Paper 5.6: Haplogroup C Report.

Abstract: Haplogroup C-M130 provides several useful downstream markers for deciphering the prehistory of Papuan, Australian, Japonic, Koreanic, Turkic, Mongolic, Tungusic and Native American languages. The haplogroups has two main branches, *C1* and *C2*. The *C1* branch is the genetic relic of human expansions during Marine Isotope Stage 3 and the initial human settlement of India, Island Southeast Asia, Australia, East Asia and Europe. The *C2* branch represents markers that expanded during Late Pleistocene.

Posted December 1, 2019
Genetic-Linguistic Interface
Dr. Michael St. Clair
Stuttgart, Germany
mstclair@genlinginterface.com

Please cite as: St. Clair, Michael R. 2019d. “Paper 5.6: haplogroup C report.” *The Genetic-Linguistic Interface.* Series 5 Papers: language prehistory from a Y-Chromosome perspective. Posted December 1, 2019. https://genlinginterface.com

Table of Contents for Paper 5.6 (Haplogroup C).

Paper 5.6. Haplogroup C-M130.

The Genetic-Linguistic Interface.

Section 1. Overview of Haplogroup C-M130.

As discussed in Paper 5.4, Section 1, the DR-M168 mutation represents the exodus of modern humans from Africa into the Levant roughly 100 thousand years ago. In this region, roughly 65 thousand years ago, three main haplogroups evolved from the M168 mutation: D-M174, E-M96, and C-M130. As discussed in the previously in Paper 5.5, E-M96 returned to Africa. However, C-M130 co-migrated out of the Levant with D-M174 about 50 thousand years ago during Marine Isotope Stage 3 (see Paper 5.4).

Focusing now on the internal phylogenetic hierarchy of the C-M130 main haplogroup, it should be noted that two different nomenclature standards are currently being used to report C-M130 data. Some studies (e.g. Wei et al. 2017) continue to use the phylogenetic tree as presented by Karafet et al. (2008). Other studies (e.g. Huang et al. 2017) utilize the current nomenclature standard that has been adopted by International Society for Genetic Genealogy (ISOGG). We utilize ISOGG (2017) and as such it should be noted that the current ISOGG perspective is far different than that presented by Karafet et al. (2008). Karafet et al. (2008) arranged the internal C-M130 phylogeny into five main branches: C1-M8, C2-M38, C3-M217, C4-M347, and C5-M356. However, ISOGG (2017) divides C-M130 into two main branches, C1-F3393 and C2-M217. For the sake of simplicity, the ISOGG perspective can be shortened and viewed as either *C1* or *C2*. Interestingly, the ISOGG (2017) perspective closely follows the archaeological record whereby *C2* expanded from a refugium in south central Siberia towards the beginning of the Holocene. *C1*, on the other hand, represents earlier human migrations during Marine Isotope State 3. These expansions signal the initial human settlement of India, Island Southeast Asia, Australia, Japan and Europe.

Section 2. Overview of *C1* Mutations.

At this point the reader is invited to review Figures 5.6.1 and 5.6.2. Both focus on downstream variation within the C1-F3393 mutation. According to Poznik et al. (2016: Supplementary Table 10), C1-F3393 separated from C-M130 about 45 thousand years ago. It would appear, based on the genetic and, archaeological and paleo-climatic data, that this may have occurred in northern India. Turning now to Figure 5.6.1, this chart focuses on C1b-F1370 mutations that expanded along a southern dispersal route during Marine Isotope Stage 3 (see Section 3 below). Figure 5.6.2, on the other hand, focuses on C1b-F1370 mutations that expanded along a northern dispersal route during the same period (see Section 4).

Section 3. Eastward Expansion of C1b-F1370 Mutation across Eurasia via a Southern Route.

3.1. Overview.

Many researcher favor the human colonization of East Asia during Marine Isotope Stage 3 (roughly 50 thousand years ago) via a single southern route, the so-called southern dispersal hypothesis (see Paper 5.4, Hg. D, Section 1). Turning now to Figure 5.6.1, the reader will notice the use of color clustering to highlight particularly important mutations within the C1b-F1370 branch. The C1b-M356 "gold cluster" mutation represents the initial human colonization of South Asia and is only found in India, Nepal and Pakistan. The C1b-

M38 "green cluster" mutation is confined to Island Southeast Asia and Oceania. This mutation stands as genetic relic of the human colonization of New Guinea and Indonesia. The human colonization of Australia is represented by the C1b-M347 "blue cluster" mutation, which is only found on this continent.

3.2. C1b-M356 Gold Cluster Mutation and Pleistocene Colonization of India.

As noted earlier, the C-M356 mutation is restricted to India, Nepal and Pakistan (see, also, Sengupta et al. 2006). Population studies reporting frequency data for C1-M356 indicate that this mutation attains a small frequency within the region, somewhere around five percent or even less. Despite the low frequency numbers, C1-M356 is still a significant mutation as it represents the genetic relic of the founding population of India (e.g. Sengupta et al. 2006; Arunkumar et al. 2012; Khurana et al. 2014). This agrees with the dating estimate provided Poznik et al. (2016: Supplementary Table 10), who report that the C1-M356 mutation evolved roughly 44 thousand years ago.

Contemporary India has a huge population with almost 1.3 billion people (CIA World Factbook 2017). The traditional social hierarchy consists of either castes or tribes. Together with a large population and complex social structure, one finds incredible linguistic diversity. *Ethnologue* (twentieth edition) lists 462 languages for India. Almost all these languages fall within one of the four language families: Dravidian, Indo-European, Austro-Asiatic or Sino-Tibetan. Thus for linguists, the C1-M356 mutation represents a starting point for gaining an understanding of the correlation between genetic and linguistic diversity in India. This analysis starts with identifying mutations that are "Paleolithic" like the C1-M356 marker, and identifying mutations that represent more recent migrations during the Mesolithic or Neolithic. Accordingly, the discussion of linguistic diversity on the Indian subcontinent continues in Paper 5.8 with the presentation of haplogroup H-M2713.

3.3. C1b-B477 and the Colonization of Sunda and Sahul.

As noted above, variants of the C1b-F1370 mutation expanded into South and East Asia roughly 50 thousand years ago during Marine Isotope Stage 3. As shown by Figure 5.6.1, C1b-F1370 splits into C1b-K281 and C1b-B477. Downstream from C1b-K281 is the C1b-M356 mutation, which represents, as discussed above, the human colonization of India. Focusing now on another C1b-F1370 variant, the C1b-B477 mutation, we find a genetic artifact of the human colonization of Australia and Island Southeast Asia. It should be noted that we define Island Southeast Asia as Indonesia and Papua New Guinea.

According to Karmin et al (2015: Table S7 and Figure S21), C1-B477 evolved about 49 thousand years ago. Downstream from C1b-B477, C1b-M38 is an informative mutation that evolved in Island Southeast Asia, C1b-M347, which also downstream from C1b-B477, evolved in Australia. In order to understand the human colonization of Island Southeast Asia and Australia, it should be noted that during the Last Ice Age sea levels were around considerably lower than present day levels (e.g. Clark et al. 2009). As a result of lower sea levels, a large landmass called "Sunda" connected the present-day Malaysian Peninsula and many of the contemporary Indonesian Islands, including Sumatra, Java, Borneo, and Bali. At the same time, the Sahul landmass connected Papua New Guinea and Australia (see Figure

5.6.4 for additional details). Since the distance between Sunda and Sahul may have been as short as ninety kilometers, a water crossing between both landmasses seems quite feasible even with primitive watercraft (Allen and O'Connell 2008). Additionally, we note the important fossil remains that support the initial human colonization Sunda and Sahul. Lake Mungo man in Australia is dated to at least 46 thousand years ago (see Bowler et al. 2003). The so-called "Deep Skull" at the Niah Cave on the Indonesian Island of Borneo is at least 35 thousand years old (Barker et al. 2007).

3.4. C1b-M38 and Colonization of Island Southeast Asia and Oceania.

C1-M38 is found almost exclusively in Island South East Asia (e.g. Mona et al. 2007; Mona et al. 2009; Karafet et al. 2010; Tumonggor et al. 2014). Karmin et al. (2015) suggest that C1-M38 evolved about 24 thousand years ago. Based on their interpretation of the data, Mona et al. (2007) suggest that C1-M38 evolved in the northwestern part of New Guinea, and the mutation expanded both to the eastern part of the island (Papua New Guinea) as well as westward to Indonesia. For linguists, the C1-M38 mutation represents an indigenous component within the linguistic tapestry of Island Southeast Asia, and as such, stands as an informative marker for deciphering the evolution of the so-called "Papuan" macro-family of languages.

Downstream variants of the C1-M38 mutation are important mutations for decipher the evolution and expansion of the Austronesian language family. Both the linguistic and archaeological evidence suggest that this language family originated among the aboriginal populations on the island of Taiwan (e.g. Diamond 2000). Beginning about six thousand years ago, Austronesian-speaking populations expanded southwards from Taiwan into the Philippines. Another expansion then carried Austronesian into Papua New Guinea and Indonesian. From Papua New Guinea, a final expansion carried Austronesian eastwards across a vast ten-thousand-kilometer expanse of ocean. We define this vast expanse of ocean as Oceana, the numerous Pacific Ocean islands that extend from the Solomon Islands to Easter Island (Rapa Nui).

The genetic inventory of the populations that colonized Oceania has both an East Asian component from Taiwan and a Papuan component from Island Southeast Asia. The Taiwanese contribution will be discussed in Paper 5.15, Hg. O, Section 6. Focusing now on the Papuan component, about 12 thousand years ago the C1-M208 mutation, which is a downstream variant of C1-M38, evolved in the highlands of West New Guinea (Delfin et al. 2012; Karmin et al. 2015). C1-M208 is rarely found in Indonesia (e.g. Mona et al. 2009; Karafet et al. 2010), which suggests a minimal westward expansion of the mutation. However, in their 2012 study Mirabel et al. report that the C1-M208 mutation exhibits an increasing frequency cline from New Guinea across Oceania.

As C1-M208 moved across Oceania, a downstream variant of the marker, the C1-P33 mutation, evolved roughly 4.5 thousand years ago, possibly in the Tongan archipelago (Cox et al. 2007). As seafarers moved further east across Oceania, the C1-P33 mutation eventually became the only genetic mutation among the populations that colonized the region, which suggest the effects of founder effect. Thus, while the Taiwanese genetic component disappeared when the Austronesian expansion ended at Easter Island, Austronesian languages survived. For linguists this is an important observation for understanding language variation. Language continuity is sometimes maintained despite population replacement.

3.5. The C1b-M347 Mutation and the Colonization of Australia.

The reader is invited to review Figure 5.6.1. As detailed above, the founding populations of Sunda and Sahul had the C1-B477 mutation. Downstream variants of C1-B477, the C1-M38 "Green Cluster" and C1-M347 "Blue Cluster" mutations, later evolved among geographically isolated populations. C1-M38 represents a founding lineage for populations that colonized Island Southeast Asia. C1-M347 represents a mutation that evolved among the aboriginal people of Australia. The most comprehensive study of indigenous Australian Y-chromosome genetic variation, Nagle et al (2016), reports an overall frequency of around forty percent within this population. Furthermore, the same study suggests that C1-M347 evolved from C1-B477 about 44 thousand years ago.

The discovery of the Australian-specific C1-M347 mutation was initially reported in 2007 by Hudjashov et al. In this report, researchers utilized the enhanced resolution of downstream variation within the main haplogroup C-M130 marker to address a rumor of geneflow between India and Australia that occurred, according to another study (Redd et al. 2002), about 10 thousand years ago. The earlier study based their findings on a type of genetic marker called Short Tandem Repeats (STRs). Hudjashov et al. (2007) disagreed with the 2002 study and asserted that the Australian aborigines had not experienced any outside geneflow for a period of roughly 45 thousand years, from time that the continent was initially colonized by modern humans until the arrival of Europeans in the late eighteenth century (see, also, Nagle et al. 2016).

A complete discussion of indigenous Y-chromosome variation among the Australian aborigines is not complete without a discussion of haplogroups M-P256, and S-B254. Accordingly this topic continues in Paper 5.13, Section 6. In the meantime the following point needs emphasis: according to the genetic evidence, the Australian language family has roots that extend not only to the human colonization of Australia, but perhaps to the human exodus out of Africa about 100 thousand years ago.

Section 4. Bi-directional Expansion of C1a-CTS11043 across Eurasia via a Northern Route.

As explained above, downstream variants of C1b-F1370 are important mutations for understanding the human colonization of India, Island Southeast Asia and Australia via a southern route during Marine Isotope Stage 3. We now focus on C1a-CTS11043, the sister clade of C1b-F1370. At this point the reader may want to review Figure 5.6.2. As shown by the figure, downstream from C1a-CTS11043 are C1a-M8 and C1a-V20. Based on data from Poznik et al. (2016: Supplementary Table 10), the C1a-M8 and C1a-V20 mutation evolved about 44 thousand years ago.

As mentioned previously in Paper 5.4, Section 4, the D-M55 and C1a-M8 mutations stand as the genetic relics of the human colonization of Japan roughly 30 thousand years ago.

Data from Sato et al (2014) suggest that about six percent of contemporary Japanese have the C1-M8 mutation. Furthermore, C1a-M8 is a Japanese-specific mutation, meaning that it is not found elsewhere, at least among contemporary populations (e.g. Hammer et al. 2006).

Surprisingly, ancient DNA results indicate a close phylogenetic relationship between modern Japanese with the C1a-M8 mutation and individuals who colonized Europe during Marine Isotope Stage 3. This is surprising because haplogroup C-M130 is rarely found among contemporary Europeans. We know that C1a-M8 and C1a-V20 were part of the genetic inventory of Paleolithic Europeans based on ancient DNA data acquired from individuals that died between 13 thousand and 35 thousand years ago. Part of the ancient DNA data stems from a 2016 report by Fu et al. This study reports a sample taken from remains found in Belgium. The sample belongs to C1a-CTS11043 and comes from the Goyet Q116-1 man who died about 35 thousand years ago. Another sample from the study, the Vestonice man, comes from the Czech Republic and belongs to C1a-V20. This individual died about 30 thousand years ago. The C1a-V20 mutation was also found in Russia. The Sunghir 1 man, who died about 34 thousand years ago near present-day Moscow (Sikora et al. 2017), has this mutation. Finally, C1a-P121, which is downstream from C1a-M8, was found in remains in Spain that date to around 13 thousand years ago (Villalba-Mouco et al. 2019).

Ancient DNA data from Paleolithic Europeans and East Asians should not be used to define a close genetic relationship between contemporary Europeans and contemporary Japanese. Rather the data supports a “northern migration” route during Marine Isotope Stage 3. In other words, about 50 thousand years ago the human tribe in the Levant split into two different groups. One group followed a southern route or a migration along the coastline of southern Asia. The genetic relics of this migration are the D-M55 and C1b-F1370 mutations. Another group, however, migrated northwards from the Levant. Somewhere in Eastern Europe or Central Asia, another split occurred. Some traveled west in the direction of contemporary Belgium, and the other group traveled eastwards in the direction of contemporary Japan. The genetic relics of this bi-directional northern expansion include C1a-M8 and C1a-V20.

Section 5. The Importance of *C1* Mutations for Linguists.

The aboriginal Australians remained isolated from the rest of the world until about two hundred years ago. Thus, the Australian language family has roots that extend to the out-of-Africa exodus. Taking this a step further, the C1b-M347 mutation, which is only found only among aboriginal Australians, supports the position that language evolved at least 100 thousand years ago. This follows the initial out-of-Africa migration and makes a huge (but plausible) assumption that the out-of-Africa tribe already had language. Turning now to C1-M38, this mutation represents an important genetic tool for deciphering the evolution of Papuan languages and the spread of Austronesian across the Pacific. The diversity of languages on New Guinea, as represented by the Papuan macro-family, is partly explained by the age of the C1b-M38 mutation. The C1b-M208, which is downstream from C1b-M38, resents an Island Southeast Asian component of Austronesian languages. Focusing now on C1b-M356, this mutation helps to explain linguistic diversity in India by defining indigenous and non-indigenous components of the gene pool. Finally, mutations downstream from C1a-CTS11043 support northern bi-directional migrations across Eurasia during Marine Isotope

Stage 3. These migrations represent an important component of the *mammoth steppe hypothesis*, a discussion that surfaces in Papers 5.14, 5.16, and 5.17. This hypothesis, in turn, helps to explain language variation in Eurasia and the Americas.

Section 6. Overview of the *C2* Mutations.

As previously detailed in Section 2, the C-M130 main haplogroup has two main branches, C1-F3393 and C2-M217, or alternatively, *C1* and *C2.* Both diverged from the C-M130 main haplogroup about 45 thousand years ago, during Marine Isotope Stage 3. Perhaps this occurred when the out-of-Africa human migration reached northern India. C1-F3393 then expanded rapidly across the Eurasian landmass. Those with C2-M217, on the other hand, appear to have "nested" in south central Siberia. Several thousand years later, after the Last Glacial Maximum, *C2* then expanded in the direction of East Asia, and then northwards into the Americas. The current distribution of C2-M217 and its variants now present a very useful tool for interdisciplinary analysis of the so-called *Transeurasian hypothesis* as presented in Section 7 (below). Additionally, the moderate frequency of C2-M217 found among Han Chinese, as discussed in Section 8, helps to decipher the evolution of Sino-Tibetan languages. Finally, *C2* mutations help to decipher the evolution of Native American languages (Section 9).

Section 7. Altaic and Transeurasian.

7.1 Overview.

Striking lexical and grammatical similarities found among the Japonic, Koreanic, Turkic, Tungusic, and Mongolic languages (e.g. Robbeets 2008) have been a topic of intense interest among linguists. The *Transeurasian hypothesis* has been formulated to explain these similarities (e.g. Robbeets 2017). An approach to this hypothesis from the perspective of historical linguistics would classify these language families as part of an Altaic or Transeurasian macro-language family (or macro-phylum). As such, linguistic similarities are explained by the evolution of Japonic, Koreanic, Turkic, Tungusic, and Mongolic from a common proto-Altaic or proto-Transeurasian language. At this point the reader is directed to Figure 5.6.5, which illustrates the *Transeurasian hypothesis* from the perspective of historical linguistics.

An alternative socio-linguistic approach to the Transeurasian hypothesis would view Japonic, Koreanic, Turkic, Tungusic, and Mongolic as part of a northeast Asian *Sprachbund*. As such, linguistic similarities stem from close geographical proximity and borrowing that has evolved over a prolonged period of time due to intense contact between the speakers of these languages. The reader is now directed to Figure 5.6.7, which illustrates the evolution of Transeurasian languages from the perspective of language contact theory.

At the Max Planck Institute for the Science of Human History, Dr. Martine Robbeets currently leads a project that explores the origins and expansion of the so-called Transeurasian languages: (http://www.shh.mpg.de/102128/eurasia3angle_group). The informal title for this the project is "Millet and Beans, Language and Genes," which reflects willingness among the researchers to employ multi-disciplinary perspectives in an effort to resolve a long-standing linguistic controversy. In a recent paper from 2017, Dr. Robbeets provides her views on the origins of Transeurasian from linguistic, archeological and genetic perspectives. One reason for citing this paper is that Dr. Robbeets suggests that haplogroup N-M231 is an informative for exploring the Transeurasian hypothesis. The evidence suggests, however, that N-M231is not a significant marker for Transeurasian languages, but rather for Uralic languages (see Paper 5.14, Section 4.1). From a Y-chromosome perspective, C2-M217 and its downstream variants seem to represent the markers of choice for exploring the *Transeurasian hypothesis*.

7.2. Origins of the C2-M217 Expansion.

The currently known C2-M217 internal phylogeny consists of four main lineages or clusters: the C2-P39 "purple cluster," the C2-M48 "red cluster," the C2-F1918 "green cluster," and the C2-M407 "blue cluster." C2-P39 is found among Native Americans. The remainder represents genetic diversity in East Eurasia. These four clusters evolved roughly fourteen thousand years ago, at the onset of the Holocene. See Zhong et al (2010) for C2-M407; Wei et al. (2017: Supplementary Figure S1 for C2-F1918; Karmin et al (2015) for C2-M48; Malyarchuk et al. (2011) for C2-P39. Additionally, available frequency data for C2-M48, C2-F1918 and C2-M407 among Turkic, Tungusic and Mongolic speaking populations reflect potential language contact among these groups. As this point the reader is directed to Figure 5.6.3 and Tables 5.6.2, 5.6.3 and 5.6.4 for additional information.

As explained in the previous paragraph, C2-M217 expanded during the Holocene. The next question seeks to identify the geographic origins of the expansion. Accordingly, a discussion of Ice Age refugia is necessary in order to provide important background information that helps to evaluate the Transeurasian hypothesis. The term "refugia" is the plural form of "refugium." For the purposes of this present discussion, both terms carry a discussion of where human populations congregated during the Last Glacial Maximum. As the reader may recall from Paper 5.4, Hg. D, Section 1, roughly 50 thousand years ago, during Marine Isotope Stage 3, the weather across southern Asia improved. This facilitated a rapid expansion of the human tribe from the Levant to Asia and Australia. However, the weather deteriorated and the ice glaciers eventually reached their maximum southern extent in the northern hemisphere about 27 thousand years ago, a point that roughly equates to the fortieth northern parallel (see, e.g., Clark 2009). Geologists and earth scientists commonly refer to this event as the Last Glacial Maximum, a final and dramatic conclusion to a long Ice Age. The advance of ice glaciers curtailed human migration and populations expansions, and in fact, forced humans into several refugia across the Eurasian landmass where they waited for better weather.

Once the glaciers retreated, some populations, such those as in present-day Japan and Australia, remained isolated. However, populations in other refugia expanded with the onset of the Holocene. For geneticists, the isolation of populations during the Last Glacial Maximum, and their subsequent post-glacial expansion during the Holocene, or alternatively, their continued isolation, represents a partial explanation for global genetic diversity. For linguists, this provides a partial explanation for global linguistic diversity. A study from 2016 by Gavashelishvili and Tarkhnishvili used computer simulation to identify the refugia where human survived during the Last Glacial Maximum. Additionally, they identified the human Y-chromosome haplogroups that expanded from these refugia with the onset of the Holocene. Their model was constructed utilizing a synthesis of climate, terrain, and hydrographic data, as well as data from fossilized pollen and plant remains. Data from Gavashelishvili and Tarkhnishvili (2016) places one of these refugia in the vicinity of south central Asia.

This research guide now proposes, based on the paleo-climatic, anthropological, and genetic evidence, that C2-M217 (present paper), Q-M242 (see Paper 5.16) and R-M207 (Paper 5.17) co-expanded from the same refugium in south central Siberia after the Last Glacial Maximum. This refugium, the so-called Altai-Sayan region, is located where China, Russia, Kazakhstan and Mongolia converge on a map. This area, along with much of Central Asia, has long been characterized by low precipitation and a vast stretch of prairie or "the steppes." These characteristics provide an explanation as to why this region may well have become an Ice Age refugium. Dolukhanov (2003a) suggests that during the Last Ice Age, northern and central Europe were depopulated because of thick layer of ice and snow. However, the Central Asian steppes were covered by just a thin layer of snow because of low precipitation in the region. As such, the steppes provided an ideal habitat for a variety of large mammals including mammoths, woolly rhinoceros, wild horses, and bison. Even during the winter months, these animals could easily forage as they simply had to scrape away a thin layer of snow to access the grass underneath. The Ice Age hunter-gatherers, in turn, hunted these large mammals that thrived in the region, and feasted on an abundant source of protein that could be harvested at a comparatively small expenditure of energy.

Y-chromosome data provided by Zhabagin et al (2017) may also identify a south central Siberian refugium for haplogroups C2-M217, Q-M242 and R-M207. This study analyzed 780 samples from the nearby Central Asian region of Transoxiana. Source populations for the data include Kazakhs, Uzbeks, Turkmen, Dungan and Karakalpak. According to data provided by Zhabagin et al (2017), among the populations of the region, C2-M217 attains an overall frequency of thirty-one percent, R1a1a-M198 attains sixteen percent, and Q-M242 attains thirteen percent.

7.3. Transeurasian and Agriculture.

In her 2017 paper, Dr. Robbeets also suggests that the origins and expansion of Transeurasian languages follow the evolution and expansion of millet cultivation that began about eight thousand years ago at Xinglonggou in Inner Mongolia. Nevertheless, it would appear as though millet cultivation carries only part of the story that explains the evolution of the Transeurasian languages. Another important crop is rice. This explains the high population density observed in Korea and Japan. According to Stevens and Fuller (2017), millet and rice cultivation began in China roughly eight thousand years ago. From southeastern Manchuria, the cultivation of foxtail and broomcorn millet eventually spread from China to Korea about 5.5 thousand years ago, followed by the spread of rice cultivation about 3.5 thousand years ago. Around 2.5 thousand years ago, millet and rice cultivation spread from Korea to Japan.

The story of agriculture in Central Asia is also an important factor in the evolution of Altaic languages (Turkic, Mongolic, and Tungusic). In this region, agriculture began about 5.5 thousand years ago when the horse was first domesticated, which appears to have occurred north-central Kazakhstan (e.g. Frachetti 2012). As mentioned earlier, horses were one of several large mammals hunted by prehistoric peoples who lived in the south central

Siberian refugium around the time of Last Glacial Maximum. Thus, the domestication of the horse should be seen as an effort to ensure its continued availability as a source of food. The adaptation of this animal as a means of transport occurred later when people began to ride horses. Soon thereafter, horses were used as draft animals to pull wagons and chariots.

It should also be emphasized that horse domestication represents only part of the success of agriculture in Central Asia. Around 4.5 thousand years ago cattle, goats and sheep appeared in the region and became part of the food economy. Then another important step in the evolution of Central Asian agriculture occurred shortly thereafter, about four thousand years ago, when mobile pastoralists began to cultivate crops such as millet, barley and wheat. China was the source of millet that was initially grown in Central Asia (Stevens and Fuller 2017). Barley and wheat, as well as goats and sheep, however, came from the Middle East (Bellwood 2005: 84-86; Spengler et al. 2014).

Robbeets (2017) suggests that an expansion of millet cultivation from China brought Transeurasian languages to Central Asia. It is difficult to find genetic support because the internal phylogeny of C2-M217 requires greater clarification. However, according to the archaeological record by around three thousand years ago nomadic pastoralism spread from Central Asia to Mongolia (Askarov et al. 1992). Perhaps this expansion carried Transeurasian languages to East Asia.

7.4. C2-M217 and Turkic.

As noted earlier, the *Transeurasian hypothesis* seeks to explain the origins of Transeurasian languages. One of the languages families within this macro-family classification is Turkic. The reader is now invited to examine Table 5.6.5, which presents a survey of Turkic-speaking populations that have appeared in published Y-chromosome studies in the last 20 years. The table illustrates the fact that Turkic-speaking populations appear over a wide geographical expanse. Examples include the following: Turks in Southwest Asia; Azerbaijani in the Caucasus; Kazakhs, Kyrgyz, Turkmen, and Uzbeks in Central Asia; Ainu (Änyu) of East Asia; and Yakuts of Siberia. Where and when Turkic languages evolved appears to still be very much a mystery (e.g. Kornfilt 2009). However, Orkun Inscriptions found in Mongolia and Old Uyghur manuscripts found in Xinjiang, China from about the eighth or ninth century, point to East Asia.

At this point the reader is directed to Table 5.6.6 which reports the frequency of C2-M217 among several different Turkic-speaking populations. Among the Central Asians Kazakhs, the frequency of C-M217 is very high. Based on the high frequency of C2-M217 among the Kazakhs, as well as the distribution and frequency of C2-M217 throughout East Eurasia, the genetic evidence potentially identifies Central Asia as the putative homeland of Turkic languages. However, the internal phylogeny of C2-M217 still remains a mystery because almost the effort devoted to refining downstream C2-M217 markers involves those that likely evolved in Mongolia or Northeastern China: the C2-M48 “red cluster,” the C2-F1918 “green cluster,” and the C2-M407 “blue cluster” mutations (see, also, Figure 5.6.3 and Tables 5.6.2, 5.6.3 and 5.6.4).

Zerjal et al. published a study in 2003 that claimed to have found a unique Y-chromosomal haplogroup C-M130 lineage based on unique pattern of Short Tandem Repeats (STRs). The study reports that this lineage was found among sixteen of Central Asian populations with an overall frequency of around eight percent. According to the study, the lineage evolved in Mongolia about 1,000 years ago and was spread to Central Asia by Genghis Khan and his descendants. This lineage became known in the literature as the "Genghis Khan Y-profile" or "Genghis Khan star-cluster" (e.g. Abilev et al. 2012, Wei et al. 2017). However, in 2017, Wei et al. determined that the C2-F1918 haplogroup is the mutation that was previously identified in the literature as the Genghis Khan star-cluster. The researchers further report that C2-F1918 evolved about fifteen thousand years ago, a date that clearly predates the Mongol Empire, which, in turn, undermines the purported reproductive success of Genghis Khan.

Abilev et al. in their 2012 study, based on their analysis of the "Genghis Khan star-cluster," report that around seventy-six percent of the Kazakh Kerey tribe have this mutation. The Kerey are the largest of the Kazakh tribes. According to the study, the Keraits, a Mongolic tribe, were defeated by Genghis Khan. Many escaped and joined the Turkic people. They adopted Turkic language and the term "Kerey" is a Turkic form of "Kerait." The study further reports that many of the contemporary Turkic ethnic groups evolved from Keraits, including Tatars, Karachays, Nogays, Bashkirs, Kazakhs, Uzbeks, Kyrgyz, and Altaians.

About thirty percent of all Turkic-speakers in the world reside in modern-day Turkey (e.g. Kornfilt 2009). The largest Y-chromosome survey of Turkey (Cinnioglu et al 2004) indicates that less than one percent of Turkish males have the haplogroup C-M130 mutation. This agrees with the historical record and follows the demise of the Byzantine Empire. Thus, language shift occurred in Anatolia without significant admixture with Turkic-speakers from Central Asia or Northern Eurasia. This underscores the following: language expansion can occur in the absence of a population expansion. Thus, language shift appears to partially explain the expansion of Turkic language. The Yakuts, a Turkic-speaking population of Siberia, provide yet another example. Their reliance on reindeer herding and the high frequency of N-M231 suggest that they initially spoke a Uralic language (e.g. Pakendorf et al. 2006).

7.5. C2-M217 and Mongolic.

Another Transeurasian language is Mongolic. *Ethnologue* (twentieth edition) classifies thirteen languages within the Mongolic language family. Twelve of the languages are spoken either in China, Russia or Mongolia. The other Mongolic language, Mogholi, is found in Afghanistan. Arguably, the earliest attestation of Mongolic languages are the so-called "Para-Mongolic" Khitan scripts dating to about the tenth century (Kane 1989: 11-37; Janhunen 2003a: 394-396), which prepared under the auspices of the Liao Dynasty. Pre-Classical Mongolic texts later emerged during the reign of Genghis Khan in the thirteenth century (Janhunen 2003b: 32-33).

At this point the reader is directed to Table 5.6.7, which provides a survey of Mongolic-speaking populations living in Russia, Mongolia and China. As shown by the table, C2-M217 attains a very high frequency among some Mongolic-speaking populations, similar to what is observed among speakers of Turkic languages (see Section 7.4) and

Tungusic languages (see Section 7.6). The reader is also directed to Tables 5.6.2, 5.6.3 and 5.6.4, which report C2-M407, C2-M48 and C2-F1918 variation among Turkic, Mongolic and Tungusic-speakers. In a recent study from 2017, Huang et al further refined the phylogeny of C2-M407 (see, also, Figures 5.6.3 and 5.6.6). According to the study, C2-F8465, a downstream variant of C2-M407, represents the genetic signature of Mongolic languages. The study further reports that this mutation evolved roughly four thousand years ago in Northeast Asia. Thus, unlike Turkic, the putative homeland of Mongolic languages seems much clearer.

7.6. C2-M217 and Tungusic.

Tungusic is another Transeurasian language. According to *Ethnologue* (twentieth edition), the Tungusic language family consists of eleven languages spoken by around 55 thousand speakers either in Northeastern China or Eastern Siberia. Tungusic languages include those spoken by the Even and Evenki people. These closely related ethnic groups consist of small populations in Siberia whose survival strategy once included the domestication of reindeer. In contrast, another Tungusic language, Manchu, stands as a former linguistic heavyweight, a relic of the Qing Dynasty of China. Thus, Manchu became a very significant East Asian language. However, the Qing Dynasty eventually collapsed in 1912, and as a result, the Manchu language rapidly became moribund.

The reader is invited to examine Table 5.6.8, which provides a survey of Tungusic-speaking populations and reported frequencies of C2-M217. Like Turkic and Mongolic, these populations also exhibit a high frequency of the mutation. The reader is also invited to examine Tables 5.6.2, 5.6.3 and 5.6.4, which provide frequency data for the known C2-M217 Holocene expansion markers. Based on these data, language contact between proto-Turkic, proto-Mongolic and proto-Mongolic populations seems possible. Nevertheless, small population size and the associated phenomenon of genetic drift limit the ability of genetic markers as a tool for identifying the geographic origins of Tungusic languages (e.g. Duggan et al. 2013). The earliest attestation of Tungusic stems from texts that appeared in the twelfth century. Under the auspices of the Jin Dynasty, these texts were written in the Jurchen language using characters borrowed from Khitan (a Mongolic language) and Chinese (Kane 1989:1-10). These texts, along with heavy frequencies of N-M231 found in the Tungusic populations of Siberia (e.g. Pakendorf et al. 2007; Fedorova et al. 2013) point to northeastern Asia as the putative homeland of Tungusic.

Variants of haplogroup N-M231 represent the genetic signature of Uralic speakers and reindeer herders (see Paper 5.14). Thus, it would appear that Tungusic speakers with C2-M217 migrated northwards from East Asia into Siberia sometime in the prehistoric past. Overtime, populations like the Even and Evenki began to herd reindeer, a subsistence strategy that is well suited to the Siberian climate. Admixture with Uralic-speaking population, as well as genetic drift, eventually produced the high frequency of N-M231 found in some Tungusic-speaking populations in Siberia (see Karafet et al. 2002 for a detailed discussion).

7.7. C2-M217 and Koreanic.

Koreanic represents another Transeurasian language. The Koreans stand among the ancient ethnic cultures of the world. A good starting point for discussing their ethnogenesis may well be the beginning of the Jeulmun pottery period about ten thousand years ago. However, Kim (2009) suggests that a reliable attestation of the Korean language emerged comparatively late in the Korean history, about six hundred years ago, when the Korean hangul script was introduced in a document called the *Hunminjeongeum*. According to the same source, classification of the Korean language has been difficult. The so-called “southern theory” attempted to associate Korean with Dravidian or Austronesian. The northern theory, on the other hand, classified Korean as part of an Altaic macro-family.

Contemporary linguistic classification of Korean has generally disassociated the language with Altaic. In their seventeenth edition, *Ethnologue* classified Korean as a language isolate. However, with the eighteenth edition, which was released in 2015, *Ethnologue* re-classified Korean within a newly created language family called Koreanic. This language family contains just two languages, with Korean having, by far, the largest number of speakers, which totals 48 million on the Korean peninsula, and 77 million worldwide. Jejueo, the other Koreanic language, has just five thousand speakers on Jeju Island in the Korean Straights.

At this point the reader is directed to Table 5.6.9. As shown by the table, around fifteen percent of Koreans have the C2-M217 mutation. The best refinement of C2-M217 variation among the Koreans emerged in 2015 with the study published by Kwon et al. At this point the reader is directed to Figure 5.6.6 and in particular, the mutations surrounded by a green border, which were reported in the study that was just cited. As shown by the figure, C2c-F1067 seems to unite the genetic history of Koreans with that of central Eurasia. Of course the genetic history of Koreans is not complete without a discussion of haplogroup O-M175. The reader is directed to Paper 5.15, Hg. O, Section 14, for additional information.

7.8. C2-M217 and Japonic.

The Transeurasian macro-family includes Japonic. Like Korean, the Japanese language has also proven difficult to classify. In the past, some linguists have placed this language within the Altaic super-family, along with Korean, Mongolic, Turkic and Tungusic (e.g. Shibatani 2009). *Ethnologue* (twentieth edition) currently places Japanese within the Japonic language family. The Japonic family has two main branches, the Japanese language, which is spoken by over 127 million people throughout Japan, and the Ryukyuan branch, which contains eleven languages, is spoken on the island of Okinawa. As detailed in Paper 5.4, Hg. D, Section 4 and Section 4 of this paper (5.6), the starting point for a discussion of Japonic languages begins roughly 30 thousand years ago with the initial human colonization of the present-day Japanese islands. As explained in these sections, the prehistoric Jomon people represent the cultural relic of this Paleolithic migration. Haplogroups D1b-M55 and C1-M8, on the other hand, provide the genetic artifacts.

The Neolithic Yayoi culture represents the second and last major human migration that settled in present-day Japan. Downstream variants of the O-M175 main haplogroup carry the bulk of the genetic evidence for this event, which clearly shows that Yayoi culture migrated to Japan from Korea (see Paper 5.15, Section 15 for additional details). Nevertheless, it should be noted that around seven percent of Japanese have the C2-M217 mutation (see Table 5.6.10). A comparison of downstream variants within C2-M217, as reported by Naitoh et al. 2013 and Kwon et al. 2015 (see, also, Figure 5.6.6), point to Korea as the source of C2-M217 variation in Japan. Thus C2-M217, like O-M175, stands as a genetic relic of the Yayoi culture.

7.9. Analysis of the Transeurasian Hypothesis.

Millet cultivation represents only a small part of the story of Transeurasian languages. The ancestral populations of contemporary groups that now speak Transeurasian languages survived and thrived due to complex combination of factors that may have begun with successful adaptation to climate change during the Last Glacial Maximum. Turning now to the Altaic component, these languages thrived and survived because of mobile pastoralism, the successful domestication of the horse in Central Asia, the successful domestication of reindeer in Northern Eurasia, the adoption of sedentary agriculture in some regions, and the expansion and demise of nomadic societies such as the Mongol Empire. Of course, the phenomenon of language shift stands as another factor that should not be neglected. Turning now to the evolution of Koreanic and Japonic, the genetic data, along with the archaeological and historical record, reflect that unlike the Altaic languages, geographical and cultural isolation played a substantial role in the evolution of both language families. Furthermore, rice cultivation clearly distinguishes the evolution of Koreanic and Japonic with that of Altaic. Korean and Japanese now occupy a huge corner of the global linguistic tapestry because their ancestors found a survival strategy that supports a very high population density.

Section 8. Han Chinese and C2-M217.

Ethnologue (twentieth edition) classifies Chinese as both a macro-language and as a branch within the Sino-Tibetan language family. With over 1.2 billion speakers, it goes without saying that Chinese plays a significant role within the global tapestry of linguistic diversity. The best source of genetic data for exploring the evolution of the Chinese macro-language comes from the Han Chinese. They are, by far, the largest ethnic group in China, representing almost ninety-two percent of the population (CIA World Factbook).

At this point the reader is directed to Table 5.6.11 which provides C2-M217 frequency data for the Han. Based on this table and data extrapolated from Zhong et al. (2011), C2-M217 attains an overall frequency of about twelve percent among this population. However, based on a review of the published data, downstream C2-M217 variation among the Chinese still remains very much a mystery. We know that the Central Asian contribution is very minimal. Based on data from Wei et al (2017: Table S1), the frequency of Central Asian C2-M217 mutations (C2-M48, C2-M407 and C2-F1918) among the Han is virtually non-existent. Thus, researcher should look elsewhere for the genetic mutations that carry the story of Sino-Tibetan language. At this point, the reader is directed to the discussion in Paper 5.15, Hg. O, Section 3.

Section 9. Native Americans and C2-M217.

C2-M217 and haplogroup Q-M242 (see Paper 5.16) mutations represent important genetic tools for deciphering the prehistory of Native American languages. Among the Native American populations of North America, haplogroup Q-M242 carries about ninety-three percent of the indigenous genetic component, whereas C2-M217 represents the remaining seven percent (e.g. Zegura et al. 2004). However, in South America Q-M242 represents almost all of the indigenous Native American genes (Geppert et al 2011; Roewer et al. 2013; Jota et al. 2016). Here, C2-M217 is extremely rare among the indigenous populations. Pinotti et al. (2019), for example, suggest that C2-M217 has only been found in thirteen indigenous South Americans. The same study also reports that indigenous South American have a C2-M217 variant that is evolutionary distant from the C2-M217 variant found among the indigenous peoples of North America, the C2b-P39 mutation. According to Pinotti et al. (2019) the unique South American C2-M217 variant and the unique North American C2-M217 variant (C2b-P39) diverged from a common ancestor roughly 22 thousand years ago.

At this point the reader should be advised that we purposely avoid a detailed discussion of the unique and recently discovered South American C2-M217 variant. Such a discussion would require analysis of the differences between the ISOGG 2017 standard and ISOGG 2019 standard. We see no need to confuse the reader in order to carry a discussion of a linguistically insignificant mutation.

The reader is now invited to examine Figure 5.6.3. The C2-P39 mutation is highlighted by a purple border. As shown by the figure, C2-FGC28881.2 is a phylogenetic sister clade of C2-P39. This mutation was reported by Wei et al. in 2017. According to the study, C2-FGC28881.2 forms part of the gene pool of contemporary Koryaks. Among the Paleo-Siberian peoples of Asia, Koryaks have traditionally lived along the Bering Sea near the Kamchatka Peninsula. They speak a language belonging to the Chukotko-Kamchatkan language family. Moreover, they have traditionally employed a hunter-gather subsistence strategy that included the harvesting of whales.

The above discussion of Koryaks serves a linguistic purpose which involves the bi-directional movement genes and culture across the Bering Sea. Such a discussion requires additional cultural context that follows the evolutionary history of the Q-M242 haplogroup among the indigenous peoples of Alaska. Accordingly, a discussion of Koryaks and C2-FGC28881.2 continues in Paper 5.16, Hg. Q, Section 8.

Section 10. The Importance of *C2* Mutations for Linguists.

Based on frequency data, the C2-M217 mutation represents a very important tool for understanding the evolution of the so-called “Altaic” macro-family of languages, which consists of the Turkic, Tungusic, and Mongolic language families. Additionally, C2-M217 attains a moderate to low frequency among the Japanese and Koreans. Thus, the mutation helps in the analysis of the so-called Transeurasian hypothesis, which advocates a common origin for the Japonic, Koreanic, Turkic, Mongolic and Tungusic language families. C2-M217 also attains a moderate frequency among the Han. However, unlike the Transeurasian languages, Central Asian variants of C2-M217 are almost non-existent among the Han. This, in turn, implies that C2-M217 is not an especially informative marker for the Sino-Tibetan language family. Finally, C2-M217 represents a useful marker for deciphering the origins of linguistic diversity among the native people of both Siberia and the Americas.

Section 11. Summary of C-M130 Data.

The evolutionary history of the *C1* and *C2* branches is so vastly different that one wonders if C1-F3393 and C2-M217 are, in fact, main haplogroups. Such a position means that C-M130 represents a higher level evolutionary step, or paragroup, within the Y-chromosome phylogeny. In terms of internal phylogeny, it is interesting to note that in the previous discussion of E-M96 (Paper 5.5), untangling the complex internal phylogeny of E-M96 presents a major undertaking. In comparison, C-M130 presents a much simpler internal phylogeny with its *C1* and *C2* branches. However, with C-M130, researchers must process and manage a tremendous amount of data that are vast in terms of time depth, geography, and linguistic significance.

Bibliography for Report

Abilev, Serikbai et al. 2012. “The Y-chromosome C3* star-cluster attributed to Genghis Khan's descendants is present at high frequency in the Kerey clan from Kazakhstan.” *Human Biology* 84(1):79-89.

Allen, Jim and James F. O'Connell 2008. “Getting from Sunda to Sahul.” In: *Islands of Inquiry: Colonisation, Seafaring and the Archaeology of Maritime Landscapes.* Geoffrey Clark, Foss Leach, and Sue O’Connor (Eds.). Canberra: ANU Press. 31-46.

Arunkumar, Ganesh Prasad et al. 2012. “Population differentiation of southern Indian male lineages correlates with agricultural expansions predating the caste system.” *Public Library of Science One* 7(11): e50269.

Askarov, A. et al. 1992. “Pastoral and nomadic tribes at the beginning of the first millennium B.C.” In: *History of Civilizations of Central Asia. Vol. 1, the Dawn of Civilization: Earliest Times to 700B.C.* A. H. Dani and V. M. Masson (Eds.). Paris: UNESCO Publishing. 450-463.

Barker, Graeme et al. 2007. “The ‘human revolution’ in lowland tropical Southeast Asia: the antiquity and behavior of anatomically modern humans at Niah Cave (Sarawak, Borneo).” *Journal of Human Evolution* 52: 243-261.

Bellwood, Peter 2005. *First farmers: the origins of agricultural societies*. Malden, MA; Oxford, UK; Victoria, Australia: Blackwell Publishing.

Bowler, James M. et al. 2003. “New ages for human occupation and climatic change at Lake Mungo, Australia.” *Nature* 421: 837-840.

CIA Work Fact Book. https://www.cia.gov/library/publications/the-world-factbook/geos/in.html Accessed December 15, 2017 and January 9, 2018.

Cinnioğlu, Cengiz et al. 2004. “Excavating Y-chromosome haplotype strata in Anatolia.” *Human Genetics* 114: 127-148.

Clark, Peter U. et al. 2009. “The Last Glacial Maximum.” *Science* 325: 710-714.

Cox, Murray P. et al. 2007. "A Polynesian motif on the Y chromosome: population structure in remote Oceania." *Human Biology* 79(5): 525-535.

Delfin, Frederick et al. 2012. "Bridging Near and Remote Oceania: mtDNA and NRY variation in the Solomon Island." *Molecular Biology and Evolution* 29(2): 545-564.

Diamond, Jared M. "Taiwan's gift to the world." *Nature* 403: 709-710.

Dolukhanov, Pavel M. 2003a. "Hunter-gathers of the Last Ice Age in Northern Eurasia." *Before Farming* 2003(2): 1-25.

Dolukhanov, Pavel M. 2003b. "Archaeology and languages in prehistoric Northern Eurasia." *Japan Review* 15: 175-186.

Duggan, Ana et al. 2013. "Investigating the prehistory of Tungusic peoples of Siberia and the Amur-Ussuri region with complete mtDNA genome sequences and Y-chromosomal markers." *Public Library of Science One* 8(12): e83570.

Dulik, Matthew C. et al. 2012. "Y-chromosome analysis reveals genetic divergence and new founding native lineages in Athabaskan- and Eskimoan-speaking populations." *Proceedings of the National Academy of Sciences of the United States of America* 109(22): 8471-8476.

Ethnologue: Languages of the World, Seventeenth edition, 2014. Lewis, M. Paul, Gary F. Simons, and Charles D. Fennig (Eds.). Dallas, Texas: SIL International. Online version: http://www.ethnologue.com/17/.

Ethnologue: Languages of the World, Eighteenth edition, 2015. Lewis, M. Paul, Gary F. Simons, and Charles D. Fennig (Eds.). Dallas, Texas: SIL International. Online version: http://www.ethnologue.com/18/.

Ethnologue: Languages of the World, Twentieth edition, 2017, Simons, Gary F. and Charles D. Fennig (Eds.). Dallas, Texas: SIL International. Online version: http://www.ethnologue.com.

Fedorova, Sardana A. et al. 2013. "Autosomal and uniparental portraits of the native populations of Sakha (Yakutia): implications for the peopling of Northeast Eurasia." *BioMed Central Evolutionary Biology* 13:127.

Frachetti, Michael D. 2012. "Multiregional emergence of mobile pastoralism and nonuniform institutional complexity across Eurasia." *Current Anthropology* 53(1): 2-38.

Fu, Qiaomei et al. 2016. "The genetic history of Ice Age Europe." *Nature* 534: 200-205.

Gavashelishvili, Alexander and David Tarkhnishvili 2016. "Biomes and human distribution during the last Ice Age." *Global Ecology and Biogeography* 25: 563-574.

Geppert, Maria et al. 2011. "Hierarchical Y-SNP assay to study the hidden diversity and phylogenetic relationship of native populations in South America." *Forensic Science International: Genetics* 5: 100-104.

Haber, Marc et al. 2012. "Afghanistan's ethnic groups share a Y-chromosomal heritage structured by historical events." *Public Library of Science One* 7(3): e34288.

Hammer, Michael F. et al. 2006. "Dual origins of the Japanese: common ground for hunter-gatherer and farmer Y-chromosomes." *Journal of Human Genetics* 51: 47-58.

Huang ,Yun Zhi et al. 2017. "Whole sequence analysis indicates a recent southern origin of Mongolian Y-chromosome C2c1a1a1-M407." *Molecular Genetics and Genomics*. doi: 10.1007/s00438-017-1403-4. [Epub ahead of print].

Hudjashov, Georgi et al. 2007. "Revealing the prehistoric settlement of Australia by Y chromosome and mtDNA analysis." *Proceeding of the National Academy of Sciences of the United States of America* 104 (21): 8726-8730.

Janhunen, Juha 2003a. "Para-Mongolic." In: *The Mongolic Languages*. Juha Janhunen (Ed.). Routledge: London. 391-402.

Janhunen, Juha 2003b. "Written Mongol." In: *The Mongolic Languages*. Juha Janhunen (Ed.). Routledge: London. 30-56.

Kane, Daniel 1989. *The Sino-Jurchen Vocabulary of the Bureau of Interpreters*. Bloomington, Indiana: Indiana University Institute for Inner Asian Studies.

Karafet, Tatiana M. et al. 2002. "High levels of Y-chromosome differentiation among Native Siberian populations and the genetic signature of a boreal hunter-gatherer way of life." *Human Biology* 74(6): 761-789.

Karafet, Tatiana et al. 2008. "New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree." *Genome Research* 18: 830-838.

Karafet, Tatiana M. et al. 2010. "Major east-west division underlies Y-Chromosome stratification across Indonesia." *Molecular Biology and Evolution* 27(8): 1833-1844.

Karmin, Monika et al. 2015. "A recent bottleneck of Y chromosome diversity coincides with a global change in culture." *Genome Research* 25: 459-466.

Khurana, Priyanka et al. 2014. "Y Chromosome haplogroup distribution in Indo-European speaking tribes of Gujarat, Western India." *Public Library of Science One* 9(3): e90414.

Kim, Nam-Kil 2009. "Korean." In: *The World's Major Languages*. Second edition. Bernard Comrie (Ed.). Oxon, UK; New York: Routledge. 765-779.

Kornfilt, Jaklin 2009. “Turkish and the Turkic languages.” In: *The World’s Major Languages*. Second edition. Bernard Comrie (Ed.). Oxon, UK; New York: Routledge 519-544.

Kwon, So Yeun et al. 2015. “Confirmation of Y haplogroup tree topologies with newly suggested Y-SNPs for the C2, O2b and O3a subhaplogroups.” *Forensic Science International: Genetics* 19: 42-46.

Malhi, Ripan Singh et al. 2008. “Distribution of Y chromosomes among native north Americans: a study of Athabaskan population history.” *American Journal of Physical Anthropology* 137: 412-424.

Malyarchuk, Boris et al. 2010. “Phylogeography of the Y-chromosome haplogroup C in northern Eurasia.” *Annals of Human Genetics* 74: 539-546.

Malyarchuk, Boris et al. 2011. “Ancient links between Siberians and Native Americans revealed by subtyping the Y chromosome haplogroup Q1a.” *Journal of Human Genetics* 56: 583-588.

Mirabal, Sheyla et al. 2012. “Increased Y-chromosome resolution of haplogroup O suggests genetic ties between the Ami aborigines of Taiwan and the Polynesian Islands of Samoa and Tonga.” *Gene* 492: 339-348.

Mona, Stefano et al. 2007. “Patterns of Y-chromosome diversity intersect with the Trans-New Guinea hypothesis.” *Molecular Biology and Evolution* 24(11): 2546-2555.

Mona, Stefano et al. 2009. “Genetic admixture history of Eastern Indonesia as revealed by Y-chromosome and mitochondrial DNA analysis.” *Molecular Biology and Evolution* 26(8): 1865-1877.

Nagle, Nano et al. 2016. “Antiquity and diversity of Aboriginal Australian Y-chromosomes.” *American Journal of Physical Anthropology* 159: 367-381.

Naitoh, Sae et al. 2013. “Assignment of Y-chromosomal SNPs found in Japanese population to Y-chromosomal haplogroup tree.” *Journal of Human Genetics* 58: 195-201.

Pakendorf, Brigitte et al. 2007. “Mating patterns amongst Siberian Reindeer herders: inferences from mtDNA and Y-chromosomal analyses.” *American Journal of Physical Anthropology* 133: 1013-1027.

Poznik, G. David et al. 2016. “Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences.” *Nature Genetics* 48(6): 593-600.

Redd, Alan J. et al. 2002. “Gene flow from the Indian Subcontinent to Australia: evidence from the Y Chromosome.” *Current Biology* 12: 673-677.

Robbeets, Martine 2008. The historical comparison of Japanese, Korean and the Trans-Eurasian languages. *Rivista degli studi orientali* 81:(1-4): 261-287.

Robbeets, Martine 2017. "Proto-Trans-Eurasian: Where and When?" *Man in India* 97(1): 19-46.

Sato, Youichi et al. 2014. "Overview of genetic variation in the Y chromosome of modern Japanese males." *Anthropological Science* 122(3): 131-136.

Sengupta, Sanghamitra et al. 2006. "Polarity and temporality of high-resolution Y-chromosome distributions in India identify both indigenous and exogenous expansions and reveals minor genetic influence of central Asian pastoralists." *American Journal of Human* Genetics 78: 201-221.

Shibatani, Masayoshi 2009. "Japanese." In: *The World's Major Languages*. Second edition. Bernard Comrie (Ed.). Oxon, UK; New York: Routledge. 741-763.

Sikora, Martin et al. 2017. "Ancient genomes show social and reproductive behavior of early Upper Paleolithic foragers." *Science* 3:358 (6363):659-662.

Spengler, Robert et al. 2014. "Early agriculture and crop transmission among Bronze Age mobile pastoralists of Central Eurasia." *Proceedings of the Royal Society* B 281: 20133382.

Stoneking, Mark and Frederick Delfin 2010. "The human genetic history of East Asia: weaving a complex tapestry." *Current Biology* 20: R188-R193.

Trejaut, Jean A. et al. 2014. "Taiwan Y-chromosomal DNA variation and its relationship with Island Southeast Asia." *BioMed Central Genetics* 15: 77.

Tumonggor, Meryanne K. et al. 2014. "Isolation, contact and social behavior shaped genetic diversity in West Timor." *Journal of Human Genetics* 59: 494-503.

Underhill, Peter A. et al. 2007. "New phylogenetic relationships for Y-chromosome haplogroup I." In: *Rethinking the Human Revolution*. P. Mellars et al. (Eds.) Cambridge, UK: McDonald Institute for Archaeological Research. 33-42.

Villalba-Mouco, Vanessa et al, 2019. "Survival of Late Pleistocene hunter-gatherer ancestry in the Iberian Peninsula." *Current Biology* 29(7): 1169-1177.

Wang, Chuan-Chao and Hui Li 2013. "Inferring human history in East Asia from Y chromosomes." *Investigative Genetics* 4:11.

Wei, Lan-Hai et al. 2017. "Phylogeny of Y-chromosome haplogroup C3b-F1756, an important paternal lineage in Altaic-speaking populations." *Journal of Human Genetics* 62: 915-918.

Zegura, Stephen L. et al. 2004. "High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y chromosomes into the Americas." *Molecular Biology and Evolution* 21(1): 164-175.

Zerjal, Tatiana et al. 2003. "The genetic legacy of the Mongols." *American Journal of Human Genetics* 72: 717-721.

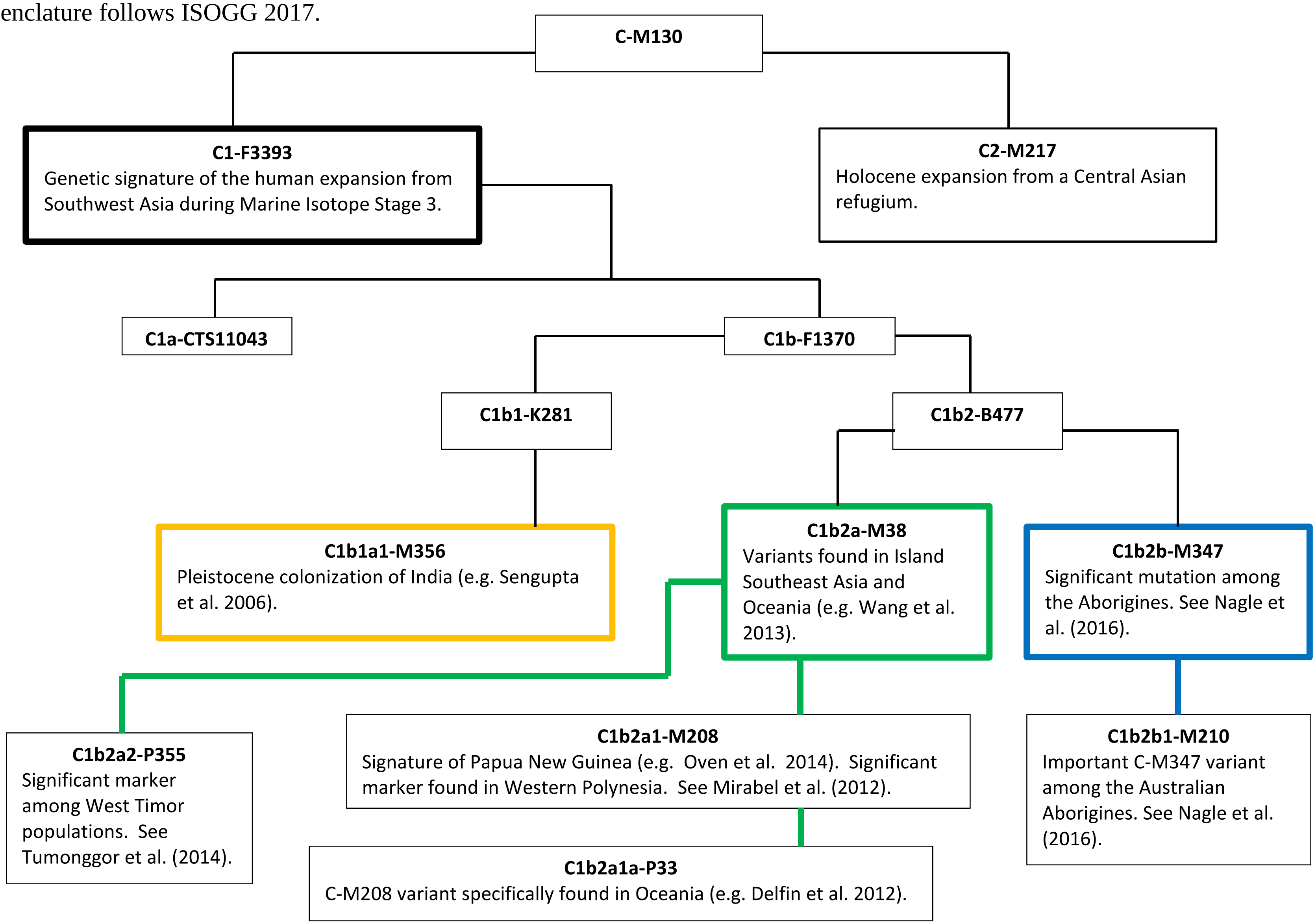
Zhabagin, Maxat et al. 2017. "The connection of the genetic, cultural and geographic landscapes of Transoxiana." *Scientific Reports* 7: 3085.

Zhong, Hua et al. 2010. "Global distribution of Y-chromosome haplogroup C reveals the prehistoric migration routes of African exodus and early settlement in East Asia." *Journal of Human Genetics* 55: 428-435.

Zhong, Hua et al. 2011. "Extended Y-chromosome investigations suggests postglacial migrations of modern humans into East Asia via the northern route." *Molecular Biology and Evolution* 28(1): 717-727.

Figure 5.6.1. Overview of C1-F3393 Southern Dispersal.

Nomenclature follows ISOGG 2017.



C-M130

C1-F3393
Genetic signature of the human expansion from Southwest Asia during Marine Isotope Stage 3.

C2-M217
Holocene expansion from a Central Asian refugium.

C1a-CTS11043

C1b-F1370

C1b1-K281

C1b2-B477

C1b1a1-M356
Pleistocene colonization of India (e.g. Sengupta et al. 2006).

C1b2a-M38
Variants found in Island Southeast Asia and Oceania (e.g. Wang et al. 2013).

C1b2b-M347
Significant mutation among the Aborigines. See Nagle et al. (2016).

C1b2a2-P355
Significant marker among West Timor populations. See Tumonggor et al. (2014).

C1b2a1-M208
Signature of Papua New Guinea (e.g. Oven et al. 2014). Significant marker found in Western Polynesia. See Mirabel et al. (2012).

C1b2b1-M210
Important C-M347 variant among the Australian Aborigines. See Nagle et al. (2016).

C1b2a1a-P33
C-M208 variant specifically found in Oceania (e.g. Delfin et al. 2012).

Figure 5.6.2. Overview of C1-F3393 Bi-Directional Northern Dispersal.

Nomenclature follows ISOGG 2017.

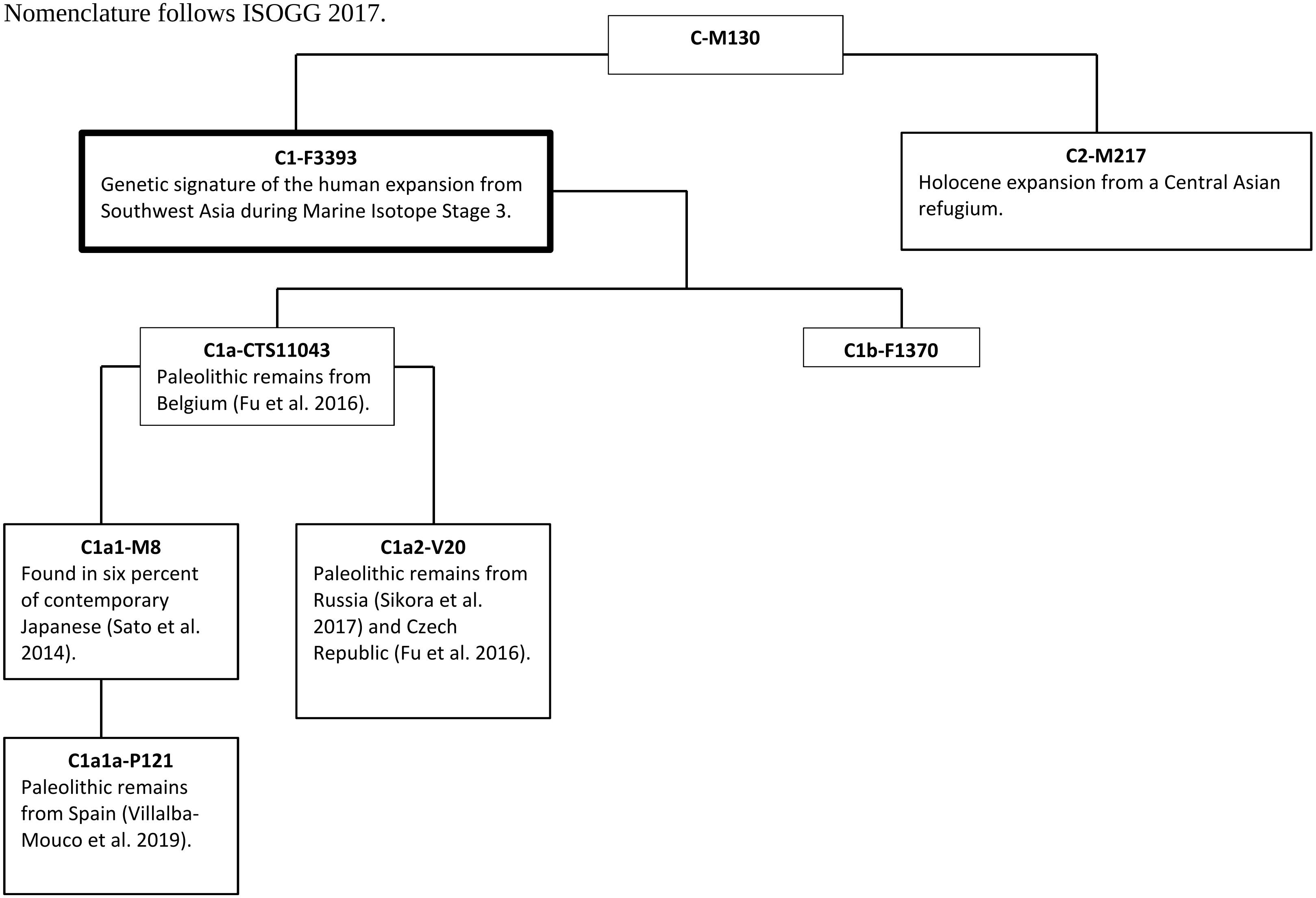
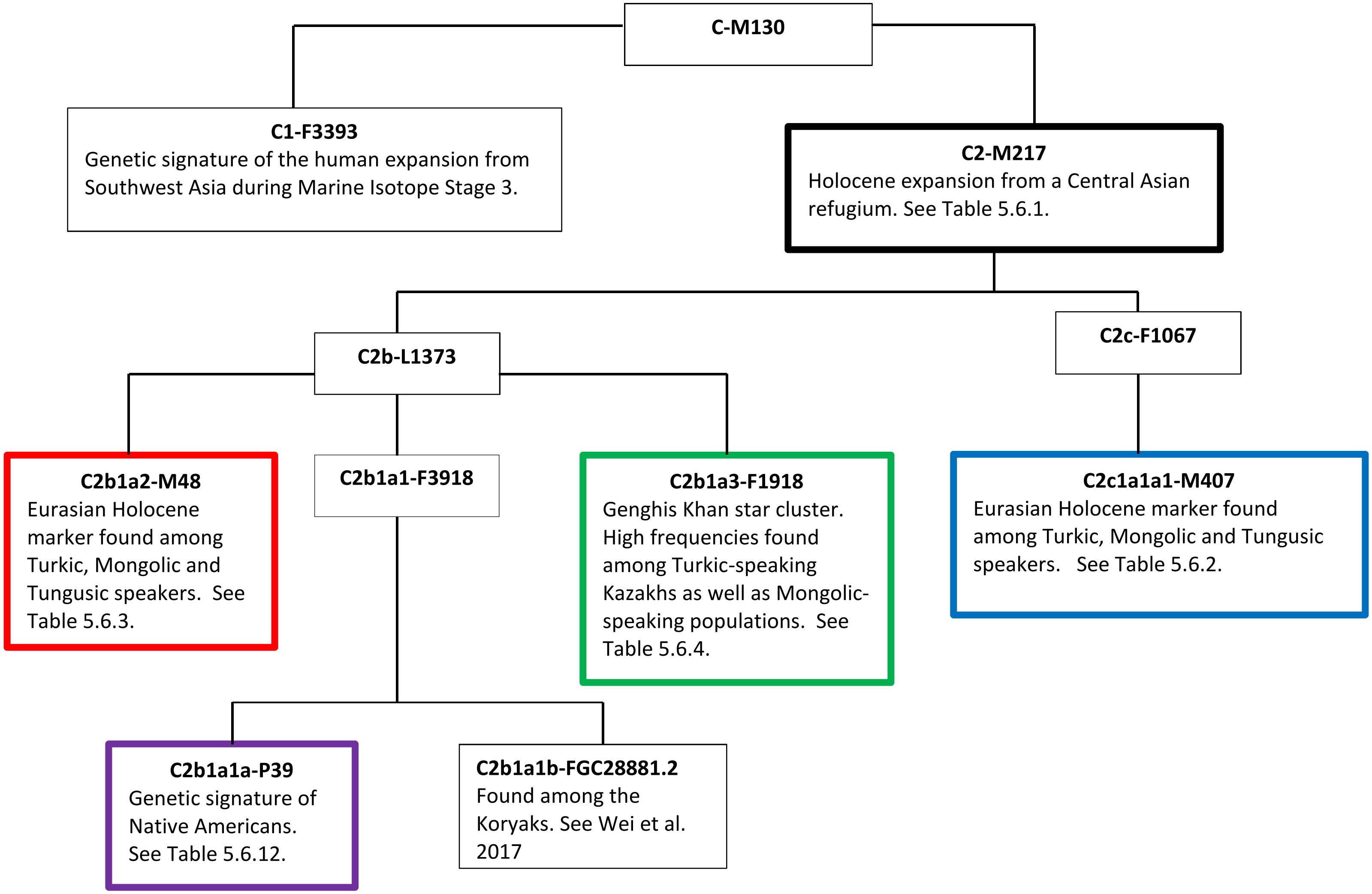


Figure 5.6.3. Overview of known C2-M217 variation in Eurasia and North Americans.

Nomenclature follows ISOGG 2017.



C-M130

C1-F3393
Genetic signature of the human expansion from Southwest Asia during Marine Isotope Stage 3.

C2-M217
Holocene expansion from a Central Asian refugium. See Table 5.6.1.

C2b-L1373

C2c-F1067

C2b1a2-M48
Eurasian Holocene marker found among Turkic, Mongolic and Tungusic speakers. See Table 5.6.3.

C2b1a1-F3918

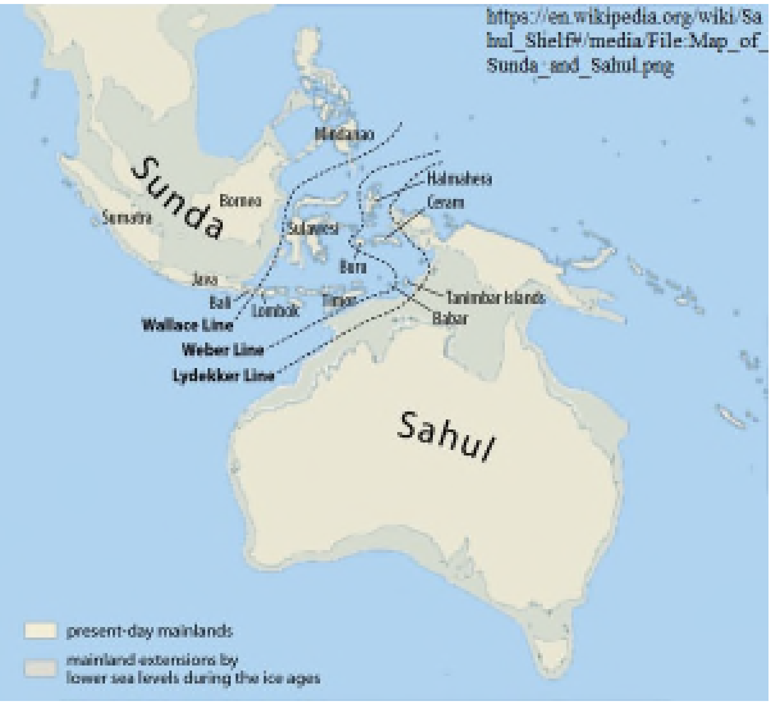
C2b1a3-F1918
Genghis Khan star cluster. High frequencies found among Turkic-speaking Kazakhs as well as Mongolic-speaking populations. See Table 5.6.4.

C2c1a1a1-M407
Eurasian Holocene marker found among Turkic, Mongolic and Tungusic speakers. See Table 5.6.2.

C2b1a1a-P39
Genetic signature of Native Americans. See Table 5.6.12.

C2b1a1b-FGC28881.2
Found among the Koryaks. See Wei et al. 2017

Figure 5.6.4. Sunda and Sahul about 50,000 Years Ago.



https://en.wikipedia.org/wiki/Sahul_Shelf#/media/File:Map_of_Sunda_and_Sahul.png

Figure 5.6.5. Transeurasian from the Perspective of Historical Linguistics.

Redrawn with permission from Eurasia3angle Project, Max Planck Institute for the Science of Human History. http://www.shh.mpg.de/102128/eurasia3angle_group

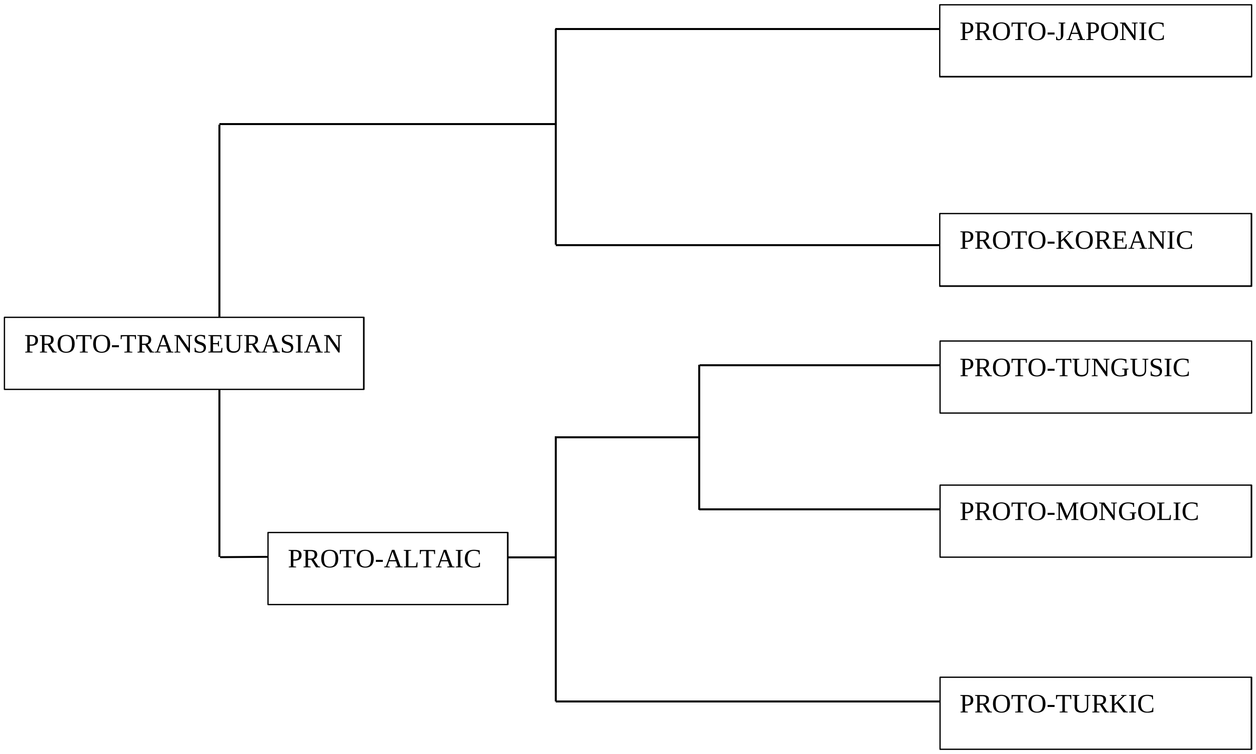
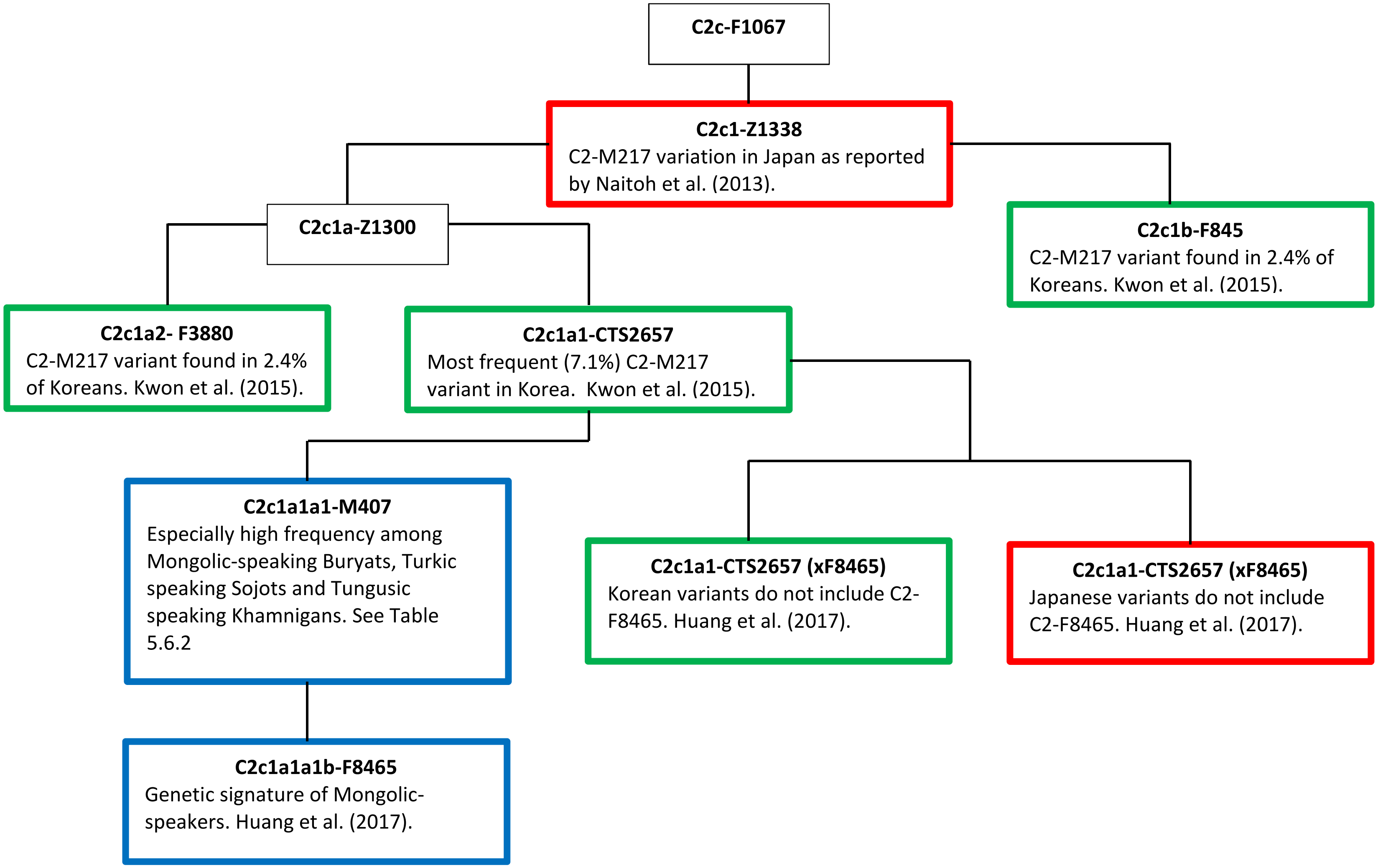


Figure 5.6.6. Overview of known C2-M217 variants among Mongolians, Koreans and Japanese.

Nomenclature generally conforms to ISOGG 2017.

Red highlights informative Japanese mutations, green is for Koreans, and blue is for Mongolians.



C2c-F1067

C2c1-Z1338
C2-M217 variation in Japan as reported by Naitoh et al. (2013).

C2c1a-Z1300

C2c1b-F845
C2-M217 variant found in 2.4% of Koreans. Kwon et al. (2015).

C2c1a2- F3880
C2-M217 variant found in 2.4% of Koreans. Kwon et al. (2015).

C2c1a1-CTS2657
Most frequent (7.1%) C2-M217 variant in Korea. Kwon et al. (2015).

C2c1a1a1-M407
Especially high frequency among Mongolic-speaking Buryats, Turkic speaking Sojots and Tungusic speaking Khamnigans. See Table 5.6.2

C2c1a1-CTS2657 (xF8465)
Korean variants do not include C2-F8465. Huang et al. (2017).

C2c1a1-CTS2657 (xF8465)
Japanese variants do not include C2-F8465. Huang et al. (2017).

C2c1a1a1b-F8465
Genetic signature of Mongolic-speakers. Huang et al. (2017).

Figure 5.6.7. Transeurasian from a Language Contact Model Perspective.

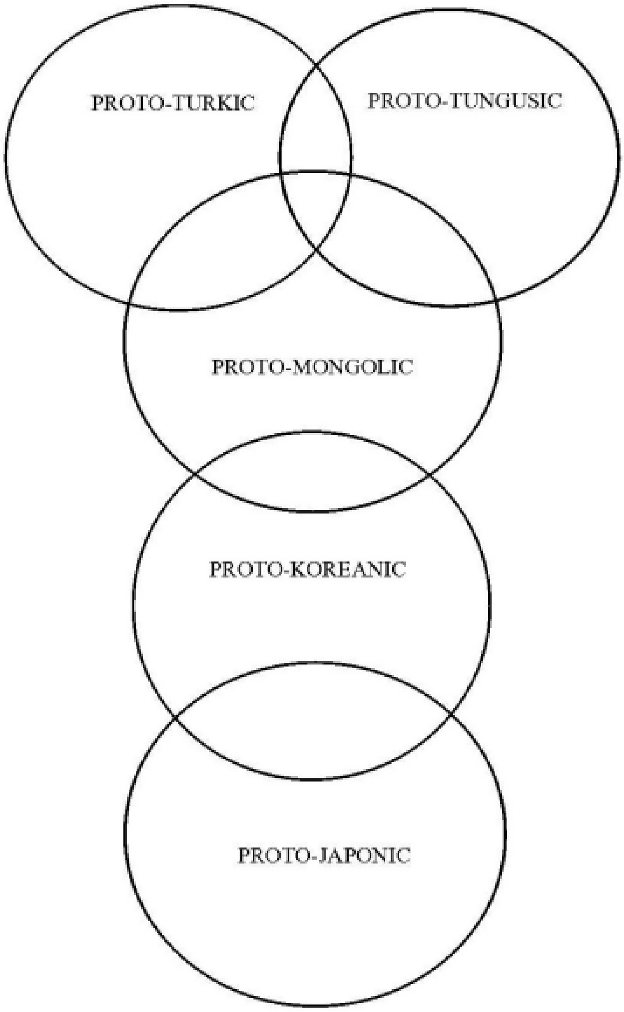


Table 5.6.1. Survey of C2-M217 in Eurasia.

Population	Location	Language Family	Branch	Sample Size	Percent C-M217	Reference
Kazakhs	Kazakhstan	Turkic		28	92.9	[1]
Kazakhs	China, Gansu, Akesai	Turkic		94	84.0	[43]
Kazakhs	China, Xinjiang, Hami, Barkol	Turkic		104	76.9	[43]
Kazakhs	Xingjiang, China	Turkic		53	75.5	[27]
Kazakhs	China, Xinjiang ,Changji, Mori	Turkic		63	74.6	[43]
Kazakhs	China, Xinjiang, Changji, Hutubi	Turkic		22	72.7	[43]
Evenki	Russia, Krasnoyarskiy	Tungusic		40	70.0	[28]
Buryats	Russia, Buryatia	Mongolic		217	68.7	[2,3]
Buryats	Russia, China	Mongolic		86	68.6	[37]
Mongols	China, Inner Mongolian	Mongolic		22	68.2	[27]
Kazakh (Kazaly)	Kazakhstan	Turkic		81	67.9	[45]
Kazakh (Arys)	Kazakhstan	Turkic		119	67.2	[45]
Mongols	Mongolia	Mongolic		46	65.2	[2,3]
Kazakhs	Kazakhstan	Turkic		20	65.0	[1]
Mongols	China, Inner Mongolia, Hailar	Mongolic		61	63.9	[43]
Kalmyks	Russia, Buryatia	Mongolic		91	62.6	[2,3]
Buryats	China, Inner Mongolia, Hailar	Mongolic		26	61.5	[43]
Oroqen	China, Inner Mongolia	Tungusic		31	61.3	[18,19]
Kazakhs	Russia, Altai	Turkic		36	58.3	[2,3]
Khamnigans	Russia, Chitinskaya?	Tungusic		51	54.9	[2,3]

Population	Location	Language Family	Branch	Sample Size	Percent C-M217	Reference
Mongols	Inner Mongolia	Mongolic		22	54.5	[44]
Even	Russia, Sakha	Tungusic		24	54.2	[29]
Koryaks	Russia, Koryanskiy	Chukotko-Kamchatkan		39	54.0	[3]
Kazakhs	Russia, Altai Republic	Turkic		89	54.0	[8]
Mongols	Mongolia, Ulaanbaatar	Mongolic		65	53.8	[18,19]
Evenki	China, Inner Mongolia	Tungusic		26	53.8	[18,19]
Sojots	Russia, Buryatia	Turkic		28	53.6	[2,3]
Mongols	Central Asia	Mongolic		97	52.6	[40]
Evenki	Russia, Krasnoyarskiy	Tungusic		41	48.8	[2]
Mongols	Central Asia	Mongolic		23	47.8	[40]
Mongols	Mongolia	Mongolic		45	46.7	[23]
Mongols	China, Inner Mongolia	Mongolic		45	46.7	[18,19]
Manchu (Man)	Heilongjiang	Tungusic		24	45.8	[44]
Mongols	Central Asia	Mongolic		20	45.0	[40]
Buryats	China, Inner Mongolia	Mongolic		36	44.4	[23]
Kazakh (Shymkent)	Kazakhstan	Turkic		55	43.6	[45]
Manchu (Man)	China, Jilin, Jiutai	Tungusic		146	41.7	[43]
Hazara	Central Asia	Indo-European	Indo-Iranian	25	40.0	[40]
Buryats	Russia, China	Mongolic		58	39.7	[37]
Kazakhs	China, Xinjiang, Yili, Zhaosu	Turkic		29	37.9	[43]
Xibe	Xinjiang	Tungusic		61	36.1	[44]
Even	Russia	Tungusic		63	34.0	[3]
Hazara-Bamiyan	Central Asia	Indo-European	Indo-Iranian	69	33.3	[40]
Hazara	Afghanstan	Indo-European	Indo-Iranian	60	33.3	[30]
Manchu (Man)	China, Liaoning	Tungusic		109	33.0	[27]

Population	Location	Language Family	Branch	Sample Size	Percent C-M217	Reference
Mongols	China, Inner Mongolia	Mongolic		46	32.6	[43]
Han	China, Shangdong, Zoucheng	Sino-Tibetan	Chinese	32	31.3	[43]
Daur	China, Heilongjiang	Mongolic		39	30.8	[18,19]
Tujia	Hubei	Sino-Tibetan	Tibeto-Burman	26	30.8	[44]
Kyrgyz-SouthWest	Central Asia	Turkic		20	30.0	[40]
Han	Jilin	Sino-Tibetan	Chinese	27	29.6	[44]
Dong	Hunan	Tai-Kadai	Kam-Tai	27	29.6	[44]
Karakalpak	Uzbekistan	Turkic		100	28.0	[45]
Kazakh (Zhanakorgan)	Kazakhstan	Turkic		94	27.7	[45]
Kyrgyz-Central	Central Asia	Turkic		40	27.5	[40]
Mongols	China, Qinghai, Delingha	Mongolic		129	27.1	[43]
Tuvans	China, Xinjiang, Altay	Turkic		148	27.0	[43]
Xibe	China, Yili,Xinjiang	Mongolic		41	26.8	[18,19]
Manchu (Man)	China	Tungusic		30	26.7	[23]
Hazara	Pakistan, Southern Baluchistan	Indo-European	Indo-Iranian	27	25.9	[18,22]
Buryats	China, Inner Mongolia, Hailar	Mongolic		54	25.9	[43]
Manchu (Man)	China, Dalian	Tungusic		35	25.7	[18,19]
Manchu (Man)	Liaoning	Tungusic		43	25.6	[44]
Han	China, Shangdong, Rizhao	Sino-Tibetan	Chinese	33	24.2	[43]
Han	Henan	Sino-Tibetan	Chinese	21	23.8	[44]
Han	China, Shaanxi, Xi'an	Sino-Tibetan	Chinese	34	23.8	[23]
Han	China, Shangdong, Qingzhou	Sino-Tibetan	Chinese	22	22.7	[43]

Population	Location	Language Family	Branch	Sample Size	Percent C-M217	Reference
Han	Liaoning	Sino-Tibetan	Chinese	40	22.5	[44]
Dong	Hunan	Tai-Kadai	Kam-Tai	45	22.2	[44]
Nanai (Hezhe)	China, Heilongjiang	Tungusic		45	22.2	[18,19]
Hani	Yunnan	Sino-Tibetan	Tibeto-Burman	60	21.7	[44]
Yugurs	China, Gansu, Sunan	Turkic		141	21.3	[43]
Tujia	Guizhou	Sino-Tibetan	Tibeto-Burman	33	21.2	[44]
Manchu (Man)	Liaoning	Tungusic		66	21.2	[44]
Han	Gansu	Sino-Tibetan	Chinese	29	20.7	[44]
Han	Gansu	Sino-Tibetan	Chinese	34	20.6	[44]
Han	Gansu	Sino-Tibetan	Chinese	34	20.6	[44]
Manchu (Man)	China, Inner Mongolia, Chifeng	Tungusic		39	20.5	[43]
Han	China, Shangdong, Zhanhua	Sino-Tibetan	Chinese	30	20.0	[43]
Altai (Altaians, Altay)	Russia, Altai Republic	Turkic		120	20.0	[16]
Han	China, Ganshu, LanZhou	Sino-Tibetan	Chinese	30	20.0	[18,19]
Han	China, Shangdong, Ningjing	Sino-Tibetan	Chinese	45	20.0	[43]
Han	Shanxi	Sino-Tibetan	Chinese	56	19.6	[44]
Tajik-Badakhshan	Central Asia	Indo-European	Indo-Iranian	37	18.9	[40]
Buryats	Russia, China	Mongolic		138	18.8	[37]
Uyghurs	Kazakstan, Almaty, Lavar	Turkic		33	18.2	[18,20]
Yao	Guangxi	Hmong-Mien	Mienic	55	18.2	[44]
Uzbek (Fergana)	Uzbekistan	Turkic		67	17.9	[45]
Hui	Ningxia	Sino-Tibetan	Chinese	62	17.7	[44]
Hani	China, Yunnan	Sino-Tibetan	Tibeto-Burman	34	17.6	[18,19]

Population	Location	Language Family	Branch	Sample Size	Percent C-M217	Reference
Han	China, Shangdong, Guangrao	Sino-Tibetan	Chinese	23	17.4	[43]
Han	China, Northeast	Sino-Tibetan	Chinese	98	17.3	[43]
Kazakhs	Russia, Altai Republic	Turkic		30	17.0	[8]
Korean	China, Jilin, Dunhua, Jiutai	Koreanic		191	16.8	[43]
Han	China, Shangdong, Yanzhou	Sino-Tibetan	Chinese	24	16.7	[43]
Han	China, Shangdong, Zhangqiu	Sino-Tibetan	Chinese	48	16.7	[43]
Han	Heilongjiang	Sino-Tibetan	Chinese	67	16.4	[44]
Korean	Korea	Koreanic		43	16.3	[19]
Korean	Korea	Koreanic		545	16.3	[51]
Altai (Altaians, Altay)	Russia, Altai	Turkic		89	15.7	[2,3]
Han	China, Gansu, Wuwei	Sino-Tibetan	Chinese	64	15.6	[43]
Han	Shandong	Sino-Tibetan	Chinese	52	15.4	[44]
Han	Jiangsu	Sino-Tibetan	Chinese	39	15.4	[44]
Evenki	Russia, Sakha	Tungusic		33	15.2	[28]
Han	Gansu	Sino-Tibetan	Chinese	20	15.0	[44]
Mongols	China, Inner Mongolia, Chifeng	Mongolic		67	14.9	[43]
Han	China, Heilongjiang, Wuchang	Sino-Tibetan	Chinese	27	14.8	[43]
Tuvans	Russia, Tuva	Turkic		55	14.5	[28]
Han	China, HeiLongJiang	Sino-Tibetan	Chinese	35	14.3	[18,19]
Uzbeks-Sar-e-Pol	Central Asia	Turkic		28	14.3	[40]
Kyrgyz-East	Central Asia	Turkic		35	14.3	[40]
Han	Heilongjiang	Sino-Tibetan	Chinese	57	14.0	[44]

Population	Location	Language Family	Branch	Sample Size	Percent C-M217	Reference
Kyrgyz	China, Xinjiang	Turkic		50	14.0	[43]
Kazakhs	China, Xinjiang, Yili, Xinyuan	Turkic		43	14.0	[43]
Han	China, Gansu, Pingliang	Sino-Tibetan	Chinese	203	13.8	[43]
Sui	Guizhou	Tai-Kadai	Kam-Tai	29	13.8	[44]
Han	China, Shangdong, Dezhou	Sino-Tibetan	Chinese	51	13.7	[43]
Han	Anhui	Sino-Tibetan	Chinese	52	13.5	[44]
Han	China, Gansu, Lanzhou	Sino-Tibetan	Chinese	83	13.3	[43]
Monguor (Tu)	China, Qinghai, Huzhu	Mongolic		121	13.2	[43]
Han	China, Gansu, Qingyang	Sino-Tibetan	Chinese	61	13.1	[43]
Hui	China, Tianjin	Turkic		46	13.0	[43]
Evenki	Inner Mongolia	Tungusic		31	12.9	[44]
Han	Guangdong	Sino-Tibetan	Chinese	31	12.9	[44]
Han	China, Gansu, Dingxi	Sino-Tibetan	Chinese	166	12.7	[43]
Vietnamese	Vietnam	Austro-Asiatic	Mon-Khmer	48	12.5	[23]
Yi (Yizu)	Sichuan	Sino-Tibetan	Tibeto-Burman	24	12.5	[44]
Han	Jilin	Sino-Tibetan	Chinese	24	12.5	[44]
Korean	Korea	Koreanic		506	12.3	[23]
Han	China, Shangdong, Qihe	Sino-Tibetan	Chinese	25	12.0	[43]
Korean	China, Jinlin	Koreanic		25	12.0	[19]
Han	China, Shangdong, lijing	Sino-Tibetan	Chinese	59	11.9	[43]
Han	China, SiChuan, ChengDu	Sino-Tibetan	Chinese	34	11.8	[18,19]
Han	China, North	Sino-Tibetan	Chinese	223	11.7	[43]
Hui	China, Haiyuan,Ningxia	Turkic		35	11.4	[18,19]
Teleuts	Russia, Altayskiy, Biysk	Turkic		44	11.4	[2,3]

Population	Location	Language Family	Branch	Sample Size	Percent C-M217	Reference
Han	China, Shangdong, linshu	Sino-Tibetan	Chinese	53	11.3	[43]
Tuvans	Russia, Tuva	Turkic		108	11.1	[2,3]
Kyrgyz-NorthWest	Central Asia	Turkic		37	10.8	[40]
Han	Shannxi	Sino-Tibetan	Chinese	56	10.7	[44]
Han	China, Shangdong, Shanghe	Sino-Tibetan	Chinese	48	10.4	[43]
Han	Gansu	Sino-Tibetan	Chinese	39	10.3	[44]
Uzbek (Xorezm)	Uzbekistan	Turkic		98	10.2	[45]
Han	Shandong	Sino-Tibetan	Chinese	40	10.0	[44]

Table 5.6.2. Survey of C2-M407 Populations.

Population	Location	Language Family	Sample Size	Percent C-M407	Reference
Buryats	Russia, China	Mongolic	86	61.6	[37]
Buryats	Russia, Buryatia	Mongolic	217	53.9	[2,3]
Sojots	Russia, Buryatia	Turkic	28	53.6	[2,3]
Khamnigans	Russia, Chitinskaya?	Tungusic	51	52.9	[2,3]
Buryats	Russia, China	Mongolic	58	19.0	[37]
Mongols	Mongolia	Mongolic	46	15.2	[2,3]
Buryats	Russia, China	Mongolic	138	13.8	[37]
Kalmyks	Russia, Buryatia	Mongolic	91	12.1	[2,3]
Mongols	Central Asia	Mongolic	20	10.0	[40]

Table 5.6.3. Survey of C2-M48 Populations.

Population	Location	Language Family	Sample Size	Percent C-M48	Reference
Evenki	Russia, Krasnoyarskiy	Tungusic	40	70.0	[28]
Kazakhs	Kazakstan, Almaty, Katon-Karagay	Turkic	38	63.2	[18,20]
Even	Russia, Sakha	Tungusic	24	50.0	[29]
Kalmyks	Russia, Buryatia	Mongolic	91	45.1	[2,3]
Evenki	Russia, Krasnoyarskiy	Tungusic	41	43.9	[2]
Oroqen	China, Inner Mongolia	Tungusic	31	41.9	[18,19]
Kazakhs	Russia, Altai	Turkic	36	41.7	[2,3]
Manchu	China, Jilin, Jiutai	Tungusic	146	41.1	[43]
Kazakhs	Russia, Altai Republic	Turkic	89	40.0	[8]
Kalmyks	Elista, Russia	Mongolic	99	37.4	[5,6]
Even	Russia	Tungusic	63	33.0	[3]
Mongols	Central Asia	Mongolic	97	29.9	[40]
Evenki	China, Inner Mongolia	Tungusic	26	26.9	[18,19]
Tuvans	China, Xinjiang, Altay	Turkic	148	22.3	[43]
Mongols	Mongolia, Ulaanbaatar	Mongolic	65	20.0	[18,19]
Mongols	China, Inner Mongolian	Mongolic	22	18.2	[27]
Evenki	Russia, Sakha	Tungusic	33	15.2	[28]
Mongols	Central Asia	Mongolic	20	15.0	[40]
Mongols	China, Qinghai, Delingha	Mongolic	129	13.2	[43]
Nanai (Hezhe)	China, Heilongjiang	Tungusic	45	11.1	[18,19]
Mongols	Mongolia	Mongolic	46	10.9	[2,3]

Table 5.6.4. Survey of C2-F1918 Populations.

Population	Location	Language Family	Branch	Sample Size	Percent C-F1918	Reference
Kazakhs	Kazakhstan	Turkic		28	89.3	[1]
Kazakhs	China, Gansu, Akesai	Turkic		94	78.7	[43]
Kazakhs	Xingjiang, China	Turkic		53	75.5	[27]
Kazakhs	China, Xinjiang ,Changji, Mori	Turkic		63	58.7	[43]
Kazakhs	China, Xinjiang, Hami, Barkol	Turkic		104	58.7	[43]
Kazakhs	Kazakhstan	Turkic		20	55.0	[1]
Buryats	China, Inner Mongolia, Hailar	Mongolic		26	46.2	[43]
Kazakhs	China, Xinjiang, Changji, Hutubi	Turkic		22	45.5	[43]
Kazakhs	China, Xinjiang, Yili, Zhaosu	Turkic		29	37.9	[43]
Kazakhs	Kazakhstan, Taraz	Turkic		181	36.5	[11]
Hazara	Central Asia	Indo-European	Indo-Iranian	25	36.0	[40]
Mongols	Mongolia	Mongolic		46	34.8	[2,3]
Mongols	China, Inner Mongolia	Mongolic		45	31.1	[18,19]
Kazakhs	China, Xinjiang, Altay	Turkic		97	30.9	[36]
Mongols	Central Asia	Mongolic		23	30.4	[40]
Hazara-Bamiyan	Central Asia	Indo-European	Indo-Iranian	69	26.1	[40]
Hazara	Pakistan, Southern Baluchistan	Indo-European	Indo-Iranian	27	25.9	[18,22]
Kyrgyz-SouthWest	Central Asia	Turkic		20	25.0	[40]
Mongols	Mongolia	Mongolic		45	24.4	[23]
Kazakhs	Kazakhstan	Turkic		99	24.2	[10]
Han	China, Inner Mongolia	Sino-Tibetan	Chinese	21	23.8	[18,21]
Karakalpaks	Uzbekistan	Turkic		51	23.5	[7]
Hazara	Afghanstan	Indo-European	Indo-Iranian	60	23.3	[30]

Population	Location	Language Family	Branch	Sample Size	Percent C-F1918	Reference
Daur	China, Heilongjiang	Mongolic		39	23.1	[18,19]
Buryats	China, Inner Mongolia	Mongolic		36	19.4	[23]
Evenki	China, Inner Mongolia	Tungusic		26	19.2	[18,19]
Mongols	China, Inner Mongolian	Mongolic		22	18.2	[27]
Mongols	Mongolia, Ulaanbaatar	Mongolic		65	16.9	[18,19]
Mongols	Northern China	Mongolic		50	16.0	[38]
Uyghurs	Kazakstan, Almaty, Lavar	Turkic		33	15.2	[18,20]
Mongols	Central Asia	Mongolic		20	15.0	[40]
Mongols	China, Inner Mongolia, Hailar	Mongolic		61	14.8	[43]
Nogais	Russia	Turkic		29	13.8	[4]
Daur	Northern China	Mongolic		30	13.3	[38]
Buryats	Russia, China	Mongolic		58	12.1	[37]
Kazakhs	Kazakhstan	Turkic		100	12.0	[9]
Mongols	Central Asia	Mongolic		97	11.3	[40]
Kazakhs	Russia, Altai Republic	Turkic		89	11.0	[8]
Xibe	Northern China	Tungusic		28	10.7	[38]
Uzbeks-Sar-e-Pol	Central Asia	Turkic		28	10.7	[40]
Kyrgyz-Central	Central Asia	Turkic		40	10.0	[40]
Kazakhs	Russia, Altai Republic	Turkic		30	10.0	[8]

Table 5.6.5. Survey of Turkic-Speaking Populations in the Integrated Linguistics Database.

Population	Language Family	Language	Region
Ainu	Turkic	Ainu	East Asia
Altai (Altaians, Altay)	Turkic	Altai (2)	Northern Eurasia
Azerbaijani	Turkic	Azerbaijani (2)	Caucasus
Azerbaijani	Turkic	Azerbaijani (2)	Middle East
Balkars	Turkic	Karachay-Balkar	Caucasus
Bashkir	Turkic	Bashkort	Eastern Europe
Chuvash	Turkic	Chuvash	Eastern Europe
Dolgan	Turkic	Dolgan	Northern Eurasia
Gagauz	Turkic	Gagauz	Eastern Europe
Karachays	Turkic	Karachay-Balkar	Caucasus
Karakalpaks	Turkic	Karakalpak	Central Asia
Kazakhs	Turkic	Kazakh	Central Asia
Kazakhs	Turkic	Kazakh	East Asia
Kazakhs	Turkic	Kazakh	Northern Eurasia
Khakas	Turkic	Khakas	Northern Eurasia
Kumyks	Turkic	Kumyk	Caucasus
Kyrgyz	Turkic	Kyrgyz	Central Asia
Kyrgyz	Turkic	Kyrgyz	East Asia
Nogais	Turkic	Nogai	Caucasus
Salar	Turkic	Salar	East Asia
Shors	Turkic	Shor	Northern Eurasia
Tatars	Turkic	Tatar	Central Asia
Tatars	Turkic	Tatar	East Asia
Tatars	Turkic	Tatar	Eastern Europe
Teleuts	Turkic	Altai (2)	Northern Eurasia
Todjins	Turkic	Tuva	Northern Eurasia
Tofalars	Turkic	Karagas	Northern Eurasia
Tubalar	Turkic	Altai (2)	Northern Eurasia
Turkmen	Turkic	Turkmen	Central Asia
Turkmen	Turkic	Turkmen	Middle East
Turks	Turkic	Turkish	Middle East
Tuvans	Turkic	Tuva	Northern Eurasia
Uyghurs	Turkic	Uyghur	Central Asia
Uyghurs	Turkic	Uyghur	East Asia

Population	Language Family	Language	Region
Uzbeks	Turkic	Uzbek (2)	Central Asia
Uzbeks	Turkic	Uzbek (2)	East Asia
Yakut	Turkic	Yakut	Northern Eurasia

Table 5.6.6. C2-M217 and Turkic Languages.

Population	Location	Sample Size	Percent C-M217	Reference
Kazakhs	Kazakhstan	28	92.9	[1]
Kazakhs	China, Gansu, Akesai	94	84.0	[43]
Kazakhs	China, Xinjiang, Hami, Barkol	104	76.9	[43]
Kazakhs	Xingjiang, China	53	75.5	[27]
Kazakhs	China, Xinjiang ,Changji, Mori	63	74.6	[43]
Kazakhs	China, Xinjiang, Changji, Hutubi	22	72.7	[43]
Kazakh (Kazaly)	Kazakhstan	81	67.9	[45]
Kazakh (Arys)	Kazakhstan	119	67.2	[45]
Kazakhs	Kazakhstan	20	65.0	[1]
Kazakhs	Russia, Altai	36	58.3	[2,3]
Kazakhs	Russia, Altai Republic	89	54.0	[8]
Sojots	Russia, Buryatia	28	53.6	[2,3]
Kazakh (Shymkent)	Kazakhstan	55	43.6	[45]
Kazakhs	China, Xinjiang, Yili, Zhaosu	29	37.9	[43]
Kyrgyz-SouthWest	Central Asia	20	30.0	[40]
Karakalpak	Uzbekistan	100	28.0	[45]
Kazakh (Zhanakorgan)	Kazakhstan	94	27.7	[45]
Kyrgyz-Central	Central Asia	40	27.5	[40]
Tuvans	China, Xinjiang, Altay	148	27.0	[43]
Yugurs	China, Gansu, Sunan	141	21.3	[43]
Altai (Altaians, Altay)	Russia, Altai Republic	120	20.0	[16]
Uyghurs	Kazakstan, Almaty, Lavar	33	18.2	[18,20]
Uzbek (Fergana)	Uzbekistan	67	17.9	[45]
Kazakhs	Russia, Altai Republic	30	17.0	[8]
Altai (Altaians, Altay)	Russia, Altai	89	15.7	[2,3]
Tuvans	Russia, Tuva	55	14.5	[28]
Kyrgyz-East	Central Asia	35	14.3	[40]
Uzbeks-Sar-e-Pol	Central Asia	28	14.3	[40]
Kyrgyz	China, Xinjiang	50	14.0	[43]
Kazakhs	China, Xinjiang, Yili, Xinyuan	43	14.0	[43]
Hui	China, Tianjin	46	13.0	[43]
Teleuts	Russia, Altayskiy, Biysk	44	11.4	[2,3]
Hui	China, Haiyuan,Ningxia	35	11.4	[18,19]
Tuvans	Russia, Tuva	108	11.1	[2,3]
Kyrgyz-NorthWest	Central Asia	37	10.8	[40]

Population	Location	Sample Size	Percent C-M217	Reference
Uzbek (Xorezm)	Uzbekistan	98	10.2	[45]

Table 5.6.7. C2-M217 and Mongolic Languages.

Population	Location	Sample Size	Percent C-M217	Reference
Buryats	Russia, Buryatia	217	68.7	[2,3]
Buryats	Russia, China	86	68.6	[37]
Mongols	China, Inner Mongolian	22	68.2	[27]
Mongols	Mongolia	46	65.2	[2,3]
Mongols	China, Inner Mongolia, Hailar	61	63.9	[43]
Kalmyks	Russia, Buryatia	91	62.6	[2,3]
Buryats	China, Inner Mongolia, Hailar	26	61.5	[43]
Mongols	Inner Mongolia	22	54.5	[44]
Mongols	Mongolia, Ulaanbaatar	65	53.8	[18,19]
Mongols	Central Asia	97	52.6	[40]
Mongols	Central Asia	23	47.8	[40]
Mongols	China, Inner Mongolia	45	46.7	[18,19]
Mongols	Mongolia	45	46.7	[23]
Mongols	Central Asia	20	45.0	[40]
Buryats	China, Inner Mongolia	36	44.4	[23]
Buryats	Russia, China	58	39.7	[37]
Mongols	China, Inner Mongolia	46	32.6	[43]
Daur	China, Heilongjiang	39	30.8	[18,19]
Mongols	China, Qinghai, Delingha	129	27.1	[43]
Xibe	China, Yili,Xinjiang	41	26.8	[18,19]
Buryats	China, Inner Mongolia, Hailar	54	25.9	[43]
Buryats	Russia, China	138	18.8	[37]
Mongols	China, Inner Mongolia, Chifeng	67	14.9	[43]
Monguor (Tu)	China, Qinghai, Huzhu	121	13.2	[43]

Table 5.6.8. C2-M217 and Tungusic Languages.

Population	Location	Sample Size	Percent C-M217	Reference
Evenki	Russia, Krasnoyarskiy	40	70.0	[28]
Oroqen	China, Inner Mongolia	31	61.3	[18,19]
Khamnigans	Russia, Chitinskaya?	51	54.9	[2,3]
Even	Russia, Sakha	24	54.2	[29]
Evenki	China, Inner Mongolia	26	53.8	[18,19]
Evenki	Russia, Krasnoyarskiy	41	48.8	[2]
Manchu (Man)	Heilongjiang	24	45.8	[44]
Manchu (Man)	China, Jilin, Jiutai	146	41.7	[43]
Xibe	Xinjiang	61	36.1	[44]
Even	Russia	63	34.0	[3]
Manchu (Man)	China, Liaoning	109	33.0	[27]
Manchu (Man)	China	30	26.7	[23]
Manchu (Man)	China, Dalian	35	25.7	[18,19]
Manchu (Man)	Liaoning	43	25.6	[44]
Nanai (Hezhe)	China, Heilongjiang	45	22.2	[18,19]
Manchu (Man)	Liaoning	66	21.2	[44]
Manchu (Man)	China, Inner Mongolia, Chifeng	39	20.5	[43]
Evenki	Russia, Sakha	33	15.2	[28]
Evenki	Inner Mongolia	31	12.9	[44]

Table 5.6.9. C2-M217 and Koreanic.

Population	Location	Sample Size	Percent C-M217	Reference
Korean	China, Jilin, Dunhua, Jiutai	191	16.8	[43]
Korean	Korea	545	16.3	[51]
Korean	Korea	43	16.3	[19]
Korean	Korea	506	12.3	[23]
Korean	China, Jinlin	25	12.0	[19]
Korean	Korea	52	9.6	[3]
Korean	Jilin	34	5.9	[44]

Table 5.6.10. C2-M217 and Japonic.

Population	Location	Sample Size	Percent C-M217	Reference
Japanese	Japan	157	7.0	[23]
Japanese	Japan	2390	6.1	[53]
Japanese	Japan	263	3.0	[52]
Japanese	Japan	47	2.1	[19]

Table 5.6.11. C2-M217 and Han.

Population	Location	Sample Size	Percent C-M217	Reference
Han	China, Shangdong, Zoucheng	32	31.3	[43]
Han	Jilin	27	29.6	[44]
Han	China, Shangdong, Rizhao	33	24.2	[43]
Han	Henan	21	23.8	[44]
Han	China, Shaanxi, Xi'an	34	23.8	[23]
Han	China, Shangdong, Qingzhou	22	22.7	[43]
Han	Liaoning	40	22.5	[44]
Han	Gansu	29	20.7	[44]
Han	Gansu	34	20.6	[44]
Han	Gansu	34	20.6	[44]
Han	China, Shangdong, Zhanhua	30	20.0	[43]
Han	China, Ganshu, LanZhou	30	20.0	[18,19]
Han	China, Shangdong, Ningjing	45	20.0	[43]
Han	Shanxi	56	19.6	[44]
Han	China, Shangdong, Guangrao	23	17.4	[43]
Han	China, Northeast	98	17.3	[43]
Han	China, Shangdong, Zhangqiu	48	16.7	[43]
Han	China, Shangdong, Yanzhou	24	16.7	[43]
Han	Heilongjiang	67	16.4	[44]
Han	China, Gansu, Wuwei	64	15.6	[43]
Han	Shandong	52	15.4	[44]
Han	Jiangsu	39	15.4	[44]
Han	Gansu	20	15.0	[44]
Han	China, Heilongjiang, Wuchang	27	14.8	[43]
Han	China, HeiLongJiang	35	14.3	[18,19]
Han	Heilongjiang	57	14.0	[44]
Han	China, Gansu, Pingliang	203	13.8	[43]
Han	China, Shangdong, Dezhou	51	13.7	[43]
Han	Anhui	52	13.5	[44]
Han	China, Gansu, Lanzhou	83	13.3	[43]
Han	China, Gansu, Qingyang	61	13.1	[43]
Han	Guangdong	31	12.9	[44]
Han	China, Gansu, Dingxi	166	12.7	[43]
Han	Jilin	24	12.5	[44]
Han	China, Shangdong, Qihe	25	12.0	[43]

Population	Location	Sample Size	Percent C-M217	Reference
Han	China, Shangdong, lijing	59	11.9	[43]
Han	China, SiChuan, ChengDu	34	11.8	[18,19]
Han	China, North	223	11.7	[43]
Han	China, Shangdong, linshu	53	11.3	[43]
Han	Shannxi	56	10.7	[44]
Han	China, Shangdong, Shanghe	48	10.4	[43]
Han	Gansu	39	10.3	[44]
Han	Shandong	40	10.0	[44]

Table 5.6.12. Native Americans.

Population	Language Family	Sample Size	Percent C2-P39	Reference
Apache	Eyak-Athabaskan	94	14.0	[47]
Dogrib	Eyak-Athabaskan	37	13.0	[51]
Gwich'in	Eyak-Athabaskan	33	8.0	[51]
SWS Sioux	Siouan-Catawban	26	8.0	[50]
Cheyenne/Arapahoe	Algic	50	8.0	[50]
Dogrib	Eyak-Athabaskan	15	5.0	[46]
Tanana	Eyak-Athabaskan	11	5.0	[47]
Chipewyan	Eyak-Athabaskan	48	3.0	[48]
SC Apache	Eyak-Athabaskan	23	2.0	[46]
Creek	Muskogean	12	1.0	[50]
Eskimos	Eskimo-Aleut	56	1.0	[51]
Navajo	Eyak-Athabaskan	75	1.0	[47]
Papago	Uto-Aztecan	13	0.0	[46]
Zapotec	Otomanguean	16	0.0	[47]
Pima	Uto-Aztecan	62	0.0	[46] [47]
Nahua Atocpan	Uto-Aztecan	7	0.0	[46]
Jemez	Kiowa-Tanoan	13	0.0	[46]
Tarahumara	Uto-Aztecan	20	0.0	[46]
Seri	Isolate	15	0.0	[46]
Nahua Cuetzalan	Uto-Aztecan	10	0.0	[46]
Cora	Uto-Aztecan	37	0.0	[46]
Huichol	Uto-Aztecan	18	0.0	[46]
Mixe	Mixe-Zoquean	12	0.0	[47]
Seminole	Muskogean	20	0.0	[49]
TM Chippewa	Eyak-Athabaskan	34	0.0	[50]
W Chippewa	Eyak-Athabaskan	29	0.0	[50]
Stillwell Cherokee	Iroquoian	30	0.0	[50]
Choctaw	Muskogean	12	0.0	[50]
Mixtec	Otomanguean	22	0.0	[46]

Bibliography for Tables

[1] Abilev, Serikbai et al. 2012. “The Y-Chromosome C3 star-cluster attributed to Genghis Khan’s descendants is present at high frequency in the Kerey clan from Kazakhstan.” *Human Biology* 84(1): 79-89.

[2] Derenko, M.V. et al. 2007a. “Distribution of the male lineages of Genghis Khan’s descendants in Northern Eurasian Populations.” *Russian Journal of Genetics* 43(3): 334-337.

[3] Malyarchuk, Boris et al. 2010. “Phylogeography of the Y-chromosome haplogroup C in northern Eurasia.” *Annals of Human Genetics* 74: 539-546.

[4] Zakharov, I. A. 2010. “A search for a Genghis Khan chromosome.” *Russian Journal of Genetics* 46(9):1130-1131.

[5] Roewer, Lutz et al. 2007. “Y-chromosomal STR haplotypes in Kalmyk population samples.” *Forensic Science International* 173: 204-209.

[6] Nasidze, Ivan et al. 2005. “Genetic evidence for the Mongolian ancestry of Kalmyks.” *American Journal of Physical Anthropology* 128: 846-854.

[7] Zhaksylyk Sabitov and Batyr Daulet 2012. “Haplogroups and haplotypes of 50 Karakalpak.” *Russian Journal of Genetic Genealogy* 4(1): 83-90.

[8] Dulik, Matthew et al. 2011. “Y-chromosome variation in Altaian Kazakhs reveals a common parental gene pool for Kazakhs and the influence of Mongolian expansions.” *Public Library of Science One* 6(3): e17548.

[9] Turuspekov, Yerlan et al. 2011. “The Kazakhstan DNA project hits first hundred Y-profiles for ethnic Kazakhs.” *Russian Journal of Genetic Genealogy* 2(1):69-84.

[10] Tarlykov, Pavel V. et al. 2013. “Mitochondrial and Y-chromosomal profile of the Kazakh population from East Kazakhstan.” *Croatian Medical Journal* 54(1): 17-24.

[11] Roewer, Lutz et al. 2007. “Y-chromosomal STR haplotypes in Kalmyk population samples.” *Forensic Science International* 173: 204-209.

[12] Kharkov, V.N et al. 2011. “Genetic diversity of the Khakass gene pool: subethnic differentiation and the structure of Y-chromosome haplogroups.” *Molecular Biology* 45(3): 404-416.

[13] Kharkov, V. N. et al. 2009. “Comparative characteristics of the gene pool of Teleuts inferred from Y-chromosomal marker data.” *Russian Journal of Genetics* 45(8) 994-1003.

[14] Bíró, A.Z. et al. 2009. “A Y-chromosomal comparison of the Madjars (Kazakhstan) and the Magyars (Hungary).” *American Journal of Physical Anthropology* 139: 305-310.

[15] Zhou, Ruixia et al. 2008. “Origins and evolution of two Yugur subclans in Northwest China: a case study in paternal genetic landscape.” *Annals of Human Biology* 35(2): 198-211.

[16] Dulik, Matthew C. et al. 2012. “Mitochondrial DNA and Y chromosome variation provide evidence for a recent common ancestry between Native Americans and indigenous Altaians.” *American Journal of Human Genetics* 90: 229-246.

[17] Fechner, Angela et al. 2008. “Boundaries and clines in the West Eurasian Y-chromosome landscape: insights from the European part of Russia.” *American Journal of Physical Anthropology* 137: 41-47.

[18] Zerjal, Tatiana et al. 2003. “The genetic legacy of the Mongols.” *American Journal of Human Genetics* 72: 717-721.

[19] Xue, Yali et al. 2006. “Male demography in East Asia: a north-south contrast in human population expansion times.” *Genetics* 172: 2413-2439.

[20] Zerjal, Tatiana et al. 2002. “A genetic landscape reshaped by recent events: Y-chromosomal insights into Central Asia.” *American Journal of Human Genetics* 71: 466-482.

[21] Xue, Yali 2001. *Research on Diversity of East Asia Y-Chromosome* [D].Harbin Medical University.

[22] Qamar, Raheel et al. 2002. “Y-chromosomal DNA variation in Pakistan.” *American Journal of Human Genetics* 70: 1107-1124.

[23] Kim, Soon-Hee et al. 2011. “High frequencies of Y-chromosome haplogroup O2b-SRY465 lineages in Korea: a genetic perspective on the peopling of Korea.” *Investigative Genetics* 2:10.

[24] Kwak, KD et al. 2006. “Y-chromosome STR haplotype profiling in the Mongolian population.” *Legal Medicine (Tokyo)* 8(1): 58-61.

[25] Zhu, B. et al. 2005. “Y-STRs haplotypes of Chinese Mongol ethnic group using Y-PLEX.” *Forensic Science International* 153(2-3): 260-263.

[26] Woźniak, Marcin et al. “Allelic and haplotypic frequencies at 11 Y-STR loci in Buryats.” *Forensic Science International* 164(2-3): 271-275.

[27] Zhong, Hua et al. 2010. “Global distribution of Y-chromosome haplogroup C reveals the prehistoric migration routes of African exodus and early settlement in East Asia.” *Journal of Human Genetics* 55: 428-435.

[28] Pakendorf, Brigitte et al. 2006. “Investigating the effects of prehistoric migrations in Siberia: genetic variation and the origins of the Yakuts.” *Human Genetics* 120: 334-353.

[29] Pakendorf, Brigitte et al. 2007. “Mating patterns amongst Siberian Reindeer herders: inferences from mtDNA and Y-chromosomal analyses.” *American Journal of Physical Anthropology* 133: 1013-1027.

[30] Haber, Marc et al. 2012. “Afghanistan’s ethnic groups share a Y-chromosomal heritage structured by historical events.” *Public Library of Science One* 7(3): e34288.

[31] Lobov, A. S. 2009. *Structure of the gene pool of the Bashkir subpopulations* . The dissertation on competition of a scientific degree of candidate of biological Sciences." Ufa, 2009.

[32] Kang, Longli et al. 2012. "Y-chromosome O3 haplogroup diversity in Sino-Tibetan populations reveals two migration routes into the Eastern Himalayas." *Annals of Human Genetics*, 76: 92-99.

[33] Cai, Xiaoyun et al. 2011. "Human migration through bottlenecks from Southeast Asia into East Asia during the last glacial maximum revealed by Y-chromosomes." *Public Library of Science One* 6(8)" e24282.

[34] Pimenoff, Ville N. 2008. "Northwest Siberian Khanty and Mansi in the junction of West and East Eurasian gene pools as revealed by uniparental markers." *European Journal of Human Genetics* 16: 1254-1264.

[35] Sengupta, Sanghamitra et al. 2006. "Polarity and temporality of high-resolution Y-chromosome distributions in India identify both indigenous and exogenous expansions and reveals minor genetic influence of central Asian pastoralists." *American Journal of Human Genetics* 78: 201-221.

[36] Niyazibiligai,Tuerxun 2011. *Polymorphisms of Y-STRs in Yugur and Kazak Ethinic in Xinjiang* [D]. Xinjiang Medical University.

[37] Kharkov, V. N. et al. 2014. "Gene pool of Buryats: clinal variability and territorial subdivision based on data of Y-chromosome markers." *Russian Journal of Genetics* 50(2): 180-190.

[38] Zheng, Lihong et al. 2009. "Y Chromosomal STR Polymorphism in Northern Chinese Populations." *Biological Research* 42: 497-504.

[39] Lippold, Sebastian et al. 2014. "Human paternal and maternal demographic histories: insights from high-resolution Y chromosome and mtDNA sequences." *Investigative Genetics* 5:13.

[40] Di Cristofaro, Julie et al. 2013. "Afghan Hindu Kush: where Eurasian Sub-Continent gene flows converge." *Public Library of Science One* 8(10): e76748.

[41] He, J. and Guo, F. 2013. "Population genetics of 17 Y-STR loci in Chinese Manchu population from Liaoning province, Northeast China." *Forensic Science International-Genetics* 7: E84–E85.

[42] Wei, Lan-Hai et al. 2017. "Genetic trail for the early migrations of Aisin Gioro, the imperial house of the Qing dynasty." *Journal of Human Genetics* 62(3):407-411.

[43] Wei, Lan-Hai et al. 2017. "Phylogeny of Y-chromosome haplogroup C3b-F1756, an important paternal lineage in Altaic-speaking populations." *Journal of Human Genetics* (2017) 62, 915-918.

[44] Zhong, Hua et al. 2011. "Extended Y-chromosome investigations suggests postglacial migrations of modern humans into East Asia via the northern route." *Molecular Biology and Evolution* 28(1): 717-727.

[45] Zhabagin, Maxat et al. 2017. "The connection of the genetic, cultural and geographic landscapes of Transoxiana." *Scientific Reports* 7: 3085

[46] Malhi, Ripan Singh et al. 2008. "Distribution of Y chromosomes among native north Americans: a study of Athabaskan population history." *American Journal of Physical Anthropology* 137: 412-424.

[47] Zegura, Stephen L. et al. 2004. "High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y chromosomes into the Americas." *Molecular Biology and Evolution* 21(1): 164-175.

[48] Bortolini, Maria-Catira et al. 2003. "Y-chromosome evidence for differing ancient demographic histories in the Americas." *American Journal of Human Genetics* 73: 524-539.

[49] Lell, Jeffrey T. et al. 2002a. "The dual origin and Siberian affinities of Native American Y chromosomes." *American Journal of Human Genetics* 70: 192-206.

[50] Bolnick, Deborah A. et al. 2006. "Asymmetric male and female genetic histories among Native Americans from Eastern North America." *Molecular Biology and Evolution* 23(11): 2161-2174.

[51] Kwon, So Yeun et al. 2015. "Confirmation of Y haplogroup tree topologies with newly suggested Y-SNPs for the C2, O2b and O3a subhaplogroups." *Forensic Science International Genetics* 19: 42-46.

[52] Naitoh, Sae et al. 2013. "Assignment of Y-chromosomal SNPs found in Japanese population to Y-chromosomal haplogroup tree." *Journal of Human Genetics* 58: 195-201.

[53] Sato, Youichi et al. 2014. "Overview of genetic variation in the Y chromosome of modern Japanese males." *Anthropological Science* 122(3): 131–136.