

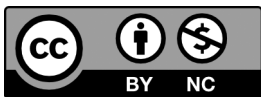
# The Typology Conversion and Integration History of Tsat and Utsat in Sanya, Hainan

Zhiquan Fan<sup>1,3</sup> Peipei Fu<sup>2</sup> Yunong Ye<sup>2</sup>

1. The Center for Studies of Fujian and Taiwan, Fujian Normal University, Fuzhou, China;
2. College of Chinese Language and Literature, Fujian Normal University, Fuzhou, China;
3. Institute of Austronesian, Fujian University of Technology, Fuzhou, China

**Abstract:** The Austronesian language group is spoken in China by the Gaoshan people of Taiwan region and the Huihui People from Sanya in Hainan (SYH). The origin of the SYH is still unclear. This article compares the Austronesian languages of China's Taiwan region with the Tsat based on research on the correlation between language and genetic evolution and explains the SYH and their evolutionary history. This article supports the view that the modern SYH originated from three ancestral sources: Central/South Asia, Southeast Asia, and East Asia, with the East Asian people being the main body. The formation of the modern Hui people in East Asia was due to both cultural and demographic diffusion.

**Keywords:** Austronesian Languages in China's Taiwan Region; Tsat; Language Type; Fusion



Copyright © 2025 by author (s) and SciScan Publishing Limited

This article is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/). <https://creativecommons.org/licenses/by-nc/4.0/>

The Huihui People from Sanya in Hainan (SYH) currently reside in communities such as Huihui and Huixin in Tianya District, Sanya City, Hainan Province, China. It represents an important part of the language and cultural heritage of the Hainan Island language group in southern China. Since its establishment during the Qin and Han dynasties, the Maritime Silk Road has served as a crucial conduit for economic and cultural exchanges between the East and the West. Located at the southern tip of Hainan Island, Sanya has historically served as a key gateway, facilitating interactions between the mainland and maritime regions. Furthermore, it has become an important town in the southern Chinese border region, serving as a crucial nexus and an integral part of the Maritime Silk Road. Since the Tang and Song dynasties, foreign merchants and immigrants from regions such as Guangdong, Fujian, and Zhejiang—regions that had established commercial ties with Arabia, Persia, and Southeast Asia—have entered China through ports such as Sanya (Huang, 2008).

The 7th National Census of 2020 revealed that 17,089 Hui people were living in Hainan Province, with 10,368 residing in

Research/Funding project: This article is a result of the National Social Science Foundation of China, “A Study on the Origin and Formation of Austronesian Languages from a Multidisciplinary Perspective” (Project No: 20&ZD248).

About author: Zhiquan Fan, male, Han ethnicity, Full-time Teacher at the Center for Studies of Fujian and Taiwan, Fujian Normal University, Institute of Austronesian, Fujian University of Technology. Research interests: Linguistic Anthropology. Peipei Fu, female, Han ethnicity, Graduate student at the College of Chinese Language and Literature, Fujian Normal University. Research interests: Linguistics. Yunong Ye, male, Han ethnicity, Graduate student at the College of Chinese Language and Literature, Fujian Normal University. Research interests: Linguistics.

Article Citation: Fan, Z. Q., Fu, P. P., & Ye, Y. N. (2025). The typology conversion and integration history of Tsat and Utsat in Sanya, Hainan. *Linguistics*, 7(4), 235-248.

Tianya District, Sanya City, making up over 60% of the province's Hui population. The SYH in Sanya speaks “*Tsat*”<sup>1</sup> and refers to their own language as (*hu*<sup>1</sup>)*tsa:n*<sup>32</sup>, which belongs to the Austronesian language family. Apart from the Gaoshan people of China's Taiwan region, they are the only two groups in China that speak Austronesian languages, They are the only two groups in China that speak Austronesian languages. The origin of the SYH remains uncertain. Despite the contributions of numerous scholars, the historical documents they reference are largely similar, leading to a lack of definitive conclusions and a shortage of corroborating evidence. Proposed origins for the SYH include Central Asia and Arabia, the Champa region of Vietnam on the Indochina Peninsula, and mainland China (Jiang & Dong, 1992).

To determine the origin and migration history of ethnic groups, it is possible to draw upon several disciplines, including genetics, linguistics, archaeology, and cultural anthropology. Genetics provides the most stable and reliable source of information, followed by language, and then by culture, which is more susceptible to borrowing. This article will therefore employ a research method that integrates linguistics and molecular anthropology, drawing on the findings of previous scholars. The objective is to clarify the origins and dispersal of the SYH using more reliable and objective research evidence, thus enabling a more comprehensive and systematic investigation into their historical origins, language, and the history of ethnic integration.

## 1 Research on the Correlation between Language and Genes

In recent years, the relationship between language and population has constituted a prominent area of academic inquiry. Darwin's theory of evolution, as outlined in *The Origin of Species* (1859), provided a foundation for interdisciplinary research in linguistics. This approach involved the use of the biological model of phylogenetic systematics to examine the relationships between languages and dialects. For example, the German scholar Schleicher (1861) employed this approach to construct the inaugural “phylogenetic tree”<sup>2</sup> of the Indo-European language family. Subsequently, phylogenetic tree theory has emerged as a pivotal theoretical model in historical-comparative linguistics (Xu, 1991). Additionally, Darwin observed the significant parallels between the evolution of populations and languages (Cavalli-Sforza, 2001). Cavalli-Sforza, a professor at Stanford University in the United States, was a pioneering figure in this field. In 1988, he and other researchers constructed a tree comparison map of human ethnic group genes and languages (Cavalli-Sforza, Piazza, & Menozzi, et al., 1988). The results demonstrate that the genetic lineage and language classification of humans have evolved in parallel, with the exception of a very small number of cases (Xu, 2015).

Furthermore, in addition to exhibiting analogous family relationships, there are numerous significant parallels between biological genetics and linguistics. These phenomena can be broadly summarized as follows by Atkinson and Gray (2005) and Pagel (2009):

**Table 1** A Comparison of Parallel Concepts in Biology and Linguistics(Xu, 2015; Deng & Gao, 2014)

Biological evolution	Linguistic evolution
Discrete characters (Nucleotides, Amino acids, and Genes)	Lexicon, Syntax, and Phonology
Mechanisms of replication	Teach, learning, and imitation
Homologies Mutation	Cognates Innovation

1 Huihui language, the exclusive language of the SYH, belongs to the Austronesian language family, with approximately 10,000 current speakers.

2 A phylogenetic tree, a tree-like diagram, is used to describe the evolutionary relationships between species or languages and is typically constructed using genetic or linguistic data.

Continued

Biological evolution	Linguistic evolution
Drift	Drift
Natural selection	Social selection
Cladogenesis	Lineage splits
Anagenesis	Linguistic change without a split
Horizontal gene transfer	Borrowing
Plant hybrids	Language Creoles
Correlated genotypes/phenotypes	Correlated cultural terms
Geographic clines	Dialects(chains)
Fossils	Ancient texts
Extinction	Language death

As shown in Table 1, the research methods employed in biology and linguistics are so closely intertwined that they are often used together. For instance, the phylogenetic tree model from biological evolution theory and quantitative analysis techniques has gradually been applied to the field of linguistics. In addition, Swadesh's "Etymological Statistics"<sup>1</sup> Represents one of the earliest quantitative methods introduced into linguistics. Despite criticism of its assumptions regarding constant lexical replacement rates, the quantitative methods from etymological statistics have paved the way for a novel approach to estimating the age of language divergence and its temporal depth.

In recent years, some scholars have started to combine theories and methods of biological phylogeny with the statistical analysis of word origins in linguistics, along with computer algorithms, to study the classification of language lineages and the construction of phylogenetic trees. The results of research in this area [e.g., Austronesian languages (Gray & Jordan, 2000; Gray, Drummond, & Greenhill., 2009), Indo-European languages (Gray & Atkinson, 2003), and Sino-Tibetan languages (Zhang, Yan, & Pan, et al., 2019; Sagart, Jacques, & Lai, et al., 2019; Zhang, Ji, & Pagel, et al., 2020)] have been published in leading international academic journals, including Nature and Science.

This research method, which reconstructs the linguistic family tree using biological phylogeny, allows for a more precise calculation of the affinity between languages. Compared to the qualitative approach based on personal experience, this method offers a novel perspective and a deeper understanding of the phylogenetic classification of languages. Linguists have applied evolutionary theory to develop various theoretical models for language evolution, which help explain the process of linguistic change. This has not only advanced the field of historical linguistics (Deng & Gao, 2014) but also established a crucial and indispensable link for interdisciplinary research at the intersection of linguistics, genetics, molecular anthropology, and other disciplines.

## 2 The Genetic Affiliation of Tsat in the SYH

The emergence of language coincides with the advent of the human species and serves as a fundamental tool for human thought and expression. Therefore, the investigation of the genesis and evolutionary trajectory of human

<sup>1</sup> Etymological Statistics, a method of comparative linguistics, identifies interlinguistic relationships by comparing the percentage of cognate words across languages.

populations is closely intertwined with the study of language (Wen, Xie, & Xu, 2013). Scholars have made significant contributions to the field of linguistics by investigating the family status and evolutionary history of the Tsat spoken by the SYH. However, due to limitations in their respective theories, methods, materials, or differing classification standards, scholars have been unable to reach a consensus on this issue, and divergent views persist. The following is a synthesis of the main perspectives concerning the family status of the Sanya Hui language: (1) Scholars such as Benedict, Haudricourt, Zheng Yiqing, Thurgood, and Meng Simu propose that Tsat is related to Vietnamese Cham and belongs to the Austronesian language family (Haudricourt, 1984; Zheng, 1986; Zhang, 2018), having developed from Proto-Austronesian (Meng, 1992, 1995); (2) Ni (1988) argues that the historical origin and structural type of Tsat are distinct and should be classified as belonging to different language families. Consequently, he proposes that the term “Austronesian-Sino-Tibetan” be used to describe them; (3) Zeng and Yin (2011) argue that Tsat is a unique language formed through long-term and intensive contact between the Cham language, Chinese, Hlai, and other languages, which has altered the basic characteristics of the Austronesian language. Some scholars even suggest that it may be a hybrid language (Li & Zhang, 1999). While scholars disagree on the specifics of Tsat’s history, it is clear from the previous analysis that there is a consensus among them that the language originated from the Cham branch of the Austronesian language family.

It is noteworthy that Tsat is situated within a linguistic context characterized by the prevalence of tonal, monosyllabic languages. In terms of linguistic classification, Tsat shares similarities with or is comparable to Chinese and Tai-Kadai languages. This is evident in the structural characteristics of monosyllabic words, the use of tones to differentiate word meanings, the basic subject-verb-object word order, and the expression of grammatical relationships through function words and word order. Fan et al. (2019) have conducted comprehensive research into the nature and evolutionary history of Tsat, examining key aspects such as its tone system, basic word order, and cognate words. This research draws on the insights of both previous and contemporary scholars in the field. However, due to the long history of contact between Hui and Chinese speakers, the linguistic characteristics of Hui are often viewed as a result of this contact. Indeed, some scholars even suggest that Hui has undergone a qualitative change as a consequence of this interaction (Zeng & Yin, 2011). Moreover, the Austronesian language spoken by the Gaoshan people of China’s Taiwan region is notable for retaining the characteristics of the original early Austronesian language. It is considered to occupy a crucial position among all Austronesian languages. It occupies a place at the top of the Austronesian family tree, closest to Proto-Austronesian, the ancestral language.

First, regarding the evolution of the basic SVO word order<sup>1</sup> in Tsat, it can be observed that in Austronesian languages, the predicate typically appears in the initial position, with the basic word order generally following the VSO or VOS pattern. However, there are instances where Austronesian languages exhibit a basic word order that deviates from the SVO structure<sup>2</sup>.

---

1 SVO word order refers to a sentence structure where the Subject, Verb, and Object are arranged in the sequence of “subject-verb-object,” and it is one of the most common word order types worldwide.

2 Data Sources: The Thao data is taken from Shih-Lang Jean. (2016) *An Introduction to the Thao Language Grammar*. Council of Indigenous Peoples; the Amis data is taken from Joy Wu. (2016) *An Introduction to Amis Grammar*. Council of Indigenous Peoples; the Saisiyat data is taken from Marie Meili YEH. (2016) *An Introduction to Saisiyat Grammar*. Council of Indigenous Peoples; the Rukai data is taken from Elizabeth ZEITOUN. (2016) *An Introduction to Rukai Grammar*. Council of Indigenous Peoples; the Seediq data is taken from Li-May Sung. (2016) *An Introduction to Seediq Grammar*, Council of Indigenous Peoples; the Indonesian data is taken from Tang Hui et al. (2017) *Indonesian Grammar*, World Publishing Company; the Tsat data is taken from Tian Xiangsheng and Chen Baoya. (2019) *Tsat of Sanya, Hainan*, Commercial Press; the South Asian languages data is taken from Yan Qixiang and Zhou Zhizhi. (2011) *Languages of the Miao-Yao Language Family in China and the South Asian Language Family*, Social Sciences Academic Press; Proto-Austronesian data is taken from: <https://www.trussel2.com/ACD>; the “Chamic languages”(Chamic languages, a subfamily of the Austronesian language family, is primarily distributed in Vietnam and Cambodia, and bears a genetic relationship with the Huihui language of Sanya.)data is taken from: <https://abvd.shh.mpg.de/austronesian/>.

Saisiyat: hi 'oya' mam 'omangang ka korkoring.

'Mother scolding child.'

Tsou: (ti)'ama amaniza'.

'Dad is going fishing.'

Indonesian: Agus mencuci piring.

'Mr. Argus washes dishes.'

The basic word order of Tsat is SVO, similar to Saisiyat, Tsou, and Indonesian. This is likely due to the fact that the subject, marked by a case marker, often moved to the initial position of the sentence to function as the topic, eventually developing into an SVO word order. In some cases, the subject acting as the topic may later become the syntactic subject (Shen, 1997). The Rukai language of China's Taiwan region offers a rich source of linguistic inspiration.

The Rukai language is an Austronesian language that does not use a case marker. Ambiguity in Rukai sentences can arise when two noun phrases can both function as the subject. For example, consider the following:

Rukai: oka'ace ta'olro 'olra'a.

(a) 'the dog bites snake.' or (b) 'the snake bites dog.'

In consequence, there are two methods of handling this in the Rukai language. The first method is to insert the marker 'i before the subject, as in:

Rukai: oka'ace 'i ta'olro 'olra'a.

'Dog bites snake.'

or oka'ace 'i 'olra'a ta'olro.

'Snake bites dog.'

The second moves the subject to the beginning of the sentence, as the subject, as in:

Rukai: dhona ta'olro oka'ace 'olra'a.

'That dog, it bites snakes.'

dhona 'olra'a oka'ace ta'olro.

'That snake, it bites dogs.'

In light of the evidence pertaining to the Saisiyat, Tsou, Rukai, and other languages, it seems reasonable to posit that the evolutionary trajectory of the fundamental SVO word order in the Austronesian languages may have unfolded as follows: V NC-S OC-O > NC-S V OC-O >  $\emptyset$ -S V OC-O >  $\emptyset$ -S V  $\emptyset$ -O > S V O. The SVO word order observed in the Tsat may have been largely unaffected by the influence of other languages with disparate structural characteristics. It is conceivable that the Proto-Austronesian language spoken by the Cham peoples in Indochina before they migrated to Sanya, and even during the Indonesian period in Southeast Asia, exhibited an SVO word order. It can thus be argued that the qualitative change of the language cannot be attributed solely to the borrowing of vocabulary. Rather, the primary factor in this change lies in the procedures and methods with distinctive national norms that the brain employs to process and organize information reflecting the objective world, thereby creating external language forms. This results in the consolidation of the aforementioned processes within the language itself, manifesting as a distinct grammatical structure (Chen, 1989). In other words, the grammatical system plays a more significant role than vocabulary in determining whether a language has undergone a qualitative change (Xu, 2018). In light of the considerations above, this article proceeds to undertake a comparative analysis of the noun phrases of the principal Tsat s with those of the Austronesian languages of China's Taiwan region (Amis, Seediq, Saisiyat, Rukai) and Indonesian. This endeavor is undertaken with a

view to facilitating academic discourse.

## 2.1 Word Order of Noun Complements and Noun Modifiers

In Tsat, when a noun (N) modifies a noun phrase [N (head)], the noun phrase is inclined and placed before the noun modifier, i.e., N (head) + N. This is also the case in Seediq and Indonesian. Concurrently, the Tsat, akin to Amis and Rukai, exhibits the placement of noun phrases subsequent to noun modifiers, specifically N + N (head). To illustrate, consider the following example:

- Tsat:        ma<sup>24</sup> za:ŋ<sup>32</sup>  
               N(head) -stuff N-people  
               ‘Someone else’s stuff.’
- Seediq:     kari Seediq  
               N(head) - language N- Seediq  
               ‘Seediq language’
- Indonesian: tantangan alam  
               N(head) -challenge N-Nature  
               ‘Challenges of nature’
- Tsat:        thoi<sup>11</sup> ŋu<sup>24</sup> ŋo<sup>42</sup>  
               N-desk N(head) - on top of  
               ‘Top of the table’
- Amis:        timolan a niyaro’  
               N-the south structural particles N(head) - tribe  
               ‘Southern tribes’
- Rukai:      ’oponoho vaha  
               N-Wanshan N(head) - language  
               ‘Rukai Wanshan dialect’

## 2.2 Word Order of Noun Complements and Possessives

In the possessive structure<sup>1</sup>, the possessive pronoun (G, genitive) of the Tsat tends to be placed after the noun head, i.e., N (head) + G, as in Amis, Seediq, and Indonesian. However, the Tsat and Amis languages also allow the possessive pronoun to be placed in front of the noun head, i.e., G + N (head). For example:

- Tsat:        ŋan<sup>33</sup> kau<sup>33</sup>  
               N(head) - hand G-me.  
               ‘My hand’
- Amis:        kafoti’an no mako  
               N (head) -room genitive case possessive case-me  
               ‘My room’

<sup>1</sup> The possessive structure, a grammatical construct in language, expresses the “relationship of ownership,” with its core function being to clarify the attributive association between the “possessor” and the “possessee.”

- Seediq: laqi=mu  
N (head) -child =genitive case.me  
'My child'
- Indonesian: usul saya  
N(head)-suggestion G-me.  
'My suggestion'
- Tsat: ha<sup>33</sup> sa<sup>33</sup> kha:n<sup>21</sup>  
G-you structural particles N (head) -book  
'Your book'
- Amis: mako a mama  
Possessive case - me structural particles N (head) - father  
'My father'
- Saisiyat: 'ansiya'a korkoring  
possessive case - he N (head) - child  
'His child'
- Rukai: Dhipolo tolropongo=ni  
Name noun N (head) - hat = genitive case. he  
'Dhipolo's hat'

### 2.3 Word Order of Noun Center and Adjective

In a modifying structure composed of an adjective (Adj, adjective) and a noun center, the Tsat, like the Seediq language and Indonesian, exhibits a tendency to postpone the adjective after the noun center, i.e., N (head) + Adj. However, the Amis language, Saisiyat language, and Rukai language demonstrate a distinct preference for preposing the adjective to the noun center, i.e., Adj + N (head). For example:

- Tsat: thoi<sup>11</sup> za<sup>55</sup>  
N(head)-table Adj-red  
'Red Table'
- Seediq: ihengo paru  
N (head) - cavern Adj - large  
'Large cavern'
- Indonesian: sikap positif  
N(head)-attitude Adj-positive  
'Positive attitude'
- Amis: kohecalay a koko'  
Adj-white structural particles N (head)-chicken  
'White chicken'
- Saisiyat: 'imahabiyalan ka pongaeh  
'ima Adj-yellow structural particles N (head)-flower  
'Yellow flower'

Rukai: toalrai-nga 'angato  
Adj-big-superlative N (head)-tree  
'Big tree'

## 2.4 Word Order of The Noun Head and Numeral

In general, the numeral (Num, numeral) of the Tsat is combined with the classifier (CL, classifier) to form a quantity structure, which then modifies the noun center. The resulting word order is Num+CL+N. This is consistent with the grammar of the Amis, Indonesian, and Saisiyat languages. The classifiers in Austronesian languages of China's Taiwan region are underdeveloped (Li, 2008). In the absence of classifiers, numerals can be directly combined with noun centers to form a modifying structure. Consequently, in the modifying structures formed by numerals and noun centers in the majority of Austronesian languages of China's Taiwan region, the numerals are typically positioned before the noun center. In most cases, structural particles or classifiers are added between the numeral and the noun center, resulting in a structure of Num + structural particles/CL + N. The following examples illustrate this point:

Tsat: ta<sup>11</sup> se<sup>33</sup> za:ŋ<sup>32</sup>  
Num - one CL - person  
'A person'

Indonesian: tiga ekor ikan  
Num-three CL-pieces fish  
'Three pieces of fish'

Amis: ta-tusa a tamdaw  
Num-two structural particles people  
'Two people'

Saisiyat: Sepat ka pongaeh  
Num-four structural particles flower  
'Four flowers'

## 2.5 Word Order of Demonstratives and Head Nouns

In Tsat, the demonstrative pronoun (Dem, demonstrative) can be placed in two different positions with respect to the noun head. It can be placed after the noun head, as in Seediq and Indonesian, that is, N (head) + dem, or it can be placed before the noun head, that is, Dem + N (head), usually with a measure phrase or measure word in between (Tian & Chen, 2019), Amis, Saisiyat and Rukai all allow demonstratives to be placed before the noun focus. For example:

Tsat: kha:n<sup>32</sup> sa<sup>33</sup> ta<sup>11</sup> phoŋ<sup>33</sup>  
N(head)-letter Dem-this Num-one CL-envelope  
'This letter'

ʔia:n<sup>32</sup> sa<sup>33</sup> ta<sup>11</sup> zo:ŋ<sup>32</sup>  
N(head)-vegetable Dem-this Num-one CL-kind  
'This kind of vegetable'

Seediq: bunga nii  
N(head)-Groundnut Dem-This.

	‘This groundnut’
Indonesian:	pistol ini N(head)-pistol. Dem-this. ‘This pistol’
Tsat:	ni <sup>33</sup> ɓe <sup>24</sup> ʔia <sup>33</sup> ɓe <sup>24</sup> Dem - this Quantifier - strip N (head) - creek ‘This creek’
Amis:	kora a wawa Dem-na structural particles N (head) - child ‘that child’
Saisiyat:	hini ka kor koring Dem-this structural particles N (head)-child ‘this child’
Rukai:	ana iroolai Dem-that N(head)-child. ‘The kid’

In terms of vocabulary, the Tsat has borrowed a substantial number of cultural words from Chinese, as well as a number of Chinese words used in religious and folk contexts. For example, the word ta33tɛiŋ33, which refers to bathing the whole body according to Islamic ritual, is derived from the Chinese word “Dajing.” The word hau33tha21ta33tɛiŋ21, which means “bring water,” is also derived from Chinese. Finally, the word kin33sa:i21 “emissary” is used to refer to the messenger of God; sɛŋ33xun33 “hadith” is used to refer to the hadith of the Prophet Muhammad. Nevertheless, Swadesh 100 core vocabulary indicates that the proportion of cognates in Tsat, Indonesian, and Cham is considerably higher than that in Chinese and Tibetan languages belonging to the Sino-Tibetan language family, as well as in Vietnamese, which is classified as a South Asian language. It is noteworthy that the Tsat shares certain fundamental vocabulary with the Hlai (or Cuengh) language, as well as with the Cham language branch. It is evident that these words were not adopted into the Tsat from other languages, such as Hlai, but rather existed prior to the arrival of the SYH. Such terms should be regarded as words in the Cham language branch (or Austronesian language) that are also found in the Tai-Kadai language family (Zheng, 1997). Furthermore, among the 100-core vocabulary, the Tsat has also incorporated select words from other South Asian languages, including:

Tsat:tsun<sup>33</sup>“Bird,” Proto-Austronesian(PAN):\*qayam; Amis:’ayam; Indonesian:buruŋ; Proto-Cham:\*cĩm; Western Cham:chĩm pãr; Jarai:chim; Cuengh:cim; Raglaiese:chip. Proto-Mon-Khmer:\*cum; Wa\_Yanshuai:sim; Bulang\_Pangpin:ʃĩm<sup>31</sup>; Khmu\_Nanqian:ʃem; Gin:tsim; Mang:θəm<sup>35</sup>.

Tsat: luai<sup>33</sup>“Swim,” PAN:\*Naŋuy; Amis: dangoy; Indonesian:berenaŋ; Proto-Cham:\*luai; Western Cham:cha luai; Jarai:loĩ; Cuengh:lua:i; Raglaiese:luai. Proto-Mon-Khmer:\*luyʔ; Wa\_Yanshuai:loi; Bulang\_Pangpin:lɔi<sup>51</sup>; Khmu\_Nanqian:vvi; Gin:boi; Mang:luan<sup>31</sup>.

Tsat: lua<sup>33</sup>“Leaf,” PAN:\*waSaw; Amis:papah; Indonesian:daun; Proto-Cham:\*sula; Western Cham:hla; Jarai: hla; Cuengh:sla:; Raglaiese:hla:q. Proto-Mon-Khmer:\*slaʔ; Wa\_Yanshuai:lhaʔ; Bulang\_Pangpin:la<sup>55</sup>; Khmu\_Nanqian:kɔlna; Gin:lá; Mang:la<sup>51</sup>.

In conclusion, regardless of whether it is phylogenetic relationship or typological characteristics, the Tsat exhibits

a close relationship with the Austronesian languages. Nevertheless, this does not preclude the possibility of other factors being involved. The typological characteristics of the Tsat have undergone evolution as a result of the combined influence of its own internal Austronesian factors and language contact with other languages, including Chinese and Hlai. The primary factors are those operating internally.

### 3 A Molecular Anthropological Analysis of the Genetic Structure of the SYH

In the early stages of population evolution research, scholars typically used phenotypic characteristics (or traits) of organisms or species, such as human morphological features, to delineate the relationships of affinity between research subjects. As science and technology have advanced, genetic classification at the molecular level has become a more accurate method for elucidating evolutionary relationships between organisms or species than the use of phenotypic traits. The field of population evolution has gradually shifted from traditional physical anthropology to molecular anthropology. Molecular anthropology is a discipline that uses the analysis of human genetic information to explore a range of complex issues related to human origins, the evolution of ethnic groups, the structure of ancient social and cultural systems, and various other multi-dimensional topics. The focus of the study is the human genome (Li & Jin, 2015).

The human genome consists of two distinct types of DNA: chromosomal DNA, which is located in the cell nucleus, and mitochondrial DNA (mtDNA), which is found in the cytoplasm. According to the principles of inheritance, the human genome can be classified into three categories: autosomal DNA, mitochondrial DNA (found in mitochondria), and Y-chromosome DNA (found in the male sex chromosome). Autosomal DNA follows inheritance patterns from both parents, undergoes recombination during transmission, and is subject to alteration by admixture. In contrast, mitochondrial DNA (mtDNA) and Y-chromosome DNA (Y-DNA) exist as single copies in cells, serve as single-lineage genetic markers, and are not influenced by admixture, which a relatively simple evolutionary tree diagram can represent.

Previous research on Hui genetics has primarily focused on elucidating the genetic marker characteristics of this culturally distinctive group. It is hypothesized that the Hui people in northern East Asia are descendants of immigrants from Central and Western Asia, within the broader context of the spread of Muslim culture. However, recent genetic findings have revealed a close genetic relationship between the Hui people, whose geographical distribution is diverse, and the East Asian population (He, Wang, & Wang, et al., 2018). On the other hand, the study of paleogenetics and whole-genome genetics, which analyzes shared alleles and haplotypes, has significantly advanced our understanding of the historical development of human populations, while also providing a wealth of novel insights into related research fields. Among these studies, three articles focus on the origins of the Hui ethnic group in central and northwestern China (Wang, Zhao, & Ren, et al., 2021; Liu, Yang, & Li, et al., 2021; Ma, Yang, & Gao, et al., 2021). Overall, the genomic diversity of the geographically dispersed Hui population suggests complex interconnections between populations in the eastern and western regions of Eurasia. It is noteworthy that the majority of these studies have focused on the Hui ethnic group in northern and southwestern China. Over the past decade, some scholars have also begun investigating the genetic characteristics of the SYH (Sun, Yang, & Ou, et al., 2007; Li, Wen, & Chen, et al., 2008). Their findings indicate that, in contrast to the Hui population on the mainland, the SYH exhibit distinctive genetic traits. This group is more closely related to the Tai-Kadai-speaking populations of Hainan (Li, Wang, & Yang, et al., 2013). These results suggest

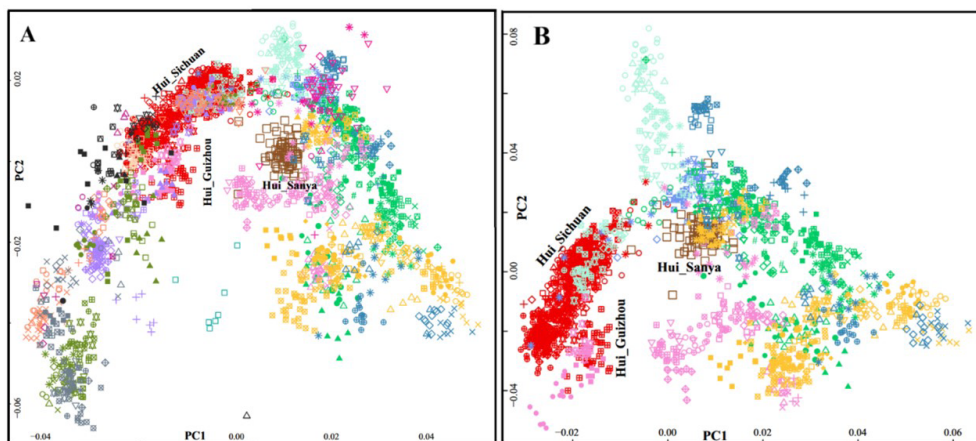
that the formation of the SYH may be linked to the large-scale assimilation of indigenous peoples. However, analyses based on single-parent genetic markers could skew our understanding of the contact and integration history of the SYH.

To achieve this objective, in July 2021, under the guidance and support of the Hainan Provincial Society of Ethnology, the Hainan Islamic Association, and the Sanya Islamic Association, our research team collected a total of 94 samples (37 females, 57 males) from the two Hui communities of Huihui and Huixin in Tianya District, Sanya City<sup>1</sup>. The remaining 57 individuals were male. More than 700,000 whole-genome single-nucleotide polymorphism (SNP) data points were genotyped, and the kinship coefficient was subsequently estimated. A population genetic analysis was then conducted, during which six individuals from three generations were excluded. This process resulted in 88 valid samples with whole-genome SNP data.

This study initially integrated the newly acquired data with three distinct single-nucleotide polymorphism (SNP) density datasets to investigate the comprehensive genetic structure through principal component analysis (PCA)<sup>2</sup>. The PCA, based on the combined HO data, revealed two distinct genetic populations: one located in northern East Asia and the other in southern East Asia. The SYH were positioned between these two populations, as illustrated in Figure 1A.

Additionally, this study examined the genetic diversity of the Han Chinese and other populations, including the Austronesian, Tibeto-Burman, Tai-Kadai, Hmong-Mien, and Austroasiatic groups. The results indicate that the Miao-Yao and Austronesian language groups (Cankanaey in the Philippines and Amis in Taiwan region, China) are located at the apex, while the Han Chinese and Tibeto-Burman groups (Bai, Naxi & Pumi) occupy the lower left quadrant. Other Austronesian groups are distributed across the remaining quadrants. The South Asian, Tibeto-Burman, and Tai-Kadai-speaking groups are situated in the bottom right, with the SYH positioned centrally among these groups and exhibiting a close relationship with the Kinh and Vietnamese, who speak South Asian languages (see Figure 1-B).

Subsequent regional PCA further corroborated the genetic similarities and differences between the Hui and Han populations. The findings conclude that the SYH exhibit a distinctive genetic structure that is notably different from that of the Han and other Hui groups from different geographical regions.



**Figure 1** Principal component analysis of the SYH and East and Southeast Asian populations (He, Fan, & Zou, et al., 2023).

1 Note: The samples must consist of direct descendants of the Hui ethnic group indigenous individuals, with no close consanguineous marriages to other ethnic groups in the past three generations. This study obtained written informed consent from each participant. It was approved by the Ethics Committee of Xiamen University (Approval No.: XDYX2019009), in accordance with the recommendations of the Declaration of Helsinki.

2 Principal Component Analysis, a statistical method, visualizes population genetic structure through dimensionality reduction and is commonly applied in molecular anthropology research.

## 4 Conclusion and reflection

The origin of the SYH has been the subject of considerable debate. The prevailing view is that the group originated through one of three processes: population diffusion, where non-East Asian populations migrated to the region; simple cultural diffusion, where the indigenous East Asian population was transformed into Hui people through assimilation into Islam; or continuous migration, where gene flow occurred between non-East Asian populations and the indigenous East Asian population. Ethnologists argue that the self-designation “hu11tsa:n32” used by the SYH represents the majority of their ancestors, or that it is an ethnic self-proclamation regarding the group’s historical origins.

The SYH speak the Tsat, which belongs to the Cham branch of the Austronesian language family. Linguistic evidence indicates that the Tsat is closely related to other Austronesian languages, both in terms of phylogenetic relationships and typological features. However, this does not rule out the involvement of other factors. The typological features of the Tsat have evolved due to the combined influence of internal linguistic factors and external influences, including contact with languages such as Chinese, Tai-Kadai languages, and Hlai languages. The presence of an “Austronesian substratum”<sup>1</sup> suggests a historical connection between its ancestors and the Austronesian peoples. This view is further supported by the genomic evidence presented in this study. The interrelationships between languages—whether through cognates, borrowings, or contact—are closely linked to the historical relationships between their peoples and the degree of affinity between ethnic groups. This provides a crucial link for interdisciplinary research that integrates linguistics, genetics, molecular anthropology, and other fields. In conclusion, this paper builds upon the work of previous scholars and presents new insights into the SYH and the historical evolution of their language. This is achieved through a comparative analysis of the Tsat and Austronesian languages of China’s Taiwan region, with a particular focus on linguistic typological characteristics, as well as integrating findings from genetic and molecular anthropological studies.

In conclusion, this paper supports the hypothesis that the modern SYH originated from three ancestral sources in Central/South Asia, Southeast Asia, and East Asia, with the genetic contribution from East Asian populations being the primary factor. Both cultural and population diffusion drove the formation of the modern Hui ethnic group in East Asia. In addition to the general origin and evolutionary history of the Hui ethnic group, which has already been published, this study provides substantial genetic evidence to challenge the controversial view that the SYH are a direct descendant of the Cham people of Vietnam.

## References

- [1] Atkinson, Q. D., & Gray, R. D. (2005). Curious parallels and curious connections—phylogenetic thinking in biology and historical linguistics. *Systematic Biology*, 54, 513–526.
- [2] Cavalli-Sforza, L. L. (2001). *Genes, peoples, and languages* (First paperback printing ed.). Berkeley, Los Angeles, London: University of California Press.
- [3] Cavalli-Sforza, L. L., Piazza, A., Menozzi, P., & Mountain, J. (1988). Reconstruction of human evolution: bringing together genetic,

---

<sup>1</sup> Austronesian substratum, in certain non-Austronesian languages, they preserve the lexical, phonological, or grammatical features of early Austronesian languages.

- archaeological, and linguistic data. *Proceedings of the National Academy of Sciences of the United States of America*, 85(16), 6002.
- [4] Chen, N. X. (1989). The verb morphology of Wutunhua. *Minority Languages of China*, 26–37.
- [5] Darwin, C. (1859). *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. London, England: John Murray.
- [6] Deng, X. H., & Gao, T. J. (2014). The theories, approaches and practice of evolutionary linguistics. *Journal of Shanxi University (Philosophy & Social Science)*, 37(2), 72–75.
- [7] Fan, Z. Q., Deng, X. H., & Wang, C. C. (2019). On the evolution history of Tsat and Utsat from the perspective of linguistics and molecular anthropology. *Journal of Hui Muslim Minority Studies*, 29, 44–52.
- [8] Gray, R. D., & Atkinson, Q. D. (2003). Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature*, 426, 435–439.
- [9] Gray, R. D., & Jordan, F. M. (2000). Language trees support the express-train sequence of Austronesian expansion. *Nature*, 405, 1052–1055.
- [10] Gray, R. D., Drummond, A. J., & Greenhill, S. J. (2009). Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science*, 323, 479–483.
- [11] Haudricourt, A. G. (1984). The classification of the Hui language of the Hui people in Sanya, Hainan Island. *Minority Languages of China*, 17–25.
- [12] He, G., Fan, Z. Q., Zou, X., Deng, X., Yeh, H. Y., Wang, Z., Liu, J., Xu, Q., Chen, L., Deng, X. H., Wang, C. C., Liu, C. H., Wang, M., & Liu, C. (2023). Demographic model and biological adaptation inferred from the genome-wide single nucleotide polymorphism data reveal tripartite origins of southernmost Chinese Huis. *American Journal of Biological Anthropology*, 180, 488–505.
- [13] He, G., Wang, Z., Wang, M., Luo, T., Liu, J., Zhou, Y., Gao, B., & Hou, Y. (2018). Forensic ancestry analysis in two Chinese minority populations using massively parallel sequencing of 165 ancestry-informative SNPs. *Electrophoresis*, 39, 2732–2742.
- [14] Huang, Y. X. (2008). Hainan “Fan Ke” —The earliest Muslim researches in China. *Humanities & Social Sciences Journal of Hainan University*, 26, 601–606.
- [15] Jiang, Y., & Dong, X. J. (1992). *Islamic culture in Hainan: A series on Hainanese ethnic cultures*. Guangzhou: Sun Yat-sen University Press.
- [16] Li, B. J., & Zhang, X. (1999). The current state of research and theoretical exploration on Chinese contact languages. *Studies in Language and Linguistics*, 190–200.
- [17] Li, D. N., Wang, C. C., Yang, K., Qin, Z. D., Lu, Y., Lin, X. J., Li, H., & The G. Consortium (2013). Substitution of Hainan indigenous genetic lineage in the Utsat people, exiles of the Champa kingdom. *Journal of Systematic and Evolutionary Botany*, 51, 287–294.
- [18] Li, H., & Jin, L. (2015). *Y chromosome and the evolution of East Asian populations*. Shanghai: Shanghai Scientific & Technical Publishers.
- [19] Li, H., Wen, B., Chen, S. J., Su, B., Pramoongjago, P., Liu, Y., Pan, S., Qin, Z., Liu, W., Cheng, X., Yang, N., Li, X., Tran, D., Lu, D., Hsu, M. T., Deka, R., Marzuki, S., Tan, C. C., & Jin, L. (2008). Paternal genetic affinity between western Austronesians and Daic populations. *BMC Evolutionary Biology*, 8, 146.
- [20] Li, Y. (2008). *A study of word order typology in the languages of southern ethnic groups of China*. Beijing: Peking University Press.
- [21] Liu, Y., Yang, J., Li, Y., Tang, R., Yuan, D., Wang, Y., Wang, P., Deng, S., Zeng, S., Li, H., Chen, G., Zou, X., Wang, M., & He, G. (2021). Significant East Asian affinity of the Sichuan Hui genomic structure suggests the predominance of the cultural diffusion model in the genetic formation process. *Frontiers in Genetics*, 12.
- [22] Ma, X., Yang, W., Gao, Y., Pan, Y., Lu, Y., Chen, H., Lu, D., & Xu, S. (2021). Genetic origins and sex-biased admixture of the Huis. *Molecular Biology and Evolution*, 38, 3804–3819.
- [23] Meng, S. M. (1992). The three historical stages of the development of Austro-Tai languages: Indonesian, Rhade, and Tsat. *Studies in Language and Linguistics*, 8, 104–109.
- [24] Meng, S. M. (1995). The three historical stages of the development of Austro-Tai languages: Phonological and grammatical differences and connections between Indonesian and Tsat (Continued). *Studies in Language and Linguistics*, 176–181.
- [25] Ni, D. B. (1988). The classification of the Hui language of the Hui people in Sanya, Hainan Island. *Minority Languages of China*, 34, 18–25.

- [26] Pagel, M. (2009). Human language as a culturally transmitted replicator. *Nature Reviews Genetics*, 10, 405–415.
- [27] Sagart, L., Jacques, G., Lai, Y., Ryder, R. J., Thouzeau, V., Greenhill, S. J., & List, J. M. (2019). Dated language phylogenies shed light on the ancestry of Sino-Tibetan. *Proceedings of the National Academy of Sciences of the United States of America*, 116, 10317–10322.
- [28] Schleicher, A. (1861). *Compendium der vergleichenden grammatik der indogermanischen sprachen*. Weimar, Germany: Hermann Böhlman.
- [29] Shen, J. X. (1997). Markedness patterns in typology. *Foreign Language Teaching and Research*, 4–13.
- [30] Sun, Y. T., Yang, B., Ou, C. Y., Zhou, Z. J., Su, Z. Y., & Li, D. N. (2007). Origins of the three minority populations in Hainan Island as seen from Y-SNP. *Science & Technology Review*, 44–47.
- [31] Tian, X. S., & Chen, B. Y. (2019). *Tsat of Sanya, Hainan*. Beijing: The Commercial Press.
- [32] Wang, Q., Zhao, J., Ren, Z., Sun, J., He, G., Guo, J., Zhang, H., Ji, J., Liu, Y., Yang, M., Yang, X., Chen, J., Zhu, K., Wang, R., Li, Y., Chen, G., Huang, J., & Wang, C. C. (2021). Male-dominated migration and massive assimilation of indigenous East Asians in the formation of Muslim Hui people in southwest China. *Frontiers in Genetics*, 11, 618614.
- [33] Wen, S. Q., Xie, X. D., & Xu, D. (2013). Contact and admixture—the relationship between Dongxiang population and their language viewed from Y chromosomes: Contact and admixture—the relationship between Dongxiang population and their language viewed from Y chromosomes. *Hereditas (Beijing)*, 35, 761–770.
- [34] Xu, D. (2015). A new perspective in linguistic studies: Coevolution of languages and genes. *Contemporary Linguistics*, 17, 215–226, 252.
- [35] Xu, D. (2018). Mixed languages in China and language mixing mechanism. *Chinese Journal of Language Policy and Planning*, 3, 59–79.
- [36] Xu, T. Q. (1991). *Historical linguistics*. Beijing: The Commercial Press.
- [37] Zeng, X. Y., & Yin, S. W. (2011). A re-examination of the nature and characteristics of Tsat. *Minority Languages of China*, 17–25.
- [38] Zhang, H. Y. (2018). The translation of your paper title would be: The main criteria for genealogical classification: A case study of Huihuihua. *Minority Languages of China*, 86–97.
- [39] Zhang, H., Ji, T., Pagel, M., & Mace, R. (2020). Dated phylogeny suggests early Neolithic origin of Sino-Tibetan languages. *Scientific Reports*, 10, 20792.
- [40] Zhang, M., Yan, S., Pan, W., & Jin, L. (2019). Phylogenetic evidence for Sino-Tibetan origin in northern China in the late Neolithic. *Nature*, 569, 112–115.
- [41] Zheng, Y. Q. (1986). Reconsidering the status of Tsat. *Minority Languages of China*, 37–42.
- [42] Zheng, Y. Q. (1997). *A study of Tsat*. Shanghai: Shanghai Far East Publishers.