The Silk Road: language and population admixture and replacement

Dan Xu, Shaoqing Wen

The Silk Road: language and population admixture and replacement

Dan XU and Shaoqing WEN

1. Introduction

This paper investigates the (non-)correlation between languages and genes on the Silk Road. It is expected to provide a clear picture of human migration history between Central Asia and East Asia. It aims to study models of language formation and replacement on the Silk Road after the 13th century AD. Combining natural sciences and human sciences enables us to overcome the limits of traditional methodology in historical linguistics. In light of biological and archeological research results, where dating is based on concrete materials, our investigation will trace language ancestry on the Silk Road over the same timescales. Diachronic linguistics needs new approaches to go beyond its chronological limits.

The Silk Road is inhabited by different ethnic groups: Chinese (Han), Hui (Muslims), Santa (Dongxiang), Bao'an (Baonan), Monguor (Tu), Eastern Yugur (Yellow Uighur), Western Yugur, Salar, Amdo Tibetan Kazakh and Uyghur. This paper will take eight representative languages (Eastern Yugur, Tu, Dongxiang and Bao'an for Mongolic language group, and Uyghur, Kazakh, Western Yugur and Salar for Turkic group) in studying language admixture and replacement with a biological background. It will show that some languages are the result of admixture over several centuries, and others are newborn languages, like their populations: their languages were completely replaced due to important historical migrations of their ancestors.

The field of linguistics was established in the 19th century. It is taken for granted that comparative studies of Indo-European languages are more advanced and mature thanks to tireless work by linguists. Their comparative method has been introduced to other language families.[1] If we take advantage of research results from other disciplines, we will have a wide-angle perspective from which we can observe many facets of a language. Molecular anthropology has seen significant development since the 1980s and provides linguistics and other domains with a powerful tool. Precise genetic dating will help us to better classify languages and language groups/families. Languages are carried by populations which are bearers of different genes. Comparing language distance based on ethnic genetic distance will help us to reconstruct different historical periods during which dispersing populations, bearers of diverse genes, also spread their languages. We must attempt to exploit new methods to improve our research and break down linguistic limits.

The paper will first present the distribution of languages and Y chromosomes in Northwest China, then analyze their (non)correlation. Two models of language formation/replacement will be studied based on these concrete data: 1) The Elite Dominance model, initially put forth by Renfrew (1987), will be revisited and refined; 2) The Culture Dominance model, proposed in this article, will complete the first model. Finally, preliminary conclusions will be given.

2. The distribution of languages and Y-chromosomes in Northwest China
2.1. Language situation at the basic word level

---

[1] But even today, there is not a strong consensus among linguists on the status of some language families (for example the Altaic family has been debated for decades), not to mention macro-families which are often even less certain.

Why are language-gene correlations often broken on the Silk Road? Based on existing historic, genetic and archeological work, we know today that populations living along the Silk Road formed at different periods from diverse ethnic sources. Every ethnic group has multiple origins. None of them come from just a single source. Their languages have undergone intense contact, admixture to different degrees, some of them were mixed and some others were replaced. The rate of change and of replacement is not homogenous. At different epochs, culture and religious factors have played an undeniable role in population and language admixture. Some languages were replaced by ones spoken by those who ruled their speakers, such as Mongolic languages on the Silk Road, but some others adopted another population's language due to cultural and religious factors, such as some Mongolian rulers like Chagatai and his descendants who underwent a Turkification in Central Asia.

The technique of glottochronology, a kind of linguistic clock proposed by Swadesh in the 1950s, is still widely used despite criticism from linguists. They sensibly rejected the view that the rate of change is even in every language. Nevertheless, linguistics researchers continue to use it in calculating language distances, keeping in mind that each language evolves with its own speed, and that each word has its own history. Unfortunately, linguists do not have other efficient means to quantify linguistic data, as do biologists who can use DNA sequences to estimate the separation of species or archeologists who can exploit radiocarbon to provide a fossil dating. Languages evolve at a much faster pace and the phonetics and semantics of words undergo uninterruptible change. A language could become unintelligible over a period of some decades due to phonetic and semantic changes. Even though the glottochronology method is not ideal, nor perfect, it is one of the appropriate ways to compare languages. A less rigorous method is always better than nothing. This paper also takes the list of 200 words established by Swadesh to study the relationship between languages on the Silk Road. In the past, cognate words between different languages have been widely studied in comparative studies. In this section we will first concentrate on the synchronic linguistic situation or the result of the phonetic evolution of the lexicon, not tracing back, for the moment, cognate words from a diachronic perspective. At this step, we will only account for phonetic differences. For example, "j-" in some languages is relative to "ʤ-" in other languages since they were cognate words and had the same source in the past, and "a-" in some languages is relative to "ha-" in other languages due to sound loss, or "-d-" in one language is connected to "-ʑ-" in another due to palatalization and so on. We count all synchronic phonetic differences as divergence in order to observe and compare their degrees of change. This method presents an advantage over traditional methods of calculating cognates. Synchronic differences are visible and objective, while diachronic differences and similarities are reconstructed differently from linguist to linguist. The criterion of cognate reconstruction could be subjective or biased. In the next section, we will also use cognate word statistics, limiting subjective reconstructions as much as possible. Actually these approaches are complementary to each other and both are needed to better understand language evolution. In the Turkic group we selected Uyghur, Tuva, Kazakh, Kirgiz, Western Yugur, Salar, and Tatar (Turkish is also used for comparison); in the Mongolic group we selected Mogolian, Dagur, Eastern Yugur, Tu (Monguor), Dongxiang (Santa) and Bao'an. Some languages of the Manchu-Tungusic group are also included as reference languages, including Manchu, Xibo (Xibe), Evenki, and Nanai[2].

After a preliminary comparative study in the Turkic group, it is noted that the lexical differences between seven languages (excluding contemporary Turkish for the moment) is not very large at the phonetic level. Concretely, in seven languages of the Turkic group within China, there exist on average less than three phonetic differences out of seven words for each entry in the 200 word list. As has been mentioned, every difference is accounted for even though these words are cognates coming from the same origin. The focus is to consider their divergence today. This means that in the Turkic group, more than half of the words have quasi-identical or similar pronunciations. In consequence, the result shows that in the Turkic group, the basic words in languages are quite stable after splitting, sharing more than 60% common words with quite similar pronunciation. When contemporary Turkish is added, the percentage is similar, keeping the same level. The literature on Turkic languages, their classifications and evolution is abundant. This paper has proposed another angle from which to observe the differences and divergence of these languages, being a complement to previous work on cognate words.

In contrast, languages in the Mongolic group present more variance. On average we find near four phonetic differences out of six words in every entry in the 200 word list. As result, the degree of similarity is about 40%. It has to be noticed that such statistics are indicative and not absolute. At least general tendencies can be seen through them. Manchu-Tungusic languages are not really spoken[3] on the Silk Road. They are conventionally classified with Turkic and Mongolic groups into the Altaic family. Their statistics serve here as a comparative reference. As has been expected, Manchu-Tungusic languages are closer to Mongolic languages but further from Turkic languages. Four Manchu-Tungusic languages are considered: Manchu, Xibo (or Xibe), Evenki and Nanai. Their phonetic difference is significant; on average, more than half of their words are divergent. Only 32% similarity is observed. Let us see the summary table.

Language group	Average difference	Variance	Similarity
Turkic group (7 languages excluding Turkish)	2,84	1,98	60%
Mongolic group (6 languages)	3,60	2,21	40%
Manchu-Tungusic group (4 languages)	2,71	0,81	32%

Table 1 Comparison of 200 basic words in Altaic languages spoken in China

Table 1 clearly suggests that a greater uniformity is observed in the Turkic group while it is much smaller in the Mongolic and Manchu-Tungusic groups. Even considering contemporary Turkish spoken outside of China, the divergence in Turkic languages is not as significant as in Mongolic languages.

In this section, 13 languages (excluding 4 Manchu-Tungusic languages and Turkish) on the Silk Road in China have been examined at a basic lexical level from a synchronic phonetic perspective. In section 3, we will focus from diachronic point of view on eight representative languages which separately belong to Mongolic and Turkic groups on the Silk Road. The linguistic data will be broadened to include not only basic words but also frequent used vocabulary. Previous statistics completed by our own data will be presented and compared with genetic data.

2.2. Paternal genetic pattern in Northwest China

Similar to intensive language contact among various linguistic families, especially between Sino-Tibetan and Altaic, population admixture had occurred at different times in history, leading to an extraordinary level of human diversity in Northwest China. Overall, four main paternal lineages, C, D, O and N, accounting for more than 90% of East Asian Y chromosomes, are also prevalent in Northwestern populations (Wang and Li, 2013). However, Western Eurasia-specific Y chromosome haplogroups E, G, H, I, J, L, Q, R and T were found at relatively high frequencies in Northwest China, reflecting the recent gene flows from the west via the ancient Silk Road.

Here, to understand the genetic structure of the eight target populations, we genotyped 17 microsatellites by using a Y-filer kit (Applied Biosystems) and hierarchically typed Y chromosomes through seven panels of 100 SNPs as described in previous studies (Wang *et al.,* 2014). A neighbor-joining (NJ) unrooted tree was then constructed based on frequencies of Y chromosome haplogroups using MEGA 5.1.

From this NJ tree, the overall clustering pattern of those populations was clearly displayed. Two Mongolic-speaking populations, Bao'an and Dongxiang from Gansu, formed a cluster with two representative Turkic-speaking populations, Uyghur and Salar. The remaining two Mongolic-speaking populations, Tu and Eastern Yugur, also contributed to this cluster. The Khazakhs and Western Yugur people, by contrast, have tended to be segregated away. Hence, at the population level, Turkic-speaking populations showed much larger genetic distances than Mongolic-speaking populations as viewed from their Y chromosomes.

Another noticeable feature from this analysis is that linguistic boundaries are often transgressed across the Altaic linguistic groups. Such a phenomenon is even more pronounced among Northwestern populations, where populations which have a common origin as recorded in historical records tend to

---

[3] The Xibo people moved from the Northeast to the Northwest in 1764.

cluster in the phylogeny. This observation is consistent with the history of Northwestern populations along the ancient Silk Road, where population migrations were substantial. In next section, we will present in detail the situation in each language and analyze the reasons for matches and mismatches.
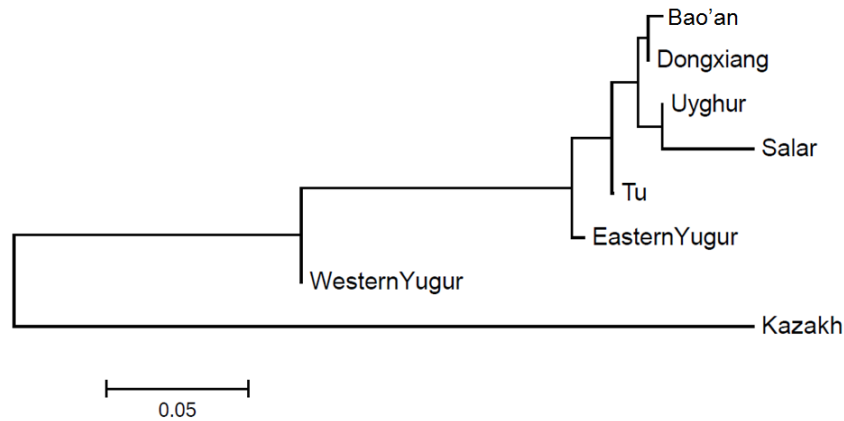

Fig 1. Neighbor-joining trees of the Mongolic and Turkic-speaking populations based on frequencies of Y chromosome haplogroups

3. The match and mismatch of two datasets

An interdisciplinary approach will be used. In particular, historical, biological and archeological research results will be taken into account to evaluate the historic frame in which languages evolved. As has been mentioned, some languages and genes are not correlated on the Silk Road. We will first compare the linguistic and genetic data we have on the Mongolic group, then on the Turkic group within China.

3.1. Mongolic languages and paternal genetic correlations

The linguistic data is composed of the *Dictionary of Mongolic languages* compiled by Sun Zhu *et al.* (1990), articles by Chen Naixiong (1990), Menhedalai and Huang Xing (1997), and our own statistics. Comparing Mongolian (Chakhar Mongolian in Inner Mongolia), Eastern Yugur, Dongxiang and Bao'an, it is seen that the percentage of cognate words to Mongolian is different and variable. Historians have shed light on populations on the Silk Road, and we now know that these four target peoples did not form in the same way. These groups each have their own characteristics concerning ethnic sources and tribal compositions. The Eastern Yugur and Tu languages correspond to Middle Mongolian, and share a significant proportion of cognates with contemporary Mongolian.

The Dongxiang and Bao'an languages however suggest another scenario. These Mongolic languages formed due to violent historical events. They are newly formed languages and most of their ancestors came from Central Asia and West Asia. Linguistic and genetic data are mismatched in these populations. The historian Liu Yingsheng (2003: 143) believes that at the beginning of the 13th century, a huge number of migrants were brought to China by Genghis Khan and his armies. There were young men forced to enroll in the army, as well as artisans of different handicrafts. This migration has left a mark on Chinese history, and even on East Asian history. It is impossible that this move did not leave any traces. These migrants were forced to learn Mongolian and today one finds few traces of Central Asian languages among these people. These Central Asians were Mongolized and their languages formed from 13th century Mongolian.

-Eastern Yugur

According to the *Dictionary of Mongolic languages* compiled by Sun Zhu *et al.* (1990), cognate words between East Yugur and Mongolian make up 54%. This statistic is based on 2093 words. Our statistics show that the cognates make up 83% of the first 100 words, and 62% of the second hundred words. On average, 72.5% cognates are found among 200 words.
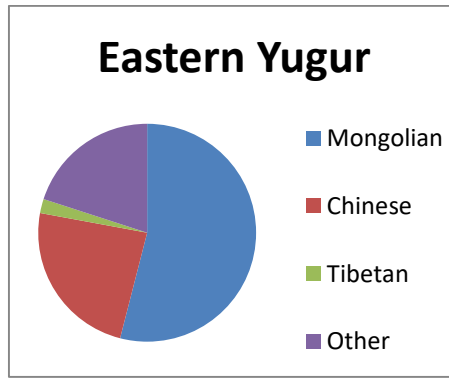
Figure 2 Different sources of the Eastern Yugur language (based on Sun Zhu *et al*. 1990)

Eighteen paternal lineages were determined from 45 individual eastern Yugur samples from Sunan Yugur Autonomous County in Gansu Province. The Yugur people are mainly comprised of two groups, Mongolic-speaking Eastern Yugur and Turkic-speaking Western Yugur. However, strong ethnic identity is shared between these two groups (Chen and Lei, 1985; Zhaonasitu, 1981). In previous genetic investigations (Shou *et al.*, 2010), it was inappropriate to employ sampling strategies that often ignored the individual's origin from Eastern Yugur or Western Yugur and thus failed to provide an insight into the relationship between language differentiation and population history. In this study, the major haplogroups C3*-M217, D1a*-N1+,N2-, D3a-P47 and J-M304 are prevalent in Eastern Yugur. Haplogroup C3*-M217, which occurs at high frequencies in Mongolic-speaking populations (Di Cristofaro *et al.*, 2013; Kharkova *et al.*, 2014; Malyarchuk *et al.*, 2010; Zerjal *et al.*, 2003; Zhong *et al.*, 2010), is also observed at a high frequency in Eastern Yugur (14.3%), suggesting a genetic link between Eastern Yugur and Mongolian populations. Typical paternal markers of Tibetan and Qiang populations (Shi *et al.*, 2008), D1-M15 and D3-P47, are also found with high frequency in Eastern Yugur. Most D1-M15 and D3-P47 haplotypes in Yugur populations are also found in surrounding populations in Gansu (such as Hui and Han) and in Tibetans from Qinghai province, exhibiting a profound influence from Tibetan populations. It is particularly noteworthy that West Eurasian haplogroup J-M304 (Zhong *et al.*, 2011) is detected at considerable frequencies in Eastern Yugur. One J-M304 haplotype is shared between Eastern Yugur, Uyghur and Hui in Gansu province, implying that they have experienced common historical events.
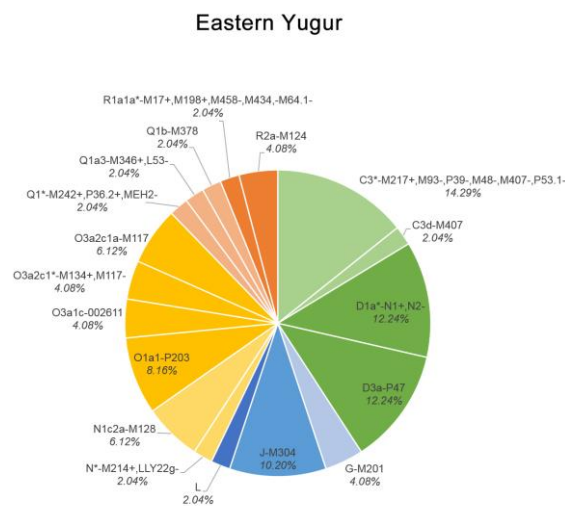

Figure 3 Observed paternal lineages in Eastern Yugur

-Tu (Monguor)

The origin of Tu has been discussed by historians and linguists (Zhou Weizhou 2004, Li Keyu 2008, Qiepai 2009 among others). The most popular idea is that the Tu people constitute one of the branches of the Mongols and their origin is from Xianbei. The core ancestors of Tu were Tuyuhun khan and his offspring, who headed Murong tribes. During the initial formation, Qiang, Tibetan and Chinese people joined this clan. After the disintegration of the Tuyuhun Kingdom (4th-7th centuries), some populations were Tibetanized and others Sinicized. The genetic data also suggests that the contribution from Central Asia is not negligible. Linguistic data shows that the Tu language shares as many cognates with Eastern Yugur as with Mongolian. Specialists think that Tu is very similar to 13th century Mongolian (Zhaonasitu 1981a, Li Keyu 2008). Mongolian vocabulary is shown to make up a major portion in Sun Zhu *et al.* (1990) with a significant percentage of Chinese loans. Other sources are not significant in number. Please see the following figures. In our statistics, in the 100 word list, Tu share cognates with Mongolian at a percentage of 78% while in the second 100 word list, it attains 65%. On average, cognates account for 71.5%. The following figure is based on Sun Zhu *et al.* (1990), with statistics based on 1845 words.
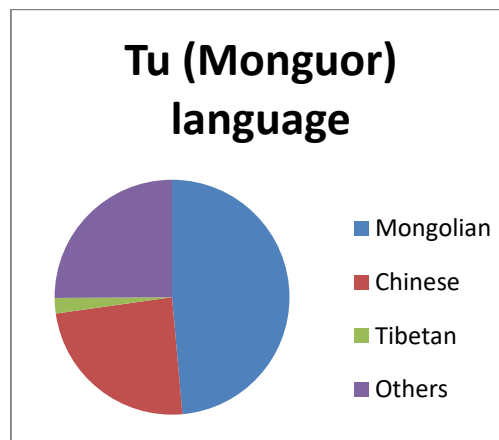

Figure 4 Different sources of the Tu language (based on Sun Zhu *et al*. 1990)

Twenty-four paternal lineages were determined from the 121 individual Tu samples from Huzhu Tu Autonomous County in Qinghai Province. Haplogroups R1a1a-M17, D1-M15, O3a2c1a-M117, O3a1c-002611 were the most frequent lineages, together comprising 44.6% of the Tu people's paternal gene pool. Haplogroup R1a1a-M17 is predominant in some Turkic-speaking populations (Di Cristofaro *et al.*, 2013; Shou *et al.*, 2010; Wells *et al.*, 2001) and appears with relatively high frequency in Tu people. As mentioned above, haplogroups D1-M15 and D3a-P47, associated with Tibeto-Burman populations (Shi *et al.*, 2008), are also observed at a high frequency and moderate frequency in Tu (10.74% and 4.13%, respectively). Main sublineages of O3-M122, O3a2c1a-M117, O3a1c-002611, accounting for 20.7% of the Tu samples, which are predominant in Han and Hui populations but are rare or absent in Mongolic and Turkic populations (Di Cristofaro *et al.*, 2013; Shou *et al.*, 2010; Wells *et al.*, 2001; Wen *et al.*, 2004). Hence these lineages can be treated as genetic influences from Han and Hui populations in the same region. Notably, haplogroup C3*-M217, as well as C3c-M48 and C3d-M407, is the common paternal lineage in Mongolic-speaking populations (Di Cristofaro *et al.*, 2013; Kharkova *et al.*, 2014; Malyarchuk *et al.*, 2010; Zerjal *et al.*, 2003; Zhong *et al.*, 2010), which are also found at moderate or relatively low levels in the Tu people, indicating the complicated genetic connection between Tu and Mongolian populations.
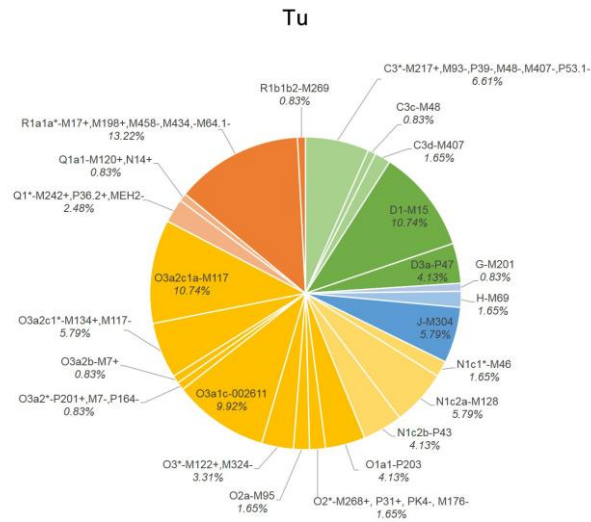
Figure 5 Observed paternal lineages in Tu

-Dongxiang

The origin of the Dongxiang people is clear. At least more than half (unpublished statistics from Fudan's laboratory) of the population came from Central Asia (cf also Shou *et al*. 2010). As has been mentioned, this language formed during the 14th century due to the large migration after the conquest by Genghis Khan of Central Asia in the 13th century. The following chart is based on our research published in 2013 showing that the Dongxiang vocabulary (based on 10994 words) has Mongolian as its main source, keeping only 5.65% of the parent language's vocabulary. Chinese loan words make up around 35%. Today, the Dongxiang language is considered to be a Mongolian variety; actually it was a replaced language, i.e. the ancestral populations might have spoken other languages of Central Asia and Western Asia, but they were forced to change their mother tongue by the ruler's language in order to survive. As for the first 100 basic words, Dongxiang has 70% cognate words with Mongolian, and in the second hundred words, the percentage is 48%. On average, 59% of words cognate with Mongolian can be found on the list of 200 words. The linguistic situation is quite similar to Bao'an, another replaced language on the Silk Road. The following graph is based on 10994 words.
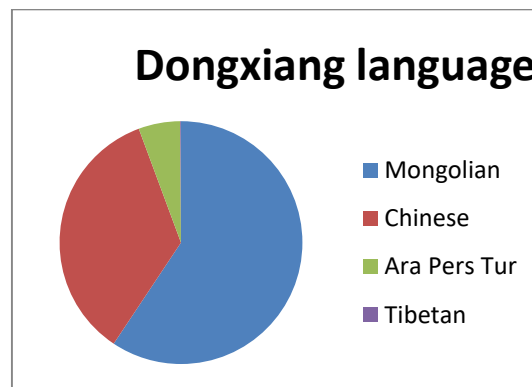

Figure 6 Different sources of the Dongxiang language (based on Dan Xu 2014)

In our previous study, we studied Y-chromosomal data drawn from the literature (Wen *et al*., 2013) and found that the Dongxiang people did not descend from Mongolians, but from Central Asian and Western Asian ethnic groups that spoke Persian or Turkic languages. Here, twenty-five paternal lineages were determined from the 107 individual Dongxiang samples from Dongxiang Autonomous County in Gansu Province. Relatively high frequencies of the western Y chromosomal haplogroups J-M304, R1a1a-M17 and R2a-M124 in Dongxiang people further strengthen our conclusion that the Dongxiang people are much closer to Central Asian ethnic groups than to the other Mongolians. Additional

admixture, in the presence of low and moderate frequencies of the typical East Asian haplogroups C, D, N and O, took place primarily on their migration route or during subsequent settlement in Northwest China.
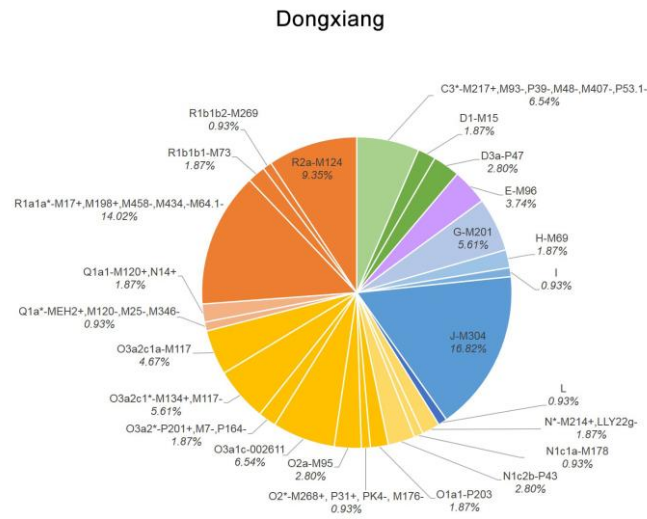

Figure 7 Observed paternal lineages in Dongxiang

-Bao'an

Like Dongxiang, the genes of Bao'an populations came from Central Asian people, but linguistically, their language belongs to the Mongolic language group. Chen Naixiong's studies (1990b: 32) indicate that Bao'an should be divided into two dialects, one in Qinghai (Nianduhu) and one in Gansu (Ganhetan). These two dialects show a different rate of change, conditioned by their geographic separation and different religious beliefs. The statistics by Chen (1990a: 17) are based on 3596 words in Nianduhu Bao'an and 3624 in Ganhetan Bao'an. According to him, 28.70% of Mongolian cognate words are found in Nianduhu and 28.12% in Ganhetan. In the basic vocabulary, we count 71% cognates with Mongolian in the first 100 word list and 47% cognates in the second 100 word list. In consequence, 59% cognate words with Mongolian are seen on average for 200 words. This result is almost identical with Dongxiang.

Chen Naixiong (1990a:17) provides interesting comparative studies. He remarks that in the two dialects of Bao'an, Nianduhu has borrowed 53.62% of words from Amdo Tibetan and Ganhetan has borrowed 58.11% of loan words from Chinese, but the consequence is different. He thinks that Bao'an in Nianduhu has maintained its syntax relatively well despite heavy borrowings while Bao'an in Ganhetan began to change its syntax due to intense language contact.
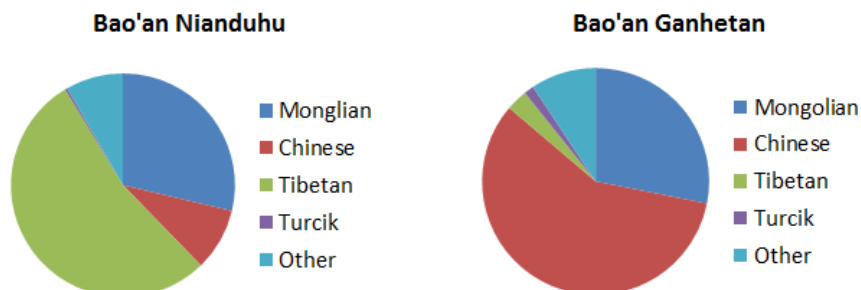

Figure 8 Comparison between two Bao'an dialects (based on Cheng Naixiong 1990)

Twenty-one paternal lineages were determined from the 64 individual Bao'an samples from Jishishan Bao'an, Dongxiang and Salar Autonomous County in Gansu Province. Three dominant lineages of the Bao'an paternal gene pool were identified based on their putative origin: West Eurasian (J-M304), Sino-Tibetan (O3a2c1a-M117) and Mongolian (C3*-M217). This result is in agreement with the hypothesis of their ancient presence in the West Eurasia and their subsequent population admixture with Sino-Tibetan and Mongolian populations in Northwest China.
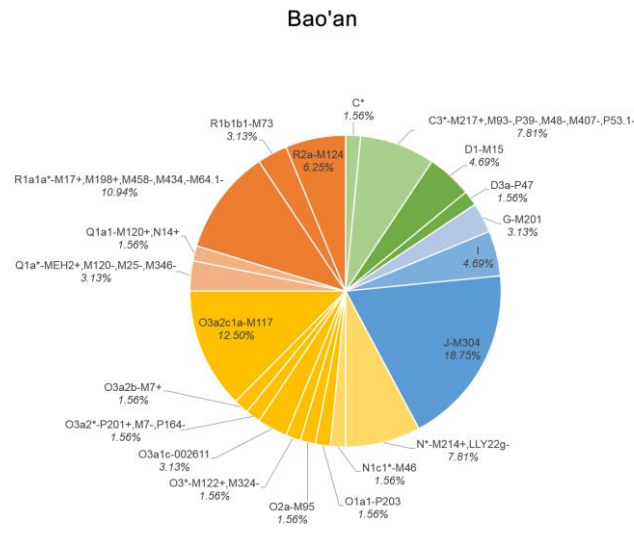

Figure 9 Observed paternal lineages in Bao'an

3.2. Turkic languages and paternal genetic correlations

The four languages in the Turkic group, Uyghur, Western Yugur, Kazakh and Salar formed at different periods and from different clans, except Uyghur and Western Yugur which are thought to be from the same ancestral tribes. According to historic documents, after the fall of the Uyghur Khanate which was defeated by Yenisei Kirghiz in 840, the Uyghur people were divided into three groups dispersing from valleys of the Selenga River and the Orkhon River. One group established the Kara-Khanid dynasty (10th-13th centuries) between today's Xinjiang and the Pamirs in Central Asia. It ended with Genghis Khan's conquest toward the West. The second group became the Uyghur of Gaochang (Xingjiang, 10th-13th centuries), which also ended with Mongol armies headed by Genghis Khan. These populations are believed to be part of the ancestors of modern Uyghur in Xinjiang. Finally the third group was the Hexi Uyghur, also called Ganzhou Uyghur (900-1036). Their rulers were overthrown by Li Yuanhao, emperor of Tangut. These peoples are identified as the forefathers of the Western Yugur. Kazakh people were composed mainly of Mongolian populations (see genetic analysis) which underwent a Turkification. Salar people came from Central Asia and their origin is clear since research from different domains such as history, anthropology, linguistics and genetics has converged without exception. During the formation of these peoples, other ethnic groups joined them forming populations with diverse origins. In these four languages, contemporary Uyghur and Kazakh are the best preserved Turkic language in China. Chinese loan words are present at the lowest level in these two languages, while in Western Yugur and Salar, Chinese borrowings at the lexical level show a significant percentage. The linguistic data is mainly based on the *Dictionary of Turkic languages in China* by Chen Zongzhen. *et al*. (1990), supplemented with published documents and our own statistics. The dictionary compiled by Chen *et al*. is the most complete dictionary of contemporary Turkic languages in China. Let us see concrete data.

-Uyghur

As mentioned earlier, due to their historical and cultural background, Uyghurs have been less influenced by Sinitic languages. Chinese loan words only make up around 3.19% in Uyghur. Its cognate words with other Turkic languages are significant, with 17.74% words of Arabic origin and 10.72% Persian

words. As we know, Uyghurs are believed to be one of the three branches that came from Old Uyghurs who migrated from valleys of the Selenga River and the Orkhon River to Xinjiang after the fall of their khanate. Xinjiang has always been a meeting place of West with East. At this site, Tocharian, Sogdian and other western languages were spoken and then extinguished due to climate conditions and cultural and religious conflicts. Old Persian and Arabic, Old Chinese and other languages had also left traces in Uyghur languages and dialects over the centuries. Modern Chinese and Russian have also influenced Uyghur but the impacts seem to be much smaller than from Arabic and Persian. These two languages can be considered as traces of language contact in an earlier stratum. The following figure is based on 4075 words.
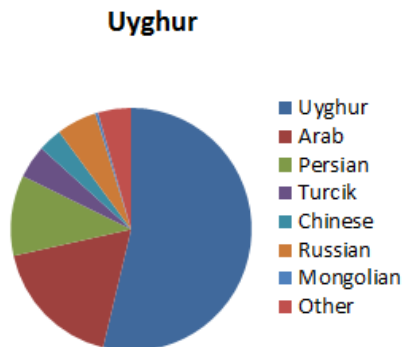

Figure 10 Different sources of the Uyghur language (based on Chen Zongzhen *et al.* 1990)

The Uyghur people, which account for about half of the total population of Xinjiang, are a population presenting a typical admixture of Eastern and Western anthropometric traits. In this article, twenty-seven paternal lineages were determined from the 134 individual Uyghur samples from Turpan City. As predicted, on the one hand, Central-South Asia and West Eurasia related haplogroups R1a1a-M17 and J-M304 have also been detected at considerable frequencies, especially R1a1a-M17, which contributes 19.83% of Uyghur; on the other hand, relatively high frequencies of North Asia and East Asia related haplogroups C3*-M217 and O3a2c1*-M134+,M117- were observed in Uyghur populations. Our result clearly indicates a typical genetic admixture pattern in the Uyghur population with ancestries from both eastern Eurasia and western Eurasia.
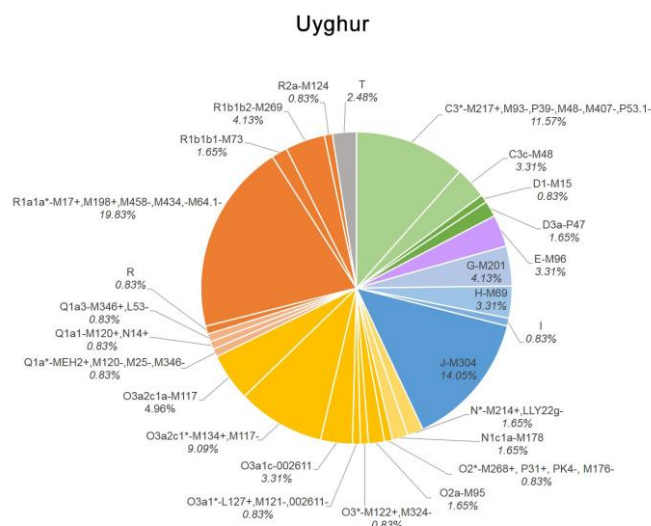

Figure 11 Observed paternal lineages in Uyghur

-Western Yugur
Here the linguistic data in Eastern Yugur is considered again in comparison with Western Yugur. It has been recognized that Eastern Yugur and Western Yugur belong to two subgroups (Mongolian and Turkic)

of one language family, Altaic. In spite of their centuries-long union, their languages remain independent and even today they cannot communicate if one is not bilingual[4]. The populations speaking these two languages are located in the vast grasslands of southern Gansu, known as Sunan. Comparing these two following Figures, it is shown that WY is heavily influenced by Chinese at the lexical level (38%). The basis of statistics for WY is 3404 words while for EY we have 2093 words. Despite these significant lexical borrowings from Chinese, it seems that the syntax of WY has not been affected. Based on 851 frequent and comparable words, WY and EY share 13.74% common words. Chinese loanwords make up almost 19% in WY while they only present 9% in EY.
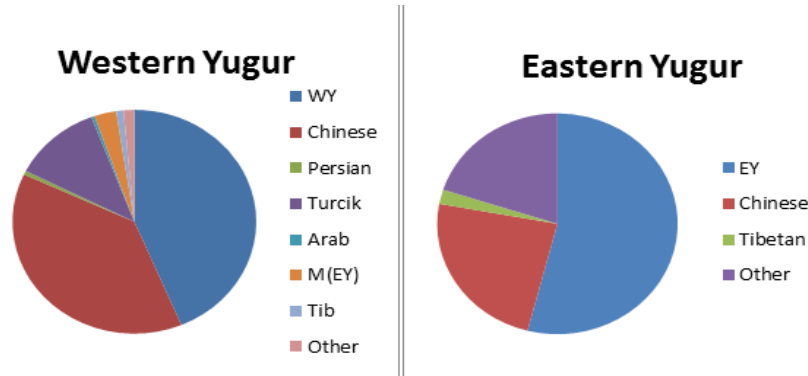

Figure 12 Comparison between WY and EY
(based on Chen Zongzhen 1990 and Sun Zhu 1990)

In total, 54 WY samples were collected from Sunan Yugur Autonomous County. Haplogroups C3*-M217, O3a2c1a-M117 and Q1a*-MEH2+,M120-,M25-,M346- were the most frequent lineages in WY. In comparison of their haplogroup frequencies, EY and WY have considerable differences. Haplogroups D3a-P47 and O1a1-P203 was significantly more frequent in EY (12.24% and 8.16%, respectively) than in WY (both absent). Additionally, C3*-M217 and Q1a2a*-L53+, M3- was significantly more frequent in WY (41.18%) than in EY. Notably, Haplogroup Q1a2a*-L53+, M3- has so far only been found in Altain, Hui and Tuvan (Dulik *et al.*, 2012 and unpublished data by the Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University). Altain and Tuvan are two Turkic-speaking populations in South Siberia. This lineage represented the close genetic relationship of the Yugur population with Tuvan and Altain in South Siberia. Interestingly, three of six Q1a2a*-L53+, M3- samples share the same family name "Yaghlaqar" (Yang in Chinese, see section 4.1.). This family name is inherited from the royal family of the ancient Uyghur Khanate (Li, 2010; Yang, 2003; Zhang, 1991). The genetic signal we observed here clearly shows a genetic tie between ancient Uygur people and present-day Tuvan and Altain. Therefore, although most haplogroups are shared between these two groups, strong paternal differentiation can also be observed between Turkic Yugur and Mongolic Yugur.

---

[4] Their language of communication is the Chinese language.

Western Yugur
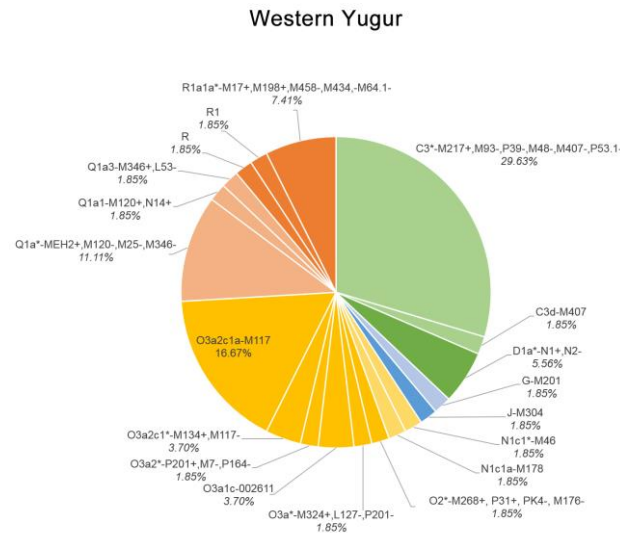

Figure 13 Observed paternal lineages in Western Yugur

-Kazakh

According to historic records, Kazakh people constitute a combination of Mongols and Kazakh people who later formed a community of Turkic tribes. Due to Islamic religion and culture, a significant number of Mongolic tribes, often headed by Genghis Khan's offspring, joined Kazakh clans. The process of Turkification lasted 150 years (Lin Gan and Gao Zihou 1994). It is amazing to observe that this Turkification was so complete that these converted Mongol tribes completely absorbed the Turkic language. To a certain extent, one can say that the mother tongue of these Mongols was replaced by a Turkic language. As in modern Uyghur, Chinese loan words in the Kazakh language are almost insignificant, while Arabic and Persian words account for nearly 21%.

Kazakh

Figure 14 Different sources of the Kazakh language (based on Chen Zongzhen *et al.* 1990)

Eleven paternal lineages were determined from 93 individual Kazakh samples from Aksai Kazakh Autonomous County in Gansu Province. This population presented the lower haplogroup and haplotype diversity values compared to the other studied populations, as they were mainly from a part of Kerey tribe that experienced a series of founder effects or strong bottlenecks when emigrating from Xinjiang Province to Gansu Province (Xie and Wang, 2009). Haplogroup C3*-M217 comprises an overwhelming majority of Gansu Kazakh people (81.72%), reflecting the complex history of interaction between Turkic-speaking populations and Mongolic-speaking populations.
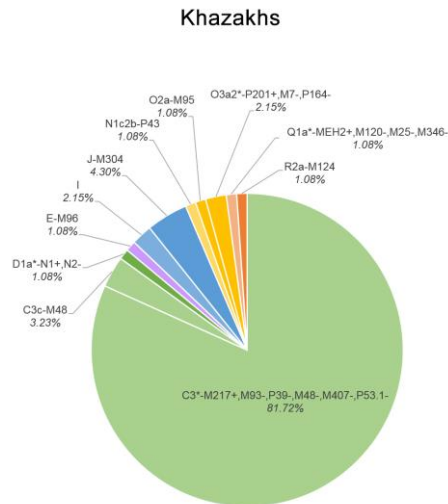
Khazakhs

Figure 15 Observed paternal lineages in Kazakh

-Salar
Historians such as Mi Yizhi (1981) think that Salars were descendants of Salor clans which belonged to Oghuz tribes. It is believed that they were brought into China by Genghis Khan from Samarkand after his victory in Central Asia. Oral legends say that these populations settled in Qinghai after a long travel led by two brothers who could not stand local repression of their family. Historical records tell us that Salars had helped Genghis Khan in beating the Tangut kingdom which fell in 1227 (see Mi Yizhi 1981). Whatever the reason for their movement from Central Asia to Xunhua in Qinghai, they were probably recruited as conscripted troops like the Santa (today called Dongxiang) and Bao'an populations. Unlike Uyghur and Kazakh, Arabic and Persian sourced words only make up 5.3% at a basis of 3147 words while Chinese loanwords reach almost 32%, comparable to the lexical borrowing level of Western Yugur.
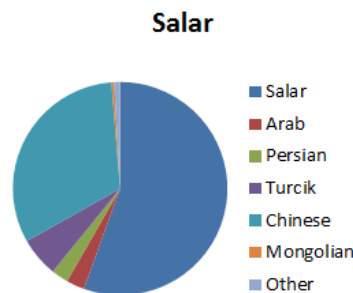
Figure 16 Different sources of the Salar language (based on Chen Zongzhen 1990)

Twenty paternal lineages were determined from the 134 individual Salar samples from Xunhua Salar Autonomous County in Gansu Province. Similar to the Y-chromosomal lineage C3*-M217 in Gansu Kazakh people, the Central Asian specific Y-chromosomal lineages R1a1a*-M17 is observed at extremely high frequency in Salar people (35.07%). The presence of haplogroups C, D, N, O and Q at lower frequency is due to local population admixture, because haplotype sharing among these populations exists, which is consistent with historical records about the formation of the Salar population.

Figure 17 Observed paternal lineages in Salar

4. Two models of language replacement on the Silk Road

Grounded on our eight linguistic and genetic samples belonging to Mongolic and Turkic groups, we will observe admixture and replacement of these languages on the Silk Road. It is certain that different models of language replacement exist and language repla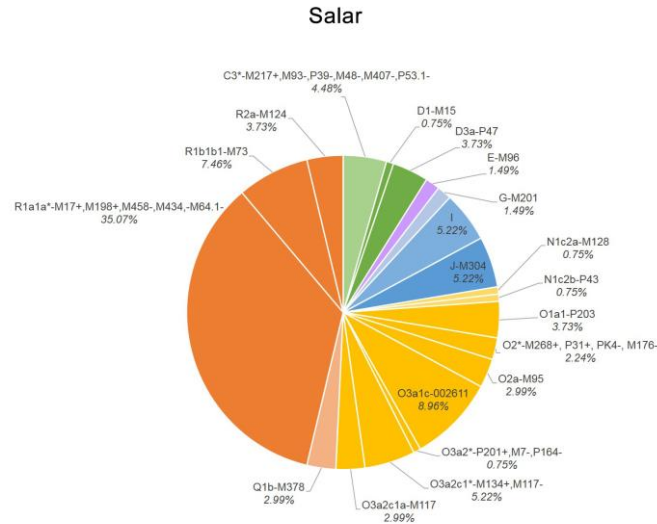cement can be provoked by numerous reasons. Here we will focus on two main models from our point of view, *Elite Dominance* model and *Cultural Dominance* model.

4.1. Elite Dominance model

The Elite Dominance model was first proposed by Renfrew (1987) in a language replacement process. It assumes that newcomers bring natives into subjection causing language change. Language coexistence may last a time and a bilingual situation subsists at the beginning stage. One possible result is that the indigenous language is replaced by that of the newcomers after a bilingual period. According to Renfrew's model, new dominant peoples possess an organized society, and/or advanced military technologies such as horse riding, wheel fabrication and so on. These small groups which have technical and political advantages over larger resident populations, will have the chance to dominate the local residents. Renfrew took Proto-Europeans as a reference from an archeological perspective.

Here we use the *Elite Dominance model,* slightly modifying it to apply to situations on the Silk Road. Within this model, two situations can be considered or refined: (1) a group with few members, generally a ruler group or military conquerors, force a larger group to speak the dominant group's language. We will call these groups RULER GROUPS; (2) a few clans which occupy a higher social position due to their clan founder status or royal offspring origin, unify scattered clans from different sources. The language spoken by the core clans is venerated as common language and followed by other less influenced tribes. We will call these groups LEADER GROUPS.

In our target region, Dongxiang and Bao'an are two typical samples in which the language formation or replacement can be explained by a RULER GROUP i.e. the first *Elite Dominance* model. As has been mentioned in previous sections, Dongxiang and Bao'an people have their main ancestral sources in Central Asia. Their core ancestors were brought by Genghis Khan and his army to China. They were forced to settle in a region corresponding to today's borders between Gansu and Qinghai to participate in military activities or produce goods for the military. Small groups of Mongolian officers governed larger immigrant populations who likely spoke different languages or dialects of their homelands. One can imagine that at the beginning they were forced to learn Mongolian taking it as an official language to communicate one to each other.

Comparing genetic and linguistic datasets shows that languages and genes present a disagreement in Dongxiang and Bao'an. These comparative studies lead us to assume that the original languages of these populations were replaced in the past. More concretely, these peoples originated after the 13th

century and their languages were newly created. Historical documents suggest that Dongxiang and Bao'an formed around the 14th century. This partly illuminates why the western branch of the Mongolic language group clearly shows different properties which are unique with respect to other Mongolic languages.

Now let us observe the LEADER GROUP situation, representing another case of the *Elite Dominance* model. In this case, populations are not forced to speak the language of dominant groups, but rather deliberately follow the tribe leaders which form core clans, and are respected by different smaller tribes. They joined the core tribes in periods of important events (military or political). This is the case of Yugur people. These clans headed by two leader families are composed of two groups: Western Yugur and Eastern Yugur. Western Yugur belongs to the Turkic language group and Eastern Yugur to the Mongolic language group. Yugur populations were nomads in the past and some of them have kept this lifestyle today. They are formed by different clans or families, with more than 14,000 persons nowadays, being composed of around fifty clans or families which each inherited their ancestral tribe name. Fifty Yugur family names reveal quite diverse sources representing each Yugur clan and allow us to trace back the multiple origins of Yugur populations embedded into two main groups, Western Yugur and Eastern Yugur. The case of Yugur clearly shows that diverse ethnic people were incorporated into these two famous clans speaking languages of these two families.

As has been mentioned in 3.2, Western Yugur and Uyghur may come from common ancestors who split in 840 after their Khanate fell due to Yenisei Kirghiz. Due to their different religious beliefs, some Old Uyghurs escaped from Xingjiang and settled in Gansu becoming Western Yugur. Among them, one family is famous for being one the founders of the Old Uyghur khanate. This clan's name is Yaklagar, with a Sinicized name *Yang* today. Actually, the name is already attested in Suji inscriptions and Terkhin inscriptions, thought to be Old Turcik from the 8th century. This name is also found in historic documents in Chinese such as *Song Shi* (Annals of Song, around 14th century). With genetic data, we observe that their ethnic sources are diverse, unifying multiple people coming from Turkic, Mongolic, Tibetan and other smaller tribes. We suppose that those people, who had joined the Yaklagar clan, must have learned the Turkic language for communication.

Another famous clan, Anjang, is the leader of the Eastern Yugur. Since the emperor of the Ming Dynasty (1368-1644) designated Buyan Tömöl as *Andingwang* (King of maintaining stability), the Anjang clan has officially been considered the leader clan of the Yugur (Western and Eastern). The historian Gao Zihou (2000: 134-136) supposes that this family consists of Genghis Khan's descendants. The genetic data tells us that ethnic sources of Eastern Yugur are as diverse as Western Yugur. From a genetic perspective, C*-M217+ is believed (Zerjal T *et al*. 2002, Serikbai Abilev *et al*. 2012) to be the characteristic lineage from Genghis Khan's clan. It is found in both Eastern and Western Yugur people. This data may correspond to the fact that a governor from the Anjang family was sent to every tribe as leader since the beginning of the Ming Dynasty around the 14th century. This suggests the same scenario within which different tribes or clans were attracted by this key family and learned Mongolian, the source language of Eastern Yugur.

In the *Elite Dominance* model, as has been seen, the RULER GROUP forces the subjected people to speak the dominators' language, while the LEADER GROUP possesses a centripetal force to attract diverse populations to learn the leader tribes' languages.

4.2. Cultural Dominance model

In the present work, a second model will be presented in the process of language formation and replacement: the *Cultural Dominance* model. In this model, populations adopting the culture and religion of another community are assimilated into a different ethnic group. This can happen to a group with a significant populations size, or to a small group of dominators.

Yugur is an interesting example. One group of Yugur residing in Huangnipu (Jiuquan county of Gansu province) has been completely Sinicized, perhaps since the Qing Dynasty[5]. This group of around 1000 people changed its lifestyle from nomadic to agrarian. Influenced by Chinese culture and language, these populations totally adopted the Chinese language and culture and do not speak Western Yugur any more. Geographically they are located very close to Jiuquan City (in Jiuquan County, Gansu) separated

---

[5] Saliyanyu (2000: 592) indicates that Yaklagar tribe-head-system was abolished by the Qing government in Huangnipu.

from other Yugurs who are still mainly nomadic. The Chinese language and cultural influences are much stronger than those exerted on the vast grassland in the south of Gansu. Their nomadic lifestyle has prevented Yugur people from losing their culture and language while an agrarian lifestyle caused the identity loss of other Yugur. This is one of the results of the *Cultural Dominance* model.

This also happened to a significant part of Kazakh populations. Haplogroup C-M130 is very frequent in Mongolian people, but occupies a high percentage (by different papers' statistics) in Kazakh populations speaking the Turkic language. Historic documents show us that Kazakh people are in fact mixed with Mongolian people due to Genghis Khan's descendants. Chagatai (the second son of Genghis Khan) and his offspring were rulers, but were surrounded by Central Asian populations which followed Islam. Over less than two centuries, these Mongolian rulers and their tribes underwent a Turkification and began to speak Turkic languages. This language replacement was due to their change of religious belief. This assimilation by Mongolic groups was voluntary and deliberate, due to cultural and religious factors. This time, the dominant culture came from dominated people, but the dominators, being converted into a new cultural system, accepted with their own volition the religion of dominated peoples. The bridge leading them to accomplish this change was the Turkic language.

The other famous example of language replacement caused by the *Cultural Dominance* model is the Manchu. The Manchurians were rulers in China from 1616 to 1911. The Manchu language is today almost extinct. This language replacement has been studied by scholars such as Ji Yonghai (2004, 2005) who indicates three stages in which Manchu gradually declined and was replaced. Manchu was spoken alongside Mandarin in Beijing at the beginning of the 18th century, but the bilingual situation changed to monolingual in the 19th century. Finally Manchu was no longer spoken even by descendants of Manchu populations. This case is another example in which a RULER GROUP accepted the culture of foreign populations dominated by it.

Here two models are presented with concrete examples. Except for the last case with Manchu, other cited examples are all attested on the Silk Road. It is possible that language replacement occurred more than once for some languages and for some people.

5. Discussion

In this article, we have compared two sets of data, linguistic and genetic in a biological sense, from populations on the Silk Road. Our analysis focused on eight languages and populations belonging to Mongolic and Turkic groups. It has shown that languages and genes often do not correlate on the Silk Road due to multiple factors. It has been assumed that the Western branch of the Mongolic group often reveals particular properties in their language due to important historical events. This branch indeed has a special status on the Silk Road since these languages were newly formed from speakers of different languages. Dongxiang, Bao'an, and a large part of Kazakh people present instances of language replacement. But the reasons for replacement are quite distinct. One can be explained by the *Elite Dominance* model, as is the case for Dongxiang and Bao'an which are newborn or replaced languages which formed, around 14th century and the other by the *Cultural Dominance* model, which is the case for Kazakh which formed when a majority of Mongolic groups underwent Turkification around the 13th and 14th centuries (see Lin and Gao 1994: 226-227). Remember that the cases of Dongxiang and Bao'an belong to the first situation, i.e. the RULER GROUP of the *Elite Dominance* model, in which dominated populations were forced to change their language.

Yugur populations have been divided into two groups, Western Yugur and Eastern Yugur. It has been shown that both groups had multiple source clans. Different ethnicities have contributed to their formation. Concretely, among Eastern Yugurs, Tibetan populations which have joined Mongolic clans in the past were Mongolized, also including an important part of Han populations. Western Yugur populations, speakers of Turkic languages, were significantly enlarged by Mongolic clans at different periods. This corresponds well to historic documents (see Lin and Gao 1994: 258) which reveal that some Mongolic clans joined Sari Uyghurs at the beginning of the Ming Dynasty (1368-1644). Yugur populations clearly reflect the *Elite Dominance* model in which the second situation, i.e. a LEADER GROUP, played a key role in their language choice. Yugurs in Huangnipu, who have been Sinicized, have followed the *Cultural Dominance* model.

It is evident that language formation and replacement can be provoked by multiple factors and reasons. Explanations and analysis provided by this paper are tentative rather than rules. Our interdisciplinary approach attempts to take advantage of research results from different domains, in

particular linguistics and molecular biology, to provide a description of a historical scenario which should be closer and more faithful to reality.

References

Abilev, S., Malyarchuk, B., Derenko, M., Wozniak, M., Grzybowski, T., Zakharov, I., 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, 79-89.

Chen, Naixiong 1990a. Tracing the evolution of Bao'an. *Minzu yuwen* [Non-Han languages] 3, 16-25.

Chen, Naixiong 1990b. Tracing the evolution of Bao'an (continuation). *Minzu yuwen* [Non-Han languages] 4, 42-48.

Chen, Zongzhen *et al*. 1990. *A Lexicon of Turkish languages in China*. Beijing: Minzu chubanshe.

Chen, Z., Lei, X., 1985. *Brief records of West Yugur language*. Beijing: Nationalities Publishing House.

Di Cristofaro, J., Pennarun, E., Mazieres, S., Myres, N.M., Lin, A.A., Temori, S.A., Metspalu, M., Metspalu, E., Witzel, M., King, R.J., Underhill, P.A., Villems, R., Chiaroni, J., 2013. Afghan Hindu Kush: where Eurasian sub-continent gene flows converge. *PloS one* 8, e76748.

Dulik, M.C., Zhadanov, S.I., Osipova, L.P., Askapuli, A., Gau, L., Gokcumen, O., Rubinstein, S., Schurr, T.G., 2012. Mitochondrial DNA and Y chromosome variation provides evidence for a recent common ancestry between Native Americans and Indigenous Altaians. *American journal of human genetics* 90, 229-246.

Gao, Zihou 2000. The source, evolution and important impacts of Mongolic subgroups of the Yugurs. In Zhong Jinwen (ed.) *Studies of the Yugur ethnic group in China.* Beijing: Minzu chubanshe. 133-137.

Ji, Yonghai 2004. From contact to admixture-on the decline of the Manchu language (1). *Manchu Studies.* 1, 24-34.

Ji, Yonghai 2005. From contact to admixture-on the decline of the Manchu language (2). *Manchu Studies.* 1, 49-55.

Kharkova, V.N., Khaminaa, K.V., Medvedevaa, O.F., Simonovaa, K.V., Ereminab, E.R., Stepanova, V.A., 2014. Gene Pool of Buryats: Clinal Variability and Territorial Subdivision Based on Data of Y-Chromosome Markers. *Russian Journal of Genetics* 50, 180-190.

Li, Keyu 2008. *Collections of studies on the Tu people's history and languages*. Beijing: Minzu chubanshe.

Li, S., 2010. *Research on Oghuz and Uighur*. Beijing: Nationalities Publishing House.

Lin, Gan and Gao Zihou 1994. History of the Old Uyghurs. Huhhot: Neimenggu renmin chubanshe.

Liu, Yingsheng 2003 A brief review of the evolution of the Hui language -from Persian to Hui Chinese. *Studies of Chinese culture*. Winter tome, 143-153.

Malyarchuk, B., Derenko, M., Denisova, G., Wozniak, M., Grzybowski, T., Dambueva, I., Zakharov, I., 2010. Phylogeography of the Y-chromosome haplogroup C in northern Eurasia. *Annals of human genetics*. 74, 539-546.

Menhedalai and Huang Xing 1997. Analysis of the cognate distribution in the basic vocabulary of Mongolic and Turkish languages. *Minzu yuwen* [Non-Han languages] 1, 45-51.

Mi, Yizhi 1981. Exploration of the origin and migration of the Salar. *Qinghai minzu xueyuan xuebao* [Journal of the Institute of Qinghai Minorities]. 3, 59-66.

Qie, Pai 2009 *Studies of peaceful cohabitation of multiple ethnicities and their trends of development in the Hexi Corridor*. Beijing: Minzu chubanshe.

Renfrew, C., 1987. *Archeology and Language*. Cambridge: Cambridge University Press.

Saliyanyu 2000. Yugurs in Jiuquan. In Zhong Jinwen (ed.) *Studies of the Yugur ethnic group in China.* Beijing: Minzu chubanshe. 592-594.

Shi, H., Zhong, H., Peng, Y., Dong, Y.L., Qi, X.B., Zhang, F., Liu, L.F., Tan, S.J., Ma, R.Z., Xiao, C.J., Wells, R.S., Jin, L., Su, B., 2008. Y chromosome evidence of earliest modern human settlement in East Asia and multiple origins of Tibetan and Japanese populations. *BMC biology* 6, 45.

Shou, W.H., Qiao, E.F., Wei, C.Y., Dong, Y.L., Tan, S.J., Shi, H., Tang, W.R., Xiao, C.J., 2010. Y-chromosome distributions among populations in Northwest China identify significant contribution from Central Asian pastoralists and lesser influence of western Eurasians. *Journal of human genetics* 55, 314-322.

Sun, Zhu *et al*. 1990. *Dictionary of Mongolic languages*. Xining: Qinghai renmin chubanshe.

Wang, C.C. and Li, H., 2013. Inferring human history in East Asia from Y chromosomes. *Investig Genet* 4 (1), 11.

Wang, C.C., Wang, L.X., Shrestha, R., Zhang, M., Huang, X.Y., Hu, K., Jin, L. and Li, H., 2014. Genetic structure of Qiangic populations residing in the western Sichuan corridor. *PLoS One* 9 (8), e103772.

Wells, R.S., Yuldasheva, N., Ruzibakiev, R., Underhill, P.A., Evseeva, I., Blue-Smith, J., Jin, L., Su, B., Pitchappan, R., Shanmugalakshmi, S., Balakrishnan, K., Read, M., Pearson, N.M., Zerjal, T., Webster, M.T., Zholoshvili, I., Jamarjashvili, E., Gambarov, S., Nikbin, B., Dostiev, A., Aknazarov, O., Zalloua, P., Tsoy, I., Kitaev, M., Mirrakhimov, M., Chariev, A., Bodmer, W.F., 2001. The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proceedings of the National Academy of Sciences of the United States of America* 98, 10244-10249.

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

Wen, S.Q., Xie, X.D. and Xu, D., 2013. Contact and admixture-the relationship between Dongxiang population and their language viewed from Y chromosomes. *Yi Chuan* 35 (6), 761-770.

Xie, G.,X., and Wang, X.,P., 2009. *History of Gansu Kazakh*. Lanzhou: Gansu Cultural Publishing House.

Yang, F., 2003. *Uighur literature and Uighur culture*. Beijing: Nationalities Publishing House.

Zerjal, T., Xue, Y., Bertorelle, G., Wells, R.S., Bao, W., Zhu, S., Qamar, R., Ayub, Q., Mohyuddin, A., Fu, S., Li, P., Yuldasheva, N., Ruzibakiev, R., Xu, J., Shu, Q., Du, R., Yang, H., Hurles, M.E., Robinson, E., Gerelsaikhan, T., Dashnyam, B., Mehdi, S.Q., Tyler-Smith, C., 2003. The genetic legacy of the Mongols. *American journal of human genetics* 72, 717-721.

Zhang, Z., 1991. *History of Gansu Yugur*. Lanzhou: Gansu Cultural Publishing House.

Zhaonasitu 1981. *Outline of the Tu language*. Beijing: Minzu chubanshe.

Zhaonasitu, 1981. *Brief records of East Yugur language (in Chinese)*. Beijing: Nationalities Publishing House.

Zhong, H., Shi, H., Qi, X.B., Duan, Z.Y., Tan, P.P., Jin, L., Su, B. and Ma, R.Z., 2011. Extended Y chromosome investigation suggests postglacial migrations of modern humans into East Asia via the northern route. *Mol Biol Evol* 28 (1), 717-727.

Zhong, H., Shi, H., Qi, X.B., Xiao, C.J., Jin, L., Ma, R.Z., Su, B., 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.

Zhou, Weizhou 2004. The past and present of Northwest China's minority multiculture. *Xibei minzu luncong* [Studies of Northwestern minorities] 3, 1-45.