Supplementary Information

The Genomic Formation of Human Populations in East Asia

Chuan-Chao Wang, Hui-Yuan Yeh, Alexander N Popov, Hu-Qin Zhang, Hirofumi Matsumura, Kendra Sirak, Olivia Cheronet, Alexey Kovalev, Nadin Rohland, Alexander M. Kim, Swapan Mallick, Rebecca Bernardos, Dashtseveg Tumen, Jing Zhao, Yi-Chang Liu, Jiun-Yu Liu, Matthew Mah, Ke Wang, Zhao Zhang, Nicole Adamski, Nasreen Broomandkhoshbacht, Kimberly Callan, Francesca Candilio, Kellie Sara Duffett Carlson, Brendan J. Culleton, Laurie Eccles, Suzanne Freilich, Denise Keating, Ann Marie Lawson, Kirsten Mandl, Megan Michel, Jonas Oppenheimer, Kadir Toykan Özdoğan, Kristin Stewardson, Shaoqing Wen, Shi Yan, Fatma Zalzala, Richard Chuang, Ching-Jung Huang, Hana Looh, Chung-Ching Shiung, Yuri G. Nikitin, Andrei V. Tabarev, Alexey A. Tishkin, Song Lin, Zhou-Yong Sun, Xiao-Ming Wu, Tie-Lin Yang, Xi Hu, Liang Chen, Hua Du, Jamsranjav Bayarsaikhan, Enkhbayar Mijiddorj, Diimaajav Erdenebaatar, Tumur-Ochir Iderkhangai, Erdene Myagmar, Hideaki Kanzawa-Kiriyama, Masato Nishino, Ken-ichi Shinoda, Olga A. Shubina, Jianxin Guo, Wangwei Cai, Qiongying Deng, Longli Kang, Dawei Li, Dongna Li, Rong Lin, Nini, Rukesh Shrestha, Ling-Xiang Wang, Lanhai Wei, Guangmao Xie, Hongbing Yao, Manfei Zhang, Guanglin He, Xiaomin Yang, Rong Hu, Martine Robbeets, Stephan Schiffels, Douglas J. Kennett, Li Jin, Hui Li, Johannes Krause, Ron Pinhasi, David Reich

TABLE OF CONTENTS

Supplementary Information section 1: Context of newly reported individuals

Newly reported samples from present-day groups

We collected blood samples from 383 healthy and unrelated individuals from Han Chinese, Tibeto-Burman, Tai-Kadai, Austroasiatic, Mongolic, Turkic, Tungusic, and Indo-European speaking groups from China and Nepal. Extended Data Table 1 lists information on these samples. The sample collection was carried out in 2014 in accordance with the human ethical research principles of The Ministry of Science and Technology of the People's Republic of China (Interim Measures for the Administration of Human Genetic Resources, June 10, 1998) and genotyping of the samples was reviewed and approved by the Ethics Committee of the School of Life Sciences, Fudan University. Study staff informed potential participants about the goals of the project, and individuals who chose to participate gave informed consent consistent with broad studies of population history. Genomic DNA was extracted using the QIAamp DNA Mini Kit (QIAGEN, Hilden, Germany). The individuals all contributed DNA samples voluntarily, and the informed consents of all newly reported individuals are consistent with the full public dissemination of de-identified genotype data. The collection of genome-wide variation data on de-identified samples was approved by the Harvard Human Research Protection Program (Protocol 11681), re-reviewed on 12 July 2016. We genotyped all the samples on the Affymetrix Human Origins single nucleotide polymorphism (SNP) array at the Shenzhen-Beijing Genomics Institute.

Ancient samples

We newly report data from 166 ancient skeletons including 82 from 43 sites in Mongolia from ~6000 BCE to ~1000 CE, 11 ~3000 BCE individuals from Shannxi in northern China (Neolithic farmers from the Wuzhuangguoliang site), 46 from Taiwan (Iron Age and earlier samples from the Hanben and Gongguan sites), 18 from the Amur River Basin in the Russian Far East dating to around ~5000 BCE (Middle Neolithic hunter-gatherers from the Boisman-2 Cemetery), one ~1000 BCE individual (Iron Age from the Yankovsky Culture), one early Medieval individual dating to ~1000 CE (from

the Roshino-4 (***Рощино-4***) Bohai Mohe culture), and 7 Jomon individuals from about 1500-1000 BCE and 2400-800 calBCE. Online Table 1 gives tabular data on the samples.

Boisman (*Бойсман*) culture[1-3]: The sites of the Middle Neolithic Boisman culture are distributed over the Pacific coastal area of the southwestern Maritime Province of Russia and are represented archaeologically by settlements with shell mounds. The subsistence of the Boisman culture was based on hunting, fishing, and gathering terrestrial and marine resources. The archaeological assemblage of the Boisman culture and other similar cultures of the Middle Neolithic of the southern region of the Russian Far East as well as some Neolithic sites of Northeast China and the northeast of North Korea are part of a distinctive archaeological complex that differs from synchronous archaeological complexes of northeastern China and the Chulmun culture of the Korean Peninsula. The radiocarbon dates of material remains from this culture range from 4900-2500 BCE, and the direct radiocarbon dates on bone obtained for this study ranged from 5368-3633 calBCE. According to some scholars, anthropometric features of skulls from Boisman-2 (***Бойсман-2***) are particularly reminiscent of Arctic populations, such as Chukchi. Morphological similarity has also been documented between the burials at Boisman-2 and two present-day populations in the Lower Amur River basin, the Nivh and Ulchi. Some scholars have suggested that the Boisman-2 remains are "the earliest known representatives" of a much broader "Pacific" branch of East Eurasian populations comprising coastal groups from Chukotka to Taiwan[4]. We successfully obtained ancient DNA data from 18 individuals from the Boisman-2 site.

Yankovsky (*Янковская*) culture[5]: Early Iron Age people of the Yankovsky culture (850-350 BCE) inhabited mainly the coastal and subcontinental zones from the northern Korea peninsula to the Eastern Primorye coast. The subsistence of the Yankovsky culture was based on hunting, millet cultivation and widespread exploitation of marine resources. Millet cultivation played some role in subsistence, but not at all sites. We successfully obtained ancient DNA from one individual assigned to

the Yankovsky culture at Pospelovo-1 (***Поспелово-1***), Burial 1, Primorsky Krai, Russia.

Heishui Mohe[6,7] (Хейшуй Мохэ, 黑水靺鞨) culture: Historical accounts indicate that in the 7th-10th centuries CE to the period of the Sui and Tang dynasties on the territory from the lower reaches of Sungari, Usury and the adjacent Amur river valley, there were 16 tribes of the Heishui Mohe, divided into northern and southern Heishui. The latter occupied the northern and northeastern parts of the present-day Maritime Province of Russia. After the decline of Bohai (Бохай) (10th century CE) they are known in the Chinese historical records as the Eastern Jurchen. They have been hypothesized to be ancestral to the Jurchen, present-day Manchu and other Tungusic peoples. We successfully obtained ancient DNA from an individual belonging to the Heishui Mohe culture from Roshino-4 (Рощино-4) Cemetery, Burial 6 Primorsky Krai, Russia and one individual belonging to the Mohe Bohai (Мохэ Бохай) culture from Chernyatino-5 (Чернятино-5) Graveyard, Primorsky Krai, Russia.

Hunter-gatherers of Japan: We obtained ancient DNA from 7 Jomon individuals from the Funadomari site (1500-1000 BCE) and Rokutsu site (2472-835 calBCE).

The Funadomari site is a shell-midden site located in the northern part of Rebun Island, Hokkaido Prefecture in the northernmost part of the Japanese archipelago. The site is elevated approximately 5 meters above sea level on a sand dune along the Funadomari Gulf. In 1998, Nishimoto led a team that uncovered 28 primary inhumation burials in good preservation in a flexed position and associated with numerous artefacts including necklaces, bracelets and anklets made of marine shell (*Mercenaria stimpsoni*)[8]. The typology of the cord marking decorations on the unearthed pottery is related to the "Kasori B" type found on the island of Honshu, Chiba Prefecture Japan, suggesting that these individuals were buried during the early to middle phase of the Late Jomon Period (ca. 1500-1000 BCE)[8]. Despite their geographical remoteness, the Funadomari Jomon crania show morphological resemblances to other known Jomon crania from Honshu

Island[9] belonging to the first layer of *Homo sapiens* that dispersed in the pre-farming period[10]. The direct 14C date for individual 23 gave a date of 4025±20 (yrBP±1σ) (3846–3644 calBP (1σ), 3960–3550 calBP (2σ)) (Paleo Labo Co., Ltd.)[11,12]. Another 14C dating of charcoal from the hearth resulted in a range between the top layer of 3635±65 calBP and lowest layer of 3850±55 cal BP[8].

The Jomon shell-midden site of "Rokutsu" is located in Ohkanza town in Chiba City, Chiba Prefecture adjacent to metropolitan Tokyo. A large-scale rescue excavation took place in 2006 by the archaeological center of Chiba city, and was led by Hideyo Tanaka and Wataru Nagahara (report unpublished). The analysis uncovered the remains of seven inhumation burials. The individuals in burial No 1 (adult female), 2 (adult male), 5 (adult female), 6 (adult male), and 7 (adult male) were laid down on the side with limbs extended, while individuals 3 and 4 (female adult) were buried with flexed joints at the elbow, hip and knee. The local typology of Jomon pottery assigned to these burials was Late Jomon in the range 2472-835 calBCE (pottery typology, Shoumyoji, Horinouchi and Kasori B types).

Wuzhuangguoliang Neolithic site (五庄果墚): The traditional archaeological culture classification in China identifies six large regional systems in the Neolithic: North, Central Plains, East, Southeast, Southwest and South[13]. Our newly reported individuals come from site of Wuzhuangguoliang (WZGL), in the contact zone between the Central Plains and the North cultural regions. The WZGL site is located at Xiaojie village, Jingbian County, Shaanxi Province, adjacent to Inner Mongolia in the north and Yan'an in the south. The excavators divided the site into three areas, A, B and C, with a total area of about 1,740 m^2, exposing 21 houses, 88 ash pits and 2 pottery kilns. The exposed human bones were all from the ash pits. A total of 20 human bones for DNA analysis were collected from the ash pits distributed in the A and B areas. The site was comprehensively excavated by the Shaanxi Provincial Institute of Archaeology in 1996 and 2001. Radiocarbon dating, based primarily on zooarchaeological remains in ash pits from this site, suggests dates of 3317-2921 calBCE (4422±29 BP, XA-8399). Two phases can be discerned in this site based on architecture, ash pits, pottery kilns, and

archaeological stratigraphy and typology, and it was suggested that the two phases might belong to the Haishengbulang culture and Ashan culture respectively, ranging from the late Yangshao period to the early Longshan period[14]. The distribution of Phase One is located in the east of the site coded as area A while Phase Two is located in the southern parts of the sites coded as areas B and C. The archaeological remains included slipped and corded pottery in various forms including jars, basins, jugs, vats, bottles and bowls, with fine clay and sand-tempered in red, gray and reddish-brown. It also included polished stone axes and elaborate houses with gardens and fireplaces. The pointed-bottomed bottles with cup-shaped mouths, which are representative of the Yangshao culture[15], suggest a connection with the Central Plains cultural region. Both hunting and gathering and raising of domesticated animals was practiced at WZGL, with some ancient animal bones from domestic animals such as *Canis familiaris* and *Sus domestica*, and others from wild animals such as *Lepus capensis*, *Procapra gutturosa*, *Equus sp*. and nearly complete felid skeletons identified as *Coun alpinus* and *Felis silvestris* in wild[16]. In the Neolithic WZGL had a steppe ecology, but today it is in a desert environment and the vegetation is accordingly very different[17]. The domestication practices of *Sus domestica* are similar to those of the Guanzhong Area which belongs to the Central Plains cultural region, further highlighting a connection between the areas[18].

Physical anthropological and molecular anthropological analyses have been carried out on ancient human bones from three ash pits (AH1, BH23 and BH35). Previous ancient mtDNA study of human bones of this site suggested that ancient WZGL people shared lineages with East Eurasians, identifying haplogroups A, B, D, R9a, N9a, F1a and Z[19]. Morphological observation and measurements highlight the close relationship of the WZGL individuals to the present-day Liuwan people (eastern Qinghai province), Yuchisi people (northern Anhui province) and Xixiahou people (middle Shandong province). The WZGL site provides comprehensively verified evidence of the use of ash pit burial in northern Shaanxi and contributes to our understanding of living habits and ideology, as well as the physical characteristics of ancient inhabitants in that period.

Hanben site (汉本遗址)[20]: The Hanben site is located at Aohua village, Nan'ao Township, Yilan County. The name "Hanben" originated from the time of Japanese political control. It is the midpoint of Suhua coastal road (the predecessor of Suhua Highway), so it was named "Ban Fen" (は ん ぶ ん; hanbun). When the North-linked road was under planning and construction, the station was named "Hanben" which is a homophone of its original name. The indigenous Atayal language - Blihun (original meaning: door, extended meaning: doorway) has been prepended to the original Hanben heritage site.

There are three prehistoric and historic cultural layers unearthed in the Blihun Hanben site. In addition to the historic cultural layer, the first cultural layer from the top to the bottom (L4 upper cultural layer) of the prehistoric cultural layer was dated to between 1600±50 BP and 1160±40 BP. The earliest and the latest calibration date is approximately between 1574 BP and 971 BP. The Metal Ages of Taiwan's prehistoric cultural sequence belong to the Puluowan type of the Shihsanhang culture, with high-temperature technology, navigational abilities, a rigorous settlement pattern and varied burial forms. The cultural relics unearthed in Hanben site include metal wares rarely seen in Taiwan, such as bronze hilts, gold foil, iron wares, and a large number of glass beads, glass rings, agate beads and other foreign products. It is speculated that at that time, the Blihun Hanben people had frequently traded with the Southeast Asian people through maritime trade routes.

According to the summary report on rescue excavations at the Hanben site[20], the second prehistoric level (L6 cultural layer) was dated to between 1972±30 BP and 1509±50 BP. The earliest and the latest calibration date are between 1992-1335 BP. In this study, we obtained direct radiocarbon dates for 14 individuals. A direct date on the individual I3618/M50 of 1406-1261 calBCE (3060±20 BP, PSUAMS-2268) is inconsistent with multiple adjacent context dates and so we ignore it pending confirmation by additional

attempts to obtain direct dates on the individual. Direct dates on other individuals from same site and context (I3611, I3615, I3620, I3732, I8081, I13695, I8075, I3619, I3728, I3735, I3621, I3616, I3617) give a range of 22-774 calCE. A debris flow caused by a large-scale disaster separates the first two cultural layers. The ages of the two cultural layers are continuous and there may even be hundreds of years of overlap. However, the cultures are slightly different. Although there are no ironmaking furnaces in the second prehistoric cultural layer, there is still metal ware made by high-temperature technology, glass and agate, along with multiple burial forms. Popular burial forms included short rectangular stone coffins, multiple body burials or limb bending burials. Their ages are earlier than the known Shihsanhang culture. Although their cultures are slightly different, they are obviously related, and thus shed light on the origins of the Shihsanhang culture.

Between the second and third prehistoric cultural layers, there is also a debris-flow, caused by another natural disaster. There are few relics unearthed in the third cultural layer, and it is speculated that its age is between 2400-2000 BP. Since the third cultural layer was buried deep in the ground exceeding the depth of pier foundation, this limited the safety depth of the engineering protection. Therefore, the archaeology team failed to carry out scientific excavation. More detailed information was obtained only from the P2 (s) pier. As the mountainous area around the settlement was burnt and cultivated, this was likely an important settlement. The second and third cultural layers mentioned above are prehistoric cultural layers, which are not found in Taiwan at present. The Hanben site has emerged as an important location for understanding Metal Age cultures in Taiwan and the movement of people along the East coast.

Green Island, Gongguan site (绿岛公馆遗址):

Green Island, a 15 square kilometer island lying 33 kilometers off the southeastern coast of Taiwan, was in a position of frequent cultural exchange and interaction with sites in Taiwan and beyond. Human beings have lived and moved around Green Island at least

since 4300-3400 years ago based on archaeological evidence for the Eastern cord-marked pottery culture evidence in the archaeological record here. The archaeological record of the island also spans the late Neolithic Peinan culture (3400-2300 years ago), the Metal Age (2300-350 years ago) to the period of settlement of the early historical Yami people (Dawu people). Green Island is also known as Sanasai Island, the legendary source of ancestors of Taiwan East Coast aboriginal groups. According to ethnographic data records, in the legends of the origin of the ancestors of Amis, Kavalan and Ketagalan, Sanasai Island was the ancestral homeland, and some groups clearly explained that this was the same as Green Island[21]. Thus, Green Island has an important status in the collective cultural memory of present-day aboriginal groups and was plausibly relevant to ancient migratory processes. The oral records of some Yami communities in Orchid Island (to the south of Green Island) and the Han people in Green Island since modern times also show that before the Han people moved into Green Island in the middle of Qing Dynasty, there were still aboriginal groups living on Green Island. From the literature and oral records, these aboriginals groups are plausibly related to the Yami ethnic people in Orchid Island today.

The Gongguan site on Green Island includes plain sand-embedded pottery, stone axes, net pendants, stone clusters, stone adzes, concave stones, whetstones and shell scrapers. The early stages of the cultural of the site are considered related to the Peinan culture, later classified as the local Gongguan type of Green Island, and indeed we obtained a direct radiocarbon date of one individual for which we obtained ancient DNA (I13721 / GG-107-M1-03) of 1366-1126 calBCE (2990±20 BP, PSUAMS-7614) in the time range expected for this culture. On the surface, ceramic pieces from the Lobusbussan Culture can be seen[22]. At present, most of the site is comprised of settlements and farmlands. Five charcoal specimens of TP1 and TP1 NW Ex unearthed in L2-L3 cultural layer were selected and sent to the BETA laboratory for carbon dating. The results indicated that the ages of the two cultural layers (after correction) were concentrated around 1540-1302 years ago, and only one of them ranges from 1702-1560 years ago. In the past, the Gongguan site was classified as being part of the late

Neolithic Age and the transitional period between the Neolithic and Metal Ages[22]. According to the results from this research, the survival of the Gongguan site should be extended to the Metal Age's Sanhe Culture Shanlangliao or Jiuxianglan type. The cultural layer of the excavation site may represent the remains of the later stage of the Gongguan site. In addition, the pottery of Lobusbussan Culture can also be seen here, some of which also appears in the cultural layer of Sanhe Culture, indicating that the Lobusbussan Culture in Green Island (Gongguan site) may have started earlier. A shell ring unearthed in the tomb contained off-island elements, possibly related to Orchid Island and Southeast Asia[23].

Mongolian sites

In what follows we describe the archaeological context for the newly reported Mongolian individuals in our study. We also justify the material culture classification that we use to group the individuals prior to genetic analysis.

Neolithic of Eastern Mongolia: The burials we analyzed probably belong to the so-called Tamsagbulag culture of the 6th to 5th millennia BCE. Until recently this culture was known from only 5 burials on 4 sites: Tamtsag-Bulak (Тамцаг булагийн суурин), Barun-olzet (Баруун Өлзийт), Dzun-olzet (Зүүн Өлзийт) and Norovlin Uul (Норовлин уул)[24, 25]. They are all found in Eastern Mongolia (Dornod aimag). The sites include several settlements and encampments consisting of Neolithic dwellings in the areas of Ovoot (Овоот), Tamtsag-Bulak (Тамцаг-Булак) and along the banks of the Kerulen (Керулен) river and its tributaries (near the city of Choybalsan (Чойбалсан)). To date, numerous sites of this culture have been investigated and permanent settlements with durable semi-subterranean houses were excavated, with the largest house recorded as 120 m^2 in area. A large number of thin stone plates were found for the manufacture of composite tools (knive-blades, daggers, arrowheads, awls, scrapers, and others), as well as jewelry made of various materials, fragments of ceramics and animal bones. It is assumed that the people of the Tamtsagbulag (тамцаг-булакская) culture were engaged in primitive hoe-agriculture, since grain grinders and pestles were

found in large numbers. But the main economic activities were hunting, fishing and gathering; whether or not stockbreeding was practiced is questionable.

In the neighboring Russian territory there have been more substantial excavations covering the Neolithic period, which have made it possible to identify several archaeological cultures in the region including Kitoi (китойская), Serovo (серовская), Isakovo (исаковская), Glazkovo (глазковская)), Transbaikalia (Fofanovo (фофановская), Ust-karengskaya (усть-каренгская), Chikoy(чикойская)), Priamurye (Gramotuhiskaya (громатухинская), Osipovskaya (осиповская), Kondonskaya (кондонская), Voznesenovskaya (вознесеновская) and others), Primorye (Boysmanskaya (бойсманская), Zaisanovskaya (зайсановская). The availability of substantial archaeological information from a Russian context makes it possible to conduct a comparative analysis with data from other regions, as well as to conclude that there were large communities of people in the South of Eastern Siberia and the Far East. In addition to the production of ceramics, there was clearly a high degree of expertise in stone processing as reflected in the creation of numerous tools and jewelry. This includes the use of composite materials. Favorable natural conditions and significant food resources contributed to the increase in population, which was reflected in numerous settlement complexes, as well as in the discovered cemeteries. New territories were also developed. Archaeologists have also noted evidence of exchanges between the south of western Siberia and Mongolia; for example, some features of the Tamsagbulag culture have analogies in the Transbaikalia region[26]. The Siberian groups in turn exchanged with peoples of the Caucasus in the Eneolithic.

Neolithic of Northern Mongolia: The 2015 study of nine Neolithic burials in the Egiyn-Gol (Эгийн-гол) river valley (Bulgan aimag (Булган аймак) of Mongolia) was an important discovery. This includes the earliest known burial ground tombs anywhere in the world including the monuments of Marzyn hytul (Марзын хөтөл) and Kharuulyn hozgor (Харуулын гозгор)[27, 28]. Among them there are individual and collective burials. Numerous beads, stone arrowheads, and composite tools similar to daggers and spears

were found with the deceased people. Well-preserved devices for weaving nets made of bone were also found. There were no ceramic vessels. Apparently, the main activities of the societies were hunting and fishing. The presented Neolithic necropolises in the Egiyn-Gol river valley (Northern Mongolia) show the greatest similarities with the burial sites of the Fofanovo culture of Transbaikalia, as well as with the materials of these cultures of the Baikal (Байкал) region[29]. This circumstance is explained by the proximity of territories and the presence of interaction routes along the Selenga (Селенга) river and its tributaries. It should be noted that individual sites of the Neolithic period are recorded in different places in Mongolia, which indicates the development of this territory, but by different groups; not enough is known as yet to fully understand the patterns.

Afanasievo culture[30,31]: People of this archaeological culture are hypothesized to have migrated from the Volga-Ural steppe region to the area of the Russian Altai in the late 4th millennium BCE and then to have spread to Middle Yenisei basin, Eastern Kazakhstan, Northern Xinjiang, and Central Mongolia between 3000-2600 BCE. The basis of the economy of this culture was sheep and cattle breeding, and small seasonal settlements have been identified. The Afanasievo is assigned to the Chalcolithic Period based on the lack of metal artifacts made from bronze, though artifacts made from copper, gold and meteorite iron have been found. Peoples of the Afanasievo culture used egg-shaped and ball-shaped pottery and so-called "censers" (incense burners) with different stamp ornamentation, similar to the pottery of the earliest Yamnaya culture (or pre-Yamnaya period) of Eastern Europe. Previous analyses have reported extremely strong genetic similarities between people of the Yamnaya culture in the western Steppe and people of the Afanasievo culture in the Altai region, suggesting they are linked. The burials we analyzed are from Bayankhongor aimag, Erdenetsogt sum, Shatar chuluu (Шатар чулуу) kurgan 2 and Bayan-Ulgi aimag, Ulaankhus sum, Kurgak govi (Хуурай говь) barrow 1 (the infant burial in the main grave).

Chemurchek (Qiemuerqieke) phenomenon: The Chemurchek phenomenon is named

for the river 切木尔切克 Qiemuerqieke in the Altai county of Xinjiang[32,33]. A complex of cultural features resembling those of contemporary peoples in Western Europe appeared in the slopes of Mongolian Altai (Northern Xinjiang and Western Mongolia) not later than 2700-2600 BCE. The basis for the culture links to Middle-Late Neolithic Western European cultures include the architecture of the burial constructions, specifically megalithic chambers constructed from huge stone slabs with collective burials surrounded by multiple stone and soil cairns overlapping one another; the form and ornamentation of vessels; the styles of stone statue-menhirs with protruding facial contours and circular protruding eyes, paintings on slabs inside burial chambers (rhombs and chevrons inscribed into each other, parallel multi-triangle festoons, sloping nets, nets with cells filled with roundish spots, meander-shaped and volute-shaped curves); and other religious objects and images. Only burial and ritual complexes of the Chemurchek phenomenon have been excavated (no settlements were identified). Finds included tin bronze and lead artifacts and domesticated animal bones. The Chemurchek people had close contact with peoples of the late Afanasievo culture who spread into Xinjiang, and later influenced the Elunino culture of Steppe Altai, the Karakol culture of Mountain Altai, and the Okunevo culture of Middle Yenisei basin.

Early Bronze Age of Ulgii: This group is represented by five sites dated to the latter half of the 3rd millennium BCE excavated near the northwest edge of Mongolia in the high-altitude Altai mountains (Bayan-Ulgii aimag). Burials were placed in the middle of rectangular fences with stone stelae clearly intended for ritual use. These fences with stone stelae were also a characteristic attribute of Chemurchek burial places in the neighboring Chinese Altai. Pottery vessels, bone tools, bone arrowheads, and small stone balls with analogs in Early Bronze Age cultures of Eastern Kazakhstan, Russian Altai and Khakassia have also been unearthed in these graves[33].

Munkhkhairkhan culture: This culture dates to ~1800-1400 BCE and ranges across the territories of Western and Northern Mongolia, Tuva, and the southern part of the Krasnoyarsk region in Russia. It is presently characterized only by funerary and ritual

complexes (no settlements have been identified). Crouched burials were arranged in small pits filled by stones inside flat round or rectangular stone platforms. Ordinary barrows are 3-5 m in diameter; however, in special cases, platforms were built measuring 30-40 m in diameter. Ritual structures include stone circles with stelae having astronomical significance. No pottery has been found. Burial goods include tin bronze knives and awls, tin bronze two-trumpet shaped rings, bone spoons, bone arrowheads, bone, and shell and stone ornaments. Knives and rings have analogs both in Western Siberia and in Chinese territories occupied by the late Qijia, Lower Xiajiadian, Siba, and Erlitou cultures. As such, it has been suggested that cultural innovation from Seima-Turbino and Andronovo complexes was introduced into the China Central Plain via the Munkhkhairkhan culture. The origin of the Munkhkhairkhan culture may be in the forest zone of Eastern Siberia in the upper Lena river basin near the Cis-Baikalian region. Bone spoons, arrowheads, and mother-of-pearl discs attributed to Cis-Baikalian Neolithic and Early Bronze Age cultures have similarities with Munkhkhairkhan, as well as funerary customs such as the filling of burial pits with stones[34,35].

Ulaanzukh type: The Ulaanzukh type is part of the "figured graves" that spread over Southeast Mongolia ~1450-1250 BCE. This group was characterized by burials under square platforms edged by masonry walls. Other groups of such graves have concave or convex sides (considered part of the Tevsh culture)[33] and are spread primarily in South Mongolia, but also partly in Central Mongolia as well as in the north of Inner Mongolia. All these burials were characterized by skeletons in prone positions with the head to the east. Burials are arranged in pits dug into the ground and filled with soil. Artifactual finds in Ulaanzukh type graves include pottery tripods, cornelian and limestone beads, gold ornaments, and stone polished tools. The prone position of buried bodies and findings of metal ornaments similar to those of northern Chinese Late Bronze Age cultures suggest contact with northern Chinese minorities during the late Shang period. It has been proposed that all graves with masonry-walled platforms and skeletons buried in the prone position be considered part of the Ulaanzukh-Tvesh

culture[36].

Late Bronze Age of Center and West Mongolia and "Mongun-Taiga type": Explorations and excavations in Center and West Mongolia and in Tuva conducted since the 19th century have revealed a large number of stone barrows of different shapes. The majority of these barrows include human burials but lack burial goods. The skeletons are arranged in stretched or in slightly crouched positions on their side or on their back in narrow earth pits or in stone cists on the ancient surface, covered by stones or organic materials[36,37]. Cairns of these barrows have circular or quadrangular shapes and may be surrounded by circular or trapezoidal fences with stelae or small stone heaps in the corners (so-called Khereksurs in Mongolian). Near to these kurgans, ritual burials of horse skins skulls were arranged within stone rings and under stone pavements. All of these barrows can be dated to the Late Bronze Age (~1500-1000 BCE). Mongolian and foreigner scholars have attempted to isolate particular types of these burial structures and designate them as belonging to specific "cultures" (e.g., "culture of Khereksurs and Deer Stones", "Sagsai culture"); however, there is not strong evidence for these cultural designations. Russian archaeologist Alexander Grach who carried out excavations in Tuva in the 1950-1970s years proposed the use of the name "Mongun-Taiga type" as a designation for burials with the body in a stretched position on the side with the head to the west. Excavations and investigations in the Mongolian Altai by A. Kovalev and D. Erdenebaatar show that such burials were not surrounded by stone fences and were not accompanied by horse sacrifices. Therefore, we considered the LBA Mongun-Taiga group independently, a classification that found support in our genetic analysis as the three individuals we analyzed from the Mongun-Taiga context were genetically distinctive from the main group of LBA Center-West individuals (Figure 3). Explorations of this region suggest that Late Bronze Age people from Central-West Mongolia and Tuva practiced sheep breeding and did not have permanent settlements. However, based on the number of burial mounds, it is likely that the population of this area of Mongolia and Tuva was very large. Physical anthropological studies have identified the people buried in these graves as having more morphological

affinity to West Eurasian populations in the west and East Asian populations in the east, a suggestion that is qualitatively supported from a genetic point of view as ancestry related to Sintashta and Andronovo people is more evident in the west than in the east. Russian anthropologist I. Gohman identified similarities between these populations and Afanasievo people of the Russian Altai and Minusinsk Basin.

Slab Grave culture: The people of this culture spread across East and Central Mongolia (up to the Gobi region) beginning in the 10th to 8th centuries BCE, and lived in this region at least until the 5th century BCE. It is likely that the center of this culture was in Transbaikalia, where rich burials of the so-called Dvortsovo type and different types of Slab graves have been discovered. Physical anthropological studies of individuals of this culture suggests they are East Asians. The economy of this culture was based on the breeding of sheep, goats, cows, and horses. Graves of this culture are characterized by rectangular stone fences oriented west to east, inside of which burials were positioned stretched out on their back with their heads toward the east. In the corners of the stone fences, vertical stone slabs were often erected. The space inside the fence was littered with stones and soil to make a flat platform. Ceramic vessels were broken and fragments were placed in the corners of the platforms. Burial goods included bronze weapons, horse harnesses, and ornaments and tools of so-called Scythian types as well of types used by peoples of Northeast China (such as the Upper Xiajiadian culture and Yuhuangmiao culture, dating to the first half of the 1st millennium BCE). It is possible that there was both demographic and cultural connections between the populations of Transbaikalia, East Mongolia, East Inner Mongolia, and North Hebei during this period.

Sagly culture: This is one of the cultures of the so-called Scythian-Siberian type[38]. This culture spread in Tuva during the 6th–3rd centuries BCE. One cemetery of this culture (Ulaangom cemetery) was excavated in Mongolia near the Tuvinian border, and Mongolian scholars designated these graves as belonging to the "Chandman culture." At present, only one cemetery in Mongolia is still attributed to this culture, while more

than 50 cemeteries of this culture were excavated from the beginning of the 20th century in Tuva. Burials of this culture were arranged in large timber chambers (approximately 4x4 m to 5x5 m and 3-4 m deep). Four to five people were buried in each chamber, likely one after the other. Burial goods included bronze and iron weapons and horse harnesses of Scythian style; bronze, gold and bone ornaments in "Animal style"; and other items attributed to the neighboring Pazyryk culture in the Altai Mountains. Ceramic vessels were determined to have similarities to Pazyryk vessels as well. People of this culture are anthropologically typed as mixed between West Eurasian and East Asian with the East Asian component increasing in presence in the later part of this culture. It is possible that an East Asian component represents matrimonial connections with peoples of North and North-West China during the Zhanguo period. Like the Pazyryk culture, the Sagly culture had strong connections with the regions of Ordos and Western Tianshan. Genetically, we found that Sagly culture individuals and Pazyryk culture individuals were similar in their deep ancestry proportions, further supporting their connection (Figure 3).

Xiongnu (Hunnu,匈奴) burials and Xianbei (鲜卑). The first evidence of Xiongnu nomadic cattle-breeding tribes in the Chinese historical sources appeared in 3rd century BCE. In this time, the Xiongnu were attested to live to the North of the bend of Yellow River (Hetao) of modern Shanxi and Hebei provinces[39]. Archaeologically, the first evidence of the Xiongnu (Hunnu) dates to the 2nd century BCE, when this cultural tradition was associated with a vast empire, and rapid material culture evolution. Xiongnu style graves in this period spread over Mongolia and over the Buryatia and Tuva regions in Russia. In Xiongnu style burials, physical anthropological evidence shows that the East Asian component increases from the West to East suggesting a mixed ancestry. The origin of Xiongnu has been hypothesized to have been in Eastern Mongolia and Manchuria. Ordinary graves of Xiongnu are arranged as deep rectangular pits with wooden coffins with buried people lying on the back with the head to the north. In the northern part of the pit, heads of goats and other livestock animals were often

laid. Elite graves of the Xiongnu were arranged as imitations of elite Han empire tombs in huge pits with a ramp from the south. Traces of shamanistic rituals and human sacrifices are evident in elite burials. In the north of Mongolia and in Buryatia, Xiongnu settlements are associated with a high level of craftsmanship. Following the 2nd century BCE, people practicing the Xiongnu culture began to use powerful weapons, most notably a long compound bow with arrows tipped with iron points; they also used Chinese iron long swords and iron horse harnesses. According to Chinese narrative sources, after Chinese offensives and the Xianbei expansion in the 2nd century CE, north Xiongnu leaders migrated out of Mongolia and local people accepted Xianbei customs. The earliest Xianbei cemeteries are thought to be from eastern Inner Mongolia. The 3-4 century CE tombs associated with this material culture in Mongolia in fact demonstrated a broad diversity of burial customs and goods so we can call them Xianbei only conditionally.

Mongol graves. Starting from the 10th century, a characteristic grave type associated with Mongols can be distinguished. These graves looks like narrow rectangular pits with burials on the back and head to the north, with a lamb shoulder blade near the head. Later, in the period of Mongol empire, buried people were laid in a special room dug into the side of burial pit. After Genghis Khan united the disparate Mongol tribes, the anthropological characteristics became somewhat different. The origin of the Mongols has been hypothesized to be associated with eastern Mongolia but Mongol ancestors settled in Central and Western Mongolia before the Turkic and Uigur expansions. Many Mongol scholars consider the Xiongnu as ancestors of the Mongols.

References

1. Popov, A.N., Chikisheva, N.F., and Shpakova, E.G. (1997). Boisman archaeological culture in southern Primorye (Novosibirsk: Institute of Archaeology and Ethnography).

2. Popov, A.N. (2008). Burial complexes of Boisman-2, multilayered site in south Primorye. Archaeology, Ethnography and Anthropology of Eurasia, *2*, 68-76,
3. Popov, A.N., Tabarev, A.V. and Mikishin, Y.A. (2014). Neolithization and ancient landscapes in southern Primorye, Russian Far East. J World Prehist, *27*, 247-261.
4. Chikisheva, T.A. (2016). On the Origin of the Neolithic Population of Northeast Asia. Archaeology, Ethnology & Anthropology of Eurasia, *44*, 148–154.
5. Sergusheva, E.A., and Vostretsov, Y.E. (2009). The advance of agriculture in the coastal zone of East Asia. In From Foragers to Farmers: Papers in Honour of Gordon C. Hillman, Fairbairn, A.S., Weiss, E., ed. (Oxford: Oxbow Books).
6. Li, D.S. (2006). A Study of the History of Heishui Mohe. Journal of Historical Science 5, 45-48.
7. Nikitin, Yu.G. (2005). Tan, Bohai and "eastern barbarians" (the eastern periphery of Bohai). The Russian Far East in antiquity and the Middle Ages: discoveries, problems, hypotheses. Vladivostok. 4, 517-541
8. Nishimoto T. (ed.). (2000). Rebuncho Funadomari Iseki Hakkutu Chousa Houkokusho. Rebun-cho KyouikuIinkai, Rebun-cho (In Japanese).
9. Matsumura H., Anezaki T., and Ishida H. (2001). A morphometric analysis of Jomon skeletons from the Funadomari site on Rebun island, Hokkaido, Japan. Anthropological Science, 109, 1–21.
10. Matsumura H., Hung H.C., Higham C., Zhang C., Yamagata M., Nguyen, L.C., Li, Z., Fan, X.C., Simanjuntak, T., Oktaviana, A.A., et al. (2019) Craniometrics reveal “two layers” of prehistoric human dispersal in eastern Eurasia. Scientific Reports 9, 1451.
11. Kanzawa-Kiriyama, H., Kryukov, K., Jinam, T.A., Hosomichi, K., Saso, A., Suwa, G., Ueda, S., Yoneda, M., Tajima, A., Shinoda, K.I., et al. (2016). A partial nuclear genome of the Jomons who lived 3000 years ago in Fukushima, Japan. J. Hum. Genet 62, 213–221.
12. Kanzawa-Kiriyama H., Jinam T.A., Kawai Y., Sato T., Hosomichi K., Tajima A., Adachi N., Matsumura H., Kryukov K., Saitou N. and Shinoda K. (2019) Late

Jomon male and female genome sequences from the Funadomari site in Hokkaido, Japan. Anthropological Science, 127: 83-108.

13. Su, B.Q., and Yin, W.Z. (1981). The regional systems, sequence and types of archaeological culture. Cultural Relics. 5:10-17.
14. Sun, Z.Y., Xu, Y.C., Li, W.H., and Shi, J. (2011). The excavated newsletter on the Wuzhuangguoliang site from Jingbian, Shaanxi. Archaeology and Cutural Relics. 6, 53–63.
15. Sun, Z.H. (1998). A restudy of the Banpo culture. Acta Archaeologica Sinica. 4, 419-446.
16. Vigne, J.D., Evin, A., Cucchi, T., Dai, L., Yu, C., Hu, S., Soulages, N., Wang, W., Sun, Z., and Gao J. (2016). Earliest “Domestic” Cats in China Identified as Leopard Cat (Prionailurus bengalensis). PLoS One 11, e0147295.
17. Hu, S., and Sun, Z. (2005). The faunal remains of the Wuzhuangguoliang site and its paleoenvironment analysis. Archaeol Cult Relics. 6, 72–84.
18. Guan, L., Hu, Y.W., Hu, S.M., Sun, Z.Y., Qin, Y., and Wang, C.S. (2008). Stable isotopic analysis on animal bones from the Wuzhuangguoliang site, Jingbian, Northern Shaanxi. Quaternary Sciences 6, 1160-1165.
19. Zhao, J., Liu, F.E., Lin, S., Liu, Z.Z., Sun, Z.Y., Wu, X.M., and Zhang, H.Q. (2017). Investigation on maternal lineage of a Neolithic group from northern Shaanxi based on ancient DNA. Mitochondrial DNA Part A. 28, 732-739.
20. Yi-Chang Liu. 2016. Summary report on rescue excavation of Hanben site-Improvement plan of Suhua highway in mountainous area of Tai 9 line. Institute of History and Philology, Academia Sinica.
21. Yi-Chang Liu, Minyong Qiu. 1995. Research Report on the Prehistoric Culture of Green Island in the Eastern Coast National Scenic Area. Institute of History and Philology, Academia Sinica.
22. Yi-Chang Liu. 2016. Research Report on Human Rights Cultural Park at Green Island. Institute of Archaeology, National Cheng Kung University.
23. Kano Tadao. 1946. Studies in the ethnology and prehistoric archaeology of Southeast Asia. Yajima Shobo, Tokyo.

24. Волков В.В. Погребение в Норовлин уула // Археология Северной и Центральной Азии / Отв. ред. А.П. Окладников и А.П. Деревянко. Новосибирск : Наука, 1975. С. 76–79. (In English: Buries in Norovlinura / Central Asian Archaeology red. a.p. (a) Claimant Trees. Novosibirsk: Science, 1975, 76-79.)
25. Дорж Д. Неолит Восточной Монголии. Улан-Батор : Изд-во АН МНР, 1971. 170 с. : ил. (In English: Neolithic Age in eastern Mongolia Ulaanbaatar: Department of International Development Press, 1971, 170.)
26. Derevianko, A.P. and Dorj, D. (1993). Neolithic tribes in Northern Parts of Central Asia // History of Civilisations of Cenral Asia. Volume I The Dawn of Civilization: Earliest Times to 700 b.c./Dani, A. H. and Masson, V.M. (ed.), UNESCO Publishing, pp. 163-183.
27. Iderkhangai, T. (2016) Neolithic burials in Northern Mongolia. Archaeological Relics of Mongolia. Volume 3. Ancient Funeral Monuments of Mongolia. Eregzen, D. (ed), Ulaanbaatar, pp. 24-30.
28. T. Idjerhangai, T., Je. Mizhiddorzh, S., Orgilbajar, B., Galbadrah, D., Jerdjenjebaatar, B. Өнөрbajar (2016) The newly discovered Neolithic burials in Northern Mongolia. Ancient cultures of Mongolia, Baikal Siberia and Northern China. Reports VII International Scientific Conference, Krasnoyarsk, October 3–7 2016. Part 1. Krasnoyarsk: SFU, pp. 126-143 (in Mongolian).
29. Асеев И.В. Юго-Восточная Сибирь в эпоху камня и металла. Новосибирск: Изд-во Института археологии и этнографии СО РАН, 2003. 208 с. (In English: Southeast Siberia in the stone and metal age. Novosibirsk: Institute of Archaeology and ethnology, Russian Academy of Sciences, 2003, 208.)
30. Vadetskaya, E., Polyakov, A., Stepanova, N. (2014) Corpus of Afanasievo culture sites. Barnaul: AZBUKA. pp. 380 (in Russian).
31. Kovalev, A. (2019) Spreading of Afanasievo culture in Xinjiang territory: chronological frames and typological particularity. Phenomenes of Cultures of Early Bronze Age of Steppe and Forest-Steppe Eurasia: roads of cultural connections in 5-3 mill. BCE. Orenburg:OGPU, pp. 188-208 (in Russian).

32. Kovalev, A. (2011) The Great Migration of the Chemurchek People from France to the Altai in the Early 3rd Millennium BCE . International Journal of Eurasian Studies. 1 (11) , pp. 1-58.
33. Kovalev, A. A., and Erdenebaatar, D. (2009). Discovery of New Cultures of the Bronze Age in Mongolia according to the Data obtained by the International Central Asian Archaeological Expedition. In Current Archaeological Research in Mongolia, J. Bemmann, H. Parzinger, E. Pohl, D. Tseveendorzh, ed. (Bonn: Vor- und Frügeschichtliche Archäologie Rheinische Friedrich-Wilhelm-Universität Bonn), pp.149–170.
34. Kovalev, A., Erdenebaatar, D. (2014) Discovery of a new Bronze Age culture in Central Eurasia (Munkh-Khairkhan culture). Russian Archaeological Annual. Saint-Petersburg: University editional consortium, pp. 194-225 (in Russian)
35. Kovalev, A. (2017) Munkh-Khairlkhan culture of Bromze Age and its connections with Neolithic-Bronze Age cultures of Eastern Siberia. Actual Problems of Archaeology and Ethnology of Central Asia.Materials of the II International conference (Ulan-Ude, 4–6th December, 2017). Ulan-Ude: The Buryat Scientific Center SB RA, pp. 58-66 (in Russian).
36. Honeychurch, W. (2015). Inner Asia and the Spatial Politics of Empire (NY: Springer), pp.122-128.
37. Amartuvshin, Ts.Tselkhagarav (2018) Investigating the khirgisuurs in Mongolia // Studia Archaeologica, XXXVI (5), pp. 56-85 (in Mongolian).
38. Mary, L., Zvénigorosky, V., Kovalev, A. et al. (2019) Genetic kinship and admixture in Iron Age Scytho-Siberians. Hum Genet 138, 411–423.
39. Yu, T. Rise of Xiongnu. Ouya xuekan (Eurasian studies). Vol. 5, 2005. (Beijing: Zhong Hua Book Company), pp. 1-8 (in Chinese) .

Supplementary Information section 2: Overview of genetic substructure

There are more than 200 languages in East Asia that are distributed into many indigenous families including Sino-Tibetan, Tai-Kadai, Austronesian, Austroasiatic, Hmong-Mien, Indo-European, Mongolic, Turkic, Tungusic, Koreanic, and Japonic, Yukaghiric, and Chukotko-Kamchatkan[1]. However, genetic structure across human populations in East Asia remains enigmatic due to limited sampling compared to other world regions, particularly in the Tibetan Plateau and in southwest China[2]. We report a fine-scale survey of the genetic structure of East Eurasians based on genome-wide variation from East Asians, Southeast Asians, and Siberians, with a specific focus on East Asia. We use a variety of methods for studying genome-wide variation data to identify qualitative patterns in the dataset. In many instances, we find that genetic clusters correlate strongly to language groupings. However, there are also exceptions where language group does not correspond to genetic patterns, and we highlight some of these patterns.

Genetic differentiation patterns (F_{st})

We computed F_{st} between all pairs of the present-day populations using smartpca (version: 13050[3], part of the EIGENSOFT package[4,5]). We found that the inbreeding corrected and uncorrected F_{st} were nearly identical (within ~0.001), and in this study, we always analyzed uncorrected F_{st}. Han Chinese are relatively homogeneous with the smallest pairwise F_{st} ranging from 0 to 0.006. The largest pairwise F_{st} values within East Asia are found between Tungusic-speakers (e.g. Nivh) and Austronesian-speakers (e.g. Atayal), where genetic differentiation reaches trans-continental levels of ~0.1 (Extended Data Fig. 3).

Qualitatively, we observe five main clusters in neighbor-joining F_{st} trees, which correspond to Altaic-speakers, Tibeto-Burman-speakers, Han Chinese, Southeast Asians (largely Tai-Kadai, Austronesian and Austroasiatic speakers), and Oceanians.

We also observe sub-clusters within these main groupings, including a sub-cluster of populations centered on the Amur River Basin (Nivh and the Tungusic-speaking Ulchi, Nanai, and Negidal), northeastern Siberians (Chukchi, Itelmen, Koryak and Eskimo), Ainu-Japanese, Nepalese groups, and Papuan-Australian. We caution that any patterns in neighbor-joining F_{st} trees should be viewed with caution. For example, Papuans and Australians are known highly divergent lineages, but they cluster with other Oceanians. It is likely that their derived position in the tree reflects admixture from Papuan-related (Melanesian) populations into nearby groups.

There are also some notable cases in which the position in the tree does not correlate perfectly to linguistic clusters or geography:

Populations with genetic affinities to West Eurasians. Three Turkic-speaking populations in northwest China—Uygur, Kazakh-China, and Kyrgyz-China—cluster with French and Russian rather than with other Turkic groups in China. The Nepalese Bahun also cluster with West Eurasians.

Austroasiatic speakers. Some Vietnamese including the Kinh cluster with mainland Tai-Kadai speakers rather than with other Austroasiatic speakers. Nicobarese and Cambodians form a separate cluster among Island Southeast Asians and Oceanians together with Thai and Borneo.

The Lahu. This Lolo-Burmese speaking population is genetically close to Tai-Kadai speakers.

The Gannan Tibetans. This northern Tibetan population clusters with Turkic and Mongolic populations.

Hmong-Mien speakers. These populations tend to be genetically intermediate between Tai-Kadai and Han.

Principal Component Analysis

We used smartpca (version: 13050), part of the EIGENSOFT package[4,5], to carry out Principal Component Analysis (PCA). We performed PCA on present-day populations from East Asia, Siberia, Europe, and Southeast Asia and then projected the ancient

samples using the lsqproject: YES option, which accounts for samples with substantial missing data. We did not perform any outlier removal iterations (numoutlieriter: 0). We set all other options to the default. We assessed statistical significance with a Tracy-Widom test using the twstats program of EIGENSOFT. All the principal components that we discuss and plot in what follows were highly statistically significant ($P<10^{-12}$).

In the first two components, there are two obvious clines from East Asians to West Eurasians, and from East Asians to Native Americans. Nepalese groups and populations in northwest China and Siberia (including Turkic speakers, Mongolic speakers and Tungusic speakers) are shifted towards the West Eurasians, while Northeast Asians are intermediate between East Asians and Native Americans. We observe that populations speaking the languages from the same language group tend to cluster, albeit with exceptions such as Turkic speakers, who are relatively scattered in the plot which is plausibly due to complex admixture with surrounding populations[6] (Extended Data Fig. 1, Extended Data Fig. 2, Figure SI2-1).

The ancient individuals from Devil's Gate, Boisman, AR_EN (Early Neolithic individuals from Amur River Basin[7]), AR_Xianbei_IA (Iron Age Xianbei samples from Amur River Basin[7]), and Yankovsky project close to present-day Tungusic speakers in the Amur River Basin, documenting a continuous presence of this ancestry profile in the Amur River Basin stretching back at least eight thousand years. Compared to this cluster, the Heishui_Mohe individual I3358 plots together with Tibetan groups and populations in northern China, documenting new genetic inputs into this region by Medieval times. Present-day people in northwest China and Siberia deviate in the direction of West Eurasians. Nepalese populations share a similar signal (although the historical processes that produced it are known to be quite different).

The Mongolian ancient individuals project on the cline from East Asians to West Eurasians. The individuals with minimum proportion of West Eurasian-related admixture (<10%), such as Mongolia_East_N, North_North_N,

Mongolia_LBA_2_Ulaanzukh, Mongolia_EIA_1_SlabGrave, and Mongolia_EIA_8 project close to the Amur River Basin populations. The individuals with about 20% West Eurasian-related ancestry from cultures such as Mongolia_LBA_4_CenterWest, Mongolia_EBA_1_Ulgii, Mongolia_MBA_1_Munkhkhairkhan, and Mongolia_LBA_6_Khovsgol project closely with published Neolithic and Early Bronze Age individuals around Lake Baikal[8]. The samples with about 40%-80% of West Eurasian related ancestry from Mongolia_EBA_2_Chemurchek, Mongolia_EIA_6_Pazyryk, Mongolia_LBA_5_CenterWest, Mongolia_EIA_4_Sagly, Mongolia_MBA_2_Munkhkhairkhan project at an intermediate position in the East-West cline. The individuals from Mongolia_Chalcolithic_1_Afanasievo cluster together with West Eurasians.

PC2 reveals a "south-north cline" within East Asia defined by the Southeast Asian Cluster (Tai-Kadai speakers in southern mainland China and Austronesian speakers in Taiwan Island) at one extreme and Eskimo-Inupiat speakers in northern Siberia and Native Americans on the other. Ancient Taiwan samples from the Hanben and Gongguan sites project closely together within the Southeast Asian Cluster including ancient populations in southern mainland China[9] and Southeast Asia. The ancient Devil's Gate, Boisman, AR_EN and the ancient Yankovsky project close to present-day populations in the Amur River Basin, at an intermediate position in the south-north cline. The Neolithic Wuzhuangguoliang samples from northern China project with present-day northern Han Chinese and other ancient populations in northern China published in Ning et al[7] and Yang et al[9]. Tibeto-Burman speaking populations on the Tibetan Plateau show complex patterns. Tibetans in Tibet plot tightly between the Amur River cluster and Han Chinese, but slightly off of the south-north cline. Northern Tibetans deviate in the West Eurasian direction. In the Tibeto-Burman Corridor, Tibetans, Lolo-Burmese speakers (Yi, Naxi, and Lahu), and Qiang are scattered along the south-north cline.

We also projected ancient samples onto the PCA inferred using only East Asian

populations after removing Uygur, Kazakh_China, Kyrgyz_China, Salar, Dongxiang, Bonan, and Yugur in northwest China from the dataset as they deviate towards West Eurasians in the East Asian-Siberian PCA (Extended Data Fig. 1, Extended Data Fig. 2, Figure SI2-1). The first two PCs correlate to the geographic map of East Asia and linguistic categories. There is a clear south-north cline from indigious populations in Taiwan and the Southeast Asia, to populations in the Amur River Basin and Tibetan Plateau. Han Chinese appear intermediate between the south and the north. We also observed genetic substructure in Han Chinese, with northern Han Chinese closer to ancient and present-day populations in the Amur River Basin, Yellow River Neolithic farmers (including Wuzhuangguoliang) and Tibetan Plateau, and southern Han Chinese deviating toward populations in Mainland Southeast Asia and southern China. Populations of the same region tend to group together, forming Taiwan Island, southern China, Han Chinese, Tibetan Plateau, and Amur Basin clusters. The ancient Devil's Gate, Boisman, AR_EN and Yankovsky samples project within the Amur Basin cluster. There is substructure within Southeast Asian groups. Mainland Southeast Asians (Vietnamese, Cambodian, and Kinh) plot in the southern China cluster, while present-day island individuals from Taiwan are outliers probably due to the genetic drift. Miao and She show closer affinity with Han Chinese. Populations in the Tibeto-Burman Corridor are intermediate between Han Chinese and Tibetans in Tibet. Our ancient Jomon samples project on the direction driven by Japanese and Amur River Basin populations, showing affinity with those populations. Mongolia ancient samples project into two separate clusters at an intermediate position between the Amur River Basin cluster and Tibetan Plateau cluster.

ADMIXTURE analysis

We carried out model-based clustering analysis using ADMIXTURE 1.23[10], combining the present-day worldwide populations with our newly reported ancient individuals. We used PLINK v1.90[11] to thin the dataset of 597,573 autosomal SNPS to remove SNPs in strong linkage disequilibrium, employing a window of 200 SNPs advanced by 25 SNPs and an r^2 threshold of 0.4 (with the flag: --indep-pairwise 200 25 0.4). A total of 335,676

SNPs remained for analysis after this procedure. We ran ADMIXTURE with default 5-fold cross-validation (--cv=5), varying the number of ancestral populations between K=2 and K=15 in 100 bootstraps with different random seeds. We used unsupervised ADMIXTURE, in which allele frequencies for non-admixed ancestral populations are unknown and are computed during the analysis. We used point estimation and terminated the block relaxation algorithm when the objective function delta < 0.0001. We chose the best run according to the highest log likelihood.

We used cross-validation to identify an "optimal" number of clusters. However, we report the full results for K=2-15, as we are interested in the genetic structure revealed by ADMIXTURE in successive models with increasing number of "ancestral populations" (increasing K) (Figure SI2-2, Extended Data Fig. 4). In what follows we describe some qualitative patterns evident in this analysis.

K=2: African and West Eurasian populations (orange) separate from East Eurasians, Native Americans and Oceanians (light blue).

K=3: An African component (yellow) separates from the West Eurasian component. The ancient Mongolia samples, populations in northwest China, northern Tibetans in Qinghai and Gansu, and populations in Siberia and Nepal are assigned some ancestry from the West Eurasian component.

K=4: Native Americans form their own component (pink), which is widely distributed in Siberian and East Asian populations, but nearly absent in populations of southern China and Taiwan Island. The Amur River cluster derives approximately half of its ancestry from the Native American component.

K=5: A component maximized in Oceanians appears (dark green), which is also distributed in ancient Southeast Asians and Jomon.

K=6: The Native American-associated component separates into two parts, with one component (purple) maximized in southern Native Americans (Karitiana and Surui) and one (pink) maximized in Siberian populations such as Nganasan. Most East Asian populations have some of the Siberian component.

K=7: A component (red) maximized in Tibetans appears, which is also widely distributed in Neolithic populations in northern China, Han Chinese, Japanese, Koreans, and some other populations in Southeast Asia, southern China, northwest China and southern Siberia.

K=8: Ancient Jomon samples are assigned their own component (light green), which is represented to a lesser extent in present-day Japanese and also detected in populations of the Amur River Basin cluster.

K=9: Onge obtains its own separate component (dark blue), which is also shown in Tibetans.

K=10: A component (dark pink) maximized in ancient Yamnaya, Afanasievo, Sintashta and Mongolian samples appears, which is also represented to a lesser extent in Siberian populations.

K=11: Far Eastern Siberians, such as Itelmen, Koryak, Chukchi, and Eskimo, are assigned their own component again (dark blue), which separate them from other Siberian populations.

K=12: Jomon related component (brown) is prevalent in Japanese and Amur River Basin groups.

K=13: A new component (cyan) that is maximized in ancient samples from Southeast Asia and Taiwan appears, which is also shown in present-day Austronesian speaking

populations and Neolithic populations in northern China. Onge obtains its own separate component.

K=14: Jomon related component separates into two parts, of which one (brown) is maximized in Jomon samples while the other (light green) is prevalent in Japanese.

K=15: A component (light salmon pink) prevalent in Tibetan and Han Chinese appears.

Genetic substructure in Tibetans

As discussed in Principal Component Analysis and ADMIXTURE analysis, northern Tibetans in Gangcha, Gannan, and Xunhua have more West Eurasian related admixture than southern Tibetans in Tibet and the Tibeto-Burman Corridor, a phenomenon that we also highlight in Table S3 and Table S4. Northern Tibetans also share less genetic drift with other Tibetan populations in outgroup f_3-statistics, which we hypothesize reflects their West Eurasian-related ancestry component (Extended Data Fig. 7). Tibetans in the Tibeto-Burman Corridor and Lolo-Burmese populations (Yi, Naxi, and Lahu) share more alleles with Southeast Asian Cluster in southern China and Taiwan than do Tibetans in Tibet itself (Extended Data Fig. 7). Based on these observations, we classify Tibetan populations into three groups based on outgroup f_3- and f_4-statistics: northern Tibetan, Tibet Tibetan, and Corridor Tibetan.

Summary

Qualitatively, East Eurasian genetic structure is strongly correlated with geography and linguistic categories, with important exceptions. We defined four genetic clusters within East Asia, correlating strongly to Amur River Basin, Han Chinese, Southeast Asian Cluster (populations in southern China, Taiwan Island, and mainland Southeast Asia), and Tibeto-Burman on the Tibetan Plateau. We observe two genetic subclusters within Mongolia: one falls closer to ancient individuals from the Amur Basin Cluster (hence termed ‘East’), and the second clusters towards ancient individuals of Afanasievo culture (henceforth called ‘West’), while a few individuals take intermediate positions

between the two.

References

1. Cavalli-Sforza, L. L. The Chinese human genome diversity project. Proc Natl Acad Sci U S A 95, 11501-11503 (1998)
2. HUGO Pan-Asian SNP Consortium. Mapping human genetic diversity in Asia. Science 326, 1541-1545 (2009)
3. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. PLoS Genet. 2, e190 (2006)
4. Price, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38, 904-909 (2006)
5. Price, A. L., Zaitlen, N. A., Reich, D. & Patterson, N. New approaches to population stratification in genome-wide association studies. Nat. Rev. Genet. 11, 459-463 (2010)
6. Raghavan, M. et al. The genetic prehistory of the New World Arctic. Science, 345, 1255832 (2014)
7. Ning, C., et al. Ancient genomes from northern China suggest links between subsistence changes and human migration. *Nat. Commun* **11**, 2700 (2020).
8. de Barros Damgaard, P., et al.. 137 ancient human genomes from across the Eurasian steppes. *Nature* **557**, 369-374 (2018).
9. Yang, M. A., et al. Ancient DNA indicates human population shifts and admixture in northern and southern China. *Science* **369**, 282-288 (2020).
10. Alexander, D. H., Novembre, J., & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res, 19, 1655-1664 (2009)
11. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4, 7 (2015)
12. Kanzawa-Kiriyama, H., et al. A partial nuclear genome of the Jomons who lived 3000 years ago in Fukushima, Japan. *J. Hum. Genet* **62**, 213–221 (2016).

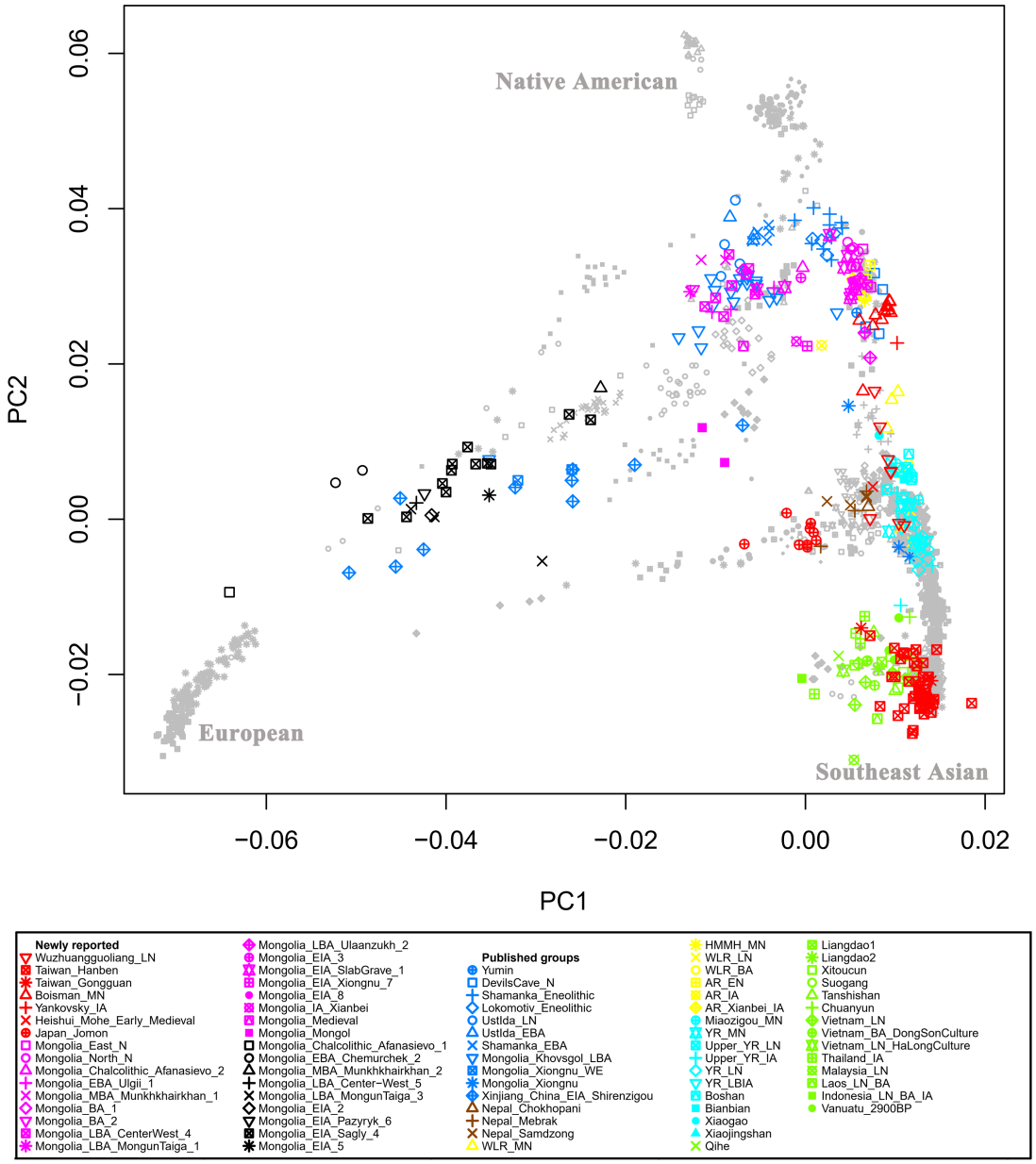
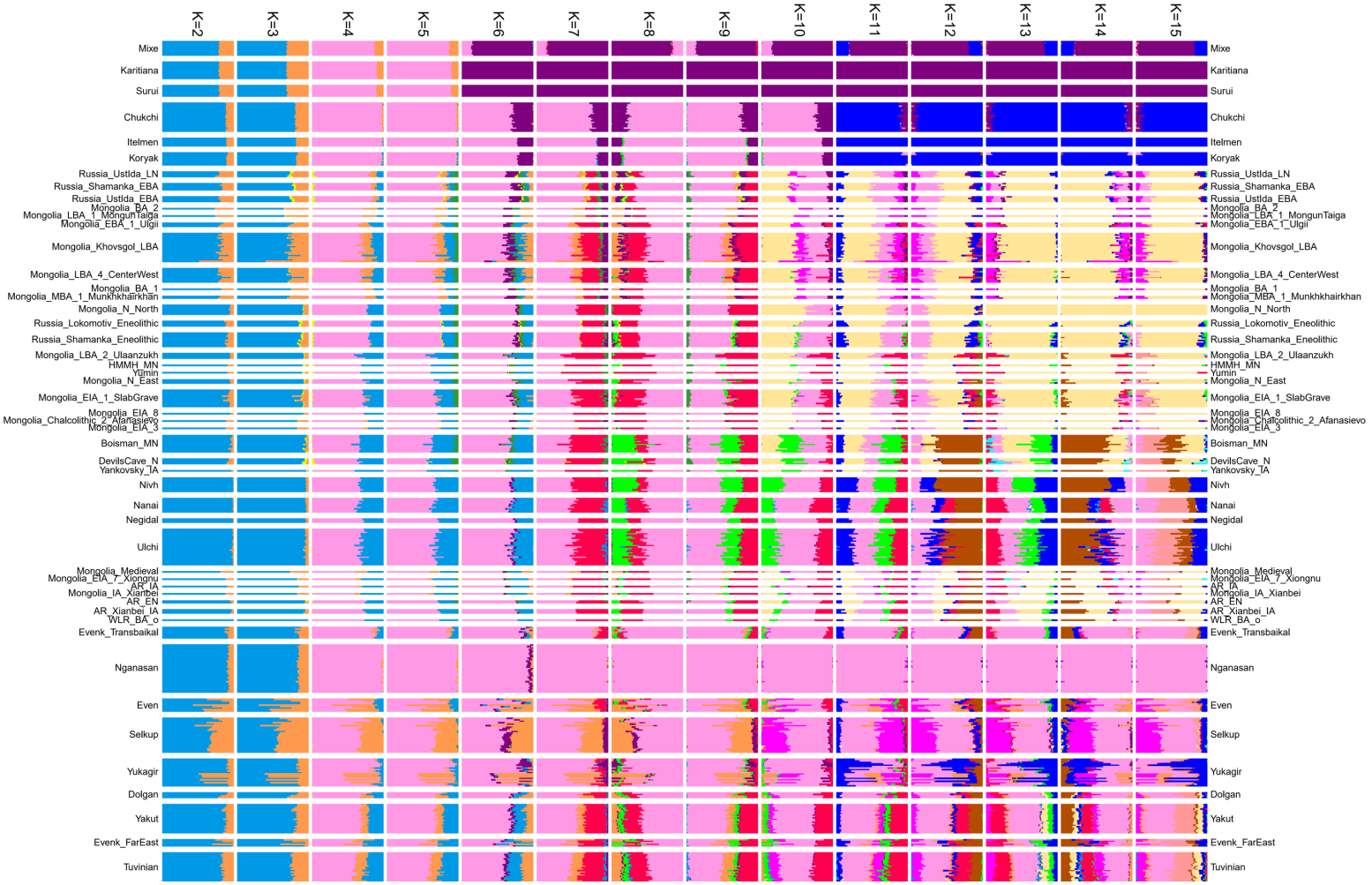
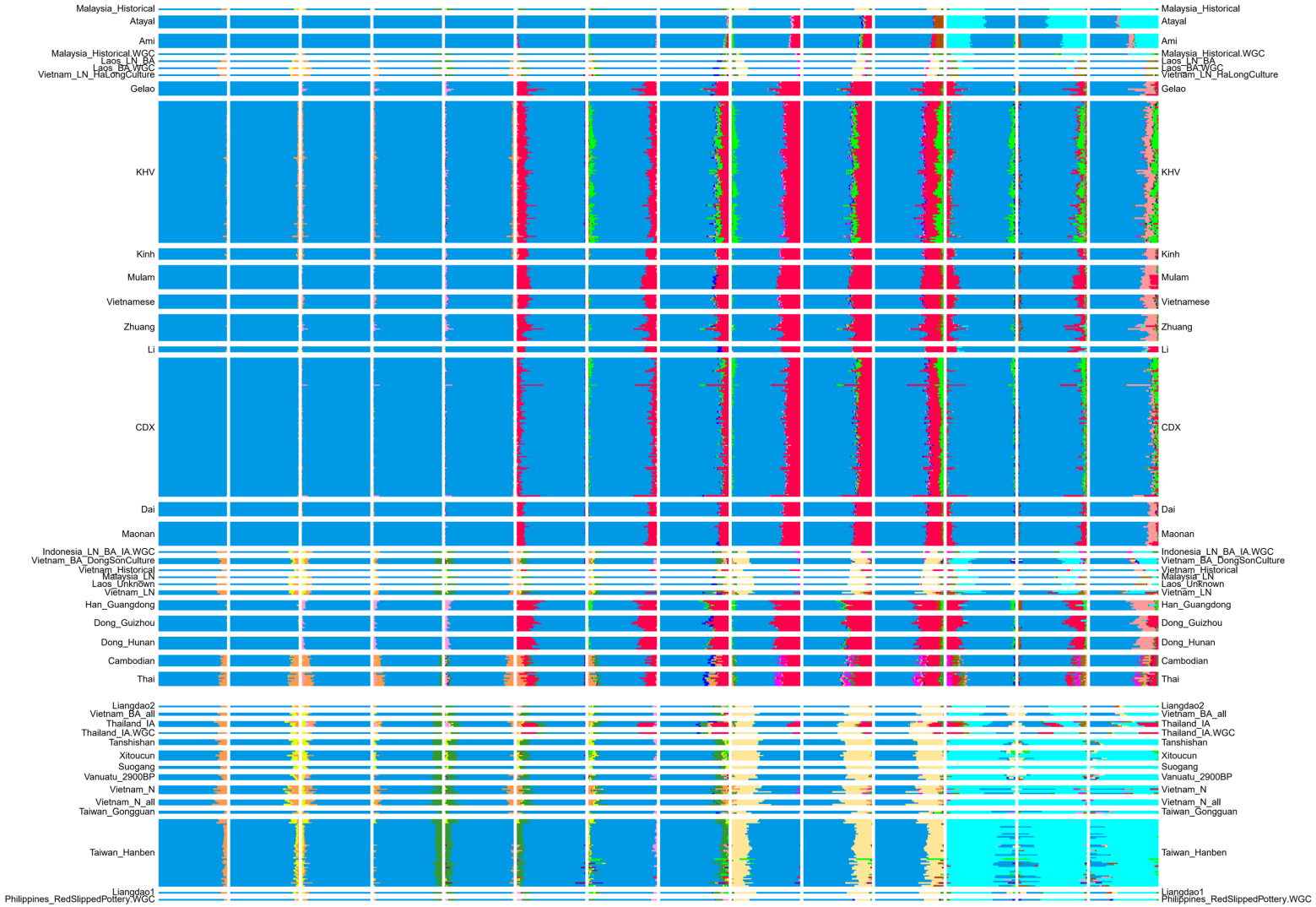
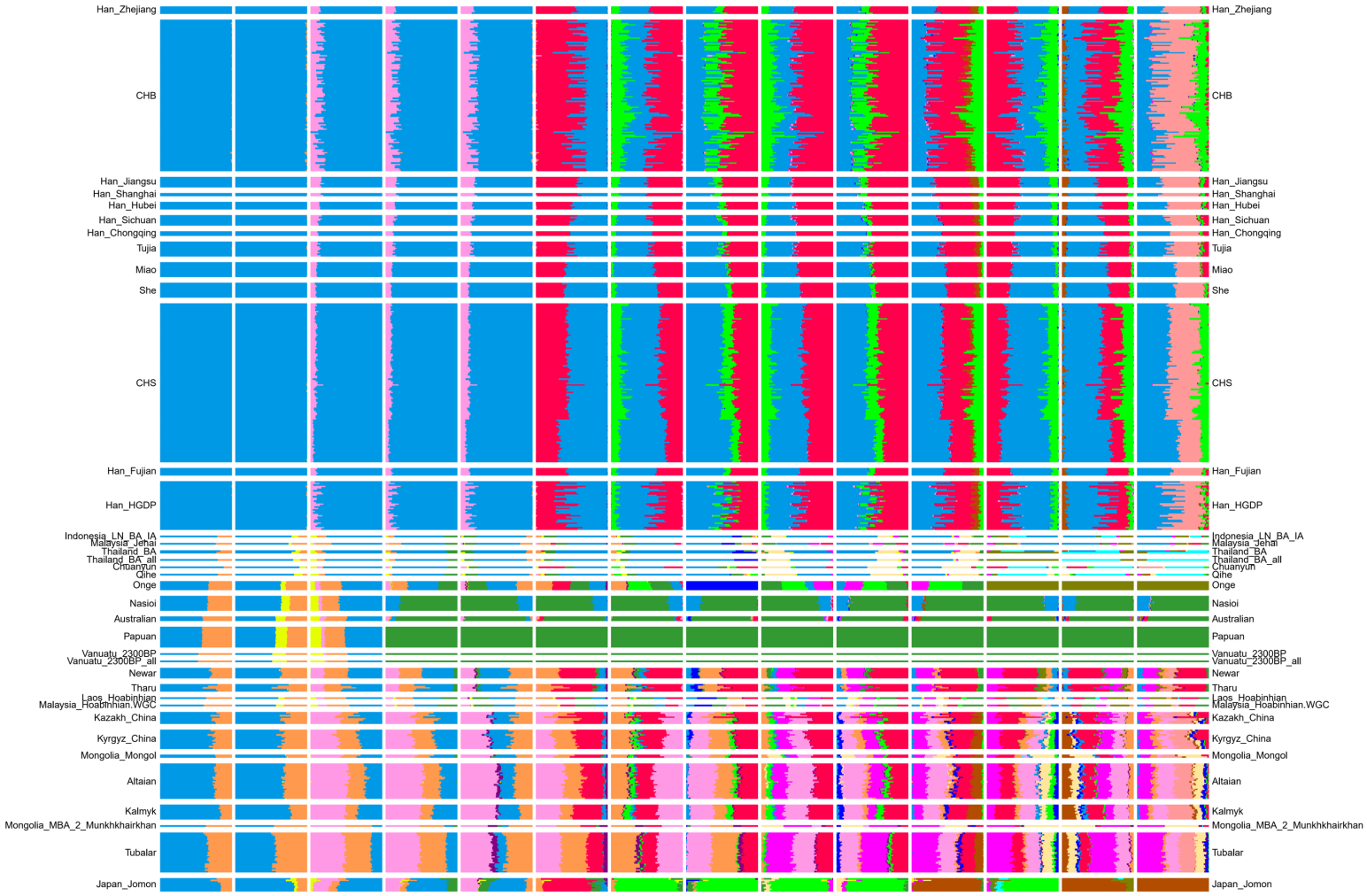
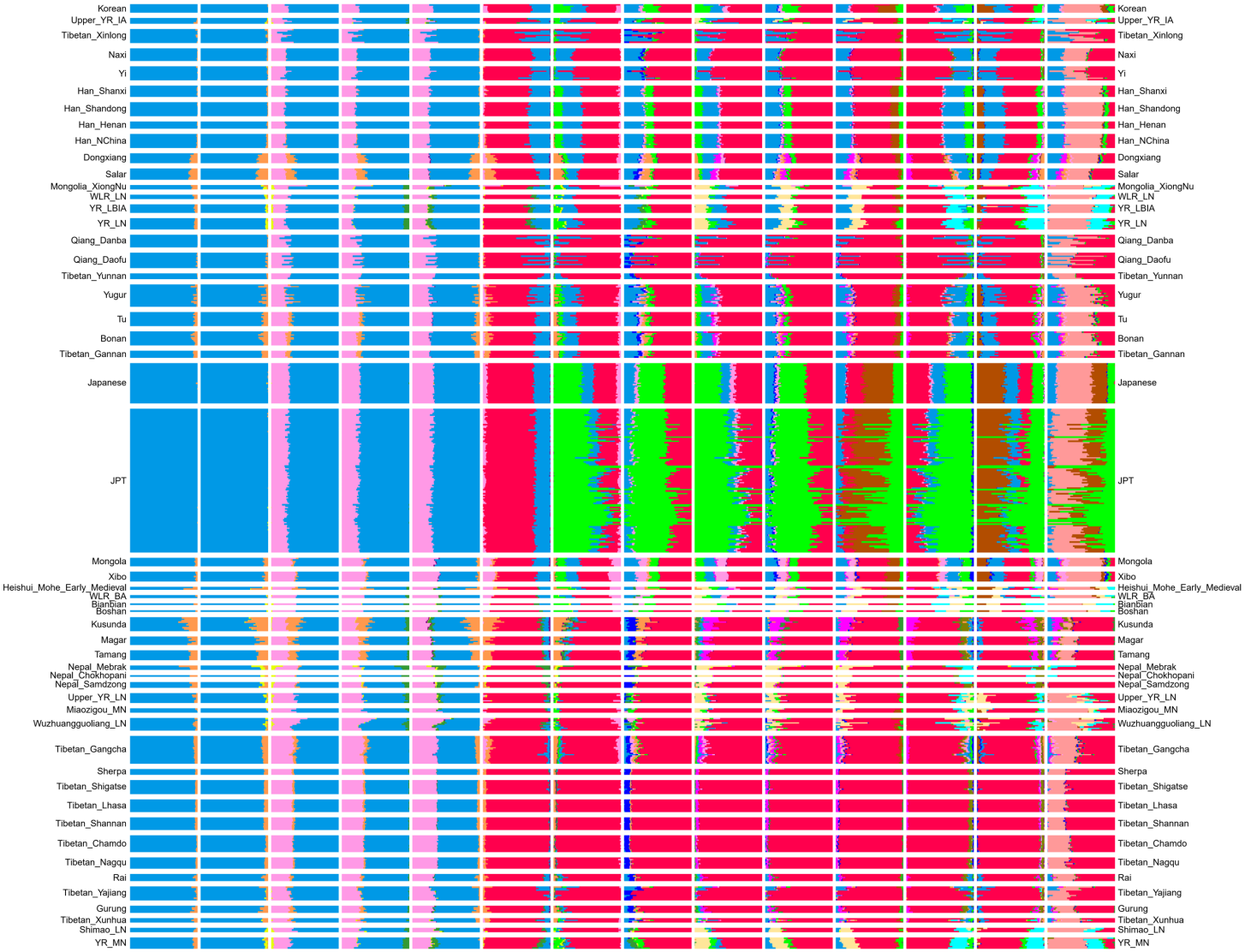


Figure SI2-1: Principal Component Analysis (PCA). PCA dimensions 1 and 2 defined by present-day East Asians, Europeans, Siberians and Native Americans. We used both the ancient and modern populations as base populations to fill the missing data with an average of the allele frequencies by removing "lsqproject: YES" and "poplistname" option from parameter setting in the smartpca program.


K=2
K=3
K=4
K=5
K=6
K=7
K=8
K=9
K=10
K=11
K=12
K=13
K=14
K=15
Mixe
Karitiana
Surui
Chukchi
Itelmen
Koryak
Russia_UstIda_LN
Russia_Shamanka_EBA
Russia_UstIda_EBA
Mongolia_BA_2
Mongolia_LBA_1_MongunTaiga
Mongolia_EBA_1_Ulgii
Mongolia_Khovsgol_LBA
Mongolia_LBA_4_CenterWest
Mongolia_BA_1
Mongolia_MBA_1_Munkhkhairkhan
Mongolia_N_North
Russia_Lokomotiv_Eneolithic
Russia_Shamanka_Eneolithic
Mongolia_LBA_2_Ulaanzukh
HMMH_MN
Yumin
Mongolia_N_East
Mongolia_EIA_1_SlabGrave
Mongolia_EIA_8
Mongolia_Chalcolithic_2_Afanasievo
Mongolia_EIA_3
Boisman_MN
DevilsCave_N
Yankovsky_TA
Nivh
Nanai
Negidal
Ulchi
Mongolia_Medieval
Mongolia_EIA_7_Xiongnu
AR_IA
Mongolia_IA_Xianbei
AR_EN
AR_Xianbei_IA
WLR_BA_o
Evenk_Transbaikal
Nganasan
Even
Selkup
Yukagir
Dolgan
Yakut
Evenk_FarEast
Tuvinian


Malaysia_Historical
Atayal
Ami
Malaysia_Historical.WGC
Laos_LN_BA
Laos_BA.WGC
Vietnam_LN_HaLongCulture
Gelao
KHV
Kinh
Mulam
Vietnamese
Zhuang
Li
CDX
Dai
Maonan
Indonesia_LN_BA_IA.WGC
Vietnam_BA_DongSonCulture
Vietnam_Historical
Malaysia_LN
Laos_Unknown
Vietnam_LN
Han_Guangdong
Dong_Guizhou
Dong_Hunan
Cambodian
Thai
Liangdao2
Vietnam_BA_all
Thailand_IA
Thailand_IA.WGC
Tanshishan
Xitoucun
Suogang
Vanuatu_2900BP
Vietnam_N
Vietnam_N_all
Taiwan_Gongguan
Taiwan_Hanben
Liangdao1
Philippines_RedSlippedPottery.WGC


Han_Zhejiang
CHB
Han_Jiangsu
Han_Shanghai
Han_Hubei
Han_Sichuan
Han_Chongqing
Tujia
Miao
She
CHS
Han_Fujian
Han_HGDP
Indonesia_LN_BA_IA
Malaysia_Jehai
Thailand_BA
Thailand_BA_all
Chuanyun
Qihe
Onge
Nasioi
Australian
Papuan
Vanuatu_2300BP
Vanuatu_2300BP_all
Newar
Tharu
Laos_Hoabinhian
Malaysia_Hoabinhian.WGC
Kazakh_China
Kyrgyz_China
Mongolia_Mongol
Altaian
Kalmyk
Mongolia_MBA_2_Munkhkhairkhan
Tubalar
Japan_Jomon


Korean
Upper_YR_IA
Tibetan_Xinlong
Naxi
Yi
Han_Shanxi
Han_Shandong
Han_Henan
Han_NChina
Dongxiang
Salar
Mongolia_XiongNu
WLR_LN
YR_LBIA
YR_LN
Qiang_Danba
Qiang_Daofu
Tibetan_Yunnan
Yugur
Tu
Bonan
Tibetan_Gannan
Japanese
JPT
Mongola
Xibo
Heishui_Mohe_Early_Medieval
WLR_BA
Bianbian
Boshan
Kusunda
Magar
Tamang
Nepal_Mebrak
Nepal_Chokhopani
Nepal_Samdzong
Upper_YR_LN
Miaozigou_MN
Wuzhuangguoliang_LN
Tibetan_Gangcha
Sherpa
Tibetan_Shigatse
Tibetan_Lhasa
Tibetan_Shannan
Tibetan_Chamdo
Tibetan_Nagqu
Rai
Tibetan_Yajiang
Gurung
Tibetan_Xunhua
Shimao_LN
YR_MN

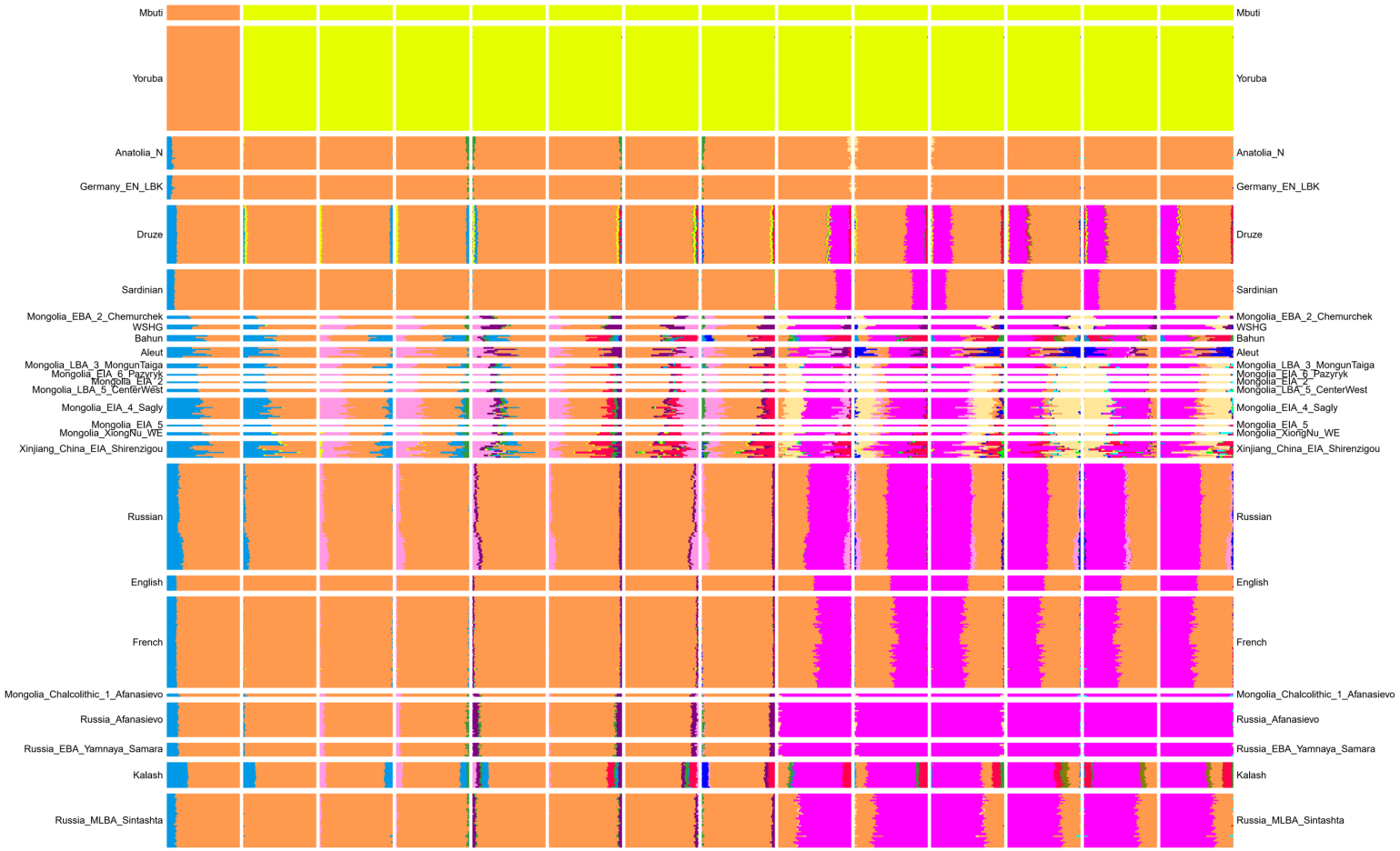


Figure SI2-2: ADMIXTURE analysis from K=2 to K=15.

Supplementary Information section 3: Admixture graph modeling

In this section, we use f_4-statistics and *qpGraph (version 6750)*,[1] as implemented in ADMIXTOOLS,[2] to investigate models of possible phylogenetic relationships for the representative East Asian populations. *qpGraph* assesses the fit of Admixture Graph models to data by comparing fitted and estimated f_2-, f_3-, and f_4-statistics. We used all sites of the 1240K dataset to explore the phylogeny. We used ancient Mongolia_East_N, Upper_YR_LN (Upper Yellow River Late Neolithic), Liangdao2, Boisman, Chokhopani, Taiwan_Hanben, and Japan_Jomon to represent the distinct genetic clusters within East Asia. We use the following parameters in running the *qpGraph* program:

outpop: NULL

blgsize: 0.05

lsqmode: NO

diag: 0.0001

hires: YES

initmix: 1000

precision: 0.0001

zthresh: 0

terse: NO

useallsnps: NO

In what follows, we only accept models as "fits" if all possible f_2-, f_3-, and f_4- statistics have a |Z|-score of less than 3 between the observed and expected values, as determined by the Block Jackknife as implemented in *qpGraph*. We also inspect the likelihood scores output by *qpGraph* to further differentiate between models, and do not accept models that have internal branch lengths of zero as these may indicate error in the graph topology. We don't set strict criteria for the likelihood scores to reject graph models, but use the likelihood scores in combination with |Z|-scores to evaluate models.

We used MSMC[3] and MSMC2[4] to obtain estimates of separation times between major branches of our model. Specifically, we inferred relative cross-coalescence rates among Ami/Atayal, Tibetan, and Ulchi to represent ancient Taiwan_Hanben, Chokhopani, and Mongolia_Neolithic/Boisman respectively to provide insight into the relative timing of the population splits within East Asia. We used Mixe as an outgroup of present-day East Asians for comparison. All genomes used in this analysis are from the SGDP dataset[5]. In the MSMC analyses (Figure SI3-1), we used 2 haplotypes (1 individual) per population. In the MSMC2 analysis (Figure SI3-2) we used 4 haplotypes per population where possible, and 2 haplotypes for the Atayal/Tibetan comparison, since there is only one Atayal genome available through the SGDP dataset[5]. While the split times between Tibetan, Ulchi and Ami/Atayal are very similar on average when compared to the split time between Tibetan and Mixe (Figures SP1 and SP2), we see a slightly more recent separation between Tibetan and Ulchi as opposed to Tibetan/Ami and Tibetan/Atayal. It is important to realize that these splits represent averages of multiple lineage separations contributing to the groups being compared. Thus, it is possible that a subset of the lineages contributing jointly to Tibetan and Ami/Atayal may be more closely related than some of the lineages contributing jointly to Tibetan and Ulchi. We used the suggestive evidence from this analysis to help constraining the *qpGraph* models, prioritizing models that imply closer relatedness of Tibetans and Ulchi.

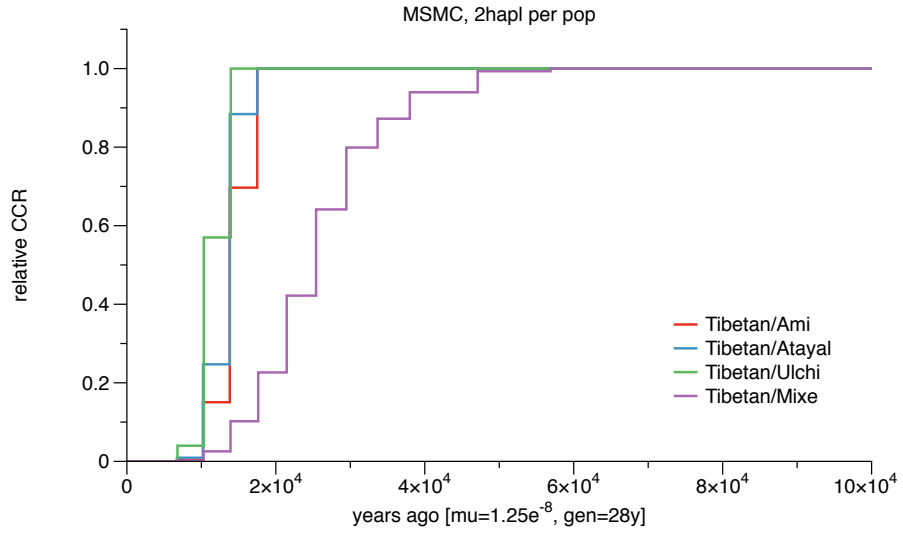


Figure SI3-1: Cross-coalescence rates for selected population pairs. We ran MSMC for four pairs of populations: Tibetan-Ami, Tibetan-Atayal, Tibetan-Ulchi and Tibetan-Mixe. We used

one individual from each population in this analysis. The genome data for those samples are from the Simons Genome Diversity Project[5]. The times are calculated based on the mutation rate and generation time specified on the x-axis.

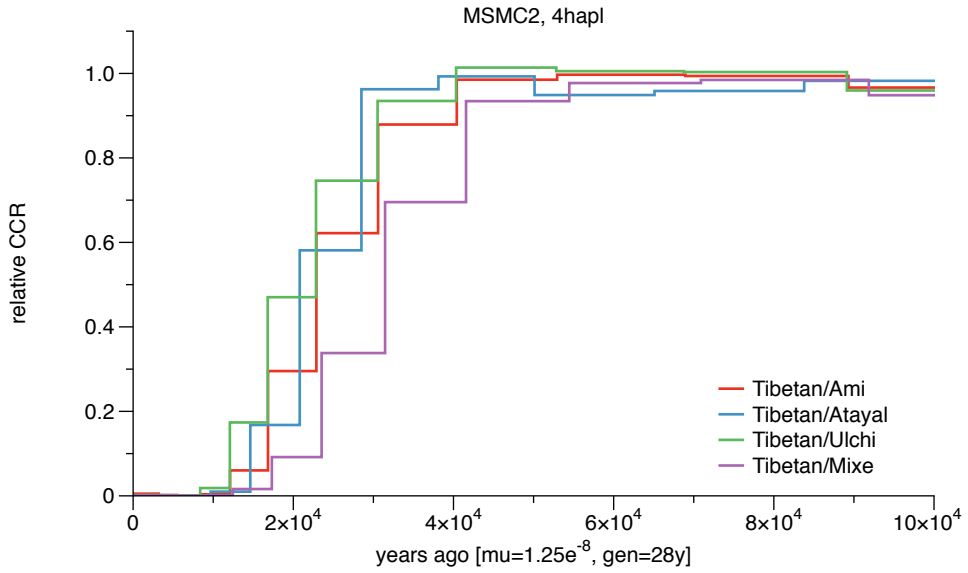


Figure SI3-2: Cross-coalescence rates for selected population pairs. Same analysis as in Figure SI3-1, but using MSMC2 instead of MSMC, and using two individuals per population except for the Tibetan-Atayal pair, where we used only one.

For the admixture graph construction, we first used the 1240K dataset. We started with a skeleton phylogenetic tree consisting of Denisovan, Mbuti, Loschbour, Tianyuan, and Onge (ONG_SG) with one admixture from Denisovan to Tianyuan (Figure SI3-3). The tree was an adequate fit with the likelihood score of 0.541 and the worst Z-score of *f*-statistic (Mbuti, Denisovan; ONG, Loschbour) = 0.666. We note that there are known features of human history that this admixture graph does not model including the Neanderthal admixture into the ancestors of all non-Africans. However, we find that this is not necessary to model in order to obtain a fit (probably because all the non-African populations have very similar proportions of Neanderthal-related ancestry) so we leave it out from the modeling for simplicity.

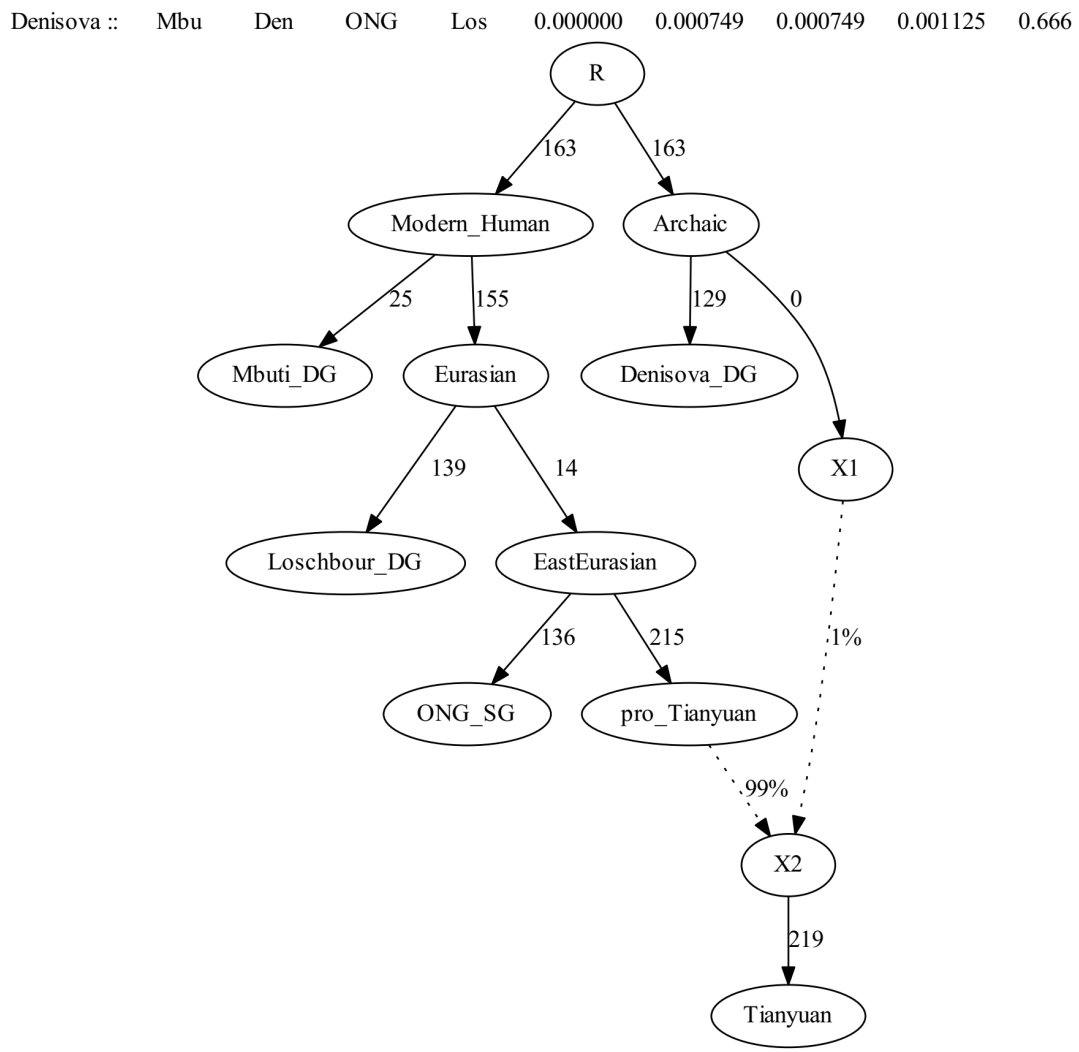


Figure SI3-3: Skeleton Admixture Graph with one admixture using 650168 SNPs, which is a fit to the genetic data in the sense that there are no *f*-statistics more than |Z|>3 different between model and expectation. The drifts along edges in this figure and the following ones are multiplied by 1000.

We then added Mongolia_East_N onto all the possible edges of the graph and found a fitted model with no additional admixture event by assigning Mongolia_East_N as a sister branch of Tianyuan (Figure SI3-4).

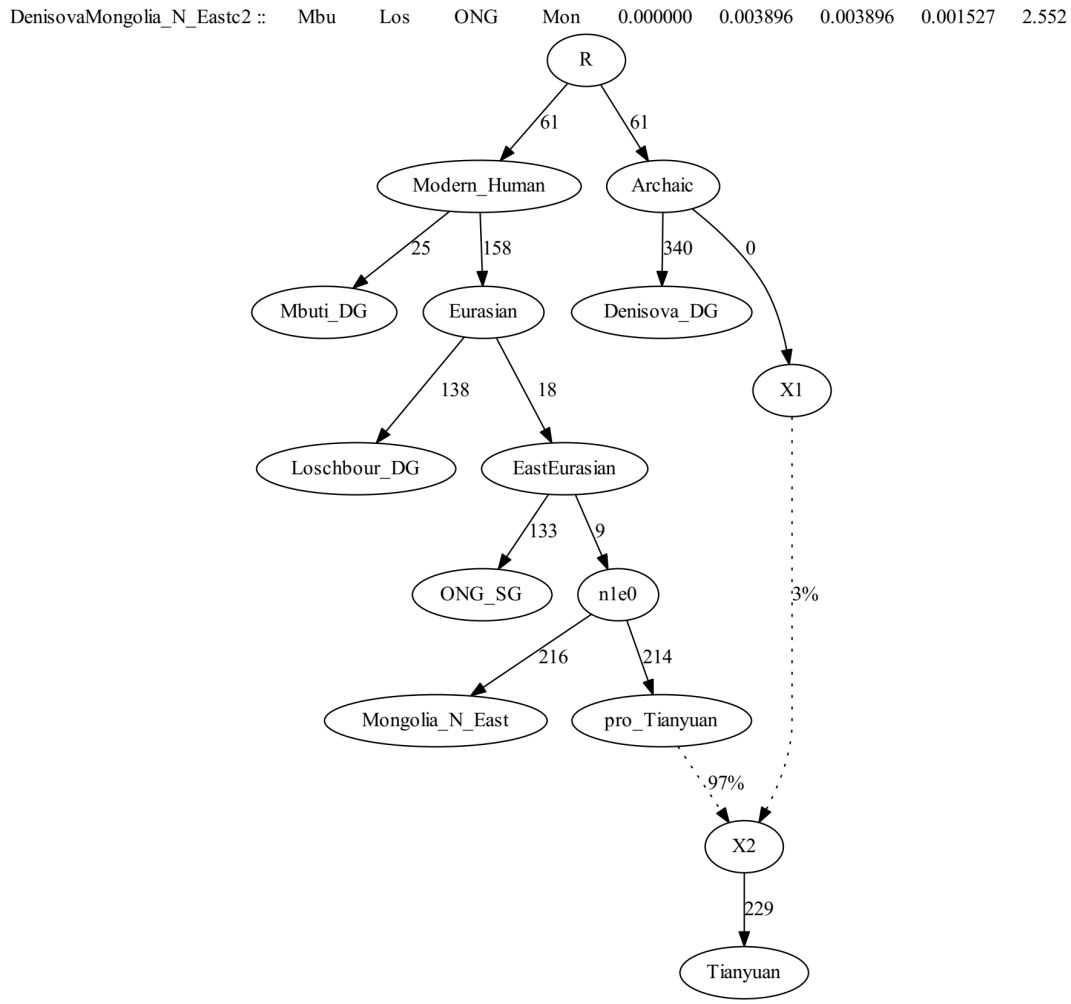


Figure SI3-4: The best fitting tree of adding Mongolia_East_N without admixture was an adequate fit as in the sense that there are no *f*-statistics more than |Z|>3 different between model and expectation. The likelihood score is 10.717 and the worst Z-score of *f*-statistic (Mbuti, Loschbour; ONG, Mongolia_East_N) = 2.552. The number of SNPs used is 620655.

We note that the worst Z-score of *f*-statistic (Mbuti, Loschbour; ONG, Mongolia_East_N) = 2.552 in Figure SI3-4 gives suggestive evidence for the possible West Eurasian related genetic influence in Mongolia_East_N. If we added one more admixture event into the graph, we found the lowest Z-score was reached by introducing 1% of West Eurasian related ancestry into Mongolia_East_N (Figure SI3-5). To account for the possible West Eurasian influence in our East Asian modeling, we next added populations onto the model as in Figure SI3-5.

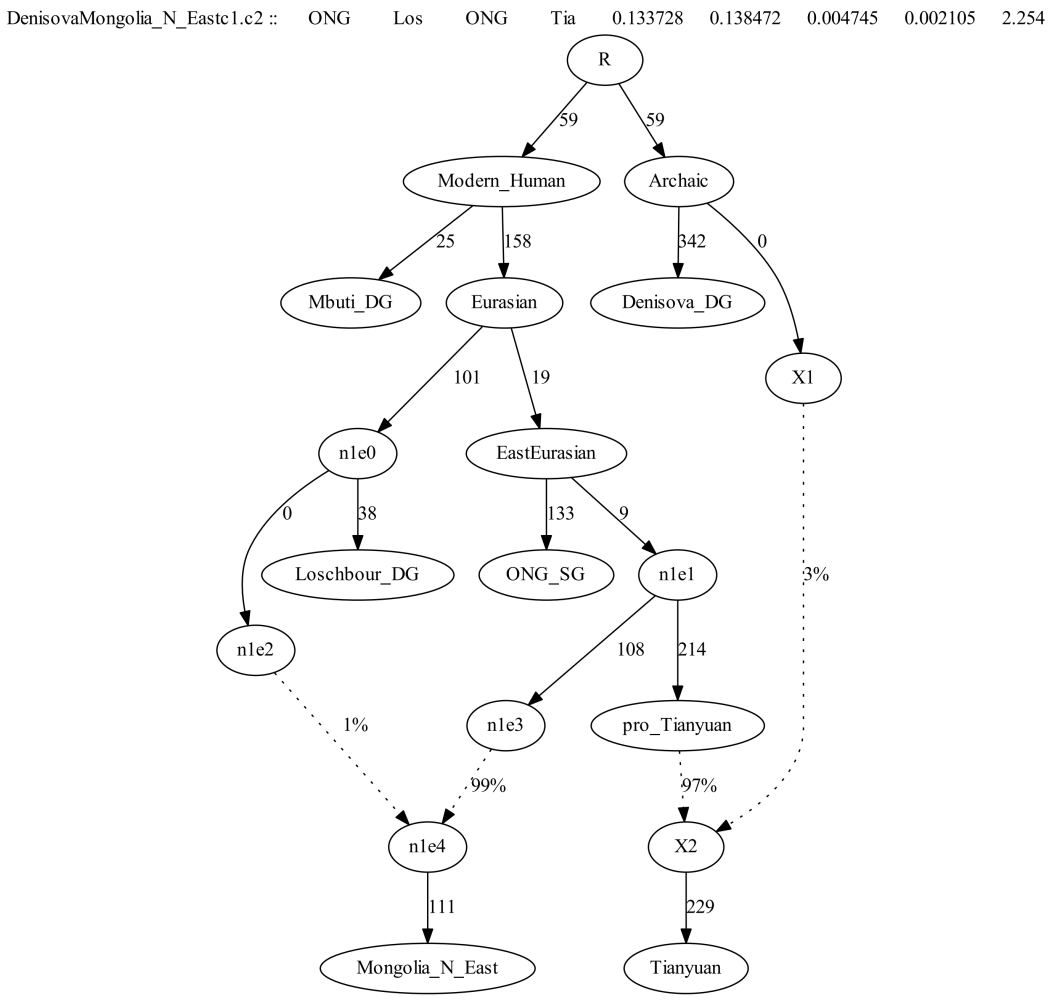


Figure SI3-5: The best fitting tree of adding Mongolia_East_N with one admixture was an adequate fit as in the sense that there are no *f*-statistics more than |Z|>3 different between model and expectation. The likelihood score is 9.641 and the worst Z-score of *f*-statistic (ONG, Loschbour; ONG, Tianyuan) = 2.254. The number of SNPs used is 620655.

We added Upper_YR_LN onto all possible edges of the graph. We obtained two fitted graphs for modeling Upper_YR_LN as unadmixed as shown below in the Figure SI3-6, in which Upper_YR_LN is a sister group of the Mongolia_East_N related lineage. However, the worst |Z|-scores for the two fitted models are larger than 2.9, suggesting there is unmodeled affinity between Upper_YR_LN and Onge related lineage. We also obtained 26 fitted graphs for modeling Upper_YR_LN samples as admixed as shown below in Table SI3-1. The two fitted models Upper_YR_LNc3.d5 and Upper_YR_LNc3.d6 with one additional admixture with the lowest Z-

scores and likelihood scores are shown in Figure SI3-7. The difference between those two models is whether Upper_YR_LN has West Eurasian related ancestry (<1%) deriving from Mongolia_East_N or not. Our current data and model could not distinguish such a low percentage of ancestry in the above competing models. We chose the model Upper_YR_LNc3.d5 to continue our population modeling.

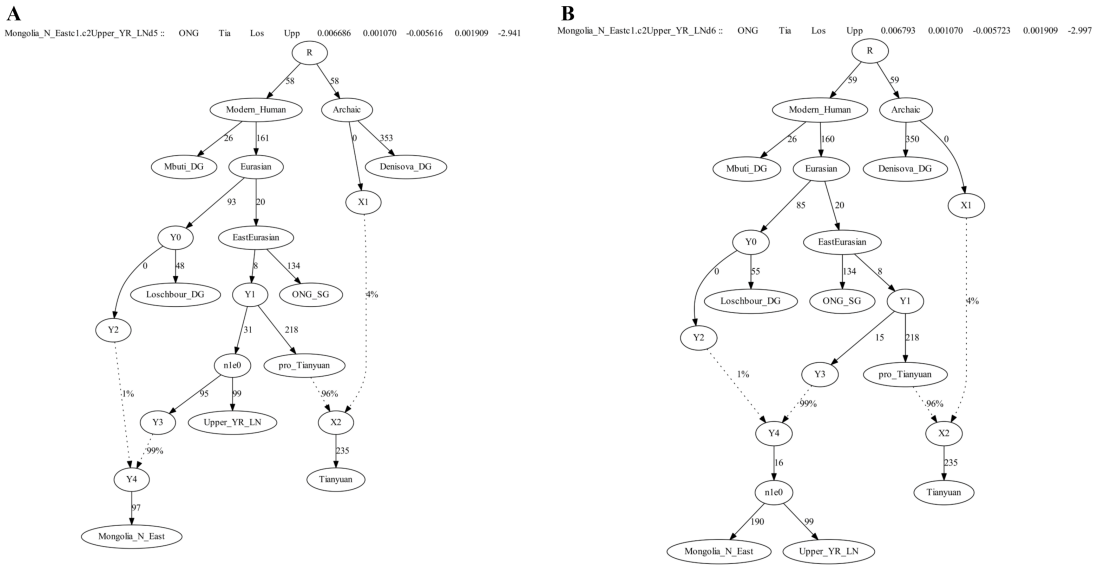


Figure SI3-6: The two fitting trees of adding Upper_YR_LN without admixture were adequate fits as in the sense that there are no *f*-statistics more than |Z|>3 different between model and expectation. Model A: The likelihood score is 20.249 and the worst Z-score of *f*-statistic (ONG,Tianyuan; Loschbour, Upper_YR_LN) = -2.941; Model B: The likelihood score is 21.684 and the worst Z-score of f-statistic (ONG,Tianyuan; Loschbour, Upper_YR_LN) = -2.997. The number of SNPs used is 629654.

Table SI3-1. The fitted models for modeling Upper_YR_LN samples onto all the possible edges of the graph with one additional admixture in Figure SI3-5.

Model	likelihood score	worst Z-score
Upper_YR_LNc3.d5	10.139	2.248
Upper_YR_LNc3.d6	10.165	2.189
Upper_YR_LNc0.d5	13.232	2.772
Upper_YR_LNd1.d5	13.254	2.772
Upper_YR_LNd1.d6	13.564	2.798

Upper_YR_LNb0.d5	19.942	-2.940
Upper_YR_LNd3.d6	20.074	-2.994
Upper_YR_LNd0.d5	20.075	-2.986
Upper_YR_LNd5.d6	20.095	-2.990
Upper_YR_LNd4.d5	20.140	-2.983
Upper_YR_LNd3.d5	20.152	-2.938
Upper_YR_LNd2.d5	20.188	-2.986
Upper_YR_LNb1.d5	20.242	-2.940
Upper_YR_LNb2.d5	20.281	-2.942
Upper_YR_LNa1.d5	20.309	-2.942
Upper_YR_LNa0.d5	20.310	-2.947
Upper_YR_LNb3.d5	20.395	-2.977
Upper_YR_LNc4.d5	20.924	-2.990
Upper_YR_LNb0.d6	21.349	-2.995
Upper_YR_LNd0.d6	21.586	-2.998
Upper_YR_LNb1.d6	21.671	-2.996
Upper_YR_LNd4.d6	21.671	-2.996
Upper_YR_LNb2.d6	21.712	-2.995
Upper_YR_LNb3.d6	21.729	-2.985
Upper_YR_LNa0.d6	21.744	-2.994
Upper_YR_LNd2.d6	21.745	-2.992

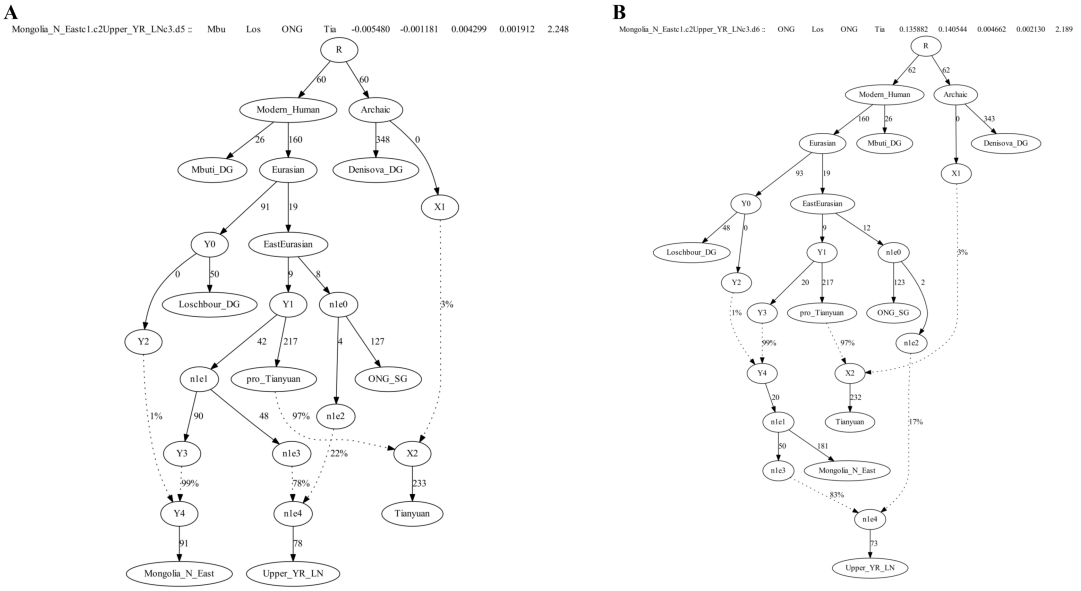


Figure SI3-7: The two fitting trees of adding Upper_YR_LN with one additional admixture were adequate fits in the sense that there are no *f*-statistics more than |Z|>3 different between model and expectation. Model A-Upper_YR_LNc3.d5: The likelihood score is 10.139 and the worst Z-score of *f*-statistic (Mbuti, Loschbour; ONG,Tianyuan) =2.248; Model B- Upper_YR_LNc3.d6: The likelihood score is 10.165 and the worst Z-score of f-statistic (ONG, Loschbour; ONG, Tianyuan) = 2.189. The number of SNPs used is 629654.

Table SI3-2. The fitted models for modeling Liangdao2 sample onto all the possible edges of the model Upper_YR_LNc3.d5 in Figure SI3-7.

Model	likelihood score	worst Z-score
Liangdao2e1.e4	11.195	2.236
Liangdao2d3.e6	12.200	2.351
Liangdao2d1.e6	12.204	2.352
Liangdao2e1.e6	12.216	2.351
Liangdao2d1.e5	13.852	2.693
Liangdao2d3.e5	14.956	-2.779
Liangdao2e1.e5	14.978	-2.777
Liangdao2e0.e1	20.673	2.749
Liangdao2e1.e2	20.786	2.764
Liangdao2b0.e1	23.557	2.847
Liangdao2d3.e3	23.815	2.847
Liangdao2d2.e1	23.849	2.843
Liangdao2d1.e3	23.853	2.847

Liangdao2e1.e3	23.854	2.847
Liangdao2d3.e1	23.855	2.848
Liangdao2b1.e1	23.864	2.846
Liangdao2d4.e1	23.864	2.845
Liangdao2d0.e1	23.871	2.847
Liangdao2b2.e1	23.892	2.846
Liangdao2d1.e1	23.893	2.847
Liangdao2c4.e1	23.896	2.845
Liangdao2c0.e1	23.907	2.846
Liangdao2a0.e1	23.927	2.846
Liangdao2a1.e1	23.927	2.845
Liangdao2e1	23.932	2.849
Liangdao2d6.e1	23.951	2.854
Liangdao2b3.e1	24.031	2.855
Liangdao2d3.d6	26.197	2.943
Liangdao2d1.d6	26.579	2.963
Liangdao2e0.e3	27.091	2.818
Liangdao2c0.e3	28.923	2.909
Liangdao2e2.e3	29.411	2.949

We then added Liangdao2 individual onto all the possible edges of the model Upper_YR_LNc3.d5 in Figure SI3-7 without additional admixture or with only one additional admixture event. We obtained only one fitted graph for modeling Liangdao2 as unadmixed but with a Z-score of 2.849 and 31 fitted graphs for modeling Liangdao2 as admixed with the lowest worst Z-score of 2.236 (Table SI3-2, Figure SI3-8). We could not add more populations onto the model with Liangdao2 as unadmixed because of the large Z-scores.

We chose the model Liangdao2e1.e4 to continue the population adding. We were not able to fit Japan_Jomon into the graph without invoking an additional admixture event. We obtained 32 fitted graphs for modeling Japan_Jomon samples as admixed as shown below in Table SI3-3, of which 25 models having internal zero-length branches. We showed the model Japan_Jomonf1.f4 with the lowest likelihood score and Z-score and without internal zero-length branches in Figure SI3-9.

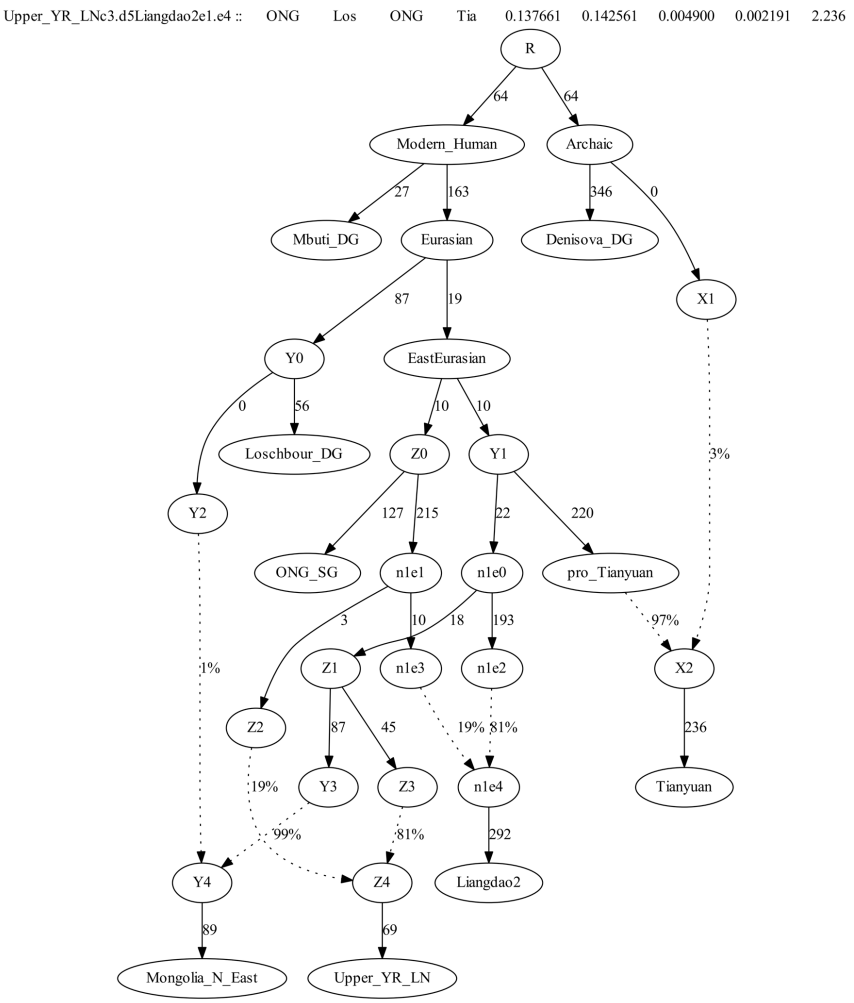


Figure SI3-8: The fitting tree Liangdao2e1.e4 with the lowest worst Z-score and likelihood score of adding Liangdao2 with one additional admixture was an adequate fit as in the sense that there are no *f*-statistics more than |Z|>3 different between model and expectation. The likelihood score is 11.195 and the worst Z-score of *f*-statistic (ONG, Loschbour; ONG,Tianyuan) =2.236. The number of SNPs used is 566071. Liangdao2 and Upper_YR_LN both can be modeled as ~19% Onge-related and ~80% Tianyuan-related ancestry, but they could not be fitted as a simple clade with each other because Upper_YR_LN is closer to Mongolia_East_N compared with Liangdao2.

Table SI3-3. The fitted models for modeling Japan_Jomon samples onto all the possible edges of the model Liangdao2e1.e4 in Figure SI3-8.

Model	likelihood score	worst Z-score	Note
Japan_Jomonf1.f4	17.400	2.385	
Japan_Jomone0.f6	17.679	2.257	internal zero-length branch
Japan_Jomonc0.f6	17.680	2.266	internal zero-length branch
Japan_Jomond1.f6	17.713	2.260	internal zero-length branch
Japan_Jomonf1.f2	19.369	2.391	internal zero-length branch
Japan_Jomonf0.f1	19.371	2.393	internal zero-length branch
Japan_Jomonf1.f6	20.959	2.552	internal zero-length branch
Japan_Jomone2.f6	20.978	2.550	internal zero-length branch
Japan_Jomone0.f4	21.553	-2.573	
Japan_Jomond6.f1	22.772	2.673	
Japan_Jomonc0.f4	22.866	-2.817	internal zero-length branch
Japan_Jomond1.f4	22.886	-2.814	internal zero-length branch
Japan_Jomone3.f1	23.040	2.675	
Japan_Jomonf0.f5	23.140	-2.863	
Japan_Jomonf0.f6	23.183	-2.859	
Japan_Jomond3.f6	23.199	-2.832	internal zero-length branch
Japan_Jomonf0.f3	23.462	-2.787	internal zero-length branch
Japan_Jomone2.f4	23.672	-2.871	internal zero-length branch
Japan_Jomone5.f1	25.817	2.653	internal zero-length branch
Japan_Jomone6.f1	25.849	2.685	
Japan_Jomone0.f2	26.827	-2.808	internal zero-length branch
Japan_Jomone0.f0	26.845	-2.810	internal zero-length branch
Japan_Jomone2.f2	27.211	2.863	internal zero-length branch
Japan_Jomone2.f0	27.227	2.865	internal zero-length branch
Japan_Jomone3.f3	28.198	2.933	internal zero-length branch
Japan_Jomone3.f5	28.272	2.931	internal zero-length branch
Japan_Jomond6.f3	28.378	2.898	internal zero-length branch
Japan_Jomond6.f5	28.378	2.898	internal zero-length branch
Japan_Jomone5.f3	29.227	2.930	internal zero-length branch
Japan_Jomone5.f5	29.346	2.960	internal zero-length branch
Japan_Jomone6.f5	29.365	2.960	internal zero-length branch
Japan_Jomone6.f3	29.934	2.929	internal zero-length branch

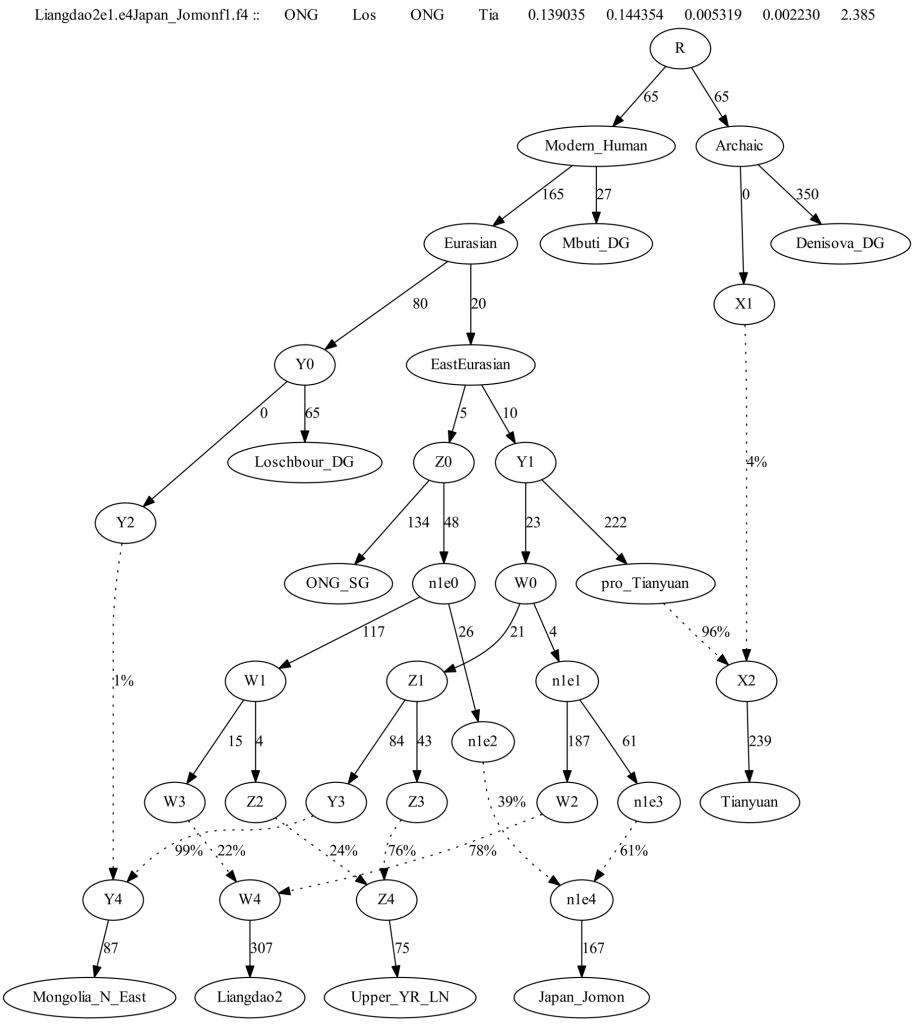


Figure SI3-9: The fitting tree Japan_Jomonf1.f4 with the lowest worst Z-score and likelihood score and without internal zero-length branches of adding Japan_Jomon with one additional admixture was an adequate fit as in the sense that there are no *f*-statistics more than |Z|>3 different between model and expectation. The likelihood score is 17.400 and the worst Z-score of *f*-statistic (ONG, Loschbour; ONG,Tianyuan) =2.385. The number of SNPs used is 545689.

We added Chokhopani onto all the possible edges of the graph Japan_Jomonf1.f4 in Figure SI3-9. We were not able to fit Chokhopani into the graph without invoking an additional admixture event. We obtained 34 fitted graphs for modeling Chokhopani sample as admixed as shown below in Table SI3-4. We showed the two models consistent with MSMC results and also with the lowest likelihood and Z-scores and without internal zero-length branches in Figure SI3-10.

Table SI3-4. The fitted models for modeling Chokhopani sample onto all the possible edges of the model Japan_Jomonf1.f4 in Figure SI3-9.

Model	likelihood score	worst Z-score	Note
Chokhopanie6.g0	20.585	-2.467	
Chokhopanie2.e6	20.949	-2.419	
Chokhopanie0.e6	20.971	-2.478	internal zero-length branch
Chokhopanic0.e6	22.238	-2.893	internal zero-length branch
Chokhopanid1.e6	22.240	-2.893	internal zero-length branch
Chokhopanie5.g2	23.642	-2.523	
Chokhopanie3.g2	24.033	-2.553	
Chokhopanie6.g2	24.128	-2.510	internal zero-length branch
Chokhopanie6.g4	24.128	-2.510	internal zero-length branch
Chokhopanic4.e6	24.341	-2.925	
Chokhopanif2.g2	24.654	-2.439	internal zero-length branch
Chokhopanid3.e6	24.903	-2.905	internal zero-length branch
Chokhopanie6.f0	24.911	-2.906	internal zero-length branch
Chokhopanid6.g2	25.624	-2.611	
Chokhopanie6.g6	28.617	-2.934	
Chokhopanie3.f3	29.145	2.777	
Chokhopanid6.f3	29.575	-2.999	
Chokhopanif0.f3	30.026	3.176	
Chokhopanie3.f5	30.403	2.795	
Chokhopanid6.f5	30.735	-2.923	internal zero-length branch
Chokhopanie5.g0	30.768	-2.775	
Chokhopanie5.g4	31.221	-2.793	internal zero-length branch
Chokhopanid0.e6	31.962	-3.658	
Chokhopanib3.e6	31.970	-3.659	
Chokhopanie5.f5	32.172	-2.849	internal zero-length branch
Chokhopanie5.f3	32.173	-2.849	
Chokhopanif2.f3	32.186	-2.850	not consistent with MSMC
Chokhopanie6.f5	32.213	-2.844	internal zero-length branch
Chokhopanie6.f3	32.214	-2.844	
Chokhopanie5.g6	32.333	-3.301	
Chokhopanie5.f0	32.687	-3.562	
Chokhopanie2.e5	32.707	-2.953	internal zero-length branch
Chokhopanif2.f5	32.830	-2.971	internal zero-length branch
Chokhopanie0.e5	33.101	-2.911	internal zero-length branch

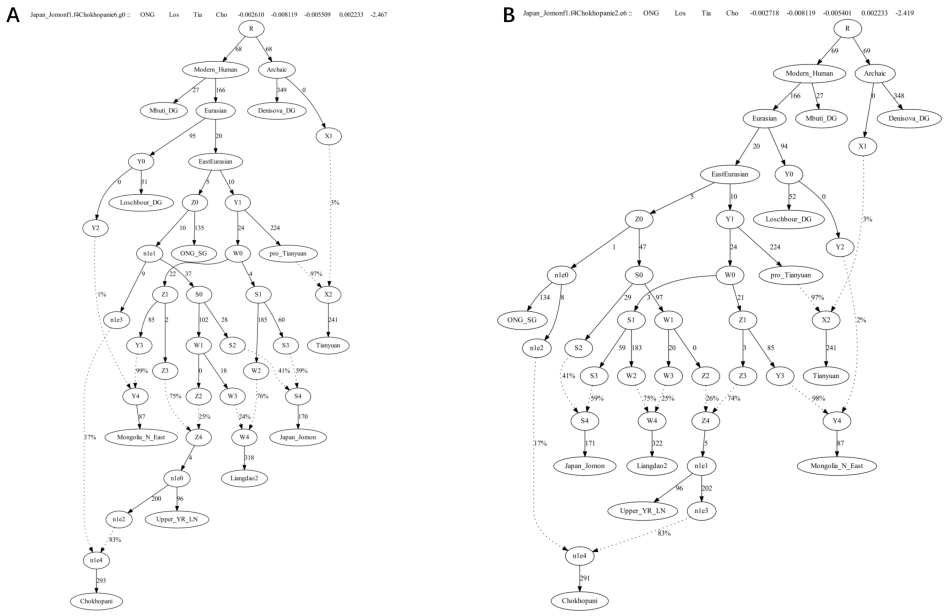


Figure SI3-10: The two fitted models consistent with MSMC results and also with the lowest likelihood scores and Z-scores and without internal zero-length branches. The likelihood score for model A-Chokhopanie6.g0 is 20.585 and the worst Z-score of *f*-statistic (ONG, Loschbour; Tianyuan, Chokhopani) = -2.467. The likelihood score for model B-Chokhopanie2.e6 is 20.949 and the worst Z-score of *f*-statistic (ONG, Loschbour; Tianyuan, Chokhopani) = -2.419. The number of SNPs used is 546547. We note that the drift connecting node Z0 and n1e0 in model B-Chokhopanie2.e6 is only 1, which may indicate the graph topology of node n1e0 is not that reliable.

We added Taiwan_Hanben samples onto all the possible edges of the graph Chokhopanie6.g0 in Figure SI3-10. We were not able to fit Taiwan_Hanben into the graph without invoking an additional admixture event. We obtained 12 fitted graphs for modeling Taiwan_Hanben samples as admixed as shown below in Table SI3-5. We showed the only two models consistent with MSMC results and also with the lowest likelihood and Z-scores and without internal zero-length branches in Figure SI3-11.

Table SI3-5. The fitted models for modeling Taiwan_Hanben samples onto all the possible edges of the model Chokhopanie6.g0 in Figure SI3-10.

Model	likelihood score	worst Z-score	Note
Taiwan_Hanbene5.f6	40.718	2.858	
Taiwan_Hanbend6.f5	41.188	-2.891	internal zero-length branch, not consistent with MSMC
Taiwan_Hanbend6.f3	41.51	-2.942	internal zero-length branch, not consistent with MSMC
Taiwan_Hanbenf2.f5	37.034	-2.949	internal zero-length branch
Taiwan_Hanbene5.f5	37.951	-2.966	internal zero-length branch, not consistent with MSMC
Taiwan_Hanbenf5.h0	37.978	-2.969	internal zero-length branch, not consistent with MSMC
Taiwan_Hanbenf2.g2	37.837	-2.975	internal zero-length branch
Taiwan_Hanbene3.f5	37.762	-2.978	internal zero-length branch, not consistent with MSMC
Taiwan_Hanbenf3.g3	31.431	-2.991	
Taiwan_Hanbeng2.h0	37.829	-2.994	internal zero-length branch, not consistent with MSMC
Taiwan_Hanbenf2.f6	50.206	-2.995	internal zero-length branch
Taiwan_Hanbene3.f6	50.147	-2.996	internal zero-length branch

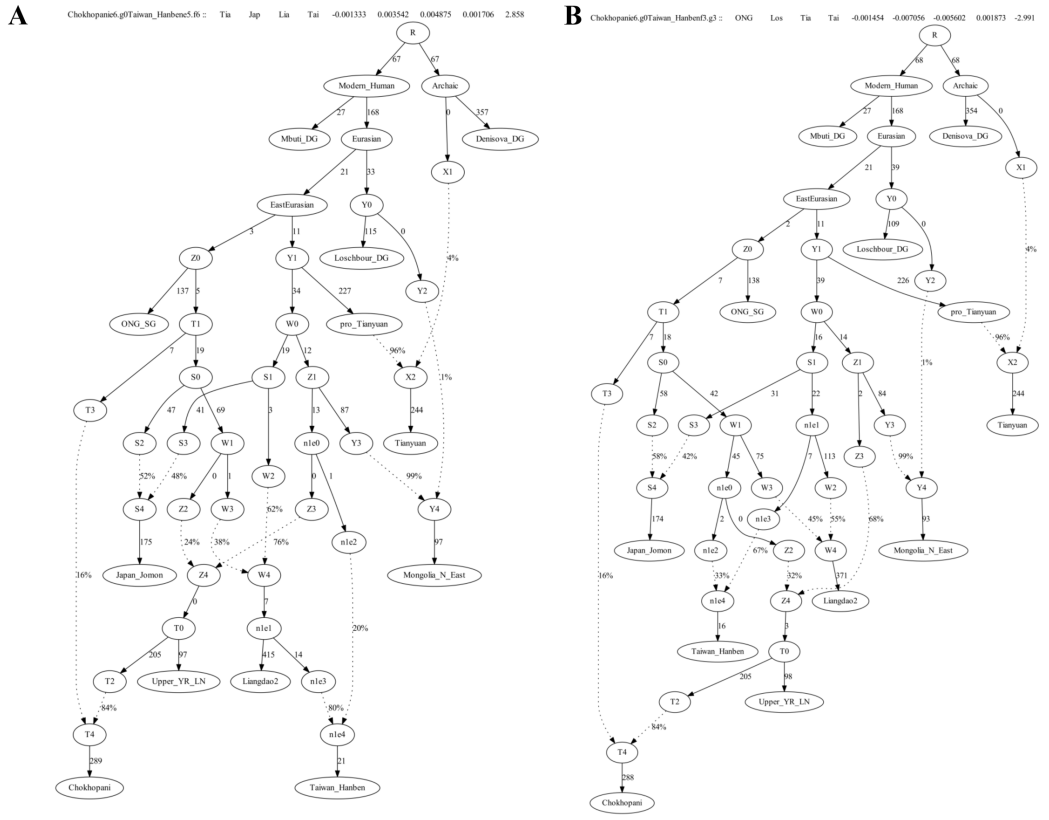


Figure SI3-11: The two fitted models consistent with MSMC results and also with the lowest likelihood scores and Z-scores and without internal zero-length branches. The likelihood score for model A-Taiwan_Hanbene5.f6 is 40.718 and the worst Z-score of *f*-statistic (Tianyuan, Japan_Jomon; Liangdao2, Taiwan_Hanben) = 2.858. The likelihood score for model B-Taiwan_Hanbenf3.g3 is 31.431 and the worst Z-score of *f*-statistic (ONG, Loschbour; Tianyuan, Taiwan_Hanben) =-2.991. The number of SNPs used is 562041.

We then added present-day Han Chinese (whole genome data from the Simons Genome Diversity Project) onto all possible edges of model Taiwan_Hanbene5.f6 and Taiwan_Hanbenf3.g3. We did not obtain any fitted model with a worst |Z|-score less than 3 by adding Han Chinese into the model Taiwan_Hanbenf3.g3. We found four fitted models with worst |Z|-score less than 3 for adding Han Chinese into the model Taiwan_Hanbene5.f6 as shown in Figure SI3-12:

Model A-Han.DGh4.i5 suggests Han Chinese derive 73% of ancestry from Upper_YR_LN related group and the remining 27% from a lineage related to Liangdao2 that also contributed substantailly to Taiwan_Hanben;

Model B-Han.DGh2.i5 suggests Han Chinese derive 67% of ancestry from Upper_YR_LN related group and the remining 33% from a lineage related to Liangdao2 that also contributed largely to Taiwan_Hanben;

Model C-Han.DGi1.i4 suggests Han Chinese derive 53% of the ancestry from Mongolia_East_N related group that also contributed largely to Upper_YR_LN and the remining 47% from a lineage related to Liangdao2 that also contributed largely to Taiwan_Hanben;

Model D-Han.DGh6.i6 suggests Han Chinese derive 62% of the ancestry from a Chokhopani related group and the remining 38% from a lineage related to Taiwan_Hanben.

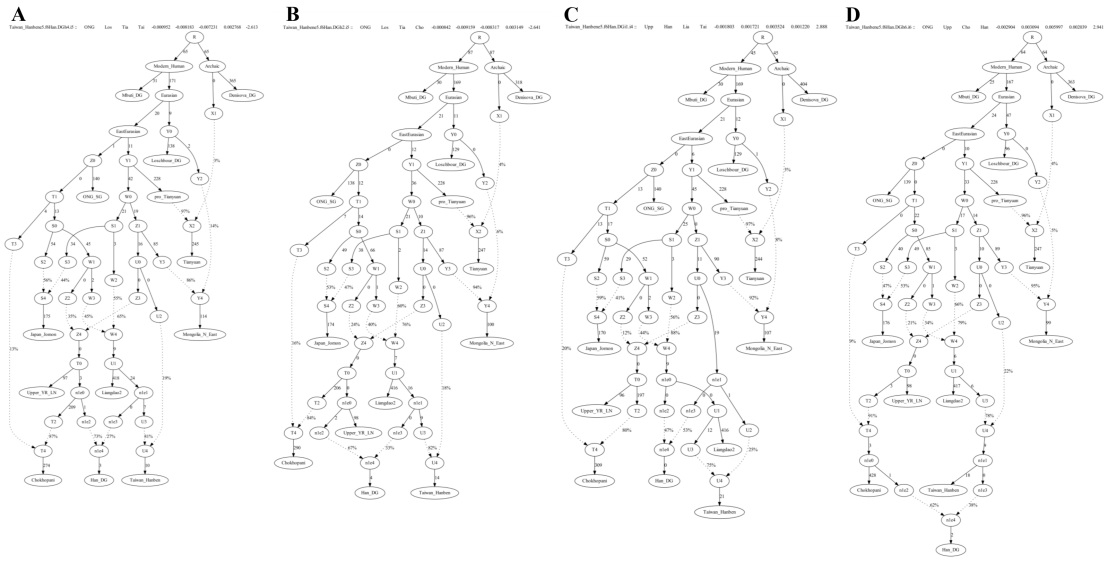


Figure SI3-12: The four fitted models with worst |Z|-score less than 3 by adding Han Chinese onto all the possible edges of model Taiwan_Hanbene5.f6. The likelihood score for model A-Han.DGh4.i5 is 31.64 and the worst Z-score is -2.613. The likelihood score for model B-Han.DGh2.i5 is 36.8 and the worst Z-score is -2.641. The likelihood score for model C-Han.DGi1.i4 is 55.977 and the worst Z-score is 2.888. The likelihood score for model D-Han.DGi1.i4 is 29.749 and the worst Z-score is 2.941. The number of SNPs used is 563480.

We added WLR_LN samples (Late Neolithic samples of West Liao River) onto all the possible edges of the graph Taiwan_Hanbene5.f6. We were not able to fit WLR_LN into the graph without invoking an additional admixture event. We obtained 2 fitted graphs with worst |Z|-scores less than

3 for modeling WLR_LN samples as admixed as shown below in Figure SI3-13. The model A-WLR_LNi1.i4 suggests WLR_LN derive 67% ancestry from a Mongolia_East_N related lineage that also contributes 90% of ancestry to Upper_YR_LN and 25% of ancestry to Taiwan_Hanben and the remining 33% from a sister clade of Liangdao2. The model B-WLR_LNg3.i4 suggests WLR_LN derive the majority of ancestry (66%) from a Mongolia_East_N related lineage that also contributes 95% of ancestry to Upper_YR_LN and 27% of ancestry to Taiwan_Hanben and the remining 34% from a lineage that also contributes largely (80%) to Liangdao2. The two models both suggest Neolithic farmers in West Liao River were closely related to Mongolia_East_N and Upper Yellow River farmers but also had relatedness to populations in southern China.

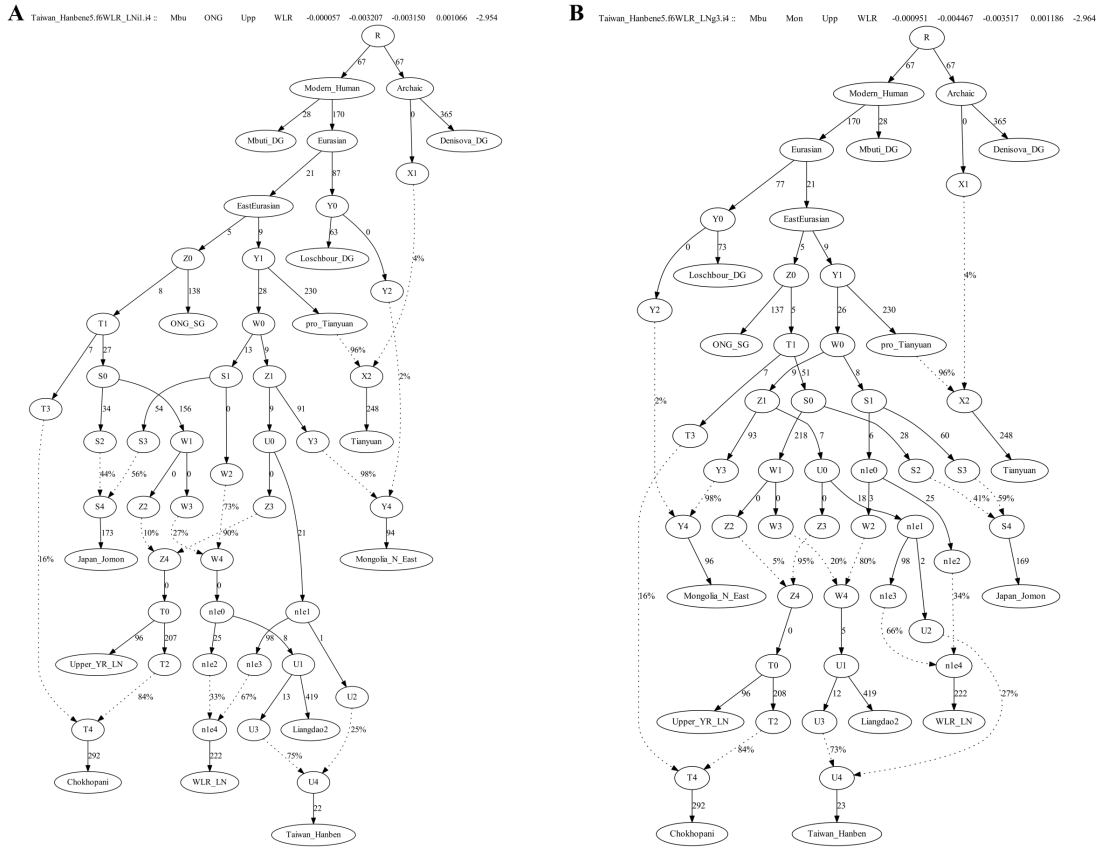


Figure SI3-13: The two fitted models with worst |Z|-score less than 3 by adding WLR_LN onto all the possible edges of model Taiwan_Hanbene5.f6. The likelihood score for model A-WLR_LNi1.i4 is 60.304 and the worst Z-score of f_4 (Mbuti, ONG.SG; Upper_YR_LN, WLR_LN) is -2.954. The likelihood score for model B-WLR_LNg3.i4 is 68.184 and the worst Z-score of f_4 (Mbuti, Mongolia_East_N; Upper_YR_LN, WLR_LN) is -2.964. The number of SNPs used is 499041.

Due to the paucity of ancient genomic data from Upper Paleolithic East Asians, there are limited constraints at present on the deep branching patterns of East Asian ancestral populations, and it is certain that this admixture graph is an oversimplification and that additional features of deep population relationships will be revealed through future work. We propose this graph modeling as a null hypothesis for further testing that can help to advance future work when more ancient DNA from Pleistocene East Asians become available.

References

1. Reich, D., Thangaraj, K., Patterson, N., Price, A.L. Singh, L., Reconstructing Indian population history. ***Nature*** 461, 489-494 (2009)
2. Patterson, N. *et al*. Ancient admixture in human history. ***Genetics*** 192, 1065-1093 (2012)
3. Schiffels, S., Durbin, R., Inferring human population size and separation history from multiple genome sequences. ***Nat Genet***. 46, 919-925 (2014)
4. Wang, K., Mathieson, I., O'Connell, J., & Schiffels, S. (2020). Tracking human population structure through time from whole genome sequences. ***PLoS Genetics***, 16(3), e1008552.
5. Mallick, S.M. *et al*. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations, ***Nature*** 538, 201-206 (2016)

Supplementary Information section 4: Y chromosomal haplogroup assignment

We performed Y-haplogroup determination by examining the state of SNPs present in ISOGG (https://isogg.org/tree/index.html) version 15.56. The different haplogroups were accessed by the following dates:
Last revision date for Haplogroup C and its Subclades: 30 December 2019;
Last revision date for Haplogroup D and its Subclades: 12 December 2019;
Last revision date for Haplogroup N and its Subclades: 1 October 2019;
Last revision date for Haplogroup O and its Subclades: 1 October 2019;
Last revision date for Haplogroup J and its Subclades: 1 October 2019;
Last revision date for Haplogroup Q and its Subclades: 1 October 2019;
Last revision date for Haplogroup R and its Subclades: 28 October 2019.

Boisman sample I1192 could be assigned as C2a1 based on the mutation C2a1-F3914: 14243251C→A. This sample also had upstream derived mutations for haplogroup C2a-F2661:17980274A→G, C2a-F3977: 18075043T→C, C2a-F4012: 21641985G→C. This sample showed ancestral alleles at C2a1a1b1-F1756: 14310844C and C2a1a3-F3791: 17853198C.

Boisman sample I1193 could be assigned as C2a1a-F1788: 14489203C→T, C2a1a-Z22209: 16376421C→G. This sample also had multiple upstream derived mutations for haplogroup C2a-F1396: 8746611C→T, C2a-F3776: 16131630G→T, C2a-F3972: 17800918G→A, C2a-F2661:17980274A→G, C2a-F3981: 18594398A→G, C2a-F3805: 21230995C→T. This sample however showed inconsistent mutations at haplogroup C2a1a3 with a derived allele at C2a1a3-F3795: 18552719C→T, but ancestral alleles at C2a1a3-F966: 7425541T, C1a1a3-F1918: 15250230G. This sample also showed ancestral allele at C2a1a1b1-F1756: 14310844C.

Boisman_MN sample I3355 could be assigned as C2a1a-F3927: 15268404C→T, C2a1a-Z22209: 16376421C→G. This sample also showed multiple upstream derived mutations for haplogroup C2a: F3851: 7431177G→C, F1396: 8746611C→T, F3904: 9521120G→T, F3914: 14243251C→A, F3923: 14774854T→C, F1906: 15062639C→T, F3972: 17800918G→A, F2661: 17980274A→G, F3977: 18075043T→C, F3980: 18413033A→C, F3805: 21230995C→T, F4012: 21641985G→C. This sample showed ancestral alleles for haplogroup C2a1a1b1-F3844: 6974428A, F3896: 9102002C, F1756: 14310844C, and F3955: 17170690T, and also for haplogroup C2a1a3-F966: 7425541T, F4141: 9445166T, F1918: 15250230G, F3939: 16243337G, F3791: 17853198C, F4002: 21131856A.

Boisman_MN sample I3356 could be assigned as C2a1a-F1788: 14489203C→T, but we caution this might be caused by ancient DNA damage. This sample also showed multiple upstream derived mutations for haplogroup C2a: F3851: 7431177G→C, F1396: 8746611C→T, F3904: 9521120G→T, F3914: 14243251C→A, F3923: 14774854T→C, F1906: 15062639C→T, F3954: 17056170A→T, F3787:

17663864C→A, F3972: 17800918G→A, F2661: 17980274A→G, F3977: 18075043T→C, F3980: 18413033A→C, F3981: 18594398A→G, F3805: 21230995C→T, F4012: 21641985G→C. However, this sample showed ancestral alleles for haplogroup C2a1a3-F966: 7425541T, F3939: 16243337G, F3791: 17853198C, F4002: 21131856A, C2a1a1b1-F3844: 6974428A, F3896: 9102002C, F3937: 16204430T, F3955: 17170690T, and C2a1a2a- M77: 21759138C, M86: 21905917T.

Boisman_MN sample I14308 could be assigned as haplogroup C2a-F3776: 16131630G->T, C2a-F3787: 17663864C->A, C2a-F3805: 21230995C->T, C2a-F3923: 14774854T->C, C2a-F3972: 17800918G->A, C2a-F6428: 9065289A->T. This sample also showed derived allele for C2a1a2-F7548: 21978177G->C. However, this sample showed ancestral alleles for haplogroup C2a1a1-F3918: 14533880A->T, C2a1a2a-F6196: 7158634C->T, C2a1a2a-F6297: 7891316A->T, C2a1a2a-F6377: 8525582G->A.

Boisman_MN sample I14771 could be assigned as haplogroup C2a-F1906: 15062639C->T, C2a-Z16720: 2864957G->A, C2a-Z16725: 8818442A->G. However, this sample showed ancestral alleles for haplogroup C2a1a1-Z18161: 8545774A->T, C2a1a2-Y12830: 21053145A->T, C2a1a2a-F6297: 7891316A->T.

Japan_Jomon sample I13883 could be assigned as haplogroup D1a2a3a1-CTS11032: 22908285T->A, Z1533-D1a2a3a1: 7283222A->C, Z1554-D1a2a3a1: 16986596C->T, D1a2a3a1-CTS9335: 18834045G->C, D1a2a3a1-Z14844: 21147701G->A, D1a2a3a1-Z1539: 8828011G->A, D1a2a3a1-Z1564: 22125569A->G. This sample also showed derive alleles for upstream haplogroup D1a2a3a-CTS3097: 14663418C->T, D1a2a3a-CTS6090: 16661643G->C, D1a2a3a-CTS7590: 17547333A->G, D1a2a3a-V2017: 8454980G->T, D1a2a3a-Z14878: 21897321T->C, D1a2a3a-Z3844: 13459848C->T, D1a2a3a-Z14872: 13841180C->T, D1a2a3a-Z14879: 22122403T->C. However, this sample showed ancestral alleles for the downstream haplogroup D1a2a3a1a-CTS621: 6912306C->G, D1a2a3a1a-CTS6303: 16813410C->T, D1a2a3a1a-Z1555: 17367959C->A, D1a2a3a1a-CTS11048: 22918861C->T.

Japan_Jomon sample I13886 could be assigned as haplogroup D1a2a3a-Z1575: 14663418C->T, D1a2a3a-CTS6090: 16661643G->C, D1a2a3a-CTS7590: 17547333A->G, D1a2a3a-CTS8666: 18138115G->A, D1a2a3a-Z14878: 21897321T->C, D1a2a3a-Y12546: 13851916G->A, D1a2a3a-Z1578: 23495419T->A, D1a2a3a-Z3840: 8400626T->G, D1a2a3a-Z3844: 13459848C->T, D1a2a3a-Z3851: 22078567G->A, D1a2a3a-Z14872: 13841180C->T, D1a2a3a-Z14879: 22122403T->C. However, this sample showed ancestral alleles for downstream haplogroup D1a2a3a1-CTS11032: 22908285T->A, D1a2a3a1-CTS321: 2890647C->A, D1a2a3a1-CTS356: 6707139T->G, D1a2a3a1-CTS1228: 7283222A->C, D1a2a3a1-CTS4617: 15736113G->T, D1a2a3a1-Z1554: 16986596C->T, D1a2a3a1-CTS7398: 17459318G->T, D1a2a3a1-CTS9335: 18834045G->C, D1a2a3a1-CTS9902:

19129296C->A, D1a2a3a1-Z14865: 8148133A->G, D1a2a3a1-Z1564: 22125569A->G.

Japan_Jomon sample I13887 could be assigned as haplogroup D1a2a3a-Z1570: 7079046C->A, D1a2a3a-Z1575: 14663418C->T, D1a2a3a-CTS6090: 16661643G->C, D1a2a3a-CTS7590: 17547333A->G, D1a2a3a-CTS8666: 18138115G->A, D1a2a3a-Z14878: 21897321T->C, D1a2a3a-Y12546: 13851916G->A, D1a2a3a-Z1571: 7950712C->T, D1a2a3a-Z3840: 8400626T->G, D1a2a3a-Z3844: 13459848C->T, D1a2a3a-Z3851: 22078567G->A. However, this sample showed ancestral alleles for downstream haplogroup D1a2a3a1-CTS11032: 22908285T->A, D1a2a3a1-CTS321: 2890647C->A, D1a2a3a1-CTS2078: 14173417A->C, D1a2a3a1-CTS6572: 16986596C->T, D1a2a3a1-CTS9335: 18834045G->C, D1a2a3a1-Y12548: 21147701G->A, D1a2a3a1-Z1564 : 22125569A->G, D1a2a3a1-CTS1670: 14013690G->A, D1a2a3a1a-CTS321: 2890647C->A.

Japan_Jomon sample I6341 showed a derived allele for haplogroup D-F1344: 8604667G->A, but we caution this might not be correct because of the low coverage on Y chromosome.

Yankovsky_IA sample I1202 could be assigned as N1a-F1206:8440417C→T, N1a-F3312:22212735A->G. This sample showed ancestral alleles at N1a1-L549: 14636086G and N1a2-F1008: 7570816G.

Heishui_Mohe sample I1209 could be assigned as O2-F633: 22931057T→C. This sample also had multiple upstream derived mutations for haplogroup O-M1740: 7257494A→T, O-P188: 23634362G→A, O-F668: 23975319C→A.

Taiwan_Hanben sample I3612 could be assigned as haplogroup O2a2b2a2-F1903: 15054271A->T, F1645: 9904495G->C. This sample also showed multiple derived upstream mutations for haplogroup O2a2b2-CTS1366: 7363966G→A, F996: 7558650C→T, F1598: 9763046C→T, F1645: 9904495G→C, F1672: 14020486T→G, F1693: 14094513G→T, F2029: 15899637C→A, F2139: 16348471C→T, F3223: 21644600T→C, F3237: 21711624G→A and haplogroup O2a2b2a-F871: 6932191G->C.

We could not be able to determine the haplogroup for Taiwan_Hanben sample I3613 because of the low coverage on Y chromosome.

Taiwan_Hanben sample I3614 could be assigned as haplogroup O2a2b2-CTS1366: 7363966G→A, F996: 7558650C→T, F1481: 9102036T→C, F1598: 9763046C→T, F1672: 14020486T→G, F1693: 14094513G→T, F2029: 15899637C→A, F2139: 16348471C→T, F2469: 17401255G→A, F2683: 18036943T→C, F3223: 21644600T→C, F3237: 21711624G→A. However, this sample showed ancestral alleles for the haplogroup O2a2b2a-F871: 6932191G and O2a2b2a2-F706: 2659661G,

O2a2b2a2-F1645: 9904495G.

Taiwan_Hanben sample I3618 could be assigned as haplogroup O2a2b2a2-F706: 2659661G→A, F1645: 9904495G→C. This sample also showed multiple derived upstream mutations for haplogroup O2a2b2-CTS1366: 7363966G→A, F996: 7558650C→T, F1481: 9102036T→C, F1598: 9763046C→T, F1672: 14020486T→G, F1693: 14094513G→T, F2029: 15899637C→A, F2083: 16201548G→A, F2139: 16348471C→T, F2469: 17401255G→A, F3223: 21644600T→C, F3237: 21711624G→A and O2a2b2a-F871: 6932191G->C.

Taiwan_Hanben sample I3731 could be assigned as haplogroup O1a1a1a1-CTS1711: 14028881G->C, O1a1a1a1-CTS3758: 15155997A->G, O1a1a1a1-CTS4478: 15666311T->A, O1a1a1a1-CTS9282: 18797957T->A, O1a1a1a1-CTS10963: 22858712A->T, O1a1a1a1-Z23467: 13862357C->T, O1a1a1a1-K642: 21694485A->G, O1a1a1a1-Y23446: 15698836A->T. This sample also showed multiple derived upstream mutations for haplogroup O1a-F589: 21622006C->T, O1a1a-L83: 22513920G->A, O1a1a1-F446: 17909248G->A, O1a1a1-CTS4588: 15723358C->T, O1a1a1-F560: 21052297A->G, O1a1a1-Z23389: 21814529C->T. O1a1a1a-F140: 8459608A->C, O1a1a1a-CTS3265: 14780370C->A, O1a1a1a-CTS11270: 23044100G->A, O1a1a1a-F157: 8621202C->T, O1a1a1a-F424: 17596016G->C, O1a1a1a-F518: 18999749G->T, O1a1a1a-F571: 21260029T->G, O1a1a1a-Z23466: 13674402G->T. However, this sample showed ancestral alleles for the haplogroup O1a1a1a1a-Z23469: 20830053C, O1a1a1a1a-Z23465: 13670454G.

Taiwan_Hanben sample I3733 could be assigned as haplogroup O1a1a1a1-CTS10963: 22858712A->T, O1a1a1a1-Z23467: 13862357C->T. This sample also showed multiple derived upstream mutations for haplogroup O1a-F589: 21622006C->T, O1a-L466: 21674068C->T, O1a1a-F31: 6678315T->A, O1a1a-F54: 6969303G->A, O1a1a-F89: 7616941C->T, O1a1a1-F560: 21052297A->G, O1a1a1a-CTS3265: 14780370C->A, O1a1a1a-F157: 8621202C->T, O1a1a1a-F424: 17596016G->C, O1a1a1a-F518: 18999749G->T, O1a1a1a-Z23466: 13674402G->T. However, this sample showed ancestral alleles for the haplogroup O1a1a1a1a-F81: 7545605T, O1a1a1a1a-Z23469: 20830053C, O1a1a1a1a-FGC15393: 13670454G.

Taiwan_Hanben sample I3734 could be assigned as haplogroup O2a2a1a2a2-Y26412: 8240967G->A, O2a2a1a2a2-Y26413: 8454269T->A, O2a2a1a2a2-Y26415: 8632430C->T, O2a2a1a2a2-Y26416: 9823261C->T, O2a2a1a2a2-Y26424: 14804338G->T, O2a2a1a2a2-Y26433: 17649939G->A, O2a2a1a2a2-Y26448: 21820235G->C, O2a2a1a2a2-Y26457: 23370512G->A, O2a2a1a2a2-Y27924: 14093880A->T, O2a2a1a2a2-Y27943: 21964078C->T, O2a2a1a2a2-Y27951: 23535405T->G. This sample also showed multiple derived upstream mutations for haplogroup O2a2a1a2a-CTS10944:22846228G->A, O2a2a1a2a-F2594:17792637G->A, O2a2a1a2a-F3177:21354797T->C, O2a2a1a2a-F3371:22920023G->A, O2a2a1a2a-M209:15575790A->G, O2a2a1a2a-

Z25245:13645205C->A.

Taiwan_Hanben sample I3736 could be assigned as haplogroup O1a1a1a1-CTS1711:14028881G->C, O1a1a1a1-CTS3758:15155997A->G, O1a1a1a1-CTS4478:15666311T->A, O1a1a1a1-CTS10963:22858712A->T, O1a1a1a1-F168:8797778C->T, O1a1a1a1-Z23467:13862357C->T, O1a1a1a1-K642:21694485A->G. This sample also showed multiple derived upstream mutations for haplogroup O1a-M119:21762685T->G, O1a-F589:21622006C->T, O1a-L466:21674068C->T, O1a1a-CTS3422:14911080G->A, O1a1a-CTS6864:17156422G->A, O1a1a-CTS8229:17889560C->G, O1a1a-CTS8934:18561615T->C, O1a1a-CTS9321:18822568G->A, O1a1a-F54:6969303G->A, O1a1a-F89:7616941C->T, O1a1a1-F446:17909248G->A, O1a1a1-CTS4588:15723358C->T, O1a1a1-F560:21052297A->G, O1a1a1-Z23389:21814529C->T, O1a1a1-Z23387:6959884G->A, O1a1a1a-CTS3269:14782408A->G, O1a1a1a-CTS11270:23044100G->A, O1a1a1a-F157:8621202C->T, O1a1a1a-F424:17596016G->C, O1a1a1a-F518:18999749G->T, O1a1a1a-F571:21260029T->G, O1a1a1a-Z23466:13674402G->T. However, this sample showed ancestral alleles for the haplogroup O1a1a1a1a-F81:7545605T->C, O1a1a1a1a-Z23469:20830053C->T.

Taiwan_Hanben sample I8072 could be assigned as haplogroup O1a-Page20: 21622006C->T, O-F45: 6858319C->T, but we note this could be caused by ancient DNA damage. This sample showed ancestral alleles for downstream haplogroup O1a1a1a1a-FGC15393: 13670454G->A, O1a1a1a1a1a1-CTS2498: 14330011A->G.

Taiwan_Hanben sample I8080 could be assigned as haplogroup O1a1a1a-F518: 18999749G->T, O-M1775: 21815957T->G, but we note this assignment may not be correct due to the low coverage of the data and only few observed mutations.

Taiwan_Hanben sample I8081 could be assigned as haplogroup O1a1a1a-Z23466: 13674402G->T, but we note this assignment may not be correct due to the low coverage of the data and only one observed mutation. This sample also showed derived alleles for upstream haplogroup O-CTS1182: 7257494A->T, O-M1741: 7556207G->A, O-F137: 8448515C->T, O-F175: 8909837G->C. This sample showed ancestral alleles for downstream haplogroup O1a1a1a-F157: 8621202C->T, O1a1a1a1-CTS10963: 22858712A->T.

Taiwan_Hanben sample I14933 could be assigned as haplogroup O1a2-F1081: 8018560T->C, O1a2-F2671: 18018434G->A. This sample also showed derived alleles for upstream haplogroup O-CTS3771: 15167624G->A, O-M1739: 6858319C->T, O-M1745: 9818338T->C. However, this sample showed ancestral alleles for downstream haplogroup O1a1a1-F560: 21052297A->G, O1a2a4-Z38625: 6860645A->C.

Taiwan_Hanben sample I14934 could be assigned as haplogroup O1a1a1a1-

CTS10963: 22858712A->T. This sample also showed derived alleles for upstream haplogroup O-P196: 15754313C->A, O1a1a-F54: 6969303G->A, O1a1a1-CTS4588: 15723358C->T, O1a1a1a-CTS11270: 23044100G->A.

Taiwan_Hanben sample I15158 could be assigned as haplogroup O1a2-F1081: 8018560T->C. This sample also showed derived alleles for upstream haplogroup O-CTS1853: 14096806G->A, O-CTS3771: 15167624G->A, O-F315: 15948265A->G, O-F540: 19384549A->T, O-F668: 23975319C->A, O-M1745: 9818338T->C, O-M1772: 21483285G->A, O-P196: 15754313C->A, O1-CTS2866: 14533362C->T. This sample showed ancestral alleles for haplogroup O1a1a1-F560: 21052297A->G, O1a2a3-Y33185: 6931073C->T.

Taiwan_Hanben sample I8078 could be assigned as haplogroup O2a2b2-F3223: 21644600T->C, O2a2b2-F1598: 9763046C->T. This sample also showed derived alleles for upstream haplogroup O2a2-F525: 19148128T->C, O-CTS7498: 17510288A->G, O-F137: 8448515C->T, O-M1773: 21604901T->C. This sample showed ancestral alleles for downstream haplogroup O2a2b2a1-F4286: 19348742A->G, O2a2b2a1b-Z25763: 22186257T->C.

Taiwan_Hanben sample I14929 could be assigned as haplogroup O2a2b-F130: 8395781T->C, O2a2b-F131: 8398931G->A. This sample also showed derived allele for upstream haplogroup O-M1780: 22880173A->G.

Taiwan_Hanben sample I14931 could be assigned as haplogroup O2a2-P201: 2828196T->C, O-L1361: 21385679A->G.

Mongolia_Chalcolithic_1_Afanasievo sample I6221 could be assigned as haplogroup J1a2-CTS11731: 23241285T->A. This sample also showed multiple derived upstream mutations for haplogroup J1-CTS426:6745512C->T, J1-CTS437:6751309A->G, J1-CTS1138:7230829A->G, J1-CTS1860:14100585T->G, J1-CTS1983:14141660C->T, J1-CTS2649:14406344T->C, J1-CTS3210:14732916G->T, J1-CTS3492:14970593G->C, J1-CTS3967:15298128A->G, J1-CTS4025:15340991C->T, J1-CTS4294:15553553C->T, J1-CTS5394:16246522A->C, J1-CTS7412:17467897G->A, J1-CTS7598:17553375G->C, J1-CTS8183:17865091T->A, J1-CTS10540:19528320G->A, J1-CTS10664:22687980C->T, J1-CTS12948:28759094C->G, J1-CTS12950:28759111A->T, J1-F4320:22741818T->G, J1-FGC1617:6644192T->C, J1-FGC1619:13816858G->C, J1-FGC3685:10033663C->T, J1-L765:21407074G->T, J1-PF4643:6873007G->T, J1-PF4653:7848030C->G, J1-PF4655:7941272A->G, J1-PF4660:8142931T->C, J1-PF4662:8406753A->G, J1-PF4665^:8673995G->A, J1-PF4666:8716192G->A, J1-PF4671:9134244T->C, J1-PF4676:9904861C->T, J1-PF4695:14363265G->A, J1-YSC0000172:14810768C->A, J1-PF4757:20835944C->T, J1-PF4759:21210798G->A, J1-PF4762:21264208G->T, J1-PF4769:21682103C->A, J1-PF4774:22086259A->G, J1-PF4780:22628779C->T, J1a-CTS4274:15537842T->C, J1a-

CTS7116:17296977T->A, J1a-PF4644:7079349T->C. However, this sample showed ancestral alleles for the haplogroup J1a2a-CTS1797: 14072309G->A, J1a2a-Z2357: 14213884T->C, J1a2b-CTS1141: 7233156A->G, J1a2b-CTS3569: 15034046G->A, J1a2b-CTS5034: 15972355G->A , J1a2b-CTS7022: 17243839C->G.

Mongolia_Chalcolithic_1_Afanasievo sample I6222 showed multiple derived mutations for haplogroup R1b1a1b-L773: 7220727A->G, R1b1a1b-PF6438: 9464078C->T, R1b1a1b-PF6497: 21222868C->G. This sample also showed a derived allele for R1b1a1b1a1-L52: 14641193C->T, but we caution this might be caused by ancient DNA damage. However, this sample showed ancestral alleles for the downstream haplogroup R1b1a1b1b-Z2103: 7186135G->C, R1b1a1b1a1a1a2-FGC14884: 15462195A->T, R1b1a1b1a1a1b1a1-S186:15668028T->C.

We could not be able to determine the haplogroup for Mongolia_EIA_4_Sagly sample I6223 because of the low coverage on Y chromosome.

Mongolia_EIA_4_Sagly sample I6224 could be assigned as haplogroup R1a1a1b2a2-Z2121: 13821017T->G. This sample also showed multiple derived upstream mutations for haplogroup R1a1a1b2-M746: 18985344C->A. R1a1a1b2a-Z95: 23956870C->T, R1a1a1b-Z647: 7683058G T->A. However, this sample showed ancestral alleles for the haplogroup R1a1a1b2a2b-Z2122: 14477397G->A, R1a1a1b2a2a1-Z2123: 16453077C->T.

Mongolia_EIA_4_Sagly sample I6225 could be assigned as haplogroup R1a1a1-CTS4259: 15525535G->A, R1a1a1-CTS5979: 16591891C->A, R1a1a1-F2957: 19160342A->G, R1a1a1-F3159: 21225537C->A, R1a1a1-F3551: 23800360T->G, R1a1a1-M649: 9407722A->G, R1a1a1-M782: 22003657C->T. This sample also showed multiple derived upstream mutations for haplogroup R1a1a-L168: 16202177A->G , R1a1a-M515: 14054623T->A, R1a1a-CTS3551: 15006352G->A, R1a1a-CTS11720: 23236856A->G, R1a1a-F989: 7545969C->A, R1a1a-F3337: 22676818G->A, R1a1a-M779: 21507919C->T.

Mongolia_EIA_7_Xiongnu sample I6228 could be assigned as haplogroup C2a1a1b1b-Y11605: 8886075C->T, C2a1a1b1b-Z30408: 9064732C->A. This sample also showed multiple derived upstream mutations for haplogroup C2a1a1b1-F3844: 6974428A->G, F3896: 9102002C->T, F1756: 14310844C->T, F3937: 16204430T->C, F3955: 17170690T->G. This sample showed ancestral alleles for the haplogroup C2a1a1b1a-F3830: 2819402T->C, C2a1a1b1a-FGC28884: 14354479C->T, C2a1a1b1b2-FT156697: 8454119G->A, C2a1a1b1b2-FT157670: 18020799C->T, C2a1a1b1b2-FT158298: 23354963C->G.

Mongolia_EIA_4_Sagly sample I6230 could be assigned as haplogroup Q1b1a3a1-L332: 16198895C->G , Q1b1a3a1-YP762: 8593849A->T, Q1b1a3a1-YP776: 17692899A->G, Q1b1a3a1-YP778: 19368913C->T, Q1b1a3a1-YP783:

22807954C->T. This sample also showed multiple derived upstream mutations for haplogroup Q1b1a-L54: 23292782G->A, Q1b1a3-Y5244: 18947607C->A, Q1b1a3-Y5263: 16377904C->T, Q1b1a3-Y11784: 23244840G->C, Q1b1a3-Z35947: 15761583C->A, Q1b1a3-Z35950: 16708573G->T. However, this sample showed ancestral alleles for the haplogroup Q1b1a3a1a-YP4540 : 15196343C->G , Q1b1a3a1a-YP4548: 21438831G->T, Q1b1a3a1a-YP4550: 21490040C->A.

Mongolia_EIA_4_Sagly sample I6231 could be assigned as haplogroup Q1b1a3a1-L332: 16198895C->G, Q1b1a3a1-YP772: 17053829G->C, Q1b1a3a1-YP776: 17692899A->G. This sample also showed multiple derived upstream mutations for haplogroup Q1b1a3-Y5235: 8548403C->A, Q1b1a3-Y5263: 16377904C->T, Q1b1a3-Y11784: 23244840G->C. However, this sample showed ancestral alleles for the haplogroup Q1b1a3a1a-L329: 21842270A->G.

Mongolia_EIA_4_Sagly sample I6232 could be assigned as haplogroup Q1b1a3a1-L332: 16198895C->G, Q1b1a3a1-YP762: 8593849A->T. This sample also showed multiple derived upstream mutations for haplogroup Q1b1a3-L334: 14482079G->A, Q1b1a3-Y5235: 8548403C->A, Q1b1a3-Y5244: 18947607C->A, Q1b1a3-Y5265: 17507994C->G, Q1b1a3-Y5273: 22630254G->A, Q1b1a3-Z35944: 14160938A->G, Q1b1a3-Z35947: 15761583C->A. However, this sample showed ancestral alleles for the haplogroup Q1b1a3a1a-L329: 21842270A->G, Q1b1a3a1a-Y22837: 22199505T->C, Q1b1a3a1a-YP4550: 21490040C->A.

Mongolia_EIA_4_Sagly sample I6233 could be assigned as haplogroup R1a1a1b2a2-Z2121: 13821017T->G. This sample also showed derived upstream mutations for haplogroup R1a1a1b2-Z93: 7552356G->A, R1a1a1b2-M746: 18985344C->A. However, this sample showed ancestral alleles for the haplogroup R1a1a1b2a2a1-Z2123: 16453077C->T

Mongolia_MBA_1_Munkhkhairkhan I6348 could be assigned as haplogroup Q1b1a3-L334: 14482079G->A, Q1b1a3-Y5235: 8548403C->A, Q1b1a3-Y5237: 14045306G->A, Q1b1a3-Y5244: 18947607C->A, Q1b1a3-Y5263: 16377904C->T, Q1b1a3-Y5265: 17507994C->G. However, this sample showed ancestral alleles for the haplogroup Q1b1a3a1a1-YP4549: 21462142G->A, Q1b1a3a1a-YP4566: 9798961A->C, Q1b1a3a1a-YP4550: 21490040C->A, Q1b1a3a1-YP772: 17053829G->C, Q1b1a3a1-YP773: 17112655C->T, Q1b1a3a1-YP776: 17692899A->G, Q1b1a3a1-YP778: 19368913C->T, Q1b1a3a1-YP783: 22807954C->T, Q1b1a3a1-YP762: 8593849A->T.

Mongolia_EIA_1_SlabGrave sample I6349 could be assigned as haplogroup Q1a1a-M265:15030650C->A, Q1a1a-F745: 2886502: T->C, Q1a1a-F750: 2889760C->T, Q1a1a-F1340: 8601548G->A. However, this sample showed ancestral alleles for the haplogroup Q1a1a1-Y535: 7355354 G->A, Q1a1a1-F4532: 2657214G->C,

Mongolia_EIA_1_SlabGrave sample I6352 could be assigned as haplogroup Q1a1a-F745: 2886502T->C, Q1a1a-F1111: 8129592T->C, Q1a1a-F1340: 8601548G->A. However, this sample showed ancestral alleles for the haplogroup Q1a1a1-F1626: 9857502G->A.

Mongolia_EIA_1_SlabGrave sample I6353 could be assigned as haplogroup Q1a1a-F1340: 8601548G->A, Q1a1a-F2043: 15941917C->T, Q1a1a-F2094: 16211687C->T. However, this sample showed ancestral alleles for the haplogroup Q1a1a1-Y535: 7355354G->A, Q1a1a1-F4935: 14228420T->G, Q1a1a1-F4532: 2657214G->C.

Mongolia_EIA_4_Sagly sample I6356 could be assigned as haplogroup Q1b1a3b-SK1944: 8550111G->A, Q1b1a3b-SK1941: 21489957G->T, Q1b1a3b-Y11948: 8318892T->A, Q1b1a3b-Y12451: 19107462G->A. This sample also showed derived upstream mutations for haplogroup Q1b1a3-SK1943: 8548403C->A, Q1b1a3-Y5239: 14849480A->G, Q1b1a3-Y5244: 18947607C->A, Q1b1a3-Y5245: 18994575A->T, Q1b1a3-Y5263: 16377904C->T, Q1b1a3-Y5265: 17507994C->G, Q1b1a3-Z35947: 15761583C->A. However, this sample showed ancestral alleles for the haplogroup Q1b1a3a1-BZ988: 18188735C->T, Q1b1a3a1-L332: 16198895C->G, Q1b1a3a1-YP762: 8593849A->T, Q1b1a3a1-YP769: 15970517G->A, Q1b1a3a1-YP776: 17692899A->G, Q1b1a3b1a-B30: 15900681C->T, Q1b1a3b1a-BZ110: 16743775T->G, Q1b1a3b2-Z35983: 2669415C->A, Q1b1a3b2-Z35986: 7321431C->T, Q1b1a3b3-Y12452: 23271973G->A, Q1b1a3b3-Y12676: 8424843A->G, Q1b1a3b4-B31: 2888380C->G.

Mongolia_EIA_1_SlabGrave sample I6359 could be assigned as haplogroup Q1b1a3a-YP779: 21463943C->T, but we caution this might be caused by ancient DNA damage. This sample also showed derived upstream mutations for haplogroup Q1b1a3-L334: 14482079G->A, Q1b1a3-Y5235: 8548403C->A, Q1b1a3-Y5241: 15914269C->G, Q1b1a3-Y5259: 8184383C->T, Q1b1a3-Y5265: 17507994C->G, Q1b1a3-Y5273: 22630254G->A, Q1b1a3-Z35944: 14160938A->G, Q1b1a3-Z35957: 21459264C->G. However, this sample showed ancestral alleles for the haplogroup Q1b1a3a1-BZ433: 15739867T->C, Q1b1a3a1-L332: 16198895C->G, Q1b1a3a1-Y18330: 18966322C->G, Q1b1a3a1-YP762: 8593849A->T, Q1b1a3a1-YP776: 17692899A->G.

Mongolia_LBA_3_MongunTaiga sample I6363 could be assigned as haplogroup R1a1a1b2a-Z95: 23956870C->T, but we caution this might be caused by ancient DNA damage. This sample also showed derived upstream mutations for haplogroup R1a1a1b-Z645: 8245045C->T, R1a1a1b-Z647: 7683058G->A, R1a1a1b-Z649: 16329760C->G, R1a1a1b-M750:19058383C->A, R1a1a1b-F3044: 19419865G->A, R1a1a1b2-M746: 18985344C->A.

Mongolia_EIA_1_SlabGrave sample I6365 could be assigned as haplogroup N1a1a1a1a-M1999: 6912974C->T, N1a1a1a1a-M2001: 7185755T->C, N1a1a1a1a-CTS3103: 14666345C->G, N1a1a1a1a-M2034: 14780417A->G, N1a1a1a1a-

CTS4260: 15525750G->A. This sample also showed derived upstream mutations for haplogroup N1a -M2013: 8440417C->T, N1a-M2042: 16329998G->A, N1a -F3361: 22859747T->C, N1a-M2099: 21568587C->T, N1a1-L395: 19048725G->A, N1a1-M2033: 14636086G->A, N1a1a-M178: 21741755C->T, N1a1a-Z1953: 14636086G->A, N1a1a1a1-M2024: 14015310C->T, N1a1a1a1-M2005: 7618972T->C. However, this sample showed ancestral alleles for the haplogroup N1a1a1a1a1a-CTS2929: 14570424T->C.

Mongolia_LBA_5_CenterWest sample I6367 could be assigned as haplogroup O2a2b1a1a1-CTS5866: 16518244C->T, O2a2b1a1a1-F42: 6840710G->A, O2a2b1a1a1-F348: 16521399C->T. This sample also showed derived upstream mutations for haplogroup O2a2b1a1-Page23: 15999244G->A, O2a2b1a1-CTS1275: 7307053A->G, O2a2b1a1-CTS5128: 16028396T->C, O2a2b1a1-CTS11742: 23246261G->A, O2a2b1a1-F141: 8461538G->T, O2a2b1a1-F235: 14077529G->A, O2a2b1a1-F342: 16472742A->G, O2a2b1a1-F476: 18209413G->C, O2a2b1a1-F579: 21353681T->C, O2a2b1a1-F581: 21359854C->T, O2a2b1a1-F613: 22558796C->T, O2a2b1a1-F649: 23578463T->C, O2a2b1a1a-CTS2338: 14266667A->G, O2a2b1a1a-F204: 9646820G->A, O2a2b1a1a-M1581: 9116739C->T.

Mongolia_East_N sample I7021 could be assigned as haplogroup C2a1a1-Z18161: 8545774A->T, C2a1a1-F4015: 22523607C->T, C2a1a1-Z18155: 22560811G->A, C2a1a1-Z18160: 19058942T->C. This sample also showed derived allele at C2a1a1b-FGC28881.2: 14218495G->A, but showed ancestral alleles at C2a1a1b2:2801848C->G, C2a1a1a-Z30542: 8233259C->T, C2a1a1a-Z30543: 8271472A->G, C2a1a1a-Z30539: 7989801C->T, C2a1a1a-Z30536: 7601110G->A, C2a1a1b1-F1756: 14310844C->T, C2a1a1b1-F3985: 18760600G->T.

Mongolia_Chalcolithic_2_Afanasievo sample I13957 could be assigned as haplogroup C2a1a1-Z18161: 8545774A->T, C2a1a1-Z18155: 22560811G->A. However, this sample showed ancestral alleles for haplogroup C2a1a1b1-F3896: 9102002C->T, C2a1a1b1-F3844: 6974428A->G, C2a1a1b1-F3843: 6950926G->C, C2a1a1b1-F1756: 14310844C->T, C2a1a1a2b-Z45646: 8023742A->G, C2a1a1a2a-Z38871: 7190078T->G, C2a1a1a2-Z45640: 9874646G->A, C2a1a1a2-Z45638: 22860774C->A, C2a1a1a2-Z45639: 8272495T->C, C2a1a1a2-NCG8: 14096488G->A, C2a1a1a1b-BY21709: 17396229A->C, C2a1a1a1a2b-Z30826: 23970251C->T, C2a1a1a1a2-Z30819: 15726112A->G, C2a1a1a1a2-Z30768: 15712563G->A, C2a1a1a1a1a1-Z30755: 23751092G->A, C2a1a1a1a1a-BY18408: 23073639G->A, C2a1a1a1a1-Z30749: 14148873C->G, C2a1a1a1a-Z30585: 21391225A->G, C2a1a1a1a-Z30563: 16615191T->C, C2a1a1a1-Z30591:23031330C->T, C2a1a1a-Z30587: 22478692A->C, C2a1a1a-Z30593: 23165402G->A, C2a1a1a-Z30544: 8397116C->T, C2a1a1a-Z30550: 9141267G->A, C2a1a1a-Z30590-BY2113: 23011980G->A, C2a1a1a-Z30592: 23043657C->T, C2a1a1a-Z30594: 23264798C->A, C2a1a1a-Z30582: 21168082T->C, C2a1a1a-Z30536: 7601110G->A.

Mongolia_EBA_1_Ulgii sample I12977 could be assigned as haplogroup C2a1a1-Z18161: 8545774A->T, C2a1a1-Z18155: 22560811G->A, C2a1a1-Z18160: 19058942T->C. However, this sample showed ancestral alleles for the haplogroup C2a1a1a-Z30536: 7601110G->A, C2a1a1a-Z30533: 2838315A->G, C2a1a1a-Z30539: 7989801C->T, C2a1a1a-Z30542: 8233259C->T, C2a1a1a-BY783: 8271472A->G, C2a1a1a-Z30545: 8418103T->C, C2a1a1a-Z30549: 8891583C->T, C2a1a1b1-F1756: 14310844C->T, C2a1a1b1-F3843: 6950926G->C, C2a1a1b1-F3844: 6974428A->G.

Mongolia_LBA_4_CenterWest sample I12975 could be assigned as haplogroup C2a1a-F1699: 14137030C->T, C2a1a-F1788: 14489203C->T, C2a1a-F3927: 15268404C->T. However, this sample showed ancestral alleles for the haplogroup C2a1a1-Z18161: 8545774A->T, C2a1a1-Z18155: 22560811G->A, C2a1a1-Z18160: 19058942T->C, C2a1a2-F6370: 8480384G->A, C2a1a2-F7225: 17772838A->G, C2a1a2-F7423: 21116585C->T, C2a1a2-F7548: 21978177G->C, C2a1a3-Z1866: 16362484G->A, C2a1a3-F914: 7153509A->G, C2a1a3-F966: 7425541T->C, C2a1a3-F4141: 9445166T->C, C2a1a4-F10261: 9062800C->T, C2a1a4-F10387: 9900580A->G.

Mongolia_North_N sample I13698 could be assigned as haplogroup C2a1a1-Z18161: 8545774A->T, C2a1a1-Z18155: 22560811G->A, C2a1a1-Z18160: 19058942T->C. However, this sample showed ancestral alleles for the haplogroup C2a1a1a-Z30536: 7601110G->A, C2a1a1a-BY738: 2838315A->G, C2a1a1a-Z30539: 7989801C->T, C2a1a1a-BY778: 8233259C->T, C2a1a1a-Z30543: 8271472A->G, C2a1a1a-Z30545: 8418103T->C, C2a1a1b1-F1756: 14310844C->T, C2a1a1b1-F3843: 6950926G->C, C2a1a1b1-F3844: 6974428A->G.

Mongolia_North_N sample I11696 could be assigned as haplogroup C2a1a1-Z18161: 8545774A->T, C2a1a1-F4015: 22523607C->T, C2a1a1-Z18155: 22560811G->A, C2a1a1-Z18160: 19058942T->C. However, this sample showed ancestral alleles for the haplogroup C2a1a1b1-F3937: 16204430T->C, C2a1a1b1-F3896: 9102002C->T, C2a1a1b1-F3844: 6974428A->G, C2a1a1b1-F3843: 6950926G->C, C2a1a1b1-F1756: 14310844C->T, C2a1a1a -Z30536: 7601110G->A, C2a1a1a-Z30539: 7989801C->T, C2a1a1a-BY778: 8233259C->T, C2a1a1a-Z30543: 8271472A->G, C2a1a1a-Z30545: 8418103T->C.

Mongolia_North_N sample I11698 could be assigned as haplogroup C2a1a1-Z18161: 8545774A->T, C2a1a1-F4015: 22523607C->T, C2a1a1-Z18155: 22560811G->A, C2a1a1-Z18160: 19058942T->C. However, this sample showed ancestral alleles for the haplogroup C2a1a1a-BY728: 2702701A->C, C2a1a1a-Z30533: 2838315A->G, C2a1a1a-Z30539: 7989801C->T, C2a1a1a-BY778: 8233259C->T, C2a1a1a-Z30543: 8271472A->G, C2a1a1a-Z30545: 8418103T->C, C2a1a1b1-F1756: 14310844C->T, C2a1a1b1-F3843: 6950926G->C, C2a1a1b1-F3844: 6974428A->G, C2a1a1b1-F3868: 8504007T->C.

Mongolia_North_N sample I11697 could be assigned as haplogroup C2a1a1-Z18161:

8545774A->T, C2a1a1-Z18155: 22560811G->A, C2a1a1-Z18160: 19058942T->C. However, this sample showed ancestral alleles for the haplogroup C2a1a1b1-F3843: 6950926G->C, C2a1a1b1-F3844: 6974428A->G, C2a1a1b1-F3868: 8504007T->C, C2a1a1b1-F3937: 16204430T->C, C2a1a1b1-F3955: 17170690T->G, C2a1a1a-Z30536: 7601110G->A, C2a1a1a-Z30539: 7989801C->T, C2a1a1a-BY778: 8233259C->T, C2a1a1a-Z30543: 8271472A->G, C2a1a1a-Z30545: 8418103T->C, C2a1a1a-Z30546: 8560348G->A.

Mongolia_MBA_1_Munkhkhairkhan sample I12955 could be assigned as haplogroup N1a1a1-CTS7728: 17632471G->C. This sample also showed derived alleles for upstream haplogroup N1a-F1206: 8440417C->T, N1a-CTS4308: 15562216C->T, N1a-F3094: 20827458T->A, N1a-F3312: 22212735A->G, N1a-F3361: 22859747T->C, N1a-M2006: 7793659A->G, N1a-M2099: 21568587C->T, N1a1-M2080: 19048725G->A, N1a1-L549: 14636086G->A, N1a1a-M178: 21741755C->T, N1a1a-L549: 14636086G->A. However, this sample showed ancestral alleles for downstream haplogroup N1a1a1a-L708: 7629512C->A, N1a1a1a1-M2005: 7618972T->C, N1a1a1a1-M2007: 7835228C->T, N1a1a1a1-M2096: 21495976G->A.

Mongolia_LBA_3_MongunTaiga sample I12976 could be assigned as haplogroup N1a-M2114: 22999882A->T, N1a-F2130: 16329998G->A, N1a-F3312: 22212735A->G, N1a-F3361: 22859747T->C, N1a-Z4845: 7793659A->G, N1a-M2099: 21568587C->T. This sample showed derived alleles for some of the mutations determining haplogroup N1a2: N1a2-CTS7235: 17369893T->C, N1a2-CTS7713: 17623551G->A, N1a2-F1360: 8624113T->C, N1a2-F4309: 21879123A->G, N1a2-FGC10768: 16525023A->T, N1a2-Y3032: 15777885C->T, N1a2-FGC10789: 7812399A->G, N1a2-Y3071 : 19498148A->T, N1a2-FGC10801: 9459830G->A, N1a2-FGC10811: 17325722A->G, N1a2-CTS4202: 15474940C->T, but also showed ancestral alleles for the following mutations determining N1a2: N1a2-F1008: 7570816G->A, N1a2-CTS10075: 19243734G->A, N1a2-CTS10895: 22823654C->A, N1a2-F864: 6911889C->T, N1a2-F1007: 7570242C->G, N1a2-F2700: 18068416C->T, N1a2-F3163: 21231863G->A, N1a2-F3290: 22160072T->C, N1a2-Y3047: 7190021G->A, N1a2-FGC10795: 23316653C->G, N1a2-Y3079: 15747238C->T. This sample showed ancestral alleles for haplogroup N1a1-Z1953: 14636086G->A, N1a1a-M178: 21741755C->T, N1a1a-L549: 14636086G->A.

Mongolia_MBA_2_Munkhkhairkhan sample I13173 could be assigned as haplogroup N-M231: 15469724G->A, N-CTS568: 6851169C->T, N-Z4838: 7264987C->G, N-CTS1653: 14006122C->T, N-M2169: 14244937C->G, N-CTS2652: 14407021A->C, N-CTS2699: 14434372C->T, N-CTS3734: 15139183A->T, N-M2184: 15309390C->T, N-CTS4082: 15382221A->C, N-CTS4127: 15412847C->T. This sample showed derived alleles for some of the mutations determining haplogroup N2: N2-Y6585: 7374679C->T, N2-FGC28395: 2739436A->G, N2-FGC28404: 7538163G->T, N2-FGC28414: 8121046G->A, N2-FGC28422: 8537238C->G, N2-FGC28427: 8880278G->A, N2-FGC28428: 8979570G->A, N2-FGC28429: 9120740C->T, N2-

FGC28431: 9397949C->T, N2-Y6480: 14332213A->T, N2-FGC28444: 14454883G->C, N2-Y6482: 14708358G->C, N2-FGC28448: 14951485T->C, N2-Y6544: 15145895A->C, N2-Y6547: 15978652T->A, N2-Y6549: 16385414C->T, N2-FGC28466: 17550925C->T, N2-FGC28492: 19503704A->T, but also showed ancestral alleles for the following mutations determining N2: N2-FGC28485: 19091959T->C, N2-FGC28417: 8299333C->G, N2-FGC28419: 8400628C->T, N2-FGC28407: 7689688C->T. This sample showed ancestral alleles for haplogroup N1-CTS10333: 19405037T->G, N1-CTS10907: 22829376G->A, N1-CTS11710: 23233272A->T, N1-M2142: 7415220G->C, N1-F1427: 8840859G->A, N1-F1840: 14792418T->C, N1-S19333: 17192238C->T, N1-F3002: 19274282G->T, N2a-FGC28405: 7644368A->C, N2a-FGC28418: 8349396C->A, N2a-FGC28426: 8840139C->T, N2a-Y6581: 24377466A->G, N2a-P189.2: 14197977G->A.

Mongolia_LBA_4_CenterWest sample I13768 could be assigned as haplogroup N-M231: 15469724G->A, N-V2948: 15437152C->T, N-CTS45: 2686422A->T, N-CTS92: 2724768G->C, N-M2137: 6851169C->T, N-CTS1200: 7264987C->G, N-CTS1653: 14006122C->T. This sample showed derived alleles for some of the mutations determining haplogroup N2: N2-Y6585: 7374679C->T, N2-FGC28390: 19266891T->C, N2-FGC28395: 2739436A->G, N2-FGC28399: 6700680C->T, N2-Y6515: 6795569G->A, N2-FGC28404: 7538163G->T, N2-FGC28408: 7702455A->G, N2-FGC28414: 8121046G->A, N2-FGC28422: 8537238C->G, N2-FGC28427: 8880278G->A, N2-FGC28428: 8979570G->A, N2-FGC28429: 9120740C->T, N2-Y6480: 14332213A->T, N2-FGC28444: 14454883G->C, N2-Y6482: 14708358G->C, N2-Y6544: 15145895A->C, N2-FGC28454: 15802738C->T, N2-FGC28457: 15978652T->A, N2-Y6549: 16385414C->T, N2-FGC28466: 17550925C->T, N2-Y6560: 18948537G->A, N2-Y6561: 19107273C->T, N2-FGC28489: 19271885G->T, N2-FGC28492: 19503704A->T, N2-Y6572: 22192736C->T, N2-FGC28509: 22214467G->A, N2-Y6530: 8086780G->A, N2-Y7890: 6931335C->T, but also showed ancestral alleles for the following mutations determining N2: N2-FGC28393: 7215561G->A, N2-FGC28401: 6932594C->G, N2-Y6524: 7689688C->T, N2-Y6532: 8299333C->G, N2-FGC28419: 8400628C->T, N2-FGC28431: 9397949C->T, N2-FGC28485: 19091959T->C. This sample showed ancestral alleles for haplogroup N2a-Y6522: 7644368A->C, N2a-Y6505: 8349396C->A, N2a-FGC28421: 8444807A->T, N2a-Y6533: 8840139C->T, N2a-FGC28433: 9829879G->A.

Mongolia_EIA_4_Sagly sample I7022 could be assigned as haplogroup Q1a2a-F4793: 8749159C->T, Q1a2a-F4775: 8545119G->A, Q1a2a-F4780: 8604771C->A, Q1a2a-F4765: 8405012C->T, Q1a2a-L712: 6851494A->G. This sample showed derived alleles for upstream haplogroup Q1a2-F4672: 6697302A->G, Q1a2-F4714: 7629052A->G, Q1a2-F4784: 8660645C->A, Q1a2-F4820: 9381034A->T, Q1a2-F4960.2: 14656738G->A, Q1a2-F4996: 15379476T->G, Q1a2-F5133.2: 18084396C->T, Q1a2-F5169: 18907907T->C, Q1a2-F5215: 21180662C->A, Q1a2-F5237: 21447397T->C, Q1a2-F5408: 23747962A->T, Q1a2-F7020: 15502461T->C.

Mongolia_LBA_2_Ulaanzukh sample I12960 could be assigned as haplogroup Q1a1a1-F1626: 9857502G->A, Q1a1a1-Y535: 7355354G->A. This sample also showed derived alleles for upstream haplogroup Q1a-F1251: 8491335T->G, Q1a-Y604: 18058751C->G, Q1a-Y627: 14275583G->A, Q1a-F1755: 14299613C->T, Q1a-F1762: 14325979A->T, Q1a1-F1202: 8438421A->T, Q1a1-F1261: 8500133G->A, Q1a1-F1298: 8547406A->C, Q1a1-F1755: 14299613C->T, Q1a1-F1762: 14325979A->T, Q1a1-Z19171: 14963804A->G, Q1a1a-F745: 2886502T->C, Q1a1a-F1908: 15077751T->C, Q1a1a-Z19174: 15941917C->T, Q1a1a-F2086: 16202425T->C, Q1a1a-SK1919: 16802392A->C, Q1a1a-F3019: 19332304C->G, Q1a1a-SK1917: 7557577C->T, Q1a1a-Z19170: 14879893G->A, Q1a1a-Z19183: 17011560G->T, Q1a1a-Z19200: 22054395G->A, Q1a1a-Z19210: 23404344G->A, Q1a1a-Y613: 7904858C->T, Q1a1a-Y697: 19431584G->C. However, this sample showed ancestral alleles for downstream haplogroup Q1a1a1a1-F4777: 8568360C->T, Q1a1a1a1-F4837: 9974534C->T, Q1a1a1a1a1-Z190: 17473966G->T, Q1a1a1a1a1-Z43857: 8433862A->G, Q1a1a1a1a1-Z43858: 8465728C->T, Q1a1a1b1-PH1003: 14207123C->T, Q1a1a1b1-PH4278: 19150649G->T.

Mongolia_EIA_1_SlabGrave sample I12969 could be assigned as haplogroup Q1a1a-Z19198: 22020659G->C, Q1a1a-F5272: 22031296G->A, Q1a1a-M265: 15030650C->A, Q1a1a-Y683.2: 2865795C->T, Q1a1a-Y697: 19431584G->C, Q1a1a-Y701: 21473854G->T, Q1a1a-Y705: 21709263G->A. This sample also showed derived allele for haplogroup Q1a1a2-M7417: 7001509C->T, but we note this might be caused by ancient DNA damage. This sample showed ancestral alleles for haplogroup Q1a1a1-F1626: 9857502G->A, Q1a1a1-Y525: 2657214G->C, Q1a1a1-Y532: 8912529G->A, Q1a1a1a-Z19217: 19299147C->T.

Mongolia_LBA_2_Ulaanzukh sample I12972 could be assigned as haplogroup Q1a1a-Y683.2: 2865795C->T, Q1a1a-Y705: 21709263G->A, Q1a1a-F1111: 8129592T->C, Q1a1a-F4710: 7557577C->T, Q1a1a-Z19183: 17011560G->T. This sample also showed derived allele for haplogroup Q1a1a1-Y527: 21523616C->T, but we note this might be caused by ancient DNA damage. This sample showed derived alleles for upstream haplogroup Q1a-F1215: 8454150A->C, Q1a-F2786: 18599094C->T, Q1a-Z19223: 17341480C->A, Q1a-F1755: 14299613C->T, Q1a-F2753: 18401339A->T, Q1a1-Z19151: 8432920A->C, Q1a1-F1755: 14299613C->T, Q1a1-F2753: 18401339A->T, Q1a1-Z19168: 14812749G->A, Q1a1-SK1918: 15744594C->T. However, this sample showed ancestral alleles for downstream haplogroup Q1a1a1a1-F4777: 8568360C->T, Q1a1a1a1-Y562: 2740226A->C.

Mongolia_LBA_4_CenterWest sample I13505 could be assigned as haplogroup Q1a1a-F5129: 18024026G->A, Q1a1a-Z19198: 22020659G->C, Q1a1a-F5272: 22031296G->A, Q1a1a-Z19206: 22675027G->A, Q1a1a-Z19210: 23404344G->A, Q1a1a-L415.2: 2888663C->T, Q1a1a-M265: 15030650C->A, Q1a1a-Y613: 7904858C->T. However, this sample showed ancestral alleles for downstream haplogroup Q1a1a1-F1626: 9857502G->A, Q1a1a1-Y525: 2657214G->C, Q1a1a1-

F4935: 14228420T->G, Q1a1a1-Y535: 7355354G->A, Q1a1a1a-Z19217: 19299147C->T, Q1a1a1a-Y558: 21872679C->A, Q1a1a2a-Z35917: 7000521G->A.

Mongolia_LBA_2_Ulaanzukh sample I14037 could be assigned as haplogroup Q1a1a1-F875: 6936716C->T, Q1a1a1-Y535: 7355354G->A. This sample also showed derived alleles for upstream haplogroup Q1a1a-F5270: 22020659G->C, Q1a1a-S435.2: 2888663C->T, Q1a1a-Y610: 7137166C->T, Q1a1a-Y683.2: 2865795C->T. However, this sample showed ancestral alleles for downstream haplogroup Q1a1a1a-F4528: 19299147C->T, Q1a1a1a1-Y541: 8568360C->T, Q1a1a1a1-Y526: 9974534C->T.

Mongolia_EIA_4_Sagly sample I7029 could be assigned as haplogroup Q1b1a3a1-BZ433: 15739867T->C, Q1b1a3a1-L332: 16198895C->G, Q1b1a3a1-YP762: 8593849A->T. This sample also showed derived alleles for upstream haplogroup Q1b1a3-L334: 14482079G->A, Q1b1a3-SK1943: 8548403C->A, Q1b1a3-Y5237: 14045306G->A, Q1b1a3-Y5244: 18947607C->A, Q1b1a3-Y5265: 17507994C->G, Q1b1a3-Y5268: 21562248A->T, Q1b1a3-Y5272: 22575911A->T, Q1b1a3-Z35944: 14160938A->G. However, this sample showed ancestral alleles for downstream haplogroup Q1b1a3a1a-L329 : 21842270A->G , Q1b1a3a1a-Y22837: 22199505T->C, Q1b1a3a1a-YP4540: 15196343C->G, Q1b1a3a1a-YP4547: 19164395C->G, Q1b1a3a1a-YP4548: 21438831G->T, Q1b1a3a1a-YP4550: 21490040C->A, Q1b1a3b-SK1944: 8550111G->A, Q1b1a3b-Y11234: 14030447G->A.

Mongolia_LBA_4_CenterWest sample I7039 could be assigned as haplogroup Q1b1a3-L334: 14482079G->A, Q1b1a3-SK1943: 8548403C->A, Q1b1a3-Y5228: 21334535G->A, Q1b1a3-Y5238: 14176657G->T, Q1b1a3-Y5244: 18947607C->A, Q1b1a3-Y5245: 18994575A->T, Q1b1a3-Y5251: 23030814C->A, Q1b1a3-Y5259: 8184383C->T, Q1b1a3-Y5261: 15913458C->G, Q1b1a3-Y5265: 17507994C->G. This sample also showed derived allele for haplogroup Q1b1a3a-YP779: 21463943C->T, but we caution this might be caused by ancient DNA damage. This sample showed ancestral alleles for downstream haplogroup Q1b1a3b-Y11234: 14030447G->A, Q1b1a3b-Z35933: 7706350C->G, Q1b1a3b-SK1944: 8550111G->A, Q1b1a3b-SK1941: 21489957G->T, Q1b1a3b-SK1942: 15880973A->G, Q1b1a3a1a-YP4566: 9798961A->C, Q1b1a3a1a-YP4567: 22627864C->T, Q1b1a3a1a-YP4548: 21438831G->T, Q1b1a3a1a-YP4550: 21490040C->A, Q1b1a3a1a-YP4551: 21724299A->G, Q1b1a3a1a-YP4552: 21800744G->A, Q1b1a3a1a-Y22837: 22199505T->C, Q1b1a3a1a-YP4540: 15196343C->G, Q1b1a3a1a-YP4541: 15573392A->C, Q1b1a3a1-YP1695: 16418022A->T, Q1b1a3a1a-L329: 21842270A->G, Q1b1a3a1-YP785: 19092789A->G, Q1b1a3a1-YP783: 22807954C->T, Q1b1a3a1-YP776: 17692899A->G, Q1b1a3a1-YP780: 22201626A->T, Q1b1a3a1-YP773: 17112655C->T.

Mongolia_EIA_4_Sagly sample I12970 could be assigned as haplogroup Q1b1a3a1-L332: 16198895C->G, Q1b1a3a1-Y18330: 18966322C->G, Q1b1a3a1-YP762: 8593849A->T, Q1b1a3a1-YP773: 17112655C->T, Q1b1a3a1-YP776: 17692899A->G.

This sample also showed derived alleles for upstream haplogroup Q1b1a3-L334: 14482079G->A, Q1b1a3-SK1943: 8548403C->A, Q1b1a3-Y5228: 21334535G->A, Q1b1a3-Y5237: 14045306G->A, Q1b1a3-Y5244: 18947607C->A, Q1b1a3-Y5254: 6631084T->A, Q1b1a3-Y5261: 15913458C->G, Q1b1a3-Y5266: 17762362C->T, Q1b1a3-Y11784: 23244840G->C, Q1b1a3-Z35944: 14160938A->G, Q1b1a3-Z35959: 22547421G->A. However, this sample showed ancestral alleles for downstream haplogroup Q1b1a3a1a-L329: 21842270A->G, Q1b1a3a1a-Y22837: 22199505T->C, Q1b1a3a1a-YP4548: 21438831G->T, Q1b1a3a1a-YP4550: 21490040C->A, Q1b1a3a1a-YP4555: 23799691C->A, Q1b1a3a1a-YP4564: 8536606C->A, Q1b1a3a1a-YP4566: 9798961A->C.

Mongolia_BA_1 sample I12973 could be assigned as haplogroup Q1b1a3-L334: 14482079G->A, Q1b1a3-Y5235: 8548403C->A, Q1b1a3-Y5228: 21334535G->A, Q1b1a3-Y5231: 2794258C->A, Q1b1a3-Y5237: 14045306G->A, Q1b1a3-Y5244: 18947607C->A, Q1b1a3-Y5263: 16377904C->T, Q1b1a3-Y5265: 17507994C->G, Q1b1a3-Y5270: 22155850T->C, Q1b1a3-Y5273: 22630254G->A, Q1b1a3-Z35944: 14160938A->G, Q1b1a3-Z35950: 16708573G->T. However, this sample showed ancestral alleles for downstream haplogroup Q1b1a3a1-BZ433: 15739867T->C, Q1b1a3a1-L332: 16198895C->G, Q1b1a3a1-YP762: 8593849A->T, Q1b1a3a1-YP773: 17112655C->T, Q1b1a3a1-YP775: 17423554C->G, Q1b1a3a1-YP776: 17692899A->G, Q1b1a3b-Z35941: 8550111G->A, Q1b1a3b-SK1941: 21489957G->T, Q1b1a3b-Y11941: 15587953T->A.

Mongolia_LBA_1_MongunTaiga sample I13174 could be assigned as haplogroup Q1b1a3-L334: 14482079G->A, Q1b1a3-Y5228: 21334535G->A, Q1b1a3-Y5265: 17507994C->G, Q1b1a3-Y5273: 22630254G->A, Q1b1a3-Z35944: 14160938A->G, Q1b1a3-Z35947: 15761583C->A. However, this sample showed ancestral alleles for downstream haplogroup Q1b1a3a1a-L329: 21842270A->G, Q1b1a3a1a-YP4547: 19164395C->G, Q1b1a3a1a-YP4564: 8536606C->A, Q1b1a3b-Z35941: 8550111G->A, Q1b1a3b-Y12451: 19107462G->A, Q1b1a3b1a-YP1691: 15900681C->T.

Mongolia_EIA_3 sample I13504 could be assigned as haplogroup Q1b1b-Y5084: 15951943G->A, Q1b1b-Y15809: 8688351G->C, Q1b1b-Y15811: 21329060T->A, Q1b1b-YP3960: 8410578C->T, Q1b1b-YP3961: 8450675G->T, Q1b1b-YP4021: 17072017A->T, Q1b1b-YP4045: 19507348T->G, Q1b1b-YP4048: 21142031T->C, Q1b1b-YP4072: 23627073G->A, Q1b1b-YP4074: 23990390G->A, Q1b1b-YP4112: 15456292C->G. This sample showed ancestral alleles for downstream haplogroup Q1b1b1-YP4042: 19195187G->A, Q1b1b1a-YP3952: 6844645A->T, Q1b1b1a-YP3978: 15172292T->C.

Mongolia_LBA_4_CenterWest sample I13766 could be assigned as haplogroup Q1b1a3a-Y20260: 16783479C->A. This sample also showed derived alleles for upstream haplogroup Q1b1a3-L330: 17766807T->C, Q1b1a3-L334: 14482079G->A,

Q1b1a3-Y5228: 21334535G->A, Q1b1a3-Y5230: 7399707G->C, Q1b1a3-Y5231: 2794258C->A, Q1b1a3-Y5238: 14176657G->T, Q1b1a3-Y5244: 18947607C->A, Q1b1a3-Y5265: 17507994C->G, Q1b1a3-Y5270: 22155850T->C, Q1b1a3-Y5273: 22630254G->A, Q1b1a3-Y11235: 7725756C->G, Q1b1a3-Y11784: 23244840G->C, Q1b1a3-Z35944: 14160938A->G. However, this sample showed ancestral alleles for downstream haplogroup Q1b1a3a1-BZ988: 18188735C->T, Q1b1a3a1-L332: 16198895C->G, Q1b1a3a1-YP762: 8593849A->T, Q1b1a3a1-YP772: 17053829G->C, Q1b1a3a1-YP773: 17112655C->T.

Mongolia_LBA_4_CenterWest sample I13767 could be assigned as haplogroup Q1b1a3-L334: 14482079G->A, Q1b1a3-SK1943: 8548403C->A, Q1b1a3-Y5228: 21334535G->A, Q1b1a3-Y5231: 2794258C->A, Q1b1a3-Y5237: 14045306G->A, Q1b1a3-Y5254: 6631084T->A, Q1b1a3-Y5263: 16377904C->T, Q1b1a3-Y5265: 17507994C->G, Q1b1a3-Y5269: 22068356C->T, Q1b1a3-Y5273: 22630254G->A, Q1b1a3-Y12111: 19216291G->A, Q1b1a3-Z35944: 14160938A->G. However, this sample showed ancestral alleles for downstream haplogroup Q1b1a3a1-L332: 16198895C->G, Q1b1a3a1-Y18330: 18966322C->G, Q1b1a3a1-YP762: 8593849A->T, Q1b1a3a1-YP763: 8617072G->A, Q1b1a3b-Z35939: 21489957G->T, Q1b1a3b-Z35942: 15880973A->G, Q1b1a3b-Y12451: 19107462G->A, Q1b1a3b1a-YP1691: 15900681C->T.

Mongolia_EIA_4_Sagly sample I7024 could be assigned as haplogroup R1a1a1b2-Z93: 7552356G->A, R1a1a1b2-M746: 18985344C->A. This sample also showed derived alleles for upstream haplogroup R1a1a1-PF6157: 7347942C->T, R1a1a1-CTS4259: 15525535G->A, R1a1a1-M695: 16414608A->G, R1a1a1-CTS5979: 16591891C->A, R1a1a1-PF6198: 17341487G->C, R1a1a1-CTS7278: 17400173C->T, R1a1a1-PF6208: 19244791G->A, R1a1a1-CTS10993: 22876368C->A, R1a1a1-V3842: 19160342A->G, R1a1a1-F3159: 21225537C->A, R1a1a1-PF6231: 23800360T->G, R1a1a1-M630: 8212820A->G, R1a1a1-PF6169: 9407722A->G, R1a1a1-PF6210: 21162924C->G, R1a1a1-PF6218: 22003657C->T, R1a1a1-PF7532: 23050018C->G, R1a1a1-PF7530: 22190201T->C, R1a1a1b-Z645: 8245045C->T, R1a1a1b-Z647: 7683058G->A, R1a1a1b-Z649: 16329760C->G, R1a1a1b-M750: 19058383C->A, R1a1a1b-Z651: 19419865G->A. However, this sample showed ancestral alleles for downstream haplogroup R1a1a1b2a-Z95: 23956870C->T, R1a1a1b2a1-M723: 17533368G->T, R1a1a1b2d-YP5324: 22653561C->T, R1a1a1b2e-YP1506: 9792228C->T, R1a1a1b2e1-YP1505: 8312442A->G, R1a1a1b2e1-YP1507: 17560160T->C, R1a1a1b2e1-YP1508: 21453767G->A, R1a1a1b2f-Y20793: 19255807T->C, R1a1a1b2g1-Y34287: 8499032C->T, R1a1a1b2h1-Y24548: 15289059G->T, R1a1a1b2h1-Y24739: 6402482G->A.

Mongolia_EIA_4_Sagly sample I7027 could be assigned as haplogroup R1a1a1b2a-F3105: 21043448T->C. This sample also showed derived alleles for upstream haplogroup R1a1a1b-Z647: 7683058G->A, R1a1a1b-CTS5508: 16329760C->G, R1a1a1b-Z650: 19058383C->A, R1a1a1b-V4100: 19419865G->A, R1a1a1b2-Z2479:

18985344C->A. However, this sample showed ancestral alleles for downstream haplogroup R1a1a1b2a1-M723: 17533368G->T, R1a1a1b2a1-F2597: 17801058T->C.

Mongolia_EIA_4_Sagly sample I7030 could be assigned as haplogroup R1a1a1b2a-Z95: 23956870C->T, R1a1a1b2a1-M780: 21610995C->T, but we caution this might be caused by ancient DNA damage. This sample also showed a derived allele for upstream haplogroup R1a1a1b2-Z2479: 18985344C->A, R1a1a1b-Z647: 7683058G->A, R1a1a1b-CTS5508: 16329760C->G, R1a1a1b-M750: 19058383C->A, R1a1a1b-Z651: 19419865G->A. However, this sample showed ancestral alleles for the following downstream haplogroup R1a1a1b2a1a1a-M787: 22679436C->G, R1a1a1b2a1a1a1-M699: 16586743C->T, R1a1a1b2a1a1a1-Y11: 6512670C->A.

Mongolia_LBA_3_MongunTaiga sample I7033 could be assigned as haplogroup R1a1a1b2a-Z95: 23956870C->T, but we caution this might be caused by ancient DNA damage. This sample also showed derived alleles for upstream haplogroup R1a1a1b2-S4582: 18985344C->A, R1a1a1b-S441: 7683058G->A, R1a1a1b-Z649: 16329760C->G, R1a1a1b-M750: 19058383C->A, R1a1a1b-Z651: 19419865G->A. However, this sample showed ancestral alleles for downstream haplogroup R1a1a1b2a1-M723: 17533368G->T, R1a1a1b2a1-M780: 21610995C->T.

Mongolia_EBA_2_Chemurchek sample I12957could be assigned as haplogroup R1b1a1b-PF6419: 6912992T->G, R1b1a1b-CTS894: 7073423G->A, R1b1a1b-M8208: 7391161C->T, R1b1a1b-M8209: 7391164G->A, R1b1a1b-PF6453: 14317555G->A, R1b1a1b-CTS2664: 14416216G->A, R1b1a1b-PF6457: 15037433C->G, R1b1a1b-CTS4608: 15732786T->C, R1b1a1b-CTS6532: 16971648T->G. However, this sample showed ancestral alleles for downstream haplogroup R1b1a1b2-Z29759: 21110969A->G, R1b1a1b2-V2381: 8627510C->T, R1b1a1b2a-PF7563: 8610246G->A, R1b1a1b1a-L51: 8502236G->A, R1b1a1b1a1-S129: 18907236A->C, R1b1a1b1a1-L52: 14641193C->T.

Mongolia_EBA_2_Chemurchek sample I12978 could be assigned as haplogroup R1b1a1b-M269: 22739367T->C, R1b1a1b-PF6419: 6912992T->G, R1b1a1b-CTS894: 7073423G->A, R1b1a1b-PF6422: 7306539C->G, R1b1a1b-PF6457: 15037433C->G, R1b1a1b-CTS4608: 15732786T->C, R1b1a1b-CTS6532: 16971648T->G, R1b1a1b-PF6469: 17461478T->C, R1b1a1b-CTS7659: 17594966C->G, R1b1a1b-CTS8052: 17813541C->T, R1b1a1b-PF6478: 18117193C->T, R1b1a1b-CTS8665: 18137831T->C, R1b1a1b-CTS8728: 18167403C->T, R1b1a1b-PF6492: 19417394A->C, R1b1a1b-CTS10451: 19462180C->T, R1b1a1b-PF6520: 23124367G->T, R1b1a1b-CTS11948: 23379254G->A, R1b1a1b-PF6529: 28590278G->A, R1b1a1b-CTS12972: 28771116C->G. This sample also showed a derived allele for R1b1a1b1-L23: 6753511G->A, but we caution this might be caused by ancient DNA damage. However, this sample showed ancestral alleles for R1b1a1b1a-L51: 8502236G->A, R1b1a1b1a1-S129: 18907236A->C, R1b1a1b2-PH1348: 14579448G->T, R1b1a1b2a-PF7563: 8610246G->A.

Mongolia_IA_Xianbei sample I13175 could be assigned as haplogroup R1b2b-BY14575: 8058964T->C, R1b2b-BY14577: 8849646G->T, R1b2b-BY14587:18977325C->T, R1b2b-SK2058: 15272992C->T, R1b2b-SK2060: 22940274C->T, R1b2b-Y23412: 8602145C->T. This sample also showed derived alleles for upstream haplogroup R1b2-BY14355: 8400716A->G, R1b2-BY14356: 8569947A->T, R1b2-BY14362: 13685191A->G, R1b2-BY14367: 15667083C->A, R1b2-BY14369: 16552474A->G, R1b2-BY14370: 16563372G->A, R1b2-BY14374: 19251241C->T, R1b2-BY14377: 22133452T->C, R1b2-BY14393: 28790180T->G, R1b2-BY14573: 18835322A->C, R1b2-PH861: 14072846C->T, R1b2-PH1030: 14231273A->G, R1b2-PH1165: 14338481C->T, R1b2-PH1417: 14657839G->C, R1b2-PH1769: 15583344A->C, R1b2-PH1840: 15831337G->A, R1b2-PH2150: 16283543T->G, R1b2-PH4622: 21082641T->G, R1b2-PH4796: 21388393T->A, R1b2-SK2056: 14773867G->A, R1b2-SK2059: 8483973G->A, R1b2-SK2061: 16389874G->T, R1b2-Y15399.2: 23203615C->T, R1b2-Z21341.2: 22625041G->C.

Mongolia_Medieval sample I13176 could be assigned as haplogroup R1b2-BY14355: 8400716A->G, R1b2-BY14369: 16552474A->G, R1b2-BY14394: 8163837T->G, R1b2-SK2056: 14773867G->A. However, this sample showed ancestral alleles for the haplogroup R1b2a-BY17637: 8285883G->A, R1b2a-Y112334: 23796662G->A, R1b2b-Y23412: 8602145C->T, R1b2b1-Y33158: 6857062C->T, R1b2b1a1-Y33152: 14978995C->T.