Annals of human genetics

doi: 10.1111/j.1469-1809.2009.00517.x

Refined Geographic Distribution of the Oriental *ALDH2*504Lys* (nee *487Lys*) Variant

Hui Li[1,2], Svetlana Borinskaya[3], Kimio Yoshimura[4,†], Nina Kal'ina[3], Andrey Marusin[5], Vadim A. Stepanov[5], Zhendong Qin[2], Shagufta Khaliq[6], Mi-Young Lee[1,‡], Yajun Yang[2], Aisha Mohyuddin[7], David Gurwitz[8], Syed Qasim Mehdi[6], Evgeny Rogaev[3,9,10], Li Jin[2], Nikolay K. Yankovsky[3], Judith R. Kidd[1] and Kenneth K. Kidd[1,*]

[1]*Department of Genetics, School of Medicine, Yale University, New Haven 06520-8005 USA*
[2]*MOE Key Laboratory of Contemporary Anthropology and State Key Laboratory of Genetic Engineering, School of Life Sciences and Institutes of Biomedical Sciences, Fudan University, Shanghai 200433 China*
[3]*Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow 119991 Russia*
[4]*Genetics Division, National Cancer Center Research Institute, Tokyo 104–0045 Japan*
[5]*Institute of Medical Genetics, Tomsk Research Center, Russian Academy of Medical Sciences, Tomsk 634050 Russia*
[6]*Sindh Institute of Urology and Transplantation (SIUT), Karachi 74200 Pakistan*
[7]*Biomedical and Genetic Engineering Laboratories; Shifa College of Medicine, Section of Biochemistry H-8/4, Islamabad 44000 Pakistan*
[8]*National Laboratory for the Genetics of Israeli Populations, Department of Human Molecular Genetics and Biochemistry, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv 69978 Israel*
[9]*Brudnick Neuropsychiatric Research Institute, Department of Psychiatry, University of Massachusetts Medical School, Worcester, Massachusetts 01604 USA*
[10]*Research Center of Mental Health, Russian Academy of Medical Sciences, Moscow 113152 Russia*

Summary

Mitochondrial aldehyde dehydrogenase (ALDH2) is one of the most important enzymes in human alcohol metabolism. The oriental *ALDH2*504Lys* variant functions as a dominant negative, greatly reducing activity in heterozygotes and abolishing activity in homozygotes. This allele is associated with serious disorders such as alcohol liver disease, late onset Alzheimer disease, colorectal cancer, and esophageal cancer, and is best known for protection against alcoholism. Many hundreds of papers in various languages have been published on this variant, providing allele frequency data for many different populations. To develop a highly refined global geographic distribution of *ALDH2*504Lys*, we have collected new data on 4,091 individuals from 86 population samples and assembled published data on a total of 80,691 individuals from 366 population samples. The allele is essentially absent in all parts of the world except East Asia. The *ALDH2*504Lys* allele has its highest frequency in Southeast China, and occurs in most areas of China, Japan, Korea, Mongolia, and Indochina with frequencies gradually declining radially from Southeast China. As the indigenous populations in South China have much lower frequencies than the southern Han migrants from Central China, we conclude that *ALDH2*504Lys* was carried by Han Chinese as they spread throughout East Asia. Esophageal cancer, with its highest incidence in East Asia, may be associated with *ALDH2*504Lys* because of a toxic effect of increased acetaldehyde in the tissue where ingested ethanol has its highest concentration. While the distributions of esophageal cancer and *ALDH2*504Lys* do not precisely correlate, that does not disprove the hypothesis. In general the study of fine scale geographic distributions of *ALDH2*504Lys* and diseases may help in understanding the multiple relationships among genes, diseases, environments, and cultures.

Keywords: East Asia, aldehyde dehydrogenase 2, alcohol associated, allele frequency, esophageal cancer

[†]Current address: Department of Health Policy and Management, School of Medicine, Keio University, Tokyo 160–8582 Japan
[‡]Current address: Department of Preventive Medicine, School of Medicine, Keimyung University, Daegu 700–712 Republic of Korea

*Corresponding author: Kenneth K. Kidd, Department of Genetics, 333 Cedar Street, New Haven, CT 06520-8005. Tel: 203-785 2654; Fax: 203-785 6568; E-mail: Kenneth.Kidd@Yale.edu

Introduction

Alcoholism is a major public health problem globally. Many genes have allele frequency variation that has been associated with risk of developing either alcoholism or complications of alcoholism. The ethanol metabolizing genes, especially the alcohol dehydrogenase (*ADH*) genes and the aldehyde dehydrogenase 2 (*ALDH2*) gene, are the strongest such associations (Couzigou et al., 1994; Long et al., 1998; Reich et al., 1998; Chen et al., 1999; Luo et al., 2006). The major metabolic pathway for ethanol is degradation by ADH enzymes to acetaldehyde followed by degradation of that intermediate metabolite to acetate by ALDH enzymes. The mitochondrial ALDH2, encoded by the *ALDH2* gene on chromosome 12, has the lowest Km (~1 μmol/L) for acetaldehyde (Algar & Holmes, 1989). The variants at ADH genes and at *ALDH2* that are associated with alcoholism appear to interact by increasing the transient levels of the toxic acetaldehyde.

The geographic distribution of the most relevant *ALDH2* variant, *ALDH2* 504Lys*, is quite dramatic being present only in East Asian populations with frequencies as high as 40% in some East Asian population samples. Flushing and discomfort, such as headache and nausea, occur in many individuals of East Asian ancestry after drinking even small amounts of ethanol; these symptoms do not occur in individuals of European ancestry after drinking equivalent, and even larger, amounts of ethanol (Wolff, 1972). That early observation led to the conclusion that alcohol metabolism is quite different between these populations. Early in 1975 the atypical form of ALDH2 isozymes was found (Samatoyannopoulos et al., 1975; Greenfield & Pietruszko, 1977; Hempel & Pietruszko, 1978) including the allele designated *ALDH2* 2* (*ALDH2* 504Lys*). The difference was identified as a substitution of the Glutamic acid at codon position 504 with Lysine (hence the reference to the variant allele as *ALDH2* 504Lys,* Ikawa et al., 1983). Position 487 of the mature protein is actually codon position 504 of the gene when the mitochondrial leader sequence is included; hence the change in this genomic era from "487" to "504" as the referent position. As this atypical form of the enzyme was seen only in East Asians, it was also referred to as the oriental variant (Yoshida et al., 1984), the variant allele acts as a dominant negative with the heterozygote having greatly reduced enzyme activity and the homozygote having no activity; in both cases acetaldehyde accumulates as ethanol is converted to acetaldehyde. The greatly reduced enzyme activity in the heterozygous state reflects the fact that the ALDH2 tetramer is essentially a dimer of a pair of subunits with an important functional interface. If that pair is a heterodimer for the 504 Glutamate and Lysine monomers, there is essentially no activity (Larson et al., 2005).

ALDH2 504Lys* appears to confer protection against some diseases such as alcohol liver disease. The unpleasant sensations experienced by heterozygotes for *ALDH2* 504Lys* (even more so by the homozygotes) after drinking, such as the flushing reaction in many East Asians, hinder/stop further heavy drinking, thus reducing serious harm to the liver and other relevant organs (Yu et al., 2002). However, some social attitudes reduce this protection by promoting drinking in many ethnic groups of East Asia (Hasituya & Su, 2007), resulting in more serious diseases in the *ALDH2* 504Lys* heterozygotes.

The accumulation of acetaldehyde in the body has many consequences in addition to the aversion to consuming ethanol. Li et al. (2006) showed that the Glu504Lys polymorphism was associated with efficacy of sublingual nitroglycerin and recently Chen et al. (2008) showed that *ALDH2* activity is critical for protection from ischemia. These findings emphasize the importance of the studies on genetic variation at *ALDH2*. The 504Lys variant is believed to increase the risk of many disorders, including many cancers. Cancer incidences increase among alcoholics in organs including esophagus, stomach, liver, upper aerodigestive tract in which acetaldehyde is produced by the alcohol dehydrogenases (Yokoyama et al., 2001). Esophageal cancer is of particular interest because studies have shown an increased risk of developing esophageal cancer in *ALDH2* 504Lys* heterozygotes in different East Asian populations (Yokoyama & Omori, 2005; Yokoyama et al., 2006; Yang et al., 2007; Li et al., 2008; Druesne-Pecollo et al., 2009). The geographic distribution of esophageal cancer, with its much higher frequencies in individuals of East Asian ancestry (Parkin et al., 1997), suggests a potential association of this cancer to *ALDH2* 504Lys*. The association is believed to be mediated through levels of acetaldehyde following drinking alcohol. The hypothesis is supported by associations of variants at two different ADH genes (Hashibe et al., 2008).

Materials and Methods

Sampling and Data Collection

A plethora of papers on *ALDH2* exists in the global literature because of its relevance for public health and human population genetics studies. It may be among the more intensively studied human genes. The large number of publications also provides allele frequency data of *ALDH2* 504Lys* (*ALDH2* 2* or ** 487Lys*) in many populations, allowing us to determine the detailed geographic distribution of this allele, with the resulting potential to study the demographic histories of populations and the multiple factors affecting the allele frequency. The allele frequency of *ALDH2* 504Lys* ranges from 0 to 40% among the East Asian populations based on the published data, constituting large variation just within East Asia and belying the common impression that the allele is common in all East Asians. We also find that some key areas or ethnic groups in East Asia have not been studied so far for the

frequencies of *ALDH2* alleles. To assess the detailed distribution of *ALDH2*504Lys*, we collected relevant data from the literature, and filled many blanks on the global map by typing relevant new population samples.

Results and Discussion

In this paper, we present new data on the *ALDH2*504Lys* frequency of 4,091 individuals in 86 populations from China, Laos, Vietnam, Russia, Japan, and other countries around the world (Table 1). Various collaborating laboratories have used different typing methods of either traditional PCR-RFLP (Oota et al., 2004) or Taqman® SNP genotyping assay (C__11703892_10). Added to the data we have extracted from the literature, the total sample size is 80,691 individuals from 366 population samples (see Table S1 and ALFRED online database for detailed data including references to the relevant geographic areas and ethnic groups; http://alfred.med.yale.edu/alfred/SiteTable1A_working.asp?siteuid=SI000734O).

Allele Frequency Map

A refined map of *ALDH2*504Lys* allele frequency was generated from these frequency data using the Surfer 8.0 program to interpolate the clinal patterns (Fig. 1). For some of the populations, more than one sample was studied. The different samples from the same population usually had similar allele frequencies, while some showed notable deviation from the common data. Some of this inconsistency may have resulted from technical problems such as typing errors or sampling bias. Here we chose either the most commonly estimated frequency for each population with multiple estimates, or the data based on the largest sample size. For instance, the frequencies are all around 17% in 11 Korean population samples; therefore, two estimates of 3% and 36% were rejected in constructing our map. Among nine Japanese samples from Tokyo, with the frequency ranging from 21.5% to 29.0%, we only chose the frequency of 26.6% with the largest sample size of 642. In some cases, we prefer random sample data to the control sample data of case-control studies, or recent data by new typing methods to the data published decades ago. All of the data are included in Table S1 with indication of which samples were included in Figure 1.

In total, the map shows a pattern of a single center of expansion within East Asia. The highest frequencies appear in a restricted area in Southeast China, among the Han Chinese in south Fujian province and east Guangdong province (the Hakka and Minnam populations), decreasing gradually to the north and west. Hakka from Changting County in Fujian have the highest frequency, 40.9%. The Hakka population samples from Taiwan and Sichuan also exhibit high frequencies, indicating that Hakka have maintained a high frequency during their migrations. The allele frequencies in other Han Chinese populations range from 9% to 40%, exhibiting a cline clearly decreasing from southeast to northwest, except for two small peaks in Shanghai in East China and Shandong in Central China.

Another high frequency area for the *ALDH2*504Lys* allele is Central Japan with 34.1% in Chiba. However, this high frequency area seems to be an extension from East China. The frequency decreases from around 30% in Honshu to around 10% in Ryukyu and Hokkaido, corresponding well to the migration history of modern Japanese (the descendants of Yayoi People, Hammer et al., 2006). Therefore, it is most probable that the *ALDH2*504Lys* allele in Japan was brought by the early Yayoi migrants from mainland East Asia.

Because the *ALDH2*504Lys* allele reduces activity in heterozygotes, though with a less severe phenotype than homozygous *ALDH2*504Lys* individuals, we have also considered the combined distribution of both homozygotes and heterozygotes (i.e., $2pq + q^2$). Figure 2 shows the distribution of this *ALDH2*504Lys* "carrier" frequency. The high frequency area of the "carriers" is much wider than the high frequency area of the allele, as expected, indicating that more populations may be at risk for the associated disorders.

Central China Origin of *ALDH2*504Lys*

The frequency decline from Southeast China to West and North China is quite smooth. The allele frequencies decrease to less than 20% in Southwest and Central China, and to less than 10% in Manchuria, Mongolia, Xinjiang, and Tibet within the broader region of China. In Central Asia and Siberia, beyond the pronounced genetic influence of Han Chinese, the *ALDH2*504Lys* allele is rare. The allele is also detected in some Iranian populations, which may be explained by diffusion along the Silk Road. We conclude that the spread of *ALDH2*504Lys* to the north and west was concomitant with the expansion of Han Chinese and diffusion of the allele into surrounding populations.

Although the *ALDH2*504Lys* allele frequency reaches a peak in Southeastern Chinese populations, we cannot draw the conclusion that this allele originated there. The population history shows clearly that Hakka and Minnam Chinese presently in Southeast China are descendants of migrants from Central China (Wen et al., 2004). The indigenous populations in South China, such as Hmong-Mien populations (Hmong and She) from the Yangtze River area, and Daic populations (Kam, Laka, Mulam, and Maonan) from the Pearl River area, exhibit much lower frequency of *ALDH2*504Lys*. *ALDH2*504Lys* is almost absent in the aboriginal populations of Hainan and Taiwan, the two largest islands in South China. Therefore, it is unlikely that the Southeast Chinese obtained

Table 1 *ALDH2* 504Lys* frequencies of the population samples typed in this project.

Region	Country	Population	2N	Frequency	Longitude	Latitude
Africa	Ethiopia	Jews[a]	42	0.000	36E	10N
Africa	Somalia	Somali	40	0.000	40–52E	12N-2S
Africa	Tanzania	Masai	40	0.000	29–42E	1N-6S
Africa	Tanzania	Sandawe	80	0.025	35–38E	4–7S
America	Mexico	Maya	102	0.000	88–90W	18–20N
America	Peru	Quechua	46	0.000	71–73W	13–14S
Central Asia	Iran	Iranians	82	0.000	57E	35N
Central Asia	Kazakhstan	Kazakh-Almaty	68	0.014	76E	43N
Central Asia	Kazakhstan	Uigur-Almaty	60	0.017	76E	43N
Central Asia	Tajikistan	Pamiri	68	0.000	71.6E	37.2N
Central Asia	Tajikistan	Tadjik	32	0.031	68.53E	38.29N
Central Asia	Turkmenistan	Turkmen	134	0.000	57E	38N
East Asia	China	Amdo	288	0.073	92–102E	33–35N
East Asia	China	Ava	120	0.025	99.46E	22.74N
East Asia	China	Baima Dee	56	0.071	104.53E	32.41N
East Asia	China	Daur	42	0.107	124E	48.5N
East Asia	China	East Mongle	36	0.036	115.5–125.5E	47.5–51N
East Asia	China	Evenki	16	0.000	124.5E	48.3N
East Asia	China	Han-Changting	60	0.409	116.35E	25.83N
East Asia	China	Han-Anyang	100	0.170	114.2E	36.1N
East Asia	China	Han-Minnam	76	0.297	117E	24N
East Asia	China	Han-Putian	72	0.329	119.06E	25.25N
East Asia	China	Han-Teochow	244	0.357	116.4E	23.2N
East Asia	China	Han-Shanghai	98	0.184	121.37E	31.11N
East Asia	China	Hezhen	48	0.068	134E	48N
East Asia	China	Hlai	108	0.028	109.52E	18.77N
East Asia	China	Hmong-Black	118	0.076	109.80E	27.88N
East Asia	China	Kam	142	0.106	109.59E	25.79N
East Asia	China	Khams-Qamdo	192	0.015	97.10E	31.08N
East Asia	China	Khams-Kangding	44	0.023	101.96E	30.05N
East Asia	China	Khazak-Balikun	72	0.028	93.01E	43.59N
East Asia	China	Korean-Jilin	78	0.171	129.30E	42.54N
East Asia	China	Laka	196	0.189	110.18E	24.13N
East Asia	China	Manchu	50	0.167	124.5E	40.4N
East Asia	China	Mongol-Shilingol	148	0.068	116.07E	43.95N
East Asia	China	Oroqen	12	0.000	126E	51N
East Asia	China	Qiang	54	0.185	103.85E	31.69N
East Asia	China	Shē-Jingning	80	0.154	119.4E	27.5N
East Asia	China	Tibetan-Lhasa	144	0.054	91.1E	29.4N
East Asia	China	Tibetan-Xigazê	144	0.030	88.53E	29.16N
East Asia	China	Tsat	108	0.120	109.27E	18.17N
East Asia	China	Uigur-Toksun	70	0.071	88.66E	42.79N
East Asia	China	Xibe	24	0.050	124E	42.3N
East Asia	China	Zhuang-North	78	0.205	108.28E	23.17N
East Asia	China	Zhuang-South	58	0.207	107.36E	22.42N
East Asia	Japan	Japanese-Akita-Yokote	154	0.149	140.3E	39.2N
East Asia	Japan	Japanese-Iwate-Ninohe	138	0.159	141.1E	40.1N
East Asia	Japan	Japanese-Nagano-Saku	196	0.189	138.3E	36.1N
East Asia	Japan	Japanese-Okinawa-Ishikawa	90	0.111	127.49E	26.25N
East Asia	Japan	Japanese-Tokyo-Katsushika	70	0.271	139.52E	35.45N
East Asia	South Korea	Koreans	102	0.216	126.57E	37.32N
Europe	Hungary	Hungarians	178	0.000	16–23E	48.5–45.5N
Europe	Russia	Russians, Archangelsk	66	0.015	41E	63N

Table 1 Continued

Region	Country	Population	2N	Frequency	Longitude	Latitude
Europe	Russia	Russians-Rostov on Don	96	0.000	39.45N	47.15E
Europe	Russia	Mari	124	0.000	48E	57N
Europe	Russia	Udmurt	168	0.000	52.3E	57.2N
Oceania	American Samoa	Samoans	16	0.000	170.42W	14.17S
Oceania	Papua New Guinea	Papuans	44	0.000	146E	6S
South Asia	India	Keralite	60	0.000	76–77E	10–8N
South Asia	India	Thoti	28	0.000	78E	18N
South Asia	Pakistan	Hazara	58	0.017	70E	33N
South Asia	Pakistan	Mohanna	104	0.000	67–68E	27–25N
South Asia	Pakistan	Pathan	244	0.000	71.3E	34.5N
South Asia	Pakistan	Negroid Makrani	56	0.036	30–41E	10.5–27S
Siberia	Russia	Buryats-UlanUde	120	0.000	107E	51N
Siberia	Russia	Buryats-Kurumkan village	134	0.000	110.18E	54.18N
Siberia	Russia	Buryats-Aginskoye	138	0.014	114.32E	51.06N
Siberia	Russia	Udeghe	134	0.037	134.2E	46.8N
Siberia	Russia	Nanai	24	0.000	132.0E	44.0N
Siberia	Russia	Tuvinians	40	0.025	90.10E	50.55N
Southeast Asia	India	Kachari	36	0.000	94E	27N
Southeast Asia	Laos	Bo	102	0.059	105.09E	18.08N
Southeast Asia	Laos	Hmong-White	112	0.071	103.54E	19.57N
Southeast Asia	Laos	Katu	96	0.000	106.39E	15.36N
Southeast Asia	Laos	Khmu	98	0.010	102.33E	20.27N
Southeast Asia	Laos	Lamet	84	0.000	101.35E	20.50N
Southeast Asia	Laos	Lao	244	0.107	102.37E	17.57N
Southeast Asia	Laos	Laven	92	0.011	106.35E	15.34N
Southeast Asia	Laos	Phunoi	38	0.000	101.05E	21.13N
Southeast Asia	Laos	Saek	114	0.035	104.55E	17.28N
Southeast Asia	Vietnam	Kinh-Hanoi	96	0.179	105.5E	21N
Southeast Asia	Vietnam	Kinh-Hue	84	0.122	107.3E	16.3N
Southeast Asia	Vietnam	Kinh-Saigon	12	0.100	106.4E	10.4N
Southwest Asia	Israel	Samaritans	82	0.000	35.2E	32.2N
Southwest Asia	Israel	Palestinian Arabs[a]	140	0.000	35.2E	31.7N
Southwest Asia	Israel	Ashkenazi Jews[a]	200	0.000	34.7E	32N

[a]These samples were obtained from The National Laboratory for the Genetics of Israeli Populations (NLGIP) at Tel-Aviv University, Israel.

the *ALDH2*504Lys* allele from the indigenous populations. Unlike the gradually decreasing frequency to the north and west, the allele frequency drops sharply to the south. The allele exists at low frequency in Peninsular Southeast Asia, and is rare in the Southeast Asian islands. If this allele originated in the Southeast Chinese populations after they arrived in the present region, the quick expansion of the allele to the north and west cannot be explained. Therefore, we conclude that the *ALDH2*504Lys* allele was most probably carried south by the Han Chinese migrants from Central China, rather than originating in the indigenous populations in the region where it now has the highest frequencies.

Understanding why the present Central China populations exhibit much lower frequency of *ALDH2*504Lys* than the Southeast China populations is crucial in the study of the history of this allele. Both the decrease in Central China and the increase in Southeast China should be accounted for. In the history of China, many Altaic populations moved from the North China to Central China after wars in the 4th, 12th, and 13th centuries which also resulted in the migration of some Chinese populations from Central China to South China. These Altaic populations later merged with the Central Chinese populations after their kingdoms or dynasties ended. The most famous examples are Sienbers (Xianbei, founders of Former Yan Kingdom, Later Yan Kingdom, Western Qin Kingdom, Southern Liang Kingdom and Southern Yan Kingdom of Sixteen Kingdoms Period, and Northern Dynasties), Huns (founders of Han-Zhao Kingdom and Northern Liang Kingdom of Sixteen Kingdoms Period), Khitans (founders of Liao Dynasty), and Jurchens (founders of Jin Dynasty). Those Altaic migrants may have included very few or no individuals carrying the *ALDH2*504Lys* allele

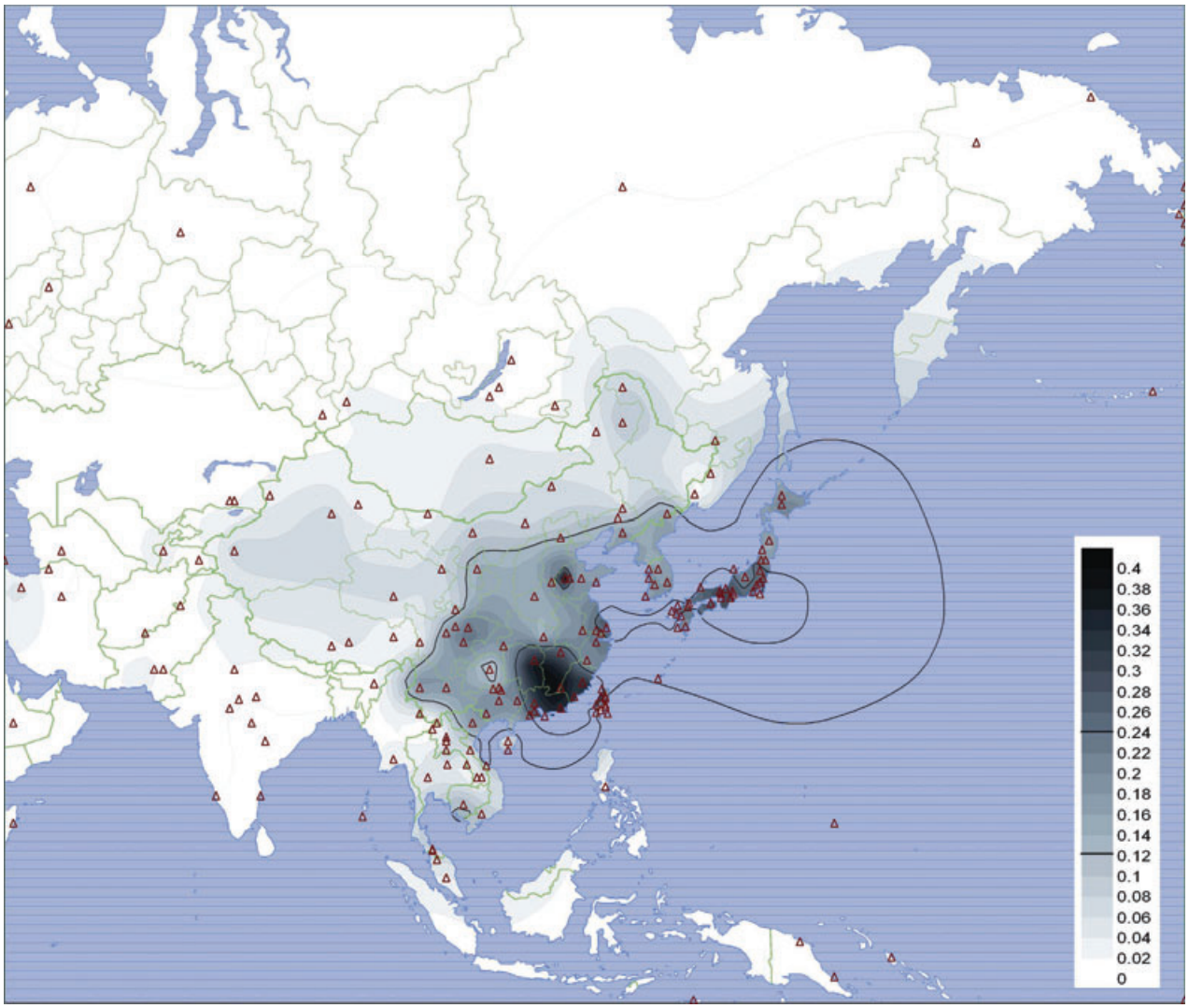

Figure 1 The geographic distribution of *ALDH2* 504Lys* allele frequency. The grey scale refers to the interpolated allele frequency and correspondences are on the right, e.g. 0.12 means an allele frequency of 12% in the region. The open red triangles represent the locations of the population samples. The encircling black lines are the 0.12 and 0.24 frequency borders. A similar format is used in Figure 2 and Figure 3.

because present Altaic populations have a low frequency of the allele. The merging of these Altaic populations could have decreased the proportion of *ALDH2* 504Lys* in the Central Chinese populations. On the other hand, some as yet unknown protective effects of *ALDH2* 504Lys* against diseases might also have contributed to the increased frequency of this allele in Southern Chinese. Since migrations to South China resulted from wars, the refugees may have been subjected to considerable stress and a selective advantage could have had great impact. We can speculate that the *ALDH2* 504Lys* heterozygotes had an advantage because they tended to drink less alcohol or had some other advantage (Chen et al., 1999). The recent appreciation of other metabolic/pharmacologic roles for ALDH2 (Li et al., 2006; Larson et al., 2007; Chen et al., 2008) suggest that if selective factors are responsible for the high ALDH2*2 frequency in East Asia, their nature may be unrelated to the current association with esophageal cancer or ethanol metabolism. Alternative hypotheses of increased resistance to some disease organisms (Goldman & Enoch, 1990; Yokoyama et al., 2001; Oota et al., 2004; Yokoyama & Omori, 2005; Yang et al., 2007; Li et al., 2008) would also explain a clear advantage to heterozygotes. However, statistically positive selection on *ALDH2* 504Lys* cannot be detected using the extended haplotype test (Sabeti et al., 2007) as very low levels of recombination exist in the genomic region of *ALDH2* locus (Oota et al., 2004). Other methods suggest positive selection on *ALDH2* 504Lys* (Long et al., 2006).

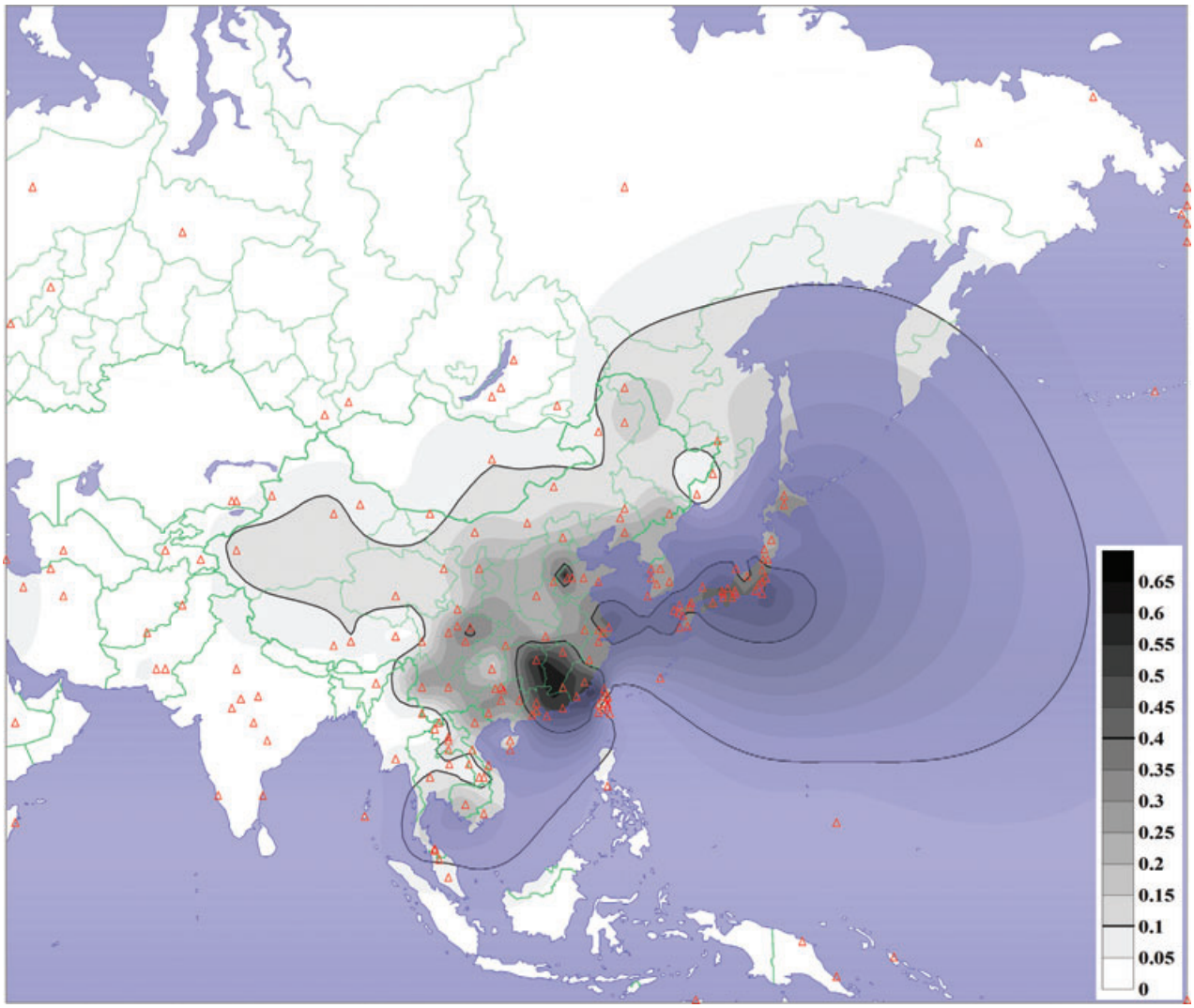

Figure 2 The geographic distribution of *ALDH2* 504Lys* allele carrier frequency. Note the difference scale compared to Figure 1.

Geographic Association with Esophageal Cancer Incidence

Whatever positive selection may have increased the frequency of *ALDH2* 504Lys*, serious diseases such as esophageal cancer or ischemia could act to decrease the *ALDH2* 504Lys* allele frequency among the populations since studies report that heavy alcohol drinkers who are heterozygotes for *ALDH2* 504Lys* have higher risk for esophageal cancer (Yokoyama & Omori, 2005; Yang et al., 2007; Li et al., 2008). In addition, ALDH2 activation was shown to reduce ischemic damage to the heart, suggesting that patients with reduced ALDH2 activity may suffer increased damage during cardiac ischemic events or coronary bypass surgery (Chen et al., 2008). The typical age of onset for esophageal cancer in the high incidence area can be earlier than 30 (He et al., 2006). We compared the geographic distribution of esophageal cancer incidence with the *ALDH2* 504Lys* allele and carrier frequency distributions. We collected the male esophageal cancer incidence data of 355 populations from the literature, covering most countries in the world (Table S2). Central and Southeast China were examined in detail. Figure 3 illustrates the world distribution of esophageal cancer incidence and the details in East Asia. The extremely high incidences only appear in East Asia and some populations in Central Asia where the frequency of *ALDH2* 504Lys* carriers is also high. However, comparison of Figure 2 and Figure 3 shows that the distributions are far from identical. However, the high cancer incidence areas mostly fall into the high frequency area of the derived allele carriers. The acetaldehyde accumulation resulting from *ALDH2* 504Lys* in those who drink alcohol is certainly not the only risk factor

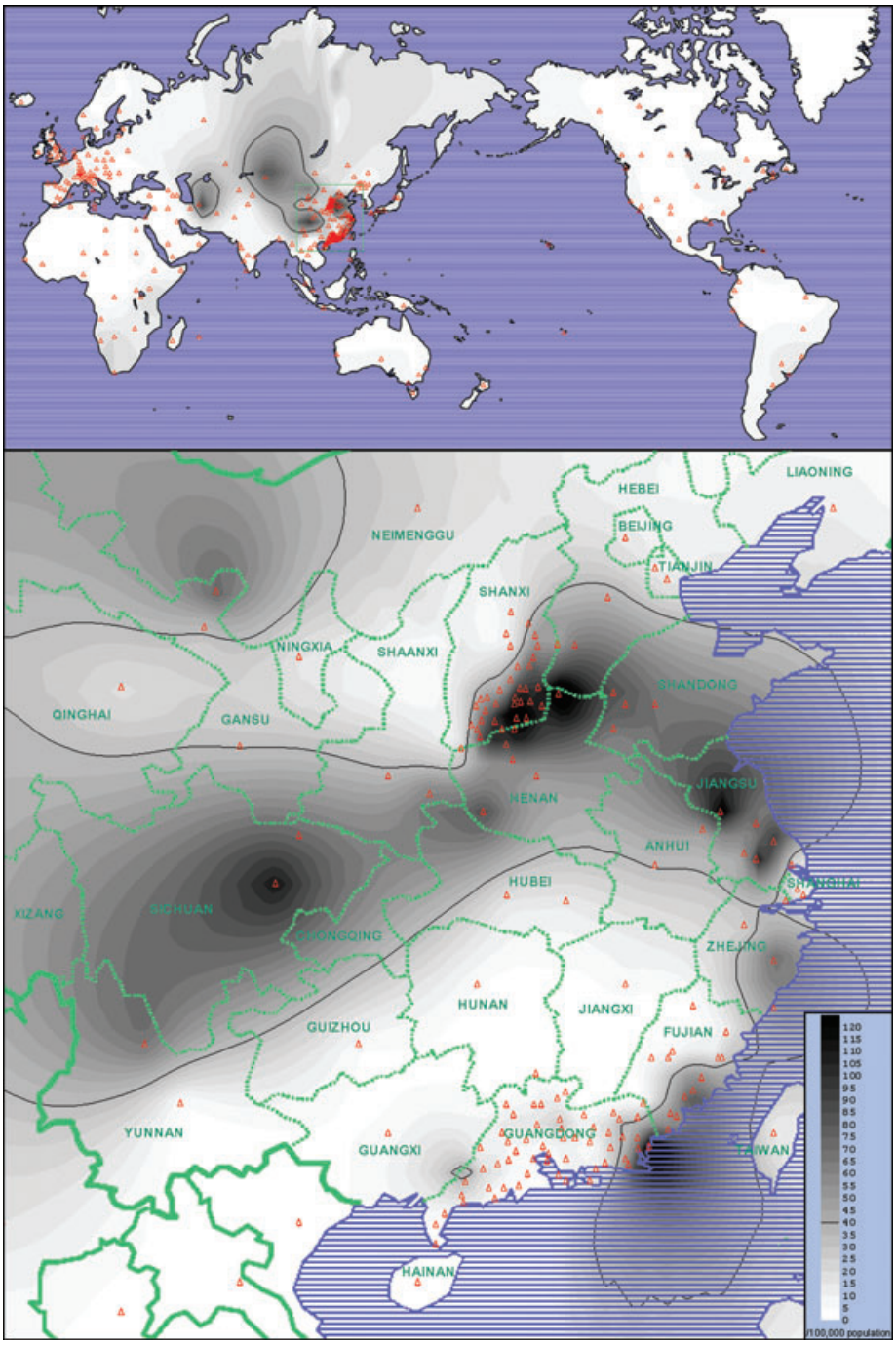

Figure 3 The esophageal cancer incidences of males in the world and in East China.

for esophageal cancer. As noted above, ALDH2 also has other metabolic functions that could be independently influencing the distribution of *ALDH2* variants (Li et al., 2006; Larson et al., 2007). Some environmental factors such as soil and vegetation characteristics and life styles may also be associated with the esophageal cancer risk (Wu et al., 2007; Fan et al., 2008; Moradi, 2008).

In Central Chinese populations, the heritability of esophageal cancer is estimated at around 49% (Han et al., 1994; Li et al., 1998). East Asian migrants in America also have a much higher esophageal cancer incidence than European Americans and African Americans (Parkin et al., 1997), indicating the pronounced heritability of esophageal cancer. Therefore, the incidence of esophageal cancer is affected by

multiple factors that interact with the *ALDH2* 504Lys* allele frequency in a complex way. That complexity could explain the differences between the distributions of esophageal cancer and the *ALDH2* 504Lys* allele carriers in East Asia.

In most areas of South China and Southeast Asia, the incidence of esophageal cancer is much lower than that observed in Central China, indicating that there are fewer environmental risk factors and lower susceptibility of esophageal cancer in South China. However, there is still a high incidence area in Southeast China, which might be associated with the highest allele frequency of *ALDH2* 504Lys* in exactly the same geographic area. In contrast to the high incidence of esophageal cancer in Southeast China being the consequence of the high *ALDH2* 504Lys* frequency, it is possible that the high incidence of esophageal cancer in Central China is working to decrease the *ALDH2* 504Lys* frequency while cultural pressure to consume ethanol increases as the impact of *ALDH2* 504Lys* decreases. The answer depends on which factors increasing risk are most important in which area and how they interact.

Conclusion

In conclusion, we hypothesize that the oriental *ALDH2* 504Lys* variant might have originated in the ancient Han Chinese population in Central China and spread to most areas of East Asia with the expansion of Han Chinese and their genetic influences on neighboring populations over the past few thousand years. Some diseases such as esophageal cancer show a complex relationship with the frequency of *ALDH2* 504Lys*. Where the *ALDH2* 504Lys* frequency is high for whatever reason, as in Southeast China, there is a clear increased risk of esophageal cancer in heterozygotes that results in higher esophageal cancer incidences in some subregions. In other areas of China there is also an increased risk of esophageal cancer in heterozygotes (Wu et al., 2001; Chen, 2005; Yang, 2005; Xiao, 2007; Yang et al., 2007) but the lower frequency of *ALDH2* 504Lys* is not sufficient to explain the high incidence of esophageal cancer. More genetic epidemiological investigations in China are required to reveal any possible reciprocal relationship between esophageal cancer and the *ALDH2* 504Lys* allele and identify the other risk factors that appear to be present.

Acknowledgements

KKK was supported in part by U.S. Public Health Service grants AA009379 and GM057672 and by National Science Foundation grant BCS0725180. NY and ER were supported by the Biodiversity and Dynamics of Gene Pools program of the Presidium of the Russian Academy of Sciences. SB, VS and AM were supported by the Russian Foundation for Basic Research grants ## 09-04-01755-a, 06-04-48274, 07-04-01629, 07-04-10173, 07-04-10177. LJ is supported by grants from the National Outstanding Youth Science Foundation of China (30625016), National Science Foundation of China (30890034), 863 Program (2007AA02Z312), and Shanghai Leading Academic Discipline Project (B111). Dr. Maria Landi from National Cancer Institute gave us important suggestion on data analyses. We especially thank the many individuals who volunteered to provide samples for this study.

Web Resource

ALDH2 504Lys* allele frequency is being updated in ALFRED, the Allele Frequency Database:
http://alfred.med.yale.edu/alfred/SiteTable1A_working.asp?siteuid=SI000734O

References

Algar, E. M. & Holmes, R. S. (1989) Kinetic properties of murine liver aldehyde dehydrogenases. In: *Enzymology and molecular biology of carbonyl metabolism 2: aldehyde dehydrogenase, alcohol dehydrogenase, and aldo-keto reductase* (eds. H. Weiner & T. G. Flynn), pp. 93–103. New York: Alan R. Liss.

Chen, C. C., Lu, R. B., Chen, Y. C., Wang, M. F., Chang, Y. C., Li, T. K. & Yin, S. J. (1999) Interaction between the functional polymorphisms of the alcohol-metabolism genes in protection against alcoholism. *Am J Hum Genet* **65**, 795–807.

Chen, H. G. (2005) A Match Case-control Study on Esophageal Cancer Susceptibility with Lifestyle Habits and Genetic Polymorphisms of *ALDH2* and *ADH2*. *Master dissertation*, Sichuan University, Sichuan, China.

Chen, C. H., Budas, G. R., Churchill, E. N., Disatnik, M. H., Hurley, T. D., Mochly-Rosen, D. (2008) Activation of aldehyde dehydrogenase-2 reduces ischemic damage to the heart. *Science* **321**, 1493–1495.

Couzigou, P., Coutelle, C., Fleury, B. & Iron, A. (1994) Alcohol and aldehyde dehydrogenase genotypes, alcoholism and alcohol related disease. *Alcohol Alcohol Suppl* **2**, 21–27.

Druesne-Pecollo, N., Hercberg, S., Latino-Martel, P., Tehard, B., Mallet, Y., Gerber, M. & Norat, T. (2009) Alcohol and genetic polymorphisms: effect on risk of alcohol-related cancer. *Lancet Oncol* **10**, 173–180.

Goldman, D. & Enoch, M. A. (1990) Genetic epidemiology of ethanol metabolic enzymes: a role for selection. *World Rev Nutr Diet* **63**, 143–160.

Fan, Y., Yuan, J. M., Wang, R., Gao, Y. T. & Yu, M. C. (2008) Alcohol, tobacco, and diet in relation to esophageal cancer: the shanghai cohort study. *Nutr Cancer* **60**, 354–363.

Greenfield, N. J. & Pietruszko, R. (1977) Two-aldehyde-dehydrogenases from human liver. Isolation via affinity chromatography and characterization of the isoenzymes. Biochem. *Biophys. Acta* **483**, 35–45.

Hammer, M. F., Karafet, T. M., Park, H., Omoto, K., Harihara, S., Stoneking, M. & Horai, S. (2006) Dual origins of the Japanese: common ground for hunter-gatherer and farmer Y chromosomes. *J Hum Genet* **51**, 47–58.

Han, C. L., Li, L. S., Jiang, Y. T. & Zhang, X. T. (1994) Study on Genetic Epidemiology of Esophageal Cancer in Shaanxi Region. *J Prev Med Chin P.L.A.* **12**, 25–28.

Hasituya, Bi, L. F. & Su, X. L. (2007) Progress on Aldehyde Dehydrogenase 2 (ALDH2) Gene. *Chin J Birth Health Hered* **15**, 3–4.

He, Y. T., Hou, J., Chen, Z. F., Song, G. H., Qiao, C. Y., Meng, F. S., Ji, H. X. & Chen, C. (2006) Study on the esophageal cancer incidence and mortality rate from 1974–2002 in Cixian, China. *Chin J Epidemiol* **27**, 127–131.

Hempel, J. D. & Pietruszko, R. (1978) Structural distinctness of the isozymes of human liver aldehyde dehydrogenase. *Fed Proc* **37**, 1798.

Hashibe, M., McKay, J. D., Curado, M. P., Oliveira, J. C., Koifman, S., Koifman, R., Zaridze, D., Shangina, O., Wünsch-Filho, V., Eluf-Neto, J., Levi, J. E., Matos, E., Lagiou, P., Lagiou, A., Benhamou, S., Bouchardy, C., Szeszenia-Dabrowska, N., Menezes, A., Dall'Agnol, M. M., Merletti, F., Richiardi, L., Fernandez, L., Lence, J., Talamini, R., Barzan, L., Mates, D., Mates, I. N., Kjaerheim, K., MacFarlane, G. J., Macfarlane, T. V., Simonato, L., Canova, C., Holcátová, I., Agudo, A., Castellsagué, X., Lowry, R., Janout, V., Kollarova, H., Conway, D. I., McKinney, P. A., Znaor, A., Fabianova, E., Bencko, V., Lissowska, J., Chabrier, A., Hung, R. J., Gaborieau, V., Goffetta, P. & Brennan, P. (2008) Multiple *ADH* genes are associated with upper aerodigestive cancers. *Nat Genet* **40**, 707–709.

Ikawa, M. C., Impraim, C., Wang, G. & Yoshida, A. (1983) Isolation and characterization of aldehyde dehydrogenase isozymes from usual and atypical human livers. *J Biol Chem* **258**, 6282–6287.

Larson, H. N., Weiner, H. & Hurley, T. D. (2005) Disruption of the coenzyme binding site and dimer interface revealed in the crystal structure of mitochondrial aldehyde dehydrogenase "Asian" variant. *J Biol Chem* **280**, 30550–30556.

Larson, H. N., Zhou, J., Chen, Z., Stamler, J. S., Weiner, H. & Hurley, T. D. (2007) Structural and functional consequences of coenzyme binding to the inactive Asian variant of mitochondrial aldehyde dehydrogenase: roles of residues 475 and 487. *J Biol Chem* **282**, 12940–12950.

Li, D. P., Dandara, C., Walther, G. & Parker, M. I. (2008) Genetic polymorphisms of alcohol metabolising enzymes: their role in susceptibility to oesophageal cancer. *Clin Chem Lab Med* **46**, 323–328.

Li, W. D., Wang, X. Q., Zhang, C. L., Han, X. Y., Chen, D. Q., Zhang, T., Pan, X. F., Jia, Y. T., Mao, X. Z., Zhang, R. & Wu, M. (1998) Esophageal carcinoma in part of population of Yangquan City. *Natl Med J China* **78**, 203–206.

Li, Y., Zhang, D., Jin, W., Shao, C., Yan, P., Xu, C., Sheng, H., Liu, Y., Yu, J., Xie, Y., Zhao, Y., Lu, D., Nebert, D. W., Harrison, D. C., Huang, W. & Jin, L. (2006) Mitochondrial aldehyde dehydrogenase-2 (ALDH2) Glu504Lys polymorphism contributes to the variation in efficacy of sublingual nitroglycerin. *J Clin Invest* **116**, 506–511.

Long, J. C., Knowler, W. C., Hanson, R. L., Robin, R. W., Urbanek, M., Moore, E., Bennett, P. H. & Goldman, D. (1998) Evidence for genetic linkage to alcohol dependence on chromosomes 4 and 11 from an autosome-wide scan in an American Indian population. *Am J Med Genet* **81**, 216–221.

Long, J., Lewis, C., Li, J., Malhi, R. & Hunley, K. (2006) Allelic Variability and Tests for Natural Selection at the Human ALDH2 Locus. The American Society of Human Genetics 56th Annual Meeting. New Orleans, Louisiana, October 9–13, 2006. No.20578/1046.

Luo, X., Kranzler, H. R., Zuo, L., Wang, S., Schork, N. J. & Gelernter, J. (2006) Diplotype trend regression analysis of the ADH gene cluster and the ALDH2 gene: multiple significant associations with alcohol dependence. *Am J Hum Genet* **78**, 973–987.

Moradi, A. (2008) Risk Factors associated with esophageal cancer in North of Iran. *Saudi Med J* **29**, 785.

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., Okonofua, F., Zhukova, O. V., Kidd, J. R. & Kidd, K. K. (2004) The evolution and population genetics of the ALDH2 locus: random genetic drift, selection, and low levels of recombination. *Ann Hum Genet* **68**, 93–109.

Parkin, D. M., Whelan, S. L., Ferlay, J., Raymond, L. & Young, J. (1997) Cancer Incidence in Five Continents, Vol VII. IARC Scientific Publication No. 143, Lyon.

Reich, T., Edenberg, H. J., Goate, A., Williams, J. T., Rice, J. P., Van Eerdewegh, P., Foroud, T., Hesselbrock, V., Schuckit, M. A., Bucholz, K., Porjesz, B., Li, T. K., Conneally, P. M., Nurnberger, J. I. Jr., Tischfield, J. A., Crowe, R. R., Cloninger, C. R., Wu, W., Shears, S., Carr, K., Crose, C., Willig, C. & Begleiter, H. (1998) Genome-wide search for genes affecting the risk for alcohol dependence. *Am J Med Genet* **81**, 207–221.

Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E. H., McCarroll, S. A., Gaudet, R., Schaffner, S. F., Lander, E. S. & The International HapMap Consortium. (2007) Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913–918.

Samatoyannopoulos, G., Chen, S. H. & Fukui, M. (1975) Liver alcohol dehydrogenase in Japanese: high population frequency of atypical form and its possible role in alcohol sensitivity. *Am J Hum Genet* **27**, 789–796.

Wen, B., Li, H., Lu, D., Song, X., Zhang, F., He, Y., Li, F., Gao, Y., Mao, X., Zhang, L., Qian, J., Tan, J., Jin, J., Huang, W., Deka, R., Su, B., Chakraborty, R. & Jin, L. (2004) Genetic evidence supports demic diffusion of Han culture. *Nature* **431**, 302–305.

Wolff, P. H. (1972) Ethnic differences in alcohol sensitivity. *Science* **175**, 449–450.

Wu, J. Z., Ding, J. H., Li, S. P., Gao, C. M., Zang, Y., Zhou, J. N., Su, P., Liu, Y. T., Zhou, X. F., Wang, R. H. & Ding, B. G. (2001) Polymorphisms of aldehyde dehydrogenase 2 genotypes and the risk of esophageal cancer. *Bull Chin Cancer* **12**, 25–27.

Wu, K. S., Huo, X. & Zhu, G. H. (2007) Soil and vegetation characteristics of high- and low-risk areas of esophageal cancer in China. *Asian Pac J Cancer Prev* **8**, 597–601.

Xiao, J. C. (2007) Study on the Population Genetic Characteristics of Esophageal Cancer. Master dissertation, Zhengzhou University, Henan, China.

Yang, C. X. (2005) Life-style and Gene Polymorphisms on Esophageal Cancer Risk. Ph.D. dissertation, Sichuan University, Sichuan, China.

Yang, S. J., Wang, H. Y., Li, X. Q., Du, H. Z., Zheng, C. J., Chen, H. G., Mu, X. Y. & Yang, C. X. (2007) Genetic polymorphisms of ADH2 and ALDH2 association with esophageal cancer risk in southwest China. *World J Gastroenterol* **13**, 5760–5764.

Yokoyama, A., Kato, H., Yokoyama, T., Igaki, H., Tsujinaka, T., Muto, M., Omori, T., Kumagai, Y., Yokoyama, M. & Watanabe, H. (2006) Esophageal squamous cell carcinoma and aldehyde dehydrogenase-2 genotypes in Japanese females. *Alcohol Clin Exp Res* **30**, 491–500.

Yokoyama, A., Muramatsu, T., Omori, T., Yokoyama, T., Matsushita, S., Higuchi, S., Maruyama, K. & Ishii, H. (2001) Alcohol and aldehyde dehydrogenase gene polymorphisms and oropharyngolaryngeal, esophageal and stomach cancers in Japanese alcoholics. *Carcinogenesis* **22**, 433–439.

Yokoyama, A. & Omori, T. (2005) Genetic polymorphisms of alcohol and aldehyde dehydrogenases and risk for esophageal and head and neck cancers. *Alcohol* **35**, 175–185.
Yoshida, A., Huang, I. Y. & Ikawa, M. (1984) Molecular abnormality of an inactive aldehyde dehydrogenase variant commonly found in Orientals. *Proc Natl Acad Sci U S A* **81**, 258–261.
Yu, C., Li, Y., Chen, W. & Yue, M. (2002) Genotype of ethanol metabolizing enzyme genes by oligonucleotide microarray in alcoholic liver disease in Chinese people. *Chin Med J (Engl.)* **115**, 1085–1087.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1 *ALDH2* 504Lys* frequencies of all the available population samples.
Table S2 Male esophageal cancer incidences in the world populations.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Received: 30 December 2008
Accepted: 18 February 2009