Correlation Between Genetic Structure and Linguistic Phylogeny in East Asia

Yunzhi Huang and Hui Li

Abbreviations

NRY   Non-recombining portion of the Y chromosome
bp   base pairs
mtDNA   mitochondrial DNA
SNP   Single nucleotide polymorphism
STR   Short tandem repeat

1   Introduction

East Asia is a quite separated region from the other parts of the world, with a vast landscape and diverse environments. The climate of East Asia is also unique as monsoon brings the moisture. That results in the unique physical characters of East Asians. The East Asians, with about 22% of the world population, can cursorily be classified into four physical styles, i.e., the Paleolithic migrants (Oceanians), the Azilian migrants (Negrito), the Neolithic migrants (major East Asians), and the Bronze Age migrants (Europeans). Many populations were mixed by some of these four waves of migrations. Since Neolithic Age, they evolved into several cultural confederates. Subsequently, each confederate might have evolved a linguistic family. Today, there are more than 1500 languages in ten linguistic families (Altaic,

Y. Huang (✉) · H. Li (✉)
MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences,
Fudan University, Shanghai 200438, People's Republic of China
e-mail: 13110700024@fudan.edu.cn

H. Li
e-mail: LiHui.Fudan@gmail.com
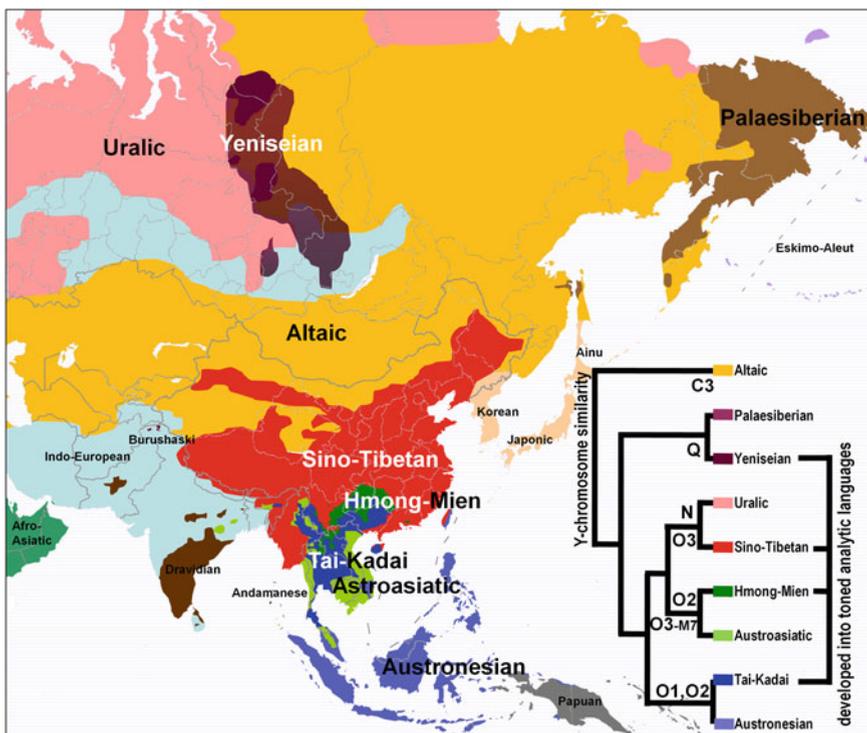
Fig. 1 Distribution of the linguistic families in East Asia and the approximate correlations between languages and Y chromosomes

Austroasiatic, Austronesian, Tai-Kadai, Hmong-Mien, Sino-Tibetan, Uralic, Palaesiberian, Yeniseian, and Indo-European) in East Asia and the flanking regions, which make this region one of the world's most important places for studying human evolution, genetic diversity, and interrelationships between genetics and cultures/languages of human populations (Cavalli-Sforza 1998) (Fig. 1).

2 Correlation Between Autosomal DNA and Linguistic Families

There are three types of genetic materials, i.e., autosomal and X chromosome DNA, paternal Y chromosome, and maternal mitochondrial DNA (mtDNA). Autosomes and X chromosome are inherited from both the parents and are always jumbled by recombination. The non-recombining portion of the Y chromosome (NRY) is

strictly inherited paternally, and therefore, is the best material to trace the paternal lineage of the populations with additional advantages of small effective population size, low mutation rate, sufficient markers, and population-specific haplotype distribution (Jobling and Tyler-Smith 1995; Underhill et al. 2000). As the linguistic families were founded in the Neolithic Age, since when people have mostly practiced traditions of patrilocal marriage, the paternal inherited Y chromosomes might be more associated to the linguistic classifications than other genetic materials. However, genetic investigations in the past decade have given more confidence for the relationships between linguistic families and genetic structures in East Asia. The whole genome diversity is also found to be well related to the languages.

Autosomal markers are not always associated to linguistic classifications. When using limited number of markers, the association is usually omitted. However, those markers which have undergone strong selection showed quite good clustering within linguistic families. For example, the *ADH* gene family, which is the most important genes for alcohol metabolism, has been selected for among many East Asian populations (Li et al. 2008a, b). The genetic diversity within *ADH* gene family is associated very well to the linguistic family (Fig. 2), even though only 30 SNP markers were included in the analysis. The Tai-Kadai populations were soundly separated from the Austro-Asiatic populations although they distributed in almost the same geographic region. Not only the genetic diversity, but also the
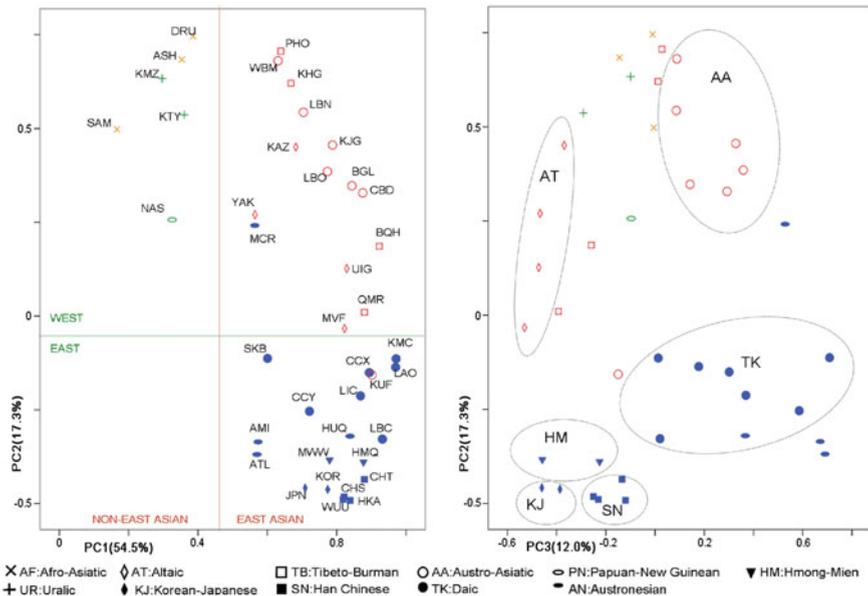
Fig. 2 Principle Analysis Plot based on 30 SNPs within ADH gene family showing clear linguistic clusters

selective signal, was associated with the linguistic families. Strong signals were detected among the Hmong-Mien and Altaic (Korean and Japanese) populations. Week signals were detected among the Han-Chinese populations. And no signal was found in other populations. This association might have resulted from the co-evolution of the *ADH* genes and linguistic families during the Neolithic and Bronze Ages. Further analyses of time estimation suggested that the ADH genes evolved rapidly in East Asia in the recent 5000 years.

When using large number of autosomal polymorphisms, the genetic structure will always be correlated to linguistic families very well among the East Asians. The best sample of such study was carried out by the HUGO Pan-Asian SNP Consortium (2009) since 2005. They analyzed 54,794 autosomal single-nucleotide polymorphisms (SNPs) in 1928 individuals representing 73 Asian and two non-Asian HapMap populations. This large-scale survey of autosomal variation from a broad geographic sample of East Asian human populations showed that genetic ancestry is strongly correlated with linguistic affiliations. Almost all populations from the same linguistic family clustered in the same clade on the genetic phylogeny (Fig. 3). Moreover, the genetic phylogeny showed a geographic association among families. That is, those families distributed in the south were more close to the root, which means that East Asians entered from the south.

The genetic relationships among the linguistic families can be judged from the Structure analysis (genetic components). In this study, 14 components were proved to be the best resolution. The continental populations in East Asia share a common component which can hardly be found in the Southeast Asian Islanders. Specifically, Koreans and Japonic populations share their major component with Sino-Tibetan. Austronesians share their major component with Tai-Kadai populations, which can also be found in Sino-Tibetan populations. Tai-Kadai also comprises a minor component which is major in Austro-Asiatic. Interestingly, the western Austronesians also comprises the Austro-Asiatic major component, while the eastern Austronesians does not. For the Tai-Kadai populations, the closer they reside to the Austro-Asiatic populations, the more Austro-Asiatic components they comprise. This structure suggests that this component originated in Austro-Asiatic populations. Among the Austronesian populations, there are some other exotic components. Those populations in eastern Indonesia where was once the area of Papuans comprise pronounced Papuan components. The Negrito people in Philippines also speak Austronesian while they have their own component besides the Austronesian major component. The Malaysian Negritos have different component from Philippine Negrito as well as a few Austro-Asiatic components. Some small Sino-Tibetan populations migrated to the south have been replaced in genetic components by Austro-Asiatic or Indo-European populations. Altogether, the linguistic families in East Asia show clear association with genetic clusters, and the language contacts are also well recorded by the genetic structure.
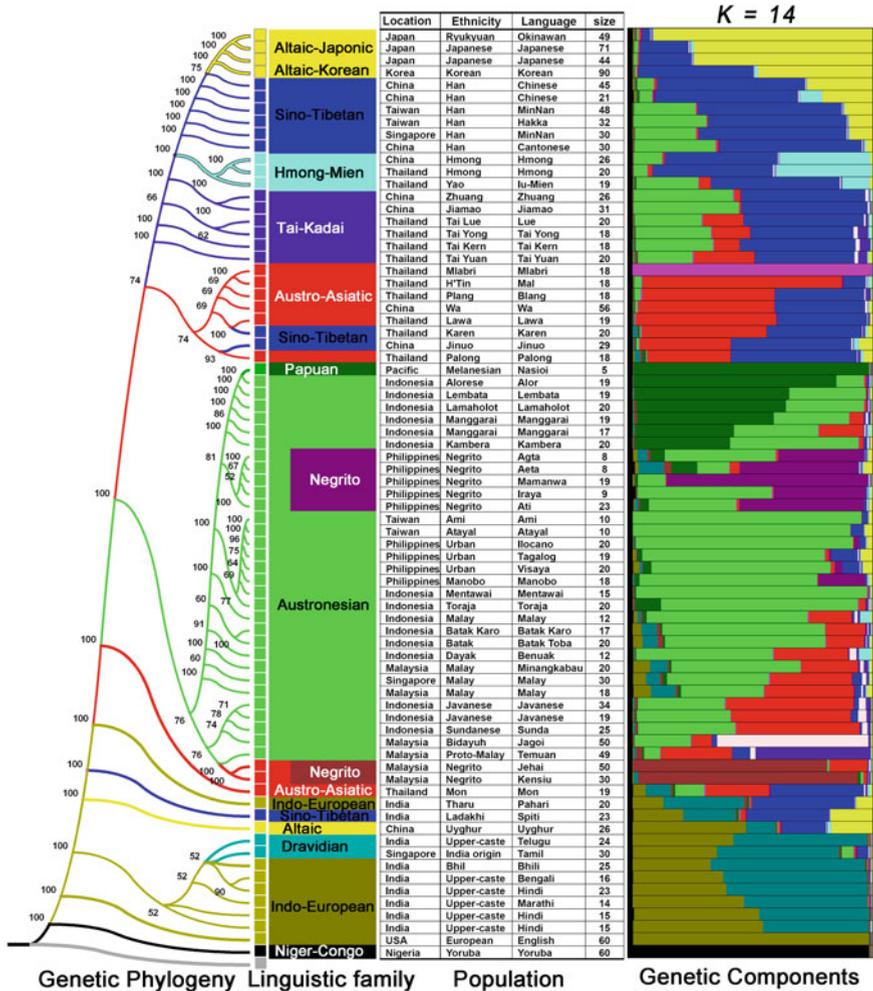
Fig. 3 Genetic structure among the linguistic families in East Asia based on 54,794 autosomal SNPs

3 Mitochondrial DNA or Y Chromosomal DNA

Mitochondrial DNA was first genetic material which came into the success application of molecular anthropology for its advantages of maternal inheritance and large copy number in the cell. The famous "African Eva" hypothesis was just suggested upon global mitochondrial DNA diversity (Cann et al. 1987). It is very clear that mtDNA shows pronounced difference among the geographic regions around world. Those mtDNA haplogroups in Africa are distinctively different from those in western Eurasia. Southern Eurasia and eastern Eurasia also have the unique

mtDNA haplogroups. However, mtDNA diversity does not show significant different among the linguistic families in the same region. In East Asia, the mtDNA diversity is clearly correlated to the climate (latitude or altitude). In a principal component analysis plot of East Asian mtDNA, all the arctic populations are in one end and the tropical populations are in the other end (Fig. 4). Populations of the same linguistic family are scattered in different parts of the plot (Qin et al. 2010). Therefore, we concluded that mtDNA is largely influenced by the climate. It is quite easy to understand that mitochondria are the energy resources of cells, and therefore, different mtDNA may have related to different efficiency of energy transformation, which can be the selective stress. In this case, when populations of the same linguistic family migrated to different climates, their mtDNA will changed to adapt
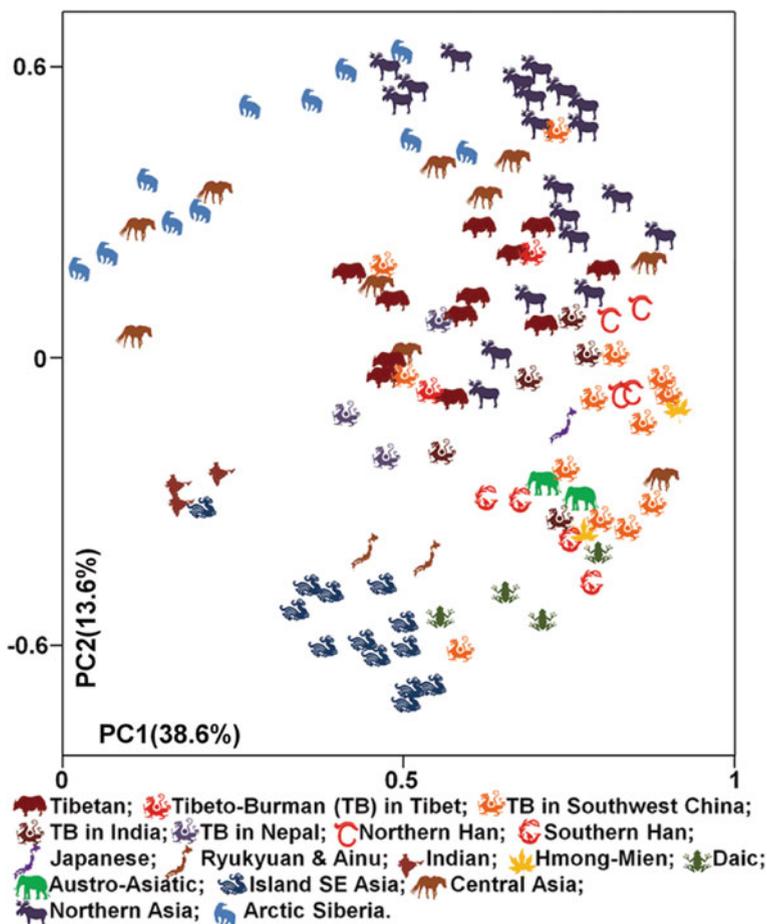
Fig. 4 Principal component analysis plot of the East Asians based on mtDNA diversity

the local climate. Moreover, mtDNA is inherited maternally. Women migrate more frequently between groups by marriages than men in the patrilocal style, which eliminates the possible genetic distance among the linguistic families. Thus, we will not expect the correlation between the linguistic family and mtDNA diversity. Different from mtDNA, the paternal Y chromosome is less frequently transmitted among populations, and therefore, is believed to be more correlated with language.

4 Y Chromosomes Suggested an African Origin for East Asians

Y chromosome studies on East Asians have been widely carried out and have answered many questions about the origin of East Asian populations. For example, the debate on the single or multiple origins of anatomically modern human has lasted for decades. In 1999, Su et al. (1999) used 19 stable and highly informative Y chromosome biallelic markers to assess the genetic structure of the paternal lineages in East Asia, and suggested that modern humans of African descent replaced the previous hominids living in East Asia. In 2001, Ke et al. (2001) examined 12,127 male individuals from 163 populations using three Y chromosome biallelic markers (DE-YAP+, F-M89T, and C-M130T). They found that all the individuals carried a mutation at one of the three sites-YAP, M89, and M130. These three mutations (DE-YAP+, F-M89T, and C-M130T) coalesced to another mutation-M168T (Fig. 5), which originated in Africa at around 64 thousand years ago (Underhill et al. 2000; Ke et al. 2001). Although there have been possible gene flows between archaic hominids and modern humans (Green et al. 2010; Reich et al. 2010; Wang et al. 2013a), it is apparent that the majority of modern human evolved recently in Africa, at least our Y chromosomes all came from Africa. The next question was how the early modern human arrived in East Asia.

Climate has played an important role in human migrations, especially the Last Glacial Period. Last Glacial Period refers to the most recent glacial period from approximately 110 to 10 thousand years ago, covering the Paleolithic and Mesolithic periods of human history (Shi et al. 1989). During this period, when the sea level was much lower than present, many of today's islands were joined to the continents, providing paths for modern human migrations. The maximum extent of glaciation (Last Glacial Maximum, LGM) was between 26.5 and 19–20 thousand years ago, when ice sheets were at their maximum extension and covered much of Asia, northern Europe, and North America (Jobling et al. 2004; Clark et al. 2009). As a consequence, the living space for human was probably very limited in the northern part of Asia. The ice sheets started to recede since 15 thousand years ago and the temperature also began to rise up. This period has really been a flourishing time for modern human migrations.

Here, we focused on the migration histories of East Asian populations achieved by studying Y chromosome and discussed the patterns, and microevolution during
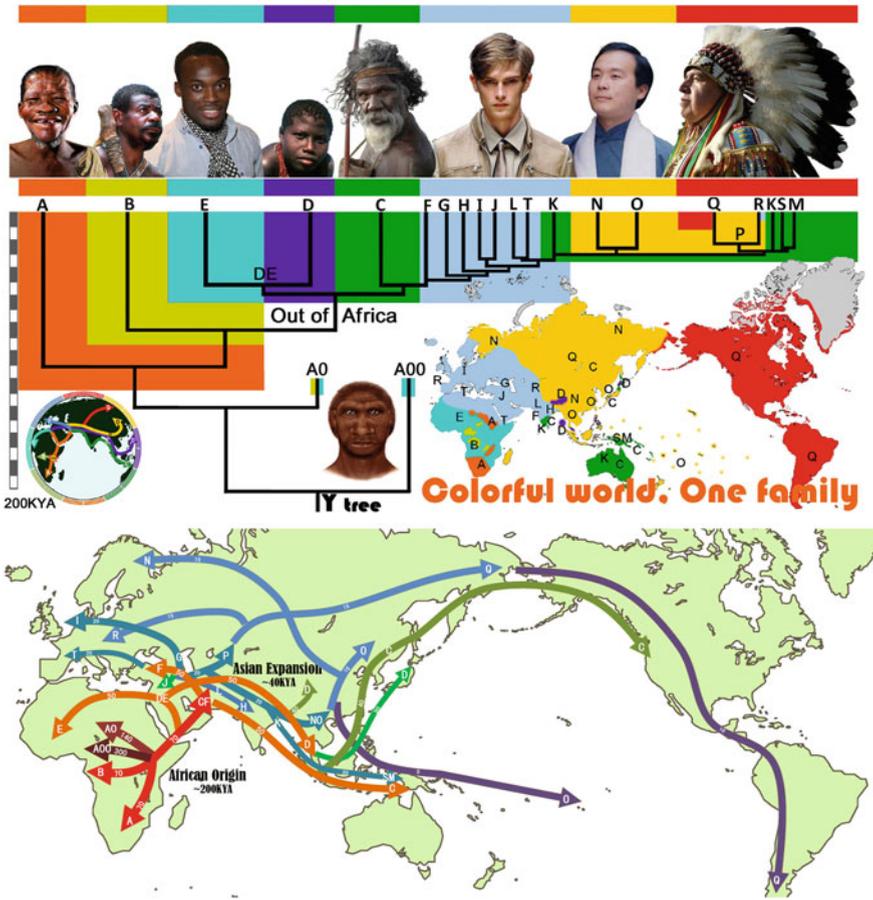
Fig. 5 Y chromosome phylogeny and its migration routes

the initial human settlement and later migrations and expansions in East Asia. It is worthy of note at the very beginning that most time estimations mentioned in this review were achieved using Y chromosome STRs. Although this approach is correct in principle, there is still many ongoing debates about the best way to use STRs in haplogroup dating. In particular, there are two popular used Y chromosome STR mutation rates, i.e., evolutionary rate (Zhivotovsky 2001; Zhivotovsky et al. 2004) and genealogical rate (Gusmão et al. 2005). To choose which mutation rate in the Y chromosome dating is controversial, as the result can be three times different. The high levels of homoplasy and varying mutational properties among loci also largely compromised the accuracy of estimation. Therefore, dates could be only intended as a rough guide for relative haplogroup ages.

5 Northern Route or Southern Route

Once it became generally accepted that modern humans evolved recently in Africa, the times and routes of migration to East Asia remained controversial. Three different models were insisted by different researchers. The first model postulated that northern populations of East Asia migrated to the south, and mixed with the Australian ancestors who had settled in Southeast Asia. The second model suggested that the northern populations of East Asia evolved from the southern settlers. However, a third model assumed that northern and southern East Asian populations have evolved independently since the late Pleistocene more than 10,000 years ago (Jobling et al. 2004; Zhong et al. 2011; Piazza 1998).

There are four dominant Y chromosome macro-haplogroups in East Asia, O-M175, C-M130, D-M174, and N-M231, accounting for about 93% of the East Asian Y chromosomes (Fig. 6). The other haplogroups, such as E-SRY4064,
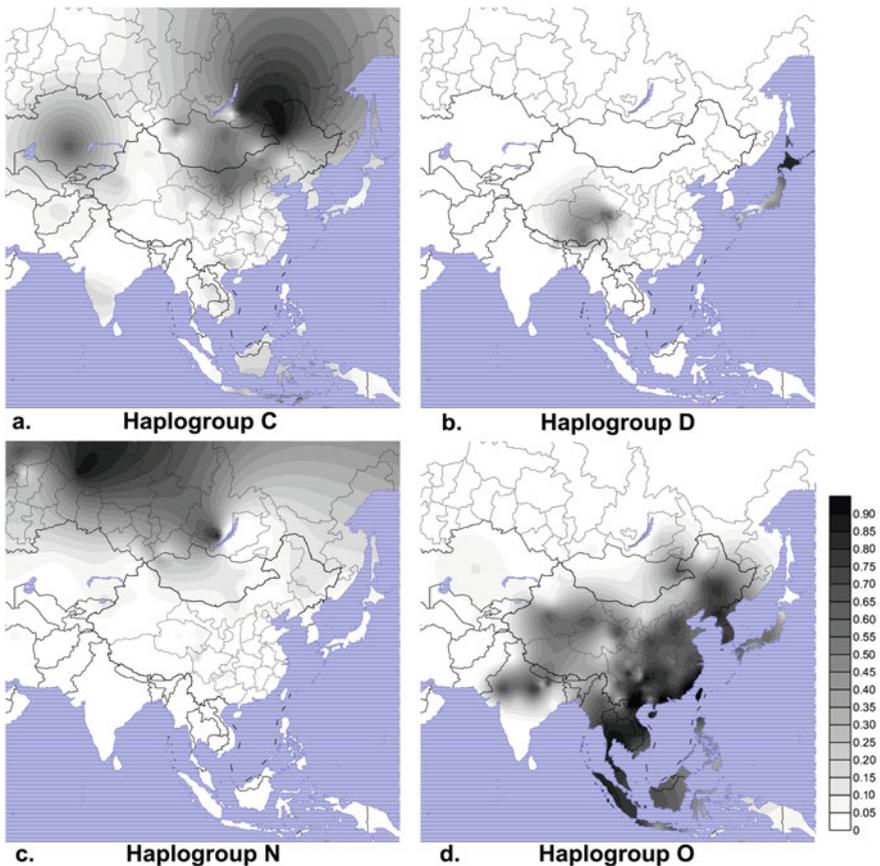
Fig. 6 Geographic distributions of Y chromosome haplogroup C, D, N, and O in East Asia

G-M201, H-M69, I-M170, J-P209, L-M20, Q-M242, R-M207, and T-M70, comprise roughly 7% of the males in East Asia (Zhong et al. 2011).

Haplogroup O-M175 is the largest haplogroup in East Asia, comprising roughly 75% of the Chinese and more than half of the Japanese population and therefore is associated to the Neolithic migrants (Fig. 6). O-M175 gave rise to three downstream haplogroups, O1a-M119, O2-M268, and O3-M122, totaling 60% of the Y chromosomes among East Asian populations (Shi et al. 2005; Yan et al. 2011). Haplogroup O1a-M119 is prevalent along the southeast coast of China, occurring at high frequencies in Tai-Kadai speaking people and Taiwan aborigines (Kayser et al. 2008). O2-M268 accounts for about 5% of the Han Chinese (Yan et al. 2011). O2a1-M95 is the most frequent subclade of O2, which is major haplogroups in the Indo-China Peninsula, and is also found in many populations located in southern China and eastern India (such as Munda) (Kayser et al. 2008; Su et al. 2000). Another subclade of O2, O2b-M176, is most frequent among Koreans and Japanese, and also occurs at very marginal frequencies in Vietnamese and Han Chinese (Ding et al. 2011; Hammer et al. 2006). O3-M122 is the most common haplogroup in China and prevalent throughout East and Southeast Asia, comprising roughly 50–60% of the Han Chinese. O3a1c-002611, O3a2c1-M134, and O3a2c1a-M117 are three main subclades of O3, each accounting for 12–17% of the Han Chinese. O3a2c1a-M117 also exhibits high frequencies in Tibeto-Burman populations. Another subclade O3a2b-M7 reaches the highest frequency in Hmong-Mien and Mon-Khmer speaking populations, but accounts for less than 5% of Han Chinese (Shi et al. 2005; Yan et al. 2011).

Su et al. examined 19 Y-SNPs (including M119, M95, and M122) and three Y chromosome short tandem repeats (STRs) in a large collection of population samples from a wide area of Asia. Principal component analysis of their study showed that all northern populations clustered together and were well included in the southern population cluster, and the southern populations were far more diversified than the northern populations. They concluded that the northern populations derived from the southern populations after the initial Palaeolithic peopling of East Asia. They also estimated the age of O3-M122 to be 18–60 thousand years using three Y-STRs under the single-step mutation model with a mutation rate of 0.18% per locus every 20 years, which might reflect the age of the bottleneck event leading to the initial settlement of East Asia (Su et al. 1999). In 2005, Shi et al. (2005) presented a systematic sampling and genetic screening of haplogroup O3-M122 in more than 2000 individuals from diverse populations in East Asia. Their data showed that the O3-M122 haplogroups in southern East Asia are more diverse than those in northern East Asia, supporting a southern origin of the O3-M122. The time of the early northward migration of O3-M122 lineages in East Asia was estimated about 25–30 thousand years ago using the average squared difference (ASD) method with an average Y-STR evolutionary mutation rate of 0.00069 per locus per 25 years (Zhivotovsky 2001; Zhivotovsky et al. 2004). Recently, Cai et al. (2011) examined the haplogroup O3a2b-M7 and O3a2c1a-M117 in Southeast Asian Mon-Khmer and Hmong-Mien speaking populations, and indicated a unidirectional diffusion through bottlenecks from
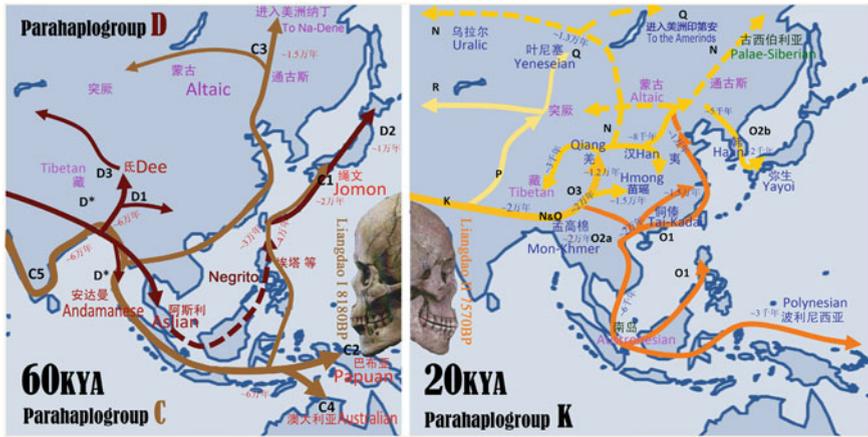
Fig. 7 Migration of the Y chromosome haplogroup C, D, N, and O in East Asia. Broken lines represent for alternative migration routes

Southeast Asia into East Asia about 19,000 years ago (also using the ASD method with an average Y-STR evolutionary mutation rate of 0.00069 per locus per 25 years) during the Last Glacial Maximum. A general south-to-north Y-STR diversity decline was also observed in haplogroup O3a1c-002611, suggesting that haplogroup O3a1c also migrated northward along with other O3-M122 lineages (Wang et al. 2013b). Therefore, the southern route of the early human migration in East Asia, taking the largest Y haplogroups O, is supported by most evidences (Fig. 7).

6   Earliest Settlement in East Asia

The age of haplogroup O in East Asia is no more than 30 thousand years when estimated from sufficient numbers (>7) of STR markers. Therefore, haplogroup O was not the earliest Y chromosome carried by modern human into East Asia. Haplogroup C-M130 may represent one of the earliest settlements in East Asia. Haplogroup C has a high to moderate frequency in Far East and Oceania, and lower frequency in Europe and the Americas, but is absent in Africa (Fig. 6). Several geographically specific subclades of haplogroup C have been identified, i.e., C1-M8, C2-M38, C3-M217, C4-M347, C5-M356, and C6-P55 (Zhong et al. 2010). Haplogroup C3-M217 is the most widespread subclade, and reaches the highest frequencies among the populations of Mongolia and Siberia. Haplogroup C1-M8 is absolutely restricted to the Japanese and Ryukyuans, appearing at a low frequency of about 5% or less. Haplogroup C2-M38 is found among certain local populations on Pacific Islands from eastern Indonesia to Polynesia. Especially among the populations of Polynesia, C2 has become the modal haplogroup due to severe

founder effects and genetic drift (Hammer et al. 2006; Kayser et al. 2006). Haplogroup C4-M347 is the most common haplogroup among Australian aborigines, and has not been found outside of the Australian continent. Haplogroup C5-M356 has been detected with low frequency in India and the neighboring regions in Pakistan and Nepal (Sengupta et al. 2006; Gayden et al. 2007). C6-P55 is geographically restricted to the highland of New Guinea (P55 has been moved to private in the latest Y chromosome tree) (Karafet et al. 2008). This wide distribution pattern of C-M130 suggests that C-M130 might arise somewhere in mainland of Asia before the modern humans arrived in Southeast Asia.

To give a clear picture about the origin and migration of haplogroup C, Zhong et al. typed twelve Y-SNPs and eight Y-STRs among 465 haplogroup C individuals from 140 East and Southeast Asian populations. A general south-to-north and east-to-west decline of C3 Y-STR diversity was observed with the highest diversity in Southeast Asia, which supports a single coastal northward expansion route of haplogroup C3 in China about 32–42 thousand years ago using the ASD method with an average Y-STR evolutionary mutation rate of 0.00069 per locus per 25 years (Zhong et al. 2010) (Fig. 7). The arrival of haplogroup C in Southeast Asia and Australia must be much earlier than that time at around 60 thousand years ago. Therefore, populations with haplogroup C must have settled in East Asia some ten thousand years earlier than those with haplogroup O.

7 Genetic Legacy of the Paleolithic Period

The migration history of haplogroup D-M174 is most mysterious. By the time, we have known little about the origin and dispersal of this haplogroup. This haplogroup was derived from African haplogroup DE-M1 (YAP insertion) and is associated with a short black Asian physical style. Haplogroups E and D are brother haplogroups. While haplogroup E was carried westwards to the Africa by the tall black people, haplogroup D might have carried eastwards to East Asia by the short black people (Fig. 5).

Haplogroup D-M174 has high frequencies in the Andaman Negritos, the northern Tibeto-Burman populations, and the Ainu of Japan, and also appears at low frequencies in other East and Southeast Asian and Central Asian populations (Fig. 6) (Su et al. 2000; Hammer et al. 2006; Karafet et al. 2001; Thangaraj et al. 2003). A northern Tibeto-Burman population, the Baima-Dee, comprises nearly 100% of haplogroup D. There are three main subclades of haplogroup D, i.e., D1-M15, D2-M55, and D3-P99, and many unclassified minor sub-haplogroups. Haplogroup D1-M15 is prevalent in the Tibetans, Tangut-Chiang, and Lolo, and also found at very low frequencies among the mainland East Asian populations (Shi et al. 2008; Wen et al. 2004a, b). Haplogroup D2-M55 is restricted to various populations of the Japanese Archipelago. Haplogroup D3-P99 is found at high frequencies among Tibetans and several Tibeto-Burman minorities in Sichuan and Yunnan provinces that reside in close proximity to the Tibetans, such as Pumi and Naxi (Shi et al.

2008). The paragroup D* is restricted to Andaman Islands (Thangaraj et al. 2003), which has been isolated for at least 20 thousand years. Some other minor haplogroups, also included in D*, can be found around Tibet. Most of the populations with haplogroup D have very dark skin color, including the Andamanese, some of the Tibeto-Burman and Mon-Khmer people. The Ainu people may have changed to pale skin to absorb more ultraviolet light in high latitude region.

For the origin of haplogroup D, Chandrasekar et al. (2007) suggested that the CT-M168 gave rise to the YAP insertion and D-M174 mutation in South Asia based their findings of the YAP insertion in northeast Indian tribes and the D-M174 in Andaman islanders. In that case, haplogroup E with YAP insertion might also have an Asian origin. However, this hypothesis is seldom supported by any evidence. If haplogroup D originated in Africa, it is most mysterious how it has traveled through the populations with haplogroups CF to East Asia.

Another mystery is how haplogroup D has migrated from southwest of East Asia all the way to Japan. It could have gone either through mainland East Asian or through Sundaland (Fig. 7). The mainland route seems to be shorter than the Sundaland route. Shi et al. proposed that the northward expansion of D-M174 to western China might predate the migrations of other major East Asian lineages at about 60 thousand years ago using ASD time estimation method with an average Y-STR evolutionary mutation rate of 0.00069 per locus per 25 years. Subsequently, these frontier populations could have traveled eastwards through a northern route via Korea or through a southern route via Taiwan and Ryukyu land bridge to Japan, where they might meet the earlier Australian style settlers. The current relic D-M174 in East Asia was probably edged out of eastern China by the later northward migration of haplogroup O and the Neolithic expansion of Han Chinese (Shi et al. 2008). However, there has never been any evidence from genetics or archaeology that haplogroup D2 or Negritos have migrated to eastern China. In contrast, there are still many Negrito populations in Sundaland from Malaya to the Philippines. It was possible that Negritos have occupied the whole Sundaland in late Paleolithic Age. Therefore, these populations might move directly from the Philippines to Taiwan and Ryukyu. The only problem is that no haplogroup D has been found in the Negritos in the Philippines. Their paternal lineages might have been replaced by the expansion of haplogroups C2 and K from Papua around 18 thousand years ago using BATWING time estimation method (Delfin et al. 2011) or much recent migration of haplogroup O from mainland East Asia (Scholes et al. 2011). However, due to the lack of data, the history of haplogroup D, as a genetic legacy of Azilian Age in East Asia, remains mysterious.

8 Recent Immigrations of Uralic and Altaic Populations

Haplogroup O has a brother haplogroup, N-M231, which reaches highest frequency in northern Eurasia, especially among most of the Uralic populations, including Finnic, Ugric, Samoyedic, and Yukaghir people, as well as some Altaic and Eskimo

populations in northern Siberia. It also appears at a low frequency in East Asia (Fig. 6) (Karafet et al. 2001; Rootsi et al. 2007). Detailed analysis of haplogroup N suggested a more recent expansion on a counter-clock northern route from inner East Asia or southern Siberia about 12–14 thousand years ago using ASD time estimation method with an average Y-STR evolutionary mutation rate of 0.00069 per locus per 25 years, which explained high frequency of haplogroup N in northeast Europe (Rootsi et al. 2007). Subclade N1a-M128 is found at low frequency among populations in northern China, such as Manchu, Xibe, Evenks, Korean, and also among some Turkic populations in Central Asia. Haplogroup N1b-P43 is approximately 6–8 thousand years old using ASD time estimation method with an average Y-STR evolutionary mutation rate of 0.00069 per locus per 25 years and probably originated in Siberia. N1b is prevalent in the Northern Samoyeds, and also occurs at low to moderate frequencies among some other Uralic and Altaic peoples (Derenko et al. 2007; Mirabal et al. 2009). The most frequent subclade N1c-Tat arose probably in China around 14 thousand years ago (ASD method with a mutation rate of 0.00069 per locus per 25 years) and subsequently experienced a serial of founder effects or strong bottlenecks in Siberia and a secondary expansion in East Europe (Rootsi et al. 2007). These studies traced the origin of haplogroup N to southwestern China and Southeast Asia. Our recent studies revealed that haplogroup N reaches the pronounced high diversity in Han Chinese populations, which suggested a Han Chinese origin of haplogroup N. Those proportions in North Asia and Southeast Asia are two distinctive clades out of Han Chinese. In this case, most paternal lineages of Uralic speakers originated in the ancestors of Han Chinese.

The migration of haplogroup N is another evidence for the southern origin of the East Asians. However, there were still studies against the southern origin of East Asians. Karafet et al. examined 52 Y-SNPs in 1383 individuals of 25 populations from East Asia and Central Asia. They found the average pairwise difference among haplogroups was noticeably smaller in southern East Asia and there was no genetic divergence between southern and northern East Asia (Karafet et al. 2001). Xue et al. (2009) applied a Bayesian full-likelihood analysis to 45 Y-SNPs and 16 Y-STRs data from 988 men of 27 populations from China, Mongolia, Korea, and Japan. They reported the Y-STRs have a higher diversity in northern East Asian populations than that in southern populations. The northern populations expanded earlier than the southern populations. However, Shi et al. (2008) pointed out that the larger diversity among Y-chromosome haplogroups observed in northern East Asia claimed by Karafet et al. (2001) is actually a false impression due to recent population admixture. The study of Xue et al. (2009) has the similar drawback. The high gene diversity observed in Mongols, Uighurs, and Manchurians was probably due to their recent extensive admixture with Central Asian, West Eurasian, and Han Chinese populations. Furthermore, the southern populations studied by Xue et al. were not sufficient and the within-population bottleneck effect caused by long-time geographic isolation might have a great impact on gene diversity estimation.

The subsequent debate focuses on how to interpret the Central Asia and West Eurasia related genetic components in East Asia. Zhong et al. (2011) sampled 3,826

males from 117 populations and performed high-resolution genotyping to address this problem. In the study of Zhong et al. haplogroups O-M175, C-M130, D-M174, and N-M231 still suggest the substantial contribution of the southern route. However, the Central Asia and West Eurasia related haplogroups, such as haplogroups R-M207 and Q-M242, occur primarily in northwestern East Asia and their frequencies gradually decrease from west to east. In addition, the Y-STR diversities of haplogroups R-M207 and Q-M242 also indicate the existence of northern route migration about 18,000 years ago (ASD method with a mutation rate of 0.00069 per locus per 25 years) from Central Asia to North Asia, and recent population admixture along the Silk Road since about 3000 years ago (Piazza 1998).

The current populations speaking the languages within the Altaic family show great genetic incoherence. Although Mongolic and Tungus speakers are similar with high frequencies of haplogroup C, Turkic speakers are not the same but admixed populations with various genetic components. Haplogroup R, C, N, Q, O, J, etc. can all be detected in Turkic speaking populations with fluctuant frequencies (Fig. 8). Therefore, genetic evidences do not support the affiliation of Turkic in Altaic family. We neither found any evidence supporting the Xiongnu ancestral for Turkic or Mongol-Tungus populations. Most of the Y haplogroups of the ancient Xiongnu samples are Q (Huang and Li 2015), while few haplogroup Q can be found in the present Turkic or Mongol-Tungus populations. On another hand, high frequencies of haplogroup Q were found among the Yeniseian and Palaesiberian speaking populations, suggesting that these two families are more likely closely related to Xiongnu in paternal lineages.
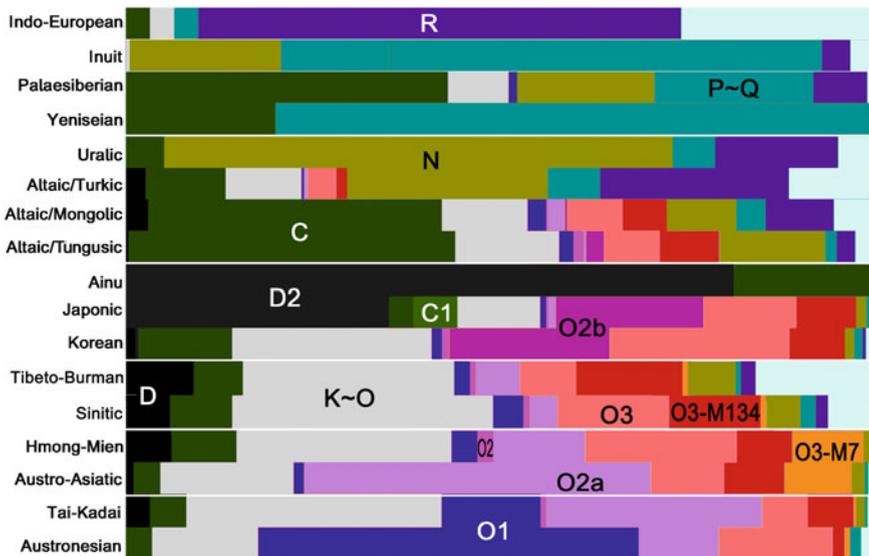
Fig. 8 Y chromosome haplogroup frequencies of the East Asian populations speaking languages of certain families

9   Genetic Coherences for Linguistic Families

The genetic patterns in human societies are often influenced by their cultural practices, such as residence patterns and subsistence strategies. Y chromosomes of East Asian populations have played an important role in documenting such influences, e.g., relationships among patrilocal populations should have stronger association with Y chromosomes than with mtDNA. East Asian languages show strong association with paternal lineages of Y chromosomes (Shi et al. 2005; Cai et al. 2011; Zhong et al. 2010; Wen et al. 2004a; Kumar et al. 2007) and whole genomic diversity (The HUGO Pan-Asian SNP Consortium 2009), but not maternal lineages of mtDNA. Y chromosome haplogroup O3-M134 is associated with Sino-Tibetan speakers (Shi et al. 2005; Wen et al. 2004a); O2-M95 is with Austro-Asiatic speakers (Kumar et al. 2007). Moreover, phylogenetic structure among the linguistic families suggested by the linguistics is also supported by Y chromosomes but not by whole genomic diversity, e.g., linguistic affinity between Hmong-Mien and Austro-Asiatic languages was proved by Y chromosome marker O3-M7 (Cai et al. 2011); and that between Tai-Kadai and Austronesian languages by O1-M119 (Li et al. 2008a, b).

Most of the Y-chromosome lineages of East Asians came from Southeast Asia. In that case, the indigenous populations of Southeast Asia, Austro-Asiatic (Mon-Khmer), should be the oldest population in East Asia. However, very few Mon-Khmer populations had been investigated, and therefore, little was known about the purported migrations from Southeast Asia into East Asia and their roles in shaping the genetic structure of East Asian populations. We analyzed the Y-chromosome data from 1652 individuals belonging to 47 Mon-Khmer and Hmong-Mien speaking populations that are distributed primarily across Southeast Asia and extend into East Asia (Cai et al. 2011). Haplogroup O3a3b-M7, which appears mainly in Mon-Khmer and Hmong-Mien, indicates a strong tie between the two groups. The short tandem repeat network of O3a3b-M7 displayed a hierarchical expansion structure (annual ring shape), with Mon-Khmer haplotypes being located at the original point, and the Hmong-Mien and the Tibeto-Burman haplotypes distributed further away from core of the network. Moreover, the East Asian dominant haplogroup O3a3c1-M117 shows a network structure similar to that of O3a3b-M7. These patterns indicate an early unidirectional diffusion from Southeast Asia into East Asia, which might have resulted from the genetic drift of East Asian ancestors carrying these two haplogroups through many small bottle-necks formed by the complicated landscape between Southeast Asia and East Asia. The ages of O3a3b-M7 and O3a3c1-M117 were estimated to be approximately 19 thousand years, followed by the emergence of the ancestors of Hmong-Mien lineages out of Mon-Khmer and the unidirectional northward migrations into East Asia.

Austronesian is a linguistic family spread in most areas of the Southeast Asia, the Pacific Ocean, and the Indian Ocean. Based on their linguistic similarity, this linguistic family included Malayo-Polynesians and Taiwan aborigines. The linguistic similarity also led to the controversial hypothesis that Taiwan is the

homeland of all the Malayo-Polynesians, a hypothesis that has been debated by ethnologists, linguists, archaeologists, and geneticists. It is well accepted that the Eastern Austronesians (Micronesians and Polynesians) derived from the Western Austronesians (Island Southeast Asians and Taiwanese), and that the Tai-Kadai populations on the mainland are supposed to be the headstream of all the Austronesian populations. To assess the relationships between the three ethnic groups, we studied 20 SNPs and 7 STRs in the non-recombining region of the 1509 Y chromosomes from 30 China Tai-Kadai populations, 23 Indonesian and Vietnam Malayo-Polynesian populations, and 11 Taiwan aboriginal populations (Li et al. 2008a, b). These three groups show many resemblances in paternal lineages. Admixture analyses demonstrated that the Tai-Kadai populations are hardly influenced by Han Chinese genetically, and that they make up the largest proportion of Indonesians. Most of the population samples contain a high frequency of haplogroup O1a-M119, which is nearly absent in other ethnic families. The STR network of haplogroup O1a* illustrated that Indonesian lineages did not derive from Taiwan aborigines as linguistic studies suggest, but from Tai-Kadai populations. We show that, in contrast to the Taiwan homeland hypothesis, the Island Southeast Asians do not have a Taiwan origin based on their paternal lineages. Furthermore, we show that both Taiwan aborigines and Indonesians likely derived from the Tai-Kadai populations based on their paternal lineages. These two populations seem to have evolved independently of each other. Our results indicate that a super-phylum, which includes Taiwan aborigines, Tai-Kadai, and Malayo-Polynesians, is genetically educible.

The Sino-Tibetan populations comprise the major population of East Asia. The geographic distribution of this group is very vast. Judging from languages and physical features, both subgroups, Han Chinese and Tibeto-Burman, exhibit pronounced differences between north and south. It has been argued whether southern Han and northern Han have the same origin and in turn the same genetic pattern. The same question was also existed for Tibeto-Burman. The spread of culture and language in human populations is explained by two alternative models: the demic diffusion model, which involves mass movement of people; and the cultural diffusion model, which refers to cultural impact between populations and involves limited genetic exchange between them. The mechanism of the peopling of Europe has long been debated, a key issue being whether the diffusion of agriculture and language from the Near East was concomitant with a large movement of farmers. By systematically analysing Y-chromosome and mitochondrial DNA variation in Han populations, we found that the pattern of the southward expansion of Han culture is consistent with the demic diffusion model, and that males played a larger role than females in this expansion (Wen et al. 2004a) (Fig. 9).

Among the ten Chinese linguistic branches, the Pinghua branch in Guangxi is genetically different. We have studied eight Pinghua populations and other ethnic groups (Zhuang, Kam, Mulam, Laka, and Mien) from Guangxi Zhuang Autonomous Region of China. Both mtDNA and the Y chromosomes were typed in these samples (Gan et al. 2008; Lu et al. 2013). High frequencies of the Y chromosome haplogroups O2a* and O*, which always present at a high frequency
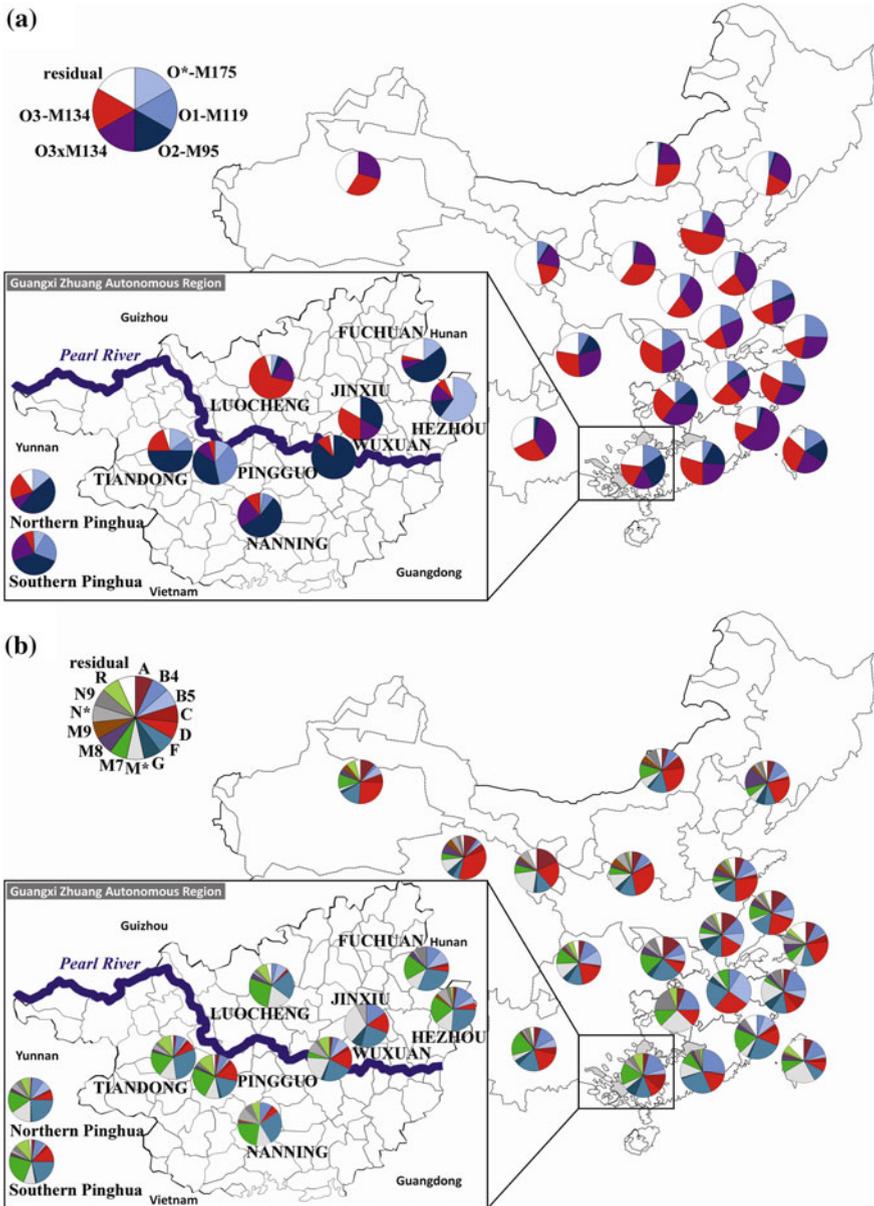
Fig. 9 Y chromosomal and mitochondrial diversity shows that Pinghua is an exception of the genetic coherence of Han Chinese

among the populations of the southern minorities, were found in Pinghua populations. Only Pinghua populations in Luocheng and Jinxiu maintain the Han frequent haplogroup O3a5a. MtDNA lineages B4a, B5a, M*, F1a, M7b1, and N* were found in Pinghua populations, exhibiting a pattern similar to the neighboring indigenous populations, especially the Tai-Kadai populations. Cluster analyses of Pinghua populations, the other Han branches, and other ethnic groups in East Asia indicated that Pinghua populations are much closer to the southern minorities than to the other Han branches. Admixture analyses confirmed this result. Therefore, we argue that Pinghua populations did not descend from Han Chinese, but from southern minorities. The ancestral populations of Pinghua people were assimilated by the Han Chinese in terms of language, culture, and self-identification and, consequently, the Pinghua people became an exceptional branch of Han Chinese's coherent genetic structure.

Tibeto-Burman populations were historically derived from ancient tribes of northwestern China and subsequently moved to the south, where they admixed with the southern natives during the past 2600 years. They are currently extensively distributed in China and Southeast Asia. We analyze the variations of 965 Y chromosomes and 754 mtDNAs in >20 Tibeto-Burman populations from China (Wen et al. 2004b). By examining the haplotype group distributions of Y-chromosome and mtDNA markers and their principal components, we show that the genetic structure of the extant southern Tibeto-Burman populations were primarily formed by two parental groups: northern immigrants and native southerners. Furthermore, the admixture has a bias between male and female lineages, with a stronger influence of northern immigrants on the male lineages (approximately 62%) and with the southern natives contributing more extensively to the female lineages (approximately 56%) in the extant southern Tibeto-Burman populations. This is the first genetic evidence revealing sex-biased admixture in southern Tibeto-Burman populations, which has genetic, historical, and anthropological implications.

The association between languages and Y chromosomes but not mtDNA might reflect sex-bias migrations due to patrilocality. Patrilocality refers to the social system that a married couple resides with or near the husband's parents. Forster et al. suggested that it may often be the language of the father that is dominant within the family group if the parents have different linguistic backgrounds (Forster and Renfrew 2011). However, as the whole genomic diversity is also associated with linguistic families, both paternal and maternal lineages must have been well kept since the linguistic families emerged. Therefore, the loss of association between mtDNA and languages might not simply be explained by a social nature of women adoption. There might be a higher effective population size in ancient population for females than for males due to frequent hunting activities and wars, and thus, the original mtDNA variation of a language group was less affected by genetic drift. Other interpretations might also be possible, such as preferential males, the number of offspring, and the different mutation rates, etc.

10 Conclusions and Perspectives

Y chromosome plays an important role in unraveling the entangled history of modern human populations in East Asia. Although many questions remain unresolved, a clear framework of the prehistory has been obtained. Four Y chromosome haplogroups C, D, O, and N, accounted for more than 90% of the East Asian Y chromosomes, are suggested to have Southeast Asian origins, carried by three waves of migrations. The distributions of western Eurasia specific Y chromosome haplogroups E, G, H, I, J, L, Q, R, and T in northwest China reflect the recent gene flows from the west and the probable northern route migration. A west-to-east decline of these western haplogroups was clearly observed. Linguistic families are well associated with the Y chromosome diversity.

However, current Y chromosome researches in East Asia are limited in two important aspects. The first limit is the poor resolution for those East Asian specific Y chromosome branches, such as haplogroup O-M175. Despite the huge population of haplogroup O, there have been much less markers defined in haplogroup O than in haplogroups R and E. For instance, three Y-SNP markers, 002611, M134, and M117, represent about 260 million people in East Asia, but downstream markers are far from enough to reveal informative genetic substructures of those populations. The second limit is inaccurate estimation of lineage and population divergence time as mentioned at the beginning.

The advent of next-generation sequencing technology made it possible to sequence the entire Y chromosome in numerous human individuals and in deep-rooting pedigrees. For instance, The 1000 Genomes Project Consortium has already sequenced the Y chromosomes at an average depth of 1.83 in 77 males in the low-coverage project, and 15.23 depth in the two trio fathers (1000 Genomes Project Consortium 2010). Further deep sequencing will offer a solution for both enhanced Y chromosome phylogenetic resolution and accurate calibration of the molecular clock in evolutionary studies.

References

1000 Genomes Project Consortium. 2010. A map of human genome variation from population-scale sequencing. *Nature* 467: 1061–1073.

Cai, X., Z. Qin, B. Wen, S. Xu, Y. Wang, Y. Lu, L. Wei, C. Wang, S. Li, X. Huang, L. Jin, and H. Li. 2011. Genographic Consortium: Human migration through bottlenecks from Southeast Asia into East Asia during last glacial maximum revealed by Y chromosomes. *PLoS ONE* 6 (8): e24282.

Cann, R.L., M. Stoneking, and A.C. Wilson. 1987. Mitochondrial DNA and human evolution. *Nature* 325 (6099): 31–36.

Cavalli-Sforza, L.L. 1998. The Chinese human genome diversity project. *Proceedings of the National Academy of Sciences of the United States of America* 95: 11501–11503.

Chandrasekar, A., S.Y. Saheb, P. Gangopadyaya, S. Gangopadyaya, A. Mukherjee, D. Basu, G.R. Lakshmi, A.K. Sahani, B. Das, S. Battacharya, S. Kumar, D. Xaviour, D. Sun, and V.R. Rao. 2007. YAP insertion signature in South Asia. *Annals of Human Biology* 34: 582–586.

Clark, P.U., A.S. Dyke, J.D. Shakun, A.E. Carlson, J. Clark, B. Wohlfarth, J.X. Mitrovica, S.W. Hostetler, and A.M. McCabe. 2009. The last glacial maximum. *Science* 325: 710–714.

Delfin, F., J.M. Salvador, G.C. Calacal, H.B. Perdigon, K.A. Tabbada, L.P. Villamor, S.C. Halos, E. Gunnarsdóttir, S. Myles, D.A. Hughes, S. Xu, L. Jin, O. Lao, M. Kayser, M.E. Hurles, M. Stoneking, and M.C. De Ungria. 2011. The Y-chromosome landscape of the Philippines: Extensive heterogeneity and varying genetic affinities of Negrito and non-Negrito groups. *European Journal of Human Genetics* 19: 224–230.

Derenko, M., B. Malyarchuk, G. Denisova, M. Wozniak, T. Grzybowski, I. Dambueva, and I. Zakharov. 2007. Y-chromosome haplogroup N dispersals from south Siberia to Europe. *Journal of Human Genetics* 52 (9): 763–770.

Ding, Q.L., C.C. Wang, S.E. Farina, and H. Li. 2011. Mapping human genetic diversity on the Japanese archipelago. *Advances in Anthropology* 1 (2): 19–25.

Forster, P., and C. Renfrew. 2011. Mother tongue and Y chromosomes. *Science* 333: 1390–1391.

Gan, R.J., S.L. Pan, L.F. Mustavich, Z.D. Qin, X.Y. Cai, J. Qian, C.W. Liu, J.H. Peng, S.L. Li, J.S. Xu, L. Jin, and H. Li. 2008. Genographic Consortium: Pinghua population as an exception of Han Chinese's coherent genetic structure. *Journal of Human Genetics* 53: 303–313.

Gayden, T., A.M. Cadenas, M. Regueiro, N.B. Singh, L.A. Zhivotovsky, P.A. Underhill, L.L. Cavalli-Sforza, and R.J. Herrera. 2007. The Himalayas as a directional barrier to gene flow. *American Journal of Human Genetics* 80: 884–894.

Green, R.E., J. Krause, A.W. Briggs, T. Maricic, U. Stenzel, M. Kircher, Nick Patterson, H. Li, W. W. Zhai, M.H.Y. Fritz, N.F. Hansen, E.Y. Durand, A.S. Malaspinas, J.D. Jensen, T. Marques-Bonet, C. Alkan, Kay Prüfer, M. Meyer, H.A. Burbano, J.M. Good, R. Schultz, A. Aximu-Petri, A. Butthof, B. Hber, B. Hffner, M. Siegemund, A. Weihmann, C. Nusbaum, E.S. Lander, C. Russ, N. Novod, J. Affourtit, M. Egholm, C. Verna, P. Rudan, Dejana Brajkovic, Ž. Kucan, I. Gušic, V.B. Doronichev, L.V. Golovanova, C. Lalueza-Fox, M. Rasilla, J. Fortea, A. Rosas, R.W. Schmitz, P.L.F. Johnson, E.E. Eichler, D. Falush, E. Birney, J.C. Mullikin, M. Slatkin, R. Nielsen, J. Kelso, M. Lachmann, D. Reich, and S. Pääbo. 2010. A draft sequence of the Neandertal genome. *Science* 328: 710–722.

Gusmão, L., P. Sánchez-Diz, F. Calafell, P. Martín, C.A. Alonso, F. Alvarez-Fernández, C. Alves, L. Borjas-Fajardo, W.R. Bozzo, M.L. Bravo, J.J. Builes, J. Capilla, M. Carvalho, C. Castillo, C.I. Catanesi, D. Corach, Lonardo A.M. Di, R. Espinheira, E. Fagundes de Carvalho, M. J. Farfán, H.P. Figueiredo, I. Gomes, M.M. Lojo, M. Marino, M.F. Pinheiro, M.L. Pontes, V. Prieto, E. Ramos-Luis, J.A. Riancho, A.C. Souza Góes, O.A. Santapa, D.R. Sumita, G. Vallejo, L. Vidal Rioja, M.C. Vide, C.I. Vieira da Silva, M.R. Whittle, W. Zabala, M.T. Zarrabeitia, A. Alonso, A. Carracedo, and A. Amorim. 2005. Mutation rates at Y chromosome specific microsatellites. *Human Mutation* 26: 520–528.

Hammer, M.F., T.M. Karafet, H. Park, K. Omoto, S. Harihara, M. Stoneking, and S. Horai. 2006. Dual origins of the Japanese: Common ground for hunter-gatherer and farmer Y chromosomes. *Journal of Human Genetics* 51: 47–58.

Huang, Y., and H. Li. 2015. Genetic and linguistic evidences suggest a Yeniseian category for Xiongnu. In *Tsinghua Journal of Mongolian History*, vol. 3, ed. Dali Yao, and Yingsheng Liu, 435–456. Beijing: The Commercial Press.

HUGO Pan-Asian SNP Consortium. 2009. Mapping human genetic diversity in Asia. *Science* 326: 1541–1545.

Jobling, M.A., M. Hurles, and C. Tyler-Smith. 2004. *Human evolutionary genetics (origins, peoples and disease)*. New York: Garland Science.

Jobling, M.A., and C. Tyler-Smith. 1995. Father and sons: The Y chromosome and human evolution. *Trends in Genetics* 11: 449–456.

Karafet, T.M., L. Xu, R. Du, W. Wang, S. Feng, R.S. Wells, A.J. Redd, S.L. Zegura, and M.F. Hammer. 2001. Paternal population history of East Asia: Sources, patterns, and microevolutionary processes. *American Journal of Human Genetics* 69 (3): 615–628.

Karafet, T.M., F.L. Mendez, M.B. Meilerman, P.A. Underhill, S.L. Zegura, and M.F. Hammer. 2008. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Research* 18: 830–838.

Kayser, M., S. Brauer, R. Cordaux, A. Casto, O. Lao, L.A. Zhivotovsky, C. Moyse-Faurie, R.B. Rutledge, W. Schiefenhoevel, D. Gil, A.A. Lin, P.A. Underhill, P.J. Oefner, R.J. Trent, and M. Stoneking. 2006. Melanesian and Asian origins of Polynesians: mtDNA and Y chromosome gradients across the Pacific. *Molecular Biology and Evolution* 23: 2234–2244.

Kayser, M., Y. Choi, M. van Oven, S. Mona, S. Brauer, R.J. Trent, D. Suarkia, W. Schiefenhövel, and M. Stoneking. 2008. The impact of the Austronesian expansion: Evidence from mtDNA and Y chromosome diversity in the Admiralty Islands of Melanesia. *Molecular Biology and Evolution* 25 (7): 1362–1374.

Ke, Y., B. Su, X. Song, D. Lu, L. Chen, H. Li, C. Qi, S. Marzuki, R. Deka, P. Underhill, C. Xiao, M. Shriver, J. Lell, D. Wallace, R.S. Wells, M. Seielstad, P. Oefner, D. Zhu, J. Jin, W. Huang, R. Chakraborty, Z. Chen, and L. Jin. 2001. African origin of modern humans in East Asia: A tale of 12,000 Y chromosomes. *Science* 292: 1151–1153.

Kumar, V., A.N. Reddy, J.P. Babu, T.N. Rao, B.T. Langstieh, K. Thangaraj, A.G. Reddy, L. Singh, and B.M. Reddy. 2007. Y-chromosome evidence suggests a common paternal heritage of Austro-Asiatic populations. *BMC Evolutionary Biology* 7: 47.

Li, H., S. Gu, X. Cai, W.C. Speed, A.J. Pakstis, E.I. Golub, J.R. Kidd, and K.K. Kidd. 2008a. Ethnic related selection for an ADH Class I variant within East Asia. *PLoS ONE* 3 (4): e1881.

Li, H., B. Wen, S.J. Chen, B. Su, P. Pramoonjago, Y. Liu, S. Pan, Z. Qin, W. Liu, X. Cheng, N. Yang, X. Li, D. Tran, D. Lu, M.T. Hsu, R. Deka, S. Marzuki, C.C. Tan, and L. Jin. 2008b. Paternal genetic affinity between Western Austronesians and Daic populations. *BMC Evolutionary Biology* 8: 146.

Lu, Y., S.L. Pan, S.M. Qin, Z.D. Qin, C.C. Wang, R.J. Gan, and H. Li. 2013. The Genographic Consortium: Genetic evidence for the multiple origins of Pinghua Chinese. *Journal of Systematics and Evolution* 51 (3): 271–279.

Mirabal, S., M. Regueiro, A.M. Cadenas, L.L. Cavalli-Sforza, P.A. Underhill, D.A. Verbenko, S. A. Limborska, and R.J. Herrera. 2009. Y-chromosome distribution within the geo-linguistic landscape of northwestern Russia. *European Journal of Human Genetics* 17 (10): 1260–1273.

Piazza, A. 1998. Towards a genetic history of China. *Nature* 395: 636–639.

Qin, Z., Y. Yang, L. Kang, S. Yan, K. Cho, X. Cai, Y. Lu, H. Zheng, D. Zhu, D. Fei, S. Li, L. Jin, and H. Li. 2010. A mitochondrial revelation of early human migrations to the Tibetan Plateau before and after the last glacial maximum. *American Journal of Physical Anthropology* 143 (4): 555–569.

Reich, D., R.E. Green, M. Kircher, J. Krause, N. Patterson, E.Y. Durand, B. Viola, A.W. Briggs, U. Stenzel, P.L. Johnson, T. Maricic, J.M. Good, T. Marques-Bonet, C. Alkan, Q. Fu, S. Mallick, H. Li, M. Meyer, E.E. Eichler, M. Stoneking, M. Richards, S. Talamo, M.V. Shunkov, A.P. Derevianko, J.J. Hublin, J. Kelso, M. Slatkin, and S. Pääbo. 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* 468: 1053–1060.

Rootsi, S., L.A. Zhivotovsky, M. Baldovic, M. Kayser, I.A. Kutuev, R. Khusainova, M.A. Bermisheva, M. Gubina, S.A. Fedorova, A.M. Ilumäe, E.K. Khusnutdinova, M.I. Voevoda, L. P. Osipova, M. Stoneking, A.A. Lin, V. Ferak, J. Parik, T. Kivisild, P.A. Underhill, and R. Villems. 2007. A counter-clockwise northern route of the Y-chromosome haplogroup N from Southeast Asia towards Europe. *European Journal of Human Genetics* 15: 204–211.

Scholes, C., K. Siddle, A. Ducourneau, F. Crivellaro, M. Järve, S. Rootsi, M. Bellatti, K. Tabbada, M. Mormina, M. Reidla, R. Villems, T. Kivisild, M.M. Lahr, and A.B. Migliano. 2011. Genetic diversity and evidence for population admixture in Batak Negritos from Palawan. *American Journal of Physical Anthropology* 146: 62–72.

Sengupta, S., L.A. Zhivotovsky, R. King, S.Q. Mehdi, C.A. Edmonds, C.E. Chow, A.A. Lin, M. Mitra, S.K. Sil, A. Ramesh, M.V. Usha Rani, C.M. Thakur, L.L. Cavalli-Sforza, P.P. Majumder, and P.A. Underhill. 2006. Polarity and temporality of high-resolution y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists. *American Journal of Human Genetics* 78: 202–221.

Shi, Y.F., Z.J. Cui, and J.J. Li. 1989. *Quaternary glacier in eastern China and the climate fluctuation*. Beijing: Science Press.

Shi, H., Y.L. Dong, B. Wen, C.J. Xiao, P.A. Underhill, P.D. Shen, R. Chakraborty, L. Jin, and B. Su. 2005. Y-chromosome evidence of southern origin of the East Asian-specific haplogroup O3-M122. *American Journal of Human Genetics* 77 (3): 408–419.

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

Su, B., J. Xiao, P. Underhill, R. Deka, W. Zhang, J. Akey, W. Huang, D. Shen, D. Lu, J. Luo, J. Chu, J. Tan, P. Shen, R. Davis, L. Cavalli-Sforza, R. Chakraborty, M. Xiong, R. Du, P. Oefner, Z. Chen, and L. Jin. 1999. Y-chromosome evidence for a northward migration of modern human into East Asia during the last ice age. *American Journal of Human Genetics* 65: 1718–1724.

Su, B., L. Jin, P. Underhill, J. Martinson, N. Saha, S.T. McGarvey, M.D. Shriver, J. Chu, P. Oefner, R. Chakraborty, and R. Deka. 2000. Polynesian origins: Insights from the Y chromosome. *Proceedings of the National Academy of Sciences of the United States of America* 97 (15): 8225–8228.

Thangaraj, K., L. Singh, A.G. Reddy, V.R. Rao, S.C. Sehgal, P.A. Underhill, M. Pierson, I.G. Frame, and E. Hagelberg. 2003. Genetic affinities of the Andaman Islanders, a vanishing human population. *Current Biology* 13 (2): 86–93.

Underhill, P.A., P. Shen, A.A. Lin, L. Jin, G. Passarino, W.H. Yang, E. Kauffman, B. Bonné-Tamir, J. Bertranpetit, P. Francalacci, M. Ibrahim, T. Jenkins, J.R. Kidd, S.Q. Mehdi, M.T. Seielstad, R.S. Wells, A. Piazza, R.W. Davis, M.W. Feldman, L.L. Cavalli-Sforza, and P. J. Oefner. 2000. Y chromosome sequence variation and the history of human populations. *Nature Genetics* 26: 358–361.

Wang, C.C., S.E. Farina, and H. Li. 2013a. Neanderthal DNA and modern human origins. *Quaternary International* 295: 126–129.

Wang, C.C., S. Yan, Z.D. Qin, Y. Lu, Q.L. Ding, L.H. Wei, S.L. Li, Y.J. Yang, L. Jin, and H. Li. 2013b. Late Neolithic expansion of ancient Chinese revealed by Y chromosome haplogroup O3a1c-002611. *Journal of Systematics and Evolution* 51 (3): 280–286.

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

Wen, B., X. Xie, S. Gao, H. Li, H. Shi, X. Song, T. Qian, C. Xiao, J. Jin, B. Su, D. Lu, R. Chakraborty, and L. Jin. 2004b. Analyses of genetic structure of Tibeto-Burman populations reveals sex-biased admixture in southern Tibeto-Burmans. *American Journal of Human Genetics* 74 (5): 856–865.

Xue, Y., Q. Wang, Q. Long, B.L. Ng, H. Swerdlow, J. Burton, C. Skuce, R. Taylor, Z. Abdellah, Y. Zhao, Asan, D.G. MacArthur, M.A. Quail, N.P. Carter, H. Yang, and C. Tyler-Smith. 2009. Human Y chromosome base-substitution mutation rate measured by direct sequencing in a deep-rooting pedigree. *Current Biology* 19 (17): 1453–1457.

Yan, S., C.C. Wang, H. Li, S.L. Li, and L. Jin. 2011. Genographic Consortium: An updated tree of Y-chromosome Haplogroup O and revised phylogenetic positions of mutations P164 and PK4. *European Journal of Human Genetics* 19 (9): 1013–1015.

Zhivotovsky, L.A. 2001. Estimating divergence time with the use of microsatellite genetic distances: Impacts of population growth and gene flow. *Molecular Biology and Evolution* 18: 700–709.

Zhivotovsky, L.A., P.A. Underhill, C. Cinnioğlu, M. Kayser, B. Morar, T. Kivisild, R. Scozzari, F. Cruciani, G. Destro-Bisol, G. Spedini, G.K. Chambers, R.J. Herrera, K.K. Yong, D. Gresham, I. Tournev, M.W. Feldman, and L. Kalaydjieva. 2004. The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *American Journal of Human Genetics* 74: 50–61.

Zhong, H., H. Shi, X.B. Qi, C.J. Xiao, L. Jin, R.Z. Ma, and B. Su. 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 (7): 428–435.

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