



# Origin of ethnic groups, linguistic families, and civilizations in China viewed from the Y chromosome

Xueer Yu<sup>1,2</sup> · Hui Li<sup>1,2</sup>

Received: 30 September 2020 / Accepted: 22 April 2021 / Published online: 26 May 2021  
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

## Abstract

East Asia, geographically extending to the Pamir Plateau in the west, to the Himalayan Mountains in the southwest, to Lake Baikal in the north and to the South China Sea in the south, harbors a variety of people, cultures, and languages. To reconstruct the natural history of East Asians is a mission of multiple disciplines, including genetics, archaeology, linguistics, and ethnology. Geneticists confirm the recent African origin of modern East Asians. Anatomically modern humans arose in Africa and immigrated into East Asia via a southern route approximately 50,000 years ago. Following the end of the Last Glacial Maximum approximately 12,000 years ago, rice and millet were domesticated in the south and north of East Asia, respectively, which allowed human populations to expand and linguistic families and ethnic groups to develop. These Neolithic populations produced a strong relation between the present genetic structures and linguistic families. The expansion of the Hongshan people from northeastern China relocated most of the ethnic populations on a large scale approximately 5300 years ago. Most of the ethnic groups migrated to remote regions, producing genetic structure differences between the edge and center of East Asia. In central China, pronounced population admixture occurred and accelerated over time, which subsequently formed the Han Chinese population and eventually the Chinese civilization. Population migration between the north and the south throughout history has left a smooth gradient in north–south changes in genetic structure. Observation of the process of shaping the genetic structure of East Asians may help in understanding the global natural history of modern humans.

**Keywords** Population Genetics · East Asia · Ancient DNA · Neolithic Age · Y chromosome · Civilization

## Introduction

East Asia mainly covers China, Mongolia, North Korea, South Korea, and Japan. This vast area has been settled by humans (the genus *Homo*) since the Pleistocene. Composed of various climates and terrains, the land has resulted in tremendous and diversified people–culture–language complexes. Researchers of anthropology, linguistics, genetics,

and other disciplines interpret this diversification process with different materials. Anthropologists and linguists focus on cultural practices attached to populations, while geneticists have made great efforts with biological materials directly from humans, including maternal mitochondrial DNA (mtDNA), paternal nonrecombining Y chromosome (NRY), and biparental whole genome. However, each of these traditional insights is limited within a single discipline and therefore cannot reveal the complete natural history of humans in East Asia. Novel studies try to form a multidisciplinary, multifaceted, and unified perspective to summarize the whole story by, for instance, establishing the connection between cultural practices and the genetic structure of an ethnic group.

Communicated by Stefan Hohmann.

✉ Hui Li  
LHCA@Fudan.edu.cn

<sup>1</sup> MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai 200438, China

<sup>2</sup> Shanxi Academy of Advanced Research and Innovation, Fudan-Datong Institute of Chinese Origin, Datong 037006, China

## Immigrating into East Asia (> 12,000 year BP)

### Indigenous descendants or African immigrants

The origin of anatomically modern humans (AMH, *Homo sapiens sapiens*) within East Asia has been debated for a long time. Most Chinese paleontologists insist on multiregionalism (origins in multiple regions) or continuity with the hybridization hypothesis (Wu 2006); that is, AMH in East Asia descended from local *Homo erectus*, who migrated out of Africa and settled in East Asia approximately 1.8 million years ago. Numerous fossils of *Homo erectus* and *Homo sapiens* have been found intermittently, filling the logical evolutionary array. These fossils share some similar morphological features unique to East Asia, such as shovel-shaped incisors and facial flatness. These morphological features also occur in modern East Asians at high frequency, but are uncommon or absent in Africa and Europe. Characteristics similar to ancient European hominins suggest interbreeding with populations from other parts of Eurasia but not a total replacement. Paleolithic archaeological aspects, such as stone tool manufacturing technology, artifact morphology, and trends of regional cultural traditions, also mainly support continuity from *Homo erectus* to archaic *Homo sapiens* (Denisovan) and AMH (Gao et al. 2017).

However, genetic studies support the dominant Recent Out-of-Africa hypothesis, that is, AMH in East Asia were exclusively from AMH in Africa. Archaic *Homo sapiens* originated from *Homo ergaster* in Africa and then spread to Eurasia, replacing *Homo erectus*. Adapted to specific geographic environments, archaic *Homo sapiens* differentiated into three subspecies: Rhodesians mainly in Africa, Neanderthals mainly in western Eurasia, and Denisovans mainly in eastern Eurasia. AMH originated from Rhodesians in Africa approximately 200,000–100,000 years ago. As the youngest Toba eruption occurred approximately 74,000 years ago in Indonesia, subsequent global volcanic winter (Rampino and Self 1992) and other effects caused a sharp decrease in human populations (Rampino and Self 1993). Other hominins nearly went extinct, while a few AMH in Africa survived and expanded. AMH left Africa several times and then occupied Eurasia, replacing Neanderthals and Denisovans. Viewed from mitochondrial DNA, all modern non-Africans derive from haplogroup L3 (Ingman et al. 2000), which emerged in East Africa 70,000–60,000 years ago (Soares et al. 2016). The mitochondrial DNA of nearly all Eurasians belong to haplogroups M and N derived from L3. Fossilized human teeth with full AMH features found in the layer deposited between 120,000 and 80,000 years ago in Fuyan Cave in South China indicate that AMH appeared in East Asia much earlier than the time proposed by the Recent Out-of-Africa theory (Liu et al. 2015); however,

additional in situ teeth were dated much younger using mtDNA sequences (Sun et al. 2021). Furthermore, viewed from the Y chromosome, all modern male East Asians carry the M168T mutation, which originated in Africa (Ke et al. 2001). Whole-genome studies also strongly support the Out-of-Africa hypothesis (Mallick et al. 2016; Pagani et al. 2016).

Both hypotheses feature some flaws. Fossils and relics are questioned as circumstantial evidence and cannot construct and confirm a complete chain of evolution. On the one hand, an individual sample of hominin fossils cannot represent the whole population, and their rarity causes analytic results to be irreproducible. The within-group variations of some characteristics might be much larger than the between-group variations. The measurements of morphological features are subjectively influenced by the measurers, as fossils are often fragmented when excavated. Moreover, specific morphological features, such as phenotypes are possibly influenced by the environment rather than by genetics. Researchers insisting on an indigenous origin have also been questioned in detail. Morphological features claimed by paleontologists to be unique to East Asia actually occurred widely in Pleistocene Africans (Groves 1989; Bräuer and Stringer 1997). There was a gap between fossils of archaic humans and AMH in East Asia between 100,000 and 50,000 years ago (Su et al. 1999), a critical period for AMH leaving Africa. Genetic results directly confirmed an exclusive origin from Africa and are more convincing than morphological studies. However, paleontologists argued that the most recent common ancestors of extant humans are not equal to ancestors of all AMH, while other native hominins made at least part of the genetic contribution (Wu 2006). However, the indigenous hypothesis is difficult to examine with DNA. Very few ancient DNA results of archaic humans, let alone *Homo erectus* of much older age, have been obtained. In contrast, paleontologists raised several questions for genetics: the basis of the Out-of-Africa hypothesis, a constant mutation rate does not exist, and the molecular clock theory is thus unreliable (Rodriguez-Trelles et al. 2001). Genetic drift and other unknown biological changes might cause bias and errors in conclusions, and the limited number of loci used leads to controversial conclusions drawn by different geneticists (Wu 2006). The Toba catastrophe theory and the supposed population decrease are also questioned by geological evidence from Africa and the continuous stone tool industry in India (Lane et al. 2013; Clarkson et al. 2020).

Recently, both hypotheses were revisited and revised by the discovery of new fossils and improved sequencing technologies, and thus, neither model was completely correct. New findings suggested that the origin and spread of AMH was a dynamic and complex process that cannot be generalized to a single model. Another advance was that genetic analyses of archaic humans confirmed gene flow between Neanderthals,

Denisovans, and AMH. The Neanderthal contribution to non-Africans was 1.5–2.1%, while the Denisovan contribution to mainland Asians and populations in Australasia (Papua New Guinea and Australia) was ~ 0.2% and ~ 5%, respectively. Intriguingly, an unknown archaic hominin contributed 2.7–5.8% of the Denisovan genome (Prüfer et al. 2013). Although the proportion varies globally, archaic humans are not the direct ancestors of AMH. Therefore, the Recent Out-of-Africa hypothesis is more widely accepted. AMH in East Asia originated in Africa and interacted with neighboring archaic humans.

### Directions of migration and diffusion

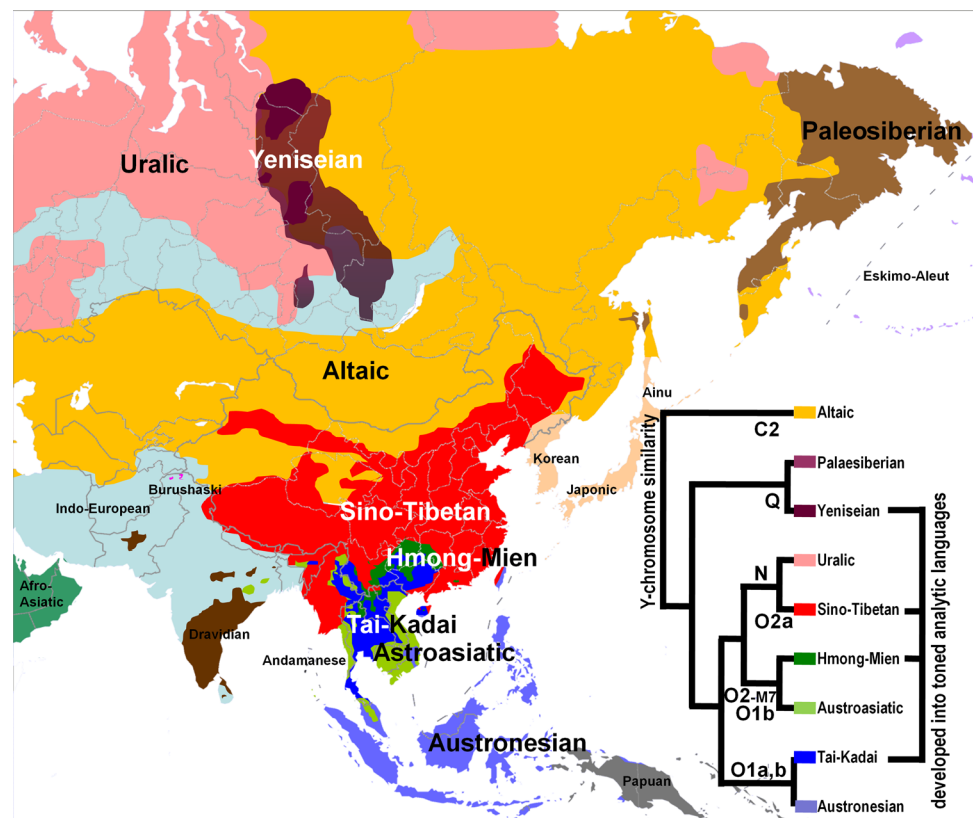
Researchers further traced the routes through which AMH arrived in East Asia. Three models were postulated, namely, a northern route, a southern route, and independent evolution of northern and southern East Asians (Piazza 1998; Jobling et al. 2003; Zhong et al. 2011). The genomic diversity of modern humans pointed to the southern route into East Asia followed by a subsequent south-to-north diffusion in East Asia. The presence and dispersal of mtDNA macro-haplogroups M and N derived from L3 suggested migration along the southern coast of Asia through India into Southeast Asia (Macaulay et al. 2005). The southern route is also supported by paternal evidence. Four Y-chromosome haplogroups, C, D, N, and O, accounting for 93% of the present East Asian males (Wang and Li 2013), all have southern origins. Males of haplogroup C went through the southern coastal route to Southeast Asia approximately 60,000 years ago and then spread to East Asia approximately 40,000 years ago (Zhong et al. 2010). Approximately 60,000 years ago, the population carrying haplogroup D spread northward to East Asia (Shi et al. 2008). These D-M174 populations surprisingly settled on the Tibetan Plateau, adapting to extreme environments, possibly benefitting from gene flow from Denisovans (Huerta-Sánchez et al. 2014). The younger haplogroups N and O migrated to East Asia in several waves 40,000 years ago (Wang and Li 2013). Furthermore, a previous study based on the autosomal single nucleotide polymorphisms (SNPs) also confirmed that Southeast Asian populations made more genetic contributions to East Asian populations than Central/South Asian populations did (HUGO Pan-Asian SNP Consortium 2009). However, new data on ancient human genomes suggested that the former southern route model needed to be revised. For instance, the 45,000-year-old Ust'-Ishim man from North Asia most likely carried haplogroup NO rather than K2 (Wong et al. 2017), probably indicating an earlier entry of this haplogroup into East Asia. In addition, migration in all directions within Asia during the early Holocene occurred more frequently than previously thought (Bai et al. 2020; Yang et al. 2020), despite primary south-to-north expansion.

### Gathering into ethnic groups (12,000–5300 year BP)

The early East Asian AMH that arrived in East Asia were hunter gatherers in a loose society. There were neither ethnic groups nor linguistic families. How they gathered into ethnic groups can also be observed in population genetics. Ethnic groups can be identified with the sharing of language, cultural traditions, and/or religion and legends of common ancestors (Anderson 2006). Present-day East Asians speak languages classified into several language families, namely, Altaic, Sino-Tibetan, Tai-Kadai, Austronesian, Austroasiatic, Hmong-Mien, and some isolated languages, such as Koreanic and Japonic languages (Fig. 1). The origin and expansion of these linguistic families are in ongoing debate as well. China, as the largest country in East Asia and with a vast territory and rich history, has had continuous interactions with adjacent regions and thus can serve as an example for studying the history of East Asia.

Among the ethnic groups in East Asia, Han Chinese has the largest population. As a typical sample, its culture may be traced to the Neolithic Age through anthropological study. Populations of Han Chinese, using Sinitic languages, which belong to one of the two subfamilies of the Sino-Tibetan family, share a common jade culture inherited from the Neolithic Age. In the previous perceptions, the oldest jade ware in China was found in the Xinglongwa culture in northeastern China, dated to 8200 years ago (Cui 2015). However, much older jade ware of more than 9000 years old was found at the Xiaonanshan site in Heilongjiang, which is north of Xinglongwa (Li and Yang 2019). Therefore, we assumed that Han Chinese originated in northeastern China. However, it was not easy to determine whether the early populations used jade ware of the same origin. The raw materials and shapes of jade pieces and their uses varied across Neolithic China. Archaeological interpretations of excavated remains sometimes introduced more mysteries. Therefore, it is necessary to introduce genetics to trace the origin of ethnic groups, including Han Chinese. The background genetic structure of the East Asian population was shaped during the Neolithic Age. Since the Neolithic Age, most East Asian societies have been patriarchal (Stoneking and Delfin 2010); therefore, maternal expansions viewed from mtDNA cannot fully reflect iterations in the dominant population (Wen et al. 2004). Moreover, the distribution of maternal lineages is strongly affected by the environment through the metabolic functions of mitochondria and thus is less related to the migration history of ethnic groups. Genome-wide data are the most precise tool to depict relationships between populations and is widely used in current population studies; however, obstacles in obtaining ancient DNA have strongly limited ancient data collection until recent years

**Fig. 1** The distribution of the linguistic families approximately correlates with Y chromosome differences in East Asia. (Huang and Li 2017)



if we want to study populations of specific time directly. NRY data, documenting patrilineal migration and expansion honestly because of nonrecombination, is convenient to acquire because of relatively short base pair length (~60 Mb) as compared to whole genome data but much larger amount of information than mitochondrial DNA (16 kb) and is strongly associated with languages (Wang and Li 2013), and; therefore, these data are suitable for reconstructing the history of East Asia when there is a lack of genome-wide data. Assumed ancestors of certain language families show a clear affinity with their descendants today from a paternal perspective (Fig. 2), with similar patterns at the genome-wide level (Fig. 3). We can also observe the expansion and interactions of ethnic groups by the distribution of certain NRY haplogroups (Fig. 4). Intriguingly, dominant NRY haplotypes coincide with legends of common ancestors of Han Chinese (Yan et al. 2014). Our study revealed that one of the most important NRY lineages of Han Chinese, O2a-M324, originated in northeastern China.

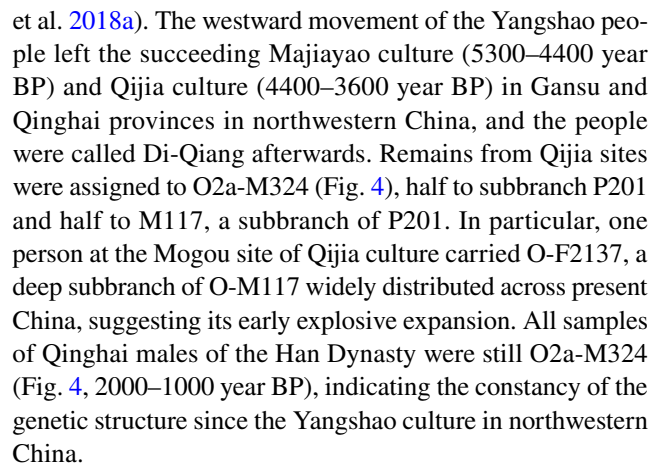
Waves of Paleolithic AMH migrated and settled across East Asia gradually but finally replaced local archaic humans. Surviving from the Last Glacial Maximum, AMH hunter-gatherers developed agriculture in a warmer climate approximately 10,000 years ago (Zhao 2010). Adapting to different environments, two centers of different agriculture appeared in East Asia. As one of the independent origins of agriculture in the world, the land of East Asia nurtured

a much larger population than ever before, especially for males (Wang et al. 2016). Agriculture provides a more comfortable, safe, and stable food supply than hunting or gathering, leading to higher population fertility and infant survival rates; thus, different populations accumulated various cultures, formed stable confederates, and eventually developed into several linguistic families. During years of development, loose communities tended to form more united tribes or even kingdoms. The increase in population led to an increasing demand for land; therefore, populations with powerful military forces started to conquer nearby regions (Liu 2005), with demic diffusion of languages and artifacts. Populations formed several regional systems and cultural types in China (Su and Yin 1981), with different dominant haplogroups (Fig. 5).

### Millet farming centralized ethnic groups

The world's oldest millet farming remains were discovered at sites of the Cishan culture in northern China along the Sanggan River near Beijing, dating back to at least 8200 years ago, with the earliest evidence of domestication since 10,000 years ago (Lu et al. 2009). Large amounts of wild and domestic animal remains were also discovered. At the Cishan site, dogs and pigs were certainly domesticated, and chickens were possibly domesticated, as suggested by early studies (Chow 1981), although there are controversial

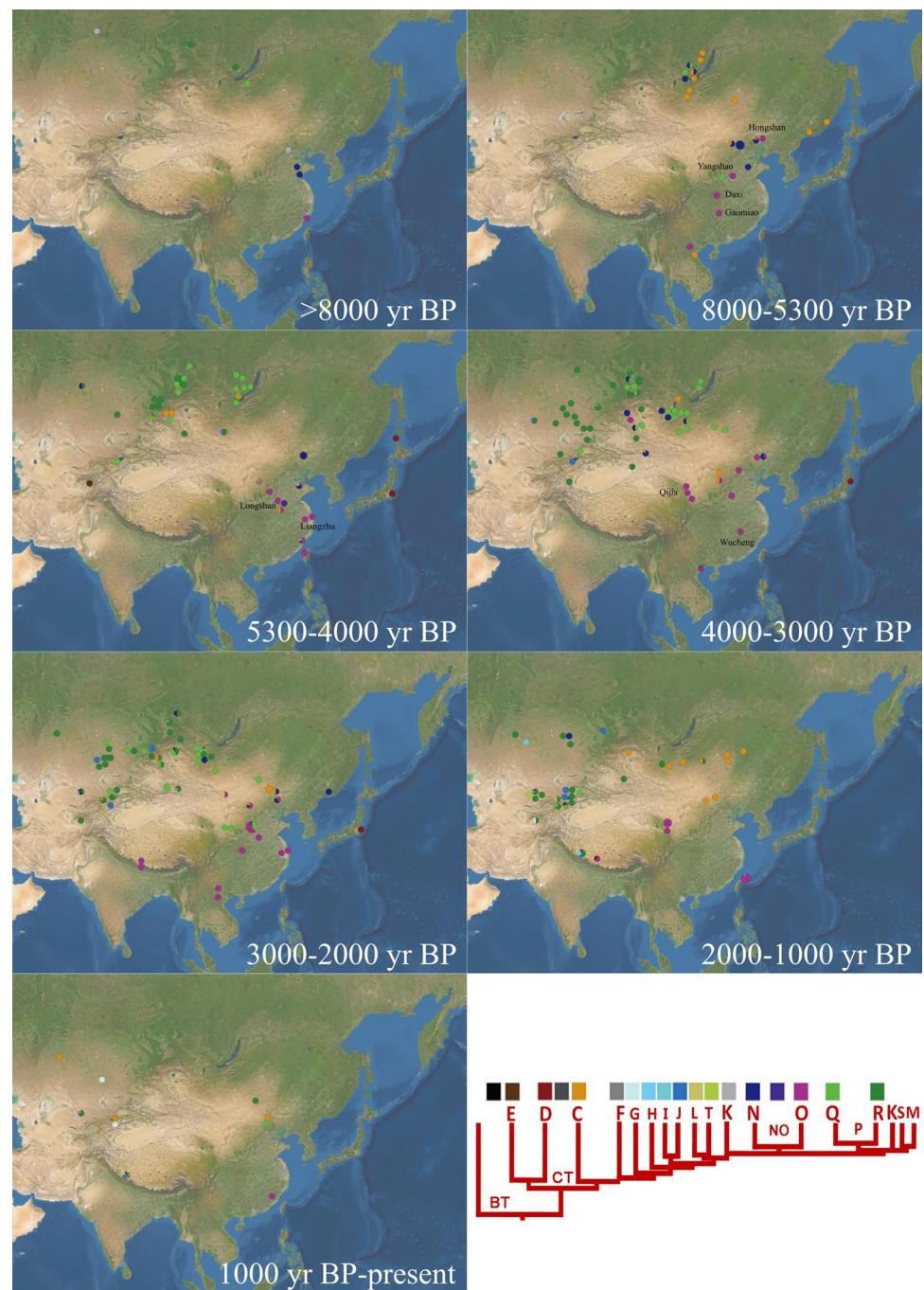
The people who created the Yangshao culture (7000–5300 year BP) are famously known as an important origin of Chinese, and their westward migration participated in the formation of Tibeto-Burman populations. Approximately 6000 years ago, a cold-dry climate event caused farmers who carried O2a2b1a1a-F5 (Oα) from Yangshao culture in the middle Yellow River basin to migrate southwestward to the Tibetan Plateau in the first wave and mixed with local Paleolithic D-M174 populations, together forming the first layer of modern Tibeto-Burman populations (Wang et al. 2018b, b). The Yangshao people possibly induced the southern/eastern haplogroup O2a1b1a1a1a-F11 (Oγ) as well, viewed from whole-genome data of a person from the Banpo site (Zhang

 Springer

**a**



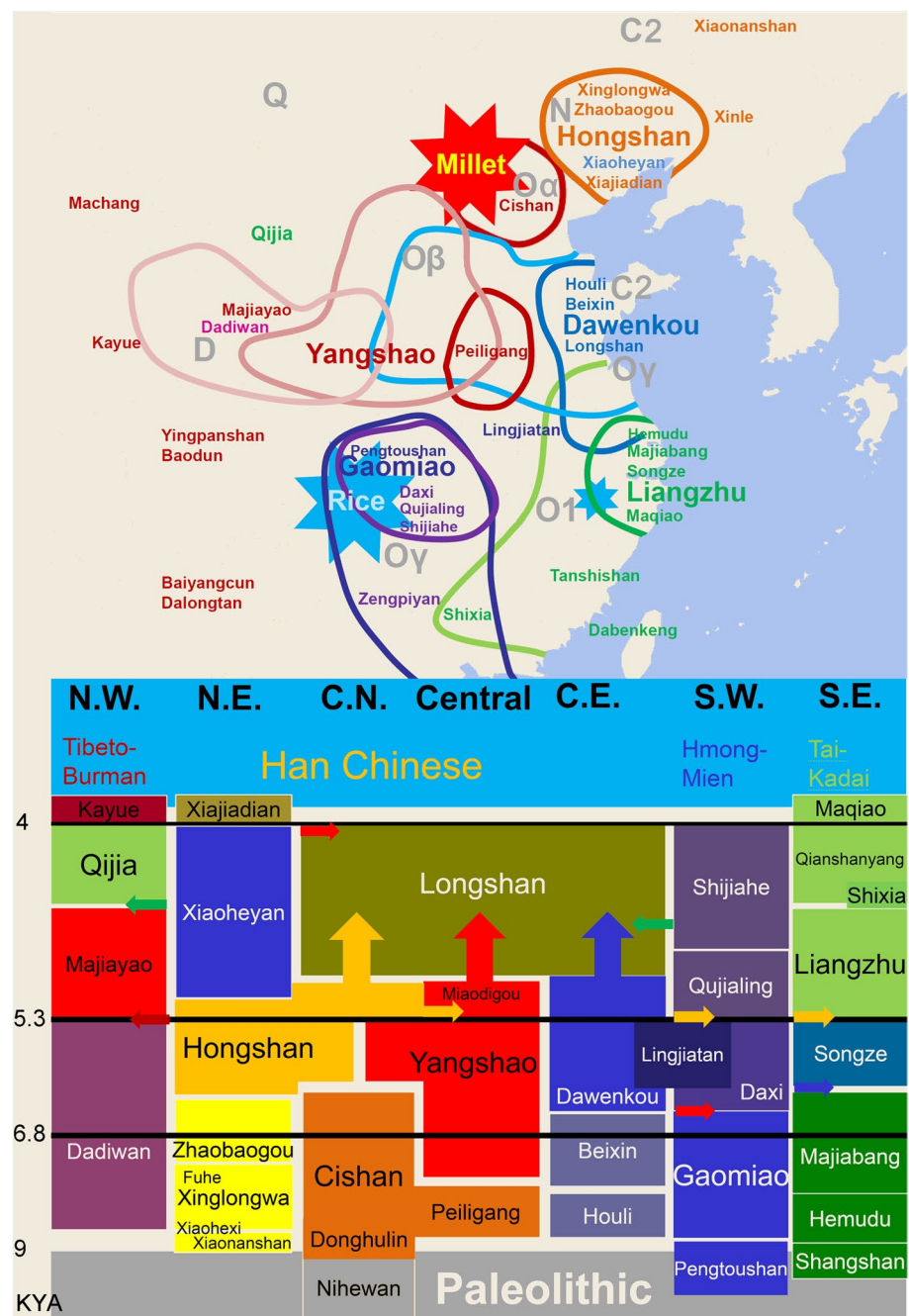
**Fig. 4** NRY haplogroup distribution changes through time. Data collected from Allen Ancient DNA Resource version 44.3 (AADR 2021), previous studies (Li et al. 2007, 2010a, 2013, 2017, 2018, 2020; Kim et al. 2010; Zhao et al. 2010b; Li 2012; Cui et al. 2013; Fu et al. 2014; Hollard et al. 2014; Raghavan et al. 2014; Allentoft et al. 2015; Cui et al. 2015; Dong et al. 2015; Gao et al. 2015; Zhao et al. 2015; Jeong et al. 2016, 2018; Liu 2016; Lkhagvasuren et al. 2016; Kanzawa-Kiriyama et al. 2017; Ning 2017; Unterländer et al. 2017; Yang et al. 2017, 2020; Zhang et al. 2017, 2018b; de Barros Damgaard et al. 2018a, b; Hollard et al. 2018; Lipson et al. 2018; McColl et al. 2018; Flegontov et al. 2019; Harney et al. 2019; Järve et al. 2019; Narasimhan et al. 2019; Ning et al. 2019, 2020; Sikora et al. 2019; Wang et al. 2020a; Yu et al. 2020), and unpublished results from the Gaomiao site of the Gaomiao culture



ancestors of Han Chinese, the Hongshan people also made genetic contributions to Han Chinese in addition to merely cultural practices. Hongshan ancestors spread haplogroup O across China approximately 5300 years ago, with diffusion of Sinitic languages. More than 40% of present Han Chinese can be assigned to clades spread by Hongshan ancestors NRY (Yan et al. 2014). Haplotype O2a-M324 occurred among aristocrats at both the Niheliang site and the Banlashan site, the two largest mausoleums of Hongshan. Lineage O2a2b1-M134 continuously existed

during later Longshan and Qijia cultures (Shimao region in the great curve of the Yellow River) at approximately 4000 year BP in central northern China, the heartland of Chinese civilization (Fig. 4). As Hongshan populations occupied northeastern China, some earlier populations of local Zhaobaogou culture carrying haplogroup N moved outward (unpublished data from Zhaobaogou site). These immigrants might have developed Finno-Uralic languages afterward (Li 2018; Li and Jin 2020). Residents who remained in the region retained haplogroup

**Fig. 5** Geography, chronology, and genetics of archaeological cultures in Neolithic East Asia



N in northeastern China. Another crucial lineage of Han Chinese is Q1a1a-M120 (Sun et al. 2019). One Yangshao person was genotyped Q1a1a-M120, suggesting its early contribution to Han Chinese at least 6000 years ago (Ning et al. 2020). Q1a1a-M120 was also found in aristocrats and commoners at the Hengbei site dated to 3000 year BP, but was not seen in slaves (Zhao et al. 2014), and it therefore participated in shaping the genetic structure of northern Han Chinese. Southern lineage C2b has made important genetic contributions to Han Chinese as well and has been

documented in the Shimao people of northern Shaanxi Province in the latest research (Ning et al. 2020).

In summary, millet farming in northern China resulted in the gathering of diverse lineages of people and in the development of the Sino-Tibetan ethnic groups in the Neolithic Age, with a leading lineage of haplogroup O2a-M324 and subgroup M117.

## Rice farming centralized ethnic groups

Rice farming in Southern China might have started earlier than millet farming in the north. Farmers near the middle and lower branches of the Yangtze River started domesticating rice more than 10,000 years ago (Zhao 2010), leading to the rise of southern Chinese civilizations. These farmers may have developed some of the linguistic families now prevalent in East Asia and Southeast Asia, as there are close linguistic relationships among Hmong-Mien, Austroasiatic, Tai-Kadai, and Austronesian languages (Pan 1995) and close genetic relationships between the Neolithic Yangtze populations and the present southern language speakers (Li et al. 2007; Zhang et al. 2015).

In the middle reaches of the Yangtze River, the ancestors of Hmong-Mien-speaking populations created the Gaomiao culture (7800–6500 year BP) and the later Daxi culture (6500–5300 year BP) (Li et al. 2007). The Gaomiao people used white pottery as ritual vessels, invented octagonal star patterns symbolizing the solar calendar, and worshiped a bird (phoenix) and the sun (Hunan Provincial Institute of Cultural Relics and Archaeology 2006). Semicircle (huang) and camber edge axe (yue) jade ware, which were important to later Chinese civilizations, were found in burials at the Gaomiao site, suggesting that initial national civilization factors occurred. The Chengtoushan site of Gaomiao culture is considered the oldest city in China due to its comprehensive layout, suggesting an early stage of civilization (Yuan 2010). These cultural factors were partially absorbed by Han Chinese and were also inherited by local minor ethnic groups, such as the Miao (Hmong). In addition to cultural similarities, high frequencies of O2a2a1a2-M7 in the present Hmong-Mien populations and the Daxi people developed out of Gaomiao provide strong evidence for shared genetic affinities. (Li et al. 2007, 2008). In addition to O2a-M7, haplogroup O $\gamma$ -002611 was found in the remains of an early leader of Chengtoushan city (unpublished data), indicating a possible relationship between the Gaomiao culture and the Dawenkou culture in central eastern China, in which O $\gamma$  was also frequent.

Close to the Hmong-Mien (Xia et al. 2019), a population living in southern China (Liao 2015; Zhang et al. 2015) developed Austroasiatic languages and migrated southward. These people were possibly located near the middle reaches of the Yangtze River Valley (Peiros and Shnirelman 1998) and the Sichuan Basin (Peiros 2011). Although ancient DNA sequences from Sichuan Province are lacking, the Wucheng people approximately 3000 years ago in Jiangxi Province, southeast of Sichuan, carried the O1b1a1a-M95 lineage (Li et al. 2007), which is distributed in Austroasiatic populations at high frequency (Zhang et al. 2015).

As a language family mostly distributed on islands, the origin of Austronesians has long been debated, with several

hypotheses from different disciplines (Fan et al. 2018). A recent study traces Austronesian speakers to ancient southern China with whole-genome sequencing (Yang et al. 2020). Considering the close relationships between Tai-Kadai populations and Austronesian populations (Li et al. 2008; Kutanan et al. 2018), the two families may have derived from the same or intimate ancestors. The linguistic hypothesis is that Tai-Kadai is derived from Austronesian (Sagart 2005). Our genetic study indicated that the earlier Majiabang populations (7000–5900 year BP) (Mou and Wei 1978; Luojiang Archaeology Team 1981) in the Yangtze Delta may have developed the Austronesian linguistic family, whereas the later in situ Songze (5900–5300 year BP) (Mou and Wei 1978) and Liangzhu populations (5300–4300 year BP) (Mou and Wei 1978), later called Bai Yue in the historical period, developed into the Tai-Kadai linguistic family (Li et al. 2007, 2008). The Liangzhu culture is known for glorious jade artifacts, especially ritual square tubes (cong), apertured plates (bi), and camber edge axes (yue). The Liangzhu culture was thought to have become one of the earliest state societies in East Asia (Renfrew and Liu 2018) due to its fortified towns and communities, advanced science and technologies, clear stratification of society, clear distinction between urban and rural areas, the presence of force, and a theocratic government (Zhao 2017). The NRY lineages observed among the Liangzhu remains were mostly O1a-M119 (Li et al. 2007), which are prevalent in Tai-Kadai and Austronesian speakers (Li 2005).

Different from the major single ethnic origin in northern China, southern China was the site of at least three ethnic origins, with haplogroups O1a-M119 in the ancestors of Austronesians and Tai-Kadai speakers in the lower reach of the Yangtze River, haplogroups O2a-M7 in the ancestors of Hmong-Mien speakers in the middle reach of the Yangtze River, and haplogroups O1b-M95 in the ancestors of Austroasiatic speakers in the upper reach.

Climate changes, expansion, and wars between ethnic groups resulted in outmigration and admixture. In the early Neolithic Age, Austronesian, Austroasiatic, Finno-Uralic, and Paleosiberian speakers migrated out of East Asia, and the genetic structures of the populations continued to drift but were isolated from those of other East Asians. Those populations remaining in East Asia, e.g., Sino-Tibetan, Hmong-Mien, and Tai-Kadai ancestors, interacted frequently, which allowed their languages to mingle extensively. Eventually, these East Asian languages evolved together in the Bronze Age, becoming analytic languages without tense or voice (Huang and Li 2017; Li 2018; Li and Jin 2020). Intensive population admixture started in East Asia approximately 5,300 years ago.

## Admixing into a nation (< 5300 year BP)

Although there is not yet a consensus of the precise start of "China" as a political entity, the traditional concept of "five thousand years of Chinese civilization" might be reasonable, judging from archaeology and genetics. Archaeological cultures dramatically shifted from approximately 5300 years ago to 4000 years ago, coinciding with the legend of wars between ancient tribes. Admixture is also reflected by excavated artifacts. Approximately 5300 years ago, jade artifacts with religious purposes belonging to the Hongshan culture spread southward to central northern China (Guo 2017). A domino effect occurred among the cultures in East Asia immediately after 5300 years ago. The Yangshao culture in the central north migrated to the west. The Dadiwan culture in the west mixed with the Yangshao culture, forming the Majiayao culture. The Lingjiatan culture, the dominant culture in southern China, was ruined. The Daxi culture in the southwest transitioned into the Qujialing culture. The Songze culture in the southeast transitioned into the Liangzhu culture. A rapid population admixture drove this process, probably along with the integration of the ethnic groups into a nation, which represent the earliest start of China as an entity. The NRY haplogroup O $\alpha$  of the Hongshan leaders was then carried to most regions of China.

Subsequently, population interaction became increasingly frequent. For example, approximately 4500 years ago, Liangzhu-style jade artifacts and their associated rituals spread northward to Shandong Province in central eastern China (Huang 1992), westward to central northern China (Han 2010), and finally arrived in northwestern China (Huang 1992) and might have accompanied by demic diffusion. Along these migration routes, southern population lineages mixed with the northern population lineages. Admixture was reflected by the continuous distribution of genetic distance as well (Wang et al. 2020a) and a smooth north-to-south gradient in genetic structure (Wen et al. 2004; Wang et al. 2020a; Yang et al. 2020). Compared with the northward migration, the initial southward spreading of haplogroup O $\alpha$  from Hongshan played the most important role. Therefore, we speculated that the Hongshan people were major participants in the establishment of the Chinese civilization.

Accelerated admixture and homogenization occurred after unification of centralized dynasties. As a nation, the history of China can be classified into three periods, i.e., idolatry dynasties (3300–2000 BC), tributary dynasties (2000–221 BC), and empery dynasties (221 BC–1912 AD). In particular, the empery dynasties markedly accelerated the population admixture, and three large-scale southward migrations (Ge et al. 1993) in the Eastern Jin, late Tang, and Southern Song dynasties caused by wars made the paternal

genetic structure of southern Han Chinese almost the same as that of northern Han Chinese but significantly different from that of the southern ethnic minorities. In contrast, the maternal genetic structure of southern Han Chinese showed signs of admixture of northern Han Chinese and southern ethnic minorities (Wen et al. 2004). The admixture occurring in Han Chinese with a north-to-south trend was not only caused by war. Nomadic proto-Mongolian populations from North Asia dispersed Altaic languages and NRY C2a lineages (Wei et al. 2017; Zhang et al. 2018b; Li et al. 2020) southward to Chinese populations. The southern East Asian lineage O2a2b1 spread to the north as well (Wang et al. 2020a). West-to-east migration is also evidenced by NRY lineage R in the Mongol Empire (Lkhagvasuren et al. 2016) and mitochondrial lineage U in central northern China during the Sui Dynasty (598 AD) (Xie et al. 2007).

The rapid development of East Asian society over the past several centuries has made migration much easier and more frequent, which has weakened the association among genetics, cultures, and languages. Although geographically distant present-day East Asian populations are still separated in the PCA plot, the distance between neighboring groups has vanished (Fig. 3a). The estimated composition of ancestry of ancient and present-day East Asians shows approximate similarity and detailed differences (Fig. 3b, c), suggesting gradual change but not substitution in populations. Major lineages in southern China and Southeast Asia, O-M95, O-M7, and O-M119, occur in Daur people in northernmost China, but the Daur people are assumed to descend from Khitan or Mongol and never visited southern China (Wang et al. 2018a). Modern ethnic groups in southwestern China, which are thought to be mostly isolated, carry a high frequency of genetic components from populations of different linguistic families (Guo et al. 2019; Xu et al. 2019; Lu et al. 2020). Homogenization makes the nation united with not only cultural identity but also genetic affinities.

## Conclusions and perspectives

We generally reviewed how ancient Africans became current East Asians by researching the chronology. AMH originated in Africa and then trekked through South Asia to Southeast Asia and spread northward to East Asia with introgression with archaic humans. Other routes of migration might also have contributed to the East Asian gene pool. Populations increased after the Last Glacial Maximum. As agricultural production of millet and rice developed, populations gathered into ethnic groups represented by uniform languages and cultures, namely, the Sino-Tibetan speakers in the Cis-han culture and the descendant Yangshao and Hongshan cultures, the Hmong-Mien speakers in the Gaomiao culture and Daxi culture, the Austroasiatic speakers in the Sichuan

Basin, the Austronesian speakers in the Majiabang culture, the Tai-Kadai speakers in the Liangzhu culture, and the ancestors of the Finnish-Uralic speakers in the Zhao-baogou culture. The Hongshan people moved southward at approximately 5300 year BP to dominate the formation of the ancient Han Chinese. Populations of other linguistic families either mixed with the ancient Han Chinese or migrated to remote regions to become modern minorities. Nomadic lineages merged into the Chinese gene pool during the empery dynasties and contributed to the structure of the modern Chinese people. Currently, population admixture is occurring on an even larger scale. In conclusion, we argue that a long history of population admixture spanning more than 5000 years has made the ethnic groups in mainland East Asia a nation.

The rapid development of molecular biological technologies makes it possible to assess genetic history on a fine scale. Data were obtained on scales ranging from one person to entire populations, from one SNP to whole-genome sequences, and from uniparental lineage to genomics and even to phenomics to depict a panoramic perspective of human beings. However, understanding the natural history of humans cannot rely solely on genetics and must also incorporate data on cultures and languages interacting with our genomes.

**Funding** This work was supported by B&R Joint Laboratory of Eurasian Anthropology (18490750300), the National Key R&D Program of China (2020YFE0201600), and the National Natural Science Foundation of China (91731303, 31671297). The European Research Council project (ERC-2019-AdG-TRAM-883700).

## Declarations

**Conflict of interest** Both authors declare that they have no conflicts of interest.

**Ethical approval** This article does not contain any studies with human participants performed by any of the authors.

## References

- AADR (2021) Allen ancient DNA resource version 44.3. <https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data>. Accessed 20 Jan 2021
- Allentoft ME, Sikora M, Sjögren K-G, Rasmussen S, Rasmussen M, Stenderup J, Damgaard PB, Schroeder H, Ahlström T, Vinner L (2015) Population genomics of bronze age Eurasia. *Nature* 522:167–172
- An ZM, Chang KC (1979) The neolithic archaeology of China a brief survey of the last thirty years. *Early China* 5:35–45
- Anderson B (2006) Imagined communities: reflections on the origin and spread of nationalism. Verso Books, London
- Bai F, Zhang XL, Ji XP, Cao P, Feng XT, Yang RW, Peng MS, Pei SW, Fu QM (2020) Paleolithic genetic link between Southern China and Mainland Southeast Asia revealed by ancient mitochondrial genomes. *J Hum Genet* 65:1125–1128
- Bergström A, McCarthy SA, Hui R, Almarri MA, Ayub Q, Danecek P, Chen Y, Felkel S, Hallast P, Kamm J (2020) Insights into human genetic variation and population history from 929 diverse genomes. *Science* 367:eaay5012
- Bräuer G, Stringer C (1997) Models, polarization, and perspectives on modern human origins. Conceptual issues in modern human origins research. Aldine de Gruyter, New York, pp 191–201
- Chen GQ (2010) A discussion on the origin of Hongshan culture. *Res China's Front Archaeol* 1:51–54
- Chen GQ (2012) Re-exploration on the provenance of Hongshan culture. *Res China's Front Archaeol* 1:83–87
- Chow BS (1981) The animal remains discovered at Cishan Village, Wu'an, Hebei Province. *Acta Archaeol Sin* 3:339–347
- Clarkson C, Harris C, Li B, Neudorf CM, Roberts RG, Lane C, Norman K, Pal J, Jones S, Shipton C (2020) Human occupation of northern India spans the Toba super-eruption ~ 74,000 years ago. *Nat Commun* 11:1–10
- Cui YQ (2015) Analysis of characteristic of Xinglongwa Culture Jade. *J Chifeng Univ (philosophy and Social Science Chinese Edition)* 36:5–10
- Cui YQ, Li HJ, Ning C, Zhang Y, Chen L, Zhao X, Hagelberg E, Zhou H (2013) Y Chromosome analysis of prehistoric human populations in the West Liao River Valley, Northeast China. *BMC Evol Biol* 13:216
- Cui YQ, Song L, Wei D, Pang YH, Wang N, Ning C, Li CM, Feng BX, Tang WT, Li HJ, Ren YS, Zhang CC, Huang YY, Hu YW, Zhou H (2015) Identification of kinship and occupant status in Mongolian noble burials of the Yuan Dynasty through a multidisciplinary approach. *Philos Trans R Soc Lond B Biol Sci* 370:20130378
- de Barros Damgaard P, Marchi N, Rasmussen S, Peyrot M, Renaud G, Korneliussen T, Moreno-Mayar JV, Pedersen MW, Goldberg A, Usmanova E (2018a) 137 ancient human genomes from across the Eurasian steppes. *Nature* 557:369–374
- de Barros Damgaard P, Martiniano R, Kamm J, Moreno-Mayar JV, Kroonen G, Peyrot M, Barjamovic G, Rasmussen S, Zacho C, Baimukhanov N (2018b) The first horse herders and the impact of early Bronze Age steppe expansions into Asia. *Science* 360:eaar7711
- Dong Y, Li CX, Luan FS, Li ZG, Li HJ, Cui YQ, Zhou H, Malhi RS (2015) Low mitochondrial DNA diversity in an ancient population from China: insight into social organization at the Fujia Site. *Hum Biol* 81:71–84
- Fan ZQ, Deng XH, Wang CC (2018) Languages and genes: comment on the origin and dispersal of Austronesian. *Acad Mon* 50:175–184
- Flegontov P, Altınışık NE, Changmai P, Rohland N, Mallick S, Adamski N, Bolnick DA, Broomandkhoshbacht N, Candilio F, Cullen BJ (2019) Palaeo-Eskimo genetic ancestry and the peopling of Chukotka and North America. *Nature* 570:236–240
- Fu QM, Li H, Moorjani P, Jay F, Slepchenko SM, Bondarev AA, Johnson PL, Aximu-Petri A, Prüfer K, de Filippo C (2014) Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* 514:445–449
- Gao SZ, Zhang Y, Wei D, Li HJ, Zhao YB, Cui YQ, Zhou H (2015) Ancient DNA reveals a migration of the ancient Di-qiang populations into Xinjiang as early as the early Bronze Age. *Am J Phys Anthropol* 157:71–80
- Gao X, Peng F, Fu QM, Li F (2017) New progress in understanding the origins of modern humans in China. *Sci China Earth Sci* 60:2160–2170

- Gayden T, Cadenas AM, Regueiro M, Singh NB, Zhivotovsky LA, Underhill PA, Cavalli-Sforza LL, Herrera RJ (2007) The Himalayas as a directional barrier to gene flow. *Am J Hum Genet* 80:884–894
- Ge JX, Cao SJ, Wu SD (1993) Jianming zhongguo yimin shi. Fujian Renmin Chubanshe, Fuzhou, Fujian
- Genomes Project Consortium (2015) A global reference for human genetic variation. *Nature* 526:68
- Groves CP (1989) A regional approach to the problem of the origin of modern humans in Australasia. *Hum Revol* 1:274–285
- Guo DS (2017) Yu yu Tao—Zhongguo Shiqian Yu Wenhua Yanjiu Tichu de Xin Ketu. Guo Da-Shun Kaogu Wenji. Liaoning People's Publishing House, Liaoning, pp 176–180
- Guo J, Xu B, Li L, He G, Zhang H, Cheng HZ, Ba J, Yang X, Wei L, Hu R, Wang CC (2019) Paternal Y chromosomal genotyping reveals multiple large-scale admixtures in the formation of Lolo-Burmese-speaking populations in southwest China. *Ann Hum Biol* 46:581–588
- Hammer MF, Karafet TM, Park H, Omoto K, Harihara S, Stoneking M, Horai S (2006) Dual origins of the Japanese: common ground for hunter-gatherer and farmer Y chromosomes. *J Hum Genet* 51:47–58
- Han JY (2010) Liangzhu, Taosi and Erlitou—the evolution route of the Early Chinese civilization. *Archaeology* 11:9
- Harney É, Nayak A, Patterson N, Joglekar P, Mushrif-Tripathy V, Mallick S, Rohland N, Sedig J, Adamski N, Bernardos R (2019) Ancient DNA from the skeletons of Roopkund Lake reveals Mediterranean migrants in India. *Nat Commun* 10:1–10
- Hollard C, Keyser C, Giscard PH, Tsagaan T, Bayarkhuu N, Bemmann J, Crubezy E, Ludes B (2014) Strong genetic admixture in the Altai at the Middle Bronze Age revealed by uniparental and ancestry informative markers. *Forensic Sci Int Genet* 12:199–207
- Hollard C, Zvenigorosky V, Kovalev A, Kiryushin Y, Tishkin A, Lazaretov I, Crubézy E, Ludes B, Keyser C (2018) New genetic evidence of affinities and discontinuities between bronze age Siberian populations. *Am J Phys Anthropol* 167:97–107
- Hong SB, Jin HJ, Kwak KD, Kim W (2006) Y-chromosome haplogroup O3–M122 variation in East Asia and its implications for the peopling of Korea. *Genes Genom* 28:1–8
- Huang TM (1992) Liangzhu—a late Neolithic jade-yielding culture in southeastern coastal China. *Antiquity* 66:75–83
- Huang YZ, Li H (2017) Correlation between genetic structure and linguistic phylogeny in East Asia. In: Xu D, Li H (eds) *Languages and genes in Northwestern China and adjacent regions*. Springer, Singapore, pp 13–36
- Huang DX, Wang GY, Yang QE (2006) Haplogroup diversity of 8 biallelic markers on human Y chromosome in Wuhan Han population. *Chin J Forensic Med* 21:13–17
- Huerta Sánchez E, Jin X, Bianba ZM, Peter BM, Vinckenbosch N, Liang Y, Yi X, He MZ, Somel M, Ni PX (2014) Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature* 512:194–197
- HUGO Pan-Asian SNP Consortium (2009) Mapping human genetic diversity in Asia. *Science* 326:1541–1545
- Hunan Provincial Institute of Cultural Relics and Archaeology (2006) Neolithic Gaomiao site in Hongjiang City. *Hunan Kaogu* 7(9–15):99–100
- Ingman M, Kaessmann H, Pääbo S, Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. *Nature* 408:708–713
- Järve M, Saag L, Scheib CL, Pathak AK, Montinaro F, Pagan L, Flores R, Guellil M, Saag L, Tambets K (2019) Shifts in the genetic landscape of the Western Eurasian Steppe associated with the beginning and end of the Scythian dominance. *Curr Biol* 29:2430–2441 (e2410)
- Jeong C, Ozga AT, Witonsky DB, Malmstrom H, Edlund H, Hofman CA, Hagan RW, Jakobsson M, Lewis CM, Aldenderfer MS, Di Rienzo A, Warinner C (2016) Long-term genetic stability and a high-altitude East Asian origin for the peoples of the high valleys of the Himalayan arc. *Proc Natl Acad Sci USA* 113:7485–7490
- Jeong C, Wilkin S, Amgalantugs T, Bouwman AS, Taylor WTT, Hagan RW, Bromage S, Tsolmon S, Trachsel C, Grossmann J (2018) Bronze Age population dynamics and the rise of dairy pastoralism on the eastern Eurasian steppe. *Proc Natl Acad Sci USA* 115:E11248–E11255
- Jeong C, Balanovsky O, Lukianova E, Kahbatkyzy N, Flegontov P, Zaporozhchenko V, Immel A, Wang C-C, Ixan O, Khussainova E (2019) The genetic history of admixture across inner Eurasia. *Nat Ecol Evol* 3:966–976
- Jobling M, Hurler M, Tyler Smith C (2003) *Human evolutionary genetics: origins, peoples & disease*. Garland Science, New York
- Kanzawa-Kiriyama H, Kryukov K, Jinam TA, Hosomichi K, Saso A, Suwa G, Ueda S, Yoneda M, Tajima A, Shinoda K-i (2017) A partial nuclear genome of the Jomons who lived 3000 years ago in Fukushima, Japan. *J Hum Genet* 62:213–221
- Karafet T, Xu L, Du R, Wang W, Feng S, Wells R, Redd AJ, Zegura SL, Hammer MF (2001) Paternal population history of East Asia: sources, patterns, and microevolutionary processes. *Am J Hum Genet* 69:615–628
- Karafet TM, Hallmark B, Cox MP, Sudoyo H, Downey S, Lansing JS, Hammer MF (2010) Major east–west division underlies Y chromosome stratification across Indonesia. *Mol Biol Evol* 27:1833–1844
- Kayser M, Choi Y, van Oven M, Mona S, Brauer S, Trent RJ, Suarkia D, Schiefenhövel W, Stoneking M (2008) The impact of the Austronesian expansion: evidence from mtDNA and Y chromosome diversity in the Admiralty Islands of Melanesia. *Mol Biol Evol* 25:1362–1374
- Ke YH, Su B, Xiao JH, Chen H, Huang W, Chen Z, Chu JY, Tan JZ, Jin L, Lu DR (2001) Y-chromosome haplotype distribution in Han Chinese populations and modern human origin in East Asians. *Sci China Ser C Life Sci* 44:225–232
- Kim W, Yoo TK, Kim SJ, Shin DJ, Tyler-Smith C, Jin HJ, Kwak KD, Kim ET, Bae Y-S (2007) Lack of association between Y-chromosomal haplogroups and prostate cancer in the Korean population. *PLoS ONE* 2:e172
- Kim K, Brenner CH, Mair VH, Lee KH, Kim JH, Gelegdorj E, Batbold N, Song YC, Yun HW, Chang EJ (2010) A western Eurasian male is found in 2000-year-old elite Xiongnu cemetery in Northeast Mongolia. *Am J Phys Anthropol* 142:429–440
- Kutanan W, Kampuansai J, Brunelli A, Ghirotto S, Pittayaporn P, Ruangchai S, Schroder R, Macholdt E, Srikummool M, Kangwanpong D, Hubner A, Arias L, Stoneking M (2018) New insights from Thailand into the maternal genetic history of Mainland Southeast Asia. *Eur J Hum Genet* 26:898–911
- Kutanan W, Kampuansai J, Srikummool M, Brunelli A, Ghirotto S, Arias L, Macholdt E, Hübner A, Schröder R, Stoneking M (2019) Contrasting paternal and maternal genetic histories of Thai and Lao populations. *Mol Biol Evol* 36:1490–1506
- Lane CS, Chorn BT, Johnson TC (2013) Ash from the Toba supereruption in Lake Malawi shows no volcanic winter in East Africa at 75 ka. *Proc Natl Acad Sci USA* 110:8025–8029
- Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, Kirsanow K, Sudmant PH, Schraiber JG, Castellano S, Lipson M (2014) Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* 513:409–413
- Li H (2005) Genetic structure of austro-tai populations. Dissertation, Fudan University
- Li HJ (2012) Y-chromosome genetic diversity of the ancient northern Chinese populations. Dissertation, Jilin University

- Li H (2018) As ilustrações mais simples da evolução humana. Ásia Associação de Estados de Humanidade e Natureza, Macau
- Li TJ (2019) Genomic study of the ancient populations from the period of YangShao to LongShan culture in Central Plain. Dissertation, Jilin University
- Li H, Jin WL (2020) Human origin and migration. Shanghai Science and Technology Education Publishing House, Shanghai
- Li H, Pan WY, Wen B, Yang NN, Jin JZ, Jin L, Lu DR (2003) Origin of Hakka and Hakkane: a genetics analysis. *Yi Chuan Xue Bao* 30:873–880
- Li H, Huang Y, Mustavich LF, Zhang F, Tan JZ, Wang LE, Qian J, Gao MH, Jin L (2007) Y chromosomes of prehistoric people along the Yangtze River. *Hum Genet* 122:383–388
- Li H, Wen B, Chen SJ, Su B, Pramoongjago P, Liu YF, Pan SL, Qin ZD, Liu WH, Cheng X, Yang N, Li X, Tran D, Lu D, Hsu MT, Deka R, Marzuki S, Tan CC, Jin L (2008) Paternal genetic affinity between Western Austronesians and Daic populations. *BMC Evol Biol* 8:146
- Li DN, Sun YT, Lu Y, Mustavich LF, Ou CY, Zhou ZJ, Li SL, Jin L, Li H (2010b) Genetic origin of Kadai-speaking Gelong people on Hainan island viewed from Y chromosomes. *J Hum Genet* 55:462–468
- Li CX, Li HJ, Cui YQ, Xie CZ, Cai DW, Li WY, Mair VH, Xu Z, Zhang QC, Abuduresule I, Jin L, Zhu H, Zhou H (2010a) Evidence that a West-East admixed population lived in the Tarim Basin as early as the early Bronze Age. *BMC Biol* 8:15
- Li HJ, Zhang Y, Zhao YB, Zhou H (2013) Ancient DNA and Kinship-based Burial Patterns at Han-Jin Dynasty Taojiazhai Site, Qinghai Province, China. *Asian Archaeol* 2:116–125
- Li JW, Zeng W, Zhang Y, Ko AM, Li CX, Zhu H, Fu QM, Zhou H (2017) Ancient DNA reveals genetic connections between early Di-Qiang and Han Chinese. *BMC Evol Biol* 17:239
- Li JW, Zhang Y, Zhao YB, Chen YZ, Ochir A, Sarenbilige ZhuH, Zhou H (2018) The genome of an ancient Rouran individual reveals an important paternal lineage in the Donghu population. *Am J Phys Anthropol* 166:895–905
- Li JW, Cai DW, Zhang Y, Zhu H, Zhou H (2020) Ancient DNA reveals two paternal lineages C2a1a1b1a/F3830 and C2b1b/F845 in past nomadic peoples distributed on the Mongolian Plateau. *Am J Phys Anthropol* 172:402–411
- Li YQ, Yang YC (2019) The excavation of zone III of the xiao nanshan site in raohe county, Heilongjiang in 2015. *Kaogu* 8:3–20
- Liao SY (2015) Genetic study on the paternal origin and prehistoric migration of Austro-Asiatic speaking populations. Dissertation, Anhui University
- Lin YW, Chia-Ling Hsu L, Kuo PL, Huang WJ, Chiang HS, Yeh SD, Hsu TY, Yu YH, Hsiao KN, Cantor RM (2007) Partial duplication at AZFc on the Y chromosome is a risk factor for impaired spermatogenesis in Han Chinese in Taiwan. *Hum Mutat* 28:486–494
- Lipson M, Cheronet O, Mallick S, Rohland N, Oxenham M, Pietrusewsky M, Pryce TO, Willis A, Matsumura H, Buckley H (2018) Ancient genomes document multiple waves of migration in Southeast Asian prehistory. *Science* 361:92–95
- Liu L (2005) The Chinese Neolithic: trajectories to early states. Cambridge University Press
- Liu M (2016) Ancient DNA research on the human remains from Dabaoshan in Inner Mongolia, China. Dissertation, Jilin University
- Liu W, Martínón Torres M, Cai YJ, Xing S, Tong HW, Pei SW, Sier MJ, Wu XH, Edwards RL, Cheng H (2015) The earliest unequivocally modern humans in southern China. *Nature* 526:696–699
- Lkhagvasuren G, Shin H, Lee SE, Tumen D, Kim JH, Kim KY, Kim K, Park AJ, Lee HW, Kim MJ, Choi J, Choi JH, Min NY, Lee KH (2016) Molecular genealogy of a mongol queen's family and her possible kinship with Genghis Khan. *PLoS ONE* 11:e0161622
- Lu HY, Zhang JP, Liu KB, Wu NQ, Li YM, Zhou KS, Ye ML, Zhang TY, Zhang HJ, Yang XY (2009) Earliest domestication of common millet (*Panicum miliaceum*) in East Asia extended to 10,000 years ago. *Proc Natl Acad Sci USA* 106:7367–7372
- Lu JN, Zhang HL, Ren Z, Wang QY, Liu YB, Li YX, He GL, Guo JX, Zhao J, Hu R, Wei LH, Chen G, Huang J, Wang CC (2020) Genome-wide analysis of unrecognised ethnic group chuanqing people revealing a close affinity with southern han chinese. *Ann Hum Biol* 47:465–471
- Luojiajiao Archaeology Team (1981) Excavations at Luojiajiao Site in Tongxiang County of Zhejiang Province. Cultural Relics Publishing House, Beijing
- Macaulay V, Hill C, Achilli A, Rengo C, Clarke D, Meehan W, Blackburn J, Semino O, Scozzari R, Cruciani F (2005) Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* 308:1034–1036
- Mallick S, Li H, Lipson M, Mathieson I, Gymrek M, Racimo F, Zhao MY, Chennagiri N, Nordenfelt S, Tandon A (2016) The Simons genome diversity project: 300 genomes from 142 diverse populations. *Nature* 538:201–206
- McColl H, Racimo F, Vinner L, Demeter F, Gakuhari T, Moreno-Mayar JV, Van Driem G, Wilken UG, Seguin-Orlando A, De la Fuente CC (2018) The prehistoric peopling of Southeast Asia. *Science* 361:88–92
- Meyer M, Kircher M, Gansauge M-T, Li H, Racimo F, Mallick S, Schraiber JG, Jay F, Prüfer K, De Filippo C (2012) A high-coverage genome sequence from an archaic Denisovan individual. *Science* 338:222–226
- Mou Y, Wei Z (1978) Majiabang wenhua he liangzhu wenhua—taihu liuyu yuanshi wenhua de fenqi wenti. *Wenwu* 4:67–73
- Narasimhan VM, Patterson N, Moorjani P, Rohland N, Bernardos R, Mallick S, Lazaridis I, Nakatsuka N, Olalde I, Lipson M (2019) The formation of human populations in south and central Asia. *Science* 365:eaat7487
- Ning C (2017) Paleogenomic research of ancient populations from northern china. Dissertation, Jilin University
- Ning C, Wang CC, Gao SZ, Yang Y, Zhang X, Wu XY, Zhang F, Nie ZZ, Tang YP, Robbeets M, Ma J, Krause J, Cui YQ (2019) Ancient genomes reveal yamaya-related ancestry and a potential source of Indo-European speakers in iron age Tianshan. *Curr Biol* 29:2526–2532
- Ning C, Li TJ, Wang K, Zhang F, Li T, Wu XY, Gao SZ, Zhang QC, Zhang H, Hudson MJ, Dong GH, Wu SH, Fang YM, Liu C, Feng CY, Li W, Han T, Li R, Wei J, Zhu YG, Zhou YW, Wang CC, Fan SY, Xiong ZL, Sun ZY, Ye ML, Sun L, Wu XH, Liang FW, Cao YP, Wei XT, Zhu H, Zhou H, Krause J, Robbeets M, Jeong C, Cui YQ (2020) Ancient genomes from northern China suggest links between subsistence changes and human migration. *Nat Commun* 11:2700
- Pagani L, Lawson DJ, Jagoda E, Mörseburg A, Eriksson A, Mitt M, Clemente F, Hudjashov G, DeGiorgio M, Saag L (2016) Genomic analyses inform on migration events during the peopling of Eurasia. *Nature* 538:238–242
- Pan WY (1995) Supporting materials to Sino-Austronesian hypothesis. *J Chin Linguist Monogr Ser* 9:113–144
- Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D (2012) Ancient admixture in human history. *Genetics* 192:1065–1093
- Peiros I (2011) Some thoughts on the problem of the Austro-Asiatic homeland. *J Lang Relatsh* 6:101–114
- Pejros I, Shnirelman V (1998) 16 Rice in Southeast Asia: a regional interdisciplinary approach. *Archaeol Lang Correl Archaeol Linguist Hypothes* 2:379

- Peng MS, Shi NN, Yao YG, Zhang YP (2015) Caveats about interpretation of ancient chicken mtDNAs from northern China. *Proc Natl Acad Sci USA* 112:E1970–E1971
- Peters J, Lebrasseur O, Best J, Miller H, Fothergill T, Dobney K, Thomas RM, Maltby M, Sykes N, Hanotte O (2015) Questioning new answers regarding Holocene chicken domestication in China. *Proc Natl Acad Sci USA* 112:E2415–E2415
- Piazza A (1998) Towards a genetic history of China. *Nature* 395:637–639
- Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, Heinze A, Renaud G, Sudmant PH, de Filippo C, Li H, Mallick S, Dannemann M, Fu QM, Kircher M, Kuhlwilm M, Lachmann M, Meyer M, Ongyerth M, Siebauer M, Theunert C, Tandon A, Moorjani P, Pickrell J, Mullikin JC, Vohr SH, Green RE, Hellmann I, Johnson PLF, Blanche H, Cann H, Kitzman JO, Shendure J, Eichler EE, Lein ES, Bakken TE, Golovanova LV, Doronichev VB, Shunkov MV, Derevianko AP, Viola B, Slatkin M, Reich D, Kelso J, Pääbo S (2013) The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* 505:43–49
- Raghavan M, Skoglund P, Graf KE, Metspalu M, Albrechtsen A, Moltke I, Rasmussen S, Stafford TW Jr, Orlando L, Metspalu E (2014) Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* 505:87–91
- Rampino MR, Self S (1992) Volcanic winter and accelerated glaciation following the Toba super-eruption. *Nature* 359:50–52
- Rampino MR, Self S (1993) Bottleneck in human evolution and the Toba eruption. *Science* 262:1955–1955
- Renfrew C, Liu B (2018) The emergence of complex society in China: the case of Liangzhu. *Antiquity* 92:975–990
- Rodriguez Trelles F, Tarrio R, Ayala FJ (2001) Erratic overdispersion of three molecular clocks: GPDH, SOD, and XDH. *Proc Natl Acad Sci USA* 98:11405–11410
- Sagart L (2005) Tai-Kadai as a subgroup of Austronesian. In: Sagart L, Blench R, Sanchez-Mazas A (eds) *The peopling of East Asia*. RoutledgeCurzon, London and New York, pp 177–181
- Sagart L, Jacques G, Lai YF, Ryder RJ, Thouzeau V, Greenhill SJ, List JM (2019) Dated language phylogenies shed light on the ancestry of Sino-Tibetan. *Proc Natl Acad Sci USA* 116:10317–10322
- Shi H, Zhong H, Peng Y, Dong YL, Qi XB, Zhang F, Liu LF, Tan SJ, Ma RZ, Xiao CJ (2008) Y chromosome evidence of earliest modern human settlement in East Asia and multiple origins of Tibetan and Japanese populations. *BMC Biol* 6:45
- Shou WH, Qiao EF, Wei CY, Dong YL, Tan SJ, Shi H, Tang WR, Xiao CJ (2010) Y-chromosome distributions among populations in Northwest China identify significant contribution from Central Asian pastoralists and lesser influence of western Eurasians. *J Hum Genet* 55:314–322
- Sikora M, Pitulko VV, Sousa VC, Allentoft ME, Vinner L, Rasmussen S, Margaryan A, de Barros DP, de la Fuente C, Renaud G (2019) The population history of northeastern Siberia since the Pleistocene. *Nature* 570:182–188
- Skoglund P, Posth C, Sirak K, Spriggs M, Valentin F, Bedford S, Clark GR, Reepmeyer C, Petchey F, Fernandes D (2016) Genomic insights into the peopling of the Southwest Pacific. *Nature* 538:510–513
- Soares P, Rito T, Pereira L, Richards MB (2016) A genetic perspective on African prehistory. In: Jones S, Stewart B (eds) *Africa from MIS 6–2*. Springer, Dordrecht, pp 383–405
- Stoneking M, Delfin F (2010) The human genetic history of East Asia: weaving a complex tapestry. *Curr Biol* 20:R188–193
- Su BQ, Yin WZ (1981) Guanyu kaoguxue wenhua de quxi leixing wenti. *Wenwu* 5:10–18
- Su B, Xiao JH, Underhill P, Deka R, Zhang WL, Akey J, Huang W, Shen D, Lu DR, Luo JC (1999) Y-Chromosome evidence for a northward migration of modern humans into Eastern Asia during the last Ice Age. *Am J Hum Genet* 65:1718–1724
- Su B, Xiao C, Deka R, Seielstad MT, Kangwanpong D, Xiao J, Lu D, Underhill P, Cavalli-Sforza L, Chakraborty R (2000) Y chromosome haplotypes reveal prehistorical migrations to the Himalayas. *Hum Genet* 107:582–590
- Sun N, Ma PC, Yan S, Wen SQ, Sun C, Du PX, Cheng HZ, Deng XH, Wang CC, Wei LH (2019) Phylogeography of Y-chromosome haplogroup Q1a1a-M120, a paternal lineage connecting populations in Siberia and East Asia. *Ann Hum Biol* 46:261–266
- Sun XF, Wen SQ, Lu CQ, Zhou BY, Curnoe D, Lu HY, Li HC, Wang W, Cheng H, Yi SW, Jia X, Du PX, Xu XH, Lu YM, Lu Y, Zheng HX, Zhang H, Sun C, Wei LH, Han F, Huang J, Edwards RL, Jin L, Li H (2021) Ancient DNA and multimethod dating confirm the late arrival of anatomically modern humans in southern China. *Proc Natl Acad Sci USA* 118:e2019158118
- Tajima A, Hayami M, Tokunaga K, Juji T, Matsuo M, Marzuki S, Omoto K, Horai S (2004) Genetic origins of the Ainu inferred from combined DNA analyses of maternal and paternal lineages. *J Hum Genet* 49:187–193
- Unterländer M, Palstra F, Lazaridis I, Pilipenko A, Hofmanová Z, Groß M, Sell C, Blöcher J, Kirsanow K, Rohland N (2017) Ancestry and demography and descendants of Iron Age nomads of the Eurasian Steppe. *Nat Commun* 8:1–10
- Wang CC, Li H (2013) Inferring human history in East Asia from Y chromosomes. *Investig Genet* 4:11
- Wang W, Wise C, Baric T, Black ML, Bittles AH (2003) The origins and genetic structure of three co-resident Chinese Muslim populations: the Salar, Bo'an and Dongxiang. *Hum Genet* 113:244–252
- Wang CC, Huang YZ, Yu XE, Chen C, Jin L, Li H (2016) Agriculture driving male expansion in neolithic time. *Sci Sin Vitae* 46:468–474
- Wang LX, Lu Y, Zhang C, Wei LH, Yan S, Huang YZ, Wang CC, Mallick S, Wen SQ, Jin L, Xu SH, Li H (2018b) Reconstruction of Y-chromosome phylogeny reveals two neolithic expansions of Tibeto-Burman populations. *Mol Genet Genom* 293:1293–1300
- Wang CZ, Shi MS, Li H (2018a) The origin of Daur from the perspective of molecular anthropology. *J North Minzu Univ* 143:110–117
- Wang MG, Zou X, Yeh HY, Wang Z, Liu Y, Liu J, Wang F, Yao HB, Chen PY, Tao RY (2020) Peopling of Tibet Plateau and multiple waves of admixture of Tibetans inferred from both modern and ancient genome-wide data. *bioRxiv*. <https://doi.org/10.1101/2020.07.03.185884>
- Wang CC, Yeh HY, Popov AN, Zhang HQ, Matsumura H, Sirak K, Cheronet O, Kovalev A, Rohland N, Kim AM, Bernardos R, Tumen D, Zhao J, Liu YC, Liu JY, Mah M, Mallick S, Wang K, Zhang Z, Adamski N, Broomandkhoshbacht N, Callan K, Culleton BJ, Eccles L, Lawson AM, Michel M, Oppenheimer J, Stewardson K, Wen S, Yan S, Zalzal F, Chuang R, Huang CJ, Shiung CC, Nikitin YG, Tabarev AV, Tishkin AA, Lin S, Sun ZY, Wu XM, Yang TL, Hu X, Chen L, Du H, Bayarsaikhan J, Mijiddorj E, Erdenebaatar D, Iderkhangai TO, Myagmar E, Kanzawa Kiriya H, Nishino M, Ki S, Shubina OA, Guo JX, Deng QY, Kang L, Li DW, Li DN, Lin R, Cai WW, Shrestha R, Wang LX, Wei LH, Xie GM, Yao HB, Zhang MF, He GL, Yang XM, Hu R, Robbeets M, Schiffels S, Kennett DJ, Jin L, Li H, Krause J, Pinhasi R, Reich D (2020) The genomic formation of human populations in East Asia. *bioRxiv*. <https://doi.org/10.1101/2020.03.25.004606>
- Wei LH, Huang YZ, Yan S, Wen SQ, Wang LX, Du PX, Yao DL, Li SL, Yang YJ, Jin L, Li H (2017) Phylogeny of Y-chromosome haplogroup C3b-F1756, an important paternal lineage in Altaic-speaking populations. *J Hum Genet* 62:915–918

- Wen B, Li H, Lu DR, Song XF, Zhang F, He YG, Li F, Gao Y, Mao XY, Zhang L, Qian J, Tan JZ, Jin JZ, Huang W, Dekar R, Su B, Chakraborty R, Jin L (2004) Genetic evidence supports demic diffusion of Han culture. *Nature* 431:302–305
- Wong EH, Khrunin A, Nichols L, Pushkarev D, Khokhrin D, Verbenko D, Evgrafov O, Knowles J, Novembre J, Limborska S (2017) Reconstructing genetic history of Siberian and Northeastern European populations. *Genome Res* 27:1–14
- Wu XZ (2006) Evidence of multiregional human evolution hypothesis from China. *Quat Sci* 26:706–709
- Xia ZY, Yan S, Wang CC, Zheng HX, Zhang F, Liu YC, Yu G, Yu BX, Shu LL, Jin L (2019) Inland-coastal bifurcation of southern East Asians revealed by Hmong-Mien genomic history. *bioRxiv*. <https://doi.org/10.1101/730903>
- Xiang H, Gao JQ, Yu BQ, Zhou H, Cai DW, Zhang YW, Chen XY, Wang X, Hofreiter M, Zhao XB (2014) Early Holocene chicken domestication in northern China. *Proc Natl Acad Sci USA* 111:17564–17569
- Xie CZ, Li CX, Cui YQ, Zhang QC, Fu YQ, Zhu H, Zhou H (2007) Evidence of ancient DNA reveals the first European lineage in Iron Age Central China. *Proc Biol Sci* 274:1597–1601
- Xu BY, Guo JX, Huang Y, Chen XY, Deng XH, Wang CC (2019) The paternal genetic structure of Jingpo and Dai in southwest China. *Ann Hum Biol* 46:279–283
- Yan WM (1979) The new discoveries of early Neolithic culture in the Yellow River basin. *Archaeology* 1:45–50
- Yan S, Wang CC, Zheng HX, Wang W, Qin ZD, Wei LH, Wang Y, Pan XD, Fu WQ, He YG, Xiong LJ, Jin WF, Li SL, An Y, Li H, Jin L (2014) Y chromosomes of 40% Chinese descend from three Neolithic super-grandfathers. *PLoS ONE* 9:e105691
- Yang MA, Gao X, Theunert C, Tong H, Aximu-Petri A, Nickel B, Slatkin M, Meyer M, Pääbo S, Kelso J (2017) 40,000-year-old individual from Asia provides insight into early population structure in Eurasia. *Curr Biol* 27:3202–3208 (e3209)
- Yang MA, Fan XC, Sun B, Chen CY, Lang JF, Ko YC, Tsang CH, Chiu HL, Wang TY, Bao QC, Wu XH, Hajdinjak M, Ko AMS, Ding MY, Cao P, Yang RW, Liu F, Nickel B, Dai QY, Feng XT, Zhang LZ, Sun CK, Ning C, Zeng W, Zhao YS, Zhang M, Gao X, Cui YQ, Reich D, Stoneking M, Fu QM (2020) ancient DNA indicates human population shifts and admixture in northern and southern China. *Science* 369:282–288
- Yu H, Spyrou MA, Karapetian M, Shnaider S, Radzevičiūtė R, Nägele K, Neumann GU, Penske S, Zech J, Lucas M (2020) Paleolithic to Bronze Age Siberians reveal connections with first Americans and across Eurasia. *Cell* 181:1232–1245 (e1220)
- Yuan JP (2010) Shilun Zhongguo Zaoqi Wengming de Chansheng—Yi Hunan Chengtoushan Diqu Gudai Wengminghua Jincheng Weili. *Cult Relics Cent China* 5(22–27):32
- Zhang XM, Liao SY, Qi XB, Liu JW, Kampuansai J, Zhang H, Yang ZH, Serey B, Sovannary T, Bunnath L, Seang Aun H, Samnom H, Kangwanpong D, Shi H, Su B (2015) Y-chromosome diversity suggests southern origin and Paleolithic backwave migration of Austro-Asiatic speakers from eastern Asia to the Indian subcontinent. *Sci Rep* 5:15486
- Zhang Y, Li JW, Zhao YB, Wu XY, Li HJ, Yao L, Zhu H, Zhou H (2017) Genetic diversity of two Neolithic populations provides evidence of farming expansions in North China. *J Hum Genet* 62:199–204
- Zhang Y, Wu XY, Li JW, Li HJ, Zhao YB, Zhou H (2018) The Y-chromosome haplogroup C3\*-F3918, likely attributed to the Mongol Empire, can be traced to a 2500-year-old nomadic group. *J Hum Genet* 63:231–238
- Zhang Y, Lei XY, Chen HY, Zhou H, Huang S (2018) Ancient DNAs and the Neolithic Chinese super-grandfather Y haplotypes. *bioRxiv*. <https://doi.org/10.1101/487918>
- Zhang MH, Yan S, Pan WY, Jin L (2019) Phylogenetic evidence for Sino-Tibetan origin in northern China in the Late Neolithic. *Nature* 569:112–115
- Zhao ZJ (2010) New data and new issues for the study of origin of rice agriculture in China. *Archaeol Anthropol Sci* 2:99–105
- Zhao H (2017) Liangzhu de Guojia Xingtai. *China Cult Herit* 3:24–30
- Zhao YB, Li HJ, Cai DW, Li CX, Zhang QC, Zhu H, Zhou H (2010) Ancient DNA from nomads in 2500-year-old archeological sites of Pengyang, China. *J Hum Genet* 55:215–218
- Zhao Q, Pan SL, Qin ZD, Cai XY, Lu Y, Xu JS, Yin RX, Li SL, Jin L, Li H (2010) The Minz Zhuang of Napo Country, Guangxi: legacy of the Thai-Lao Migration South. *Commun Contemp Anthropol* 4:22–26
- Zhao YB, Li HJ, Li SN, Yu CC, Gao SZ, Xu Z, Jin L, Zhu H, Zhou H (2011) Ancient DNA evidence supports the contribution of Di-Qiang people to the Han Chinese gene pool. *Am J Phys Anthropol* 144:258–268
- Zhao YB, Zhang Y, Li HJ, Cui YQ, Zhu H, Zhou H (2014) Ancient DNA evidence reveals that the Y chromosome haplogroup Q1a1 admixed into the Han Chinese 3,000 years ago. *Am J Hum Biol* 26:813–821
- Zhao YB, Zhang Y, Zhang QC, Li HJ, Cui YQ, Xu Z, Jin L, Zhou H, Zhu H (2015) Ancient DNA reveals that the genetic structure of the northern Han Chinese was shaped prior to 3,000 years ago. *PLoS ONE* 10:e0125676
- Zhong H, Shi H, Qi XB, Xiao CJ, Jin L, Ma RZ, Su B (2010) Global distribution of Y-chromosome haplogroup C reveals the prehistoric migration routes of African exodus and early settlement in East Asia. *J Hum Genet* 55:428–435
- Zhong H, Shi H, Qi XB, Duan ZY, Tan PP, Jin L, Su B, Ma RZ (2011) Extended Y chromosome investigation suggests postglacial migrations of modern humans into East Asia via the northern route. *Mol Biol Evol* 28:717–727
- Zhou R, An L, Wang X, Shao W, Lin G, Yu W, Yi L, Xu S, Xu J, Xie X (2007) Testing the hypothesis of an ancient Roman soldier origin of the Liqian people in northwest China: a Y-chromosome perspective. *J Hum Genet* 52:584–591
- Zhou R, Yang D, Zhang H, Yu W, An L, Wang X, Li H, Xu J, Xie X (2008) Origin and evolution of two Yugur sub-clans in Northwest China: a case study in paternal genetic landscape. *Ann Hum Biol* 35:198–211

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.