Human Biology Open Access Pre-Prints

WSU Press

3-15-2024

The Origin and Dispersal of Austroasiatic Languages from the Perspectives of Linguistics, Archeology and Genetics

Jian Yu
Xiamen University

Xiaohua Deng
XiamenUniversity

Zhiquan Fan
Fujian Normal University

Wenjiao Yang
Xiamen University, Ningxia University

Zhi Ji
XiamenUniversity

See next page for additional authors

Recommended Citation

Yu, Jian; Deng, Xiaohua; Fan, Zhiquan; Yang, Wenjiao; Ji, Zhi; and Wang, Chuanchao, "The Origin and Dispersal of Austroasiatic Languages from the Perspectives of Linguistics, Archeology and Genetics" (2024). *Human Biology Open Access Pre-Prints*. 211.
https://digitalcommons.wayne.edu/humbiol_preprints/211

Authors

Jian Yu, Xiaohua Deng, Zhiquan Fan, Wenjiao Yang, Zhi Ji, and Chuanchao Wang

The Origin and Dispersal of Austroasiatic Languages from the Perspectives of Linguistics, Archeology and Genetics

Jian Yu[1], Xiaohua Deng[1], Zhiquan Fan[2], Wenjiao Yang[1,3], Zhi Ji[1], Chuan-Chao Wang[1,4,5*]

[1]Department of Anthropology and Ethnology, Institute of Anthropology, Fujian Provincial Key Laboratory of Philosophy and Social Sciences in Bioanthropology, School of Sociology and Anthropology, Xiamen University, Xiamen, China

[2]Fujian Normal University Fujian-Taiwan Regional Research Center, School of Letters, Fuzhou, China

[3]School of Foreign Languages and Cultures, Ningxia University, Yinchuan, China.

[4]State Key Laboratory of Cellular Stress Biology, School of Life Sciences, Xiamen University, Xiamen, China

[5]Ministry of Education Key Laboratory of Contemporary Anthropology, Department of Anthropology and Human Genetics, School of Life Sciences, Fudan University, Shanghai, China

*Correspondence to: Chuan-Chao Wang, wang@xmu.edu.cn

The Austroasiatic (AA) languages are a large language family in Mainland Southeast and South Asia. Theoretical, methodological, and material constraints have limited research on the origin and dispersal of AA-speaking populations within historical-comparative linguistics. With the deepening of archaeological and genetic studies, interdisciplinary collaboration has become the key to solving this problem. Based on the latest achievements of linguistics, archaeology and molecular anthropology, we review the hypotheses and propose insights on the origin and dispersal of AA languages. The ancestors of the AA-speaking populations were suggested to be the rice farmers living in the Neolithic Age in southern China. During 3000–4500 BP, some of these ancestors who spoke Proto-AA migrated from southern China to northern Vietnam together with shouldered stone tools and domesticated rice. They mixed with local hunter-gatherers and expanded to the south of Mainland Southeast Asia (MSEA), giving rise to the Mon-Khmer, Aslian, and Nicobarese populations. They also spread to the

Preprint version. Visit http://digitalcommons.wayne.edu/humbiol/ after publication to acquire the final version.

northeast of India to form the Munda-speaking populations. Another group arrived near Dian Lake in Yunnan about 2500 BP, where they created the Bronze Drum culture with the Proto-Tai-Kadai (TK)-speaking populations and later spread eastward to northern Vietnam via Guangxi. Finally, the Proto-AA-speaking people who remained in southern China mixed with the Proto-TK-speaking groups from Fujian and Guangdong, leading to a language shift, which we hypothesized was one of the main reasons for the "disappearance" of the AA in southern China.

Keywords: Language, Archaeology, Genetics, Austroasiatic-speaking populations, Evolution.

1 INTRODUCTION

The AA family consists of over 150 languages and dialects, with a total population of over 65 million speakers. It is widely distributed throughout MSEA, Northeast India, and the Nicobar Islands (Diffloth, 2018). There are 12 AA languages in China, mainly distributed in Yunnan and Guangxi provinces (Dao J., 2018).

The AA family, the oldest language family in MSEA and Northeast India, is considered the underlying structure of all languages in Southeast Asia (Li Y., 2016), making this language family critically important. However, the history of AA studies is relatively short, with the earliest research tracing back to J. R. Logan's *The Ethnology of the Indian Archipelago* in 1852. It was not until the early 20th century that German anthropologist W. Schmidt defined the AA family. After that, research on AA was mired in a long-standing debate over its internal relationships and subgrouping. Schmidt (1907–1908, 1926), Maspero (1929), Pinnow (1959), Thomas (1966), Headley (1976), Diffloth (1979, 2005), Sidwell (2010), and others have proposed different classification schemes based on different types of wordlists and research methods.

Given the chaotic classification of AA languages, the Ethnologue website divides the AA languages into two major branches: the Mon-Khmer branch in MSEA and the Munda branch in eastern India. Meanwhile, the confusion in the classification of AA has seriously hindered the determination of the original homeland of Proto-AA (Li Y., 2016).

Previous scholars have proposed various hypotheses based on their classification frameworks and theories, but the issues of the origin, the timing of differentiation, and dispersal pathways of the AA family have not been resolved to this day. Additionally, the evolution of AA-speaking populations is closely related to the TK and Austronesian (AN) populations in Southern China and Southeast Asia. Resolving the issues of the origin and dispersal of AA-speaking populations is crucial to understanding the evolutionary relationship between these three language families. To this end, this article attempts to combine the latest research findings in linguistics, archaeology, and molecular anthropology to propose new implications for the issues of origin and dispersal of the AA family.

2 Hypotheses on the Origin and Dispersal of the AA Family

Currently, there are mainly four hypotheses on the homeland of the AA family, namely, (1) northern Myanmar, (2) northeastern India or the Bay of Bengal, (3) the Mekong River basin in Southeast Asia and (4) the middle Yangtze River or Southern China.

2.1 Northern Myanmar

Schmidt (1906) believed there was a cognate relationship between the AA and AN languages and proposed including them in a larger language family called “Austric.” This viewpoint was supported by scholars such as Diffloth (1994) and Reid (1994). Blust (1996) provided a clearer explanation, suggesting that around 8000 BP, the ancient homeland of the AA and AN language families, was located in northern Myanmar. Later, the Mon-Khmer branch of the AA family spread southward along the Mekong River to MSEA, while the Munda branch spread westward along the Brahmaputra River to northern India. However, linguists have widely questioned the Austric hypothesis due to the small number of common lexicons between AA and AN, even Blust (2013) himself stated that he no longer supports this hypothesis. Therefore, the hypothesis of the AA family originating in northern Myanmar has been largely excluded from the current discussion on the origin of the AA languages.

2.2 Northeastern India or the Bay of Bengal

American scholars Donegan and Stampe (2004) believe that the AA in Northeast India and the Bay of Bengal appeared earlier than those in Southeast Asia based on profound phonetic structural changes within the Munda language family. For the first time, they proposed that the AA originated from the Indian subcontinent and spread eastward to Southeast Asia. However, this hypothesis of the origin of AA in the West was established by American scholar Diffloth (2005), who used linguistic paleontology as a theoretical basis. He used reconstructed words for animal species to infer the homeland of the proto-AA language family. The reconstructed proto-AA roots for animals like monitor lizard, pangolin, water buffalo, goat, civet, elephant, peacock, rhinoceros, and bamboo rat suggest that the homeland of the original AA family cannot be in temperate regions (i.e., China) but in tropical regions. However, the overlapping area based on reproducible faunas remains large, specifically covering Northeast India, the India-Myanmar border and Myanmar. Based on this, Diffloth proposed that the deepest language differentiation in the AA occurred between the western Munda and eastern Khasi-Aslian regions. He also suggested that the vicinity of the Bay of Bengal is the origin of the AA. George van Driem (2008, 2012) has been actively promoting this view, arguing that linguistic paleontology and the geographical center of the language family based on phylogenetic analysis are the two most important criteria for precisely locating the origin of a given language family.

2.3 Mekong River Basin

The earliest proposal that the AA family originated in central Southeast Asia came from Austrian cultural historian Heine-Geldern(1923). This hypothesis was further elaborated by Sidwell (2010). Based on the hypothesis that the region with the greatest linguistic diversity is most likely the homeland of that language family (Sapir, 1916), Sidwell reckoned that the possibility of a western or northern origin for the AA should be excluded. All three independent lines of inquiry, morphological, phonological, and lexical, have failed to provide decisive indications of nested sub-groupings among AA branches. This refutes the tripartite phylogenetic topology of the AA system proposed by Diffloth (2005). Instead, Sidwell (2008) proposed a rake-like classification framework for the AA family with 13 branches. Subsequently, Sidwell and Blench (2011) conducted a Bayesian phylogenetic analysis on a vocabulary matrix of 28 AA languages. The resulting phylogenetic network showed a strong tree signal with 12 branches clearly distinguishable. Accordingly, most of the branches of the AA are distributed approximately along the central Mekong River from southeast to northwest, suggesting that the AA dispersed and spread along this axis.

In addition, Sidwell and Blench (2011) and Blench (2015) reconstructed the early livelihood of the AA-speaking populations based on the vocabulary of modern AA languages and constructed the early history of the AA-speaking populations. The early AA-speaking populations belonged to an agricultural society along the central Mekong River, yet they also practiced fishing and gathering. They planted taro, rice, and millet crops and raised livestock such as cows, pigs, goats, dogs, and chickens. They also had access to improved types of boats. Based on this, Sidwell and Blench proposed a model of the origin and dispersal of the AA family; the Proto-AA-speaking populations spread by boat along the Mekong River and dispersed westward to the parallel basins to expand agriculture. A significant westward migration allowed the AA to develop in the southwest, which led to the origin of the Mon, Nicobaric, and Aslian languages. Later, due to the dispersal of the TK, the Sino-Tibetan (ST), and the AN languages, the geographical continuity of the AA languages was disrupted, resulting in their relative isolation in many remote areas. There was a typological shift when Monda languages came into contact with South Asian languages. However, this was limited to a single branch rather than indicative of an early two-way division in the phylum.

2.4 Middle Yangtze River or Southern China

The article “The Austroasiatics in Ancient South China: Some Lexical Evidence,” by Jerry Norman and Tsu-Lin Mei (1976), laid the foundation for the hypothesis that the AA originated from the central or southern regions of the Yangtze River. One of its main arguments is that the word “江*krung” (river) was borrowed from the AA language. It is inferred that the ancient inhabitants of the Yangtze River were of the AA ethnic group and that the Han people arrived at the north bank of the Yangtze River during the Shang and Zhou dynasties and borrowed the word *krung(river) from the AA language. The article also argues that ancient Chinese roots such as “[illegible],” “虎,” “牙,” “札,” “獲,” and “弩” have corresponding meanings and sounds in the AA language, indicating that these words were also borrowed from the AA. Jerry Norman(1983) further discovered that the Chinese words “浦” and “澳” were also borrowed from the AA language. From then on, the hypothesis on the origin of the AA family in the central Yangtze River or southern China has attracted widespread attention internationally, and it is the most supported and controversial hypothesis. Harry L. Shorto (1979) was the first to accept the opinion of the Yangtze River origin and affirmed that “tiger” and “river” are borrowed words from AA into Chinese, and there is more AA vocabulary borrowed into Tibeto-Burman languages. Pulleyblank (1983) boldly proposed that there are AA-speaking populations from northern Vietnam to eastern coastal areas of China, including Shandong. Schuessler (2007) provided more old Chinese words related to AA and believed that the ancient Yi (夷) people in the Shandong Peninsula and the Vietnamese people in the Yangtze River Delta are likely to be ancestors of the AA-speaking populations. Sagart (2008) quoted Chang, K.C. (1986) to point out that the climate in central China during the Holocene was 2° to 5° higher than it is today. This implies that tropical flora and fauna like monitor lizards, pangolins, water buffalo, civets, elephants, peacocks, rhinoceros, etc., can live in southern China. The potential area of Proto-AA tropical flora and fauna reconstructed by Diffloth (2005) cannot rule out the possibility of a central Yangtze

River or southern China origin.

Research on the origin of AA languages by Chinese linguistics is relatively scarce. Wang J.L. (1985) found a large number of correspondences between phonetic words of the TK and AA families in China and concluded that the two are genetically related. Chen B.Y. (1997) further analyzed 200 core words in the TK and AA language families in China and found that the related words resulted from contact, forming a "Mon-Tai language alliance." Wu A.Q. (2000) demonstrated the existence of phonetic and semantic correspondences between words such as "water," "cry," "blood," "you," and "swallow" in the Hmong-Mien (HM) and the Mon-Khmer languages, suggesting that the ancestors of the Mon-Khmer speakers and the HM speakers lived in adjacent regions in ancient times. The presence of these words in the HM languages may have been borrowed from the ancestral language of the Mon-Khmer speakers. Tu L.J. (2009) verified the name "Kunming (a place name in Yunnan)" as an AA word. It is etymologically related to the ethnic group names such as "Kəmu," "Khmer," and "Khmer," whose original meaning is "human." The name of the Mon people, who are Mon-Khmer language speakers, as well as other ethnic names "Mang," "Meng," and "Mang," all derived from the ethnic group names mentioned above with an abscission of the first syllable and also mean "human." Ye X.F. (2014) analyzed some characteristic words in the ancient Chu language, such as "观" (son), "邛"(mountain), "危"(sit), "淈"(stir), "篁"(bamboo grove), "党"(know), and "凭"(full). He found that these characteristic words in the ancient Chu language have no relationship with the TB, HM, and TK languages; instead, they have some correspondences in sound and meaning to AA languages. This indicates that these characteristic words in the ancient Chu language originate from the AA language. The cognates in modern TK, HM, and AA, as well as AA loanwords and substratum place names in the ancient Chu language, old Chinese language, and modern southern Chinese dialects, all suggest that AA were once widely distributed in southern China, supporting the hypothesis that the homeland of the AA family is middle Yangtze River or Southern China.

On the other hand, the hypothesis of the Yangtze River Basin or southern

China origin also faces strong criticism. Zhang H.M.(1998) used interdisciplinary evidence from linguistics, ethnology, historical literature, and geography to prove that "江"(river) is an inherent basic word in Chinese, not a loan word from AA languages. Sidwell and Blench (2010, 2011) questioned the correspondence between the sound and meaning of "江" (river) in Chinese and AA languages as a coincidence of similarity. Moreover, "虎" (tiger) is freely borrowed among TK, TB, and AA languages and cannot be used as evidence for locating the origin of the language family. Most importantly, no AA is present in the middle reaches of the Yangtze River or southern China, which does not meet the linguistic diversity criterion for a language's origin.

The above discussion shows that previous scholars have made many achievements in exploring the origin and dispersal of the AA family, but there are severe disagreements and controversies. This is mainly due to the problems in the theories, methods, and materials scholars adopt. Specifically, these problems include:

First, in terms of theory, linguistic paleobiology is not an ideal tool for locating the origin of a language family. Colin Renfrew (1987) expressed doubts about the validity and reliability of finding the origin of language using linguistic paleobiology. Moreover, based on the geographic scope of the reconstructed paleoenvironment of the AA family by Diffloth (2005), it includes three potential origin areas, i.e., ancient southern China, Southeast Asia, northeast India and the Bay of Bengal; however, it cannot give a precise conclusion about the origin of the language family. Secondly, the idea that "the region with the most linguistic diversity is most likely the origin of that language" is borrowed from the theory of the relationship between biological phenotypic diversity and the origin of species in genetics. For example, human genetic and phenotypic diversity decreases as the distance from Africa increases since the founder effect reduces diversity gradually as humans experience continuous population bottlenecks during expansion, which lays the foundation for the African origin of humans (Atkinson 2011). However, language diversity cannot be defined simply based on the number of differences. The causes for language diversity are related to population migration, differentiation and language contact among populations. Therefore, diversity cannot be directly related to the origin of language families simply based on the number of differences. Finally, the linguistic substratum theory is proposed to explain linguistic phenomena such as fusion and transfer caused by language contact that cannot be solved by comparative historical linguistics. The substratum vocabulary is often scattered and isolated and cannot be used for phylogenetic classification; thus, it cannot locate a language family's origin. The substratum vocabulary of the AA language in southern China can only prove that the AA once existed in southern China but cannot directly prove that southern China is the origin of the AA family.

Second, with regard to methodology, the confusion in the classification of AA languages is one of the reasons for the ongoing dispute over their origin. The root of the problem lies in the limitations of language classification methods. The classical comparative historical linguistics classification criteria make it difficult to determine whether the common linguistic features between languages are due to parallel evolution from a common ancestral language or the contact between languages, which have led to ongoing controversies in research results. Similarly, the contact signals conveyed by phylogenetic networks (a quantitative analysis method based on etymological statistics) are limited to only diachronic contact signals that still exist at the synchronous level. Whereas the deep contact between two language groups (such as language fusion, language conversion, etc.) cannot be included in the quantitative analysis of etymological statistics.

Third, in terms of materials, there are misunderstandings in the previous studies on the origin of AA languages. Due to the limitation of linguistic materials, the origin of AA languages is limited to the present-day AA family area, which ignores the changes in the settlements of ancient and modern AA speakers. Diffloth and Sidwill, for example, overlooked the changes in historical and cultural contexts while searching for the homeland of AA populations in Southeast Asia and the Indian subcontinent, where the present-day AA speakers are active and using the present-day AA languages to reconstruct the culture of the homeland of AA. The linguistic and cultural factors of the AA populations have accumulated in the indigenous people of southern China and the southern Han society. Thus, they cannot be ignored as they are objectively present. Although this region is no longer the settlement area of the current AA-speaking populations, it is still an important activity area or homeland of the “original” AA-speaking populations. Therefore, a lot of etymologically or substratum-related vocabulary shared between AA and TK, HM, the ancient Chu language, old Chinese, and Chinese dialects should not be excluded from the multidisciplinary practice of studying the origin of the AA family.

In summary, a purely linguistic perspective encounters a bottleneck in explaining the origin and dispersal of the AA family, and linguistics itself has not been able to provide convincing dates for the differentiation of the AA family. Archaeology and molecular anthropology can often provide a time frame for population evolution. However, research on the origin and evolutionary history of the AA family and its populations is still relatively inadequate. No academic consensus has integrated evidence from multiple disciplines.

3 Archaeological Anthropology analysis of the origin and spread of the AA family

Archaeological Anthropology utilizes material remains to reconstruct, describe, and interpret human behavior and cultural patterns. The materials studied include artifacts, material products, and animal and plant remains humans have made, used, or modified. The origins and spread of the AA language family are explored by studying shouldered stone tools, rice domestication, and the Bronze Drum Culture.

3.1 Shouldered Stone Tools and the AA-speaking Populations

Many Shouldered stone axes or segmented stone adzes have been discovered in Southeast Asian countries, India, Bangladesh, and Pacific islands. Although they belong to different cultural systems and have slight variations in form, they are similar to those found in southern China’s archaeological discoveries. Scholars often link these two archaeological cultures to the migration and cultural dispersal of indigenous populations in southern China. The distribution of shouldered stone axe and Segmented stone adze has different centers, with the former centered on the Pearl River Delta and closely related to the AA-speaking populations, while the latter is centered on the lower reaches of the Yangtze River and closely related to the AN-speaking populations. (Fu X.G. 1988)

Academic circles have two hypotheses about the origin of the shouldered stone axe, one being that it originated in Vietnam and spread to southern China (Duff, 1970), and the other being that it originated in the Pearl River Delta and spread across southern China (Zeng Q., 2015), Southeast Asia, and India. Archaeological evidence shows that the Shouldered Stone Axe first appeared in the middle Neolithic period of Guangdong, such as in the Xiqiao (西樵) Mountain site (MO Z., 1959) and the lower layer of the Jinlan (金兰) Temple site in Zengcheng(Mo Z, 1961). Fu X.G. (1988) conducted a zone-by-zone survey of all the segmented stone adze and shouldered stone axe in Guangdong province and found that the Pearl River Delta region of Guangdong was the center of Shouldered Stone Axe, while the eastern and northern regions were peripheral. During the middle Neolithic period, the shouldered stone axe spread upstream along the Xi River into southern Guangxi (such as the Geshouyan 歌寿岩 site (Jiang Y.Y., 1981)), and then arrived in Vietnam via the Zuojiang River during the Bronze Age at the Bach Ho Site (Tao. W.Y.,1959). In addition, the Shouldered Stone Axe found in India is believed to have spread to the Assam and Bay of Bengal regions from southern China via the coastal areas of Southeast Asia around 200 BC (Dani, H., 1960). It can be seen that the Shouldered Stone Axe in Southeast Asia and India was younger than that in southern China due to the dispersal of the Shouldered Stone Axe from southern China.

It is worth noting that after the segmented stone adze production technology was introduced to Guangdong in the late Neolithic period, the shouldered-and-segmented stone axes with strong local characteristics were produced. The coexistence of segmented axes and shouldered axes in Guangxi, Yunnan, and southern Guizhou (Fu X.G., 1988) indicates that the proto-AN-speaking populations and the proto-AA-speaking populations had extensive cultural contacts and integration in the mainland of southern China.

3.2 Origin of Rice Domestication and the AA-speaking Populations

Linguistic evidence suggests that the AA-language family, besides the HM-language, is the second linguistic group to have a large number of reconstructable rice farming terms (George van Driem, 2012). Both ancient HM-speaking and AA-speaking populations have been identified as the domesticators of Asian rice (Fuller, 2007). Gutaker R. M. et al. (2020) traced rice dispersal in Asia using over 1,400 landraces' whole-genome sequences, coupled with various data, revealing its origin in the Yangtze Valley 9,000 years ago, followed by diversification and rapid spread to Southeast Asia around 4,200 and 2,500 years ago, respectively. Therefore, the origin and domestication timeline of Asian rice would reflect the origin of AA-speaking populations. The international community has put forward three hypotheses about the origin of Asian rice: India, Thailand, and southern China. Based on the results of modern genetics and archaeology, the Yangtze River Basin has the oldest domesticated rice site, with two sub-sites in the lower and middle reaches of the Yangtze River (Fuller, 2015). The domestication of Japonica rice in the middle reaches of the Yangtze River was completed at about 8,000 BP. In comparison, the domestication of Japonica rice in the lower reaches of the Yangtze River has completed nearly 7,000 BP (Zheng Y.F. et al., 2016). The spread of rice cultivation from the middle and lower reaches of the Yangtze River to southern and southwestern China occurred around 4,500 BP. The earliest rice agriculture site in Guangxi, YNNDS4, with rice phytoliths, dates back to earlier than 4,500 BP (Zhang C. & Hong X.C., 2009), while the earliest agricultural site in Guizhou, JGS, unearthed rice dates back to 3,300–2,700 BP. The earliest unearthed rice remains in Yunnan date back to 4,235±150 BP. The archaeological evidence for planting rice in Southeast Asia is between 3000 BCE and 2000 BCE, which is also the result of the spread of rice cultivation from the Yangtze River Basin (Fuller, 2008). The northeastern part of India and the India-Myanmar border have been considered the origin of Asian rice domestication due to the high diversity of wild rice types (Hazarika, 2006). The middle and lower reaches of the Ganges River may represent another region where the Indica rice subtype was independently domesticated, but there is currently no evidence of rice domestication in that area earlier than 3000 BCE (Fuller, 2006). Therefore, the AA-speaking populations most likely originated from the rice domesticators in the Yangtze River Basin. Sagart (2011) believes that the proto-AA population may be a sub-group of the pre-AA population participating in early rice domestication in the Yangtze River Basin. The pre-AA people brought rice along the Xiangjiang or Yuanjiang River valleys and expanded south or southwest. After arriving in Southeast Asia, they became the ancestral population of the modern AA people and migrated to northeastern India from Southeast Asia.

3.3 The Bronze Drum Culture Circle and the AA-speaking populations

The bronze drum culture circle is a unique cultural phenomenon in southern China and Southeast Asia. Peng C, L. (2016) summarized the evolution process of the formation, development, prosperity, dissemination, and inheritance of the Bronze Drum Culture Circle. The earliest bronze drums unearthed belong to the Wanjia Ba Culture (600 B.C.) in Chuxiong, Yunnan (Li K.S., 1990). The Wanjia-Ba type bronze drums later developed into different subtypes belonging to the Shizhaishan Culture (400 B.C.) in Jinning, Yunnan, and the Dong Son Culture (500 B.C.) in northern Vietnam (Fan M.X., 1996). After the Bronze Drum Culture matured in the Dian Lake and Dong Son regions of Vietnam, it formed the early Bronze Drum Culture Circle. Then it spread to provinces in southern China such as Sichuan, Chongqing, Guizhou, Guangxi, Guangdong, and Hainan, as well as countries in Southeast Asia such as Laos, Vietnam, Thailand, Myanmar, Cambodia, Malaysia, and Indonesia (Wan F.B.

& Wei T.F., 2015).

Although the bronze drum culture originated in Yunnan, China, the creators of this culture, the Pu people, were migrants from the Yangtze River basin. The ancient Pu people have developed into the modern-day Mon-Khmer ethnic group, and the ancient Yue people have developed into the modern-day TK ethnic group. The Pu culture had already spread west of Hunan and Hubei during the Xia and Shang dynasties. According to archaeological excavations, the Pu people and Pu culture migrated southward along the Yuan River, entered Guizhou during the late Shang and early Zhou dynasties, and then entered Yunnan during the middle and late Spring and Autumn Periods, forming the Shizhaishan culture (Xi K.D. &Zhu X.S.,1987). Later, in the late Warring States period, they spread to the north of Vietnam via the Red River, forming the Dong Son culture.

According to the presence of the shouldered stone axe, the domestication of rice, and the origin, chronology, and dispersal of the Bronze Drum culture, the Yangtze River basin and southern China are the most likely homeland of the AA-speaking populations. The AA-speaking populations probably migrated from southern China to MSEA and northeastern India. In addition, the huge differences in the chronology of the three archaeological cultures may indicate the multi-layered and complex nature of the migration and dispersal events of AA-speaking populations.

4 Molecular anthropology studies the origin and dispersal of AA-speaking populations

Molecular anthropology studies the history of human evolution by analyzing population differences through genomic materials. The human genome comprises chromosomes in the cell nucleus and mitochondrial DNA in the cytoplasm. Mitochondrial DNA is maternally inherited, while the Y chromosome is paternally inherited (Li H. & Jin L., 2015). The Y chromosome haplotypes include two types of genetic markers: single nucleotide polymorphism (Y-SNP) and short tandem repeat (Y-STR). The former is considered to faithfully record the migration and evolutionary history of populations due to its low mutation rate, while the latter can record more recent events in population history, such as estimating the time of origin and divergence of populations (Kayser M. et al., 2000). It is worth noting that the linguistic classification can also be reflected in the Y chromosome DNA (Wang C.C. et al., 2013). This provides a theoretical basis for the interdisciplinary research between molecular anthropology and linguistics. Molecular anthropology and linguistics both agree on the ancient nature of the AA family. The AA-speaking population is located at the root of the phylogenetic tree of multiple Asian populations, indicating that the AA-speaking population may be one of the oldest populations in Asia (HUGO Pan-Asian SNP Consortium et al., 2009). However, similar to the debates in linguistics, the molecular anthropological studies on the South Asian, Southeast Asian, and southern East Asian AA-speaking populations and related groups have led to three different conclusions: Indian origin, MSEAn origin, and southern China origin.

4.1 Mitochondrial DNA and AA-speaking populations

Basu et al. (2003) reported a comprehensive statistical analysis of 58 DNA markers (including mitochondrial, Y-chromosomal, and autosomal markers) from a large number of ethnic groups in India. The study found that the mitochondrial DNA of the Munda-speaking populations in India exhibited very high levels of genetic diversity and clearly distinguished AA-speaking groups in India from those in Southeast Asia. This is consistent with the classification of AA languages proposed by Diffloth (2005). Additionally, an independent evaluation of the Y-STR of the Indian haplogroup O2a revealed an origin that can be traced back to 65,000 BP, leading to the hypothesis that AA language groups may have been the earliest inhabitants of India. However, this conclusion has been widely questioned due to the estimated age of the Indian O2a haplogroup being much older than its ancestor haplogroups K and NO. This discrepancy may be due to the use of Y-STR evolution mutation rates by

Basu et al.

Chaubey et al. (2011) systematically collected samples from AA-speaking populations in India and a few samples from Southeast Asia. It is found that the autosomal, Y-chromosomal, and mitochondrial DNA haplogroups of the Munda-speaking populations in India showed significant overlap with their neighboring Dravidian and Indo-European-speaking populations. Furthermore, high-resolution analysis of maternal mitochondrial DNA in the Munda-speaking populations revealed that all seven haplogroups were indigenous to India, and their mitochondrial DNA lineages did not cluster at the root (ancestral haplogroups M, N, or R) of the tree but rather were distributed among derived branches traceable to less than 10,000 BP. This suggests that the mitochondrial DNA diversity in contemporary Munda-speaking populations results from gender-biased admixture with neighboring populations in India.

Studies by Basu (2003) and Chaubey (2011) suggest that mitochondrial DNA diversity is not a reliable criterion for determining the origin of AA-speaking populations, while Y-chromosome diversity and coalescence time depth are often more informative in this regard.

4.2 The Y-chromosome haplogroups and AA-speaking populations

Kumar et al. (2007) analyzed 1222 Y-chromosome SNPs and STR data from 25 Indian populations, covering AA language groups such as the Munda, Khasi, and Mon-Khmer. The results showed that the average frequency of haplogroup O-M95 was the highest (52%). Among all three branches of AA language groups, an average of 55% of Munda-speaking populations, 41% of Khasi-speaking populations, and all 11 Nicobarese samples belonged to haplogroup O-M95. This suggests a strong patrilineal genetic connection between Indian AA populations and those in Southeast Asia. However, Kumar et al. (2007) also estimated the age of haplogroup O-M95 to be 65,000 BP, concluding that the AA-speaking populations in South Asia originated in India and spread towards Southeast Asia.

Unlike Kumar et al.'s (2007) study, the results of Chaubey et al.'s (2011) research indicate that the Y chromosome haplogroup O-M95 has significantly higher diversity and a more ancient time to the most recent common ancestor (1.7–2.8 thousand BP) in Southeast Asia. This suggests that AA-speaking populations originated in Southeast Asia during the Neolithic period and spread to South Asia, where they intermixed extensively with the local Indian populations. Similarly, in the study by Kutanan et al. (2019), analysis of 92 male Y chromosome haplogroups in 59 ethnic groups from Thailand and Laos showed that haplogroup O-M95*, with a frequency of over 70% in AA-speaking populations, accounted for 50.54% of the paternal genetic contribution to the populations in Thailand and Laos.

According to comprehensive data, the O2a1-M95 lineage dominates almost all AA-speaking populations, including those in MSEA and southern China, making it a reliable genetic marker for tracking the paternal history of AA-speaking populations. Given the sparse sampling of AA-speaking populations in MSEA and southern China, and the high-frequency of the O2a1-M95 sublineage among TK-speaking populations (with an average frequency of about 45%), Zhang et al. (2015) sampled AA-speaking and TK-speaking populations from Cambodia, Thailand, and southern China, then combined these samples with data from Chaubey et al. (2011), and estimated the age of the O2a1-M95 lineage in different populations. The results showed that the average coalescence time of the O2a1-M95 lineage in the TK-speaking populations in southern China and the AN-speaking populations in Taiwan (mainly the TK-speaking populations) was earlier than 30,000 BP, which is significantly older than the coalescence times of the lineage in MSEA (16,000BP), India (10,000BP), and Southeast Asian islands (11,000BP). Moreover, the comparison of Y-STR haplotype diversity of the O2a1-M95 lineage among different geographic populations also showed the same distribution pattern. The Y-STR haplotype diversity of the O2a1-M95 lineage was the highest in southern China populations (~0.5017), especially among the TK-speaking populations, followed by populations in MSEA (~0.3858), Southeast Asian islands (~0.3680), and finally India (~0.3168). These pieces of evidence suggest that the O2a1-M95 lineage originated in the TK-speaking populations in southern China, then expanded southward to MSEA, westward to India, and migrated to Southeast Asian islands after the Last Glacial Maximum (LGM). In addition, Singh et al. (2020) recently identified four sub-branches of the O2a-M95 lineage based on 1,437 samples from Munda-speaking populations, namely, M95x (B418, B419, B426, M1284). Their research also located the origin of the ancestral paternal populations of these founders to East Asia and Southeast Asia (>12,000 BP), from where they migrated to South Asia around 5,000 BP.

4.3 Whole Genome Sequencing and AA-speaking Populations

In recent years, the academic community has studied single-lineage genetic markers and begun sequencing the entire genomes of different populations. Liu et al. (2020) analyzed genome-wide single-nucleotide polymorphism data from the Kinh ethnic group in Vietnam and 21 other populations. The results showed that genetic diversity in Vietnam is not largely indigenous. They found that Vietnamese ethnolinguistic groups harbor multiple sources of genetic diversity likely to reflect different sources of the ancestry associated with each language family. This suggests that linguistic diversity does not fully match genetic diversity among populations; for instance, there is broad interaction between the HM and TK-speaking populations, while different AA-speaking groups show varying degrees of affinity with other language groups. The study also found AA-speaking populations in Vietnam shifting to use AN languages over the past 2500 years.

Kutanan et al. (2021) reported the whole-genome SNP data of 452 samples from 33 populations of five different linguistic families from Thailand and Laos. They compared the data with modern Asian populations and ancient Southeast Asian samples. The study found that local genetic structure was driven by varying levels of interaction with other populations in the same geographic region, for example, the TK-speaking populations. Some Thai populations showed genetic admixture with

AA-speaking populations, likely due to trade networks. It is noteworthy that the AN-speaking populations in southern Thailand not only showed admixture with South Asians but also exhibited overall genetic affinities with the AA-speaking populations.

While whole-genome studies have not provided clues to the origin and migration routes of AA-speaking populations, they have revealed admixture events between AA and HM, AN populations during their migration and dispersal in MSEA.

4.4 Ancient DNA and AA-speaking populations

In recent years, the study of ancient DNA has also provided some potential evidence for the origin and dispersal of different populations and their languages. McColl et al. (2018) sequenced 26 ancient human genomes (25 from MSEA and one from the Japanese Jomon period). They found that the previously held views on agricultural development in the sampled regions were inaccurate. Some scholars believed that the agricultural development in these regions was indigenous. In contrast, others supported the “dual-layer” hypothesis, which posits that the expansion of farmers to the south caused the genetic diversity in Southeast Asia today (Jinam, T. A., et al. 2017). However, neither of these explanations accurately captures the complexity of the evolutionary history of Southeast Asian populations. In reality, both hunting-gathering populations and agricultural populations from East Asia have contributed to the genetic diversity of present-day Southeast Asia.

Lipson, M et al. (2018) conducted a study on the complete genome data of 18 individuals from five ancient sites in Southeast Asia, spanning from the Neolithic to the Iron Age (4100 to 1700 BP). They observed at least two waves of migration from southern China to MSEA, one during the Neolithic period and the other during the Bronze Age. Principal component analysis showed that these ancient individuals were genetically close to modern Chinese and Vietnamese populations. Moreover, individuals from Man Bac (Neolithic; 4100–3600 BP) in Vietnam, Ban Chiang (late Neolithic to Iron Age; 3500–2400 BP) in Thailand, and Vat Komnou (Iron Age; 1900–1700 BP) in Cambodia clustered together with modern AA-speaking populations and moved slightly towards the direction of the Andamanese and Papuan populations. This suggests that the Southeast Asian farmers during the Neolithic period were a mixture of southern Chinese agricultural populations and early Southeast Asian hunter-gatherers genetically similar to the Andamanese and Papuan populations. Individuals from Oakaie in Myanmar (Neolithic to late Bronze Age; 3200–2700 BP) were genetically closer to the Burmese and other Sino-Tibetan populations, indicating that from the late Neolithic to Bronze Age individuals from Oakaie did not have genetic introgression from ancestor of AA-speaking populations and were more closely related to Sino-Tibetan populations. Meanwhile, individuals from Nui Nap in Vietnam (Bronze Age; 2100–1900 BP) were genetically closer to the Kinh and TK-speaking populations, suggesting a close genetic relationship between the TK and Vietnam populations during the Bronze Age. Additionally, by evaluating the relationship between these ancient and modern populations using admixture models, the authors identified two genetic admixture events outside Mainland Southeast Asia, one involving admixture between Western Eurasian, Eastern Eurasian (South Asian aboriginal ancestry), and AA-speaking populations among the AA speakers in eastern India, and the other involving admixture between AN, AA, and Papuan populations in the AN populations in western Indonesia.

The study by Zhang X. et al. (2020) analyzed the complete mitochondrial genomes of 41 human remains from 13 hanging-coffin sites in southern China and northern Thailand dating from 2500 to 660 BP. The research found a genetic connection between populations with hanging-coffin customs in different geographical regions. Notably, the maternal genetic diversity of hanging-coffin populations in southern China is much higher than in northern Thailand, which is consistent with the theory of a single origin of the hanging-coffin custom in southern China around 3600 BP. This imply that the people with hanging-coffin customs spread from southern China to Thailand.

In summary, research on the AA-speaking populations' origin, dispersal routes, and divergence time have been constantly revised as the spatiotemporal range of genomic sampling and sample size expands, ultimately supporting the hypothesis that the AA-speaking populations originated from southern China. Molecular anthropology research has broken through the limitation of population and language affiliation, using the same Y-chromosome haplotype genetic markers to detect the relationship and divergence time of different linguistic populations also has inspiring implications for historical-comparative linguistics. As historical-comparative linguistics advances to the stage of superfamily construction, it indicates that research on the “homogeneity” of languages will shift to the detection of “relatedness” between language families (Su N., &Wang C.C.,2020).

5 CONCLUSIONS

In the three-dimensional perspective of language, archaeology, and genetics, the origin and dispersal of the AA language are no longer limited to simple historical linguistic comparisons. The phylogenetic topology of the AA language family, the chronology of linguistic and ethnic group differentiation, the reconstruction of ancient living environments, archaeological sites, etc., should be seamlessly integrated into a network with the origin and dispersal of the AA-speaking populations. By combining the evidence from the vocabularies of indigenous languages in southern China, archaeological materials such as Shouldered axes, rice domestication, and bronze drums, as well as genomic data, the origin of the AA-speaking populations can be traced back to southern China. In addition, based on the “allied” relationship between AA and TK languages (Chen B.Y., 1994), the coexistence and fusion of the Shouldered stone axe and Segmented stone adze in southern China (Fu X.G., 1988), the profound relationship between the Pu-Yue people and the Bronze Drum culture (Wan F.B. et al. 2015), and genomic evidence that the present-day TK-speaking population in southern China is the primary ancestral source of the AA-speaking populations (Zhang X.M. et al. 2015 & 2020, Lipson 2018), it is reasonable to infer that the ancestors of the Proto-AA and Proto-TK-speaking populations had an extensive genetic exchange and language contact in southern China, which led to language transformation and ultimately the “disappearance” of AA languages in southern China.

The ancestors of the AA populations were most likely agricultural populations living in the southern region of China during the Neolithic period. Around 7,000 BP, some original AA-speaking populations in the middle and lower reaches of the Yangtze River first domesticated rice and transmitted it to other AA-speaking populations in South China. Around 4,000–5,000 BP, the original AA populations (BaiPu) in the middle reaches of the Yangtze River were influenced by the expansion of the northern Han group and Chu culture, hence migrated southward or southwestward along the river valleys in southern China. Recently published genomic research on Tibeto-Burman and AA populations by Guo Jianxin et al. (2021) has shown new progress. The study suggests that the ancient "TK and AN" populations spread along two routes, one along the middle and lower reaches of the Yangtze River to the southwestern hinterland and the other along the southeastern coast to northern Vietnam. The dispersal and pathway of this population probably further promoted the spread of the AA-speaking population and, at the same time, led to the interweaving of the migration history of the two populations and even the occurrence of genetic mixing and language conversion.

During the late Neolithic period in southern China, the pre-AN-speaking populations in the Yangtze River Delta had cultural contact and fusion with the pre-AA-speaking populations in the Pearl River Delta. Around 3,000–4,500 BP, some of the ancestral AA speakers migrated southward to northern Vietnam with shouldered stone axes and domesticated rice, where they mixed with local hunter-gatherers and spread southward and southwestward along the river valleys of MSEA, developing related populations such as the Mon-Khmer, Aslian, and Nicobarese people. During their westward migration, they left behind the Khmuic and Khasi populations in Myanmar and Bangladesh. They ultimately reached the Northeast Indian region, mixing with local indigenous populations to form the Munda-speaking populations. The AA-speaking populations that arrived in southern Vietnam, southern Thailand, and Indonesia mixed with the AN-speaking populations and underwent language shifts. Another group of AA-speaking people migrated through Guizhou about 3,000

BP and arrived near Lake Dian in Yunnan about 2,500 BP, where they created the Bronze Drum Culture together with the incoming pre-TK-speaking populations. This also laid the foundation for the distribution of AA languages in Yunnan. About 2,300 BP, the AA-speaking populations in Yunnan, with the bronze drum culture, spread southeastward through Guangxi to northern Vietnam, leaving remnants of the AA languages in Guangxi. Finally, the pre-AA-speaking ancestors who remained in southern China continued to intermix with the pre-TK-speaking ancestors from Fujian and Guangdong and underwent language shifts. This was one of the main reasons for the "disappearance" of AA-speaking populations in southern China.

Received 27 May 2023; accepted for publication 6 March 2024.

Conflicts of interest

The authors declare no conflict of interest.

Acknowledgments

This study was supported by the Nanqiang Outstanding Young Talents Program of Xiamen University(X2123302), the Major project of the National Social Science Foundation of China(20&ZD248), the Major project of the National Social Science Foundation of China(21&ZD285), the National Social Science Foundation of China (22CYY049).

References

Atkinson, Q. D. (2011). Phonemic Diversity Supports a Serial Founder Effect Model of Language Expansion from Africa. *Science*, 332(6027), 346–349.

Ban, T. (2009). Substrate theory and the relationship between Chinese and Tai languages. *Journal of Guangdong University of Foreign Studies* (in Chinese), 5, 92–96.

Basu, Analabha et al. (2003). Ethnic India: a genomic view, with special reference to peopling and structure. *Genome Research*, 13(10), 2277–2290.

Blench, Roger. (2015). Reconstructing Austroasiatic prehistory. In Jenny, M., & Sidwell, P. (Eds.), *The Handbook of Austroasiatic Languages*. The Netherlands: Brill.

Blust, R. A. (1996). Beyond the Austronesian homeland: The Austric hypothesis and its implications for archaeology. In W. H. Goodenough (Ed.), *Prehistoric Settlement of the Pacific* (pp. 117–160). Transactions of the American Philosophical Society, 86(5).

Blust, R. A. (2013). *The Austronesian Languages*. Canberra: Pacific Linguistics.

Cai X, Qin Z, Wen B, et al. (2011). Human migration through bottlenecks from Southeast Asia into East Asia during Last Glacial Maximum revealed by Y chromosomes. *PloS one*, 6(8), e24282.

Chang, K. C. (1986). *The Archaeology of Ancient China*. Yale University Press.

Chaubey, G.et al. (2011). Population genetic structure in Indian Austroasiatic speakers: the role of landscape barriers and sex-specific admixture. *Molecular Biology and Evolution*, 28(2), 1013–1024.

Chen, B. Y. (1997). The etymological relationship between Dongtai languages and South Asian languages—on the relationship between the ancient Yue and Pu ethnic origins. *Journal of Yunnan Nationalities University* (Social Science Edition) (in Chinese), 1.

Coupe, Alexander. (2021). Linguistic diversity and language contact in Nagaland. In Critical Conversations in Naga Society, ed. Yanbeni Yanthan & Hewasa Lorin Khing. Heritage Publishing House, p. 9–32.

Cysouw, M., Dediu, D., Moran, S., & Bickel, B. (2012). Comment on Phonemic Diversity Supports a Serial Founder Effect Model of Language Expansion from Africa. *Science*, 335(6069), 657.

Dani, H. (1961). Prehistory and Protohistory of Eastern India. *Bulletin of the School of Oriental and African Studies*, 24(3), 596–598.

Dao, J. (2018). Distribution Overview of Chinese South Asian Languages. *Journal of Honghe University*, 6, 126–132.

Deng, X. H., & Gao, T. J. (2014). Theory, methodology, and practice of evolutionary linguistics. *Journal of Shanxi University* (Philosophy and Social Sciences Edition) (in Chinese), 2, 66–73.

Diffloth, G. (1974). Austro-Asiatic languages. In M. Adler (Ed.), *Encyclopaedia Britannica* (pp. 480–484). Chicago, IL/London/Toronto/Geneva: Encyclopaedia Britannica Inc.

Diffloth, G. (1994). The lexical evidence for Austric. *Oceanic Linguistics*, 33(2).

Diffloth, G. (2005). The contribution of linguistic palaeontology to the homeland of Austroasiatic. In L. Sagart, R. Blench, & A. Sanchez-Mazas (Eds.), *The Peopling of East Asia: Putting Together the Archaeology, Linguistics and Genetics* (pp. 77–80). London: Routledge Curzon.

Diffloth, G. (2018). Austroasiatic languages. *Encyclopedia Britannica*.

Donegan, P. J., & Stampe, D. (2004). Rhythm and the Synthetic Drift of Munda. In R. Singh (Ed.), *The Yearbook of South Asian Languages and Linguistics* (pp. 3–36).

Driem, G. van. (2008). The Shompen of Great Nicobar Island: new linguistic and genetic data, and the Austroasiatic homeland revisited. *Mother Tongue*, (13), 227–247.

Duff, R. (1970). Stone adzes of Southeast Asia: an illustrated typology (No. 3). Christchurch: Canterbury Museum Trust Board.

Fan, Z. Q., Deng, X. H., & Wang, C. C. (2018). Language and genes: On the origin and dispersal of the Austronesian language family. *Academic Monthly* (in Chinese), 10, 90–96.

Fu, X. G. (1988). On Youduan stone adzes and shouldered stone implements. *Archaeology* (in Chinese), (1), 23–32.

Fuller, D. Q. (2006). Agricultural origins and frontiers in South Asia: A working synthesis. *Journal of World Prehistory*, 20(1), 1–86.

Fuller, D. Q., Harvey, E., & Qin, L. (2007). Presumed domestication? Evidence for wild rice cultivation and domestication in the fifth millennium BC of the Lower Yangtze region. *Antiquity*, 81(312), 316–331.

Fuller, D., Qin, L., Zheng, Y., Zhao, Z., Chen, X., Hosoya, L. A., Sun, G., & Li, J. (2008). Japonica rice carried to, not from, Southeast Asia. *Nature Genetics*, 40(11), 1264–1265.

Guo, J. X., et al. (2021). Genomic insights into Neolithic farming-related migrations in the junction of east and southeast Asia. *American Journal of Biological Anthropology*, 33(1), 1–15.

Gutaker, R. M., et al. (2020). Genomic history and ecology of the geographic spread of rice. *Nature plants*, *6*(5), 492–502.

Hazarika, M. (2006). Neolithic culture of Northeast India: A recent perspective on the origins of pottery and agriculture. *Ancient Asia*, 1, 25–43.

Heine-Geldern, Robert von. (1923). Südostasien. In Georg Buschan (Ed.), *Illustrierte Völkerkunde*. Stuttgart: Strecker und Schröder.

Headley, R.K. Jr. (1976). Some Considerations on the classification of Khmer. In P.N. Jenner, L.C. Thompson, & S. Starosta (Eds.), *Austroasiatic Studies* (pp. 431–451). Honolulu, HI: The University of Hawaii Press.

Thomas, D. (1966). Mon-Khmer subgroupings in Vietnam. In N.H. Zide (Ed.), *Studies in Comparative Austroasiatic Linguistics* (pp. 194–202). The Hague: Mouton.

HUGO Pan-Asian SNP Consortium, Abdulla, M. A., Ahmed, I., Assawamakin, A., Bhak, J., Brahmachari, S. K. & Zilfalil, B. A. (2009). Mapping human genetic diversity in Asia. *Science*, 326(5959), 1541–1545.

Jiang, T. Y. (1981). *An outline of the archaeology of the Neolithic period in Guangxi*. Beijing: Wenwu Press. (in Chinese)

Jinam, T. A., et al. (2017). Discerning the origins of the Negritos, first Sundaland people: deep divergence and archaic admixture. *Genome Biology and Evolution*, 9(8), 2013–2022.

Kayser, M., et al. (2000). Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. *American Journal of Human Genetics*, 66(5), 1580–1588.

Kottak, C. P. (2005). *Window on Humanity: A Concise Introduction to General Anthropology*. McGraw-Hill Higher Education.

Kumar, V., Reddy, A. N., Babu, J. P., Rao, T. N., Langstieh, B. T., Thangaraj, K., & Reddy, B. M. (2007). Y-chromosome evidence suggests a common paternal heritage of Austro-Asiatic populations. *BMC Evolutionary Biology*, 7(1), 1–14.

Kutanan, W., Kampuansai, J., Srikummool, M., Brunelli, A., Ghirotto, S., Arias, L., & Stoneking, M. (2019). Contrasting paternal and maternal genetic histories of Thai and Lao populations. *Molecular Biology and Evolution*, 36(7), 1490–1506.

Kutanan, et al. (2021). Reconstructing the human genetic history of mainland Southeast Asia: insights from genome-wide data from Thailand and Laos. *bioRxiv*.

Li, H., & Jin, L. (2015). *Y Chromosome and East Asian Populations Evolution.* Shanghai: Shanghai Scientific and Technical Press. (in Chinese)

Li, K. S., & Huang, D. R. (1990). On the Wanjiaba type of Dong Son drums. *KaoGu* (in Chinese), 5, 427–433.

Li, Y. (2016). New Progress in Research on the Genetic Relationships of South Asian Languages. *Contemporary Linguistics*, 1, 76–86.

Lipson, M., et al. (2018). Ancient genomes document multiple waves of migration in Southeast Asian prehistory. *Science*, 361(6397), 92–95.

Liu, D., Duong, N. T., Ton, N. D., Van Phong, N., Pakendorf, B., Van Hai, N., & Stoneking, M. (2020). Extensive Ethnolinguistic Diversity in Vietnam Reflects Multiple Sources of Genetic Diversity. *Molecular biology and evolution*, 37(9), 2503–2519.

Logan, J. R. (1854). Ethnology of the Indo-Pacific Islands. *Journal of the Indian Archipelago and Eastern Asia*, 8, 28–200.

Maspero, H. (1929). Langues. In G. Maspero (Ed.), *Un empire colonial francais: L'Indochine*, Vol. 1 (pp. 63–82). Paris:Bruxelles: Les éditions G. van Oest.

McColl H, et al. (2018). The prehistoric peopling of Southeast Asia. *Science*, 361(6397), 88–92.

Mo, Z. (1959). Stone Tools Unearthed from Xiqiaoshan, Nanhai, Guangdong. *Kaogu Xuebao* (in Chinese), 4.

Mo, Z. (1961). New discoveries from archaeological surveys and excavations in Guangdong. *KaoGu* (in Chinese), 2.

Norman, J. & Mei, T.-L. (1976). The Austroasiatics in Ancient South China: Some Lexical Evidence. *Monumenta Serica*, 32(1), 274–301.

Norman, J. (1983). Ancient dialectal characters in Min dialects. *Fangyan* (in Chinese), 3, 159–165.

Osada, T. (1995). The Rice and Food Culture of Munda in Eastern India: An Ethnolinguistic Study. In (Ed.), *Mundajin no Nōkō Bunka to Shokuji Bunka: Minzoku Gengogakuteki Kōsatsu*. Kyoto: Kokusai Nihon Bunka Kenkyū Sentā.

Peng, C. L. (2016). The evolution of the Dong Son culture area. *Guangxi Ethnic Studies* (in Chinese), 1, 54–59.

Pham, M. H. (1996). *Dong Son culture: Unity and diversity*. (W. H. Nguyen, Trans.). Hanoi: Vietnam Social Sciences Publisher.

Pinnow, H.J. (1959). *Versuch einer historischen Lautlehre der Kharia-Sprache*. Wiesbaden: Otto Harrassowitz.

Pulleyblank, E. G. (1983). The Chinese and their neighbours in prehistoric and early historic times. In Keightley, D. N. (Ed.), *The Origins of Chinese Civilization* (pp. 411–466). Berkeley: University of California Press.

Renfrew, C. (1987). *Archaeology and Language: The Puzzle of Indo-European Origins*. London: Jonathan Cape. pp. 97–98.

Reid, L. (1994). Morphological evidence for the Austric. *Oceanic Linguistics*, 33(2).

Sapir, E. (1916). Time perspective in aboriginal American culture: A study in method. *Journal of American Folklore*, 29(114), 233–264.

Sagart, L. (2008). The expansion of Setaria farmers in East Asia: a linguistic and archaeological model. In A. Sanchez-Mazas, R. M. Blench, M. D. Ross, I. Peiros, & M. Lin (Eds.), *Human migrations in continental East Asia and*

Taiwan. Matching archaeology, linguistics and genetics (pp. 133–158). London: Routledge.

Sagart, L. (2011). The Austroasiatics: East to West or West to East? In Enfield, N.J. (Eds.) *Dynamics of human diversity: The case of mainland Southeast Asia*. (pp. 345–361). Canberra: Pacific Linguistics.

Schmidt, P. W. (1906). Die Mon-Khmer-Völker: ein Bindeglied zwischen Völkern Zentralasiens und Austronesiens. *Archiv für Anthropologie*, 5, 59–109.

Schmidt, P. W. (1907–1908). Les peuples Mon-Khmer, trait d' union entre les peoples de l' Asie central et del' Austronesie. *BEFEO*, 7, 213–250, 8, 1–35.

Schuessler, A. (2007). *ABC Etymological Dictionary of Old Chinese*. Honolulu: University of Hawaii Press.

Shorto, H. L. (1979). The Linguistic Protohistory of Mainland South East Asia. In R. B. Smith & W. Watson (Eds.), *Early South East Asia* (pp. 273–278). Oxford: Oxford University Press.

Sidwell, P. (2008). Is Mon-Khmer dead? Long live Austroasiatic! Paper read at the 18th Annual Meeting of the Southeast Asian Linguistics Society, Bangi: Universiti Kebangsaan Malaysia.

Sidwell, P. (2010). The Austroasiatic central riverine hypothesis. *Journal of Language Relationship, Вопросы языкового родства*, 4, 117–134.

Sidwell, P., & Blench, R. M. (2011). The Austroasiatic Urheimat: The Southeastern riverine hypothesis. In N. Enfield (Ed.), *Dynamics of human diversity in Mainland SE Asia* (pp. 317–345). Canberra: Pacific Linguistics.

Singh, P. P., Vishwakarma, S., Sultana, G. N. N., Pilvar, A., Karmin, M., Rootsi, S., & Chaubey, G. (2021). Dissecting the paternal founders of Mundari (Austroasiatic) speakers associated with the language dispersal in South Asia. *European journal of human genetics*, 29(3), 528–532.

Sun, N., & Wang, C.C. (2020). The dispersal of language from a three-dimensional perspective. *China Social Science News* (in Chinese), p. 3.

Tao, W. Y. (1959). *Ancient History of Vietnam* (in Chinese). Beijing: Kexue Press.

Tu, L. J. (2009). A study on the origin of the name "Kunming." *Yunnan Normal*

University Journal (Philosophy and Social Sciences Edition) (in Chinese), 6.

Van Driem, G. (2012). The ethnolinguistic identity of the domesticators of Asian rice. *Comptes Rendus Palevol*, 11(2–3), 117–132.

Wan, F. B., & Wei, D. F. (2015). On the Circle of Bronze Drum Culture. *Guangxi Ethnic Studies* (in Chinese), (1), 118–125.

Wang, C. C., Wang, L. X., Shrestha, R., Zhang, M., Huang, X. L., Hu, K. Q., Jin, L., Li, H. (2013). Inferring Human History in East Asia from Y Chromosomes. *Investigative Genetics*, 4(1).

Wang, J., & Chen, X. (1982). On the Cognate Relationship of "Village," "Surname," and "Home" in the Mon-Khmer and Tai-Kadai Languages. *Minzu Yuwen* (in Chinese), (03), 12–24.

Wu, A. Q. (2000). The dissemination of ancient South Asian culture. *Shijie Minzu* (World Ethno-National Studies) (in Chinese), 2.

Wu, C. M., & Cao, J. (2005). Four misconceptions in the research on the origin of the Austronesian language family. *Journal of Xiamen University* (Philosophy and Social Sciences Edition) (in Chinese), 3, 110–115.

Xi, K. D., & Zhu, X. S. (1987). A Report on the Excavation of Bi Jie Wao Yao Site, Guizhou. *KaoGu* (in Chinese), (4).

Xi, K. D., & Zhu, X. S. (1987). Excavation report on the Wayoutan site in Bijie, Guizhou. *Archaeology*, 4, 372–392.

Xiao, M. H. (2001). A brief account of Yunnan archaeology. *KaoGu* (in Chinese), 12, 1–10.

Xu, J. X., & Li, H. (2017). Molecular anthropology perspective: a new discussion on the origin of Guangxi's indigenous ethnic groups. *Journal of Guangxi Normal University* (Philosophy and Social Sciences Edition) (in Chinese), 4, 85–91.

Ye, X. F. (2014). South Austroasiatic components in ancient Chu language. *Minzu Yuwen* (in Chinese), 3.

Zeng, Q. (2015). On the Issues Related to the Segmented Stone, Double-shouldered Stone Tools, and "Geometric Pattern Impressed Pottery." In Y. Xibing (Ed.), *Archaeological Research on the Xiqiaoshan Site*. Guilin: Guangxi Normal

University Press.

Zhang, C., & Hong, X. C. (2009). The emergence of agriculture in South China and Southwest China and related issues. *Nanfang Wenwu* (in Chinese), 3, 65–74.

Zhang, H. R., Wang, L., & Luo, E. H. (2006). Brief report on the excavation of Jigongshan site in Weining County, Guizhou in 2004. *KaoGu* (in Chinese), 8, 78–85.

Zhang, H.M. (1998). Chinese Etyma for River. *Journal of Chinese Linguistics*, 26, 1–47.

Zhang, X., et al. An updated phylogeny of the human Y-chromosome lineage O2a-M95 with novel SNPs. *PLoS One*, 2014, 9(9), pp.e108691.

Zhang, X.M., et al. Y-chromosome diversity suggests southern origin and Paleolithic backwave migration of Austro-Asiatic speakers from eastern Asia to the Indian subcontinent. *Sci Rep*, 2015, 5, pp.15486.

Zhang, X.M., et al. (2020). A Matrilineal Genetic Perspective of Hanging Coffin Custom in Southern China and Northern Thailand. *iScience*, 23(4).

Zheng, Y. F., Jiang, L. P., & Crawford, G. W. (2016). Rice grain shattering and the origin and domestication of rice in the lower Yangtze River Valley. *Nanfang Wenwu* (in Chinese), 3, 103–111.