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

Authors: Yu, Jian, Deng, Xiaohua, Fan, Zhiquan, Yang, Wenjiao, Ji, Zhi, et al.

Source: Human Biology, 94(3) : 141-154

Published By: Wayne State University Press

URL: https://doi.org/10.1353/hub.2017.a925564

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]*

ABSTRACT

The Austroasiatic (AA) languages comprise a large language family in mainland Southeast Asia (MSEA) 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 key to solving this problem. We review the latest hypotheses in linguistics, archaeology, and molecular anthropology and propose insights on the origin and dispersal of AA languages. The ancestors of the AA-speaking populations were suggested to be rice farmers living in the Neolithic Age in southern China. Between 3,000 and 4,500 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 MSEA, giving rise to the Mon-Khmer, Aslian, and Nicobarese populations. They also spread to the northeast of India to form the Munda-speaking populations. Another group arrived near Dian Lake in Yunnan about 2,500 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 hypothesize was one of the main reasons for the "disappearance" of AA in southern China.

The Austroasiatic (AA) language family consists of over 150 languages and dialects, with a total population of over 65 million speakers. It is widely distributed throughout mainland Southeast Asia (MSEA), Northeast India, and the Nicobar Islands (Diffloth 2024). The 12 AA languages in China are distributed mainly in Yunnan and Guangxi Provinces (Dao 2018).

The AA family, the oldest language family in MSEA and Northeast India, is considered the underlying structure of all languages in Southeast Asia (Y. Li 2016), making this language family critically important. However, the history of AA studies is relatively short, with the earliest research tracing back to Logan's *The Ethnology of the Indian Archipelago* in 1852 (see Logan 1845). Not until the early 20th century did German anthropologist Wilhelm Schmidt define the AA family. After that, research on AA was mired in a long-standing debate over its internal relationships and subgrouping. Schmidt

---

[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

KEYWORDS: language, archaeology, genetics, Austroasiatic-speaking populations, evolution.

Human Biology, Summer 2022, v. 94, no. 3, pp. 141–154. Copyright © 2024 Wayne State University Press, Detroit, Michigan 48201

(1907–1908), Maspero (1929), Pinnow (1959), Thomas (1966), Headley (1976), Diffloth (1974, 2005), Sidwell (2010), and others have proposed different classification schemes based on different types of word lists and research methods.

Given the chaotic classification of AA languages, the Ethnologue website (https://www.ethnologue.com/subgroup/231/) divides the AA languages into two major branches: the Mon-Khmer branch in MSEA and the Munda branch in eastern India. Meanwhile, confusion in the classification of AA languages has seriously hindered determination of the original homeland of Proto-AA (Y. Li 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 Tai-Kadai (TK) and Austronesian (AN) populations in southern China and southeastern Asia. Resolving issues of the origin and dispersal of AA-speaking populations is crucial to understanding the evolutionary relationship among these three language families. To this end, in this article we attempt 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.

Hypotheses on the Origin and Dispersal of the AA Language Family

Four main homelands have been hypothesized for the AA family: (a) northern Myanmar, (b) northeastern India or the Bay of Bengal, (c) the Mekong River basin in Southeast Asia, and (d) the middle Yangtze River or southern China.

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 8,000 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.

Northeastern India or the Bay of Bengal

American scholars Donegan and Stampe (2004) believe that the AA languages 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. They proposed that AA originated from the Indian subcontinent and spread eastward to Southeast Asia. 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 such animals as 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 must be in tropical regions. However, the overlapping area based on reproducible fauna remains large, specifically covering Northeast India, the India-Myanmar border, and Myanmar. Based on this, Diffloth (2005) proposed that the deepest language differentiation in 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 AA. George van Driem (2008, 2012) has actively promoted 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 its origin.

Mekong River Basin

The earliest proposal that the AA family originated in central Southeast Asia came from Austrian cultural historian Yon Heine-Geldern (1917), a hypothesis 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 AA should be excluded. All three independent lines of inquiry, morphological, phonological, and lexical, have failed to provide decisive indications of nested subgroupings 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. Most of the branches of the AA languages 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 (2011) 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 AA to develop to the southwest, which led to the origin of the Mon, Nicobaric, and Aslian languages. Later, due to the dispersal of TK, Sino-Tibetan, and 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, but this was limited to a single branch and does not indicate an early two-way division in the phylum.

Middle Yangtze River or Southern China

Norman and Mei (1976) laid the foundation for the hypothesis that AA originated from the central or southern regions of the Yangtze River. One of their main arguments is that the word “*krung” 江 (river) was borrowed from AA. They infer 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” from AA. They also argue that ancient Chinese roots such as 蠵, 虎, 牙, 札, 獲, and 弩 have corresponding meanings and sounds in AA languages, indicating they were also borrowed from AA. Jerry Norman (1983) further discovered that the Chinese words 浦 and 澳 were also borrowed from AA. 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 controversial hypothesis. 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 more AA vocabulary was borrowed into Tibeto-Burman languages. Pulleyblank (1983) boldly proposed 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 Yue 越 people in the Yangtze River Delta are likely to be ancestors of the AA-speaking populations. Sagart (2008) quoted Chang (1986) to point out that the climate in central China during the Holocene was 2–5°C warmer than it is today. This implies that such tropical flora and fauna as monitor lizards, pangolins, water buffalo, civets, elephants, peacocks, and rhinoceros could 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 and (1982) found a several correspondences between phonetic words of the TK and AA language families in China and concluded that the two are genetically related. Chen (1997) 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." A. Q. Wu (2000) demonstrated the existence of phonetic and semantic correspondences between such words as "water," "cry," "blood," "you," and "swallow" in the Hmong-Mien (HM) and Mon-Khmer languages, suggesting that the ancestors of Mon-Khmer and HM speakers lived in adjacent regions in ancient times. These words in HM languages may have been borrowed from the ancestral language of the Mon-Khmer speakers. Tu (2009) verified the name "Kunming" ("a place name in Yunnan") as an AA word. It is etymologically related to ethnic group names such as "Kəmu," "Khmer," and "Khmer," whose original meaning is "human." The name of the Mon people, who speak Mon-Khmer languages, as well as ethnic names "Mang," "Meng," and "Mang," all derive from the ethnic group names mentioned above with an abscission of the first syllable, and they also mean "human." Ye (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 Tibeto-Burman (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. H. M. Zhang (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 languages are present in the middle reaches of the Yangtze River or in southern China, so the linguistic diversity criterion for a language's origin is not met.

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

First, in terms of theory, linguistic paleobiology is not an ideal tool for locating the origin of a language family. Renfrew (1987: 97–98) expressed doubts about the validity and reliability of finding the origin of language using linguistic paleobiology. Moreover, the geographic scope of the reconstructed paleoenvironment of the AA family by Diffloth (2005) includes three potential origin areas: 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. In addition, the idea that the region with the most linguistic diversity is most likely the origin of that language is borrowed from genetics, specifically, the theory of the relationship between biological phenotypic diversity and the origin of species. 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; this lays the foundation for the African origin of humans (Atkinson 2011). However, language diversity cannot be defined simply based on the number of differences (Cysouw et al. 2012). 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 (Fan et al. 2018). 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 (Ban 2009). 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 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, confusion in the classification of AA languages is one reason 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 due to contact between languages (Deng and Gao 2014), which has led to ongoing controversies in interpreting 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. In contrast, deep contact between two language groups (e.g., language fusion and language conversion) cannot be included in the quantitative analysis of etymological statistics.

Third, due to limitations of linguistic materials, the origin of AA languages is limited to the present-day AA family, which ignores changes in settlements of ancient and modern AA speakers (Wu and Cao 2005). Diffloth (2005) and Sidwell (2010), for example, overlooked 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 AA homeland. 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 among 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 language family and its populations is still incomplete. No academic consensus has integrated evidence from multiple disciplines.

Archaeological Anthropological Analyses 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 made, used, or modified by humans. The origins and spread of the AA language family are explored by studying shouldered stone tools, rice domestication, and the Bronze Drum culture.

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. Scholars often link these two archaeological cultures to the migration and cultural dispersal of indigenous populations in southern China. The distributions of shouldered stone axes and segmented stone adzes have different centers, respectively, the Pearl River Delta, closely related to the AA-speaking populations; and the lower reaches of the Yangtze River, closely related to the AN-speaking populations (Fu 1988).

Two hypotheses on the origin of the shouldered stone axe are that it originated in Vietnam and spread to southern China (Duff 1970), and that it originated in the Pearl River Delta and spread across southern China (Zeng 1983), Southeast Asia, and India. Archaeological evidence shows that the shouldered stone axe first appeared in the middle Neolithic period in Guangdong, such as at the Xiqiao 西樵 Mountain site (Mo 1959) and in the lower layer of the Jinlan 金兰 Temple site in Zengcheng (Mo 1961). Fu's (1988) zone-by-zone survey of all the segmented stone adzes and shouldered stone axes in Guangdong Province found that the Pearl River Delta region of Guangdong was the center of the 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 1981) and then arrived in Vietnam via the Zuojiang River during the Bronze Age at the Bach Ho Site (Tao 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 BCE (Dani 1955). Thus, the shouldered stone axes in Southeast Asia and India were younger than those in southern China due to the dispersal of the shouldered stone axe from southern China.

After the segmented stone adze was introduced to Guangdong in the late Neolithic period, 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 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.

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 (van Driem 2012). Both ancient HM-speaking and AA-speaking populations have been identified as the domesticators of Asian rice (Fuller et al. 2007). Gutaker 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 BP, followed by diversification and rapid spread to Southeast Asia around 4,200 and 2,500 BP, 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 subsites in the lower and middle reaches of the Yangtze River (Fuller et al. 2007). The domestication of Japonica rice in the middle reaches of the Yangtze River was completed about 8,000 BP, and in the lower reaches of the Yangtze River, nearly 7,000 BP (Zheng et al. 2016). Rice cultivation spread from the middle and lower reaches of the Yangtze River to southern and southwestern China around 4,500 BP. The earliest rice agriculture site in Guangxi, YNNDS4, with rice phytoliths, dates back to earlier than 4,500 BP (C. Zhang and Hong 2009), while at the earliest agricultural site in Guizhou, JGS, unearthed rice dates back to 3,300–2,700 BP (Zhang et al. 2006). The earliest unearthed rice remains in Yunnan date back to 4,235 ± 150 BP (Xiao 2001). Archaeological evidence dates planting rice in Southeast Asia between 3000 and 2000 BCE, which also spread from the Yangtze River basin (Fuller et al. 2008). The northeastern part of India and the India-Myanmar border have been considered the origin of Asian rice domestication due to their 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 subgroup 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 valley and expanded south or southwest. After arriving in Southeast Asia, they became the ancestral population of the modern AA-speaking people and migrated to northeastern India from Southeast Asia.

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 (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 BCE) in Chuxiong, Yunnan (K. S. Li and Huang 1990). The Wanjia-Ba–type bronze drums later developed into different subtypes belonging to the Shizhaishan culture (400 BCE) in Jinning, Yunnan, and to the Dong Son culture (500 BCE) in northern Vietnam (Peng 2016). 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 and Wei 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 Period, forming the Shizhaishan culture (Xi and Zhu 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 multilayered and complex nature of the migration and dispersal events of AA-speaking populations.

Molecular Anthropological Studies on 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 (mDNA) in the cytoplasm. mDNA is maternally inherited, while the Y chromosome is paternally inherited (H. Li and Jin 2015). Y chromosome haplotypes include two types of genetic markers: single-nucleotide polymorphisms (SNPs) and short tandem repeat (STRs). Y-SNPs are considered to faithfully record the migration and evolutionary history of populations due to its low mutation rate, while Y-STRs can record more recent events in population history, allowing estimates of the times of origin and divergence of populations (Kayser et al. 2000). Linguistic classification can also be reflected in Y chromosome DNA (C. C. Wang et al. 2013), providing a theoretical basis for interdisciplinary research between molecular anthropology and linguistics. Both disciplines 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 in Asia (HUGO Pan-Asian SNP Consortium et al. 2009). However, similar to debates in linguistics, molecular anthropological studies on the origins of South Asian, Southeast Asian, and southern East Asian AA-speaking populations and related groups have led to three different conclusions: India, MSEA, and southern China.

mDNA and AA-Speaking Populations

Basu et al. (2003), in a comprehensive statistical analysis of 58 DNA markers (including mitochondrial, Y-chromosomal, and autosomal markers) from a large number of ethnic groups in India, found that the mDNA 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 Y-STRs of the Indian haplogroup O2a revealed an origin that can be traced back to 65,000 BP, leading to the hypothesis that AA-speaking 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 from the Y-STR evolution mutation rates used by Basu et al. (2003).

Chaubey et al. (2011) systematically collected samples from AA-speaking populations in India and a few samples from Southeast Asia. The autosomal, Y-chromosomal, and mDNA 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 mDNA in the Munda-speaking populations revealed that all seven haplogroups were indigenous to India, and that their mDNA 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 mDNA diversity in contemporary Munda-speaking populations results from gender-biased admixture with neighboring populations in India.

Studies by Basu et al. (2003) and Chaubey et al. (2011) suggest that mDNA diversity is not a reliable criterion for determining the origin of AA-speaking populations.

Y-Chromosome Haplogroups and AA-Speaking Populations

Compared with mDNA, Y-chromosome diversity and coalescence time depth are often more informative. Kumar et al. (2007) analyzed 1,222 Y-SNPs and Y-STR data from 25 Indian populations, covering such AA language groups as the Munda, Khasi, and Mon-Khmer. Haplogroup O-M95 had the highest average frequency (52%): on average, 55% of Munda-speaking populations and 41% of Khasi-speaking populations, and all 11 Nicobarese samples, belonged to haplogroup O-M95. This suggests a strong patrilineal genetic connection between Indian Southeast Asian AA-speaking populations. 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 toward 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 greater time to the most recent common ancestor (1,700–2,800 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 (averaging about 45%), X. M. Zhang et al. (2015) sampled AA-speaking and TK-speaking populations from Cambodia, Thailand, and southern China, combined these samples with data from Chaubey et al. (2011), and estimated the age of the O2a1-M95 lineage in different populations. 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 (16,000 BP), India (10,000 BP), and Southeast Asian islands (11,000 BP). Moreover, the comparison of Y-STR haplotype diversity of the O2a1-M95 lineage among different geographic populations also showed the same distribution pattern: 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). This evidence suggests that the O2a1-M95 lineage originated in the TK-speaking populations in southern China and then expanded southward to MSEA and westward to India and migrated to Southeast Asian islands after the Last Glacial Maximum. In addition, Singh et al. (2021) recently identified four subbranches of the O2a-M95 lineage based on 1,437 samples from Munda-speaking populations: 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.

Whole-Genome Sequencing and AA-Speaking Populations

In recent years, the academic community has studied single-lineage genetic markers and sequenced entire genomes of different populations. Liu et al. (2020) analyzed genome-wide SNP data from the Kinh ethnic group in Vietnam and 21 other populations and found that genetic diversity in Vietnam is not largely indigenous: Vietnamese ethnolinguistic groups harbor multiple sources of genetic diversity, likely reflecting different ancestries 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. Liu et al. (2020) also found AA-speaking populations in Vietnam shifting to AN languages over the past 2,500 years.

Kutanan et al. (2021), compared whole-genome SNP data of 452 samples from 33 populations of five different linguistic families from Thailand and Laos with data from modern Asian populations and ancient Southeast Asian samples. 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. 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.

Thus, 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 and AN populations during their migration and dispersal in MSEA.

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 agricultural development in these regions was indigenous. In contrast, others supported the "dual-layer" hypothesis: expansion of farmers to the south caused the genetic diversity in Southeast Asia today (Jinam 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 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 (from 4,100 to 1,700 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, 4,100–3,600 BP) in Vietnam, Ban Chiang (late Neolithic to Iron Age, 3,500–2,400 BP) in Thailand, and Vat Komnou (Iron Age, 1,900–1,700 BP) in Cambodia clustered together with modern AA-speaking populations and moved slightly toward 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, 3,200–2,700 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 ancestors of AA-speaking populations and were more closely related to Sino-Tibetan populations. Meanwhile, individuals from Nui Nap in Vietnam (Bronze Age, 2,100–1,900 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, Lipson et al. (2018) identified two genetic admixture events outside MSEA, one involving admixture of Western Eurasian, Eastern Eurasian (South Asian aboriginal ancestry), and AA-speaking populations among the AA speakers in eastern India, and the other involving admixture among AN, AA, and Papuan populations in the AN populations in western Indonesia.

The study by X. M. Zhang 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 2,500 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 3,600 BP. This implies 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, which also has inspiring implications for historical-comparative linguistics. As the field of historical-comparative linguistics advances toward construction of superfamilies, research on the "homogeneity" of languages will shift to the detection of "relatedness" between language families (Sun and Wang 2020).

Conclusions

In the multidisciplinary perspective of language, archaeology, and genetics, examinations of the origin and dispersal of the AA language family 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, and the reconstruction of ancient living environments, archaeological sites, and so on, 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; from archaeological materials such as shouldered axes, rice domestication, and bronze drums; and from 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 1997), the coexistence and fusion of the shouldered stone axe and segmented stone adze in southern China (Fu 1988), the profound relationship between the Pu-Yue people and the Bronze Drum culture (Wan and Wei 2015), and genomic evidence that the present-day TK-speaking population in southern China is the primary ancestral source of the AA-speaking populations (X. M. Zhang et al. 2015, 2020; Lipson et al. 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, influenced by the expansion of the northern Han group and Chu culture, migrated southward or southwestward along the river valleys in southern China. Guo et al.'s (2021) genomic research on Tibeto-Burman and AA populations 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 populations in the Yangtze River Delta had cultural contact and fusion with the pre-AA 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 into 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 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 AA languages in Guangxi. Finally, the pre-AA ancestors who remained in southern China continued to intermix with the pre-TK ancestors from Fujian and Guangdong (Xu and Li 2017) 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.

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), and 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 [in Chinese]. *J. Guangdong Univ. Foreign Stud.* 5:92–96.

Basu, A., N. Mukherjee, S. Roy et al. 2003. Ethnic India: A genomic view, with special reference to peopling and structure. *Genome Res.* 13(10):2:277–2,290.

Blench, R. 2015. Reconstructing Austroasiatic prehistory. In *The Handbook of Austroasiatic Languages,* M. Jenny and P. Sidwell, eds. Leiden: Brill, 1–13.

Blust, R. A. 1996. Beyond the Austronesian homeland: The Austric hypothesis and its implications for archaeology. *Trans. Am. Philos. Soc.* 86(5):117–160.

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

Chang, K. C. 1986. *The Archaeology of Ancient China*. New Haven, CT: Yale University Press.

Chaubey, G., M. Metspalu, Y Choi. et al. 2011. Population genetic structure in Indian Austroasiatic speakers: The role of landscape barriers and sex-specific admixture. *Mol. Biol. Evol.* 28(2):1:013–1,024.

Chen, B. Y. 1997. The etymological relationship between Kam-Tai languages and Austro-Asiatic—on the relationship between the ancient Yue and Pu ethnic origins [in Chinese]. *J. Yunnan Nationalities Univ.* (Social Sci. Ed.) 1:40–44.

Cysouw, M., D. Dediu, S. Moran et al. 2012. Comment on phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science* 335(6069):657.

Dani, H. 1955. Prehistory and protohistory of eastern India. *Bull. Sch. Orient. Afr. Stud.* 24(3):596–598.

Dao, J. 2018. Distribution overview of Chinese South Asian languages. *J. Honghe Univ.* 6:126–132.

Deng, X. H., and T. J. Gao. 2014. Theory, methodology, and practice of evolutionary linguistics [in Chinese]. *J. Shanxi Univ.* (Philos. Social Sci. Ed.) 2:66–73.

Diffloth, G. 1974. Austro-Asiatic languages. In *Encyclopaedia Britannica,* M. Adler, ed. Chicago: Encyclopaedia Britannica, 480–484.

Diffloth, G. 1994. The lexical evidence for Austric. *Oceanic Linguist.* 33(2):309–321.

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

Diffloth, G. 2024. Austroasiatic languages. *Encyclopedia Britannica*. https://www.britannica.com/topic/Austroasiatic-languages.

Donegan, P. J., and D. Stampe. 2004. Rhythm and the synthetic drift of Munda. In *The Yearbook of South Asian Languages and Linguistics,* R. Singh, ed. Berlin: De Gruyter Mouton, 3–36.

Duff, R. 1970. *Stone Adzes of Southeast Asia: An Illustrated Typology* (*no. 3*). Christchurch: Canterbury Museum Trust Board.

Fan, Z. Q., X. H. Deng, and C. C. Wang. 2018. Language and genes: On the origin and dispersal of the Austronesian language family [in Chinese]. *Acad. Mthly.* 10:90–96.

Fu, X. G. 1988. On Youduan stone adzes and shouldered stone implements [in Chinese]. *Archaeology*, no. 1:23–32.

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

Fuller, D. Q., E. Harvey, and L. Qin. 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., L. Qin, Y. Zheng et al. 2008. Japonica rice carried to, not from, Southeast Asia. *Nat. Genet.* 40(11):1:264–1,265.

Guo, J. X., W. Wang, K. Zhao et al. 2021. Genomic insights into Neolithic farming-related migrations in the junction of East and Southeast Asia. *Am. J. Biol. Anthropol.* 33(1):1–15.

Gutaker, R. M., S. C. Groen, E. S. Bellis et al. 2020. Genomic history and ecology of the geographic spread of rice. *Nat. 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.

Headley, R. K., Jr. 1976. Some considerations on the classification of Khmer. In *Austroasiatic Studies,* P. N. Jenner, L. C. Thompson, and S. Starosta, eds. Honolulu: University of Hawai'i Press, 431–451.

HUGO Pan-Asian SNP Consortium, M. A. Abdulla, I. Ahmed et al. 2009. Mapping human genetic diversity in Asia. *Science* 326(5959):1:541–1,545.

Jiang, T. Y. 1981. *An Outline of the Archaeology of the Neolithic Period in Guangxi* [in Chinese]. Beijing: Wenwu Press.

Jinam, T. A., M. E. Phipps, F. Aghakhanian et al. 2017. Discerning the origins of the Negritos, first Sundaland people: Deep divergence and archaic admixture. *Genome Biol. Evol.* 9(8):2,013–2,022.

Kayser, M., L. Roewer, M. Hedman 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. *Am. J. Human Genet.* 66(5):1:580–1,588.

Kumar, V., A. N. Reddy, J. P. Babu et al. 2007. Y-chromosome evidence suggests a common paternal heritage of Austro-Asiatic populations. *BMC Evol. Biol.* 7(1):1–14.

Kutanan, W., J. Kampuansai, M. Srikummool et al. 2019. Contrasting paternal and maternal genetic histories of Thai and Lao populations. *Mol. Biol. Evol.* 36(7): 1:490–1,506.

Kutanan, W., D. Liu, J. Kampuansai et al. (2020). Reconstructing the human genetic history of mainland Southeast Asia: Insights from genome-wide data from Thailand and Laos. *Mol. Biol. Evol. 38*(8):3,459–3,477.

Li, G., G. He, J. Zhao et al. (2022). Genomic insights into Neolithic farming–related migrations in the junction of East and Southeast Asia. *Am. J. Biol. Anthropol* 177(2):328–342.

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

Li, K. S., and D. R. Huang. 1990. On the Wanjiaba type of Dong Son drums [in Chinese]. *KaoGu* 5:427–433.

Li, Y. 2016. New progress in research on the genetic relationships of South Asian languages. *Contemp. Linguist.* 1:76–86.

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

Liu, D., N. T. Duong, N. D. Ton et al. 2020. Extensive ethnolinguistic diversity in Vietnam reflects multiple sources of genetic diversity. *Mol. Biol. Evol.* 37(9):2:503–2,519.

Logan, J. R. 1854. Ethnology of the Indo-Pacific islands. *J. Indian Archipelago East. Asia* 8:28–200.

Maspero, H. 1929. Langues. In *Un empire colonial francais:*

L'Indochine, Vol. 1, G. Maspero, ed. Paris: Bruxelles: Les éditions G. van Oest, 63–82.

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

Mo, Z. 1959. Stone tools unearthed from Xiqiaoshan, Nanhai, Guangdong [in Chinese]. *Kaogu Xuebao* 4:1–17.

Mo, Z. 1961. New discoveries from archaeological surveys and excavations in Guangdong [in Chinese]. *KaoGu* 2:666–668.

Norman, J. 1983. Ancient dialectal characters in Min dialects [in Chinese]. *Fangyan* 3:159–165.

Norman, J., and, T.-L. Mei. 1976. The Austroasiatics in ancient South China: Some lexical evidence. *Monumenta Ser.* 32(1):274–301.

Peng, C. L. 2016. The evolution of the Dong Son culture area [in Chinese]. *Guangxi Ethnic Stud.* 1:54–59.

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 *The Origins of Chinese Civilization*, D. N. Keightley, ed. Berkeley: University of California Press, 411–466.

Renfrew, C. 1987. *Archaeology and Language: The Puzzle of Indo-European Origins*. London: Jonathan Cape.

Reid, L. 1994. Morphological evidence for the Austric. *Oceanic Linguist.* 33(2):323–344..

Sagart, L. 2008. The expansion of Setaria farmers in East Asia: A linguistic and archaeological model. In *Human Migrations in Continental East Asia and Taiwan: Matching Archaeology, Linguistics, and Genetics,* A. Sanchez-Mazas, R. M. Blench, M. D. Ross et al., eds. London: Routledge, 133–158.

Sagart, L. 2011. The Austroasiatics: East to west or west to east? In *Dynamics of Human Diversity: The Case of Mainland Southeast Asia*, N. J. Enfield, ed. Canberra: Pacific Linguistics, 345–361.

Sapir, E. 1916. Time perspective in aboriginal American culture: A study in method. *J. Am. Folkl.* 29(114):233–264.

Schmidt, P. W. 1906. Die Mon-Khmer-Völker: Ein Bindeglied zwischen Völkern Zentralasiens und Austronesiens. *Arch. Anthropol.* 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 *Early South East Asia*, R. B. Smith and W. Watson, eds. Oxford: Oxford University Press, 273–278.

Sidwell, P. 2008. Is Mon-Khmer dead? Long live Austroasiatic! Paper presented at the 18th annual meeting of the Southeast Asian Linguistics Society, Bangi, Malaysia.

Sidwell, P. 2010. The Austroasiatic central riverine hypothesis. *J. Lang. Relat.* 4:117–134.

Sidwell, P., and R. M. Blench. 2011. The Austroasiatic Urheimat: The Southeastern riverine hypothesis. In *Dynamics of Human Diversity in Mainland SE Asia*, N. Enfield, ed. Canberra: Pacific Linguistics, 317–345.

Singh, P. P., S. Vishwakarma, G. N. N. Sultana et al. 2021. Dissecting the paternal founders of Mundari (Austroasiatic) speakers associated with the language dispersal in South Asia. *Eur. J. Hum. Genet.* 29(3):528–532.

Sun, N., and, C. C. Wang. 2020. The dispersal of language from a three-dimensional perspective [in Chinese]. *China Social Sci. News*, 15 December.

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

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

Tu, L. J. 2009. A study on the origin of the name "Kunming" [in Chinese]. *Yunnan Normal Univ. J.* (Philos. Social Sci. Ed.), no. 6:124–131.

van Driem, G. 2008. The Shompen of Great Nicobar Island: New linguistic and genetic data, and the Austroasiatic homeland revisited. *Mother Tongue*, no. 13:227–247.

von Heine-Geldern, R. 1917. Kopfjagd und Menschenopfer in Assam und Birma und ihre Ausstrahlungen nach Vorderindien. *Mitt. Anthropol. Ges. Wien* 37:1–65.

van Driem, G. 2012. The ethnolinguistic identity of the domesticators of Asian rice. *CR Palevol.* 11(2–3):117–132.

Wan, F. B., and D. F. Wei. 2015. On the circle of Bronze Drum culture [in Chinese]. *Guangxi Ethnic Stud.* 1:118–125.

Wang, C. C., L. X. Wang, R. Shrestha et al. 2013. Inferring human history in East Asia from Y chromosomes. *Invest. Genet.* 4(1).

Wang, J. L., and X. Chen. 1982. On the cognate relationship of "village," "surname," and "home" in the Mon-Khmer and Tai-Kadai languages [in Chinese]. *Minzu Yuwen*, no. 3:12–24.

Wu, A. Q. 2000. The dissemination of ancient South Asian culture [in Chinese]. *Shijie Minzu*, 2:66–73.

Wu, C. M., and J. Cao. 2005. Four misconceptions in the research on the origin of the Austronesian language family [in Chinese]. *J. Xiamen Univ.* (Philos. Social Sci. Ed.) 3:110–115.

Xi, K. D., and X. S. Zhu. 1987. A report on the excavation of Bi Jie Wao Yao site, Guizhou [in Chinese]. *KaoGu*, no. 4:303.

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

Xu, J. X., and, H. Li. 2017. Molecular anthropology perspective: a new discussion on the origin of Guangxi's indigenous ethnic groups [in Chinese]. *J. Guangxi Normal Univ.* (Philos. Social Sci. Ed.) 4:85–91.

Ye, X. F. 2014. South Austroasiatic components in ancient Chu language [in Chinese]. *Minzu Yuwen*, no. 3:28–36.

Zeng, Q. 1983. On the Neolithic culture in South China [in Chinese]. *Prehist. Res.* 1:57–57.

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

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

Zhang, H. M. 1998. Chinese etyma for river. *J. Chin. Linguist.* 26:1–47.

Zhang, X. M., S. Liao, X. Qi et al. 2015. Y-chromosome diversity suggests southern origin and Paleolithic backwave migration of Austro-Asiatic speakers from eastern Asia to the Indian subcontinent. *Sci. Rep*. 5:15,486.

Zhang, X. M., .C. Li, Y. Zhou et al. 2020. A matrilineal genetic perspective of hanging coffin custom in southern China and northern Thailand. *iScience* 23(4):1–8.

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