Spring 2007

The Genetic History of the Karachay

Aslihan Sen
University of Pennsylvania

The Genetic History of the Karachay

Disciplines

Anthropology

THE GENETIC HISTORY OF THE KARACHAY

By

Aslihan Sen

AN UNDERGRADUATE RESEARCH PAPER

In

Anthropology

Submitted to the
Department of Anthropology
University of Pennsylvania

Thesis Advisor: Theodore G. Schurr

2007

Table of Contents *page*

Index of Supplementary Material

Abstract

The Karachay-Malkar population of the Northwestern Caucasus Mountains has an unclear history. Oral traditions say that they are descended from Alans, who were ancient Iranian tribes. The language they speak is a Kipchak Turkic language, and was supposedly brought to the Caucasus by the invading Qumans from the Minusinsk Basin (Yenisei River-Altai Mountains). They are also supposedly related to the Bulgars, and the name Malkar/Balkar is evidence for this affiliation.

To elucidate their genetic past, this study compares the frequency of Karachay mitochondrial DNA haplogroups from 67 individuals with previously published results from other populations nearby. In general, it was hypothesized that the Karachays would be more similar to other Caucasus populations and Central Asian Turkic populations.

Results indicate that the most similar population was indeed a Caucasus population, the Adygei, but other Caucasus populations and geographically close populations in the region were not similar in haplogroup frequency. Eastern Iranians were the next most similar, and indicate that the theory involving ancestry to the Scythio-Iranians (Alans) may be true. Also, Altaians showed similar haplogroup types, and somewhat similar frequencies. They might indicate the ancient Turkic origins of the Karachays from the Altay Mountains.

This study was helpful in clarifying some aspects of the Karachay past, but further research is needed in both the fields of mitochondrial DNA analysis and Y-chromosome analysis to identify paternally inherited genetic influences on the Karachay population.

I. Background

I. 1 History of Turkic Peoples and the Karachay-Malkar

The diverse clans and groups that represent the Turkic peoples have dispersed from their original homeland in the Altay Mountains in present-day Mongolia and Kazakhstan. Turkic peoples now inhabit most of the Central Asian steppes and Eurasia west to Turkey. The earliest known records of Turks, the Huns, come from China in the 3rd century BC (Menges, 1968, 17).

The Huns are well known as the tribes who warred against the Chinese Empires for 500 years after this date. They had begun westward expansion in the first century BC, and became a threat to the West and the Roman Empire in the 4th century AD. Around this time also, other Hunnic tribes arrived on the steppes between the Danube and Volga. They were the Saraguroi, Urogoi, and Onoguroi, and were mentioned by Byzantine historians at about 463 AD. These tribes joined the Hun confederacy led by Irnak, Attila's son, and after 482 AD, the confederacy was called Bulgaroi by historians, and the name has led to some confusion about their ethnicity (Menges, 1968, 18-9).

The name Bulgar is connected historically to Bulgaria, as we shall see later, but it is also related to Malkar or Balkar (m and b interchangeable phonetically in the language). This could be a kinship relation or it could also be just a coincidence, since the Balkar/Malkar live near the Malqa River in the Caucasus. The Malkar are ethnically the same as the Karachay and speak the same language, with a few phonetic differences, separated in the Caucasus by the mountains. They distinguish each other simply as if they are different clan names, and frequently inter-marry. The names Alan, Ars/As, Kipcak, and Bulgar are all related to the Karachay ethnic identity in different ways (vide infra).

The confederacy stretched from modern northeastern Bulgaria to the Kuban River area in the northwest Caucasus, which the Karachay people call home, and which was the center of the confederacy. It is unknown whether the Bulgaroi were Hunnic in origin or merely connected with the Hunnic confederacy, but a separate tribe. However, tribes under the name of Bulgaroi were among the first Turkic speaking immigrants in Eastern Europe (Menges, 1968, 20).

In 555 AD, the Mongolian Avars came from East Central Asia, following the same path as the Huns had earlier, and rode through Eurasia to the Balkans to conquer the Roman Empire. They probably were the first to introduce the title Qaghan (Kaan) to mean the chief or ruler, in place of the older term "Jabghu." Despite this, the ethnic and linguistic composition of the area does not seem to have changed, since the Avars were eager to move on and conquer (Menges, 1968, 20).

Meanwhile, in Central Asia, two other Turkic tribes united in the late 5th century to form the Turkut (plural of Turk) confederacy, as mentioned in Chinese sources. The Turkut expanded to the West and soon gained much power. They controlled the steppes between Kazakhstan and the Volga region, and came into contact with the Bulgaroi, Hunnic tribes, and Xazars, who were incorporated into the western Turkut confederacy (Menges, 1968, 21).

The Chinese defeated the Turkut in the east in 581, which caused the Turkut confederacy to divide into the Western Turkut, centered in Kazakhstan, and Eastern Turkut, or the Gokturk, centered in northern Mongolia (Menges, 1968, 22). From this split on, the history of the Karachay is mostly concerned with the Western Turks until the Middle Ages.

Before the Altaic invasions began in the 3rd century BC, there were many Iranian and Iranized tribes living in the Central Asian region around the Caspian Sea at different time periods, such as the Sogdians, Sarmatai, Medes, and Kurds. Among them, one comes across the

names Alanoi (Alans) and Arsoi (Ars) (Menges, 1968, 24-5). Karachays call each other "Alan" informally in personal conversation. For example, they would say, "Alan, where have you been?" Karachay oral tradition says that they have descended from the Alans or Ars (As), which means "real" or "original." However, if these nomadic tribes were supposedly Persian-speaking, it would be difficult to explain why the Karachays today speak a Kipchak Turkic language (as opposed to Oguz Turkic or other Turkic). The Alans may have been originally Turkic, and spoke a Turkic language. Some of the texts attributed to them have been analyzed and seem to make more sense if interpreted using the Karachay-Balkar language rather than Persian (Zakiev, 1995, 38-57).

Many of the tribes that made up the Western Turkut confederacy were Oguz tribes, and had a different dialect from other Turkic tribes and from those in the East. These Oguz tribes became more powerful and conquered more of the Middle East, forming the Seljuk Confederacy, and the Islamic Seljuk Empire (Menges, 1968, 26). The Oguz tribal groups in the Aralo-Caspian region began to occasionally clash with the Pecenegs, who were part of the later Northwest Turkic tribal confederacies, the Kipchaks, along with Bulgars, Xazars, and others. They were eventually incorporated into the Oguz confederacy and were all converted to Islam by the end of the 11th century (Menges, 1968, 27).

When Seljuk power declined in the Western part of the empire into small regional dynasties, one of them from Anatolia, the Osman tribe, eventually produced the leaders who conquered the Byzantine Empire in 1453 to form the Ottoman Empire (Menges, 1968, 28).

The above mentioned Oguz tribes were the progenitors of the modern Turkic populations of Turkmenistan, Iran, Trans-Caucasia, Anatolia and the Balkans. These peoples show physical

features of admixture with the local Iranian, Arab, and other populations, unlike the other Altaic peoples (Menges, 1968, 21-9).

Going back further in time, in the 7th century, the Xazars, tribes who represented the westernmost of the Western Turkut, became independent and moved into the Kuban River, where Karachays are located, and even the Dnepr River area. They ruled their state using ancient Turkic traditions such as the double rule of Qagans (Chiefs) and Beys (Local Princes). They kept other religious influences out of their area until about 740, when they accepted Judaism as the state religion (Menges, 1968, 29). The Xazars had control over all of Eastern Europe from the Dnepr to the Ural River from 680 until they were defeated by the Turkic Qumans in 1030, exhibiting control over the Normano-Russian Great Princes of Kijev as well as the Volga Bulgars (Menges, 1968, 30).

The Xazars were not uniformly Turkic. They had various groups of different ethnic and linguistic affiliation. It is said that there were two distinct groups within these. One group was dark skinned (called Black Bulgaria) and spoke a language different from Turks but similar to Bulgars, possibly Hunnic, Hunno-Bulgarian, or Bulgaro-Turkic. The other was fair skinned and possibly Finno-Ugric or Ugric and Proto-Hungarian. This is confusing, though, since some Turkic groups were also fair skinned, blue-eyes, and blonde-haired, like the Qumans. This could be a result of North-Iranian, Alanic, influence (Menges, 1968, 30-1). Today, the Karachays also have differing physical traits. Some look like the Central Asian Altaic peoples, who exhibit more Asian traits, and others have blond hair, blue eyes, and very fair skin!

During Xazar rule, there were two tribal movements of the Bulgars. A ruler named Qubrat, who was baptized in Constantinople in 619, was able to establish a center of rule in the Kuban area, and called it Great Bulgaria. Later, one of his sons migrated into western Byzantine

territory, and settled between the Balkans and the Danube. The Bulgarians spread out, and over time, their language, Altaic-Bulgarian, became more Slavic as their culture and religion also were influenced by Christianity and the Greek civilization. Today, however, the present -day Bulgarian populations exhibit "very strong features of the High Asiatic or Altaic racial type, much more so than their immediate and eastern neighbors, the Osman-speaking Oguz Turks of the Balkans and even Anatolia" (Menges, 1968, 31).

The second tribal migration was up the Volga River, facilitated by the Xazars, who wanted the Hunno-Bulgarian tribes in their realm further from their center of state. There, the Bulgars converted to Islam in 922 AD. When the Qumans and Pecenegs defeated the Xazars in the 11th century, they spread the Kipcak Turkic language in Eastern Europe. This is the Turkic language type that the Karachays speak.

In 1236, the Mongols of the Golden Horde conquered the Volga region, destroying the Hunno-Bulgar or Volga-Bulgar realm (Menges, 1968, 32).

It is well-known among the Karachays that this is also the conquest that forced the Karacay-Balkar to flee to the higher elevations to avoid conflict. The Mongols began the Turkicization of the area. They absorbed the Hunno-Bulgarians who were still living along the Kuban River, "Black Bulgaria", after the Xazars fell in the late 11th century. The Turkicized descendants of these "Black Bulgarians" are said to be the Karachay-Balkar in the Caucasus! The name Balkar could be, as stated earlier, a continuation of Bulgar or it could be related to the name of the Malqa River in the area (Menges, 1968, 32). Although the Mongols "Turkicized" the Bulgars, it is likely that the Karachay-Malkar were already speaking Kipcak Turkic, influenced by Qumans.

The history of the Qumans is complicated. The tribal name Quman is probably derived from the Turkic quba, meaning "yellow-red, reddish, yellowish," referring to the color of their hair, since they were fair skinned and haired, as previously stated, and which distinguished them from the other Altaic peoples (Menges, 1968, 35). Before moving west, the Qumans were originally part of a larger Mongolian tribal confederacy, the Qaj, whose older name was Komagqaj, an older version of "Quman." They lived in northeast Mongolia, just east of the Qirgiz (whose language today is one of the Turkic languages most similar to Karachay) (Menges, 1968, 35).

The Yenisej Qirgiz are named in Chinese sources the Gjan-kun (Kun is also the name of the Quman in Hungarian). These Qirgiz are also described by the Chinese as "green-eyed, fair-skinned, and blond," like the Quman. They are the only two Altaic groups which physically stand out in this way. Today, there are still people with these features in the Minusinsk Basin. It seems these un-Asiatic features are remnants of the far eastward migration of Eurapoid Scythian or Scythoid populations (perhaps even those Iranian Arsoi and Alanoi originally from the Caucasus) or of even older Paleo-European populations (Menges, 1968, 36). If the former theory is correct, it seems like the ancestors of the Karachay may have originally been from the Caucasus after all, and went eastward only to come back westward and settle in the Caucasus again as Qumans (**Figure 1**)!

There is also archaeological evidence of a Europoid migration of Scythian or later Iranian or Indo-European group into the Altay and Minusinsk region. Evidence has been found of Caucasian style metallurgy, "Animal Style." The Qirgiz who descended from these peoples also acquired the metallurgical culture, and their weapons and metal work were held in high esteem

by the Chinese. Traces of their industry can still be found in the forests and mountain of the Minusinsk region (Menges, 1968, 48-9).

In the Tien-Shan Mountains of Central Asia, these Qirgiz were able to preserve their culture and separate themselves from outside influence, just like the Karachay-Balkar preserved their culture and people in the Caucasus Mountains (Menges, 1968, 43). The mountains protected many small groups of people in this way, safeguarding languages and cultures to this day, though many are dying out quickly. Perhaps the ancestors of the Karachay, already having good experiences in the eastern mountains, knew that the best place to settle to save their culture was again in another mountainous country!

I. 2 The Karachay-Malkar People and Culture Today

Although many Karachay are devout Muslims today, it seems that their religious heritage has been varied and not well documented. It is clear, however, that they were originally pagans, as all the Turkic peoples have been, and worshipped deities of nature, with Teyri (Tanri) as the chief god.

The Karachay may have converted to Christianity, Judaism, and Islam back and forth throughout the history of the various peoples who inhabited the region, but they always kept elements of their pagan traditions alive (Musukaev and Musukaev, 2007). Even today, there is mention of Teyri Terek, the "Sacred Tree God" among the elder members of Karachay society. There are also still several superstitions and names referring to the deities and spiritual beings of nature that controlled the universe of the ancient Karachay.

The Karachay and Malkar were definitively Muslim by the 1800s, though, and by this time had mostly given up their former mixed religious beliefs, even the eating of pork

(Curtubayev, 2006). There is a story of one of my mother's family clans, Chomalari or informally, Chochkalari, which means "baby pigs." According to our oral history, they were supposedly very successful in the rearing of pigs, and when they were converted to Islam, one woman of this clan would eat pork in secret, declaring "May God curse those who separated us from you [referring to their beloved pigs!]"

Despite this story, it is well known in history that the Karachay and Malkar have been the most successful group in the Caucasus in developing hardy sheep and horse breeds, often having thousands at a time in their individual sheep flocks (Karca, 1954). They have even won medals for their good breeding stock in livestock shows in London, Russia, and elsewhere.

It is also interesting to note that the Caucasus may have been one of the earliest sites of domestication of sheep, if not the earliest, and a mitochondrial lineage not found elsewhere has been found in Karachay sheep, indicating the antiquity of domestication in the Caucasus (Tapio et. al. 2006). The Karachay/Malkar have long been producing high quality livestock as an integral part of the regional economy.

To be able to maintain these large numbers of sheep, Karachays employed many laborers, and distinguished them as the "kuls," or servant class. The wealthier owners of the stock were called "biy," or "bey" in Turkish, and the society was somewhat feudal in nature (Karca, 1954). It is said that one of the Karachay villages in Turkey consists almost entirely of those that were of the laboring class in the Caucasus homelands, and it has often been remarked that they "do things differently" than other Karacays. This type of class differentiation among the Karacays and their former "kul" has only recently become less pronounced, since the migrations to Turkey and then the United States. The following is a short account of how they came to those parts.

The Karachay and Malkar people had long been living in the mountains when the Russian Tsars decided to take over their land in the late 16th century. The Caucasus mountain peoples, especially the Muslim groups, defended their homelands bravely for hundreds of years, keeping the Russian armies at bay temporarily. Imam Shamil was one of the most famous leaders of the freedom loving mountaineers who rebelled against the Russian conquests.

However, the lengthy Caucasus Wars took their toll on the populations. The area of the north Caucasus by the Kuban River, the Circassian and Karachay lands, were the last to finally fall to Russian power in 1864. Tsar Nicholas I expelled and killed many of the native inhabitants who were left, including 600,000 "Circassians" (which probably comprises more than just Circassians, but also the other North Caucasus peoples in the same area, like Karachay-Balkar) and settled Russians in their place. In this time, many Karachays, Balkars, and others fled to Turkey and the Middle East to escape the repression. When Russia and Turkey started fighting again in 1877, revolts were revived, but were fruitless. In 1917, the Tsarist regime fell in the October revolution (Conquest, 1960, 2-9).

When the Soviets came into power, the Caucasian ethnic minorities were given promises of land and freedom, many of which were not kept (Conquest, 1960, 19). Various borders were drawn and dissolved, and drawn again. Different peoples were united and broken up, united with other groups, and so on.

In WWII, the Karachay/Balkars, Ingush, Chechens, and others were accused by Stalin of helping the Nazis and were deported from their lands by the thousands in cattle trains. Many thousands died of starvation, disease, and exposure, mostly old people and children. The Karachays alone lost over 25% of their potential population when at least 70,000 of them (nearly all of the total population of Karachays) were deported to many locations in Central Asia and

Siberia to work in concentration camps (Comins-Richmond, 2002). After these deportations, most returned to their homelands when access was returned to them, and some migrated to other countries like Turkey.

Today, there is a total Karacay/Balkar speaking world population of about 241,000 (Seegmiller, 1996). Most are in Karachay-Cherkessia, but there are significant numbers in Turkey and other countries of the Middle East, where they took refuge from persecution in the late 1800s and again after WWII.

From Turkey, a number of Karachay migrated to the United States beginning in the 1960s and 70s. They are centered in Paterson, NJ, and are the subjects of this genetic study. A few members of the population have come directly from the Caucasus as well. They constitute in NJ a closely-knit ethnic group which still maintains its language, traditional folkdances, and religious beliefs, rarely marrying with non-Karachays, though there are several cases. There may also have been intermarriage with Turks in Turkey, though this is also probably not too common, since the Karachays settled in separate villages in Turkey, where the entire village population was Karachay. It is still frowned upon to marry a Turk, rather than a Karachay, let alone someone of a different language group and religion, mostly for reasons of good marriage relations based on common culture.

I. 3 History of Genetic Studies in the Caucasus

Many different ethnic and linguistic groups inhabit the Caucasus Mountains. The diversity and uniqueness of the languages is such that the Caucasus has been known as the Mountain of Tongues (Colarusso, 1992, 1). The origins of these groups of people have been recorded in their oral histories, but over time, discrepancies, mixing of accounts, and loss of oral

history has led to an unclear picture of how each group came to settle in these mountains. Several of the groups trace their ancestry to the Alans, about whom little is known. Some sources describe Alans as Turkic, while others say they are Iranian in origin.

For these reasons, people have focused on the genetics of Caucasus populations in a few studies. Early ones were conducted using blood groups, immunologic, and biochemical markers. There were several of these done on the Karachay, Balkar, and nearby populations in the early 1990s (Nasidze et. al. 1990). Some studies indicated that these populations were not restricted by the mountains or geographic barriers in gene flow, but more through linguistic or ethnic subdivision. However, those distinctions were sometimes superficial, and there were patterns of language replacement not associated with genetic and migratory movements (Barbujani et. al. 1994). Other studies showed that the Caucasus populations are not descendants of Neolithic farmers who expanded across Europe, but may be earlier migrants (Bulayeva et. al. 2003).

Recently there have been quite a few genetic studies on the Caucasus, but none have solid evidence on the Karachays, or if they do, they are inaccessible to me (in Russian). The literature that I found had very few Karachay samples and did not list the important sequence data. Therefore, the interesting history and lack of genetic studies make Karachays good candidates for mitochondrial genetic research.

I.4 Specific Aims of My Study

Mitochondrial DNA is passed down only through the mother of each person. It is not in the nucleus of each cell, but in the tiny mitochondria in the cytoplasm that produce energy for the cell. Instead of being in chromosome form, as nuclear DNA is, mtDNA is circular and only

about 16,500 base pairs long. The length, maternal inheritance, and lack of recombination make mtDNA ideal for quick molecular analyses.

I used this mitochondrial DNA to study the maternal gene pool of the Karachays, comparing my results to those of previous studies involving populations in the Caucasus, Russia, the Middle East, and Central Asia. By looking at the frequencies of West and East Eurasian maternal haplogroups in each of the populations and the Karachays, I was able to draw some conclusions about which of them have influenced the Karachay population more than others (relative genetic contributions), and where geographically these influencing factors may have come from. My historical analysis will also be helpful in this regard.

II. Research Design and Methods

II. 1 My Discovery of Molecular Anthropology and the Beginning of My Research

Before stumbling upon a certain research poster on a Penn Museum wall, I had no idea it was possible to trace the genetic history of a population using DNA. The research was Dr. Schurr's, and it was about Native Americans' genetic relationship with Altaian peoples. I had known from our oral history that my ancestors were Turkic, and they had originally come from the Altay Mountains, so I felt an immediate personal connection to this unique research. I wondered if the same kind of study could be done on my particular community of ethnic Karachays in northern NJ. I was satisfied with an answer within a few moments, noticing that Dr. Schurr's office was just in front of the poster. I knocked on the door to discover him inside. He was kind enough to invite me in and listened to my enthusiastic plea to do this study with him. I was immediately comforted to know that I would soon be starting my spur of the moment research!

Dr. Schurr and I wrote up a proposal for my study to be reviewed by the University of Pennsylvania Institutional Review Board. Soon after, the approval was issued, and I began my project.

That summer, I started collecting samples for my project. Because I am a member of a closely-knit Karachay community in northern NJ, I was able to relatively easily approach people for sample donation. Karachays are easily offended and perplexed by non-direct contact from people they know personally, so I had to meet people in person to collect samples, which made the process more accurate and secure as well. Because social rules about talking to unrelated members of the opposite sex are so strict in our community, my mother and father were a great help in the collection of samples. Also, that they are older and more well-known than I am was a benefit.

In general, our closer relatives were eager to donate, while people who were less related to my family were more suspicious. Relatives are considered to be those who have a common ancestor from up to seven generations back. Thus, the sample may have been somewhat biased, but I do not think it had a large impact. The final results show that the samples, from 67 participants, were diverse.

Because most of the people I was sampling had no idea what DNA even was, or how genetics could be used to give information about their origins, I explained to each of them personally what the project actually meant. I asked them about themselves and their families (date of birth, birthplace, ethnicity) as far back in generations as they could remember. Most could only remember up to their grandparents. I wrote this information down myself, and also wrote down to which clans their relatives belonged (Appendix I-Genealogical Form). There are

many clans in the Karachay, perhaps over thirty, but most of the participants seemed to belong to the major ones.

This was not a problem because marriage rules in the Karachay community dictate that one can only marry someone if they had no common ancestors for seven generations back. In other words, exogamy prevailed, and I would expect more diversity even if participants were representative of a few clans.

In addition to the interviews, participants read and signed Informed Consent forms (Appendix II) to protect their privacy, with the option to withdraw if they wanted to later. Their identities were kept anonymous in all parts of the research project; only my advisor, Dr. Schurr, and I have this information. The individual results I have uncovered will be sent to each participant, along with a general and anonymous analysis of the Karachay population as a whole.

II.2 Laboratory Methods and Procedures

Sample Preparation and DNA Extraction

I collected buccal (cheek) cells from each participant through a cheek scraping method using an Omni Swab applicator. The swabs containing the buccal cells were ejected into small cryovials containing 2.0 ml sterile 1X TE buffer, pH 8.0, and vial caps wrapped in parafilm to prevent loss of buffer volume.

At the molecular anthropology lab at Penn, I vortexed the buccal samples to resuspend the cells in solution, and then centrifuged (15 min, 14K rpm) the buffer solution with suspended cells to collect the cells in the pellet at the bottom. The cells were lysed with detergent and the DNA was extracted with phenol-chloroform (Maniatis et. al. 1982). Purification with ethanol

followed, and dried DNA samples were re-suspended in water as stock solutions. The stocks were used to create 1:10 dilutions that were used in all the following tests.

Mitochondrial DNA Analysis

The mtDNA studies involved single nucleotide polymorphism (SNP) analysis and control region (CR) sequencing to characterize maternal haplogroups (more general) and haplotypes (specific lineages within haplogroups) (Schurr et al. 1999). Previous studies have resulted in the knowledge of a universal mitochondrial tree, in which the many individual sequences of mutations have diverged from a single mitochondrial "Eve." Haplogroups are those sequences (haplotypes) that have common originating points of divergence, and are grouped together phylogenetically. Haplogroups and subhaplogroups are defined using letter and number designations.

Hypervariable region I (HVR1) (nucleotide positions 16000-16400) of the control region (CR) was sequenced using the primers and polymerase chain reaction (PCR) conditions reported in Schurr et al. (1999). HVS-I sequencing provides an individual pattern of mutations that have occurred in this region of the mitochondrial genome within each individual's mtDNA, and also identifies key nucleotide mutations that define its broader mtDNA "haplogroup." These defining changes have been characterized from extensive previous studies.

All sequencing reactions were read on ABI 3130x1 DNA Analyzers in the Laboratory of Molecular Anthropology, and the resulting sequences aligned and compared using the Sequencher 3.1 software tool (Gene Codes Corporation).

The SNP analysis gives information about whether a particular mutation that is expected for a specific haplogroup (from previous published work) is present or not. The region where the

mutation is expected is amplified using PCR and the mutation is discovered by trying to cut it with restriction endonucleases, which cut at known base pair sequences. Based on whether the target mutation is present, the resulting base pair pattern either conforms to this cut location or not, and the DNA is cut accordingly. The restriction enzyme digested samples are run on agarose gels to separate the different lengths of DNA fragments. The number of fragments, and their size indicates whether the enzyme cut or not, and thus whether the mutation which defines the haplogroup is present or not.

The SNP analysis involves screening Karachay samples for mutations that define West Eurasian [H-K, N, R, T-X] (Torroni et al. 1994b, 1996, 1998; Macaulay et al. 1999), East Eurasian [A-G, M, Y, Z] (Torroni et al. 1993, 1994a; Schurr et al. 1999; Kivisild et al. 2002), and Indian (Kivisild et al. 1999, 2003; Bamshad et al. 2001; Palanichamy et al. 2004) haplogroups (lineages). These are the ones most pertinent to the genetic history of the Karachay.

Phylogenetic Analysis

I constructed a tree using the median-joining network method (Bandelt et al. 1995, 1999). This allowed me to generate phylogenies or networks of mtDNA haplotypes (the specific sequences of mutations within haplogroups) that reveal the phylogenetic connections between each of the haplotypes present in the Karachays.

To assess different population origin models, using the relative genetic contributions of West and East Eurasian source populations to the Karachays, I compared my data with published data to see what portion of the Karachay's genotypes come from a hypothesized source population, or which surrounding populations could have similar roots. Data from the putative source populations was taken from the published literature (e.g., Bermisheva et. al. 2004;

Derenko et al. 2003; Malyarchuk et. al. 2002a, 2002b; Al-Zahery et al. 2003).

III. Results and Analysis

III. 1 The Karachays

When I had obtained the genealogical information from participants, I had asked about their birth locations and their parents' and grandparents' birth locations. This would clarify whether there may have been intermarriage with non-Karachays in their history. Almost all of the participants had been born in Turkey and immigrated to the United States later in life, so the fact that they were collected in the US had no bearing on their genetic history. The participants born in Turkey were also mostly born in Karachay villages, furthering the hypothesis that they have not intermarried with Turks in Turkey. The first in their families to have come from the Caucasus were their grandparents, who came in the 1870s-1890s. This has produced a sample set which is indicative of the genetic lineages that made it to Turkey when the Karachays fled from Russian persecution in the late 1800s. For the sake of clarity in **Figure 2**, I have placed their geographic location in the Caucasus, where present-day Karachay-Cherkessia exists.

All of the 67 samples collected were able to be amplified and analyzed using the methods described, except for one which could not be sequenced, but was successful for RFLP analysis. The resulting data yielded 31 different haplotypes; about every other person had a unique haplotype. These were grouped into broader haplogroups, and some into subhaplogroups. This information is in **Table 1**, and the visual representation of the table is **Figure 3**.

The network tree that was constructed using the specific sequences is shown in **Figure 4**. This represents only the Karachay population and is a specific form of the more general tree of mitochondrial diversity seen in **Figure 5** (Kivisild et. al. 2004). The haplotypes are

characterized by the sequence of mutations leading up to the yellow circles. The size of the yellow circles and their locations represent the number of samples with a given haplotype and where these particular haplotypes are found in the network.

The high number of haplotypes with respect to the sample size indicates that the sample is diverse. Also, some of the branches are very long, perhaps indicating old haplotypes that have not diversified in the recent past, meaning they could have been the result of cultural or geographical isolation (Schurr and Wallace, 2002). The maternal gene pool may thus not have been affected by outside influences, but the paternal side may give a different story. Further analysis would have to be done on the Y-chromosome to clarify this issue.

III.2 Haplogroup Diversity

The diversity of the haplogroups represented is used to understand the extent of the variation in the individual sequences found. The Karachays had representatives of only the M and R branches of the mtDNA tree, with only a single sample in the N macrohaplogroup family. These results indicate a lower diversity than other populations which have representatives of all macrohaplogroup families (**Figures 6a-6o**). However, the number of different U types is interesting, and indicates high diversity in this haplogroup. Although there was not a high percentage of M types, those that were present indicated a wide diversity as well, with the haplogroups D, C, and M represented.

Overall the Karachays had 9% M types, which are East Eurasian lineages as represented in the Eastern Eurasian network tree in **Figure 7** (Kivisild et. al. 2002). Only 2% were N types, and 89% were R types, both of which are West Eurasian haplogroups. Of all the haplogroups, the most represented was U, which was 41% of all samples, with U1 and U5 being the most

prevalent. The individual breakdowns of these percentages is shown in **Figure 3**.

III.3 Comparisons with Other Populations

To understand which populations may have contributed to the Karachay maternal gene pool, I compared the frequencies of haplogroups to other populations found in the published literature. I chose populations from the Caucasus, the Middle East and Iran (Scythian influence), Russia, and Central Asia (Turkic origins). Their specific haplogroup frequencies are represented individually in **Table 2**, and those which are the most unique are shown in **Figures 6a-6o** and geographically in **Figure 2**. To show more clearly where ethnic groups are located in the Caucasus, a separate map is given in **Figure 8**.

The most similar and most dissimilar populations deserve closer analysis. Overall, the Karachays seemed most similar to the Adygei, which are a Circassian population just west of the Karachay and who speak a non-Turkic language. Their cultural traditions and folk dances are similar to the Karachay, and most are Muslim as well. The frequencies of West Eurasian (U, K, H, HV, J, R, T, W) haplogroups is almost identical to that of the Karachays. The only significant differences are the absence of W in Karachays, and the absence of I and HV in Adygei, which are present in Karachays. The Adygei also have many more H's and less M's than Karachays, indicating more Western and less Eastern genetic influence. Overall, the similarities present an expected outcome, since the populations have very little geographic distance between them in the Caucasus, and significant admixture is likely.

The next most similar population in haplogroup frequency is the Eastern Iranian population. This was a surprise at first, since they are separated by a large geographic distance, but when the history of the Karachays as related to the Iranian Scythians was considered, the

outcome was supportive of this theory. The frequency in both populations of M types (Eastern Eurasian) is the same, 9%, while there are only slight differences in the frequencies of the West Eurasian types present. The presence of W (West Eurasian), A, and B (East Eurasian), and the absence of H and R (West Eurasian) in Eastern Iranians are the only discrepancies with the Karachay population haplogroups. Karachays have in place of these types more U's instead.

To test whether Karachays have affinities with other Turkic populations in Central Asia, where Turks originated, I compared with Bashirs (North of Kazakhstan) and Altaians (mountains NW of Mongolia and NE of Kazakhstan).

Bashkirs have all the same haplogroups present in Karachays, except for R, and some are even in the same frequencies (H, K, J, I), but Bashkirs have much more diversity, and have both more Western and Eastern Eurasian types that are not present in Karachays (V, W, F, A, B, N1, N9, Y). Also, the East Eurasian types (M) prevail in overall frequency, as expected, since they are closer to East Asia than Karachays. While Karachays have only 9% M's, Bashkirs have 32% M's.

Altaians were less diverse than Bashkirs, and more similar to Karachays in haplogroup *composition* but less in frequency to the Karachays. They had even more M's than Bashkirs (45%), but less T's and H's. In these ways they were more dissimilar to the Karachays than Bashkirs. This is expected, because the Altaians are closer to East Asia than the Bashkirs, who are in between Altaians and West Eurasia.

These results could indicate that the Turkic or Central Asian population from which some of the maternal ancestors of the Karachays originated could have been from the Altay region, but underwent significant admixture as they moved westward into the Caucasus. In addition, this admixture could have occurred when Iranian Scythians encountered the Qumans and Yenisei

Kirgiz from this region near the Altay Mountains before the Qumans moved west toward the Caucasus.

Interesting to note is that while Karachays are most similar to Adygei, Eastern Iranians, and somewhat similar to Altaian peoples, they do not seem to have any great affinities, at least according to haplogroup frequencies, with the Russians (in the west, near Moscow), Ukrainians, Western Iranians, Turks, Azerbaijanis, and Nogays (eastern Caucasus, near Karachays), all of which are closer in geographic and political affiliations with the Karachays. The types and frequencies of haplogroups in these populations is very different from the Karachay, and they can be excluded on this basis from having had much genetic contribution on the part of the maternal gene pool. Georgians and Iraqis, which are also close in geographic terms to the Karachay, are not typically very similar or dissimilar to the Karachay, and can be considered neutral populations that may have had some genetic contributions, but the evidence is not overwhelming.

Despite evidence from haplogroup frequencies for the genetic contributions of certain populations to the Karachay maternal gene pool, these results may not be indicative of actual migration patterns and true origins. This is because haplogroups may be similar in broad terms, but the populations in question could have totally different and non-overlapping haplotypes within those haplogroups. Also, although two populations could have different frequencies of haplogroups, one population may have received most of a specific haplotype of a less prominent haplogroup from a neighboring population with many haplotypes representative of that group. Thus, these results are only preliminary, and to elucidate further the contributions of various populations, more statistical analysis on this data, and further research on the Y-chromosome patterns must be done.

IV. Conclusions

The Karachays, according to history, were descended from a variety of peoples, including the Scythians, Alans, Hunno-Bulgars, and Qumans. To help clarify which populations currently represent those that have contributed to the genetic identity of the Karachays in the Caucasus, I have analyzed the mitochondrial DNA from this population, and compared the resulting haplogroup frequencies to those of the pertinent populations.

Results indicate that Karachays are not related closely to those groups that are closest in geographic distance, except for the Adygei, who are located right to the West of the Karachay. Instead, the closest in haplogroup types and frequencies seem to be the Eastern Iranians and Altaians.

These outcomes are supportive of the theory that Karachays are descendants of ancient Turkic peoples from the Altay Mountains (Huns-Bulgars), who have moved westward into the Caucasus and later admixed with Turkic Qumans (also from near Altay and Yenisei River Basin), who also had previously admixed with ancient Iranian peoples from the Caspian Region (Scythians and Alans).

Thus, this study indicates that the Karachay do not have one clear picture of their genetic past. Indeed, it is very complex, and to contribute more to the understanding of this intriguing problem, it is necessary to complete further research on the genetics of the Karachay people.

References Cited

Al-Zahery N, Semino O, Benuzzi G, et al. (2003) Y-chromosome and mtDNA polymorphisms in Iraq, a crossroad of the early human dispersal and of post-Neolithic migrations. *Mol. Phylogenet. Evol.* 28(3):458-72.

Bamshad M, Kivisild T, Watkins WS et al (2001) Genetic evidence on the origins of Indian caste populations. *Genome Res.* 11:994–1004.

Bandelt HJ, Forster P, and Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* 16:37-48.

Bandelt HJ, Forster P, Sykes BC, et al. (1995) Mitochondrial portraits of human populations using median networks. *Genetics* 141:743-53.

Barbujani G, Nasidze IS, Whitehead GN. (1994) Genetic Diversity in the Caucasus. *Hum. Biol.* 66(4):639-68.

Bermisheva MA, Kutuev IA, Korshunova TY, et. al. (2004) Phylogeographic Analysis of Mitochondrial DNA in the Nogays: A Strong Mixture of Maternal Lineages from Eastern and Western Eurasia. *Mol. Biol.* 38(4):516-23.

Bulayeva K, Jorde LB, Ostler C, et. al. (2003) Genetics and Population History of Caucasus Populations. *Hum. Biol.* 75(6):837-53.

Colarusso, John. (1992) A Grammar of the Kabardian Language. Calgary, Alberta: University of Calgary Press.

Comins-Richmond. (2002) The Deportation of the Karachays. *Journal of Genocide Research.* 4(3):431-9.

Conquest, Robert. (1960) The Soviet Deportation of Nationalities. London: Macmillan.

Curtubayev, Mahti. (2006). Karacay-Malkar Kulturunde Hristiyanlik Inanci (online website article-Christian Beliefs in the Karachay-Malkar Culture). http://www.kamatur.org/content/view/66/108/lang,tr/

Derenko MV, Grzybowski T, Malyarchuk BA, et. al. (2003) Diversity of Mitochondrial DNA Lineages in South Siberia. *Annals of Human Genetics.* 67:391-411.

Karca, Ramazan. (1954) Karacay-Malkar Turklerinde Hayvancilik ve Bununla Ilgili Gelenekler (Turkish-Husbandry and Associated Cultural Practices in the Karacay-Malkar People). Ankara: Turk Tarih Kurumu Basimevi.

Kivisild T, Bamshad MJ, Kaldma K, et al. (1999) Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineage. *Curr. Biol.* 9:1331-334.

Kivisild T, Tolk HV, Parik J, et al. (2002) The emerging limbs and twigs of the East Asian mtDNA tree. *Mol. Biol. Evol.* 19(10):1737-51

Kivisild T, Rootsi S, Metspalu M, et al. (2003) The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *Am. J. Hum. Genet.* 72:313–32.

Kivisild T, Reidla M, Metspalu E, et. al. (2004) Ethiopian Mitochondrial DNA Heritage: Tracking Gene Flow Across and Around the Gate of Tears. *Am. J. Hum. Genet.* 75(5):752-70.

Macaulay V, Richards M, Hickey E, et al. (1999) The emerging tree of West Eurasian mtDNAs: A synthesis of control-region sequences and RFLPs. *Am. J. Hum. Genet.* 64:232-49.

Malyarchuk BA, Grzybowski T, Derenko MV, et. al. (2002a) Mitochondrial DNA Variability in Poles and Russians. *Ann. Hum. Genet.* 66:261-83.

Malyarchuk BA, Derenko MV, Denisova GA, et. al. (2002b) Mitochondrial DNA Polymorphisms in Populations of the Caspian Region and Eastern Europe (Russian). *Genetika.* 38(2):1-5.

Maniatis T, Fritsch E, and Sambrook J. (1982) Molecular Cloning: A Laboratory Manual. New York: Cold Spring Harbor Laboratory Press.

Menges, Karl H. (1968) The Turkic Languages and Peoples: An Introduction to Turkic Studies. Germany: Otto Harrassowitz.

Musukaev, A.A. and Musukaev A.I. (2007) Kabardey ve Balkar Kulturunde Kehanet Sembolleri (Turkish- online website article). http://www.kamatur.org/content/view/487/108/lang,tr/

Nasidze IS, Inasaridze ZP, Shengeliia LA, et. al. (1990) Genetics of Caucasian Ethnic Groups: Distribution of Various Immunological and Biochemical Markers in the Adyge and Karachai-Cherkess Autonomous Oblast. *Genetika.* 26(7):1300-8.

Palanichamy M, Sun C, Agrawal S, et al. (2004) Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: Implications for the peopling of South Asia. *Am. J. Hum. Genet.* 75:000–000.

Schurr TG, Sukernik RI, Starikovskaya YB, et al. (1999) Mitochondrial DNA diversity in Koryaks and Itel'men: Population replacement in the Okhotsk-Bering Sea region during the Neolithic. *Am. J. Phys. Anthropol.* 101:1-40.

Schurr T and Wallace D. (2002) Mitochondrial DNA Diversity in Southeast Asian Populations.

Hum. Biol. 74:431-52.

Seegmiller, Steve. (1996) Karachay. Munchen-Newcastle: Lincom Europa.

Tapio M, Marzanov N, Ozerov M, et. al. (2006) Sheep Mitochondrial DNA in European, Caucasian, and Central Asian Areas. *Mol. Biol. Evol.* 23(9): 1776-83.

Torroni A, Huoponen K, Francalacci P, et al. (1996) Classification of European mtDNAs from an analysis of three European populations. *Genetics* 144:1835-850.

Torroni A, Lott MT, Cabell MF, et al. (1994b) mtDNA and the origin of Caucasians: Identification of ancient Caucasian-specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-loop region. *Am. J. Hum. Genet.* 55:760-76.

Torroni A, Miller JA, Moore LG, et al. (1994a) Mitochondrial DNA analysis in Tibet: Implications for the origin of the Tibetan population and its adaptation to high altitude. *Am. J. Phys. Anthropol.* 93:189-99.

Torroni A, Sukernik RI, Schurr TG, et al. (1993) mtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans. *Am. J. Hum. Genet.* 53:591-608.

Torroni A, Bandelt HJ, et. al. (1998) MtDNA Analysis Reveals a Major Late Paleolithic Population Expansion from Southwestern to Northeastern Europe. *Am. J. Hum. Genet.* 62(5):1137-52.

Zakiev, Mirfatykh Z. (1995) Who Were The Alans? In: Tatars: Problems of the History and Language. Kazan. Pp. 38-57. (from website: http://www.hunmagyar.org/turan/tatar/alan.html).

Map: Ethnolinguistic Groups in the Caucasus Region (from website: http://upload.wikimedia.org/wikipedia/en/f/fa/Caucasus-ethnic.jpg)

Map: Caucasus and Central Asia (from website: http://www.lib.utexas.edu/maps/kazakhstan.html)

Table 1: Karachay MtDNA Results

RFLP Test Results

Sample	HG	Sub HG	HVS I	M/N 10398 DdeI	M/N 10400 AluI	R(+) 12705 MboII	F(-) 12406 HpaI	F(-) 12406 HincII	H/HV(-) 14766 MseI	H(-) 7025 AluI	D(-) 5176 AluI	C(-) 13259 HincII	C(+) 13262 AluI	U(+) 12308 HinfI	K(-) 9052 HaeII	K(-) 9053 HhaI	J(-) 13704 BstNI	T(+) 15606 AluI	T(-) 15925 HpaII	I(-) 4529 HaeII
1	U		93-129-183Ad-186-189-249-365	-	-	+	+	+	+	+				+				+		
2	U	U5a1	256-270-293-399	-	-	+								+						
3	I		129-223-391	+	-															
4	R		189-193-311	-	-	+			+	+				-						
5	H		304	-	-	+	+	+	-	-										
6	T	T2	126-294-296	-	-	+								-				+	+	
7	D		223-362	+	+	-					-	+	-							
8	H		278-293-311	-	-	+			-	-										
9	U	U1	111-214A-249-290-327	-	-	+								+						
10	HV		168-192-343	-	-	+			-	+										
11	C		129-223-298-327	+	+	-					+	-	+							
12	U	U2	51-129C-183C-190Ci-362	-	-	+								+						
13	H		304	-	-	+			-	-										
14	U		362	-	-	+			+	+				+						
15	T	T2	126-294-296	-	-	+								-				+	+	
16	U	U1	111-214A-249-290-327	-	-	+								+						
17	U	U1	111-214A-249-290-327	-	-	+								+						
18	R		304	-	-															
19	D		223-362	+	+	-					-	+	-							
20	K		167-224-311-359	+	-	+			+	+										
21	U	U1	111-214A-249-290-327	-	-	+														
22	U	U5a	114A-192-256-270-292-294	-	-	+								+						
23	U	U1	111-214A-249-290-327	-	-	+								+						
24	U	U5a1	256-270-293-399	-	-	+								+						
25	H		304	-	-	+		+	-	-										
26	R		304	-	-	+		+		-				-						
27	R	R1	278-311	-	-	+			+	+				-						
28	T	T2	126-153-294-296	-	-	+								-				+		
29	HV		168-192-343	-	-	+			-	+										
30	D		223-362	+	+	-					-	+	-							
31	U		93-129-183Ad-186-189-249-365	-	-	+		+	+	+										
32	HV		184-189-343-390	-	-	+			-	+										
33	U	U1	111-214A-249-290-327	-	-	+														
34	H		311	-	-	+			-	-										
35	J	J1b	69-126-145-189-215-222-261	+	-	+											-			
36	T	T2	126-294-296	-	-	+								-				+		
37	U	U2	51-129C-183C-190Ci-362	-	-	+								+						
38	HV		168-192-343	-	-	+			-	+										
39	T	T2	126-294-296	-	-	+												+	+	
40	M	M7c	223-254-295	+	+	-					+	+	-							
41	R		184-189-311	-	-	+			+	+				-						
42	H		311	-	-	+			-	-										
43	U	U5a1	256-270-399	-	-	+								+						
44	H		311	-	-	+			-	-										
45	T		51-126-296	-	-	+			+	+				-				+		
46	T		126-183-184-189-294-311	-	-	+			+	+				-						
47	U	U3	343	-	-	+														
48	U	U2	51-129C-183C-190Ci-362	-	-	+														
49	K		?	+	-	+									-	-				
50	U	U5a1	256-270-293-399	-	-	+								+						
51	U	U3	343	-	-	+								+						
52	H		188	-	-	+			-	-										
53	HV		168-192-343	-	-	+			-	+										
54	H		162	-	-	+			-	-										
55	C		129-223-298-327	+	+	+					+	-	+							
56	U	U1	111-214A-249-290-327	-	-									+						
57	R		304	-	-															
58	HV		220C-292	-	-	+			-	+										
59	U		179A-183C-189-343-390	-	-	+			+	+				+						
60	U	U1	111-214A-249-290-327	-	-	+														
61	H		304	-	-	+		+	-	-										
62	U	U1	111-214A-249-290-327-362	-	-	+								+						
63	U	U1	111-214A-249-290-327	-	-	+														
64	U		93-129-183Ad-186-189-249-365	-	-	+	+	+	+	+				+						
65	R	R1	278-311	-	-	+			+	+				-						
66	U	U5a1	256-270-293-399	-	-	+								+						
67	U	U1	111-214A-249-290-327	-	-	+								+						

Figure 1: Flow chart of important ethnic groups and migrations that may have influenced Karachay genetic identity.

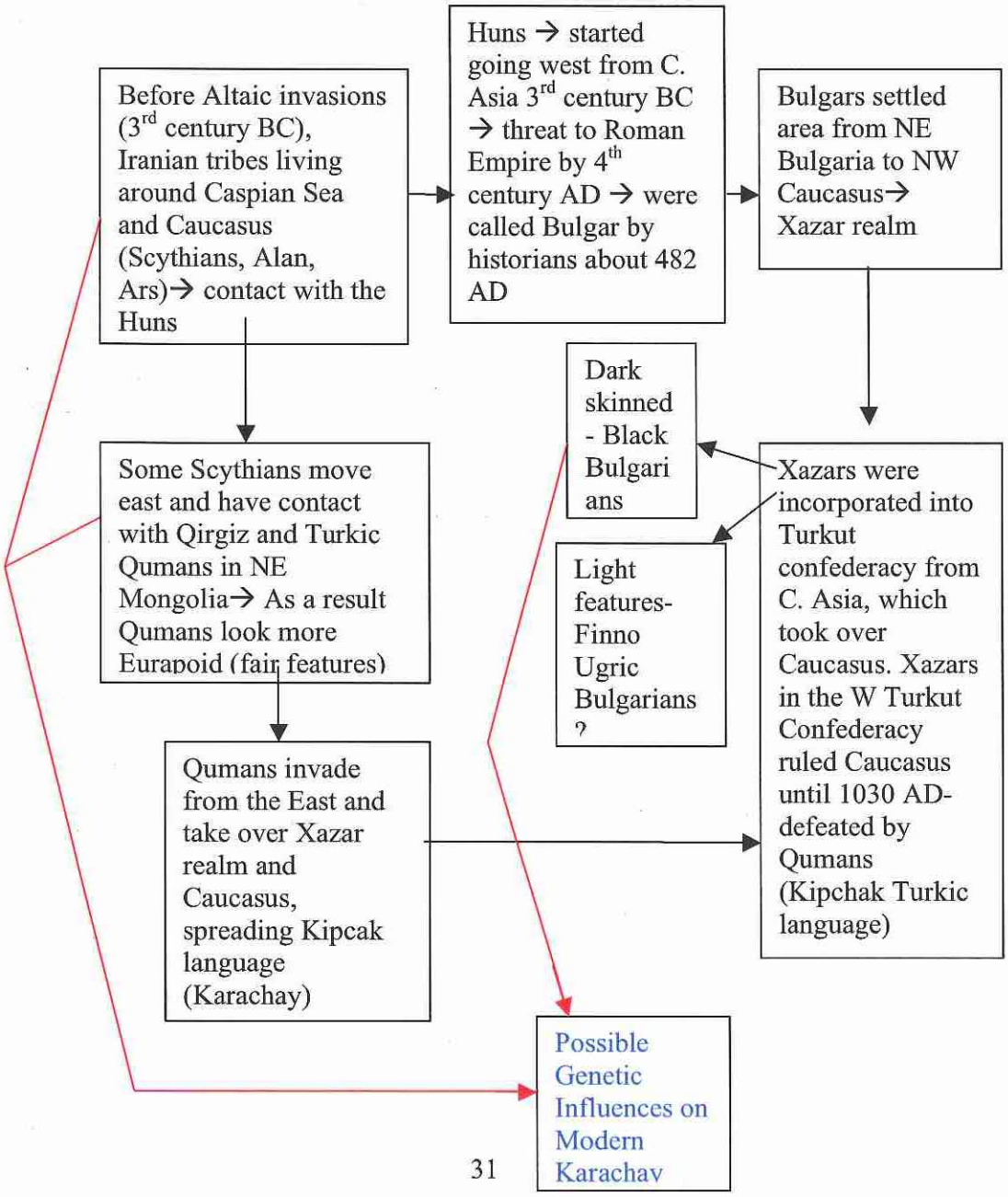

Figure 2: This map shows where the comparison populations are, and their frequencies of haplogroups. (from website: http://www.lib.utexas.edu/maps/kazakhstan.html)

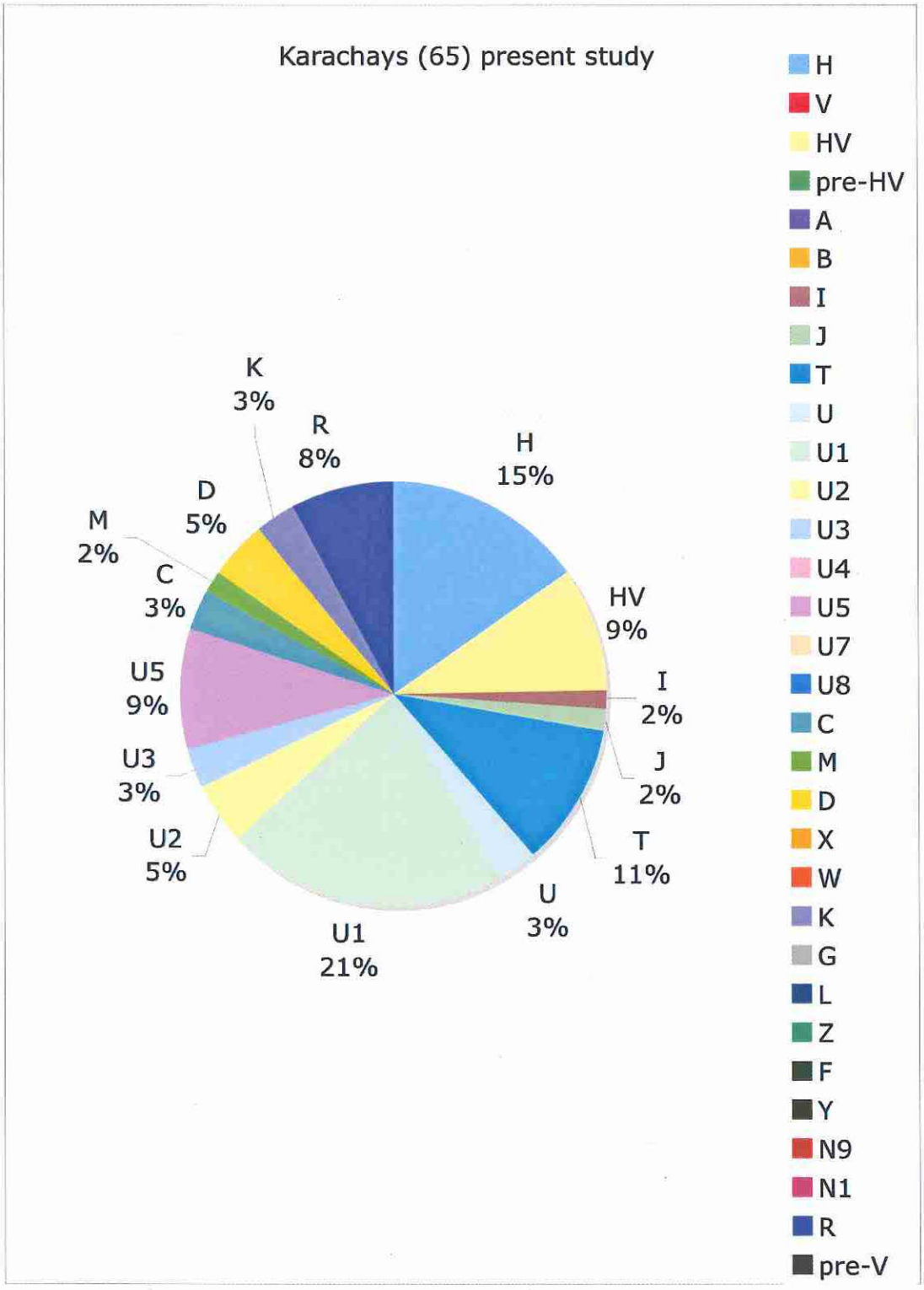

Figure 3: The different haplogroups represented by the Karachay population results.

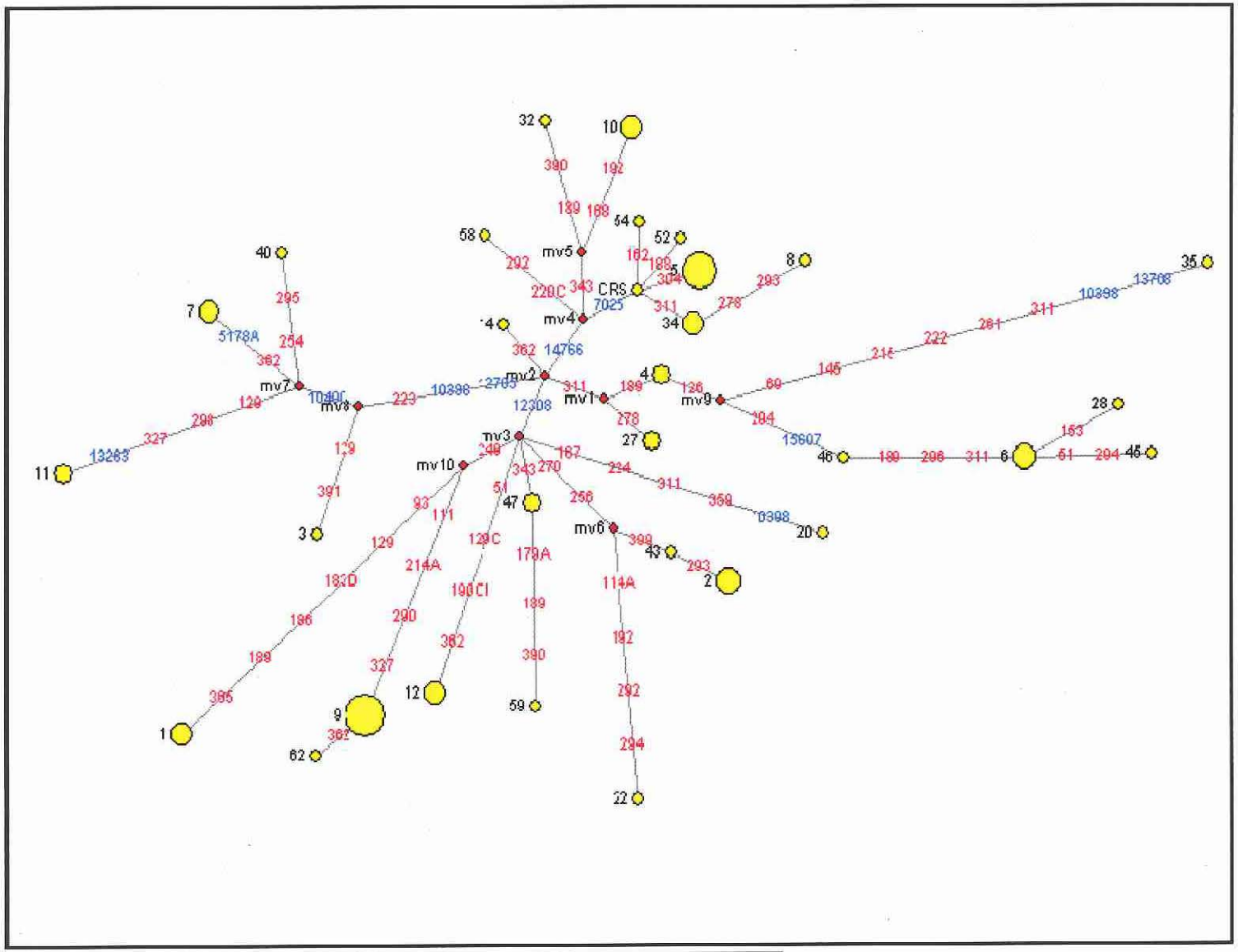

Figure 4: Karachay Mitochondrial DNA Tree

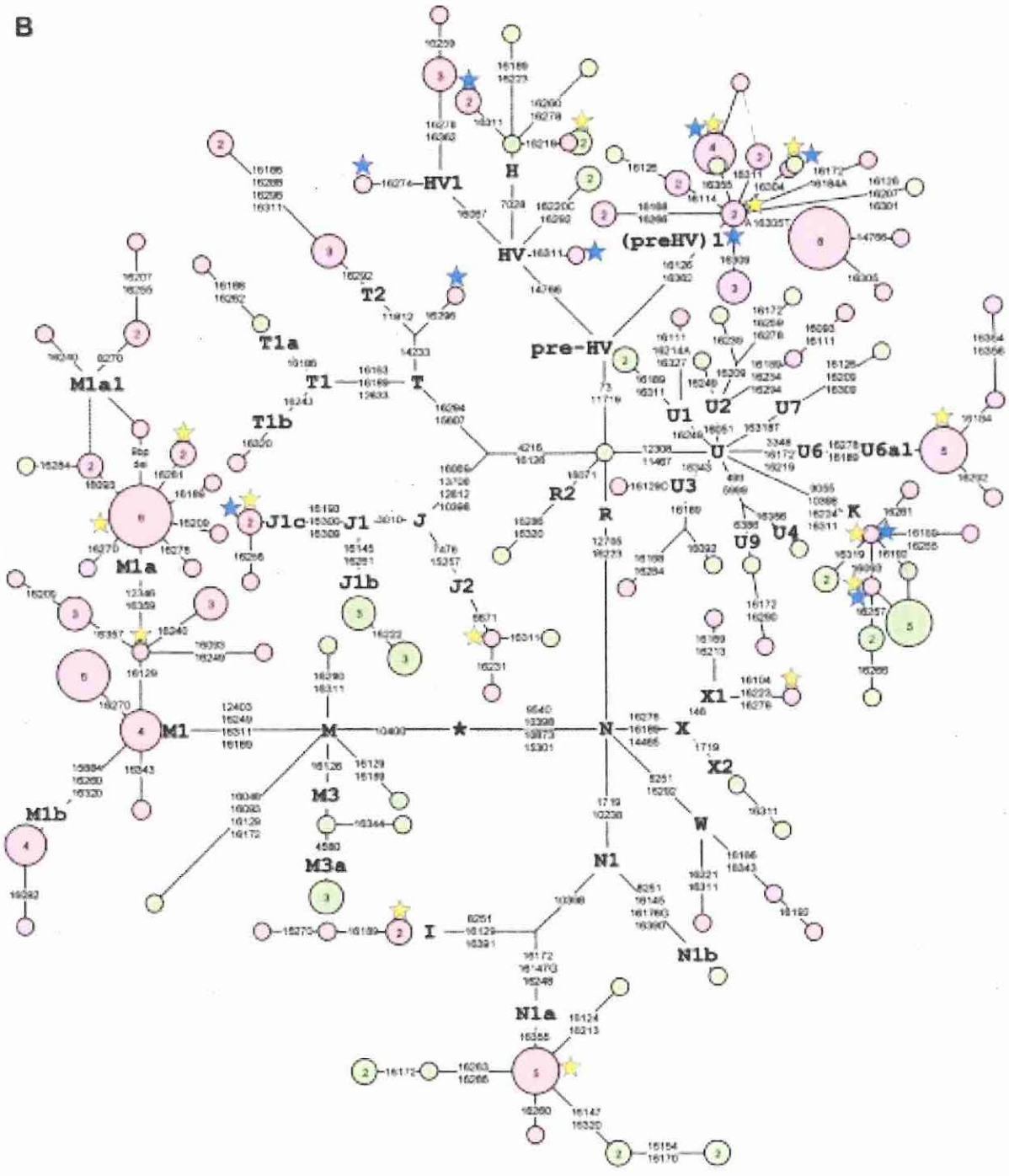
B
HV1
H
HV
(preHV) 1
T2
T1a
pre-HV
T1
T
M1a1
T1b
U1
U2
U7
U
U6
U6a1
R2
R
U3
J1c
J1
J
U9
U4
K
M1a
J1b
J2
X1
M1
M
N
X
X2
M3
W
M1b
N1
M3a
I
N1b
N1a

Figure 5: Mitochondrial DNA Tree of Non-African Lineages (Kivisild, 2004).

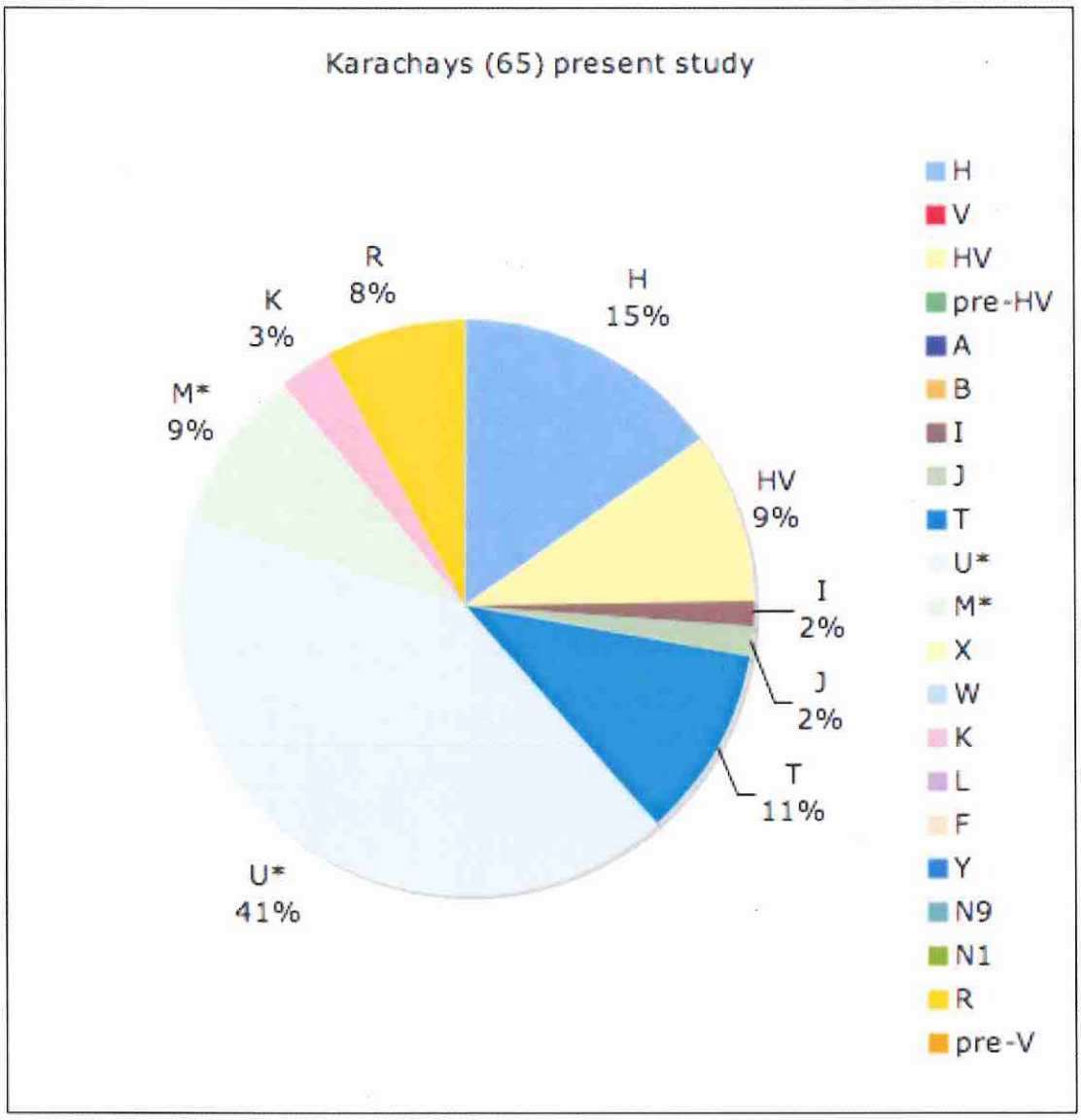

Figure 6a: Karachay Haplogroups with U's and M's grouped.

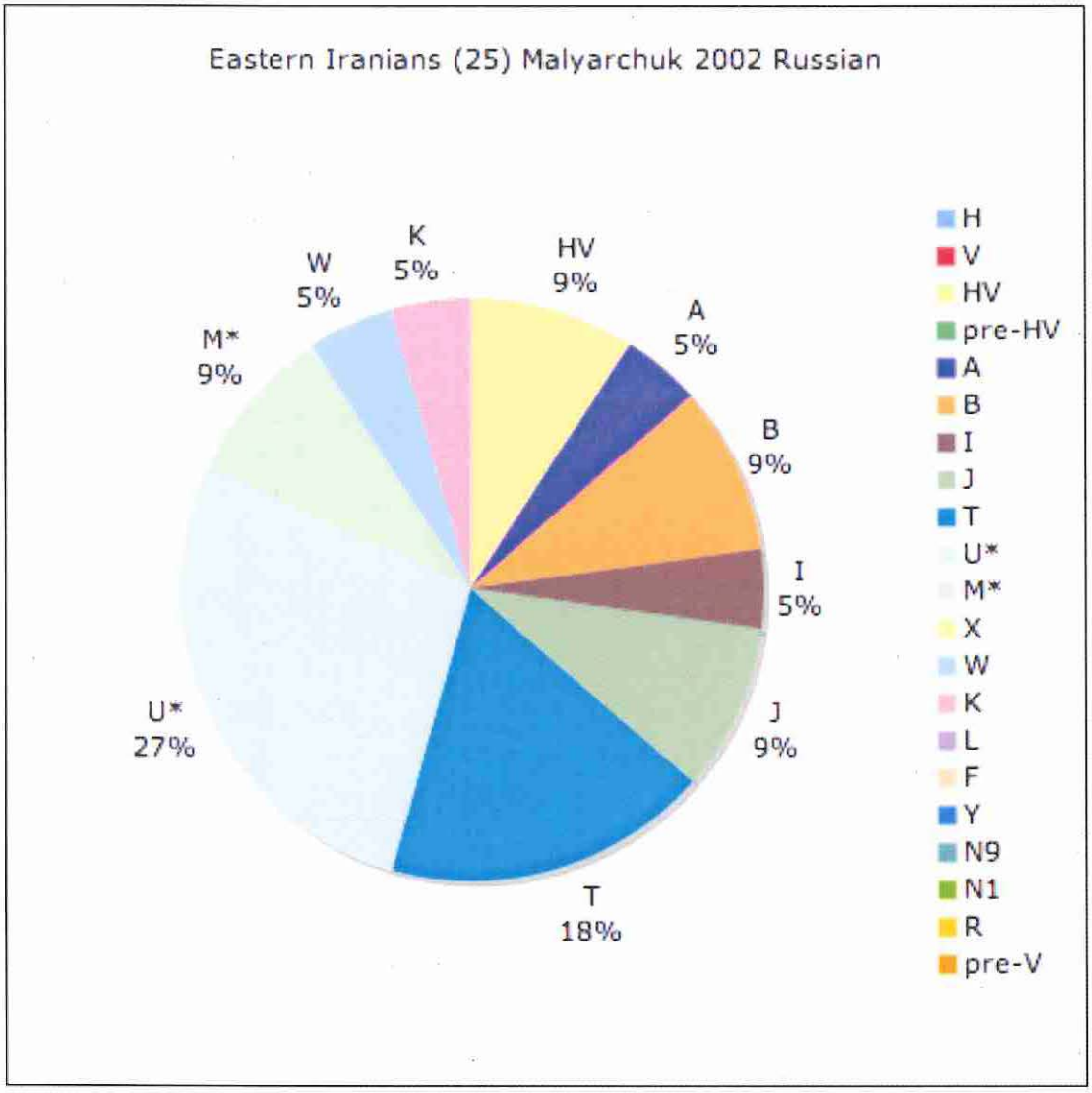

Figure 6b: Eastern Iranians

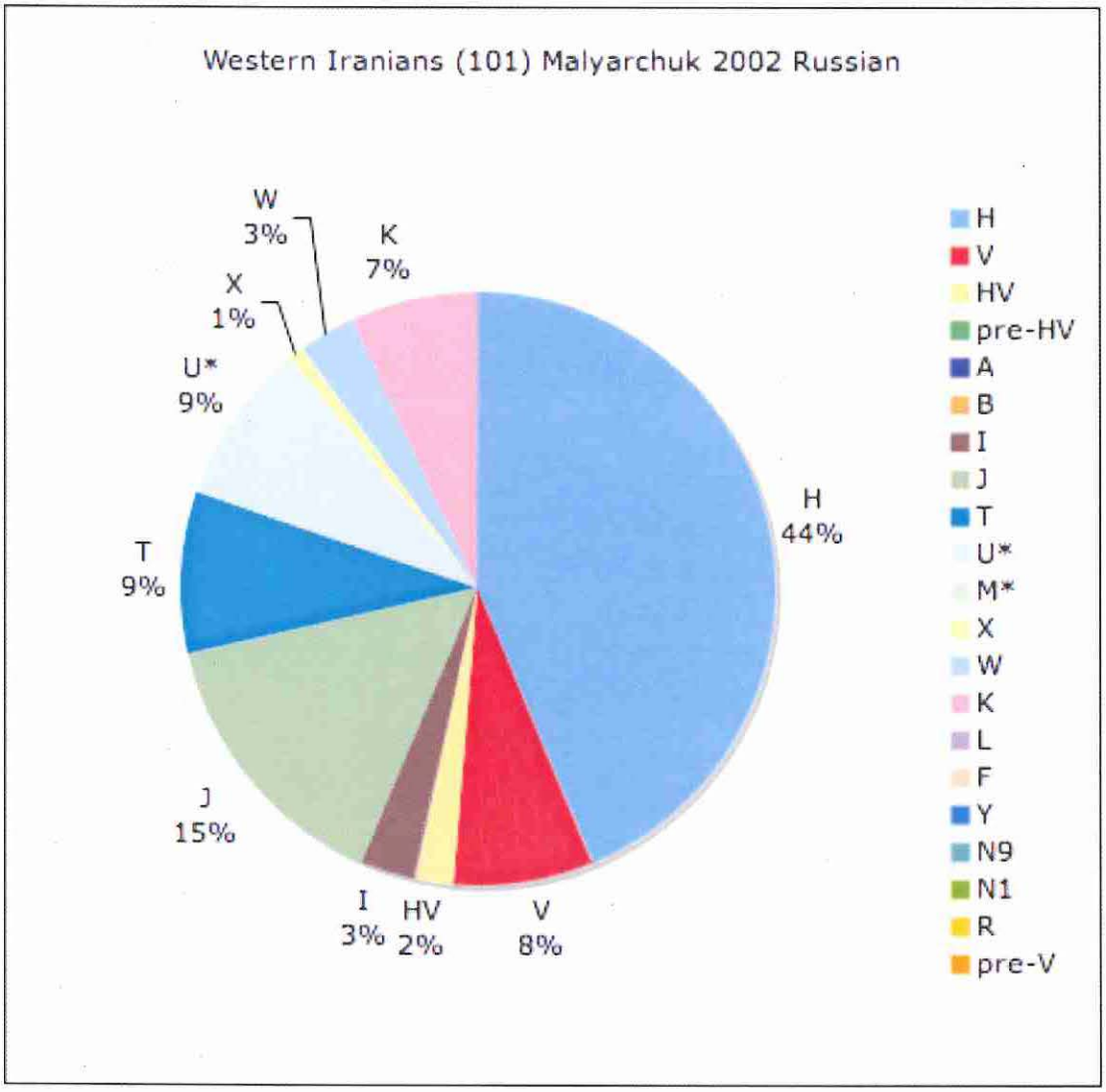

Figure 6c: Western Iranians

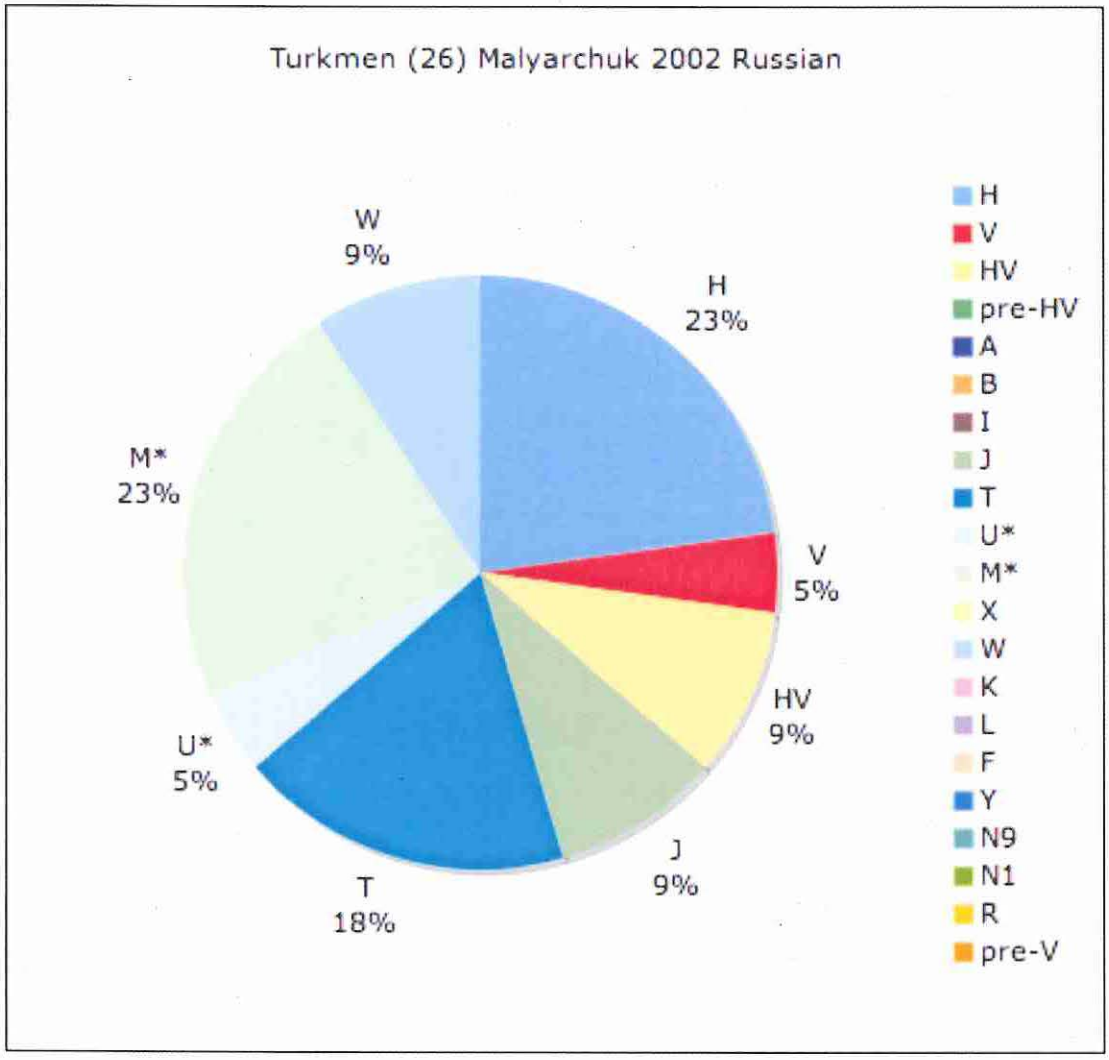

Figure 6d: Turkmen

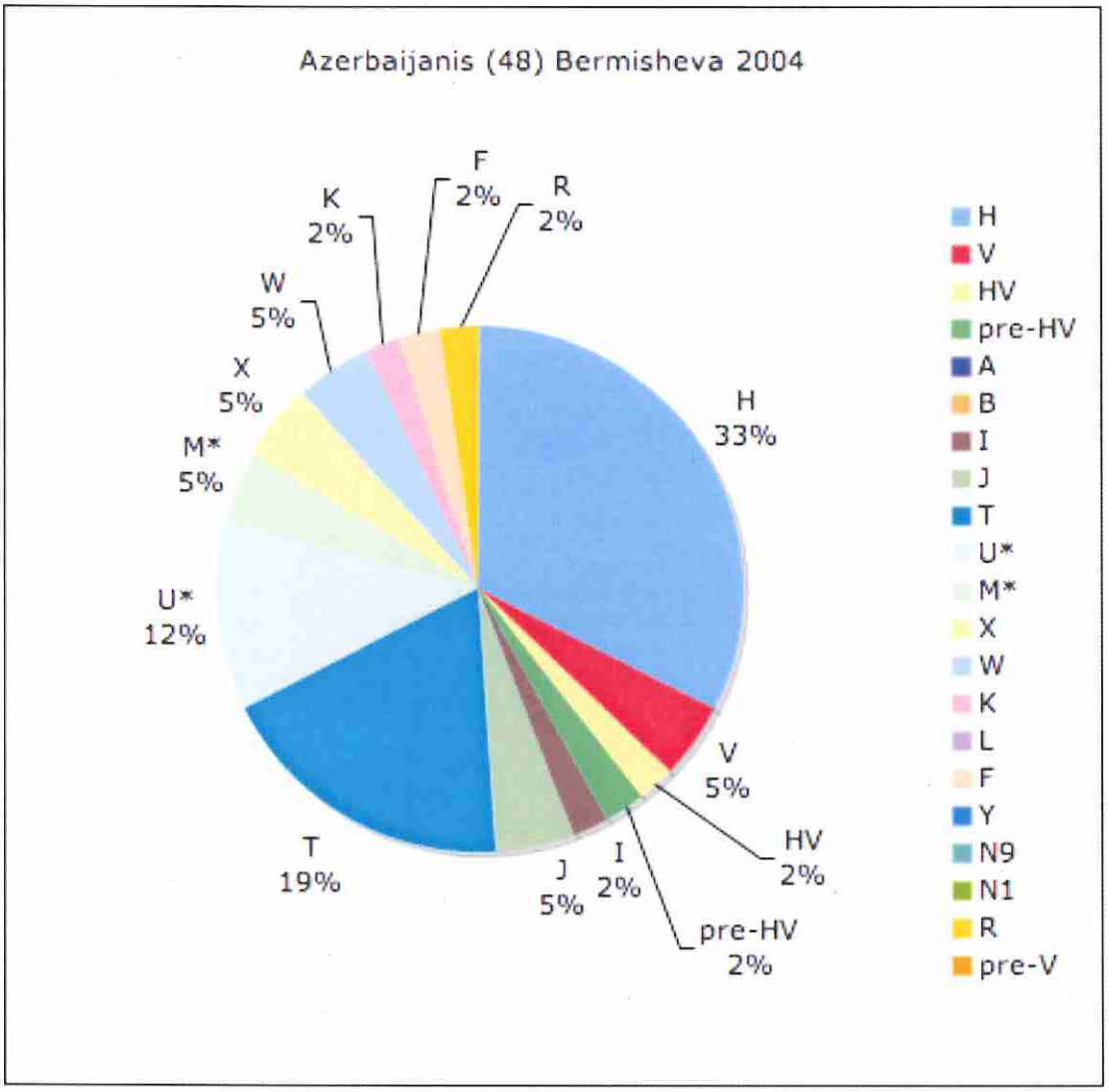

Figure 6e: Azerbailjanis

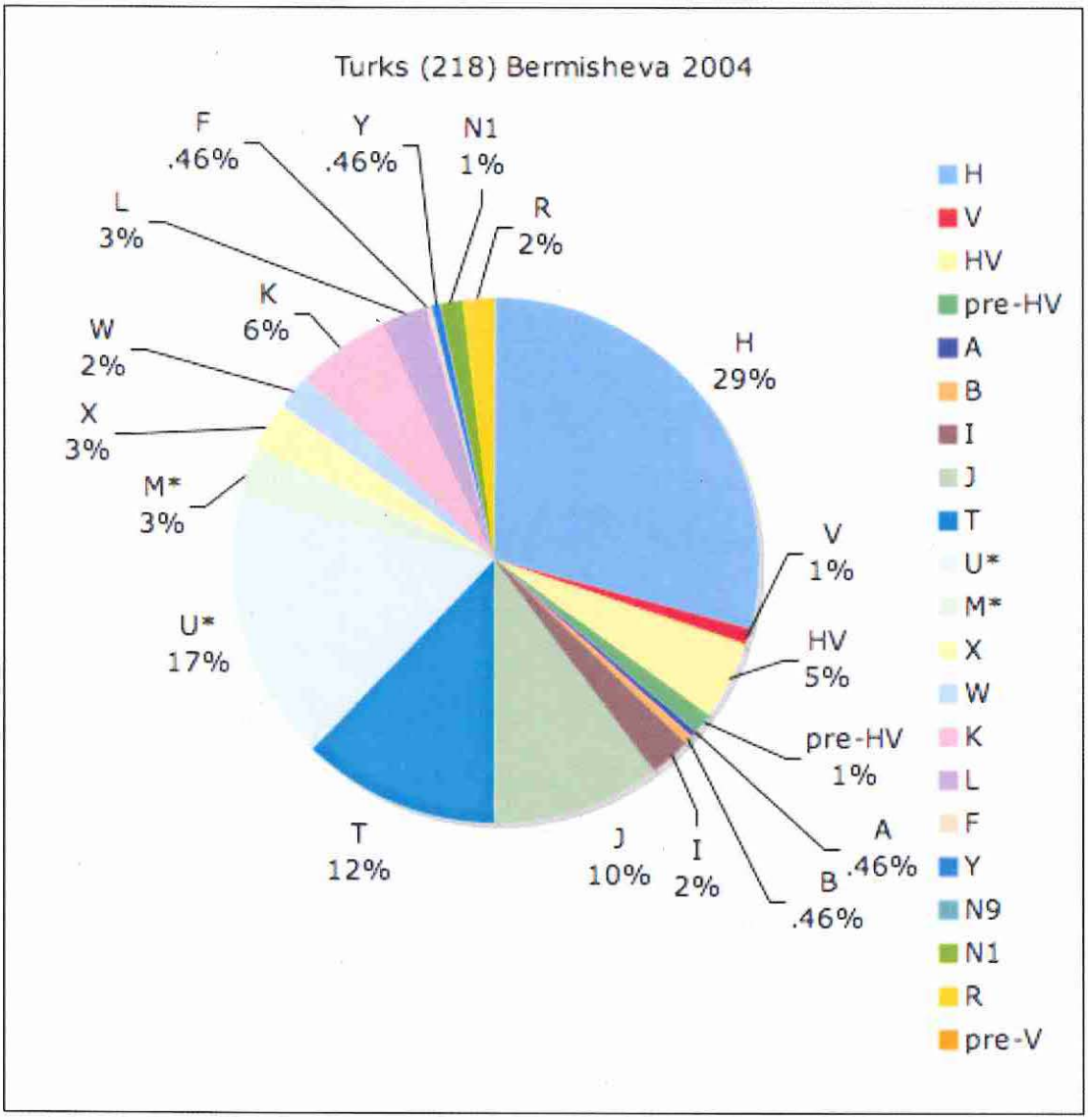

Figure 6f: Turks

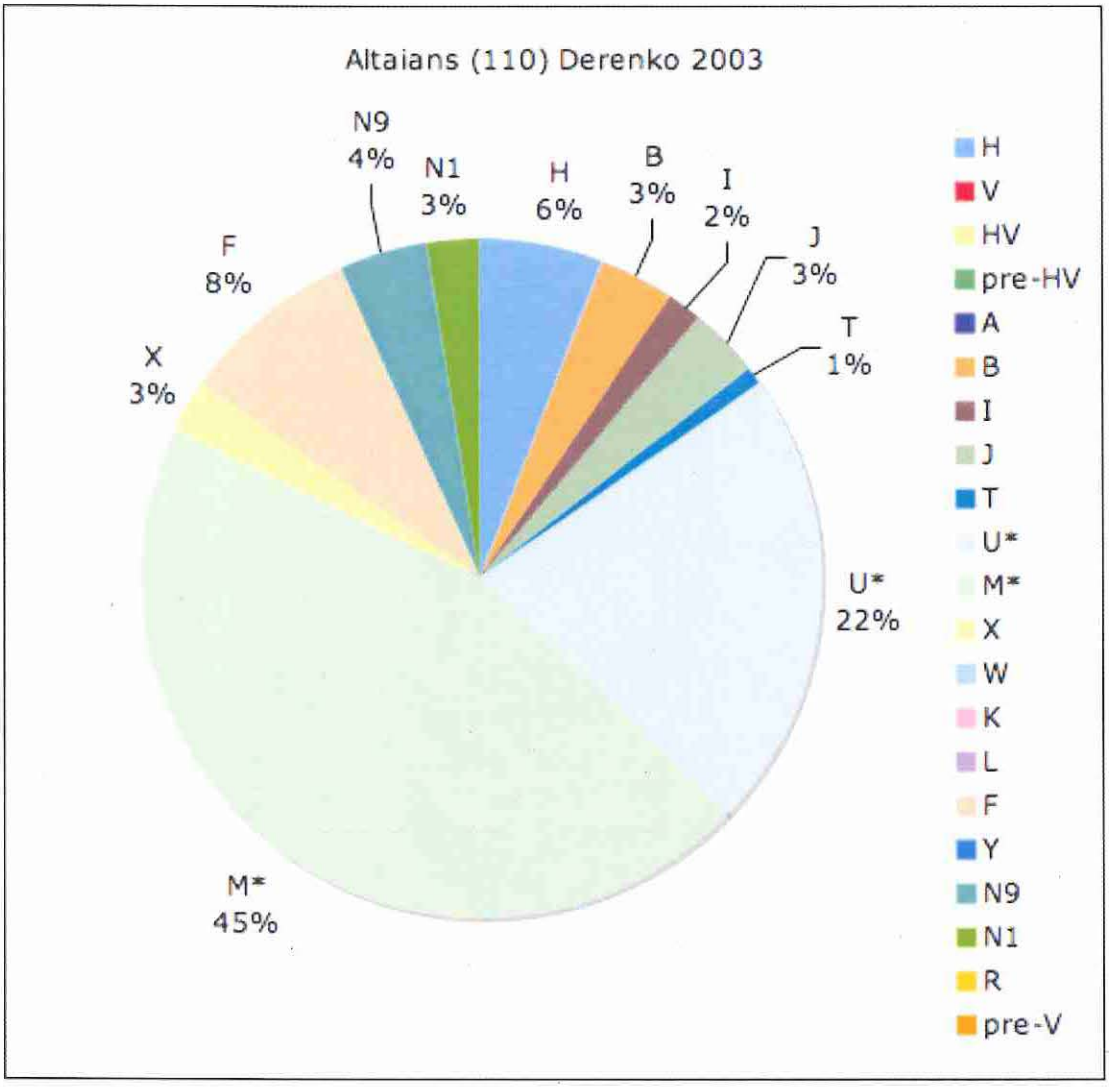

Figure 6g: Altaians

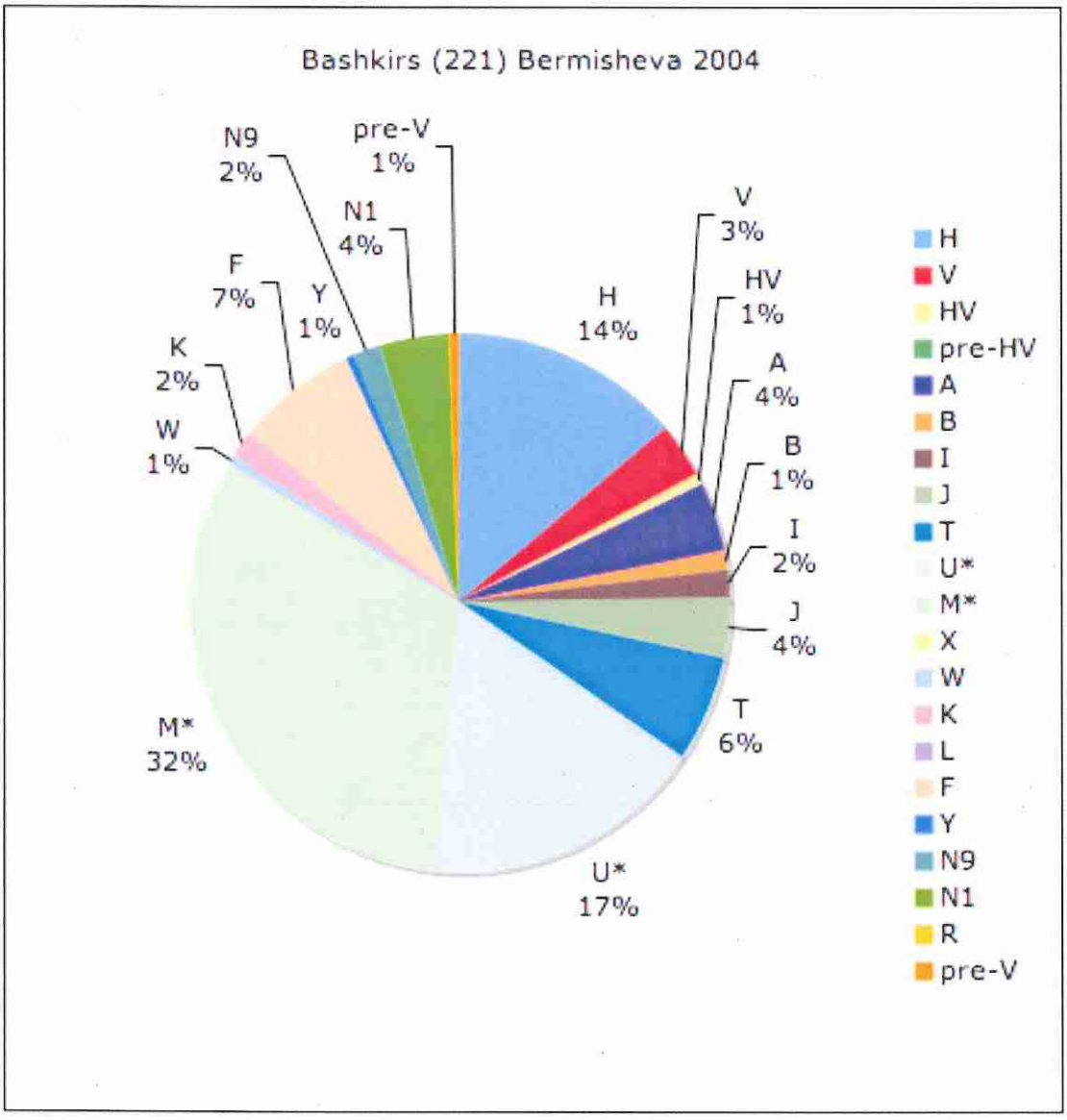

Figure 6h: Bashkirs

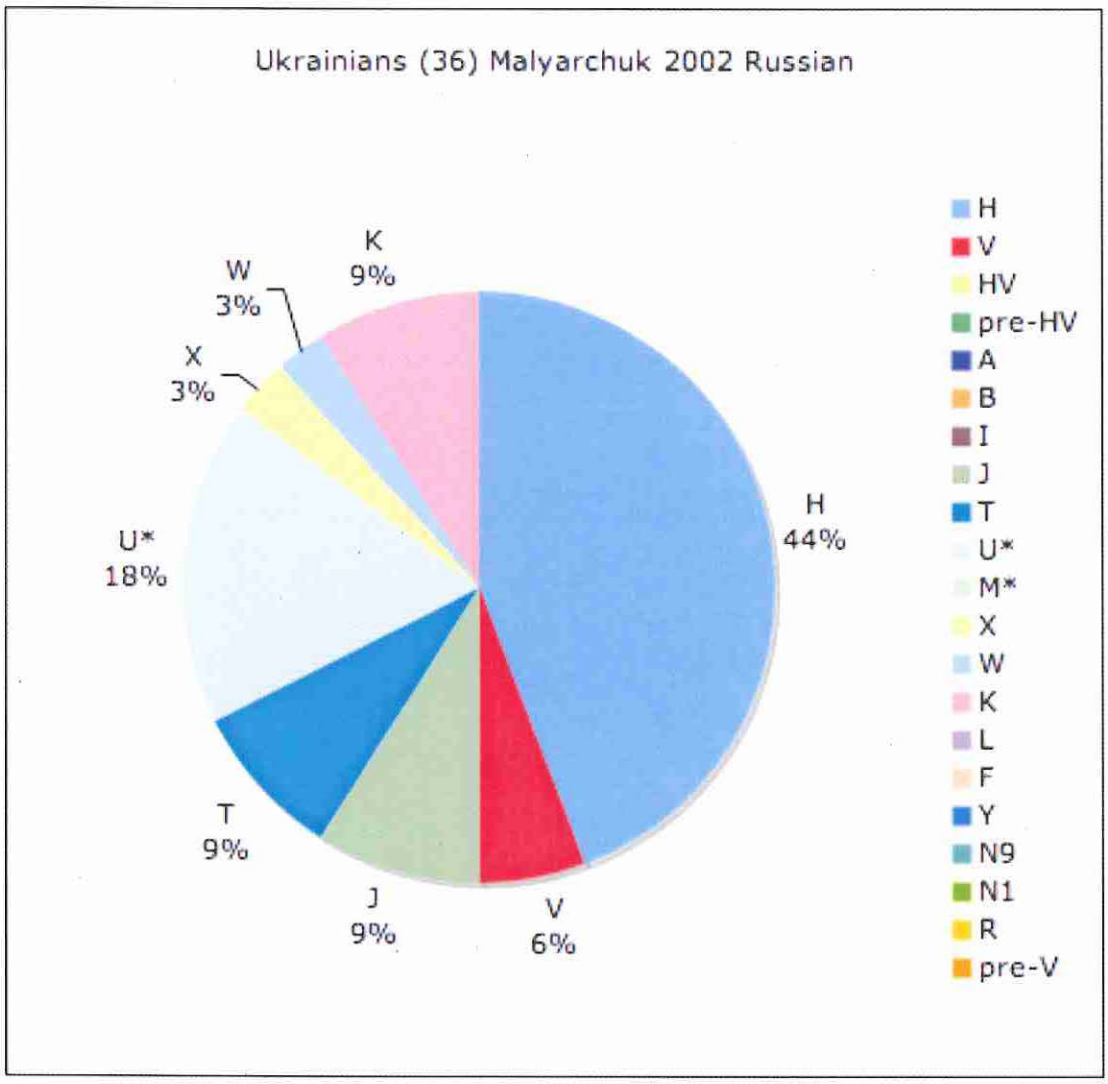

Figure 6i: Ukrainians

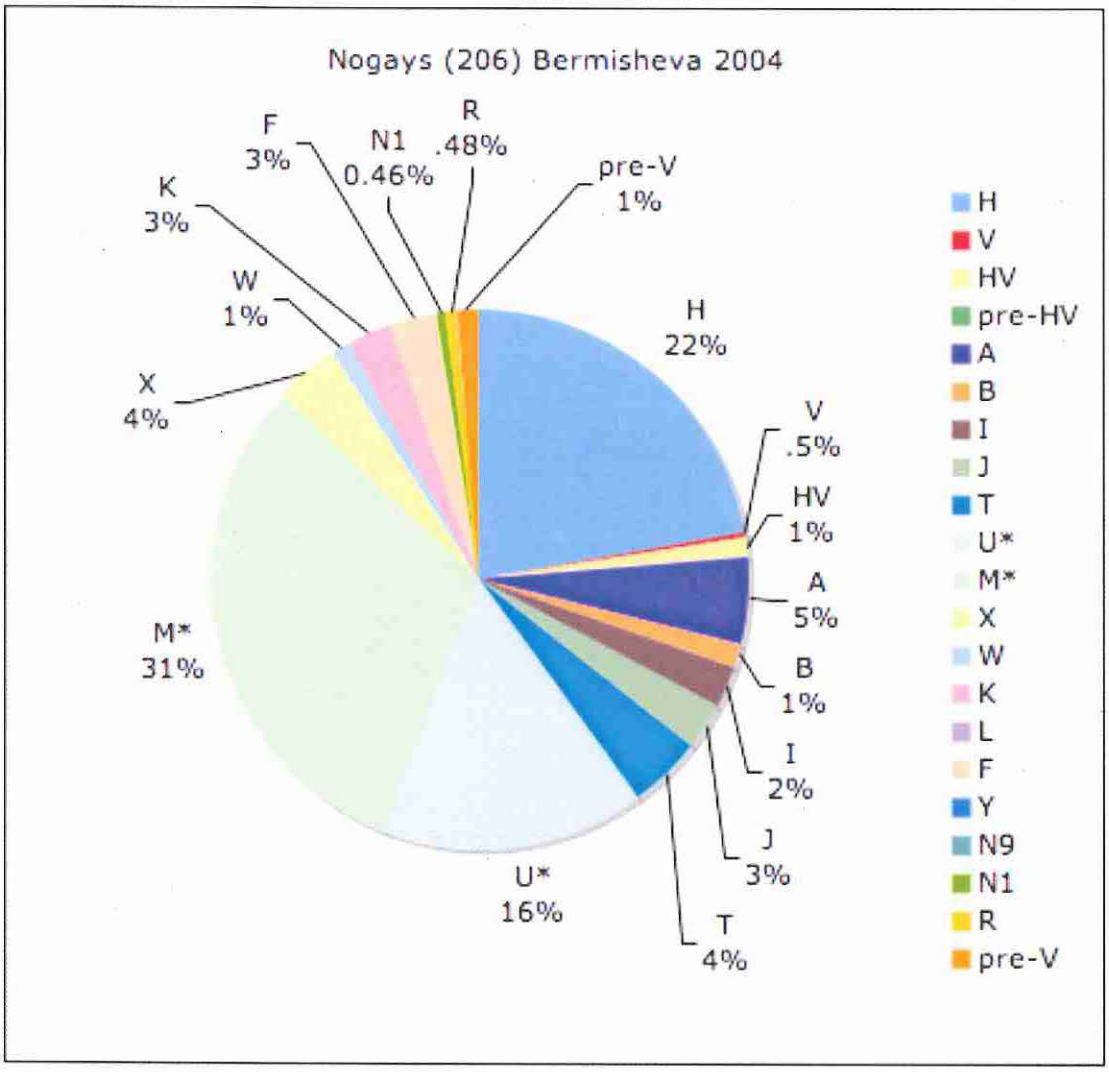

Figure 6j: Nogays

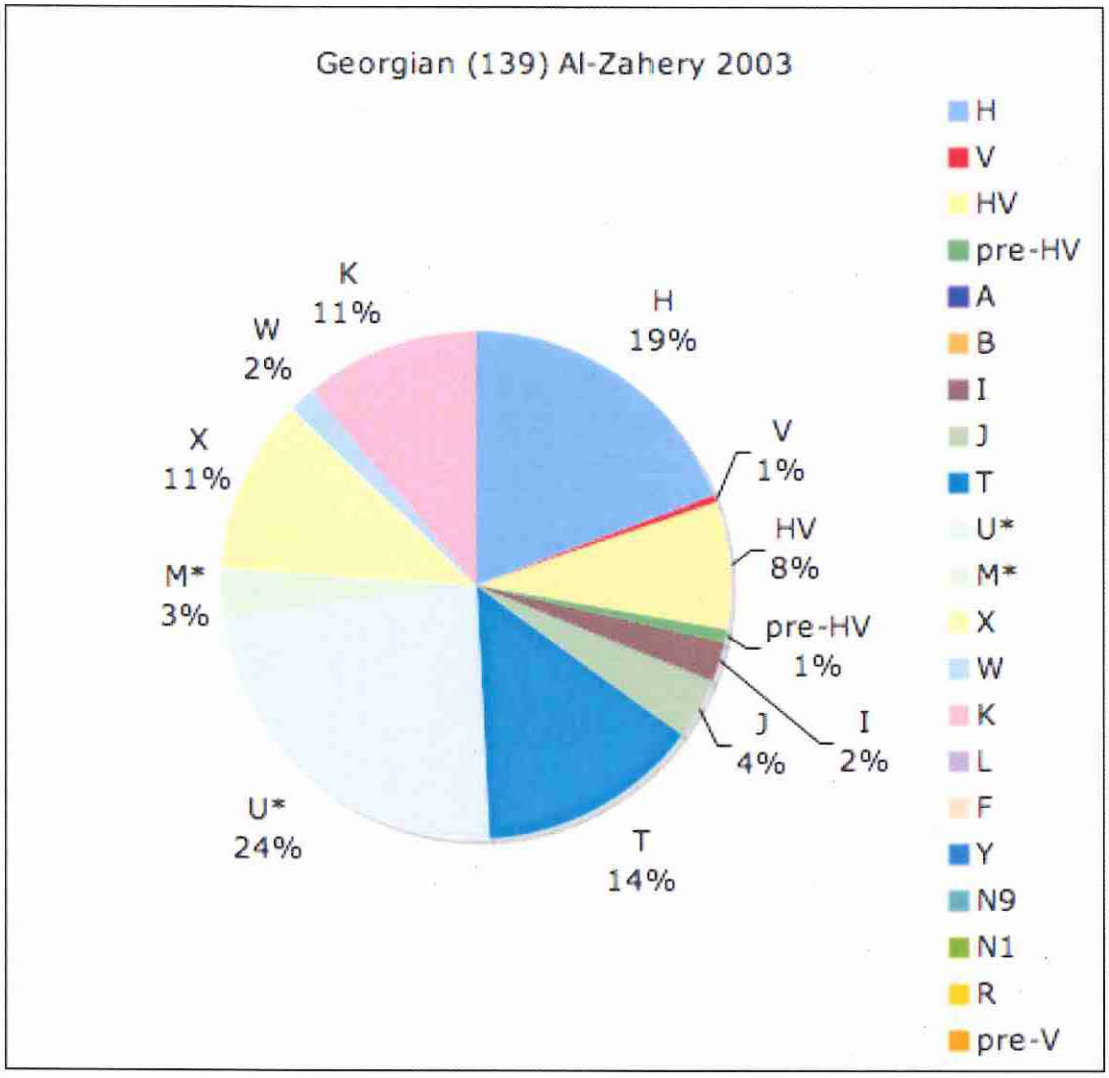

Figure 6k: Georgians

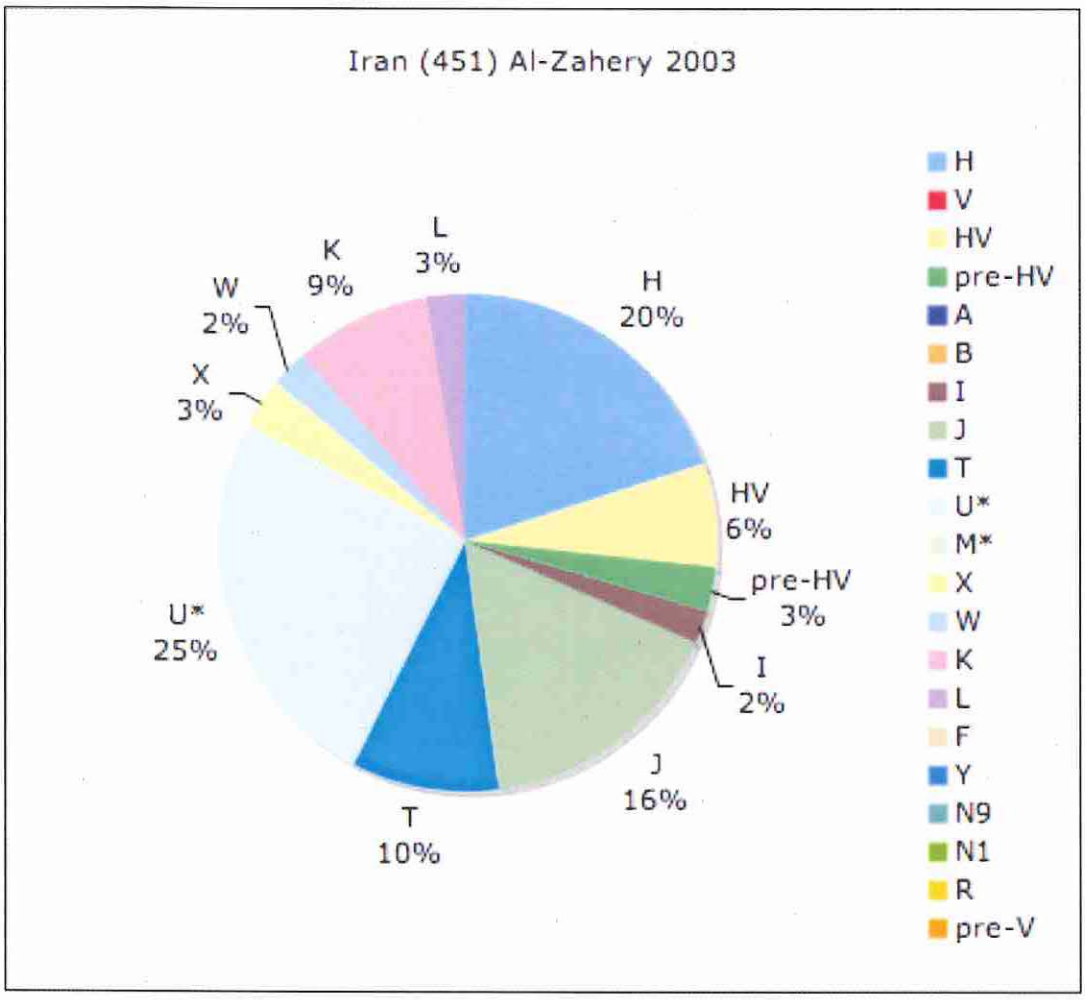

Figure 61: Iranians (no specific location in Iran)

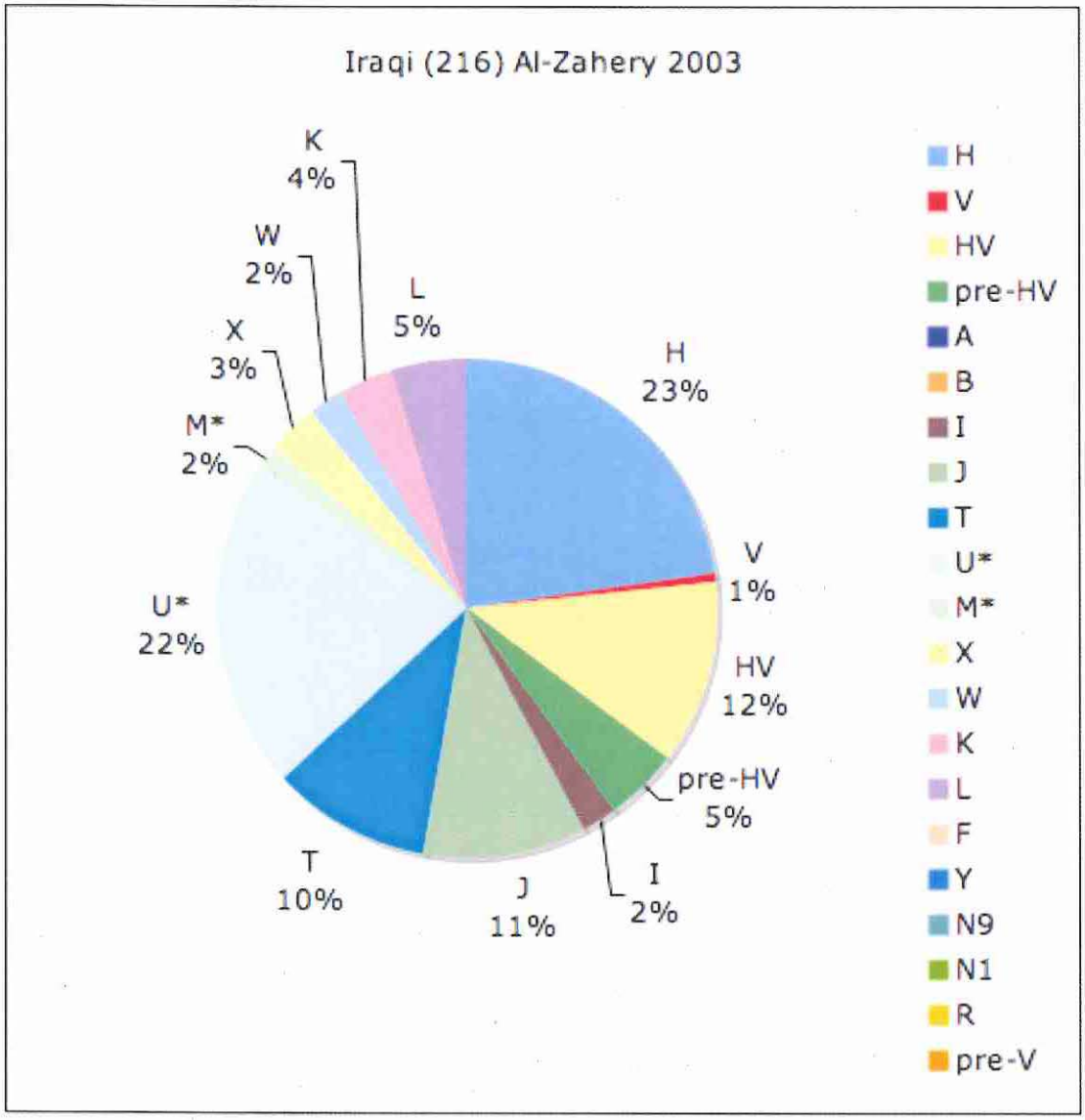

Figure 6m: Iraqis

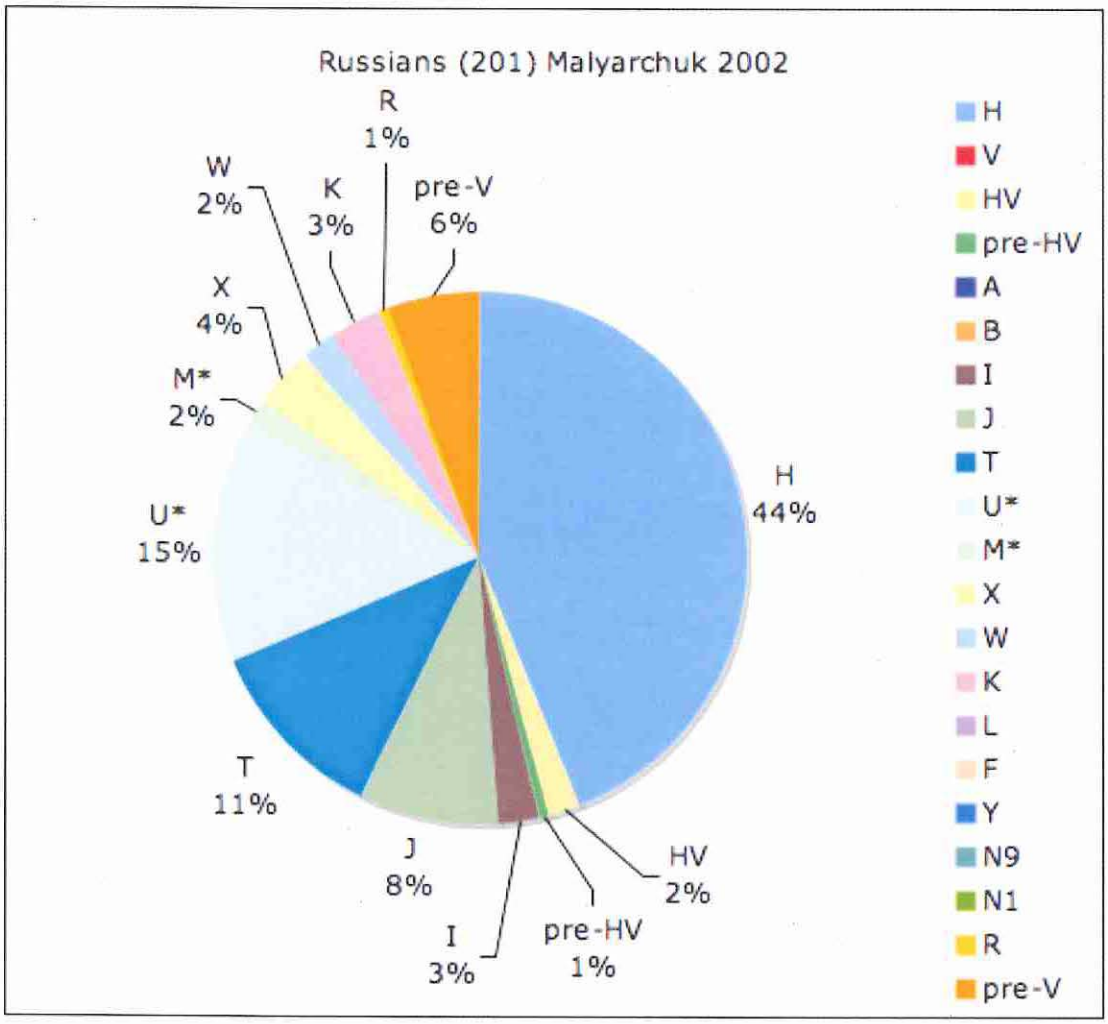

Figure 6n: Russians

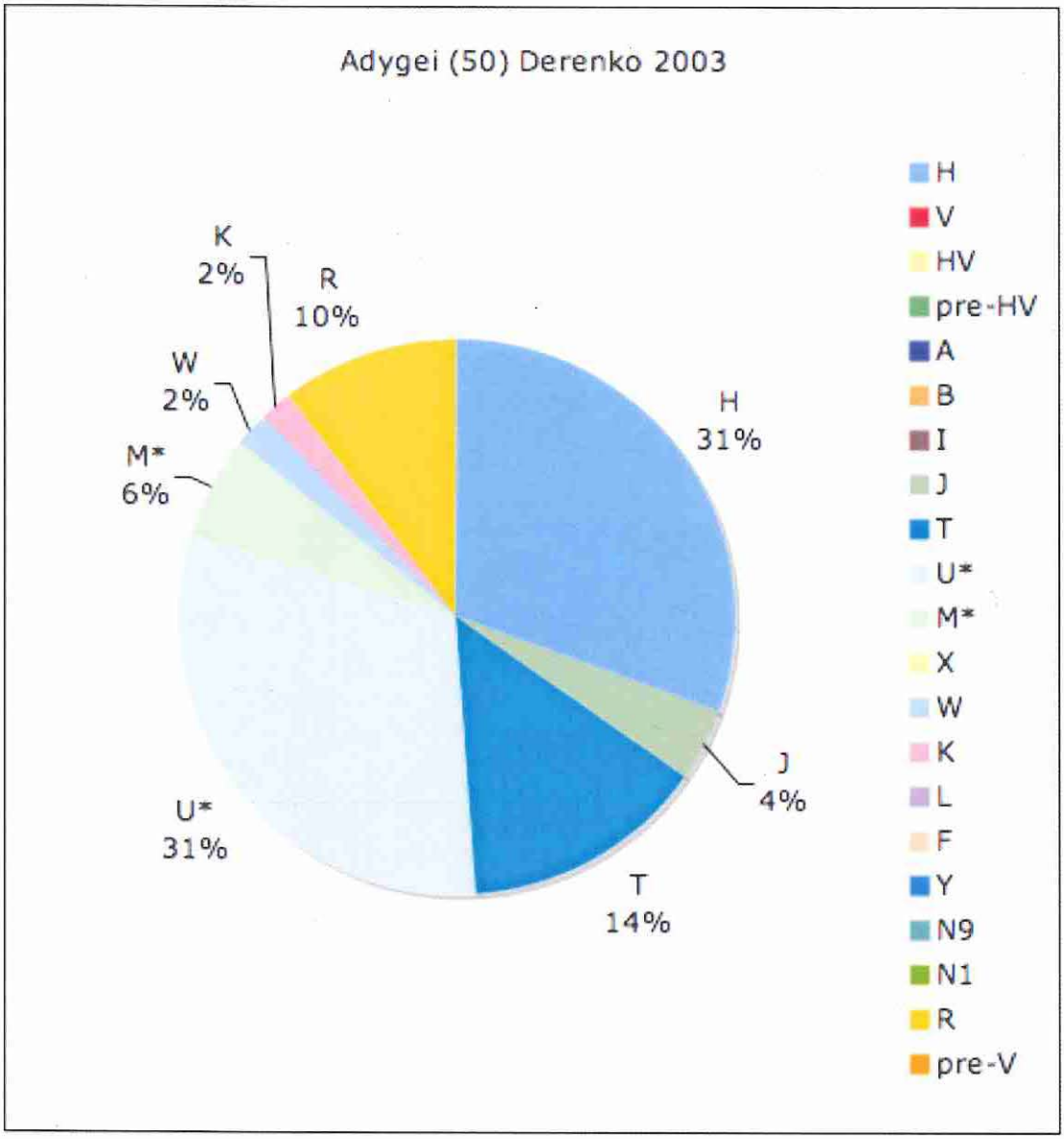

Figure 6o: Adygei

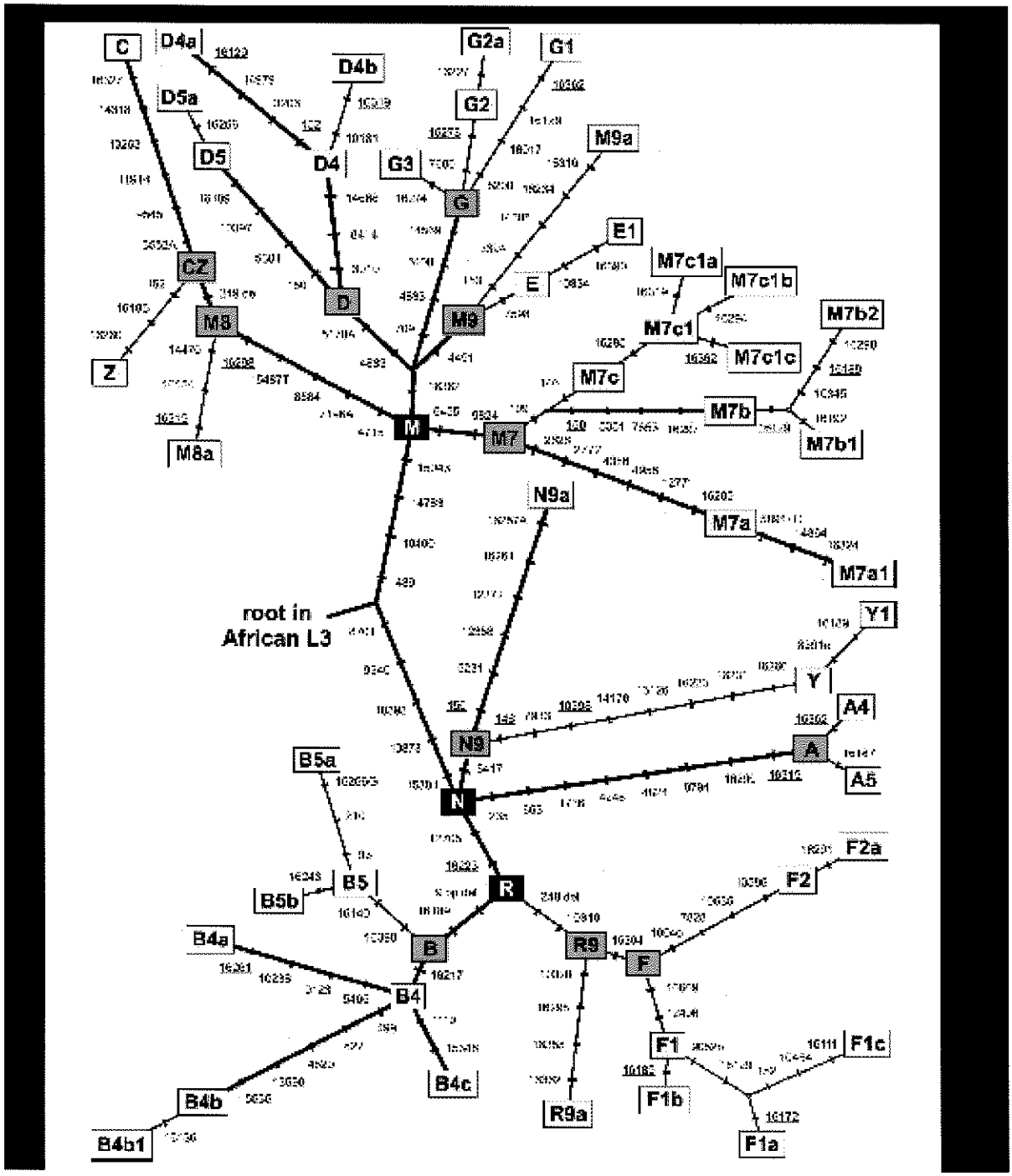

Figure 7: East Asian Mitochondrial DNA Tree Showing East Asian Specific Haplogroups (Kivisild, 2002).

Ethnolinguistic Groups in the Caucasus Region

Figure 8: Map of Ethnolinguistic Groups in Caucasus Region (from website: http://upload.wikimedia.org/wikipedia/en/f/fa/Caucasus-ethnic.jpg)

Appendix I (1 page): Genealogical Form

Genealogical Information

The Genetic History of the Karachays

PI: Theodore G. Schurr, Ph.D., Assistant Professor
Department of Anthropology, University of Pennsylvania,
325 University Museum, 3260 South Street, Philadelphia, PA. 19104-6398
Tel#: 215-573-7632; Email: tgschurr@sas.upenn.edu

Collaborator: Ms. Aslihan Sen
669 Franklin Lakes Road, Franklin Lakes, NJ. 07417
Tel#: (201) 485-8825; Email: aslihan@sas.upenn.edu

Sample #_____ Date Collected_____

Name_____ Date of Birth_____ Sex_____

Place of Birth_____ Ethnicity_____

Mother's Name_____ Date of Birth_____

Place of Birth_____ Ethnicity_____

Father's Name_____ Date of Birth_____

Place of Birth_____ Ethnicity_____

Maternal Grandmother_____ Date of Birth_____

Place of Birth_____ Ethnicity_____

Maternal Grandfather_____ Date of Birth_____

Place of Birth_____ Ethnicity_____

Paternal Grandmother_____ Date of Birth_____

Place of Birth_____ Ethnicity_____

Paternal Grandfather_____ Date of Birth_____

Place of Birth_____ Ethnicity_____

Additional Notes:

Appendix II (2 pages): Informed Consent Form

Informed Consent Form

The Genetic History of the Karachays

PI: Theodore G. Schurr, Ph.D., Assistant Professor
Department of Anthropology, University of Pennsylvania,
325 University Museum, 3260 South Street, Philadelphia, PA. 19104-6398
Tel#: 215-573-7632; Email: tgschurr@sas.upenn.edu

Collaborator: Ms. Aslihan Sen
669 Franklin Lakes Road, Franklin Lakes, NJ. 07417
Tel#: (201) 485-8825; Email: aslihan@sas.upenn.edu

Invitation to Participate: You are invited to participate in a study of genetic variation of the Karaçays to elucidate their history and origins in West Asia and the Caucasus.

Purpose: In this research project, we will survey your mitochondrial DNA (mtDNA) and Y chromosome genes for sequence variation. Small mutations in these portions of your genome create genetic markers that can be used to characterize human populations. These changes identify specific genetic lineages within human populations that can be traced through maternal (mtDNA) and paternal (Y chromosome) lineages. For these reasons, we can trace these mutations through human families from the present to the distant past with considerable accuracy. We can also reconstruct patterns of human movement through geographic areas by tracking the spread of these lineages in different human groups.

Procedures: If you decide to participate, we will obtain buccal (cheek) samples from you. This procedure requires only 5-10 minutes, and will not subject you to any significant health risks. We will also ask you to complete a short questionnaire about your family history before providing the buccal samples. The whole process will take no more than 30-60 minutes. The buccal cells will be taken back to the PI's (Schurr) laboratory at the University of Pennsylvania, and DNA will be extracted from them. The DNA will then be analyzed for mtDNA and Y-chromosome variation using molecular biology methods. Upon the completion of the study, the DNA samples will be destroyed to prevent further use of them outside of the parameters of this project.

Potential Risks or Discomforts to Participant: The buccal cell collection procedure requires only a few minutes to complete, and may result in very minor discomfort. Beyond the temporary discomfort, this procedure will subject you to no other significant health risks.

Benefits to Participant: There are no direct benefits to you as a result of your participation in this study. However, it is our hope that, by analyzing the sequence variation in the DNA from your buccal cells, we will more fully elucidate the history of the Karaçays.

Confidentiality: During the research project, all genetic and genealogical data will be kept in computerized form in the PI's laboratory, and accessed by only those persons involved in the

project. Any data under the researchers' control will be disclosed in scientific reports or public presentations in a manner that does not reveal your identity. The data obtained by the analysis of your DNA will also be made available to you upon your request.

Compensation and Medical Treatment: While there is minimal risk to you in taking buccal cell samples, should any discomfort arise during this process, the medical personnel involved in sample collection will provide the requisite treatment. There will also be no financial compensation for your participation.

Research Contacts: Should you have any questions about this project or your rights as a research participant, please contact the PI (Schurr) at the mailing address, telephone number, or email address provided above.

Terms of Participation: Your participation in this research project is completely voluntary, and you may withdraw from it at any time.

Conclusion: You have read and understand the consent form. You agree to participate in this research study.

Name of Subject	Signature of Subject	Date
Name of Person Obtaining Consent	Signature of Person Obtaining Consent	Date

Table 2: Population Haplogroup Frequencies

Population (#individuals)	Source	H	V	HV	pre-HV	A	B	I	J	T	U	U1	U2	U3	U4	U5	U7	U8	C	M	D	X	W	K	G	L	Z	F	Y	N9	N1	R	pre-V
Poles (436)	Malyarchuk 2002	45.18		0.92	0			1.83	7.8	11.46	0.23	0	0.92	0.46	5.05	8.72	0.23	0.46		1.83		1.83	3.67	3.44							0.46	0.46	4.82
Russians (201)	Malyarchuk 2002	42.29		1.99	0.5			2.49	7.96	10.95	0	1	1.49	1	3.48	10.45	0.5	0		1.49		3.48	1.99	2.99							0	0.5	5.47
Altaians (110)	Derenko 2003	6.4				0	3.6	1.8	3.6	0.9	16.4		7.3						19.1	7.3	15.5	2.7		0	1.8		4.6	9.1	0	4.5	2.7	0	
Khakassians (53)	Derenko 2003	3.8				3.8	3.8	0	1.9	1.9	11.3		0						35.9	0	13.2	0		0	0		0	22.6	0	1.9	0	0	
Buryats (91)	Derenko 2003	2.2				2.2	6.6	1.1	2.2	1.1	1.1		3.3						28.6	3.3	33	0		0	14.3		1.1	1.1	2.2	0	0	0	
Sojots (30)	Derenko 2003	0				10	3.3	0	0	0	3.3		0						20	0	46.7	0		3.3	6.7		0	0	3.3	0	0	3.3	
Todjins (48)	Derenko 2003	2.1				4.2	4.2	0	0	0	6.3		4.2						47.9	4.2	4.2	0		0	18.8		0	2.1	2.1	0	0	4.2	
Tuvinians (90)	Derenko 2003	1.1				1.1	0	0	5.6	1.1	3.3		0						47.8	0	17.8	0		0	4.4		1.1	2.2	1.1	1.1	0	2.2	
Tofalars (58)	Derenko 2003	6.9				5.2	0	0	8.6	5.2	0		0						62.1	0	0	0		0	0		5.2	0	0	0	0	1.7	
Adygei (50)	Derenko 2003	30		0	0			0	4	14	0	6	2	14	2	8			6	0		0	2	2		0					10	
Druze (45)	Derenko 2003	13.33		8.88	4.44			2.22	6.66	4.44	4.44	6.66	0	0	0	0			0	2.22		26.7	0	15.56		4.44					0	
Iraqi (216)	Al-Zahery 2003	19.9	0.5	10.6	4.2			1.9	9.3	8.8	19									1.4		2.8	1.9	3.2		4.2						
Iran (451)	Al-Zahery 2003	17.1		5.5	2.4			2	13.5	8.4	21.5											2.9	2	7.5		2.2						
Arabia (389)	Al-Zahery 2003	12.9		3.6	15.2			0.8	20.8	4.6	10.5											1.8	1.8	3.6		10.5						
Syrian (69)	Al-Zahery 2003	24.6	2.9	4.3	5.8				10.1	10.1	15.9									1.4			2.9	4.3		5.8						
Palestinian (117)	Al-Zahery 2003	30.8		1.7	2.6				9.4	12.8	7.6									1.7		3.4	2.6	6.8		5.2						
Georgian (139)	Al-Zahery 2003	17.3	0.7	7.2	0.7			2.2	3.6	12.9	21.6									2.9		10.1	1.4	10.1								
Armenian (192)	Al-Zahery 2003	30.9		7.3	0.5			1.6	8.9	11.5	22.5											2.1	1	7.9								
Anatolia (388)	Al-Zahery 2003	25		3.6	2.8			2.3	10.9	11.9	19.3									4.4		4.4	3.9	5.9		0.3						
Nogays (206)	Bermisheva 2004	22.33	0.48	0.97		5.34	1.46	2.43	2.91	4.37		1.94	2.42	5.34	1.94	2.91	2.91	0.48	12.1	2.9	7.77	3.88	0.97	2.91	6.31		2.4	2.91			0.48	0.48	1.46
Kazakhs (82)	Bermisheva 2004	14.63		4.83		7.32	3.66		1.22	8.54			1.22		1.22	2.44			11	4.88	17.1		2.44		3.66	1.22	8.5	3.66	1.2				
Kirghiz (92)	Bermisheva 2004	16.3		1.09	1.09	3.26	6.52	2.17	5.43	3.26		2.17				1.09			14.1	6.52	19.6				8.69	1.09	1.1	3.26		3.3			
Uighurs (99)	Bermisheva 2004	16.16			2.02	6.06	4.04		2.02	2.02		1.01	5.05		5.05		4.04		4.04	10.1	15.2	1.01	3.03	2.02	5.05	2.02		7.07		1		2.02	
Mongols (103)	Bermisheva 2004	7.77				3.88	9.71			0.97		0.97	2.91						14.6	11.64	30.1			0.97	2.91		3.9	5.83	1.9	1.9			
Bashkirs (221)	Bermisheva 2004	12.2	2.87	0.48		3.6	0.9	1.44	3.2	5.4			0.5		12.7	13.6		0.48	11.8	1.88	9		0.48	1.44	4.5		0.9	6.32	0.5	1.4	3.6		0.48
Turks (218)	Bermisheva 2004	28.7	0.93	4.63	1.39	0.46	0.46	2.31	10.2	12.04	0.46	4.63	1.39	7.41	1.85	0.93	0.93	0.46	1.39		1.85	2.78	2.31	5.56		2.78		0.46	0.5		1.39	1.85	
Azerbaijanis (48)	Bermisheva 2004	29.17	4.17	2.08	2.08			2.08	4.17	16.67				2.08	8.33	8.33			4.17			4.17	4.17	2.08				2.08				2.08	
Russians-Belgrade (69)	Malyarchuk 2002	37.68	7.25	2.9		0	0	4.35	11.6	8.69	18.84								1.45	0	0	1.45	1.45	1.45	0								
Russians-Krasnodar (49)	Malyarchuk 2002	38.78	8.16	4.08		0	0	2.04	18.4	12.24	10.2								0	0	0	0	0	0	0								
Ukrainians (36)	Malyarchuk 2002	41.7	5.6	0		0	0	0	8.3	8.3	16.7								0	0	0	2.8	2.8	8.3	0								
Western Iranians (101)	Malyarchuk 2002	43.56	7.92	1.98		0	0	2.97	14.9	8.91	8.91								0	0	0	0.99	2.97	6.93	0								
Eastern Iranians (25)	Malyarchuk 2002	0	0	8		4	8	4	8	16	24								0	4	4	0	4	4	0								
Turkmen (26)	Malyarchuk 2002	19.23	3.85	7.69		0	0	0	7.69	15.38	3.85								7.69	0	7.69	0	7.69	0	3.85								
Karachays (67)	Malyarchuk 2002	15.4	0	9.23	0	0	0	1.54	1.54	10.8	3.08	21.5	4.62	3.08	0	9.23	0	0	3.08	1.54	4.62	0	0	3.08	0	0	0	0	0	0	7.69	0	

95