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.

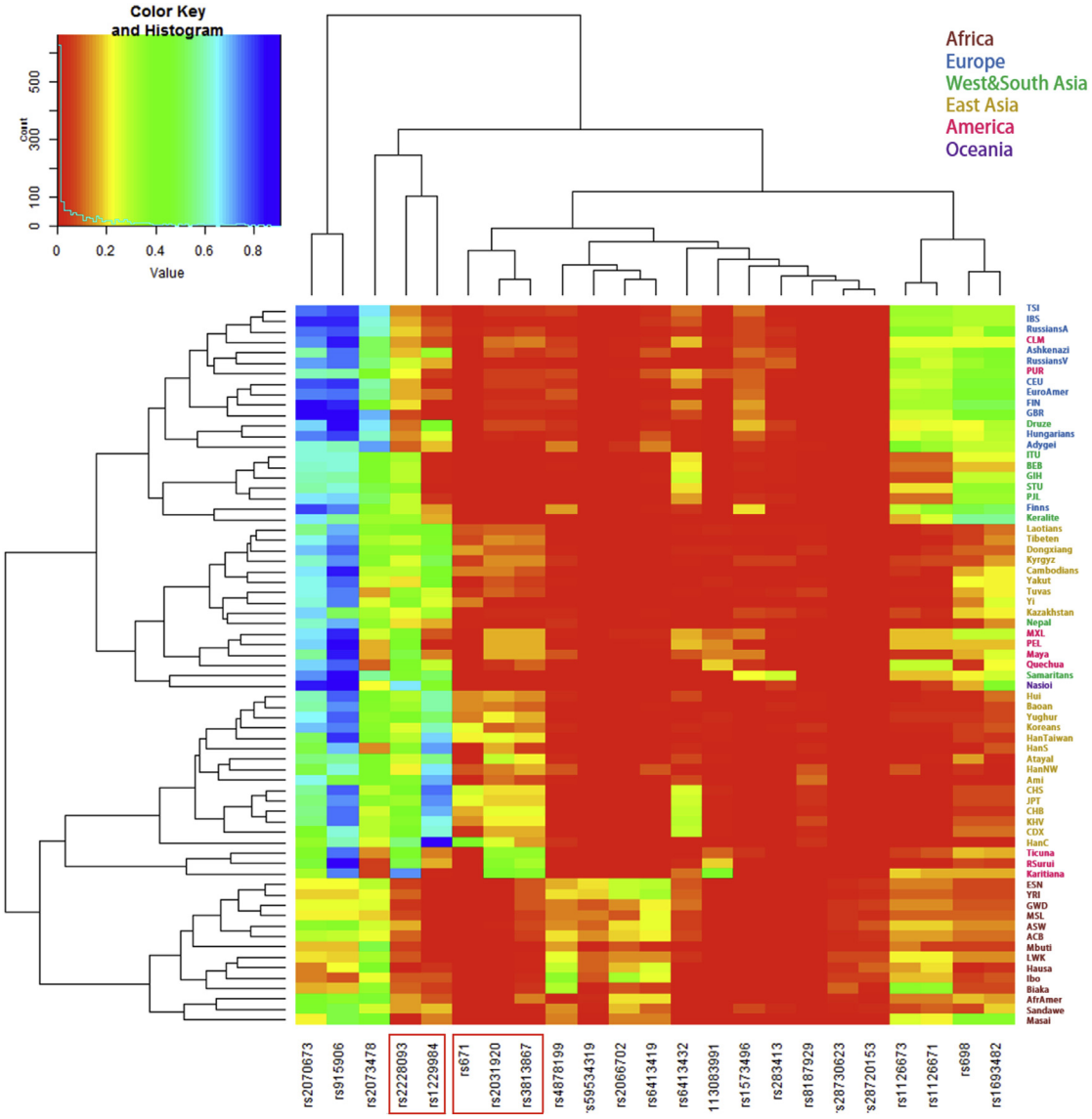

Fig. 1. The heatmap of the clustering of the 69 populations and the 23 functional SNPs. This pattern reflected the geographic clustering of the major populations being studied and indicate the worldwide differentiation of alcohol metabolism. East Asians are mainly clustered in two groups. SNPs in the red boxes are close related with ADH1B*47His (rs1229984) and ALDH2*504Lys (rs671) respectively according to the K-means clustering. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The similarities/differences in the allele frequency patterns across all populations (Fig. 1) and 1000 Genomes populations can be separately quantified as correlation coefficients between each SNP pair (Table S1). Table 1 shows the values of the 3 close-related SNPs discovered in heatmap with rs1229984 (*ADH1B*) and rs671 (*ALDH2*) across 1000 Genomes populations. In addition to the 3 close-related SNPs, a new SNP rs8187929 (*ALDH1A1*) stood out with the highest coefficient (0.9566) between rs1229984 and the second-highest coefficient (0.8555) between rs671. The frequency of rs8187929 was detected to be near 0.05 among East Asians while almost absent in any other populations. The scenario is similar in the correlation coefficients matrix of all the populations (Table S1).

Table 1
Pearson product moment correlation of allele frequencies across 1000 Genomes populations. Here list the values of the 3 close-related SNPs: rs2228093 (ALDH1B1), rs3813867, rs2031920 (CYP2E1), discovered in heatmap with rs1229984 (ADH1B) and rs671 (ALDH2), along with the value of a new detected SNP rs8187929 (Bold).1

	ADH1B rs1229984	*ALDH2* rs671	*ALDH1A1* **rs8187929**	*ALDH1B1* rs2228093	*CYP2E1* rs3813867	*CYP2E1* rs2031920
rs1229984	1.0000					
rs671	0.9112	1.0000				
rs8187929	0.9566	0.8555	1.0000			
rs2228093	0.6175	0.4995	0.5716	1.0000		
rs3813867	0.8130	0.7420	0.7887	0.5267	1.0000	
rs2031920	0.8523	0.7596	0.8121	0.7201	0.9281	1.0000

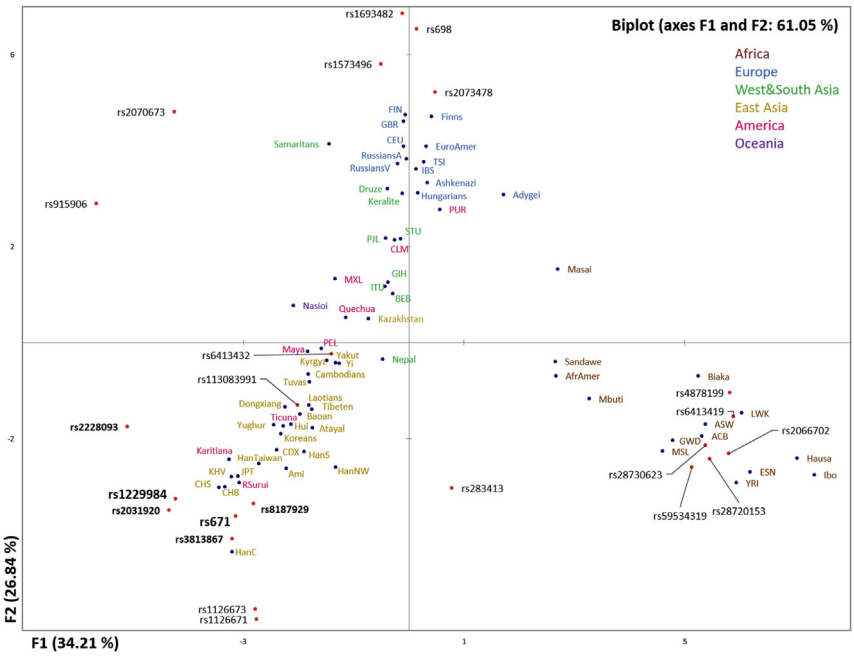

Fig. 2. Principal Component Analysis of the 69 populations and 23 functional SNPs shown in biplot. The first PC accounts for 34.21% of the variance and primarily separates African populations from the rest of the world. The second PC accounts for 26.84% of the variance and primarily separates Europe from East Asia. The dispersion of the SNP reflects their contributions to the distinction of populations.

PCA on the allele frequencies in the populations showed four distinct groupings of populations based on the first two components (Fig. 2): a highly distributed African group, a more tightly clustered East Asian group, a modestly clustered European group, and an intermediate American-Southwest Asian group. This pattern reflected the geographic clustering of the majority of the populations being studied. In the East Asian cluster, Chinese Han, Japanese (JPT), and Vietnamese (KHV) were the most representative populations, while some nomadic populations such as Kazakhstan, Yakut, Kyrgyz, and Tuvas tended to be placed in more intermediate positions. PCA analysis shown in biplot can visually tell the contribution of the SNPs to the distinction of populations. The combine of new detected SNPs rs3813867 (*CYP2E1*), rs2031920 (*CYP2E1*), rs8187929 (*ALDH1A1*), and rs2228093 (*ALDH1B1*), together with rs1229984 (*ADH1B*), rs671 (*ALDH2*), contributed to the major resolution of East Asians. The two *ADH4* SNPs rs1126673, rs1126671 in the bottom of the plot also made contributions. These two SNPs are widespread, but nearly fixed in East Asians, and will not be involved in the following analysis.

Taken together, the PCA analysis validates the results from the heatmap and correlation analysis. Four functional SNPs in *ALDH1A1*, *ALDH1B1*, *CYP2E1* were found to be closely related to rs1229984 (*ADH1B*) and rs671 (*ALDH2*) in East Asia.

3.2. Selection on the genes

To detect the molecular signature of recent selection on the three genes polymorphism, we applied the LRH method (Sabeti et al., 2002) using haplotype data of East Asian populations (CDX: Chinese Dai in Xishuangbanna, China; CHB: Han Chinese in Beijing, China; JPT: Japanese in Tokyo, Japan; KHV: Kinh in Ho Chi Minh City, Vietnam; CHS: Southern Han Chinese, China) from the 1000 Genome Project. We scan the 100 Kb regions covering each gene and detect the decay of LD from the core haplotype. Core haplotype involving the functional SNP (in bold) for each gene is as following, *ALDH1A1*: rs3739962-rs2161811-**rs8187929**; *ALDH1B1*: rs139891405-rs2073477-**rs2228093**; *CYP2E1*: **rs3813867**-**rs2031920**-rs2031921. The SNPs in core haplotypes are all within 1 Kb regions and in high linkage with the functional SNPs, which reflects the true situation of the functional SNPs in selection. We plotted the haplotype-bifurcation diagrams and EHH/REHH test results in Fig. 3. In *ALDH1A1* and *ALDH1B1*, three types of core haplotypes were found, while in *CYP2E1* only two. The core haplotypes covering the derived allele in three genes all show a long-range homozygosity through a slower decline rate of haplotype bifurcation. The EHH and REHH values are plotted against the distance away from the core. Both the EHH and REHH values of the derived allele are much higher than these of the ancestral allele, suggesting that the derived allele is the targets of positive selection. In summary, the distribution pattern of rs8187929 (*ALDH1A1*), rs2228093 (*ALDH1B1*), rs3813867 (*CYP2E1*), and rs2031920 (*CYP2E1*) in the East Asian populations studied cannot be explained by random genetic drift, and recent selection needs to be invoked.

3.3. The time of selection

To see if the initial increase of the three genes polymorphisms in East Asia occurred during the same period as rs1229984 (*ADH1B*) and rs671 (*ALDH2*) did in Neolithic time (Li et al., 2008; Li et al., 2009, 2011; Peng et al., 2010). We estimate the age of the four associated functional SNPs based on the haplotype data of East Asian populations from the 1000 Genome Project. With the fine-scale genetic map, we selected series of contiguous polymorphic SNPs over the three genes in a density of 1SNP/3 Kb to estimate the age (Table 2). As the method based on the moments estimator (Slatkin and Rannala, 2000) is not suitable for the region of low average recombination rates, Regions with <0.1 cM/Mb average recombination rate were excluded. Therefore, the age estimated based on the whole gene regions reflect the real age of allele. The time for rs8187929 (*ALDH1A1*) is 8608 (±1338) YBP, the time for rs2228093 (*ALDH1B1*) is 10,937 (±7461) YBP, and the time for rs3813867 (*CYP2E1*) is 4682 (±2004) YBP. Because rs3813867 and

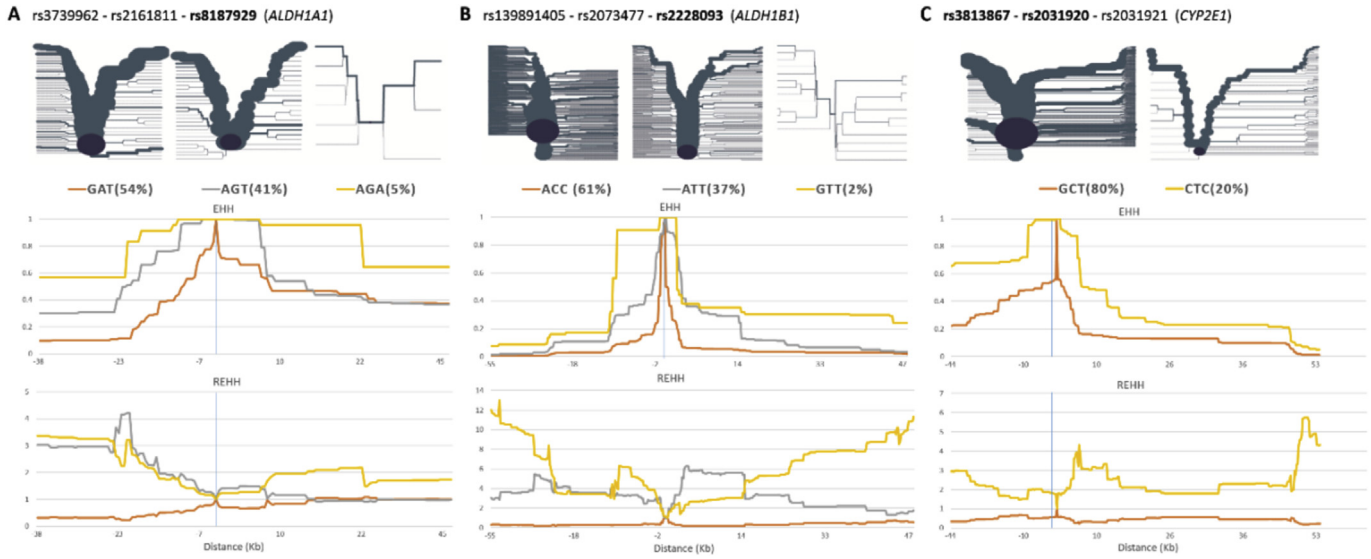

Fig. 3. The haplotype-bifurcation diagrams and EHH/REHH test for the three core haplotypes covering the functional SNP (in bold) in the East Asian populations. (A) *ALDH1A1*: rs3739962-rs2161811-**rs8187929** (B) *ALDH1B1*: rs139891405-rs2073477-**rs2228093** (C) *CYP2E1*: **rs3813867**-**rs2031920**-rs2031921. Decline rate of haplotype bifurcation is much slower for the derived allele than the ancestral allele, which suggests a long-range homozygosity. The EHH and REHH values are plotted against the physical distances extending both upstream and downstream of the selected core region. EHH and REHH values of the derived allele are much higher than these of the ancestral allele, suggesting that the derived allele is the targets of positive selection.

rs2031920 are highly linked and are in very near physical distance, we only calculate the age for rs3813867 to represent the other SNP. Taken together, the age of the derived allele in these three genes falls in the range of 11,000–4600 year before present.

4. Discussion

Our analyses revealed that ethnic populations with different life styles have different evolutionary histories of alcohol metabolic

Table 2
Allele ages estimation for rs8187929 (*ALDH1A1*), rs2228093 (*ALDH1B1*), and rs3813867 (*CYP2E1*).

rs8187929 (*ALDH1A1*)					
Polymorphism	Distance to rs8187929 (bp)	Genetic distance to rs8187929 (cM)	SNP frequency (%)		Allele age (years)
			T	A	
rs10869198	−32,734	0.055186	56.24	93.48	7308
rs3764435	−23,628	0.051667	56.34	93.48	7826
rs4646547	−22,922	0.047571	56.13	93.48	8457
rs348471	−20,529	0.013846	45.53	97.83	7353
rs8187981	−18,566	0.011928	41.79	97.83	7977
rs10781106	−17,582	0.011031	45.74	97.83	9266
rs63319	−15,720	0.009262	45.43	97.83	10,972
rs2210103	11,629	0.009009	44.70	97.83	11,129
rs1330286	12,449	0.009523	40.44	97.83	9760
rs348452	12,881	0.009700	44.70	97.83	10,336
rs7027604	14,448	0.010238	38.36	97.83	8766
rs2309779	18,534	0.012493	44.80	97.83	8040
rs3909559	19,065	0.012493	40.64	97.83	7466
rs647880	19,314	0.013045	44.80	97.83	7700
rs595958	20,370	0.013766	40.54	97.83	6764
Average					8608 (SD = 1338)
rs2228093 (*ALDH1B1*)					
Polymorphism	Distance to rs8187929 (bp)	Genetic distance to rs8187929 (cM)	SNP frequency (%)		Allele age (years)
			C	T	
rs12554425	−31,858	−0.210269	0.86	0.92	10,225
rs12352220	−28,632	−0.208904	0.91	0.97	4808
rs911755	−24,568	−0.171417	0.51	0.58	29,229
rs1923421	−22,172	−0.127087	0.85	0.91	16,095
rs4878797	−15,652	−0.112405	0.72	0.85	17,068
rs28759073	−14,714	−0.098640	0.70	0.84	18,484
rs7848268	−13,662	−0.074792	0.58	0.79	23,778
rs1543495	−12,654	−0.057012	0.80	0.96	9839
rs73646178	−8105	−0.035104	0.88	0.98	9533

(*continued on next page*)

Table 2 (*continued*)

rs8187929 (*ALDH1A1*)					
Polymorphism	Distance to rs8187929 (bp)	Genetic distance to rs8187929 (cM)	SNP frequency (%)		Allele age (years)
			T	A	
rs4878802	−4784	−0.020797	0.66	0.94	22,407
rs7853954	3719	0.058621	0.86	0.99	4063
rs77795946	4381	0.064744	0.85	0.99	2685
rs12555438	6827	0.077016	0.88	0.97	10,616
rs56673731	9235	0.081893	0.85	0.99	2647
rs4878811	12,523	0.086054	0.86	0.99	2115
rs74562401	16,480	0.090330	0.86	0.98	3080
rs4878812	19,006	0.093436	0.82	0.97	4794
rs7029468	23,899	0.100201	0.35	0.70	15,105
rs7039990	27824	0.104359	0.65	0.84	14,626
rs200332009	30,453	0.106432	0.90	0.97	6859
rs10973799	31,811	0.107518	0.27	0.61	17,673
rs77012313	33,880	0.109198	0.81	0.96	5101
rs1028627	37,465	0.112834	0.81	0.96	4937
rs12551065	43,353	0.117108	0.90	0.97	6710
Average					10,937 (SD = 7461)
rs3813867 (*CYP2E1*)					
Polymorphism	Distance to rs8187929 (bp)	Genetic distance to rs8187929 (cM)	SNP frequency (%)		Allele age (years)
			G	C	
rs11101794	−31,938	0.020600	27.99	98.04	3350
rs4543900	−28,797	0.020254	27.99	98.04	3407
rs1854458	−25,790	0.019922	35.82	98.04	3893
rs9418982	−18,983	0.019152	35.95	98.04	4058
rs9418984	−17,249	0.017140	35.95	98.04	4534
rs6537609	−16,521	0.015881	36.07	98.04	4903
rs7911139	−16,080	0.015561	35.95	98.04	4994
rs7077457	−15,235	0.014955	35.95	98.04	5197
rs9418987	−15,196	0.014928	35.82	98.04	5196
rs9418988	−11,973	0.012643	33.46	98.53	4419
rs7095379	−11,073	0.011284	33.46	98.53	4951
rs2026047	−9989	0.008585	33.21	98.53	6483
rs751946	−9501	0.007768	7.96	98.53	5183
rs2070674	5735	0.005144	1.49	98.53	7310
rs2864987	8974	0.016181	9.33	98.04	3377
rs743534	9621	0.017035	74.50	98.04	11,742
rs8192780	14,520	0.024847	33.96	98.04	3032
rs10857741	18,963	0.028235	9.33	98.04	1935
rs1581935	20,843	0.029124	75.00	98.53	5203
rs11593189	25,540	0.029979	9.08	96.57	3208
rs5029271	27,454	0.030665	9.08	98.04	1777
rs2987791	30,455	0.031120	74.88	98.53	4845
Average					4682 (SD = 2004)

system. In East Asia, long-term farming ethnic groups such as Han, differ from the nomadic populations in the pattern of alcohol-related genetic polymorphism. This divergence was mainly attributed to the 6 close related functional SNPs, namely, 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 coincide with the Neolithic time. Therefore, we suggest that the ethnic-related distribution of the alcohol genetic polymorphism in East Asia was due to the emergence and expansion of agriculture in Neolithic time. Naturally, the next question is to explain the impact of the farming-related polymorphism pattern or how these gene variations help human adapt to the new artificial environment of agriculture.

Alcohol is unique among substance abuse drugs, as it is a natural by-product of fermentation. Alcohol consumption dates back to the Neolithic period, and alcoholic beverages may even have predated bread as a staple food (Hanson, 1995). Rice has been used as the material to produce fermented food and beverages for a long time in China since early Neolithic time (McGovern et al., 2004). However, ingestion of ethanol will induce a series of physiological effects on body, including euphoria, flushed skin, and decreased social inhibition at lower doses, with larger doses producing progressively severe impairments of balance, muscle coordination (ataxia), and decision-making ability (potentially leading to violent or erratic behavior) as well as nausea or vomiting from disruptive effect of alcohol on the semicircular canals of the inner ear and chemical irritation of the gastric mucosa. Sufficiently high levels of blood-borne alcohol will cause coma and death from the depressive effects of alcohol upon the central nervous system (Olive et al., 2001; Ridderinkhof et al., 2002; Watanabe et al., 2004; Gu et al., 2005; Goral et al., 2008; Miller and Gold, 2011). Therefore, it is reasonable to see that genetic polymorphisms involved in the ethanol metabolic pathway be the targets of selection. In addition to the proved functional SNP ADH1B*47His (rs1229984) and ALDH2*504Lys (rs671), the new detected putative functional SNPs rs8187929, rs2228093, rs3813867 and rs2031920 in this paper belong to three genes, *ALDH1A1*, *ALDH1B1*, and *CYP2E1*.

ALDH1A1, also known as *ALDH1*, is of cytosolic origin, associated with a low Km for NAD and a high Km for acetaldehyde, and strongly inactivated by disulfiram, while *ALDH2* is of mitochondrial origin, associated with a high Km for NAD and a low Km for acetaldehyde, and insensitive to disulfiram. For individuals with

deficient *ALDH2*, *ALDH1A1* isozyme plays a major part in acetaldehyde metabolism (Hsu et al., 1985). Up to now, research on rs8187929 (*ALDH1A1*) is limited.

ALDH1B1, also known as *ALDH5* or *ALDHX*, is another important kind of aldehyde dehydrogenases, encoding a deduced 517-amino acid protein that shares 70.6% and 62.8% sequence identity with the *ALDH2* and *ALDH1* proteins, respectively (Hiraoka et al., 1995). The rs2228093 (*ALDH1B1*) was found to be significantly associated with alcohol-induced hypersensitivity among Scandinavians (Linneberg et al., 2010).

Unlike ADH pathway, *CYP2E1* displays a special characteristic in ethanol metabolism, that is, its inducibility (Lieber, 1997). Previous studies have shown that the activity of *CYP2E1* is increased up to 10-fold after chronic alcohol consumption. Therefore, while the largest part of ingested ethanol is normally metabolized by *ADH*, *CYP2E1* is significantly involved in ethanol metabolism at high ethanol concentrations or after long-term ethanol intake (Lieber and DeCarli, 1968; Lieber et al., 1988; Tsutsumi et al., 1989; Takahashi et al., 1993; Lieber, 1999). The abundance of expression of *CYP2E1* in liver and extrahepatic tissues holds importance keeping in view its role in generating oxidative stress which contribute to liver injury due to alcohol consumption (Cederbaum, 1990; Joshi and Tyndale, 2006). Further, due to its ability to modulate the effects of drugs, *CYP2E1* plays a crucial role in drug metabolism (Lieber, 2004; Joshi and Tyndale, 2006).

The polymorphism rs3813867 (-1293G-C transversion) and rs2031920 (-1053C-T transition) in the 5-prime transcriptional regulatory region of the *CYP2E1* gene, is referred to as CYP2E1*5B, associated with increased transcription of *CYP2E1* (Wang et al., 2009). Study in the Japanese found that Individuals with the derived homozygous genotype at the CYP2E1*5B could consume more ethanol on average than those with the ancestral homozygous genotype in subjects for the *ALDH2* (rs671) ancestral homozygous genotype, suggesting an interactive effect between *ADLH2* and *CYP2E1* on alcohol consumption (Tanaka et al., 1997; Sun et al., 1999).

There may exist complex interactions among these genes, compensatory or synergistic, jointly regulating the metabolism of ethanol, acetaldehyde and other related metabolites in human body. Our findings reflect one side of the interactions, yet the real functional mechanism of these polymorphisms in different populations needs further studies.

Acknowledgements

This work was supported by NSFC (91131002), NSFC Excellent Youth Funding (31222030), Shanghai Rising-Star Program (12QA1400300), Shanghai Shuguang Project (14SG05), MOE University Doctoral Research Supervisor's Funds (20120071110021), and MOE Scientific Research Project (113022A).

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.quaint.2016.03.008.

References

Armelagos, G.J., Harper, K.N., 2005. Genomics at the origins of agriculture, part one. Evolutionary Anthropology: Issues, News, and Reviews 14 (2), 68–77.
Bersaglieri, T., Sabeti, P.C., Patterson, N., Vanderploeg, T., Schaffner, S.F., Drake, J.A., Rhodes, M., Reich, D.E., Hirschhorn, J.N., 2004. Genetic signatures of strong recent positive selection at the lactase gene. The American Journal of Human Genetics 74 (6), 1111–1120.
Bocquet-Appel, J., 2011. When the world's population took off: the springboard of the Neolithic Demographic Transition. Science 333 (6042), 560–561.
Browning, S.R., Browning, B.L., 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. The American Journal of Human Genetics 81 (5), 1084–1097.
Cederbaum, A.I., 1990. Microsomal generation of reactive oxygen species and their possible role in alcohol hepatotoxicity. Alcohol and Alcoholism (Oxford, Oxfordshire). Supplement 1, 291–296.
Gong, Z., Chen, H., Yuan, D., Zhao, Y., Wu, Y., Zhang, G., 2007. The temporal and spatial distribution of ancient rice in China and its implications. Chinese Science Bulletin 52 (8), 1071–1079.
Goral, J., Karavitis, J., Kovacs, E.J., 2008. Exposure-dependent effects of ethanol on the innate immune system. Alcohol 42 (4), 237–247.
Gross, B.L., Zhao, Z., 2014. Archaeological and genetic insights into the origins of domesticated rice. Proceedings of the National Academy of Sciences 111 (17), 6190–6197.
Gu, J.W., Bailey, A.P., Sartin, A., Makey, I., Brady, A.L., 2005. Ethanol stimulates tumor progression and expression of vascular endothelial growth factor in chick embryos. Cancer 103 (2), 422–431.
Hanson, D.J., 1995. Preventing Alcohol Abuse: Alcohol, Culture, and Control. Greenwood Publishing Group.
Harada, S., Okubo, T., Nakamura, T., Fujii, C., Nomura, F., Higuchi, S., Tsutsumi, M., 1999. A novel polymorphism (-357 G/A) of the ALDH2 gene: linkage disequilibrium and an association with alcoholism. Alcoholism: Clinical and Experimental Research 23 (6), 958–962.
Hawks, J., Wang, E.T., Cochran, G.M., Harpending, H.C., Moyzis, R.K., 2007. Recent acceleration of human adaptive evolution. Proceedings of the National Academy of Sciences 104 (52), 20753–20758.
Hiraoka, L.R., Hsu, L., Hsieh, C., 1995. Assignment of ALDH3 to human chromosome 17p11. 2 and ALDH5 to human chromosome 9p13. Genomics 25 (1), 323–325.
Hsu, L.C., Tani, K., Fujiyoshi, T., Kurachi, K., Yoshida, A., 1985. Cloning of cDNAs for human aldehyde dehydrogenases 1 and 2. Proceedings of the National Academy of Sciences 82 (11), 3771–3775.
Joshi, M., Tyndale, R.F., 2006. Regional and cellular distribution of CYP2E1 in monkey brain and its induction by chronic nicotine. Neuropharmacology 50 (5), 568–575.
Li, H., Borinskaya, S., Yoshimura, K., Kal Ina, N., Marusin, A., Stepanov, V.A., Qin, Z., Khaliq, S., Lee, M.Y., Yang, Y., 2009. Refined geographic distribution of the oriental ALDH2* 504Lys (nee 487Lys) variant. Annals of Human Genetics 73 (3), 335–345.
Li, H., Gu, S., Cai, X., Speed, W.C., Pakstis, A.J., Golub, E.I., Kidd, J.R., Kidd, K.K., 2008. Ethnic related selection for an ADH class I variant within East Asia. PLoS One 3 (4), e1881.
Li, H., Gu, S., Han, Y., Xu, Z., Pakstis, A.J., Jin, L., Kidd, J.R., Kidd, K.K., 2011. Diversification of the ADH1B gene during expansion of modern humans. Annals of Human Genetics 75 (4), 497–507.
Li, T.K., Yin, S.J., Crabb, D.W., O'Connor, S., Ramchandani, V.A., 2001. Genetic and environmental influences on alcohol metabolism in humans. Alcoholism: Clinical and Experimental Research 25 (1), 136–144.
Lieber, C.S., 1997. Ethanol metabolism, cirrhosis and alcoholism. Clinica Chimica Acta 257 (1), 59–84.
Lieber, C.S., 1999. Microsomal ethanol-oxidizing system (MEOS): the first 30 years (1968-1998)–a review. Alcoholism: Clinical and Experimental Research 23 (6), 991–1007.
Lieber, C.S., 2004. The discovery of the microsomal ethanol oxidizing system and its physiologic and pathologic role. Drug Metabolism Reviews 36 (3–4), 511–529.
Lieber, C.S., DeCarli, L.M., 1968. Ethanol oxidation by hepatic microsomes: adaptive increase after ethanol feeding. Science 162 (3856), 917–918.
Lieber, C.S., Lasker, J.M., DeCarli, L.M., Saeli, J., Wojtowicz, T., 1988. Role of acetone, dietary fat and total energy intake in induction of hepatic microsomal ethanol oxidizing system. Journal of Pharmacology and Experimental Therapeutics 247 (2), 791–795.
Linneberg, A., Gonzalez Quintela, A., Vidal, C., Jørgensen, T., Fenger, M., Hansen, T., Pedersen, O., Husemoen, L., 2010. Genetic determinants of both ethanol and acetaldehyde metabolism influence alcohol hypersensitivity and drinking behaviour among Scandinavians. Clinical & Experimental Allergy 40 (1), 123–130.
Lu, R.B., Ko, H.C., Lee, J.F., Lin, W.W., Huang, S.Y., Wang, T.J., Wu, Y.S., Lu, T.E., Chou, Y.H., 2005. No alcoholism-protection effects of ADH1B* 2 allele in antisocial alcoholics among Han Chinese in Taiwan. Alcoholism: Clinical and Experimental Research 29 (12), 2101–2107.
McGovern, P.E., Zhang, J., Tang, J., Zhang, Z., Hall, G.R., Moreau, R.A., Nuñez, A., Butrym, E.D., Richards, M.P., Wang, C., 2004. Fermented beverages of pre-and proto-historic China. Proceedings of the National Academy of Sciences of the United States of America 101 (51), 17593–17598.
McNeill, W., 2010. Plagues and Peoples. Anchor.
Miller, N.S., Gold, M.S., 2011. Addictive Disorders in Medical Populations. John Wiley & Sons.
Olive, M.F., Koenig, H.N., Nannini, M.A., Hodge, C.W., 2001. Stimulation of endorphin neurotransmission in the nucleus accumbens by ethanol, cocaine, and amphetamine. The Journal of Neuroscience 21 (23), C184.
Oota, H., Pakstis, A.J., Bonne Tamir, B., Goldman, D., Grigorenko, E., Kajuna, S.L., Karoma, N.J., Kungulilo, S., Lu, R.B., Odunsi, K., 2004. The evolution and population genetics of the ALDH2 locus: random genetic drift, selection, and low levels of recombination. Annals of Human Genetics 68 (2), 93–109.

Osier, M., Pakstis, A.J., Kidd, J.R., Lee, J., Yin, S., Ko, H., Edenberg, H.J., Lu, R., Kidd, K.K., 1999. Linkage disequilibrium at the ADH2 and ADH3 loci and risk of alcoholism. The American Journal of Human Genetics 64 (4), 1147–1157.
Osier, M.V., Pakstis, A.J., Goldman, D., Edenberg, H.J., Kidd, J.R., Kidd, K.K., 2002. A proline-threonine substitution in codon 351 of ADH1C is common in Native Americans. Alcoholism: Clinical and Experimental Research 26 (12), 1759–1763.
Peng, Y., Shi, H., Qi, X., Xiao, C., Zhong, H., Run-lin, Z.M., Su, B., 2010. The ADH1B Arg47His polymorphism in East Asian populations and expansion of rice domestication in history. BMC Evolutionary Biology 10 (1), 15.
Ridderinkhof, K.R., de Vlugt, Y., Bramlage, A., Spaan, M., Elton, M., Snel, J., Band, G.P., 2002. Alcohol consumption impairs detection of performance errors in mediofrontal cortex. Science 298 (5601), 2209–2211.
Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z., Richter, D.J., Schaffner, S.F., Gabriel, S.B., Platko, J.V., Patterson, N.J., McDonald, G.J., 2002. Detecting recent positive selection in the human genome from haplotype structure. Nature 419 (6909), 832–837.
Slatkin, M., Rannala, B., 2000. Estimating allele age. Annual Review of Genomics and Human Genetics 1 (1), 225–249.
Sun, F., Tsuritani, I., Honda, R., Ma, Z., Yamada, Y., 1999. Association of genetic polymorphisms of alcohol-metabolizing enzymes with excessive alcohol consumption in Japanese men. Human Genetics 105 (4), 295–300.
Takahashi, T., Lasker, J.M., Rosman, A.S., Lieber, C.S., 1993. Induction of cytochrome P-4502E1 in the human liver by ethanol is caused by a corresponding increase in encoding messenger RNA. Hepatology 17 (2), 236–245.
Tanaka, F., Shiratori, Y., Yokosuka, O., Imazeki, F., Tsukada, Y., Omata, M., 1997. Polymorphism of alcohol-metabolizing genes affects drinking behavior and alcoholic liver disease in Japanese men. Alcoholism: Clinical and Experimental Research 21 (4), 596–601.
Tsutsumi, M., Lasker, J.M., Shimizu, M., Rosman, A.S., Lieber, C.S., 1989. The intralobular distribution of ethanol-inducible P450IIE1 in rat and human liver. Hepatology 10 (4), 437–446.
Voight, B.F., Kudaravalli, S., Wen, X., Pritchard, J.K., 2006. A map of recent positive selection in the human genome. PLoS Biology 4 (3), 446.
Wang, E., Ding, Y., Flodman, P., Kidd, J.R., Kidd, K.K., Grady, D.L., Ryder, O.A., Spence, M.A., Swanson, J.M., Moyzis, R.K., 2004. The genetic architecture of selection at the human dopamine receptor D4 (DRD4) gene locus. The American Journal of Human Genetics 74 (5), 931–944.
Wang, E.T., Kodama, G., Baldi, P., Moyzis, R.K., 2006. Global landscape of recent inferred Darwinian selection for Homo sapiens. Proceedings of the National Academy of Sciences of the United States of America 103 (1), 135–140.
Wang, S., Zhu, A., Li, D., Wang, Z., Zhang, P., Zhang, G., 2009. Frequencies of genotypes and alleles of the functional SNPs in CYP2C19 and CYP2E1 in mainland Chinese Kazakh, Uygur and Han populations. Journal of Human Genetics 54 (6), 372–375.
Watanabe, H., Yamazaki, M., Miyazaki, H., Arikawa, C., Itoh, K., Sasaki, T., Maehama, T., Frohman, M.A., Kanaho, Y., 2004. Phospholipase D2 functions as a downstream signaling molecule of MAP kinase pathway in L1-stimulated neurite outgrowth of cerebellar granule neurons. Journal of Neurochemistry 89 (1), 142–151.
Yang, X., Wan, Z., Perry, L., Lu, H., Wang, Q., Zhao, C., Li, J., Xie, F., Yu, J., Cui, T., 2012. Early millet use in northern China. Proceedings of the National Academy of Sciences 109 (10), 3726–3730.