examined the selection signature on the genetic polymorphism in East Asia and estimated the allele ages. Then we hypothesized that the alcohol metabolic system was mainly selected during the Neolithic time, and the driving force tended to be the emergence and expansion of agriculture, which led to the current ethnic-related distribution of the alcohol-related genetic polymorphism in East Asia.

2. Material and method

2.1. Samples

We typed 1211 individuals from a global sample of 44 populations (Table S1). According to population ancestry and geographic locations, these 44 populations are categorized into six groups. The populations and sample sizes are as follows: Africa: Biaka Pygmy 35, Sandawe 26, African American 17, Hausa 20, Mbuti Pygmy 17, Masai 12, and Ibo 17; Europe: Adygei 20, European American 42, Finnish 20, Hungarian 43, Russian from Archangelsk 21, Russian from Vologda 21, and Ashkenazi Jews 23; West and South Asia: Druze 43, Samaritan 26, Keralite 6, and Nepal 22; East Asia: Yakut 28, Atayal 30, Ami 27, Cambodians 24, Laotians 120, Han (Taiwan) 42, Koreans 34, Yughur (China) 60, Baoan (China) 44, Dongxiang (China) 42, Han (Central China) 10, Han (Northwest China) 12, Han (South China) 18, Hui (Northwest China) 46, Kazakhstan (China) 33, Kyrgyz (China) 46, Tuvas (China) 48, Tibetan (China) 76, and Yi (China) 4; Oceania: Nasioi Melanesian 9; America: Maya 12, Quechua 11, Karitiana 24, Ticuna 25, and Surui 27. Apart from the new added population data in China and Nepal, the detailed information of the other populations can be found in the Allele Frequency Database (ALFRED). All samples were collected with informed consent under protocols approved by the relevant institutional review boards. The haplotype data of 2504 individuals of 26 populations from 1000 Genome project phase 3 data set was merged together.

2.2. Markers and genotyping

According to the dbSNP annotation, we scanned all the missense mutations or pathogenic SNPs across the gene cluster *ADH1A*, *ADH1B*, *ADH1C*, *ADH4*, *ADH5*, *ADH6*, *ADH7*, *ALDH1A1*, *ALDH1B1*, *ALDH2*, *CYP2E1*. The rs numbers of the 23 SNPs were as follows: rs28730623, rs1126673, rs1126671, rs28720153, rs2066702, rs1229984, rs698, rs1693482, rs283413, rs59534319, rs1573496, rs671, rs8187929, rs2228093, rs2073478, rs113083991, rs4878199, rs3813867, rs2031920, rs2070673, rs915906, rs6413419, and rs6413432. Genotyping was conducted by the NGS method on Illumina Hiseq2000. We designed a series of primers for covering the candidate SNPs regions. SNPs were called with an average allele depth >200×. Overall, missing genotypes account for 1.5% of the total, with no SNP exceeding 5% missing genotypes among the 1211 individuals.

2.3. Statistics

Haplotype estimation and imputation were conducted using Beagle version 4.1 (Browning and Browning, 2007). Heatmap was calculated using the public program in R. Principal Component Analysis (PCA) of population allele frequencies used XLSTAT (Version 2014.4.04; Addinsoft SARL, http://www.xlstat.com/). The recent positive selection was detected using the LRH method (Sabeti et al., 2002). The extended haplotype homozygosity (EHH) and the relative EHH (REHH) were examined using the phased haplotype data of East Asians (CDX, CHB, JPT, KHV, and CHS) from the 1000 Genome project.

Allele age calculations were conducted by the standard methods published previously (Slatkin and Rannala, 2000; Wang et al., 2004; Hawks et al., 2007). In brief: $t = [1/\ln(1-c)]\ln[(x(t)-y)/(1-y)]$, where t = allele age (in generations), c = recombination rate, x(t) = frequency in generation t, and y = frequency on ancestral chromosomes. We assumed that the origin of the derived allele is on the background of the ancestral allele haplotype, and the calculation utilizes the value of c, determined from the 1000 Genome project phase 3 data set. The phased haplotype data of East Asians (CDX, CHB, JPT, KHV, and CHS) is obtained from the 1000 Genome project. Regions with <0.1 cM/Mb average recombination rate were excluded. For conversion of time in generations, t, into time in years, a generation time of 25 years was assumed. This method is a method-of-moments estimator (Slatkin and Rannala, 2000), because the estimate results from equating the observed proportion of non-recombinant chromosomes with the proportion expected if the true value of t is the estimated value. It requires no population genetic or demographic assumptions, only the exponential decay of initially perfect LD because of recombination.

3. Result

3.1. The pattern of allele frequencies

We analyzed a total of 1211 individuals from 44 worldwide populations, merged with data of 26 populations from the 1000 Genomes Project Phase 3. The heatmap in Fig. 1 was based on the population allele frequencies for the 23 functional SNPs in all the alcohol metabolism related genes. It allows a very quick visualization (1) of the relationship of each SNP in the data set to the others, and (2) of how each SNP contributes to distinguishing among populations. The marginal dendrograms showed the relationships of the SNPs and of the populations graphically. The 69 worldwide populations clustered well according to their geographical ethnicity. The diverse patterns of allele frequency among continental populations indicate the worldwide differentiation of alcohol metabolism. East Asians are clustered in two groups. One group consists of the nomadic populations such as Kazakhstan, Yakut, Kyrgyz, Tuvas etc., while another contains the farming ethnic groups with relatively high frequency of ADH1B*47His (rs1229984) and ALDH2*504Lys (rs671). According to the K-means clustering, 3 SNPs were discovered to be close related with rs1229984 and rs671 respectively. That is, rs2228093 (*ALDH1B1*) clusters with rs1229984, two pathogenic SNPs rs3813867, rs2031920 (*CYP2E1*) cluster with rs671.