



# The formation of proto-austronesians: insights from a revised phylogeography of the paternal founder lineage

Hui-Xin Yu<sup>1</sup> · Xian-Peng Zhang<sup>1</sup> · Yun Bai<sup>2</sup> · Kai-Jun Liu<sup>3</sup> · Hui Li<sup>4,5</sup> · Yong-Lan Li<sup>1</sup> · Jin Sun<sup>6</sup> · Lan-Hai Wei<sup>5</sup> 

Received: 3 March 2023 / Accepted: 1 July 2023

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

## Abstract

**Objectives** Previous studies suggested that the Y-chromosome haplogroups O2-N6-B451-AM01756 and O1a-M119 are two founder lineages of proto-Austronesians at about five thousand years ago. The objective of this study was to investigate the formation of proto-Austronesians from the perspective of the paternal gene pool.

**Materials and methods** In this study, we developed a highly revised phylogenetic tree with age estimates for haplogroup O2-N6 and early branches of O1a-M119 (M110, F1036, and F819). In addition, we also explored the geographical distribution of eight sub-branches of O2-N6 and O1a-M119, and spatial autocorrelation analysis was conducted for each sub-branch.

**Results** The paternal lineage combination of proto-Austronesians is a small subset of a diverse gene pool of populations from the mainland of East Asia. The distribution map and results of the spatial autocorrelation analysis suggested that the eastern coastal region of northern China is likely the source of lineage O2-N6 while the coastal region of southeastern China is likely the diffusion center of early branches of O1a-M119. We developed an evolutionary diagram for Austronesians and their ancestors in the past 18,000 years.

**Discussion** We proposed that the millet farming community in North China is the common ancestor group of the Austronesians and the Han people, while the diverse ancient people in the southeast coastal areas of East Asia form the common ancestor group of the Austronesians and the East Asian mainland population. The demographic history of multiple ancestral groups of the most recent common ancestor group itself in the more ancient period is helpful to understand the deep roots of the genetic components and cultural traditions of Austronesians.

**Keywords** Austronesians · Phylogeography · Founder lineage · O2-N6 · O1a-M119

---

Communicated by Shuhua Xu.

---

Hui-Xin Yu, Xian-Peng Zhang and Yun Bai authors contributed equally to the study.

---

✉ Jin Sun  
sunjin19851228@gmail.com

✉ Lan-Hai Wei  
Ryan.lh.wei@gmail.com

<sup>1</sup> Institute of Humanities and Human Sciences, School of Ethnology and Anthropology, Inner Mongolia Normal University, Hohhot 010028, China

<sup>2</sup> College of Life Science and Technology, Inner Mongolia Normal University, Hohhot 021002, China

## Introduction

The origin and differentiation history of Austronesian-speaking populations is an important topic and has long fascinated linguists, archaeologists, and geneticists. It is generally believed that the recent common ancestors of Austronesian speakers resided on both sides of the Taiwan Strait, including Taiwan, around 5.0–4.0 thousand years

<sup>3</sup> School of International Tourism and Culture, Guizhou Normal University, Guiyang 550001, China

<sup>4</sup> MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai 200438, China

<sup>5</sup> B&R International Joint Laboratory for Eurasian Anthropology, Fudan University, Shanghai 200438, China

<sup>6</sup> School of Literature and Media, Xingyi Normal University for Nationalities, Xingyi 562400, China

ago (kya), but the more ancient common ancestors of the Austronesians are ambiguous (Ko et al. 2014; Yang et al. 2020). Archaeological evidence indicates that Neolithic cultural relics appearing in the southeastern coastal region of China (from Fujian to Zhejiang) may be related to the common ancestral groups of the Austronesian population. There are also some specific cultural factors of Taiwanese Aborigines, such as millet planting and jade, that are suggestive of a relationship with millet agricultural culture (Sagart et al. 2018). Therefore, linguists and ethnologists believed that part of the ancestral groups of Austronesians may have migrated from the eastern coastal region of North China. However, there is not enough archaeological evidence to support this theory of ancient population migration.

Paleogenomic research revealed the complex differentiation history of the Austronesian population. The Early Neolithic individuals in the coastal region of Fujian, such as Liangdao and Qihedong, may be one of the ancestral groups of Austronesian speakers (Yang et al. 2020). For the ancient people in Fujian and Taiwan after the Neolithic Age, such as Hanben individuals (1.6 kya) show a closer genetic relationship with modern Austronesian populations (Wang et al. 2021). However, ancient DNA research has not found obvious admixture signals related to ancient populations in North China in the Early and Middle Neolithic Ages in ancestral groups of Austronesians. This may be attributed to the limited number of ancient samples or the low SNP coverage of ancient DNA. Extensive maternal mitochondrial DNA (mtDNA) research indicates that Austronesian speakers share most of their maternal lineages with people in the southeast of the East Asian mainland.

The research on the Y chromosome lineage provides paternal insight to explore the origin and diversification process of Austronesian populations. Previous studies have found that haplogroup O2-N6 and varied sub-branches of haplogroup O1a-M119 are founding paternal lineages of Austronesian speakers (Sun et al. 2021; Wei et al. 2017). The higher frequency of haplogroups O1b1a1a-M95, C2-M38, S-B254, and M-P256 in Austronesian populations in Southeast Asian Island and Oceania may be related to recent population admixture, and ancient DNA evidence favors this view. Our previous studies explored deeply the sub-lineages of O1a-M119 (especially O1a1a-M307.1) and proposed a shared paternal ancestry of the Han, Tai-Kadai-speaking (Sun et al. 2021). And this study provided many haplogroup O2-N6 sequences of Austronesian populations to Karmin et al. (Karmin et al. 2022), and finally revealed the formation process of Austronesian populations in Southeast Asia and Oceania. In general, the genetic structure and differentiation process of the common ancestor of Austronesians was determined based on different genetic makers. However, the origin and admixture process of common ancestral

groups of Austronesian populations in ancient times remain ambiguous.

In this study, we focused on the differentiation process of the founding paternal lineages of Austronesian speakers in the past 18 kya to 3 kya. First, we generated a highly-revised phylogenetic tree for O2-N6 and varied early sub-branches of O1a-M119 and attempted to find some paternal lineages in the mainland populations that are closer to the founding paternal lineage of Austronesians (O2-N6-B451-AM01756). Second, we divided the sub-lineages of haplogroups O2-N6 and O1a-M119 into eight categories and explored the geographic distribution of each of the sub-branches. Finally, we constructed the evolutionary diagram of Austronesian ancestral groups in the past 18 kya to 3 kya.

## Materials and methods

### Samples and sequencing

Saliva samples were collected from unrelated healthy males in East Asian populations over the past few decades. All of the participants provided written informed consent prior to participating. The study and sample collection process were reviewed and approved by the Medical Ethics Committee of Fudan University and Inner Mongolian Normal University and complied with the ethical principles of the 2013 Helsinki Declaration of the World Medical Association. Following the steps in our previous study, the frequency data of O2-N6 and early sub-branches of O1a-M119 among Chinese participants were drawn from our in-house database, which contains various types of information for each sample. The frequency data was used to generate the distribution map (Supplementary Table S1). The revised phylogenetic tree was adopted from public resources and the genetic genealogy community (Supplementary Table S2). We explain the data process and age estimation of public resources in the Supplementary text.

### Data analysis

We followed the steps in our previous study to draw the distribution map. We used ArcGIS software to conduct the spatial autocorrelation analysis. Moran's I Index indicated the degree of correlation between the distribution frequency and spatial geographical distribution. The HotSpots plot indicated the clustering region of hot spots and cold spots, which generally correspond to the center of diffusion and its mirror image region shown by the general distribution pattern. The Lisa Cluster shows the cluster and outliers of the general distribution pattern of high-low frequency.

## Results

A schematic representation of the revised phylogeny of the studied lineages with age estimations is shown in Fig. 1. The detailed tree is in Supplementary Table S1. We detected a complex downstream structure and many new SNPs and sub-branches. The distribution map and the results of the spatial autocorrelation analysis (HotSpots plot) are shown in Fig. 2. The results of the Lisa Cluster map and Moran's I Index are shown in Figure S1.

The flourishing pattern of the paternal haplogroup O2-N6 (since 15 kya) is shown in Fig. 1. There are 60 sub-lineages of O2-N6 with a differentiation age of over 5,000 years, and O2-N6-B451-AM01756 which is a specific paternal lineage of Austronesian-speaking populations, is just one of them (Table S1). The results of the geographic distribution map and HotSpots plot of the spatial autocorrelation analysis supported the theory that the circum-Bohai-sea region in the East of North China is the diffusion center of four sub-branches of O2-N6, including O2a2b2\*-N6x (N7, F706) (Fig. 2A), O2a2b2a1-N7 (Fig. 2B), O2a2b2a2\*-F706xB451 (Fig. 2C), and O2a2b2a2b-B451 (Fig. 2D). There is an absence of haplogroup AM01756 in the mainland population, so the geographic distribution of O2a2b2a2b-B451 shown in Fig. 2D corresponds to the geographic distribution of O2a2b2a2b\*-B451xAM01756. Therefore, we can consider that the geographic distribution on the mainland shown in Fig. 2D corresponds to the branches that are closer to the founding paternal lineage AM01756. The eastern coastal region of North China is the diffusion center of O2a2b2a2b-B451, but the result of the Lisa Cluster analysis showed that there is an abnormal high-frequency distribution in Fujian. The phylogenetic tree of AM01756 showed this haplogroup experienced a 2,400-year-bottleneck period before its expansion in 4.8 kya. These results strongly support that the distribution of AM01756 in Fujian and Taiwan is related to a long-distance migration from the eastern coastal regions of North China to the coastal region of southeastern China.

In this work, we focused on three early branches of O1a-M119, including O1a-F819, O1a-M110, and O1a-F1036. Paleogenomic research showed that the paternal haplogroup of the ~ 8 kya Liangdao individual (LD1) is O1a-F819, and for the ~ 4.3 kya Tanshishan individual (ID: L7415), it is O1a-M110. The haplogroup O1a-M110 is one of the dominant paternal lineages in modern Austronesian- and Tai-Kadai-speaking populations, O1a-F819 has a lower frequency, and O1a-F1036 has not been found in modern Austronesian and Tai-Kadai-speaking populations. The results of the geographic distribution map and spatial autocorrelation analysis indicated the diffusion centers of

these three paternal lineages are southeastern China, the South China coastal region, and the eastern coastal region of North China. The different distribution patterns of these three paternal lineages in modern and ancient populations proposed different roles in the formation process of Austronesian populations.

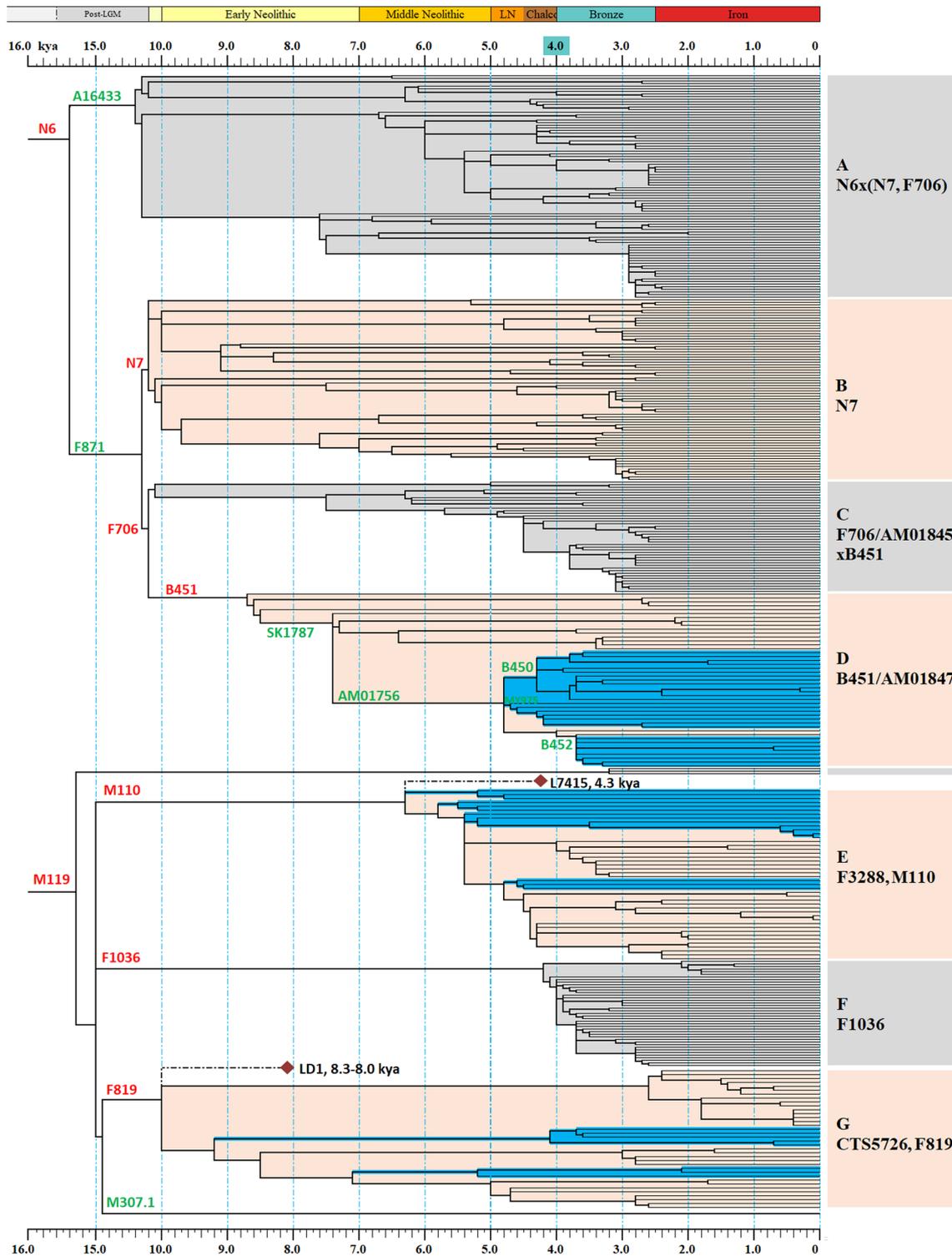
## Discussion

### Cultural affiliation between Austronesians and Han populations

The demographic history of the varied ancestor groups of the most recent common ancestors in the more ancient historical period is helpful in understanding the remote affiliations of the various attributes between different groups of modern populations. The diffusion center of modern Austronesians is remote to the middle and lower Yellow River Basin, which is the early cradle of the Han people (Ning et al. 2020; Wang et al. 2021), but the remote affiliations of the two ethnic groups can be found in cultural traditions, such as millet agriculture, jade, and linguistic factors (Sagart et al. 2018). It is generally believed that the recent common ancestors of Austronesians resided in the southeast coastal region of the mainland, including Taiwan, around 5.0–4.0 kya. However, we proposed a long-distance migration of O2-N6-B451-AM01756, the founding paternal lineage of Austronesians, from the eastern coastal region of North China to the southeast coastal region of China. Therefore, Austronesians and Han people have a partial common ancestry related to the millet farming communities in North China during the early and middle Neolithic Age. The paternal lineage O2-N6-B451-AM01756 is the founding paternal lineage of the Austronesians, but other sub-branches of O2-N6 can be found in modern Han people. We suggest that the origin and demographic history of the founding lineage (O2-N6-B451-AM01756 here) in the more ancient historical period contributed profoundly to the remote cultural affiliation between the two different population groups (Austronesians and Han populations here).

### The differentiation of O1a-M110

Previous studies proposed paternal lineage O1a-M110 was shared by the Han people, Austronesian-speaking population, and Tai-Kadai-speaking population (Sun et al. 2021). In this study, our higher-resolution phylogenetic tree of O1a-M110 is shown in Fig. 1 and Figure S1. There are five specific sub-lineages in Austronesians (B392, Z38625, Y33185, MY252, and Y180671), and four of them are early branches. The ages of Y175056 and M3444, which are shared by Han people and Tai-Kadai speakers, are younger than that of the

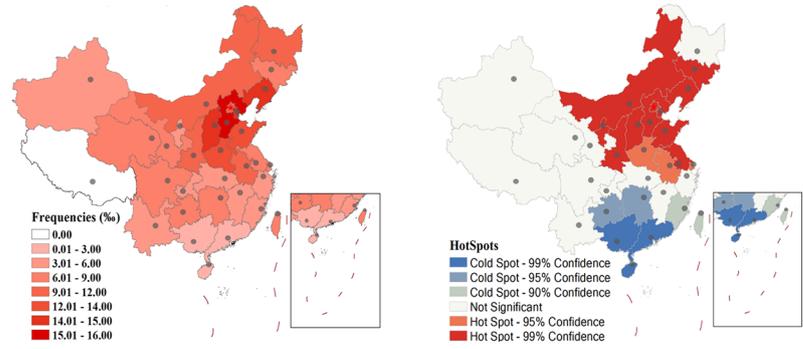


**Fig. 1** Schematic diagram of the revised phylogeny of haplogroups O2-N6 and O1a-M119. The red SNP labels indicate the definition marker of categories of sub-branches in this study. The green SNP

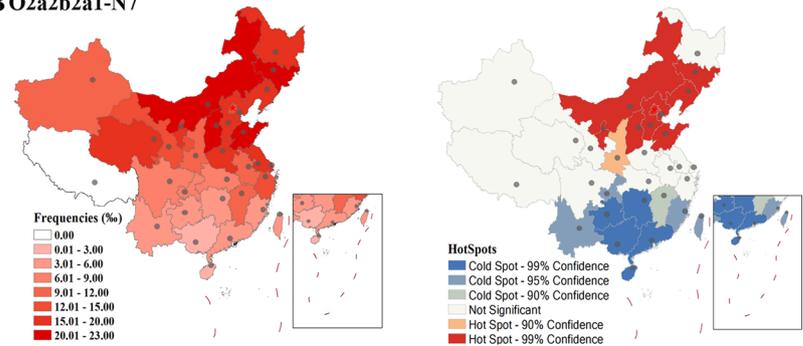
labels indicate other important markers of the phylogenetic tree. The blue color bar represents the unique branch of the Austronesians. Detailed information can be found in Table S1

**Fig. 2** Distribution and HotSpots analysis of eight sub-branches of O2-N6 and O1a-M119. The hot spot area of the HotSpots plot indicates the most likely region for the start of diffusion

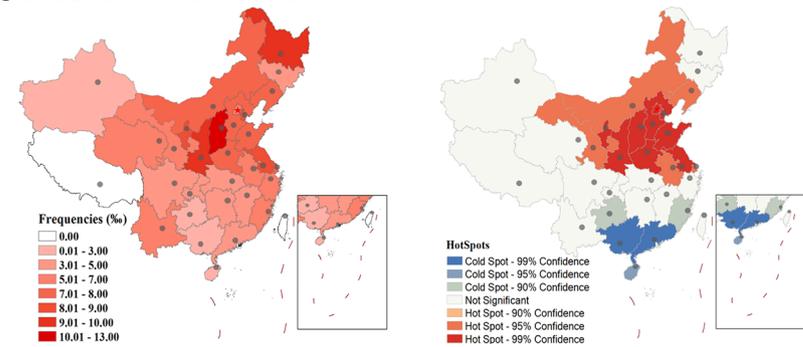
**A O2a2b2\*-N6x(N7, F706)**



**B O2a2b2a1-N7**



**C O2a2b2a2\*-F706/AM01845xB451**



**D O2a2b2a2b-B451/AM01847**

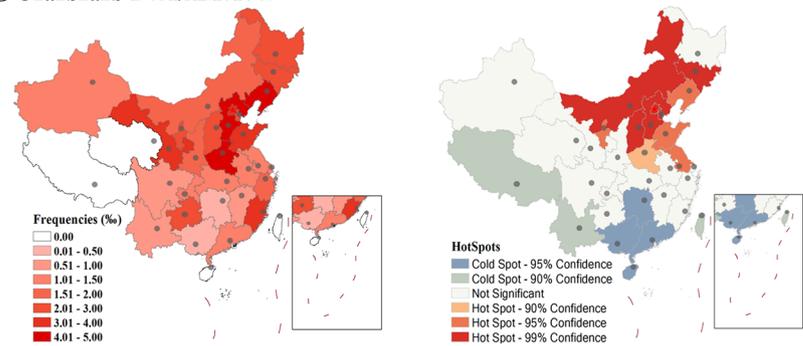
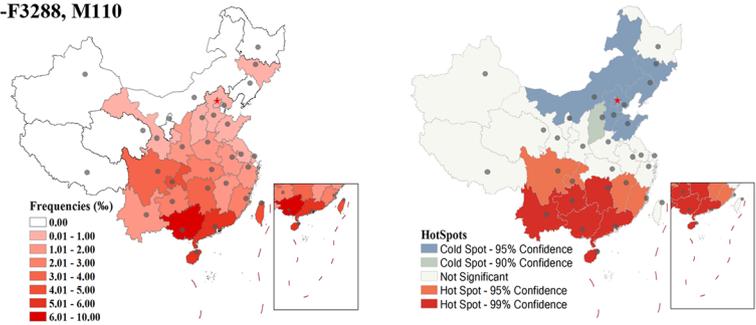
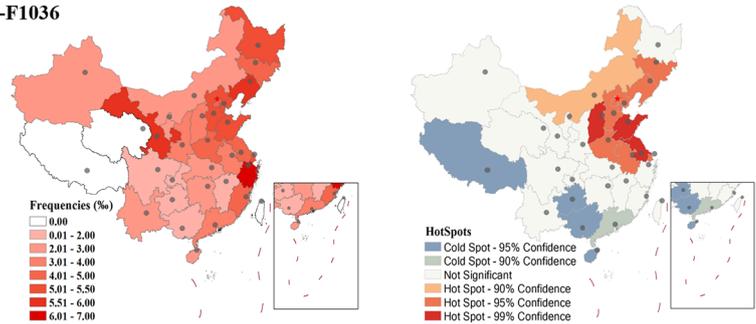
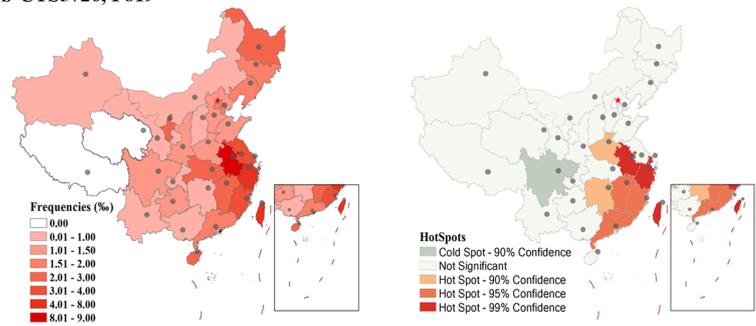
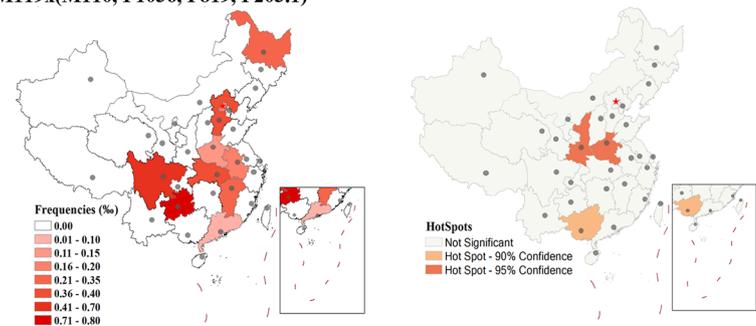


Fig. 2 (continued)

**E O1a2-F3288, M110****F O1a3-F1036****G O1a1b-CTS5726, F819****HO1a-M119x(M110, F1036, F819, P203.1)**

five specific sub-branches of Austronesians. Overall, M110 in Tai-Kadai speakers can be considered a branch of sub-lineages of Austronesians from the whole differentiation pattern. The ~4.3 kya Tanshishan individual (ID: L7415) is the most ancient people belonging to O1a-M110. The extension

period of the Tanshishan culture (5.5–4.0 kya) is consistent with the key differentiation period of O1a-M110, during which the special sub-branches emerged for Austronesians and Tai-Kadai-speaking populations. However, the high frequency of O1a-M110 in a group of people in Shanghai may

be the result of the northward expansion of this haplogroup in a relatively late historical period.

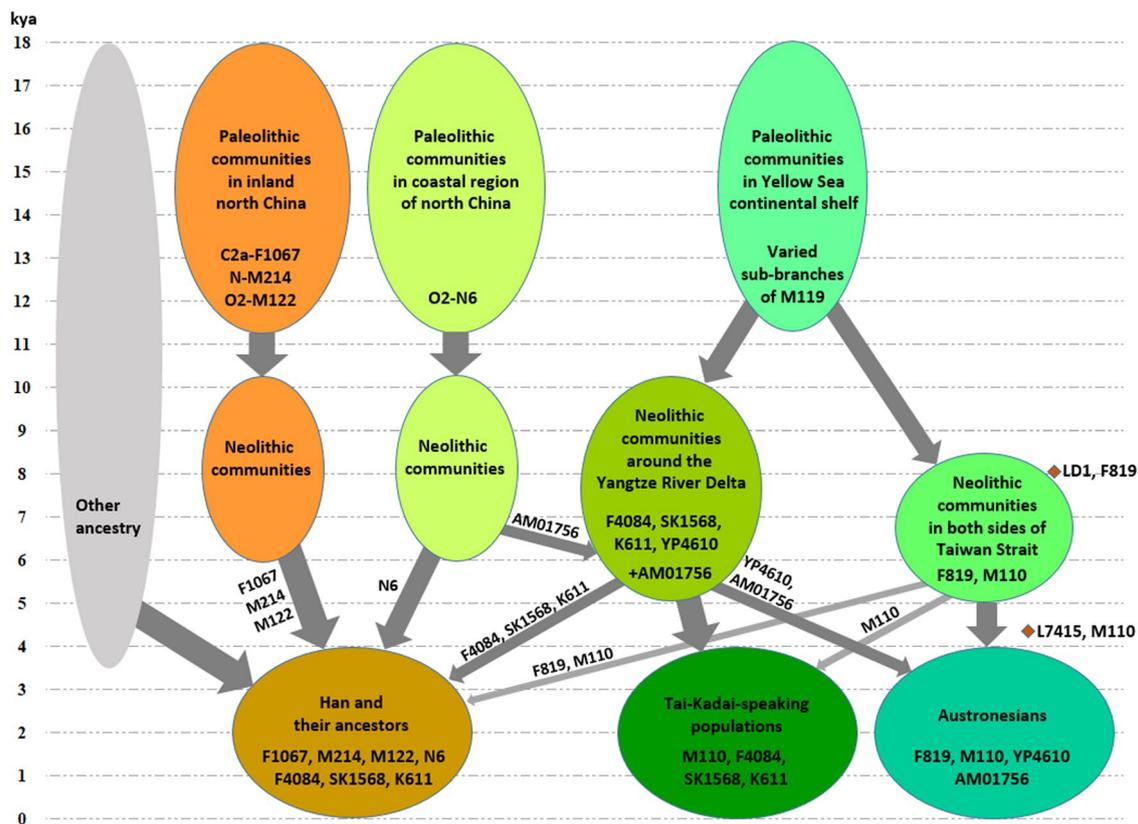
Since haplogroup O2-N6 is rare in Tai-Kadai-speaking populations, haplogroup O1a-M110 represents an important founder lineage that is shared by Austronesians and Tai-Kadai-speaking populations. Using the available evidence of ancient DNA and modern distribution, we propose that the expansion of O1a-M110, as well as other sub-branches of O1a-M307.1 in the mainland of East Asia, may have contributed significantly to the cultural affiliation of the ancient Baiyue people in southern China. More sequences are needed to provide a better resolution of the phylogeny that can represent the expansion and migration that directly led to the formation of ancient populations in the historical records of South China.

### Diversification history of Proto-Austronesians in the past 15,000 years

Population admixture is a common characteristic of modern and ancient populations that is revealed by obtaining ancient DNA and genomic data of modern populations. Previous studies considered the Han people, Tai-Kadai-speaking population, and Austronesians as three ethnic groups with

independent diversification histories. Researchers explored the cultural relationship among these three ethnic groups, but there are many controversial topics related to these studies. Population genetic ancestry is divided into multiple components, the differentiation history of every component has been explored in modern genetic research, and different genetic components may show disparate diversification processes and population relationships of ethnic groups. For example, the Han people, Tai-Kadai-speaking population, and Austronesians show a closer population relationship, which is brother branches in the phylogenetic tree of haplogroup O1a-F140. Tai-Kadai speakers are a sub-branch of Austronesian speakers, which is supported by paternal lineage O1a-M110. F4084 is shared by the Han people and Tai-Kadai speakers, O2-N6 is shared by the Han people and Austronesians. There are also specific genetic ancestries in these three ethnic groups. Therefore, if we accept population admixture as common, then it is too simple to explore only the genetic affinity between two ethnic groups in a study.

In this work, we mapped the evolutionary diagram of ancestral groups of Austronesian speakers in the past 150,000 years by combining ancient DNA and paternal lineage information (Fig. 3). We propose that there are two different ancestral groups of Austronesian and



**Fig. 3** The proposed evolution pattern of common ancestor groups of Austronesians and their relationship with that of the Han people and the Tai-Kadai-speaking populations

Tai-Kadai-speaking populations. The first ancestral group is ancient populations on both coasts of the Taiwan Strait during the Early Holocene to the early and middle Neolithic Age. These people are represented by the Liangdao individual, Qihedong individual, and Tanshishan individual, and the main paternal lineages may be O1a-M110 and O1a-M819. The second ancestral group is the Neolithic populations on the Yangtze River Delta, and their paternal lineages belong to sub-branches of O1a-M307 including F4084, SK1568, K611, and YP4610. In the past 7 kya to 5 kya, the people who migrated from the coastal region of North China took the O2-N6 lineage to the Yangtze River Basin, and then the mixed populations further migrated southward to the coastal region of Fujian and mixed with the first ancestral group to form the recent common ancestor of Austronesian speakers. The complex admixture mode of the Han people, Tai-Kadai speakers, and Austronesian speakers is shown in Fig. 3. In general, we believe the differentiation and admixture process in Fig. 3 is helpful in understanding the complex diversification process and population relationship in Austronesian populations.

In conclusion, we explored the founding paternal lineages of Austronesian-speaking populations. The results of our high-resolution geographic distribution analysis strongly support that O2-N6-B451-AM01756 in Austronesian speakers is related to a long-distance southward migration from the eastern coastal region of North China. The phylogenetic framework of O1a-M110 is helpful to understanding the recent common ancestors between Tai-Kadai-speaking and Austronesian-speaking populations. Finally, we provided a simple mode related to the evolutionary pattern of Proto-Austronesian groups since the end of the Last Glacial Maximum. We proposed there are two different ancestral groups in Tai-Kadai- and Austronesian-speaking populations. In the future, more paleogenomic research and a higher-resolution phylogenetic tree may provide a clearer profile of population origin and evolutionary history.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00438-023-02054-w>.

**Acknowledgements** We thank all of the donors for providing DNA samples and/or DNA sequences. This study was supported by \*\*. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Funding** This study was funded by the The Fundamental Research Funds for the Universities of Inner Mongolia Autonomous Region (Grand no: 2022JBBJ004 to LHW). Major Projects of the National Social Science Foundation (Grand No: 20&ZD248 to XHD). The Key projects of strategic international scientific and technological

innovation cooperation of the Chinese Ministry of Science and Technology (Grand no: 2020YFE0201600).

**Data availability** The sources of raw data used in this study had been fully list in supplementary Table S2.

## Declarations

**Conflict of interest** The authors declare that they do not have any conflicts of interest.

## References

- Karmin M, Flores R, Saag L, Hudjashov G, Brucato N, Crenna-Darussallam C, Cox MP (2022) Episodes of diversification and isolation in Island Southeast Asian and near Oceanian Male Lineages. *Mol Biol Evol*. <https://doi.org/10.1093/molbev/msac045>
- Ko AM, Chen CY, Fu Q, Delfin F, Li M, Chiu HL, Ko YC (2014) Early Austronesians: into and out of Taiwan. *Am J Hum Genet* 94(3):426–436. <https://doi.org/10.1016/j.ajhg.2014.02.003>
- Ning C, Li T, Wang K, Zhang F, Li T, Wu X, Cui Y (2020) Ancient genomes from northern China suggest links between subsistence changes and human migration. *Nat Commun* 11(1):2700. <https://doi.org/10.1038/s41467-020-16557-2>
- Sagart L, Hsu TF, Tsai YC, Wu CC, Huang LT, Chen YC, Hsing YC (2018) A Northern Chinese origin of Austronesian agriculture: new evidence on traditional Formosan cereals. *Rice (n Y)* 11(1):57. <https://doi.org/10.1186/s12284-018-0247-9>
- Sun J, Li YX, Ma PC, Yan S, Cheng HZ, Fan ZQ, Wei LH (2021) Shared paternal ancestry of Han, Tai-Kadai-speaking, and Austronesian-speaking populations as revealed by the high resolution phylogeny of O1a-M119 and distribution of its sub-lineages within China. *Am J Phys Anthropol* 174(4):686–700. <https://doi.org/10.1002/ajpa.24240>
- Wang CC, Yeh HY, Popov AN, Zhang HQ, Matsumura H, Sirak K, Reich D (2021) Genomic insights into the formation of human populations in East Asia. *Nature* 591(7850):413–419. <https://doi.org/10.1038/s41586-021-03336-2>
- Wei LH, Yan S, Teo YY, Huang YZ, Wang LX, Yu G, Li H (2017) Phylogeography of Y-chromosome haplogroup O3a2b2-N6 reveals patrilineal traces of Austronesian populations on the eastern coastal regions of Asia. *PLoS ONE* 12(4):e0175080. <https://doi.org/10.1371/journal.pone.0175080>
- Yang MA, Fan X, Sun B, Chen C, Lang J, Ko YC, Fu Q (2020) Ancient DNA indicates human population shifts and admixture in northern and southern China. *Science* 369(6501):282–288. <https://doi.org/10.1126/science.aba0909>

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.