Y-chromosome-based genetic pattern in East Asia affected by Neolithic transition

Shao-Qing Wen, Xin-Zhu Tong, Hui Li*
MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai 200438, China

Article info
Article history: Available online 6 May 2016
Keywords: East Asia Neolithic transition Y chromosomal lineages Demographic history Linguistic family Archeology culture

ABSTRACT
The Neolithic transition, defined as the shift from a hunter-gatherer economy into the one based on agricultural activities, is assumed to have resulted in extensive human population growths. Despite major progress has been made by archaeologists in the use of archaeobotanic data to reconstruct a reliable time frame of Neolithic transition in China, the roles played by Neolithic transition in East Asian demographic history are not yet well understood. This paper offers a perspective on the issues regarding when and how the East Asian population expanded and its consequences. Considering diverse genetic evidences, we revealed that, in East Asia, there were at least two population expansion events in the Paleolithic Era and notably, the latter Paleolithic expansion and climate improvement after Last Glacial Maximum (LGM, about 15 ka) might together facilitate the emergence of agriculture. In comparison to the Paleolithic expansion, Neolithic expansion had its own characteristics, such as stronger expansion dynamics and sex-specific expansion pattern. According to the Y-chromosome-based tree in our recent studies, we found strong signals of partial paternal lineage expansions during or after the Neolithic transition, for instance, O3a2c1a-F5 (Oa), O3a2c1-F46 (Ob), O3a1c1-F11 (Og), C3a1-F2613 (Ca), Q1a1-M120 (Qa), and O1a1a1-F78 (Od), suggesting the influence of a strong bottleneck caused by cultural changes. Finally, we discussed the relationship between major East Asian paternal lineages and linguistic families, as well as early archeology cultures, linking the Daxi culture to the modern Hmong-Mien populations, the Liangzhu Culture and/or Songze Culture to modern Austronesian and Daic populations, and Yangshao Culture, Dawenkou Culture and/or Longshan Culture, and Hongshan Culture to modern Sino-Tibetan populations.

1. Introduction
Out-of-Africa hypothesis was first proposed in the late 1980s. With new methods for producing and analyzing genome-wide data, the scenarios of the global peopling of modern humans become clearer. Around 60 ka (Henn et al., 2012), modern humans expanded out of Africa, and by approximately 15 ka, they had colonized all inhabitable continents (Jobling et al., 2013). During most of this period, especially in the Last Glacial Period (~110–10 ka), the climate was both cold and unstable (Shi et al., 1989). Nevertheless, the long-term cold weather, on one hand, limited the living space for human beings; on another hand, it made the sea level much lower than present, and as a result, many of today’s islands were joined to the continents, providing paths for modern human migrations (Jobling et al., 2003). After approximately 15 ka, the ice sheets, which covered northern and central Asia, northern Europe, and North America, started to recede (Clark et al., 2009), and by approximately 10 ka the temperature began to rise and became stable as we know today. This period was a flourishing time for the expansion of modern humans. Significantly, it was just then that the development of agriculture began in the Fertile Crescent in the Near East, with domesticating few wild plant and animal species. During the subsequent times between 9 and 4 ka, agriculture arose independently in different geographical regions except Australia (Diamond and Bellwood, 2003). Albeit it had important consequences for rapid changes in selective pressures, in particular a spread of infectious diseases catalyzed by sedentarization and the increase of population density, this ‘Neolithic transition’, from a foraging subsistence strategy to a sedentary farming society, affected many aspects of lifestyle and led to the sedentarization of many populations (Bocquet-Appel and Bar-Yosef, 2008;
As such, Neolithic transition must be seen as a major turning point in the human prehistory.

Human prehistory was studied by different means in various disciplines. Archaeologists analyzed the data drawn from archaeological remains to understand the culture evolution of ancient populations; historians dissected the records obtained from ancient books to explore the human activities; linguists compared the features gained from survivor languages to study the diversification of languages. Similarly, studying the pattern of genetic diversity among human populations can tell us about the demographic history of populations, i.e., migrations, expansions and colonization. Compared to autosomal DNA, two uniparental genetic materials, mitochondrial DNA (mtDNA) and Y chromosome, are susceptible to the effects of genetic drift, and hence, are more likely to generate population-specific haplotype distribution, which allows us to trace the population history easily and clearly (Wang and Li, 2013). Moreover, because of the sex-specific pattern of inheritance of the Y chromosome, drift can be greatly accentuated by social selection, such as social organization and technological innovation (Jobling, 2012). Y chromosome has several properties that make it potentially very informative about historical events, including the Neolithic transition.

We investigate the demographic history of populations (Cavalli-Sforza, 1998), and comparing the benefits that might be possible. Although all these societies might loosely be called chiefdoms, with significantly different subsistence preferences, sequences of change, modes of leadership, and even symbolic representations, they have participated in underpinning the early Chinese civilization (Fang et al., 2015). Consequently, as quoted above, the geographical distribution of initial complex societies with representative archaeological sites and a general time frame of Neolithic transition in China are summarized in Figs. 1 and 2.

3. Cultural change and Neolithic super-grandfathers

Demographic changes are known to leave footprints on genetic polymorphism. Approaches to inferring human demographic history have included inferences from the biparentally inherited autosomes and X chromosomes with high-density genotyping or sequencing data (e.g. Schaffner et al., 2005; Liu and Fu, 2015), from the whole mtDNA genomes (e.g. Gignoux et al., 2011; Zheng et al., 2012), and from the whole Y chromosome genomes (e.g. Hallast et al., 2015; Karmin et al., 2015). However, current analyses are often limited and seem to produce conflicting results. Estimates based on the analysis of nuclear polymorphism suggested the upper Paleolithic expansions (Aime et al., 2013) or, paradoxically, the more recent Neolithic expansions (Aime et al., 2014) in both African and Eurasian populations with contrasted lifestyles. Estimates based on the analysis of mitochondrial DNA in worldwide populations show that most major maternal lineage expansions began before the Neolithic transition (Zheng et al., 2012; Karmin et al., 2015). Yet a series of recent studies (Wang et al., 2013; Hallast et al., 2015; Karmin et al., 2015) based on globally distributed whole Y chromosomes found that the most major paternal lineage expansions coalesced in the Holocene after the emergence of agriculture. These inconsistent results were mainly due to either ascertainment bias (Hallast et al., 2015; Liu and Fu, 2015), such as choosing appropriate mutation rate and employing various demographic model, or, most importantly, the nature of diverse genetic materials themselves. For instance, autosomal genome might have important implications for our understanding of early dispersals; mtDNA and Y chromosome provide matrilineal and patrilineal demographic history, respectively, revealing pictures of sex-specific processes in the past. Thus, a combination of insights derived from different genetic data may be the most realistic postulate.

Around 50 ka, one Paleolithic expansion caused by major non-African founder population after the out-of-Africa bottleneck has been found by autosomes (Aime et al., 2013; Liu and Fu, 2015), mtDNA (Karmin et al., 2015) and Y chromosome genomes (Hallast et al., 2015; Karmin et al., 2015), corresponding to a rapid initial colonization model of Eurasia and Oceania. In eastern Asia, this Paleolithic expansion happened at approximately 30 ka (Wang et al., 2013; Yan et al., 2014; Hallast et al., 2015), which was
coincident with the estimated time of the second wave peopling of this continent and the evidence from the genetics of human parasites and molecular adaptations that, almost simultaneously, the East Asian related genotype of Tuberculosis and Ectodysplasin A receptor (EDAR) emerged in central China (Kamberov et al., 2013; Luo et al., 2015). Another inferred Paleolithic expansion was triggered by climate improvement after Last Glacial Maximum (LGM, about 15 ka), which has not been detected by autosomes (Liu and Fu, 2015) and Y chromosome genomes (Karmin et al., 2015), but obviously observed by mtDNA genomes (Zheng et al., 2011, 2012; Karmin et al., 2015). This finding best fitted the “demographic theory” proposed by Sauer (1952) that Paleolithic expansions in some populations may have ultimately favored their shift toward farming. Thus, to some extent, climate improvement and modest population pressure might together facilitate the emergence of agriculture.

**Fig. 1.** The geographical distribution of initial complex societies in China with representative archaeological sites at approximately 6000 cal BP. The yellow dots represent the sites of Hongshan societies, centered in the West Liao River Valley in northeast China. The pink and brown dots represent the sites of Middle to Late Yangshao societies and Dawenkou societies, respectively, spread through the middle and farther downstream of the Yellow River valley. The purple and blue dots represent the sites of Daxi societies and Songze societies, respectively, situated along the middle Yangzi River valley and the Yangzi River delta in the Middle and Lower Yangtze River regions in southern China. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

**Fig. 2.** The simplified phylogenetic tree of East Asian-specific Y-chromosome haplogroups. The genetic markers used to classify the haplogroup are labeled on the branches. The blue band represents the time framework of Neolithic transition in China. The isosceles triangle represents the structure of lineage expansion. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
The autosomal (Aimé et al., 2014; Liu and Fu, 2015), mtDNA (Gignoux et al., 2011; Karmin et al., 2015) and Y chromosome (Hallast et al., 2015; Karmin et al., 2015) lines of evidence all suggest a strong signal of global population booming after the advent and spread of agriculture in the Holocene. This population genetic finding seem consistent with the archaeological evidence that there was a high birth rate phase in the skeletal record in Eurasian burial sites during the Neolithic period (Bocquet-Appel, 2011), and the molecular epidemiological evidence that, at that time, Tuberculosis rapidly expanded as a consequence of increases in human population density (Comas et al., 2013). In comparison to the Paleolithic expansion, Neolithic expansion has its own characteristics. Firstly, as agriculture has provided a much more stable food supply than hunting and foraging, the Neolithic expansion had significantly stronger expansion dynamics (i.e., higher growth rates and effective population sizes) than Paleolithic expansion (Wang et al., 2013; Karmin et al., 2015; Liu and Fu, 2015). Secondly, a clear difference in male and female population history was observed during the Neolithic expansion; that is, the female effective population size is several-fold higher than the male effective population size when the male lineages suffered a post-Neolithic bottleneck (Rasteiro and Chikhi, 2013; Wang et al., 2015; Karmin et al., 2015). This sex-biased expansion might truly reflect the influence of social selection and male-specific behavior during the Neolithic expansion (Rasteiro and Chikhi, 2013), which means that the prehistoric men possessing social prestige had more chance to enjoy enhanced reproductive success and the prehistoric women preferred those males who had stronger military prowess or higher social status associated with farming or metalworking. As a result, only a few of paternal lineages with a 10- to 100-fold increase in population size, rather than most of them (Wang et al., 2013; Karmin et al., 2015), which can potentially have a large impact on the Y-chromosomal gene pool.

Application of next-generation sequencing (NGS) to whole Y chromosomes enabled informative phylogenetic tree to be constructed in which branch-lengths are proportional to time, allowing direct assessment of the times-to-most-recent-common-ancestor (TMRCAs) of nodes. In addition, according to the NGS-based tree, we can systematically analyze how the various factors have influenced the human population size and structure. For instance, those primary sub-branches with the long internal branch lengths indicate low population growth and frequent bottlenecks, whereas those clades with short internal branch lengths seen as star-like structures are interpreted as a strong signal of rapid population expansions. Apart from star-like phylogenies, sometimes, there are bifurcating structures in phylogenetic tree, which is also associated with population expansions. However, the detailed demographic history reflected by two contrasting structures is quite different. For example, by using coalescent simulations, Sikora et al. (2013) found star-like and bifurcating phylogenies for the major lineages R1b and E1b1a associated with the Neolithic expansions in Western Europe and sub-Saharan Africa, respectively. In concrete terms, R1b has a star-like structure indicative of a successful expansion within a very short period as soon as it entered the continent, while E1b1a has a more regular bifurcating structure, indicating that the E1b1a expansion began from a larger population size, took thousands of years and ended only recently.

Regarding paternal demographic history in East Asia, in our recent studies (Wang et al., 2013; Yan et al., 2014), we have observed a strong bottleneck and population expansions in Neolithic Era. These paternal population expansions can be easily detected via NGS-based tree. Among the Y-chromosomal lineages, there are five star-like phylogenies, O3a2c1-F5 (Oz), O3a2c1-F46 (Oβ), O3a1c1-F11 (Oγ), C3a1-F2613 (Cx) and Q1a1-M120 (Qz), and one bifurcating phylogenies, O1a1a1-F78 (Oδ), in East Asian-specific phylogenetic tree. Notably, we found that up to 65% of modern Chinese are patrilineal descendants of these six super-grandfathers. The expansion dates are estimated 5.4 ka for Oz, 6.5 ka for Oβ, 6.8 ka for Oγ, 6.4 ka for Cx, 5.2 ka for Qz and 5.0 ka for Oδ (see Supporting Discussions in Yan et al., 2014), during or after the Neolithic transition in China (Fig. 2). As mentioned above, we revealed that a majority of Han Chinese derived from just a few patrilineal ancestors in the Neolithic Era, but now comes the question, who were they? Whether each of the super-grandfathers could be related to the legendary chieftains of the initial complex societies, like the well-known cases of Genghis Khan’s star-cluster and Giocangga’s Manchu cluster (Balaresque et al., 2015)? In the near future, the joint analyses of ancient DNA and modern humans can throw light on these issues, which might be a key to understand the origin of Chinese populations.

4. Farming/language dispersal and its implication

Languages, similar to genes, serve as an important clue to trace the past of human beings (Pagel, 2009). Most linguists have recognized that languages within a certain linguistic family have a common origin in the range of 10 ka ago, which also fits perfectly with archaeological evidence for the advent of agriculture in each region (Diamond and Bellwood, 2003; Henn et al., 2012). Thus, Bellwood (2009) proposed a farming/language dispersal hypothesis that farmers and their culture replace neighboring hunter-gatherers and the latter’s culture, leading to a sex-specific language replacement. This phenomenon has been observed by genetic investigations, which means that linguistic families show strong association with Y-chromosomal DNA variation but not with mtDNA variation (Poloni et al., 1997; Kemp et al., 2010; Forster and Renfrew, 2011; Wang and Li, 2013). Such farming/language dispersals have already been hypothesized for East Asia, including the Austro-Asiatic, Tai, Sino-Tibetan and Austronesian language families (Diamond and Bellwood, 2003).

Like elsewhere, in East Asia, linguistic families are well associated with Y chromosome diversity (Wang and Li, 2013). Most families have their dominant Y lineages, for instance, the Y chromosome haplogroup O3-M134 in Sino-Tibetan speakers (Wen et al., 2004), O2-M95 and O3-M7 in Hmong-Mien and Austro-Asiatic speakers (Kumar et al., 2007; Cai et al., 2011), O1-M119 in Tai-Kadai and Austronesian speakers (Li et al., 2008), and N-TAT in Uralic speakers (Lappalainen et al., 2006). Thus, Y-chromosomal lineages sometimes may be seen as a predictor of language (Forster and Renfrew, 2011). Furthermore, according to the similarity in the Y chromosome profiles, there are four pairs of congenetic families, that is, Austro-Asiatic and Tai Kadai (Li et al., 2008), Mon Khmer and Hmong Mien (Cai et al., 2011; Lu et al., 2011), Sino-Tibetan and Uralic (Shi et al., 2013; Hu et al., 2015), Yeniseian and Palaesiberian (Lell et al., 2002; Dulik et al., 2012; Flegontov et al., 2016), demonstrating the substantial early language contact and gene flow between these ancient populations. Notably, between approximately 4000–2000 ka (Zhui, 2009), Tai-Kadai, Hmong-Mien, Sino-Tibetan, and Yeniseian languages transformed into tonic and analytic languages, becoming quite different from the rest four. There must be some social or biological factors induced the transformations at that time, which is worth doing more linguistic and genetic researches.

5. Ancient DNA evidence and its implication

Apart from the indirect means of analyzing the genomes of modern populations, another method to address the past is the relatively direct means of studying ancient DNA, which has
advantages that it provides ‘real’ information. Nevertheless, the ancient DNA studies associated with the East Asian archeology sites during the Neolithic transition were extremely limited. Based on the findings of three relevant ancient DNA studies (Li et al., 2007; Cui et al., 2013; Dong et al., 2015), we had a hazy idea of the distribution of Y chromosomal lineages during the Neolithic transition.

Haplogroup O3-M7, found at high frequency in the Daxi site (Daxi Culture) in the middle Yangzi River valley (Li et al., 2007), is seen as the characteristic lineage of Mon-Khmer and Hmong-Mien (Cai et al., 2011), linking the Daxi culture to the modern Hmong-Mien populations. Haplogroup O1-M119, which accounts for about 60% in Xindili site and Maqiao site (Lianzhu Culture, which succeeded the Songze Culture) in the Yangzi River valley (Li et al., 2007; Dong et al., 2015), respectively. In addition, Haplogroup N-M231, a sister haplogroup of O-M175, which exhibited an extremely high frequency in Niuheilang site (Hongsan Culture) in the West Liao River valley (Cui et al., 2013), has both a unique and widespread distribution throughout northern Eurasia, Tibeto-Burman and Chinese-speaking populations and thought to be originated in China (Hu et al., 2015). These findings linked Late Yangshao Culture and/Longshan Culture, Dawenkou Culture and Hongsan Culture to modern Sino-Tibetan populations. It is worthwhile to note that a clear genetic pattern was observed among the prehistoric cultures, indicating the genetic foundation of the multiple origins of the Chinese Civilization.

Further insights into the detailed Y-chromosome-based genetic pattern during the Neolithic transition will benefit from high-density genotyping or population-scale sequencing data from ancient DNA studies and their interpretation in an interdisciplinary framework, including historical, archaeological, linguistic, and paleoecological evidence.

6. Perspectives

In our previous studies, Y chromosome analyses of present-day East Asian populations have revealed that four principal haplogroups C, D, N, and O, which accounted for more than 90% of the East Asian Y chromosomes, are suggested to have Southeast Asian origins in Paleolithic Era. In this article, we found that the extant genetic pattern was deeply affected by cultural change during the Neolithic transition, suggesting partial male lineage expansions, in particular, O3a2c1-F5 (Ox), O3a2c1-F46 (Oj), O3a1c1-F11 (Oy), C3a1-F2613 (Cq), Q1a1-M120 (Qs), and O1a1a1-F78 (O6). However, the initial expanding regions of these clans remain unknown.

From the archaeological perspective, an additional center related to origin of ancient tropical agriculture was located in the region of Zhjiang River, which deserves thorough investigation (Zhao, 2011). From our perspective, this center may be associated with the origins of some southern ancient ethnic groups. Therefore, the ancient DNA studies of this region might well repay investigation. Moreover, as emphasized above, the Neolithic transition has played an important role in formation of the genetic and cultural pattern in East Asia but there remain many unsolved questions due to the lack of ancient DNA researches.

Acknowledgments

This work was supported by NSF for Excellent Young Scholar (31222030), MOE Scientific Research Project (113022A), Shanghai Shuguang Project (14SG05) and The French National Research Agency (No. ANR-12-BSH2-0004-01).

References


