ADH VARIATION IN INDIGENOUS ALTAIAN POPULATIONS

By

Lydia Gau

In

Anthropology

Submitted to the
Department of Anthropology
University of Pennsylvania

Thesis Advisor: Dr. Theodore Schurr

2011

Abstract

Recent studies have revealed the high frequency of a specific alcohol dehydrogenase (ADH) allele (ADH1B*47His) that is associated with a decreased risk of alcoholism in East Asian populations. It has been speculated that selection on this gene occurred with the expansion of rice agriculture in East Asia, suggesting that cultural practices have acted to increase the frequency of the allele. The purpose of my study is to extend these findings and examine ADH variation in indigenous Altaian populations. These groups live in southern Siberia, a region that has not yet been surveyed for variation at this locus, and have traditionally been nomadic hunter-gatherers, not agriculturalists. Using PCR amplification, DNA sequencing and SNP genotyping methods, I discovered that the ADH1B*47His allele is present in these Altaian populations, albeit at lower frequencies that seen in previously studied East Asian populations. This finding suggests several possible explanations for the presence of this allele in southern Siberia, including other protective effects of this allele, contact with neighboring populations, genetic drift, or some combination of these factors. The new data from this study will add to our understanding of the evolution of the ADH gene in Central and East Asia, and the history of Altaian populations themselves.

Table of Contents

Table of Figures & Tables

Introduction

The purpose of this study is to broaden our understanding of the evolution of the alcohol dehydrogenase (ADH) gene in populations from the Altai-Sayan region. These regions surround East Asia, where research on this gene has already been conducted, but have not yet been surveyed for variation at this locus. One particular allele (ADH1B*47His) is highly frequent in East Asia, and thought to have undergone selection in the recent past (Han et al. 2007). This frequency is particularly interesting because there are few other high allelic frequencies of other types of genes associated with a particular function throughout the world.

As shown in Figure 1, this allele has been found at statistically significant frequencies in some East Asian populations, and exists in other East Asian populations, some in the Near East and Europe. While the ADH1B*47His allele has been associated with protective effects against alcoholism, it has been speculated that selection on this allele occurred with the expansion of agriculture of rice in East Asia. This suggests that culturally related forces have acted to increase the frequency of this allele over the past several thousand years.

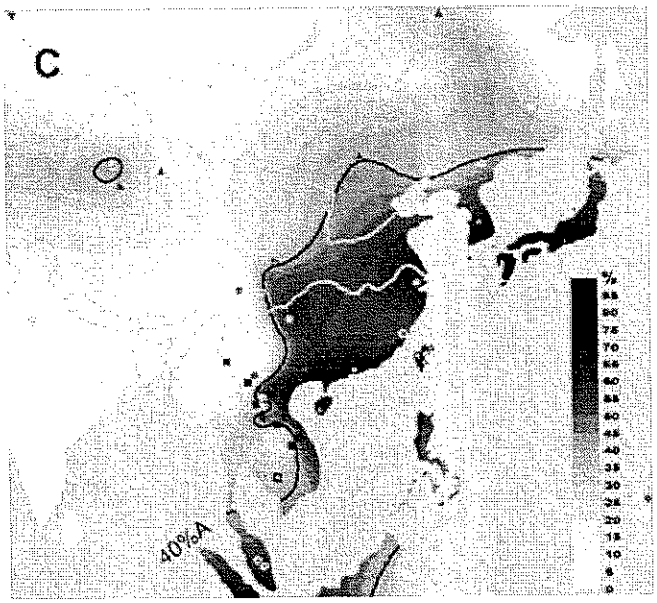

Figure 1. ADH1B*47His frequencies spread in East Asia (Li et al. 2008)

Background to the Research Problem

According to Ramchandani et al. (2001), alcohol metabolism influences human drinking behaviors and the development of alcoholism. Alcohol is broken down in a two-step process. (Figure 2). The first step involves these alcohol dehydrogenase genes. ADH genes code for the ADH enzymes that break down alcohol to a synthesizable form through this process occurring mostly in the liver.

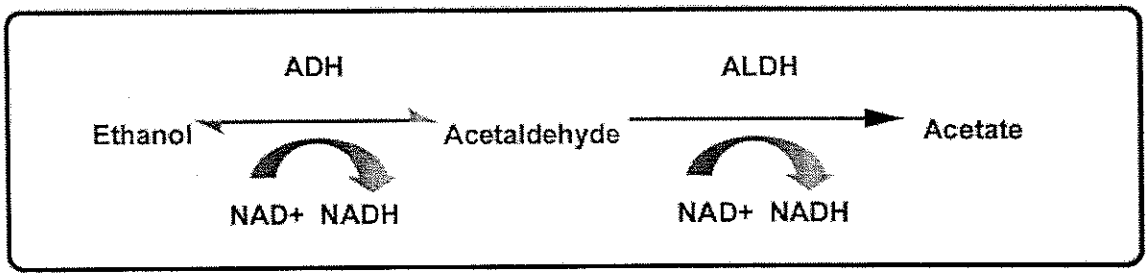

Figure 2. Ethanol Metabolism Process

There are seven known ADH genes that encode enzymes necessary for ethanol metabolism (Osier et al. 2002). These seven genes exist in one area in a ~380-kb cluster on the long arm of chromosome 4: The Class I ADH genes (ADH1A, ADH1B, ADH1C), ADH7, ADH6, ADH4, and ADH 5 (Figure 3). It has been observed that ADH1B and ADH1C genes have alleles that influence the activity of ADH enzymes that catalyze the metabolic reaction (Edenberg and Bosron 1997).

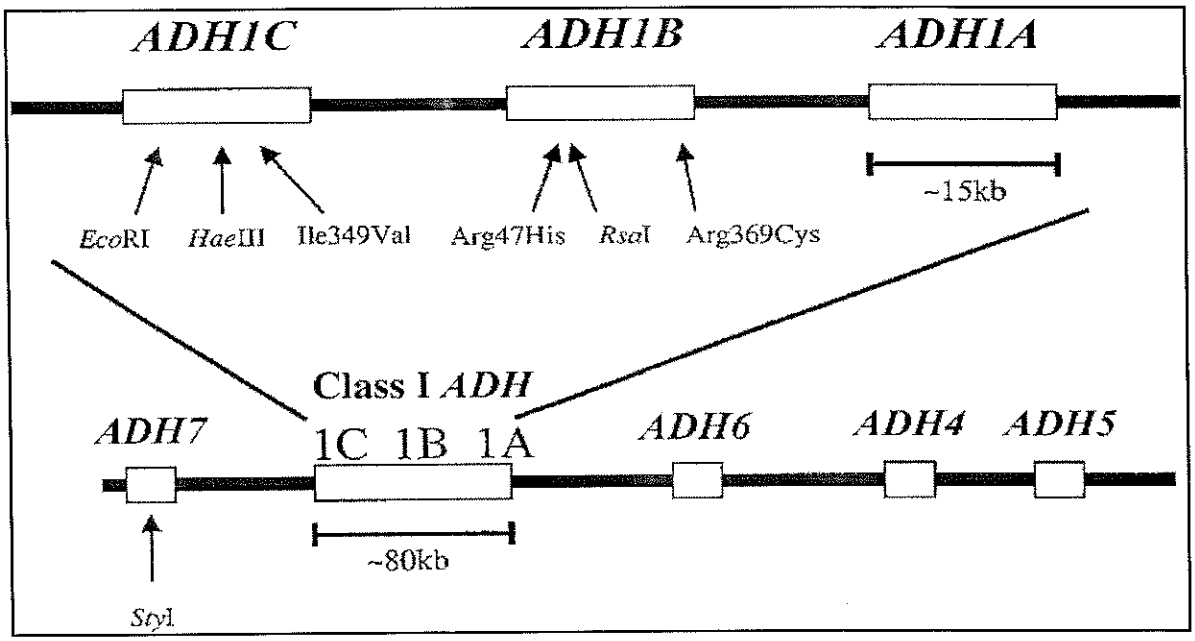

Figure 3. Map of ADH genes (Osier et. al. 2002)

Studies have shown the genetic association between this area on chromosome 4 and alcoholism (Reich et al. 1998). The combination of ADH genes (alleles or haplotypes) that a person has affects the rapidity and efficacy with which alcohol is broken down. There are two functional variants associated with a high activity level: ADH1B*47His at the Arg47His (exon 3) SNP, and ADH1C*349Val at the Ile349Val (exon 8) SNP. These two alleles have been found to be in strong linkage disequilibrium (Osier et al. 1999).

As noted above, the frequency of these alleles is noticeably different in different world populations. According to Newmark et al. (1998), certain East Asian populations have frequencies of the ADH1B*47His allele that are greater than 0.33, while other populations have less than 0.25. These two alleles are found at significantly lower frequencies in alcoholic individuals than non-alcoholic controls in East Asian samples, suggesting the effects of selection, given that this rare allele is so heavily present in one geographical area (Li et al. 2001).

The byproduct of this first step is the compound acetaldehyde. Acetaldehyde is further broken down into acetate by aldehyde dehydrogenase (ALDH), which is encoded by the gene *ALDH2*. This is an important step because acetaldehyde is carcinogenic to humans. Thus, the efficiency of the breakdown of alcohol depends on the activity of the ADH and ALDH enzymes. Different forms of these enzymes function at different activity levels, thus leading to differences in exposure time to acetaldehyde in alcohol drinkers (Osier et al. 2002).

Along with a lower risk of alcoholism, research has revealed that the ADH1B*47His is associated with the flushing response in East Asian populations, such as Koreans, Japanese, Han Chinese, Hmong-Mien, Daic, and Austronesians (Li et al. 2008). The ADH1B*47His allele increases blood levels of acetaldehyde because it causes a reduced efficiency in its breakdown, thereby causing this flushing response (Eng et al. 2007). Physiological responses can include facial flushing, nausea, and tachycardia (Harada et al. 1981). This response is more commonly known as the "Asian flush."

The protein produced by ALDH2, ALDH2*2 (or ALDH2-deficient), is also associated with the flushing response in East Asian populations. According to Osier et al. (2002:96), it "acts as a dominant null allele that dramatically reduces the ability to break down acetaldehyde." ALDH2*2 paired with ADH1B*47His results in a high level of acetaldehyde in the body following ethanol ingestion. This observation suggests that the first step of ethanol metabolism is particularly fast, while the second conversion is not. Thus, these effects deter these individuals from drinking heavily resulting in lower risk of alcoholism.

It is believed that selection has occurred on these alleles because mutations affect these two sequential steps in ethanol metabolism, it is only common in East Asia, and that these mutations both reinforce the effect of increased acetaldehyde (Goldman and Enoch 1990). Interestingly, studies suggest that selection on this allele was not recent. In fact, it has been

speculated that selection on this gene occurred with the expansion of rice agriculture in East Asia, suggesting that culturally related forces have acted to increase the frequency of the allele. More specifically, it is suggested that the allelic variants of the *ALDH* and *ADH* genes were selected prior to alcohol production.

In this regard, Goldman and Enoch (1990) have suggested that mycotoxins and infectious diseases have led to the selection of this allele in this region (Han et al. 2007). Mycotoxins, which are produced by toxin-producing fungi that infest rice, can be protected against by the deficient ALDH2*2 variant of these enzymes. The further dangerous effects of mycotoxins can be provoked through the consumption of alcohol. Having alcohol more quickly converted into a different organic compound would be favorable in mitigating the effects of mycotoxins. Thus, individuals carrying the favored allele that metabolizes alcohol faster would be selectively favored.

Infectious disease agents are also likely to be harmed by acetaldehyde. Individuals with deficient *ALDH2*2* can produce high enough levels of acetaldehyde to inhibit the growth of pathogens. Similarly, protective *ADH* alleles may have been selected for similar reasons to these *ALDH* alleles, i.e., they help reduce infection by pathogens.

Specific Aims

Given this history of the Class I ADH gene cluster and the protective effects of ADH and ALDH2 alleles, I undertook this project to test the hypothesis that selection on this allele correlates with this specific sustenance method, and also determine the range of its spread through North-East Asia. By surveying these populations and using historical data about the

spread of rice agriculture, we can open up the discussion of this hypothesis. If we were to find the ADH1B*47His allele to be statistically significant in this area, this information would disprove this hypothesis as rice agriculture did not reach this region until much later during Russian conquest.

To achieve these aims, I genotyped 180 individuals for seven SNPs in four ADH genes (ADH1A, ADH1B, ADH1C, ADH7) in two indigenous Altaian populations. These SNPs were chosen based on previous research on the ADH genes in different world populations. In addition, indigenous Altaians live in a region that flanks East Asia, and have not yet been surveyed for variation at this locus. This analysis yielded new information about Class I ADH gene variation in indigenous Siberians, and allows assess of the possible evolutionary and cultural forces that influenced its spread throughout East Asia.

Materials and Methods

Samples

The populations in which I have studied ADH variation are indigenous Northern and Southern Altaians. These samples were previously collected through fieldwork conducted in the Altai-Sayan region between 1993-2003. These populations are indigenous to the Altai Mountains, shown in Figure 4. They both are Turkic-speaking populations that have been traditionally nomadic hunter-gatherers, although having historically different backgrounds.

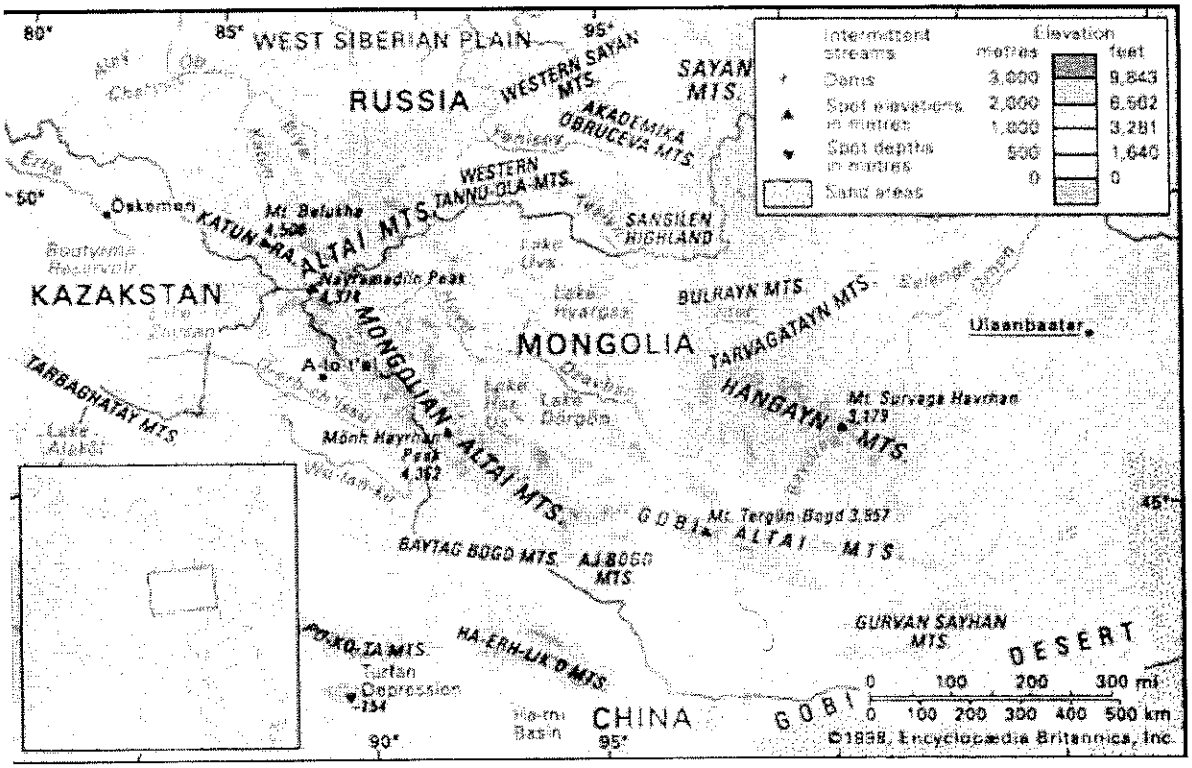

Figure 4. Map of Altai Region

The Northern Altaians are actually comprised of a number of different ethnic groups. These include Chelkan, Tubular, and Kumandin. They lived in either the Kebezen, Kurmach-Baigol, or Biika villages. The Southern Altaian samples all came from Altai-kizhi individuals living the Mendur-Sokkon village (**Figure 5**). A minority of the samples that were analyzed came from ethnic Russians, Kazakhs, and Ukrainians living in these locations. In total, I analyzed 92 indigenous Southern Altaian and 88 indigenous Northern Altaian samples for variation at the ADH locus. I also received control samples from Dr. Ken Kidd's lab at Yale University to help confirm the presence or absence of specific mutations in the ADH genes.

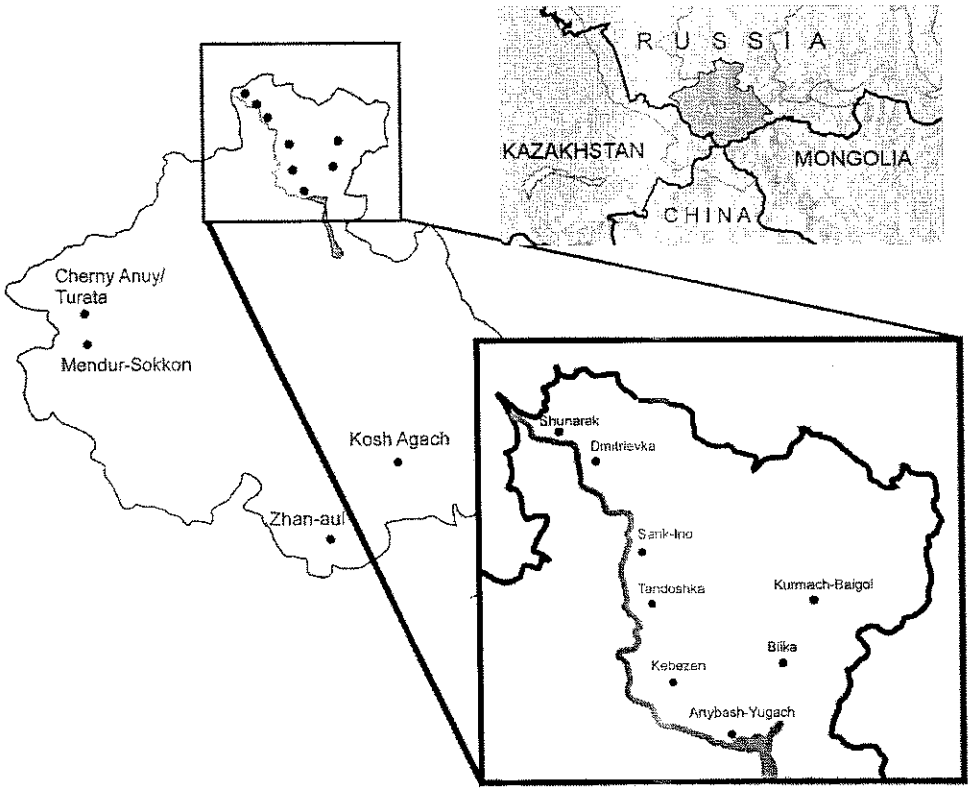

Figure 5. Map of Indigenous Altaian villages

The data generated from this study were compared to those obtained in the previously mentioned studies of ADH variation in East and Central Asian populations (Osier et al. 2002, Li et al 2007). All comparative data from these studies were retrieved from the Allele Frequency Database (ALFRED, http://alfred.med.yale.edu/alfred/). These data include each specific single nucleotide polymorphism frequency data from other populations around the world.

Genotyping Methods

As described in Osier et al. (1999) and Osier et al. (2002), all polymorphisms were first typed through PCR-RFLP analysis. The ADH Class I genes consists of ADH1A, ADH1B and ADH1C. Given that the entire Class I ADH cluster consists of multiple loci, I focused on the ADH1B and ADH1C genes, as well as the ADH7 loci. ADH7 is upstream of the Class I ADH

genes. The single nucleotide polymorphism (SNP) I analyzed on this gene is StyI (intron 5). I analyzed SNPs that were present in different ADH Class I genes. I looked at three sites on each ADH1B and ADH1C, for a total of six single nucleotide polymorphisms. For ADH1C, these sites were EcoRI (intron 2), HaeIII (exon 5), and Ile349Val. For ADH1B, these sites were Arg47His, RsaI (intron 3), and Arg369Cys.

The enzyme used for the RFLP analysis for ADH7 StyI analysis was StyI. If the site was absent, the fragment size was 477 base pairs (bp). If the site was present, the two fragments would be 263bp + 214bp. In this case, the site present indicates the ancestral state. Accordingly, the enzymes used for each polymorphic site and the rest of the fragment sizes are listed in Table 1. For ADH1C EcoRI and ADH1C Ile349Val a site present indicates the ancestral state. For ADH1C HaeIII, ADH1B Arg47His, ADH1B RsaI, and ADH1B Arg369Cys, the site absent indicates the ancestral state.

I used the primers and PCR conditions from Osier et al. (2002) to amplify each section of ADH gene in which the six SNPs in ADH1B and ADH1C occurred (Table 1). I analyzed 64 Southern Altaian and 44 Northern Altaian samples using this method. PCR amplifications were done in 96-well plates (total volume: 25 μl) using conditions stated in Table 1. The PCR products were then digested with the appropriate restriction enzymes for the six different SNPs. Digestion reactions took place over a 12-hour period in a 37° incubator. RFLPs were detected through gel electrophoresis using 3% Nusieve agarose gels. Gels were imaged using KODAK Image Analysis Software. The raw data from these analyses are shown in the Appendix. In the raw data, a site present is indicated by (+). A site absent is indicated by (-). due to the ADH genes being autosomal in nature, i.e., present in two copies in each individual, I have indicated whether they are heterozygous (+, -) or homozygous (+, +) or homozygous (-, -).

Table 1. PCR Reagents and Conditions (Osier et al. 2002)

Polymorphic Site (Location)	Forward and Reverse Primers	Cycling Conditions[a] (No. of Cycles)	Enzyme[b]	Fragment Size(s) (bp)	
				Site Absent	Site Present
ADH7 StyI (intron 5)	A7IN5DW2 (5'-TAT TAA ATT ATT GCT TAA TAA CTG G-3'), A7IN5UP1 (5'-TTC CTG TGT CTC TTA CAG TG-3')	95°C (15 s), 54°C (15 s), 72°C (60 s) [40]	*Sty*I	477	263 + 214
ADH1C:					
*Eco*RI (intron 2)[c]	A3EX2DW (5'-TTG CAC CTC CTA AGG CTC-3'), A3*Eco*UP2 (5'-TCT AAT GCA AAT TGA TTG TGA AC-3')	95°C (15 s), 51°C (15 s), 72°C (75 s) [40]	*Eco*RI	323	242 + 81
*Hae*III (exon 5)	A3EX5FOR2 (5'-TGA GTT TGC ACA TTA GTT ATG G-3'), A3EX5REV1 (5'-TGC TCT CAG TTC TTT CTG GG-3')	95°C (30 s), 56°C (30 s), 72°C (60 s) [35]	*Hae*III	435	193 + 242
Ile349Val[c]	A3FXNFOR1 (5'-TTG TTT ATC TGT GAT TTT TTT TGT-3'), A3FXNREV3 (5'-CGT TAC TGT AGA ATA CAA AGC-3')	95°C (15 s), 51°C (15 s), 72°C (75 s) [40]	*Ssp*I	378	274 + 104
ADH1B:					
Arg47His	A2FXNFOR (5'-ATT CTA AAT TGT TTA ATT CAA GAA G-3'), A2FXNREV (5'-ACT AAC ACA GAA TTA CTG GAC-3')	95°C (30 s), 56°C (30 s), 72°C (60 s) [35]	*Msl*I	685	443 + 242
Arg369Cys[c]	HE39[d] (5'-TGG ACT TCA CAA CAA GCA TGT-3'), HE40[d] (5'-TTG ATA ACA TCT CTG AAG AGC TGA-3')	95°C (15 s), 58°C (15 s), 72°C (60 s) [40]	*Alw*NI	201	130 + 71

Note.—All PCR was performed using 100 ng genomic template, 100 ng each primer, 200 μM dNTP, 2.0 mM $MgCl_2$, 50 mM KCl, 10 mM Tris HCl (pH 8.4), and 0.5 U AmpliTaq DNA polymerase (Perkin Elmer) in a total volume of 25 μl.

[a] All cycling protocols were performed on a Perkin Elmer 9600, with an initial hold at 95°C, for 5 min, and a final hold at 72°C, for 10 min.

[b] PCR products were digested with 5 U of the appropriate restriction enzyme by use of the buffer that was recommended by the manufacturer.

[c] Dimethyl sulfoxide was added to a final concentration of 5% by volume.

[d] Xu et al. (1988).

In parallel, I used Custom TaqMan assays (ABI) for all seven of the ADH SNPs to screen all of the Altaian samples (180 total) for the same mutations identified through the PCR-RFLP method. These included the ADH1B*47His allele and the other markers in the Class I ADH and ADH7 genes. TaqMan analysis performs an allelic discrimination without having to directly sequence the samples. It provides information about the presence or absence of ancestral and derived alleles, as defined to Table 2, with all samples being homozygous or heterozygous for these alleles. The raw data are shown in the Appendix.

Table 2. Sequences of Ancestral States (Osier et al. 2002)

Polymorphic Site (Location)	Sequence[a]				Consensus Ancestral State (Symbol)
	Human	Chimpanzee	Gorilla	Orangutan	
ADH7 StyI (intron 5)	G/C CATGG	C CATGG	C CATGG	Unknown	Site present (2)
ADH1C:					
EcoRI (intron 2)	GAAT T/G C	GAAT T C	GAAT T C	AAAC T C	Site present[b] (2)
HaeIII (exon 5)	A/G GCC	A GCC	A GCC	A GCC	Site absent (1)
Ile349Val	AAT A/G TT	AAT G TT	AAT G TT	AAT G TT	349Val (2)
ADH1B:					
Arg47His	C A/G CACAGATG	C G CACAGATG	C G CACAGATG	C G CACAGATG	47Arg (1)
RsaI (intron 3)	G T/C AC	G C AC	G C AC	G C AC	Site absent (1)
Arg369Cys	CAGTATC C/T G	CAGTATC C G	CAGTATC C G	CAGTATC C G	369Arg (1)

Note.—SNP are indicated in boldface, and the sequence differences are italicized and underlined.

[a] Oriented in the 5'→3' direction in the gene cluster.

[b] With the exception of orangutan, in which two different base changes result in the restriction site being absent.

The TaqMan assays were read on an ABI 7900HT Fast Real-Time PCR system. Amplifications were done in 369-well plates (total volume: 5 μl) using conditions stated in the ABI Allelic Discrimination Guide. Reactions took place over a 3-hour period. The TaqMan genotyping was read on the SDS computer program. I repeated the screenings that failed or were unclear.

Statistical Analyses

Genotype and allele frequencies were calculated by direct gene counting. I used the computer program PHASE (Stephens et al. 2001) to calculate the maximum-likelihood estimates of haplotype frequencies. Only samples with data from all seven ADH SNPs were used to calculate the haplotype frequencies. This included 77 Southern Altaian individuals and 70 Northern Altaian individuals. With these haplotypes, I used the Arlequin v3.11 (Excoffier et al. 2005) to estimate descriptive statistics with the ADH data, including F_{st} values used for distance analysis. Multidimensional scaling (MDS) were created through the computer program SPSS to visualize the information and explore the similarities and dissimilarities in the data.

Results

Allelic Data

The locations of the genes on chromosome 4 and the assay IDs are listed in Table 7 of the Appendix. Using the ancestral states of the Class I ADH sites determined by Osier et al. (2002), I compared the genotype frequencies of each single nucleotide polymorphism of each sample and the allelic frequency in each population.

Individual Genotype Frequencies

ADH genotype frequencies for each population are listed in Table 3. The individual-site allele-frequency data are all listed in Table 4. All data collected that is confirmed is listed in the table, regardless if I was able to screen all seven sites for a particular sample.

Table 3. ADH Genotype Frequencies

SNP	ADH1B Arg369Cys			ADH1B RsaI			ADH1B Arg47His			ADH1C Ile349Val			ADH1C HaeIII			ADH1C EcoRI			ADH7 StyI	
	G/G	A/G	A/A	G/G	A/G	A/A	C/C	C/T	T/T	C/C	C/T	T/T	C/C	C/T	T/T	A/A	A/C	C/C	G/G	G/C
SALT	81	0	0	31	35	12	39	33	8	4	26	50	45	13	5	73	16	2	16	42
NALT	86	0	0	19	40	6	44	34	4	2	29	53	36	41	10	62	21	2	13	33

	Ancestral	Derived
ADH1B Arg369Cys	G	A
ADH1B RsaI	G	A
ADH1B Arg47His	C	T
ADH1C Ile349Val	C	T
ADH1C HaeIII	T	C
ADH1C EcoRI	A	C
ADH7 StyI	C	G

Table 4. ADH Allele Frequencies (%)

SNP	ADH1B Arg369Cys		ADH1B RsaI		ADH1B Arg47His		ADH1C Ile349Val		ADH1C HaeIII		ADH1C EcoRI		ADH7 StyI	
	G	A	G	A	C	T	C	T	C	T	A	C	G	C
SALT	100	0	62.18	37.82	69.38	30.63	21.3	78.8	81.75	18.25	89.01	10.99	44	56
NALT	100	0	40	60	74.39	25.61	19.6	80.4	64.94	35.06	85.29	14.71	55.7	44.3

Using the frequency data for the ADH1B*47His allele in 82 Southern Altaian samples and 85 Northern Altaian samples, I compared the Altaian data to those from the populations in Table 5. Some of the ADH1B*47His frequencies are shown in Figure 6.

Table 5. ADH1B*47His Frequency comparison with different populations

COUNTRY	POPULATION	2N	*ADH1B*47HIS* FREQUENCY (%)
Mozambique	Negroid Makrani	56	1.8
Tanzania	Sandawe	80	.0
Somalia	Somali	36	2.8
Hungary	Hungarian	178	10.1
Pakistan	Mohanna	104	7.7
India	Keralite	58	.0
Pakistan	Hazara	58	24.1
Russia	Southern Altaian	164	30.6
Russia	Northern Altaian	170	25.6
Laos	Oy	100	48.0
Laos	Brao	100	20.0
Laos	Talieng	96	38.5
Laos	Alak	92	35.9
Laos	Jeh	100	34.0
Laos	Ngeq	124	39.5
Laos	Taoih	10	50.0
Laos	Kataang	92	54.3
Laos	Suy	104	26.0
Laos	Inh	98	56.1
Laos	So	108	25.9
Laos	Phuthai	100	14.0
Laos	Aheu	90	31.1
Laos	Bo	104	31.7
Laos	Tai Mene	104	35.6
Laos	Phuan	106	39.6
Laos	Rien	100	26.0
Laos	Mal	100	21.0
Laos	Kang	36	36.1
Laos	Tai Deang	96	41.7
Laos	Tai Dam	102	33.3
Laos	Puoc	82	25.6
Laos	Bit	86	10.5
China	Bugan	46	87.0
China	Lachi	40	100.0
China	Yerong	14	42.9
China	Cao Lan	12	91.7
China	Pahng	20	100.0
China	Pou	18	83.3
China	Tujia	48	60.4

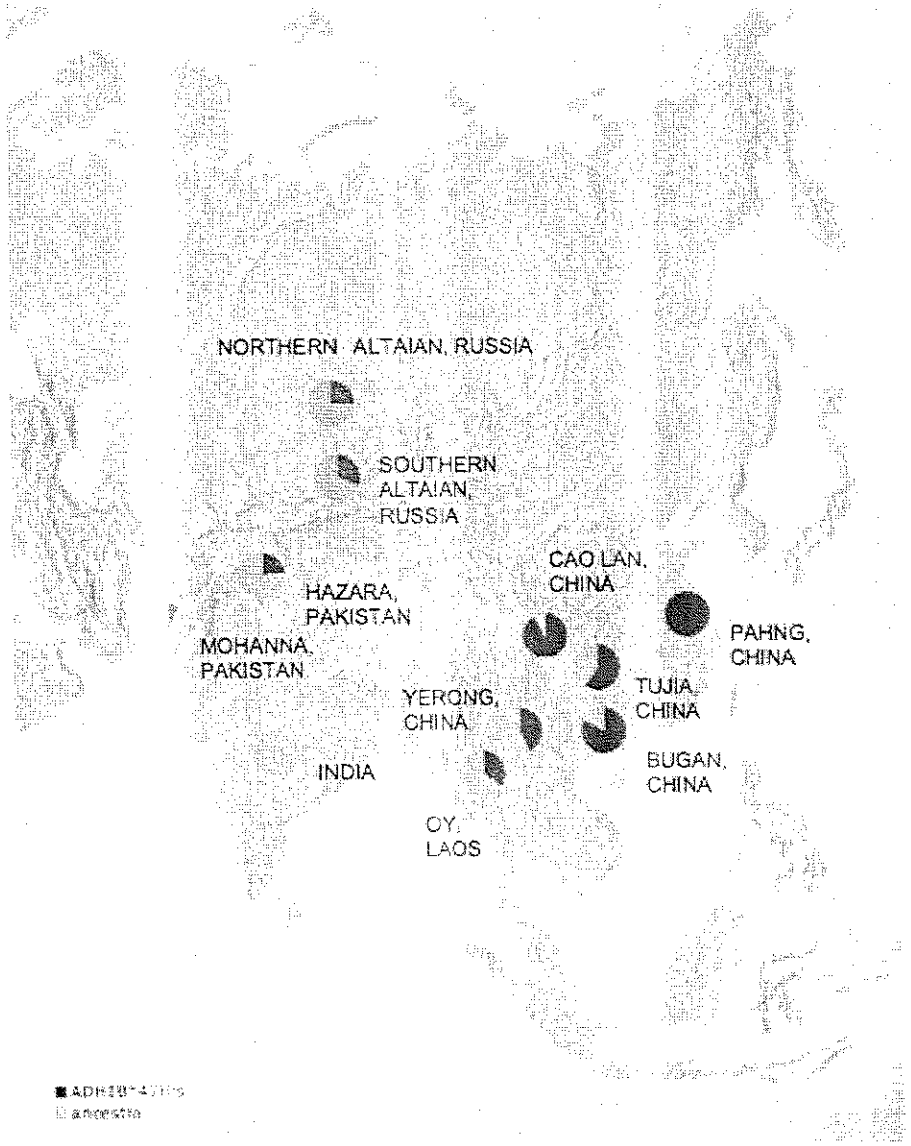

Figure 6. Map of varying ADH1B*47His frequencies in Asia

Haplotype Frequencies

Altaian haplotype diversity and *frequencies* of the haplotypes defined by the six SNPs in Class I ADH genes are presented in Table 6. They are listed in the order of ADH1C coRI, ADH1C HaeIII, ADH1C Ile349Val, ADH1B Arg47His, ADH1B RsaI, and ADH1B Arg369Cys. Each SNP is indicated as "1" being the ancestral and "2" being the derived allele. The data from other populations that appear in this table were taken from Osier et al. 2002. The "Residual Column" accounts for the total frequencies of haplotypes found in the populations, but have not exceeded 5% of the population.

The ancestral haplotype is indicated as being 212111. This haplotype does not occur often in African samples, and is more frequent in European and Native American samples. Most East Asian samples (San Francisco Chinese, Taiwanese Chinese, Hakka, Japanese, Ami, and Atayal) had high frequencies of the 221221 haplotype. According to Osier et al. (2002:88) "[t]his haplotype consists of the ADH1C EcoRI site-present allele, the ADH1C HaeII (exon 5) site-present allele, the ADH1C*349Ile allele (high-activity enzyme), the ADH1B*47His allele (high activity enzyme), and the ADH1B RsaI site-present allele." The 221211, 221221, 222211 haplotypes also have the derived ADH1B*47His allele.

Both the 212111 and 221221 haplotypes are present in the indigenous Altaian populations. However, the 221111 haplotype is the most frequent one in both populations. This haplotype does not have the derived ADH1B*47His allele. The second most frequent haplotype in Southern Altaians, 221221, includes this derived ADH1B*47His allele. The second most frequent haplotype occurring in Northern Altaians, 111111, does not include the derived ADH1B*47His allele.

Table 6. Class I ADH Six-Site Haplotype Frequencies

			ESTIMATED FREQUENCY OF HAPLOTYPE														
			1	1	1	1	2	2	2	2	2	2	2	2	2	2	2
			1	1	1	2	1	1	1	1	2	2	2	2	2	2	2
			1	2	2	1	1	2	2	2	1	1	1	1	1	2	2
			1	1	1	1	1	1	1	2	1	1	1	2	2	1	2
			1	1	2	1	1	1	2	1	1	1	2	1	2	1	1
POPULATION	*2N*	RESIDUAL	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1
Sub-Saharan African:																	
SE Bantu	96	1.5	1.4	2.8	0	4.2	0	0	0	0	76.1	9.2	4.9	0	0	0	0
!Kung San	82	0.4	2.3	0	0	2.6	8.6	10.6	0.3	0	67.2	1.2	6.8	0	0	0	0
Biaka	134	4.8	0	2.4	0	0	10.8	0	0	0	69.4	5.9	6.7	0	0	0	0
Mbuti	74	0	0	2.7	1.3	0	0	0	0	0	87.5	5.7	1.4	0	0	1.4	0
Yoruba	152	0	0	14.7	1.3	0	0	1.6	0	0	52.8	29.4	0	0	0	0	0
Ibo	94	0	0	9.6	0	0	0	3.1	0	0	50.1	35	2.1	0	0	0	0
Hausa	78	0	0	5.1	0	1.3	0	0	0	0	63	22.5	8.1	0	0	0	0
African Americans	176	4.7	0	13.6	0.3	0	1.2	2.8	0	0	50.3	19.8	6.7	0	0.6	0	0
Northern African:																	
Ethiopians	60	0	0	13.1	0	0	0	1.9	0	0	37.4	0	6.8	34	3.5	3.3	0
Northern Moroccans	186	9.5	2.2	10.7	0	0.6	0	6.1	1.9	0	35.7	1.9	22.1	0	3.7	5.5	0
Central Moroccans	174	4.9	0	17.2	1.4	1.5	0	5.3	0.6	0	35.8	2.8	22.7	0	6.9	1	0
Saharans	118	6.6	0	13.1	3	8.9	0	4.7	0.7	0	43.1	1.2	9.2	0	7.1	2.3	0
European/Middle Eastern:																	
Yemenite Jews	76	4.1	0	19.5	0.5	0	0	5	0	0	26.9	0	7.8	34	2.2	0	0
Samaritans	80	1.3	0	5	0	0	0	0.7	0	13.1	1.9	0	25	43.9	0	0	9.2
Druze	140	3.7	0	17.6	1.6	0	0	7.3	0	0	19.1	0	23	22.8	0	2.1	2.9
Adygei	106	0	0	19.8	0	0	0	5.7	0	0	39	0	22.3	7.8	5.4	0	0
Catalans	176	3.4	0	29.3	1.9	0.7	0	5.2	0	0	25.6	1.1	27	0	4.6	0.6	0.6
Basque	190	5.7	1.6	30.8	5.4	1.6	0	7	0	0	20.6	0.1	23.5	0	3.7	0.5	0
Russians	92	0	0	30.4	0	0	0	14.8	0	0.4	24.5	0	23.7	3.8	2.4	0	0
Finns	70	1.4	0	29.3	0	0	0	30.7	0	0	20.9	0	17.7	0	0	0	0
Danes	96	3.4	0	29	1.2	0	1	12.5	0	0	25.6	0	27.5	0	0	0	0
Irish	136	0	0	29.4	0	0	0.8	15.5	0	0	31.7	0	21.9	0	0	0.7	0
Caucasians	176	3.5	0	29.2	3.3	0	0.7	7.4	0.4	0	27.2	0.6	23.2	4.6	0	0	0
Eastern Asian:																	
SF Chinese	96	2.1	0	2.1	0	0	1	6.3	0	0	9.4	0	10.4	0	68.8	0	0
Taiwanese Han	94	2.2	0	0	0	0	0	3.2	0	0	10.6	0	10.7	1.1	72.3	0	0
Hakka	82	0	0	0	1.2	0	0	9.8	1.2	0	9.8	0	9.8	0	68.3	0	0
Japanese	84	1.4	0	2.4	0	0	0	3.4	0	0	9.7	0	8.3	0	74.8	0	0
Ami	80	1.3	0	0	0	0	0	0	0	0	19.6	0	2.9	15.4	60.9	0	0
Atayal	82	0	0	0	0	0	0	0	0	0	13.3	0	2.5	2.5	81.6	0	0
Cambodians	46	6.5	0	14	0	0	0	5.6	0	0	20.5	0	25.1	0	28.3	0	0
Kachari	30	3.3	0	13.3	0	0	0	0	0	0	17.4	0	59.2	6.7	0	0	0
Pacific:																	
Nasioi	44	2.3	0	0	0	0	0	38	7.4	0	23.4	0	6.2	0	18.2	4.5	0
Micronesians	66	4.7	0	0	0	0	3	14.6	2	0	42.3	0	11	0	22.2	0	0
Siberia																	
Southern Altaian	154	1.2	0	11	0	0	3.2	9.1	0.6	0	32.5	0	11.7	5.2	24.7	0.6	0
Northern Altaian	140	13.4	2.9	4.3	4.3	1.4	4.3	2.9	3.6	0	36.4	0	9.3	7.9	9.3	0	0
Yakut	98	1.3	0	14	0	0	1.3	13.1	0	0	42.4	0	9.7	2.4	15.8	0	0
North American:																	
Cheyenne	110	0	0	9.1	0	0	0	29.1	0	0	56.4	0	0.9	0	0	4.5	0
Arizona Pima	88	0	1.1	21	0	0	0	24.6	0	0	51.1	0	0	0	0	2.3	0
Mexican Pima	106	0.9	0	16.6	1.3	0	0	48.1	1	0	28.6	0	3.5	0	0	0	0
Maya	94	1.3	9.2	2.3	0	0	0	11.7	0	0	68.3	1.1	1.1	5.1	0	0	0
South American:																	
Ticuna	130	0	0	6.9	0	0	0	10.8	0	0	82.3	0	0	0	0	0	0
Rondonian Surui	76	0	0	2.6	0	1.5	0	6.6	0	0	89.3	0	0	0	0	0	0
Karitiana	100	1.2	1.1	22.7	0	1.1	0	3.2	0	0	64.9	0	0	0	0	5.8	0

F_{st} Values

F_{st} values were estimated from the 6-SNP haplotypes in Altaian and comparative populations (40), including 7 East Asian groups. The similarities and differences between populations are shown graphically through multidimensional scaling plots (SPSS) (Figures 7 and 8). The closer that two points are, the greater genetic similarity that they have.

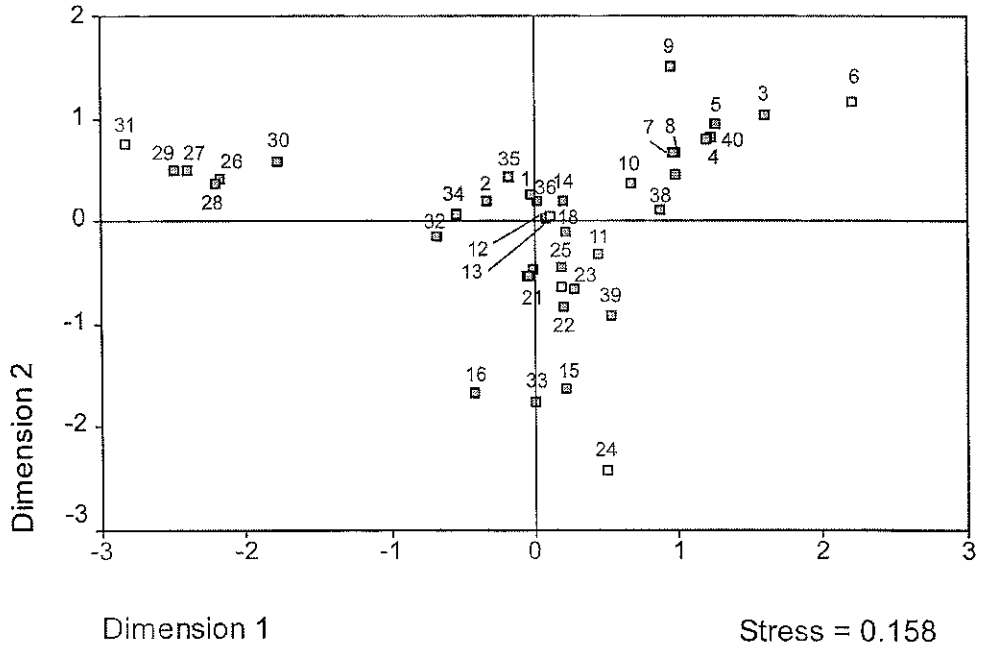

Figure 7. Distance plots for Altaians and Comparative Populations

Label	Population name
1:	Northern Altaian
2:	Southern Altaian
3:	Southeastern Bantu-speakers
4:	!Kung San
5:	Biaka
6:	Mbuti
7:	Yoruba
8:	Ibo
9:	Hausa
10:	African Americans
11:	Ethiopians
12:	Northern Moroccans
13:	Central Moroccans
14:	Saharans
15:	Yemenite Jews
16:	Samaritans
17:	Druze
18:	Adygei
19:	Catalans
20:	Basque
21:	Russians
22:	Finns
23:	Danes
24:	Irish
25:	European North Americans
26:	San Francisco Chinese
27:	Taiwanese Chinese
28:	Hakka
29:	Japanese
30:	Ami
31:	Atayal
32:	Cambodians
33:	Kachari
34:	Nasioi
35:	Micronesians
36:	Yakut
37:	Cheyenne
38:	Arizona Pima
39:	Mexican Pima
40:	Maya

My results show that the ADH1B*47His allele is present in these two populations, but appears at lower frequency than seen in East Asian populations. The ADH1B*47His allele occurs at only a slightly higher frequency in Southern Altaians compared to Northern Altaians. The Southern Altaian have a haplotype occurring second most frequently with this ADH1B*47His allele. The Southern Altaian haplotype frequencies indicate that there are more haplotypes in this populations with the ADH1B*47His derived allele than the Northern Altaian haplotype frequencies. Most Northern Altaian haplotypes have the ADH1B*47His ancestral allele.

Discussion

My findings indicate the presence of this ADH1B*47His allele in the Northern and Southern Altaian populations, but at lower frequencies than other East Asian populations. Looking at the Class I ADH Six-Site Haplotype Frequencies, we find a difference between the Northern and Southern Altaians. The Southern Altaians have more haplotypes occurring that include this ADH1B*47His derived allele. Most Northern Altaian haplotypes include the ADH1B*47His ancestral allele. This difference is significant, given the historical and ethnic differences between these two populations. Slight differences in frequencies between the two populations may reflect greater gene flow from Mongolians to Southern Altaians, or possibly genetic drift in small aboriginal populations.

The Northern and Southern Altaians are historically hunter-gatherers. Rice agriculture is not present in Altai-Sayan region until Russian conquest. Evidence has shown that agriculture started more than 8,000 years ago in East Asia (Gu et al. 2001). These findings indicate that the

ADH1B*47His allele must have entered the region well before the Russian conquest, but after it had already been well established in East Asia. Thus, the presence of rice agriculture much later would explain why the ADH*1B47His derived allele would not have been selected for. These findings suggest that there may also be other unknown protective effects of this allele beneficial, other than mitigating the effects of mycotoxins found in rice, and thus maintained in their gene pool. The maintenance of this derived allele in these populations may reflect the favorable affects of producing acetaldehyde quickly to defend against infectious diseases.

The findings indicating the presence of this ADH1B*47His allele in these populations is not surprising. Gene flow could have spread the allele without selection, since there are benefits of having this allele. The frequency of this allele can be due to genetic drift. The presence of this allele could be due to contact with neighboring populations, such as the Mongols shown historically, or the maintenance shown by the other protective effects of the allele. Given the low frequency of this allele, there is no evidence suggesting selection in these populations. Thus, selective forces that shaped ADH allelic variation in East Asian cultures are found to be different than in this Altai-Sayan region since the Altaians are different from other East Asian groups.

Acknowledgments

I would like to thank the Indigenous Altaian participants for their involvement in this research, Dr. Ludmila Osipova from the Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, and Matt Dulik from the Laboratory of Molecular Anthropology at Penn.

My research was supported through funds provided by grants from the College Alumni Fund, the Department of Anthropology Undergraduate Research Fund, and the SSHRC Major Collaborative Research Initiative 412-2005-1004.

References Cited

Edenberg HJ, Bosron WF (1997) Alcohol dehydrogenases. In: Guengerich FP (ed) Comprehensive toxicology. Volume 3: Bio-transmormation. Pergamon Press, New York, pp 119-131.

Eng MY, Luczak SE, Wall TL (2007) ALDH2, ADH1B, and ADH1C genotypes in Asians: A literature review. Alcohol Res Health 30: 22-27.

Excoffier L, Laval G, Schneider S. 2005. Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evol Bioinform Online 1:47–50.

Gokcumen O, Dulik MC, Schurr TG, Pai AA, Zhadanov SI, Rubinstein S, Osipova LP, Andreekov OV, Tabikhanova LE, Gubina MA, and Labuda D (2008) Genetic Variation in the Enigmatic Altaian Kazakhs of South-Central Russia: Insights into Turkic Population History. *American Journal of Physical Anthropology* 136: 278-293.

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

Gu WN (2001) Agricultural Archeology in China its research development and main results. Agricultural Archaeology 21(1): 1–16.

Han Y, Gu S, Oota H, Osier MV, Pakstis AJ, Speed WC, Kidd JR, Kidd KK (2007) Evidence of positive selection on a class I ADH locus. *Am J Hum Genet* 80:441–456.

Harada S, Agarwal DP, Goedde HW (1981) Aldehyde dehydrogenase deficiency as cause of facial flushing reaction to alcohol in Japanese. Lancet 2: 982.

Li H, Mukherjee N, Soundararajan U, Ta´rnok Z, Barta C, et al. (2007) Geographically Separate Increases in the Frequency of the Derived AD- H1B*47His Allele in Eastern and Western Asia. *Am J Hum Genet* 81: 842–846.

Li H, Gu S, Cai X, Speed WC, Pakstis AJ, Golub EI, Kidd JR, and Kidd KK (2008) Ethnic Related Selection for an ADH Class I Variant within East Asia. *PLoS ONE* 3.4: e1881.

Li TK, Yin SJ, Crabb DW, O'Connor S, Ramchandani VA (2001) Genetic and environmental influences on alcohol metabolism in humans. Alcohol Clin Exp Res 25:136–144.

Neumark YD, Friedlander Y, Thomasson HR, Li T-K (1998) Association of the ADH2*2 allele with reduced ethanol consumption in Jewish men in Israel: a pilot study. J Stud Alcohol 59:133–139.

Osier MV, Pakstis AJ, Soodyall H, Comas D, Goldman D, Odunsi K, Okonofua F, Parnas J, Schulz L, Bertranpetit J, et al (2002) A global perspective on genetic variation at the *ADH* genes reveals unusual patterns of linkage disequilibrium and diversity. *Am J Hum Genet* 71:84–99.

Osier M, Pakstis AJ, Kidd JR, Lee J-F, Yin S-J, Ko H-C, Edenberg HJ, Lu R-B, Kidd KK (1999)

Linkage disequilibrium at the ADH2 and ADH3 loci and risk of alcoholism. Am J Hum Genet 64:1147–1157.

Ramchandani VA, Bosron WF, Li TK (2001) Research advances in ethanol metabolism. Pathol Biol (Paris) 49:676–682.

Reich T, Edenberg HJ, Goate A, Williams JT, Rice JP, Van Eerdewegh P, Foroud T, et al (1998) Genome-wide search for genes affecting the risk for alcohol dependence. Am J Med Genet 81:207–215.

Stephens, M., Smith, N., and Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics,* 68, 978--989.

Appendix

Table 7. Raw Data

Assay ID		ss2978361		ss2978360		ss2978361		ss2978363		ss2978364		ss2978359		ss2978365
		ADH1C		ADH1B		ADH1C		ADH1C		ADH1B		ADH1B		ADH7
Position:	Chr.4	1E+08		100239319		100268856		100260789		100229017		1E+08		100340522
Ancestral		(-)	T	(-)	C	(+)	A	(-)	C	(-)	G	(-)	G	C
Village	Sample	HaeIII	HaeIII Taq	Arg47His(MslI)	Arg47His Taq	EcoRI	EcoRI Taq	Ile349Val(SspI)	Ile349Val Taq	Arg369Cys(AlwNI)	Arg369Cys Taq	RsaI	RsaI Taq	StyI Taq
MC	3	(+/+)	C/C	(+/-)	C/T	(+/+)	A/A	(+/+)	T/T	(-/-)	G/G	(+/+)	A/A	G/C
	4	(+/+)	C/C	(-/-)	C/C	(+/+)	A/A	(+/+)	T/T	(-/-)	G/G	(-/-)	G/G	G/C
	6	(-/-)	C/T	(+/-)	C/T	(+/+)	A/A	(+/-)	C/T	(-/-)	G/G	(+/-)	A/G	G/C
	7					(+/	A/	(+/+)		(-/		(+/		
	11					(+/	A/	(+/+)		(-/		(+/		
	13	(+/-)	C/T	(+/-)	C/T	(+/-)	A/C	(+/-)	C/T	(-/-)	G/G	(+/-)	A/G	G/C
	14	(+/+)				(+/	A/	(+/+)		(-/		(+/		
	19	(+/-)	C/T	(-/-)	C/C	(+/-)	A/C	(+/-)	C/T	(-/-)	G/G	(-/-)	G/G	G/C
	22	(+/+)	C/C	(-/-)	C/C	(+/+)	A/A		T/T	(-/-)	G/G	(+/+)	A/A	C/C
	25	(+/+)	C/C	(+/-)	C/T	(+/+)	A/A	(+/+)	T/T	(-/-)	G/G	(+/-)	A/G	C/C
	27	(+/+)	C/C	(+/+)	T/T	(+/+)	A/A	(+/+)	T/T	(-/-)	G/G	(+/+)	A/A	C/C
	29	(+/-)		(+/-		(+/	A/	(+/+)				(+/		
	31	(+/-)		(+/-		(+/	A/			(-/		(+/		
	32	(+/+)	C/C	(+/-)	C/T	(+/+)	A/A	(+/+)	T/T	(-/-)	G/G	(+/-)	A/G	G/C
	33	(+/+)	C/C	(+/+)	T/T	(+/+)	A/A		T/T		G/G	(+/+)	A/A	C/C
	37		C/C		C/C	(+/+)	A/A	(+/+)	T/T		G/G	(-/-)	A/G	G/C
	40	(+/+)	C/C	(+/-)	C/T	(+/+)	A/A	(+/+)	T/T		G/G		A/G	C/C
	42	(-/-)	T/T	(-/-)	C/C	(-/-)	A/C	(-/-)	C/C	(-/-)	G/G	(-/-)	A/G	G/C
	45	(+/+)		(+/-		(+/	A/			(-/		(+/		
	52	(+/+)				(+/	A/					(-/	G/	
	53	(+/+)	C/C		C/C	(+/	A/		T/T		G/G	(-/		
	55	(-/-)	C/T	(-/-)	C/C	(+/+)	A/A	(+/-)	C/T	(-/-)	G/G	(-/-)	G/G	C/C
	57		C/T		C/C		A/A	(+/+)	T/T	(-/-)	G/G		G/G	
	64	(+/-)	C/T	(+/-)	C/T	(+/-)	A/C	(+/+)	C/T		G/G	(-/-)	G/G	G/C
	65	(+/+)	C/C	(-/-)	C/C	(+/+)	A/A	(+/+)	T/T		G/G	(+/-)	A/G	G/C
	66	(+/-)	C/T	(-/-)	C/C	(+/+)	A/A	(+/-)	C/T		G/G	(+/-)	A/G	G/G
	68	(+/+)	C/C	(-/-)	C/C	(+/+)	A/A	(+/+)	C/T	(-/-)	G/G	(+/-)	A/G	G/G
	69	(+/+)	T/T ?	(-/-)	C/C	(+/+)	A/A	(+/+)	T/T	(-/-)	G/G	(+/-)	A/G	G/G
	71	(+/+)	C/C	(-/-)	C/C	(+/+)	A/A	(+/+)	T/T		G/G	(-/-)	G/G	C/C
	73	(+/+)	C/T	(+/+)	T/T	(+/+)	A/A	(+/+)	T/T	(-/-)	G/G	(+/+)	A/A	C/C
	74	(+/+)	C/C	(+/-)	C/C	(+/+)	A/A	(+/+)	T/T	(-/-)	G/G	(+/-)	A/G	G/G
	78			(-/-		(+/	A/					(+/		
	79		C/C		C/C	(+/+)	A/A		T/T		G/G	(+/-)	A/A	G/C
	81	(+/+)	C/C	(+/-)	C/T	(+/+)	A/A	(+/+)	T/T	(-/-)	G/G	(+/-)	A/G	G/C
	83	(-/-)	T/T	(-/-)	C/C	(-/-)	C/C	(-/-)	C/C	(-/-)	G/G	(-/-)	G/G	C/C
	87	(+/+)	C/C	(+/-)	C/T	(+/+)	A/A	(+/+)	T/T	(-/-)	G/G	(+/-)	A/G	G/C
	94	(+/+)				(+/	A/			(-/-		(+/-		

	96	(+/+)	C/C	(+/-)	C/T	(+/+)	A/A		T/T	(-/-)	G/G	(+/-)	G/G	G/C
	99	(+/+)	C/T	(+/-)	C/T	(+/+)	A/A	(+/-)	C/T	(-/-)	G/G	(-/-)	A/G	G/C
	104	(+/-)	C/T	(-/-)	C/C	(+/-)	A/C	(+/-)	C/T	(-/-)	G/G	(-/-)	G/G	G/C
	105	(+/-)	C/T	(-/-)	C/C	(+/+)	A/A	(+/-)	C/T	(-/-)	G/G	(-/-)		C/C
	113		C/T		C/C	(+/+)	A/A	(+/-)	C/T	(-/-)	G/G		G/G	C/C
	116		C/C		C/T	(+/+)	A/A		T/T	(-/-)	G/G	(-/-)	A/G	G/G
	119	(+/-)	C/T	(-/-)	C/C	(+/+)	A/A	(+/-)	C/T	(-/-)	G/G	(-/-)	G/G	G/G
MC	125	(+/-)	C/C		C/T	(+/+)	A/A	(+/+)	T/T		G/G	(-/-)	A/G	C/C
	130	(+/-)		(+/-)	C/C	(+/-)	A/C	(+/-)	T/T	(-/-)	G/G	(-/-)		C/C
	136	(+/+)	C/C	(-/-)	C/C	(+/+)	A/A	(+/+)	T/T	(-/-)	G/G	(-/-)	A/A	G/G
	137	(+/+)	C/C	(-/-)	C/T	(+/+)	A/A	(+/+)	T/T	(-/-)	G/G	(-/-)	G/G	G/C
	138	(+/+)	C/C	(+/-)	C/C	(+/+)	A/A	(+/+)	T/T	(-/-)	G/G		A/G	G/C
	143	(+/+)	T/T		C/C		A/A		T/T		G/G	(-/-)	G/G	C/C
	145	(+/+)	C/C		C/T		A/A		T/T		G/G		A/G	G/C
	148	(-/-)					A/C			(-/-)		(+/-)		C/C
	149	(+/-)			T/T		A/C			(-/-)		(+/-)		
	151	(+/+)	C/C		T/T		A/A		T/T	(-/-)	G/G	(+/-)	A/A	C/C
	156		C/T		C/T		A/C		C/T	(-/-)	G/G	(+/-)	A/G	G/C
	161	(+/+)	C/C		C/T	(+/+)	A/A	(+/+)	T/T	(-/-)	G/G	(+/-)	G/G	G/C
	165	(+/-)	C/C		C/T		A/A		T/T	(-/-)	G/G	(+/+)	A/G	G/G
	168	(+/+)	C/C		C/T	(+/+)	A/A	(+/+)	T/T	(-/-)	G/G	(+/-)	A/G	G/C
	169	(-/-)	C/T		C/C	(-/-)	A/A		T/T	(-/-)	G/G		A/G	G/G
	170	(+/-)	C/T		C/T	(+/-)	A/C	(+/-)	C/T	(-/-)	G/G	(+/-)	A/G	C/C
	171	(+/-)	C/T		C/C	(+/+)	A/A	(+/-)	C/T	(-/-)	G/G	(-/-)	G/G	G/G
	172	(+/+)	C/C		C/T	(+/+)	A/A	(+/+)	T/T	(-/-)	G/G	(-/-)	G/G	G/G
	176	(+/-)	C/T		C/T	(+/-)	A/C	(-/-)	C/T	(-/-)	G/G	(-/-)	A/G	G/G
	186		C/C		C/C		A/A		T/T		G/G		A/G	C/C
	187	(-/-)	C/T		C/C		A/C	(+/-)	C/T	(-/-)	G/G	(-/-)	G/G	G/C
	189		C/C		C/T		A/A		T/T	(-/-)	G/G		G/G	G/C
	192		C/C		T/T		A/A		T/T	(-/-)	G/G		A/A	G/C
	193		C/T		C/C		A/A		C/T	(-/-)	G/G		G/G	G/C
	199		C/T		C/T		A/C		C/T	(-/-)	G/G		A/G	G/C
	201		C/C		C/C		A/A		T/T	(-/-)	G/G		G/G	G/C
	203		C/C		C/C		A/A		T/T	(-/-)	G/G		A/G	G/G
	204		C/C		C/T		A/A		T/T	(-/-)	G/G		A/G	G/C
	206		C/C		C/T		A/A		T/T	(-/-)	G/G		G/G	G/G?
	207		C/T		C/T		A/A		C/T	(-/-)	G/G		A/G?	C/C
	210		C/T		C/T		A/A		C/T	(-/-)	G/G		A/G	C/C
	212		C/T		C/T		A/C		C/T	(-/-)	G/G		A/G	G/G
	213		C/T		C/C		A/A		C/T	(-/-)	G/G		G/G	G/C
	218		C/C		T/T		A/A		T/T	(-/-)	G/G		A/A	G/C
	222		C/C		C/T		A/A		T/T	(-/-)	G/G		G/G	C/C
	228		C/T		C/C		A/A		C/T	(-/-)	G/G		G/G	G/C
	230		C/C		C/T		A/A		T/T	(-/-)	G/G		A/G	C/C
	233		T/T		C/C		C/C		C/C	(-/-)	G/G		G/G	C/C
MC	236		T/T		C/T		A/C		C/C	(-/-)	G/G		G/G	C/C
	238		C/C		C/T		A/A		T/T	(-/-)	G/G		A/A	C/C
	240		C/T		C/C		A/A		C/T	(-/-)	G/G		G/G	G/C
	242		C/C		T/T		A/A		T/T	(-/-)	G/G		A/A	C/C
	244		C/C		C/T		A/A		T/T	(-/-)	G/G		A/G	G/C
	248		C/C		C/T		A/A		T/T	(-/-)	G/G		G/G	G/G
	250		C/C		C/C		A/A		T/T		G/G		A/G	G/C
	251		C/C		C/C		A/A		T/T		G/G		G/G	C/C
	256		C/C ?		C/T				T/T		G/G		A/G	
	257		C/T		C/C		A/C		C/T		G/G		G/G	G/C

Assay ID		ss2978361		ss2978360		ss2978361		ss2978363		ss2978364		ss2978359		ss2978365
		ADH1C		ADH1B		ADH1C		ADH1C		ADH1B		ADH1B		ADH7
Position:	Chr. 4	1E+08		100239319		100268856		100260789		100229017		1E+08		100340522
Ancestra l		(-)	T	(-)	C	(+)	A	(-)	C	(-)	G	(-)	G	C
Village	Sample	**HaeIII**	**HaeIII Taq**	**Arg47His(MslI)**	**47His Taq**	**EcoRI**	**EcoRI Taq**	**Ile349Val(SspI)**	**Ile349Val Taq**	**Arg369Cys(AlwNI)**	**Arg369Cys Taq**	**RsaI**	**RsaI Taq**	**StyI Taq**
NALT	1	(+/-)	T/T	(-/-)	C/	-/-)	A/A	(+/-)	T/T	(-/-)	G/G	(+/+)	A/G	G/G
	2	(+/+)	C/C	(+/-)	T/	+/+)	A/A	(+/+)	T/T	(-/-)	G/G	(+/+)	A/G	G/G
	3	(+/-)	T/C	(-/-)	T/	+/-)	A/C	(+/-)	T/C	(-/-)	G/G	(+/+)	A/G	G/C
	4	(+/-)	C/C	(-/-)	C/	+)	A/A	(+/-)	T/T	(-/-)	G/G	(-/-)	G/G	G/C
	5	(+/+)	C/C	(-/-)	C/	+)	A/A	(+/+)	T/T	(-/-)	G/G		G/G	G/C
	6	(+/-)	T/C	(+/-)	C/	+)	A/C	(+/-)	T/T	(-/-)	G/G		G/G	G/C
	7	(+/+)	T/C	(-/-)	C/	+)	A/A	(+/+)	T/T	(-/-)	G/G	(-/-)	G/G	G/C
	8	(+/+)	C/C	(-/-)	T/	+)	A/A	(+/+)	T/T	(-/-)	G/G	(+/-)	A/A	G/C
	9	(+/-)	T/T	(-/-)	T/	+/-)	A/C	(+/-)	T/T	(-/-)	G/G		G/G	G/C
	10	(+/-)	T/C	(+/-)	T/	+)	A/A	(+/-)	T/T	(-/-)	G/G	(+/-)	A/A	G/C
	12	(+/-)	T/C	(-/-)	C/	-/-)	A/C	(+/-)	T/T	(-/-)	G/G		G/G	G/C
	13	(+/+)	T/C	(+/-)	C/	/+)	A/A	(+/+)	T/T	(-/-)	G/G		A/G	C/C
	14	(+/+)	T/C	(+/-)	C/	/+)	A/C	(+/+)	T/C	(-/-)	G/G	(+/+)	A/G	G/C
	15	(-/-)	T/C	(-/-)	T/	/-)	A/A	(-/-)	T/T	(-/-)	G/G	(+/+)	A/A	G/C
	16	(+/+)	C/C	(-/-)	C/	+)	A/A	(+/+)	T/T	(-/-)	G/G	(+/+)	A/G	G/C
	18	(+/-)	T/C	(-/-)	T/	+)	C/C	(+/-)	T/T	(-/-)	G/G		G/G	C/C
	19	(+/+)	T/C	(-/-)	C	+)	A/C	(+/+)	T/C	(-/-)	G/G	(+/+)	A/G	G/G
	20	(+/-)	T/T	(-/-)	T/	+)	A/A	(+/+)	T/T	(-/-)	G/G	(+/+)	A/G	G/G
	22	(+/-)	T/C	(-/-)	C/	-)	A/A	(-/-)	T/C	(-/-)	G/G		A/G	C/C
	23	(+/-)	T/C	(-/-)	T/	/+)	A/A	(+/-)	T/T	(-/-)	G/G			
	25	(+/-)	T/C	(-/-)	C	/+)	A/A	(+/-)	T/C	(-/-)	G/G		A/G	G/C
	26	(+/-)	T/C	(-/-)	T/	/+)	A/A	(+/-)	T/C	(-/-)	G/G		G/G	G/G
	27	(+/-)	T/C	(+/-)	T	+)	A/A	(+/-)	T/T	(-/-)	G/G	(-/-)	G/G	C/C
	28	(+/-)	T/C	(-/-)	C/	/+)		(+/-)	T/C	(-/-)	G/G	(-/-)	G/G	G/C
	29	(+/+)	T/C	(-/-)	C/	+)	A/A	(+/+)	T/C	(-/-)	G/G		A/G	G/C
	30	(+/+)	T/T	(-/-)	T	+)	A/A	(+/+)	T/T	(-/-)	G/G		A/G	
	31	(-/-)	T/T	(-/-)	C/	+)	A/A	(-/-)	T/C	(-/-)	G/G		A/G	G/C
	32	(-/-)	T/T	(-/-)	T	+)		(-/-)	T/T	(-/-)	G/G	(+/+)	A/G	G/G
	33	(+/-)	T/C	(-/-)	C/	/+)	A/C	(+/-)	T/C	(-/-)	G/G		G/G	C/C
	34	(+/-)	T/C	(-/-)	C/C	(+/+)	A/C	(+/+)	T/C	(-/-)	G/G	(+/+)	A/G	G/C
	35	(+/-)	T/C	(-/-)	C/C	/-)	C/C	(+/-)	T/T	(-/-)	G/G	(-/-)	G/G	G/C
	36	(+/-)	T/C	(-/-)	T/	/-)	A/A	(+/-)	T/C	(-/-)	G/G		A/G	G/C
	37	(+/-)	C/C	(-/-)	T/	-/-)	A/A	(+/-)	T/T	(-/-)	G/G		A/A	G/C
	38	(-/-)	C/C			/+)	A/A	(+/-)	T/T	(-/-)	G/G	(-/-)	G/G	G/C
	39	(+/-)	T/T	(+/-)	T	/+)	A/A	(+/-)	T/T	(-/-)	G/G	(+/-)	A/A	G/C
	40	(+/+)	T/C		T	+)	A/A		T/C	(-/-)	G/G	(+/+)	A/G	G/C
	41	(+/+)	T/C	(-/-)	T	+)	A/A	(+/+)	T/T	(-/-)	G/G	(+/+)	A/G	G/G
	42	(+/+)	C/C	(-/-)		+)	A/A	(+/+)		(-/-)		(-/-)	G/G	G/C
	43	(+/+)	C/C		C	+)	A/A	(+/+)	T/T	(-/-)	G/G		A/G	G/C
	44	(+/-)	C/C		T	+)	A/A	(+/-)	T/T	(-/-)	G/G		A/G	G/G
	45	(-/-)	C/C	(-/-)	T/C	-/-)	A/A	(-/-)	T/T	(-/-)	G/G		G/G	G/C

	46		C/C		C		A/A		T/T	(-/-)	G/G		G/G	G/C
	47		C/C		C		A/A		T/T	(-/-)	G/G	(-/-)	G/G	G/C
	48		C/C		T		A/A		T/T	(-/-)	G/G	(-/-)	G/G	G/C
	49		C/C		C		A/A		T/T	(-/-)	G/G	(-/-)	G/G	C/C
	50		T/C		C		A/C		T/C	(-/-)	G/G		A/G	G/G
	51		T/C		T		A/C		T/C	(-/-)	G/G		A/G	G/C
	52		T/C		T		A/A		T/T	(-/-)	G/G		A/G	G/C
	53		T/C		T		A/A		T/T	(-/-)	G/G		G/G	G/C
	54		C/C		C		A/A		T/T	(-/-)	G/G		[illegible]	G/G
	55		C/C		T		A/A		T/C	(-/-)	G/G		A/G	G/G
	56		[illegible]				A/A		T/C	(-/-)	[illegible]		G/G	G
	57		T/C		C		[illegible]		T/T	(-/-)	G/G		A/G	G/G
	58		C/C		T		[illegible]		T/T	(-/-)	G/G		G/G	G/C
	59		C/C		T		[illegible]		C/C	(-/-)	G/G		A/A	G/
	60		T/C		T		[illegible]		T/C	(-/-)	G/G		A/G	G/G
	61		T/C		C		[illegible]		T/C	(-/-)	G/G		G/G	G/C
	63		T/T		T		[illegible]		T/C	(-/-)	G/G		A/G	G/G?
	64		T/C		T		[illegible]		T/C	(-/-)	G/G		G/G	G/G
	66		T/C		C		[illegible]		T/C	(-/-)	G/G		A/G?	G/C
	67		C/C		T		[illegible]		T/T	(-/-)	G/G		G/G	G/C
	68		C/C		C		[illegible]		T/T	(-/-)	G/G		[illegible]	G/G
	69		C/C		C		A/A		T/T	(-/-)	G/G		A/A	G/G
	70		T/C		C		[illegible]		T/C	(-/-)	G/G		A/G	G/G
	71		T/C		[illegible]		[illegible]		T/C	(-/-)	G/G		A/G?	G/C
	72		C/C		C		[illegible]		T/T	(-/-)	G/G		A/G	G/C
	73		C/C		C		[illegible]		T/T	(-/-)	G/G		G/G	G/C
	74		C/C		[illegible]		[illegible]		T/T	(-/-)	G/G		G/G	G/C
	75		C/C		T		[illegible]		T/T	(-/-)	G/G		A/G	C/C
	77		T/C		C		[illegible]		T/C	(-/-)	G/G		A/G	G/G
	78		C/C		T		[illegible]		T/T	(-/-)	G/G		A/G	G/G
	79		C/C		T		[illegible]		T/T	(-/-)	G/G		G/G	G/G
	80		C/C		T		[illegible]		T/T	(-/-)	G/G		G/G	G/G
	81		T/C		C	C	[illegible]		T/T	(-/-)	G/G		[illegible]	G/C?
	83		T/C		C	C	[illegible]		T/C	(-/-)	G/G		G/G	G/G
	84		T/C		C		[illegible]		T/T	(-/-)	G/G		G/G	G/G
	85		T/T		C		[illegible]		[illegible]	(-/-)	G/G		[illegible]	G/C
	86		T/T		C		[illegible]		C/C	(-/-)	G/G		A/G?	C/C
	92		C/C		T		[illegible]		T/T	(-/-)	G/G		[illegible]	[illegible]
	93		C/C		C		[illegible]		T/T	(-/-)	G/G		A/G	G/C
	94		C/C		T		[illegible]		T/T	(-/-)	G/G		A/A	G/G
	95		T/C		C		[illegible]		T/C	(-/-)	G/G		A/G	G/C
	96		T/C		C		[illegible]		T/C	(-/-)	G/G		A/G?	G/C
	97		C/C		T		[illegible]		T/T	(-/-)	G/G		A/G	G/C
	98		C/C		T	C	[illegible]		T/T	(-/-)	G/G		A/G	G/C
	99		C/C		C	C	[illegible]		T/T	(-/-)	G/G		A/G?	G/G
	100		C/C		[illegible]		[illegible]		T/T	(-/-)	G/G		A/G	G/G
	111		T/C		[illegible]		[illegible]		T/C	(-/-)	G/G		A/G	G/C