

Paternal gene pool of Malays in Southeast Asia and its applications for the early expansion of Austronesians

Jin Sun¹  | Lan-Hai Wei^{1,2} | Ling-Xiang Wang³ | Yun-Zhi Huang⁴ | Shi Yan⁵ | Hui-Zhen Cheng¹  | Rick Tzee-Hee Ong⁶ | Woei-Yuh Saw^{6,7} | Zhi-Quan Fan¹ | Xiao-Hua Deng^{1,8} | Yan Lu⁹ | Chao Zhang^{9,10} | Shu-Hua Xu^{9,10,11} | Li Jin^{4,5} | Yik-Ying Teo^{6,7,12,13,14} | Hui Li^{4,2,5}

¹Department of Anthropology and Ethnology, Institute of Anthropology, Xiamen University, Xiamen, China

²B&R International Joint Laboratory for Eurasian Anthropology, Fudan University, Shanghai, China

³Enlighten Co., Ltd., Shanghai, China

⁴MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai, China

⁵Human Phenome Institute, Fudan University, Shanghai, China

⁶Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore

⁷Life Sciences Institute, National University of Singapore, Singapore, Singapore

⁸Center for collation and studies of Fujian local literature, Fujian University of Technology, Fuzhou, China

⁹Chinese Academy of Sciences (CAS) Key Laboratory of Computational Biology, Max Planck Independent Research Group on Population Genomics, CAS-MPG Partner Institute for Computational Biology (PICB), Shanghai Institute of Nutrition and Health, Shanghai Institutes for Biological Sciences, University of Chinese Academy of Sciences, CAS, Shanghai, China

¹⁰School of Life Science and Technology, Shanghai Tech University, Shanghai, China

¹¹Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, China

¹²Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore, Singapore

¹³NUS Graduate School for Integrative Science and Engineering, National University of Singapore, Singapore, Singapore

¹⁴Department of Statistics and Applied Probability, National University of Singapore, Singapore, Singapore

Correspondence

Yik-Ying Teo, Life Sciences Institute, National University of Singapore, 21 Lower Kent Ridge Road, Singapore 119077, Singapore.
Email: statyy@nus.edu.sg

Hui Li, MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, #2005 Songhu Rds., Shanghai, China.
Email: lhca@fudan.edu.cn

Funding information

Key Research Program of Frontier Sciences, Grant/Award Number: QYZDJ-SSW-SYS009; National Key Research and Development Project, Grant/Award Number: 2016YFC0900300; Chinese Academy of Sciences, the National Key

Abstract

Objectives: The origin and differentiation of Austronesian populations and their languages have long fascinated linguists, archeologists, and geneticists. However, the founding process of Austronesians and when they separated from their close relatives, such as the Daic and Austro-Asiatic populations in the mainland of Asia, remain unclear. In this study, we explored the paternal origin of Malays in Southeast Asia and the early differentiation of Austronesians.

Materials and Methods: We generated whole Y-chromosome sequences of 50 Malays and co-analyzed 200 sequences from other Austronesians and related populations. We generated a revised phylogenetic tree with time estimation.

Results: We identified six founding paternal lineages among the studied Malays samples. These founding lineages showed a surprisingly coincident

Jin Sun, Lan-Hai Wei, and Ling-Xiang Wang contributed equally to the study.

Research and Development Program, Grant/Award Number: 2016YFC0906403; Fundamental Research Funds for the Central Universities, Grant/Award Number: 20720191047; National Natural Science Foundation of China, Grant/Award Numbers: 91731303, 31525014, 31671297, 31771388, 31900406, 31961130380; Scientific and Technology Committee of Shanghai Municipality, Grant/Award Number: 18490750300; Strategic Priority Research Program, Grant/Award Number: XDB13040100; UK Royal Society-Newton Advanced Fellowship, Grant/Award Number: NAF\R1\191094

expansion age at 5000 to 6000 years ago. We also found numerous mostly close related samples of the founding lineages of Malays among populations from Mainland of Asia.

Conclusion: Our analyses provided a refined phylogenetic resolution for the dominant paternal lineages of Austronesians found by previous studies. We suggested that the co-expansion of numerous founding paternal lineages corresponds to the initial differentiation of the most recent common ancestor of modern Austronesians. The splitting time and divergence pattern in perspective of paternal Y-chromosome evidence are highly consistent with the previous theories of ethnologists, linguists, and archeologists.

1 | INTRODUCTION

The origin and diffusion process of Austronesians has long fascinated ethnologists, linguists, archeologists, and geneticists. The Malays are one of the Austronesian groups predominantly present in Southeast Asia. It is generally believed that direct ancestral groups of Austronesians appeared in the western coastal region of Taiwan about 6000 to 5000 years ago (Bellwood, 2017). Scholars have suggested that the Dabengkeng culture (about 3000-2500 BC), the earliest archeological culture in Taiwan, and a series of subsequent archeological cultures were created by the ancestor groups of modern Austronesians (Bellwood, 2000; Blust, 1976). Between 5000 and 4000 years before the present, the ancestors of Austronesian-speaking populations spread from Taiwan to Southeast Asia and the western Pacific (Gray, Drummond, & Greenhill, 2009; Lipson et al., 2014; Ross, 2005). Eventually, they scattered in a vast region ranging from the Madagascar Island of Africa in the west to the Easter Island (Rapa Nui) of the Pacific Ocean in the east (Bellwood, 1986). Some scholars speculate that the ancient Austronesian groups might have originated from the Neolithic populations in the southeast coastal areas of China. The remains of the Dabengkeng culture show some affiliation to the cultural elements of the Hemudu culture in Zhejiang, China and some archeological sites along the coast of Fujian province (Chang & Goodenough, 1996; Li, 2013). However, current archeological evidence is not yet clear enough to describe in detail the origin of the Austronesian ancestor groups in the southeastern coastal areas of mainland China (Jiao, 2007).

Previous studies have revealed the high diversity in the paternal gene pool of Austronesians (Delfin et al., 2012; Heyer, Georges, Pachner, & Endicott, 2013; Karafet et al., 2010; Kayser et al., 2003; Kayser

et al., 2006; Kayser et al., 2008; Li et al., 2008; Loo et al., 2011; Mirabal et al., 2012; Scheinfeldt et al., 2006; Scholes et al., 2011; Su et al., 2000; Trejaut et al., 2014; Trejaut et al., 2019; Wei et al., 2017). Haplogroups O1a-M119 and O2a2b2-N6 are the predominant paternal lineages of Taiwanese aborigines (Trejaut et al., 2014). Particularly, O1a-M119 contributes to a very high proportion (>90%) among some populations of Taiwanese aborigines. For the Austronesian-speaking populations in Southeast Asia, there are many diverse lineages that can be traced to local communities since the Paleolithic Age (Karafet et al., 2010). However, the sub-lineages of O-M175 contribute to the majority of the paternal gene pool of these populations and these lineages were likely introduced by migrations from East Asia since the Neolithic Age (Lipson et al., 2018). For the Austronesians from the Pacific Ocean region, there are two predominant paternal lineages. One of these is the haplogroup O2a2b2-N6 which may have originated in the eastern coastal region of East Asia (Wei et al., 2017). This haplogroup was determined as O3*-M122+, M324+, P201+, P164+, 002611-, M7-, and M134- in most previous studies (Delfin et al., 2012; Karafet et al., 2010; Kayser et al., 2003; Kayser et al., 2006; Kayser et al., 2008; Loo et al., 2011; Mirabal et al., 2012; Scheinfeldt et al., 2006; Scholes et al., 2011; Trejaut et al., 2014). The other haplogroups, like C2-M38 and K-M9, may be a result of admixture between ancient populations diffused from Taiwan and local populations across the island region of Southeast Asia and the Pacific Ocean. Besides, the analyses of Soares et al. (2016) had identified some candidate founding paternal lineages for Austronesians and discussed their demographic history, like haplogroup O1a2-M50 (M110), a sub-lineage of O1a-M119 (Soares et al., 2016). Ancient DNA study revealed paternal lineages O1a-M119, O2a2b2-N6, K-M9, and M-P256 among ancient South Pacific individuals that may be ancestor of

modern Austronesians in the same region (Posth et al., 2018). These results strongly support the observation from modern DNA. In general, scholars have revealed that there are many sharing components in gene pool of Austronesians from different geographic regions. However, significant difference can also be detected in gene pool of these populations.

In the past decades, numerous works had been to reveal the maternal genetic heritage of Austronesians and the demographic process of their early origin and differentiation (Brandão et al., 2016; Delfin et al., 2012; Loo et al., 2014; Posth et al., 2018; Pugach et al., 2018; Soares et al., 2008; Soares et al., 2016; Tabbada et al., 2010; Trejaut et al., 2005). Importantly, the analyses of Brandão et al. and Soares et al. had identified numerous founding maternal lineages for Austronesians and the diffusion pattern of these founding lineages insisted to reveal a clear demographic history of Austronesians in perspective of maternal genetic components (Brandão et al., 2016; Soares et al., 2016). In summary, most previous studies of Y-chromosome were based on the genotyping results of a small number of Y-single nucleotide polymorphism (Y-SNP) markers, rather than sequencing data. Additional studies with sequence data are required to explore the detailed origin of proto-Austronesian communities in the coastal region of East Asia, as well as the early differentiation process of paternal lineages of Austronesians.

In this study, we generated whole Y-chromosome sequences of 50 Malays and co-analyzed 200 sequences from other Austronesians and related populations. Our objectives were to (a) reconstruct a phylogeny with age estimation for paternal lineages of Malay and (b) identify the founding paternal lineages of Malay population and their expansion time. In particular, we explored the formation of proto-Austronesians and the initial expansion process of Austronesians. Overall, we generated a refined phylogenetic tree with accurate dating for founding paternal lineages of Malay to explore the early history of Austronesians.

2 | MATERIALS AND METHODS

Blood samples from populations across East Asia were collected from unrelated healthy males after obtaining informed consent (Table S1). The ethics committee for biological research at the School of Life Sciences in Fudan University approved the study. Full genome sequencing and sequencing of target Y-chromosomes of 15 new samples were performed using the Illumina HiSeq 2000 (San Diego, California). We used the procedures that we previously described for the other steps

prior to next-generation sequencing for DNA shearing, adding an adaptor, gel electrophoresis, library preparation, bait design, and liquid-phase capture technology (Yan et al., 2014). The raw sequence data reported in this article has been deposited in the Genome Sequence Archive (Y. Wang et al., 2017) in the Beijing Institute of Genomics (BIG) Data Center, (Members, 2017), Chinese Academy of Sciences, under accession number PRJCA001717 that is publicly accessible at <http://bigd.big.ac.cn/gsa>. The raw data of 50 Malay males are available per the request to the corresponding authors.

The Y-chromosome sequences of 50 Malays were extracted from whole genomes sequence of SGVP project we reported in previous publications (Wong et al., 2013). To obtain a comprehensive overview of this paternal lineage, available Y-chromosome data from publications were also included. In total, 200 sequences were analyzed, including 15 sequences came from our lab, 61 from SGVP (38) (Wong et al., 2013), 73 from 1000 Genomes Project (Poznik et al., 2016), 44 from SGDP (Mallick et al., 2016), and 7 from PICB (Lu et al., 2016). For the samples from SGDP project, all samples from East Asia and Southeast Asia were included. For the samples from 1000 Genomes Project and PICB, we only selected the samples that are related to major lineages of Austronesians (O1a-M119, O1b1a1a-M95, O2a2a1a2-M7, and O2a2b2-N6), so as to generate a revised phylogenetic tree with better resolution. All analyses in this study were conducted by using reference hg19. We referred to www.isogg.org (www.isogg.org 15.34 Date: February 3, 2020) for the assignment of the SNP and haplogroup name (Y Chromosome Consortium, 2002). The genotype of variants of studied sequences can be found in Table S2. The analyzed region of Y-chromosome in this study was listed in Table S4 and the depth and coverage length of newly generated sequences are listed in Table S5.

Mapping of reads and SNP (single-nucleotide polymorphism) calling were conducted with the human reference genome hg19 using GATK packages and standard procedures for next-generation sequencing (Li et al., 2009; Li & Durbin, 2010). To obtain a confident Y-SNP dataset for reconstruction of phylogenetic tree and age estimation, we applied a series of strict filters on the original variants file, including: (a) restriction to confident 8.4 Mb region (Adamov, Guryanov, Karzhavin, Tagankin, & Urasin, 2015) which is a subset of the studied region of Poznik et al. (2016) and Karmin et al. (2015); (b) restriction to variants that are Y-SNP; (c) removal of all positions with call rate <80% on all samples with sequences; (d) removal of position with heterozygosity call rate >5% on all samples; (e) base coverage ≥ 3 , base quality >20, and distance between SNPs >10 bp; and (f) removal of recurrent or triadic mutations.

Bayesian evolutionary analyses and age estimation were implemented in BEAST (v2.0.0) (Bouckaert et al., 2014). We selected a Bayesian skyline coalescent tree prior and a strict clock. Traditional substitution models, like GTR (Rodríguez, Oliver, Marín, & Medina, 1990) and HKY + G (Hasegawa, Kishino, & Yano, 1985), were widely used in previous studies of mtDNA sequences (Macaulay et al., 2005). However, we applied bModelTest package which allows the BEAST program to infer the most optimistic substitution model for input sequences (Bouckaert & Drummond, 2017).

In most of previous studies about Y-chromosome sequences, a universal mutation rate was applied for evolutionary history of all haplogroups in the past 100 000 years (Poznik et al., 2016). In particular, there are several popularly used Y chromosomal substitution rates, such as the evolutionary rates measured from human-chimpanzee comparisons (Kuroki et al., 2006; Thomson, Pritchard, Shen, Oefner, & Feldman, 2000), the genealogical rate observed in a deep-rooting pedigree (Xue et al., 2009), the rate adjusted from autosomal mutation rates (Mendez et al., 2013), the rates based on archeological evidence of founding migrations (Francalacci et al., 2013; Poznik et al., 2016), and a mass collection of deep-rooting pedigrees from Iceland populations (Helgason et al., 2015). Recently, the ancient sequences of 45 000-year-old Ust'Ishim male from western Siberia (Fu et al., 2014), the 24-ky-old Siberian individual with haplogroup R (Raghavan et al., 2014), and the 12.6-ky-old Anzick infant of Q-L54 (Rasmussen et al., 2014) were used as recalibration point of the mutation rate of the whole human Y-chromosome phylogenetic tree (Bergstrom et al., 2016; Ilumäe et al., 2016; Karmin et al., 2015; Poznik et al., 2016). Although using the Y chromosomal molecular clock in time estimation is widely accepted, there are still many ongoing debates about the most suitable substitution rate in demographic and lineage dating (Elhaik, Tatarinova, Klyosov, & Graur, 2014; Wang, Jin, & Li, 2014; Wilson Sayres, Lohmueller, & Nielsen, 2014). Furthermore, we found that the mutation rates of some haplogroups are faster than that of other haplogroups. The different mutation among different paternal lineages, mtDNA lineages, and whole genomes may be universal for human populations from different geographic region of the world (Harris, 2015). Therefore, a recalibration point at the root of tree may help to provide more accurate results of splitting time for different haplogroups in different history period. Thus, an age of 41 900 years for haplogroup NO (95% CI = 40 175-43 359 years) was used for calibration of age estimation in this study (Karmin et al., 2015). The calculation was performed with 20 million iterations and sampling every 10 000 steps. Results were visualized in Tracer

v.1.6 (Rambaut, Drummond, Xie, Baele, & Suchard, 2018) and FigTree v1.4.2 (Rambaut, 2018), with a burn-in of 30% and over 200 effective sample sizes.

3 | RESULTS

3.1 | Paternal gene pool of Malays and Southeast Asia populations

In this study, we analyzed Y-chromosome sequences of 200 males from Malays and related populations (Figure 1 and Table S1). In total, the Malays samples contributed to 50 individuals among the 56 Austronesians samples. We constructed a refined phylogenetic tree for all studied samples (Figure S1) and used 127 samples with good sequencing depth from SGVP, SGDP, PICB, and our lab to calculate the diverging time of downstream sub-lineages (Tables S1 and S2, Figure S2). The revised phylogenetic tree and age estimation may help to explore the paternal origin and founding process of Malays and Austronesians populations.

We found that a number of rare lineages diverged from each other before 45 kya, including C1b1a2*-B65, C1b1a2 b-F725, C1b2a-M38, F2-M427, S-B254, and P*-M45 (Figures 2 and S2, and Table S1). Samples from these lineages contribute to 18% (9/50) of the paternal gene pool of modern Malays (Tables S1 and S2). We categorized the rare lineages observed in modern Southeast Asia populations as "Paleolithic lineages." Our observation is consistent with the conclusion of previous studies, which demonstrated that the haplogroups C-M130, F-M89, and K-M9 diverged rapidly in a short interval between 47 and 52 kya (Bergstrom et al., 2016; Karmin et al., 2015; Malaspinas et al., 2016; Poznik et al., 2016). The newly emerging lineages during this period eventually became the dominant paternal lineages of modern human in Eurasia, Oceania, and America (Bergstrom et al., 2016; Karmin et al., 2015; Malaspinas et al., 2016; Poznik et al., 2016). Four samples (8%, 4/50) belong to haplogroup J1a2b-P58, J2a4b2-M166, J2b-M102, and R2-M124, respectively (Table S1). Previous studies showed that these lineages originated in either Middle East or South Asia (Poznik et al., 2016; Sengupta et al., 2006). These lineages may have been introduced into Southeast Asia in a historical age (Karafet et al., 2005).

The major part of the paternal gene pool of Malays consisted of various sub-lineages of haplogroup O-M175 (84%, 42/50, see Table S1). Samples from sub-lineages O-M175 contributed to 100% of the paternal gene pool of

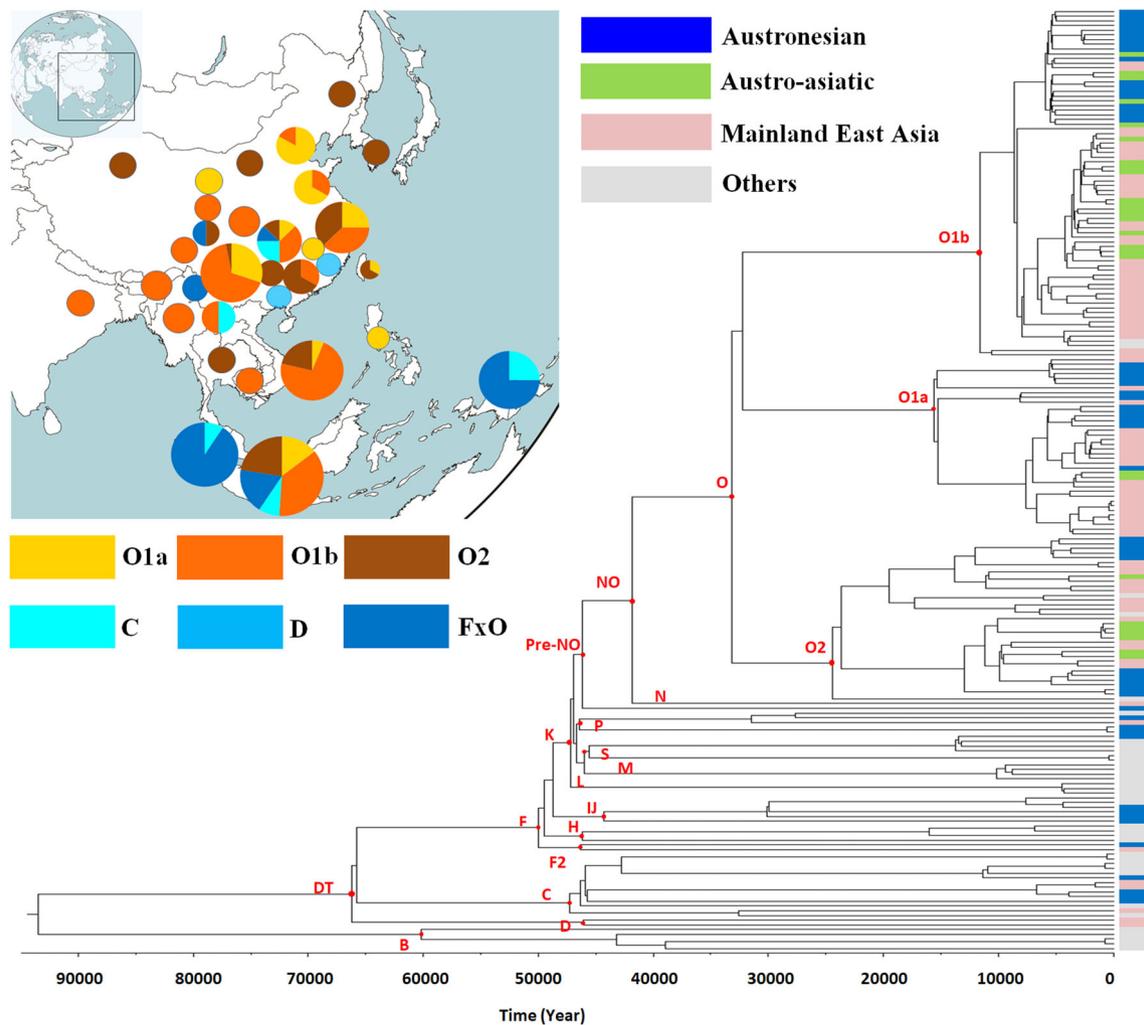


FIGURE 1 Phylogenetic tree of Y-chromosome sequences and location of samples used in this study. Note: The size of circle on the map corresponds to the number of samples in each place. Pie charts were used to indicate the brief distribution of Y haplogroups in the analyzed samples on the map

Austro-Asiatic populations in this study. The refined phylogenetic tree in this study helps to break down the samples of O-M175 into 32 sub-lineages (Table S1). As shown in Figure 2, Figure S1, and Figure S2, we identified a number of special paternal lineages for Malays, Austro-Asiatic populations, and other populations from the mainland of East Asia. The top six frequent lineages of Malays were O2a2b2-N6-B451, O2a2a1a2-M7-F14807, O1a2-M110-AM01446, O1a-M119-YP4610, O1b1a1a-M95-M1283-B418, and O1b1a1a-M95-M1283-Y9322. For Austro-Asiatic populations, O2a2a1a2-M7-V3237, O2a1-M88, and O2a-M95-M1283-Y9322 were the top four frequent lineages. Previous studies indicated that sub-lineages of O-M175 may have been introduced into Southeast Asia by migrations in Neolithic Age from Southern East Asia (Li et al., 2008; Lipson et al., 2018). Therefore, we categorized all sub-lineages of O-M175 as “Neolithic lineages” for Southeast Asian populations.

3.2 | Founding paternal lineages for Malays and Austronesians

According to the age estimations of this study (Figure 2, Figure S2 and Table S3), a surprising consistency was observed for the time to the most recent common ancestor (TMRCA) of the top six frequent lineages of Malays, including O2a2b2-N6-B451, O2a2a1a2-M7-F14807, O1a-M119-YP4610, O1a2-M110-AM01446, O1b1a1a-M95-M1283-B418, and O1b1a1a-M95-M1283-Y9322. As shown in Figure 2 and Table S3, except for the lineage O1a-M119-YP4610, the TMRCA of the other five lineages falls in a short interval between 5.5 and 5.8 kya. Additionally, lineage O1a-M119-YP4610 (6956, 95% CI = 6162-7797) contains two sub-lineages O1a-M119-B386 (4179, 95% CI = 3400-5066) and O1a-M119-SK1555 (3340, 95% CI = 2560-4159). Therefore, lineage O1a-M119-B386 and O1a-SK1555 may have emerged in any age between 3.3 and 6.9 kya. Given the fact

3.3 | Sharing of young lineages between Malays and other populations

As shown in Figure S1, there are also many lineages that are shared between Austronesians (mainly Malays), Austro-Asiatic populations, and other populations from the mainland of East Asia (see also Figure S2 and Table S2). For lineage O1a2-M110-AM01446 in Austronesians, the closest sample on the phylogenetic tree is sample HN1-95 from the Dong ethnic group in the Hunan province in South China. Sample S_Igorot-2 from the Philippines is close to lineage O1a-M119-F492 which flourished in the southeast region of China. Figure S1 also shows that samples of O1b1a1a-M95-M1283 in Austronesians are very close to that of Austro-Asiatic populations. Additionally, many samples of Austro-Asiatic populations have recent common ancestors with that of other populations from the mainland of East Asia (Figures S1 and S2, and Table S2). The sharing of young lineages will help to explore the recent differentiation process of modern populations.

4 | DISCUSSION

4.1 | Paternal origin of Malays and Austronesians

In this study, we identified several unique paternal lineages for Malays and discovered a surprising consistency between the expansion time of major parts of these unique lineages at about 5 to 6 kya. The age estimation and the distribution of these lineages in modern populations helped us investigate the paternal origin of Malays and Austronesians. Our analyses are consistent with the finding of a peak of migration to Island Southeast Asia at 4 to 5 kya by Soares et al., 2016 (Soares et al., 2016), after the initial differentiation of Austronesians.

For O2a2a1a2-M7-F14807, O1b1a1a-M95-M1383-B418, and O1b1a1a-M95-M1283-Y9322 in Malays and Austronesians, close relatives can be found in populations in the mainland of Southeast Asia and mainland of East Asia (Figures S1 and S2, and Table S1). Previous studies have indicated that O2a2a1a2-M7 and O1b1a1a-M95 are two predominant lineages in Austro-Asiatic populations (Cai et al., 2011; Li et al., 2008; Lipson et al., 2018). In addition, these two lineages are rare or absent in other Austronesians from Taiwan and Oceania (Delfin et al., 2012; Trejaut et al., 2014). The genetic scenario of O2a2a1a2-M7 and O1b1a1a-M95 is consistent with previous archeological, linguistic, and genetic studies. The sharing of a recent common ancestor of O2a2a1a2-M7 and O1b1a1a-M95 between Austronesian and Austro-Asiatic populations is also

consistent with previous studies on the close relationship between west Austronesians and populations from mainland Southeast Asia (Anderson, 2005; Blench, 2010; Higham, 2004; Lipson et al., 2014; Lipson et al., 2018; The HUGO Pan-Asian SNP Consortium et al., 2009; Vallée, Luciani, & Cox, 2016). Therefore, we proposed that lineages O2a2a1a2-M7 and O1b1a1a-M95 were introduced into west Austronesians through mainland Southeast Asia by ancient populations who are close to ancestors of Austro-Asiatic populations (Lipson et al., 2018). During the later expansion process of ancient populations from Taiwan, populations with high frequencies of O2a2a1a2-M7 and O1b1a1a-M95 in the west Indonesia Islands may have been assimilated and eventually became an important part of modern Malays and Austronesians.

Based on a comparison with other Y-chromosome studies, we proposed that lineages O2a2b2-N6-B451, O1a2-M110-AM01446, and O1a-M119-YP4610 may be the dominant paternal lineages of a direct ancestor of Austronesians in the Neolithic Age of Taiwan. Our results were supported by the findings of O2a2b2-N6 and O1a-M119 lineages among ancient South Pacific individuals that may be ancestor of modern Austronesians in the same region (Posth et al., 2018). The age estimation of TMRCA of O2a2b2-N6-B451, O1a2-M110-AM01446, and O1a-M119-YP4610 generally fall in the same historical period (Figure 2 and Table S3). Previous studies have suggested that the appearance of O2a2b2-N6 in Austronesians may have originated from the introduction of fox-tail millet into Neolithic Age Cultures in Taiwan. (Delfin et al., 2012; Mirabal et al., 2012; Trejaut et al., 2014; Wei et al., 2017). Furthermore, lineages O1a*-M119 and O1a2-M110 are predominant paternal lineages in Taiwan Aborigines and other Austronesians from the Philippines, Indonesia, and Near Oceania (Delfin et al., 2012; Karafet et al., 2010; Trejaut et al., 2014). However, the frequencies of O2a2b*-P164 (nearly equal to O2a2b2-N6-B451), O1a2-M110, and O1a*-M119 varied largely among all Austronesians. This pattern can be explained by the possible matriarchal society structure of ancient Austronesians and the continuous bottleneck effects during the expansion of Austronesians in a vast geographic region.

4.2 | Possible origin of proto-Austronesians

As discussed in the above section, there may be four dominant lineages in the ancestor group of Austronesians, including O2a2b2-N6-B451, O1a2-M110-AM01446, O1a-M119-YP4610, and O1a-M119-F819. As proposed in our previous study, lineage O2a2c2-N6-B451 was likely brought by an ancient population from the east coastal

region of North China, where the foxtail millet was the dominant plant-based food since the beginning of the Neolithic Age (Wei et al., 2017). Moreover, the delta region of the Yangtze River is a vital region on the migration route from the eastern coastal region of North China to Taiwan. Furthermore, archeology studies discovered a great affiliation between the Majiabang Culture and Dabenkeng Culture in Taiwan (Chang & Goodenough, 1996). Therefore, we suggested that the ancient populations diffused from north of the Yangtze River with the dominant paternal lineage O2a2b2-N6-B451 may have intensively admixed with local populations in the delta region of the Yangtze River who consisted of varied sub-branches of O1a-M119. The mixed population continued to migrate to the South and reached the coastal areas of Fujian Province between 6 and 5 kya. The proposed origins and migration routes of Austronesians and their ancestors here were showed on Figure 3.

On the other hand, it is worth noting the consistency of similar TMRCA between sub-lineages of O1b1a1a-M95 and other founding lineages of Austronesians. Ancient DNA studies suggested that haplogroup O1b1a1a-M95 is one of the most predominant paternal lineages in the Neolithic age of South China and Southeast Asia (Li et al., 2007; Lipson et al., 2018). According to archeological studies, developed rice farming with well-

organized farmland appeared firstly in the middle reaches of the Yangtze River at about 6.5 kya (Pei, Church, & Chung, 1998) and spread widely across Southern China between 6.5 and 5 kya (Gao, Dong, Yang, & Chen, 2020; Nasu, Gu, Momohara, & Yasuda, 2012), including the Fujian Province (Yang et al., 2018). This process may correspond to the diffusion of paternal lineage O1b1a1a-M95 across South China and further to Southeast Asia in later age, thus bringing paternal lineage O1b1a1a-M95 to Fujian province.

Two waves of immigrants, with developed rice farming and millet farming, respectively, may have coincidentally reached Fujian province the same period between 6 and 5 kya. On the other hand, a recent ancient DNA study suggested the early Neolithic communities, representing by the 7.4 thousand-year-old Qihe2 individual and two Liangdao individuals (8.2 and 7.5 thousand years old, respectively), had also contributed to the genetic pool of modern Austronesians (Ko et al., 2014; Yang et al., 2020). Therefore, we proposed that the admixture between ancient populations diffused from two different regions of South China and local populations across may have eventually given birth to the proto-Austronesians with paternal lineages O1a-M119, O2a2b2-N6, and O1b1a1a-M95 in coastal region of Fujian (Figure 3). Parts of this proto-Austronesians

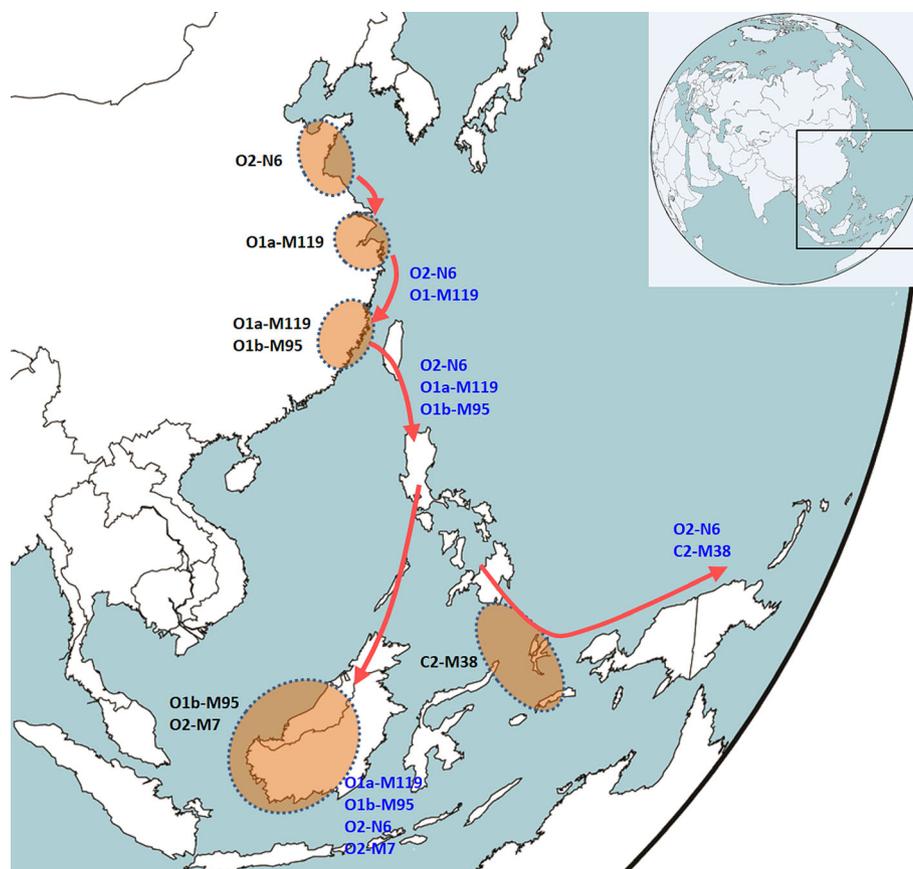


FIGURE 3 The proposed origins and migration routes of Austronesians and their ancestors. The dashed circles were used to show the possible origin or admixture events. The Y-single nucleotide polymorphisms (Y-SNPs) in black indicate the possible dominant lineages of local populations while the Y-SNPs in purple indicate the possible primary lineages of Austronesians or their ancestors after the admixture. Short names were applied for all lineages

community moved to Taiwan and triggered the great dispersal of Austronesians in the next thousands years (Yang et al., 2018).

The scenario we described here help to explain the similar TMRCA between sub-lineages of O1b1a1a-M95 and other founding lineages of Austronesians, as well the low frequencies, but widely distributed, of O1b1a1a-M95 in Taiwan aborigines, the earliest branches of Austronesians. The source of O1b1a1a-M95 in Taiwan aborigines may be different from that of O1b1a1a-M95 among Austronesians in Maritime Southeast Asia which are likely results of admixture in later age (Figure 3). However, more analyses on ancient DNA and modern DNA are needed to verify our hypothesis here.

In summary, we constructed a high-resolution phylogeny of paternal lineages in Malays. We determined several founding paternal lineages that corresponded directly to the origin and early differentiation of Malays and Austronesians. We discovered a surprising consistency between the expansion times of major parts of these founding lineages in a short period between 6 and 5 kya. We also found a number of samples from the mainland of East Asia, which are the most closely related to the founding paternal lineages of Austronesians. Furthermore, we discussed the possible ancestor groups, admixture process, and migration routes for proto-Austronesians. Studies on maternal mitochondrial DNA (mtDNA), ancient DNA, and whole genome data in the future will help reveal a clear scenario for the origin, migration, and admixture history of populations in East Asia and Southeast Asia.

ACKNOWLEDGMENTS

The authors are grateful to all sample donors. Y.-Y. T. acknowledges the support of the Saw Swee Hock School of Public Health, the National University Health System, and the Life Sciences Institute at the National University of Singapore. This work was supported by the National Natural Science Foundation of China (31900406 to L.-H. W.; 31671297 and 91731303 to H. L.; 91731303, 31525014, 31771388, and 31961130380 to S.-H. X.). S.-H. X. was also supported by the UK Royal Society-Newton Advanced Fellowship (NAF\R1\191094); the Strategic Priority Research Program (XDB13040100); and Key Research Program of Frontier Sciences (QYZDJ-SSW-SYS009) of the Chinese Academy of Sciences, the National Key Research and Development Program (2016YFC0906403). L.-H. W. was also supported by the Fundamental Research Funds for the Central Universities (No. 20720191047). This study was also supported by the Scientific and Technology Committee of Shanghai Municipality (No. 18490750300). H. L. was also supported by National Key Research and Development Project (2016YFC0900300).

CONFLICTS OF INTEREST

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

AUTHOR CONTRIBUTIONS

Lan-Hai Wei: Conceptualization; data curation; formal analysis; software; visualization; writing-original draft; writing-review and editing. **Ling-Xiang Wang:** Conceptualization; data curation; formal analysis; software; visualization. **Yun-Zhi Huang:** Data curation; formal analysis. **Shi Yan:** Data curation; formal analysis. **Hui-Zhen Cheng:** Data curation; formal analysis; software; validation; visualization. **Rick Twee-Hee Ong:** Data curation; formal analysis; resources. **Woei-Yuh Saw:** Data curation; formal analysis; resources. **Zhi-Quan Fan:** Investigation; methodology. **Xiao-Hua Deng:** Conceptualization; investigation; methodology. **Yan Lu:** Data curation; formal analysis. **Chao Zhang:** Data curation; formal analysis. **Shu-Hua Xu:** Data curation; formal analysis. **Li Jin:** Formal analysis; methodology. **Hui Li:** Conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; project administration; resources; validation; writing-original draft; writing-review and editing.

ORCID

Jin Sun  <https://orcid.org/0000-0003-0348-6715>

Hui-Zhen Cheng  <https://orcid.org/0000-0002-3965-9895>

REFERENCES

- Adamov, D., Guryanov, V., Karzhavin, S., Tagankin, V., & Urasin, V. (2015). Defining a new rate constant for Y-chromosome SNPs based on full sequencing data. *The Russian Journal of Genetic Genealogy*, 7(1), 68–89.
- Anderson, A. (2005). Crossing the Luzon Strait: Archaeological chronology in the Batanes Islands, Philippines and the regional sequence of Neolithic dispersal. *Journal of Austronesian Studies*, 1(2), 25–45.
- Bellwood, P. (1986). *Prehistory of the Indo-Malaysian archipelago*. Orlando, FL: Academic Press.
- Bellwood, P. (2000). Formosan prehistory and Austronesian dispersal. In D. Blundell (Ed.), *Austronesian Taiwan: Linguistics, history, ethnology, prehistory* (pp. 337–365). Berkeley, CA: The Regents of the University of California.
- Bellwood, P. (2017). *First islanders: Prehistory and human migration in Island Southeast Asia*. Hoboken, NJ: Wiley-Blackwell.
- Bergstrom, A., Nagle, N., Chen, Y., McCarthy, S., Pollard, M. O., Ayub, Q., ... Tyler-Smith, C. (2016). Deep roots for aboriginal Australian Y chromosomes. *Current Biology*, 26(6), 809–813. <https://doi.org/10.1016/j.cub.2016.01.028>
- BIG Data Center Members. (2017). The BIG data center: From deposition to integration to translation. *Nucleic Acids Research*, 45(D1), D18–D24. <https://doi.org/10.1093/nar/gkw1060>
- Blench, R. (2010). Was there an Austroasiatic presence in Island Southeast Asia prior to the Austronesian expansion? *Bulletin of the Indo-Pacific Prehistory Association*, 30, 133–144.

- Blust, R. (1976). Austronesian culture history: Some linguistic inferences and their relations to the archaeological record. *World Archaeology*, 8, 19–43.
- Bouckaert, R., Heled, J., Kuhnert, D., Vaughan, T., Wu, C. H., Xie, D., ... Drummond, A. J. (2014). BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, 10(4), e1003537. <https://doi.org/10.1371/journal.pcbi.1003537>
- Bouckaert, R. R., & Drummond, A. J. (2017). bModelTest: Bayesian phylogenetic site model averaging and model comparison. *BMC Evolutionary Biology*, 17(1), 42. <https://doi.org/10.1186/s12862-017-0890-6>
- Brandão, A., Eng, K. K., Rito, T., Cavadas, B., Bulbeck, D., Gandini, F., ... Soares, P. (2016). Quantifying the legacy of the Chinese Neolithic on the maternal genetic heritage of Taiwan and Island Southeast Asia. *Human Genetics*, 135(4), 363–376. <https://doi.org/10.1007/s00439-016-1640-3>
- Cai, X., Qin, Z., Wen, B., Xu, S., Wang, Y., Lu, Y., ... Genographic, C. (2011). Human migration through bottlenecks from Southeast Asia into East Asia during last glacial maximum revealed by Y chromosomes. *PLoS One*, 6(8), e24282. <https://doi.org/10.1371/journal.pone.0024282>
- Chang, K.-C., & Goodenough, W. H. (1996). Archaeology of Southeastern Coastal China and Its Bearing on the Austronesian Homeland. In *Prehistoric settlement of the Pacific Transactions of the American Philosophical Society new series* (Vol. 85, pp. 36–56). Philadelphia, PA: American Philosophical Society.
- Delfin, F., Myles, S., Choi, Y., Hughes, D., Illek, R., van Oven, M., ... Stoneking, M. (2012). Bridging near and remote Oceania: mtDNA and NRY variation in the Solomon Islands. *Molecular Biology and Evolution*, 29(2), 545–564. <https://doi.org/10.1093/molbev/msr186>
- Elhaik, E., Tatarinova, T. V., Klyosov, A. A., & Graur, D. (2014). The 'extremely ancient' chromosome that isn't: A forensic bioinformatic investigation of Albert Perry's X-degenerate portion of the Y chromosome. *European Journal of Human Genetics*, 22(9), 1111–1116. <https://doi.org/10.1038/ejhg.2013.303>
- Francaacci, P., Morelli, L., Angius, A., Berutti, R., Reinier, F., Atzeni, R., ... Cucca, F. (2013). Low-pass DNA sequencing of 1200 Sardinians reconstructs European Y-chromosome phylogeny. *Science*, 341(6145), 565–569. <https://doi.org/10.1126/science.1237947>
- Fu, Q., Li, H., Moorjani, P., Jay, F., Slepchenko, S. M., Bondarev, A. A., ... Paabo, S. (2014). Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature*, 514(7523), 445–449. <https://doi.org/10.1038/nature13810>
- Gao, Y., Dong, G., Yang, X., & Chen, F. (2020). A review on the spread of prehistoric agriculture from southern China to mainland Southeast Asia. *Science China Earth Sciences*, 63(5), 615–625. <https://doi.org/10.1007/s11430-019-9552-5>
- Gray, R. D., Drummond, A. J., & Greenhill, S. J. (2009). Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science*, 323(5913), 479–483. <https://doi.org/10.1126/science.1166858>
- Harris, K. (2015). Evidence for recent, population-specific evolution of the human mutation rate. *Proceedings of the National Academy of Sciences of the United States of America*, 112(11), 3439–3444. <https://doi.org/10.1073/pnas.1418652112>
- Hasegawa, M., Kishino, H., & Yano, T. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, 22(2), 160–174. <https://doi.org/10.1007/bf02101694>
- Helgason, A., Einarsson, A. W., Guethmundsdottir, V. B., Sigurethsson, A., Gunnarsdottir, E. D., Jagadeesan, A., ... Stefansson, K. (2015). The Y-chromosome point mutation rate in humans. *Nature Genetics*, 47(5), 453–457. <https://doi.org/10.1038/ng.3171>
- Heyer, E., Georges, M., Pachner, M., & Endicott, P. (2013). Genetic diversity of four Filipino Negrito populations from Luzon: Comparison of male and female effective population sizes and differential integration of immigrants into Aeta and Agta communities. *Human Biology*, 85(1–3), 189–208. <https://doi.org/10.3378/027.085.0310>
- Higham, C. (2004). Mainland Southeast Asia from the Neolithic to the Bronze Age in Southeast Asia: From prehistory to history. In I. Glover & P. Bellwood (Eds.), *Southeast Asia: From prehistory to history* (pp. 41–67). London, England: Routledge-Curzon.
- Illumäe, A. M., Reidla, M., Chukhryaeva, M., Jarve, M., Post, H., Karmin, M., ... Rootsi, S. (2016). Human Y chromosome haplogroup N: A non-trivial time-resolved phylogeography that cuts across language families. *American Journal of Human Genetics*, 99(1), 163–173. <https://doi.org/10.1016/j.ajhg.2016.05.025>
- Jiao, T.-L. (2007). *The Neolithic of Southeast China: Cultural transformation and regional interaction on the coast*. New York, NY: Cambria Press.
- Karafet, T. M., Hallmark, B., Cox, M. P., Sudoyo, H., Downey, S., Lansing, J. S., & Hammer, M. F. (2010). Major east-west division underlies Y chromosome stratification across Indonesia. *Molecular Biology and Evolution*, 27(8), 1833–1844. <https://doi.org/10.1093/molbev/msq063>
- Karafet, T. M., Lansing, J. S., Redd, A. J., Reznikova, S., Watkins, J. C., Surata, S. P., ... Hammer, M. F. (2005). Balinese Y-chromosome perspective on the peopling of Indonesia: Genetic contributions from pre-Neolithic hunter-gatherers, Austronesian farmers, and Indian traders. *Human Biology*, 77(1), 93–114.
- Karmin, M., Saag, L., Vicente, M., Wilson Sayres, M. A., Jarve, M., Talas, U. G., ... Kivisild, T. (2015). A recent bottleneck of Y chromosome diversity coincides with a global change in culture. *Genome Research*, 25(4), 459–466. <https://doi.org/10.1101/gr.186684.114>
- Kayser, M., Brauer, S., Cordaux, R., Casto, A., Lao, O., Zhivotovsky, L. A., ... Stoneking, M. (2006). Melanesian and Asian origins of Polynesians: mtDNA and Y chromosome gradients across the Pacific. *Molecular Biology and Evolution*, 23(11), 2234–2244. <https://doi.org/10.1093/molbev/msl093>
- Kayser, M., Brauer, S., Weiss, G., Schiefenhovel, W., Underhill, P., Shen, P., ... Stoneking, M. (2003). Reduced Y-chromosome, but not mitochondrial DNA, diversity in human populations from West New Guinea. *American Journal of Human Genetics*, 72(2), 281–302. <https://doi.org/10.1086/346065>
- Kayser, M., Choi, Y., van Oven, M., Mona, S., Brauer, S., Trent, R. J., ... Stoneking, M. (2008). The impact of the Austronesian expansion: Evidence from mtDNA and Y chromosome diversity in the Admiralty Islands of Melanesia. *Molecular Biology and Evolution*, 25(7), 1362–1374. <https://doi.org/10.1093/molbev/msn078>

- Ko, A. M., Chen, C. Y., Fu, Q., Delfin, F., Li, M., Chiu, H. L., ... Ko, Y. C. (2014). Early Austronesians: Into and out of Taiwan. *American Journal of Human Genetics*, *94*(3), 426–436. <https://doi.org/10.1016/j.ajhg.2014.02.003>
- Kuroki, Y., Toyoda, A., Noguchi, H., Taylor, T. D., Itoh, T., Kim, D. S., ... Fujiyama, A. (2006). Comparative analysis of chimpanzee and human Y chromosomes unveils complex evolutionary pathway. *Nature Genetics*, *38*(2), 158–167. <https://doi.org/10.1038/ng1729>
- Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, *26*(5), 589–595. <https://doi.org/10.1093/bioinformatics/btp698>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., & Homer, N. (2009). Genome project data processing S: The Sequence Alignment/Map Format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Li, H., Huang, Y., Mustavich, L. F., Zhang, F., Tan, J. Z., Wang, L. E., ... Jin, L. (2007). Y chromosomes of prehistoric people along the Yangtze River. *Human Genetics*, *122*(3–4), 383–388. <https://doi.org/10.1007/s00439-007-0407-2>
- Li, H., Wen, B., Chen, S. J., Su, B., Pramoongjago, P., Liu, Y., ... Jin, L. (2008). Paternal genetic affinity between Western Austronesians and Daic populations. *BMC Evolutionary Biology*, *8*, 146. <https://doi.org/10.1186/1471-2148-8-146>
- Li, K.-T. (2013). First farmers and their adaptation in coastal Taiwan. In A. P. Underhill (Ed.), *A companion to Chinese archaeology* (pp. 612–633). Hoboken, NY: John Wiley & Sons.
- Lipson, M., Cheronet, O., Mallick, S., Rohland, N., Oxenham, M., Pietruszewski, M., ... Reich, D. (2018). Ancient genomes document multiple waves of migration in southeast Asian prehistory. *Science*, *361*(6397), 92–95. <https://doi.org/10.1126/science.aat3188>
- Lipson, M., Loh, P. R., Patterson, N., Moorjani, P., Ko, Y. C., Stoneking, M., ... Reich, D. (2014). Reconstructing Austronesian population history in Island Southeast Asia. *Nature Communications*, *5*, 4689. <https://doi.org/10.1038/ncomms5689>
- Loo, J. H., Trejaut, J. A., Yen, J. C., Chen, Z. S., Lee, C. L., & Lin, M. (2011). Genetic affinities between the Yami tribe people of Orchid Island and the Philippine islanders of the Batanes archipelago. *BMC Genetics*, *12*, 21. <https://doi.org/10.1186/1471-2156-12-21>
- Loo, J. H., Trejaut, J. A., Yen, J. C., Chen, Z. S., Ng, W. M., Huang, C. Y., ... Lin, M. (2014). Mitochondrial DNA association study of type 2 diabetes with or without ischemic stroke in Taiwan. *BMC Research Notes*, *7*, 223. <https://doi.org/10.1186/1756-0500-7-223>
- Lu, D., Lou, H., Yuan, K., Wang, X., Wang, Y., Zhang, C., ... Xu, S. (2016). Ancestral origins and genetic history of Tibetan highlanders. *American Journal of Human Genetics*, *99*(3), 580–594. <https://doi.org/10.1016/j.ajhg.2016.07.002>
- Macaulay, V., Hill, C., Achilli, A., Rengo, C., Clarke, D., Meehan, W., ... Richards, M. (2005). Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science*, *308*(5724), 1034–1036. <https://doi.org/10.1126/science.1109792>
- Malaspina, A. S., Westaway, M. C., Muller, C., Sousa, V. C., Lao, O., Alves, I., ... Willerslev, E. (2016). A genomic history of aboriginal Australia. *Nature*, *538*(7624), 207–214. <https://doi.org/10.1038/nature18299>
- Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., ... Reich, D. (2016). The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature*, *538*(7624), 201–206. <https://doi.org/10.1038/nature18964>
- Mendez, F. L., Krahn, T., Schrack, B., Krahn, A. M., Veeramah, K. R., Woerner, A. E., ... Hammer, M. F. (2013). An African American paternal lineage adds an extremely ancient root to the human Y chromosome phylogenetic tree. *American Journal of Human Genetics*, *92*(3), 454–459. <https://doi.org/10.1016/j.ajhg.2013.02.002>
- Mirabal, S., Herrera, K. J., Gayden, T., Regueiro, M., Underhill, P. A., Garcia-Bertrand, R. L., & Herrera, R. J. (2012). Increased Y-chromosome resolution of haplogroup O suggests genetic ties between the Ami aborigines of Taiwan and the Polynesian Islands of Samoa and Tonga. *Gene*, *492*(2), 339–348. <https://doi.org/10.1016/j.gene.2011.10.042>
- Nasu, H., Gu, H. B., Momohara, A., & Yasuda, Y. (2012). Land-use change for rice and foxtail millet cultivation in the Chengtoushan site, Central China, reconstructed from weed seed assemblages. *Archaeological and Anthropological Sciences*, *4*(1), 1–14.
- Pei, A. P., Church, S., & Chung, G. (1998). Notes on new advancements and revelations in the agricultural archaeology of early rice domestication in the Dongting Lake region. *Antiquity*, *72*(278), 878–885.
- Posth, C., Nagele, K., Colleran, H., Valentin, F., Bedford, S., Kami, K. W., ... Powell, A. (2018). Language continuity despite population replacement in remote Oceania. *Nature Ecology & Evolution*, *2*(4), 731–740. <https://doi.org/10.1038/s41559-018-0498-2>
- Poznik, G. D., Xue, Y., Mendez, F. L., Willems, T. F., Massaia, A., Wilson Sayres, M. A., ... Tyler-Smith, C. (2016). Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. *Nature Genetics*, *48*(6), 593–599. <https://doi.org/10.1038/ng.3559>
- Pugach, I., Duggan, A. T., Merriwether, D. A., Friedlaender, F. R., Friedlaender, J. S., & Stoneking, M. (2018). The gateway from near into remote Oceania: New insights from genome-wide data. *Molecular Biology and Evolution*, *35*(4), 871–886. <https://doi.org/10.1093/molbev/msx333>
- Raghavan, M., Skoglund, P., Graf, K. E., Metspalu, M., Albrechtsen, A., Moltke, I., ... Willerslev, E. (2014). Upper Paleolithic Siberian genome reveals dual ancestry of native Americans. *Nature*, *505*(7481), 87–91. <https://doi.org/10.1038/nature12736>
- Rambaut, A. (2018). FigTree download page. Retrieved from <http://tree.bio.ed.ac.uk/software/figtree/>.
- Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, *67*(5), 901–904. <https://doi.org/10.1093/sysbio/syy032>
- Rasmussen, M., Anzick, S. L., Waters, M. R., Skoglund, P., DeGiorgio, M., Stafford, T. W., Jr., ... Willerslev, E. (2014). The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature*, *506*(7487), 225–229. <https://doi.org/10.1038/nature13025>
- Rodríguez, F., Oliver, J. L., Marín, A., & Medina, J. R. (1990). The general stochastic model of nucleotide substitution. *Journal of*

- Theoretical Biology*, 142(4), 485–501. [https://doi.org/10.1016/S0022-5193\(05\)80104-3](https://doi.org/10.1016/S0022-5193(05)80104-3)
- Ross, M. (2005). The Batic languages in relation to the early history of the Malayo-Polynesian subgroup of Austronesian. *Journal of Austronesian Studies*, 1(2), 1–24.
- Scheinfeldt, L., Friedlaender, F., Friedlaender, J., Latham, K., Koki, G., Karafet, T., ... Lorenz, J. (2006). Unexpected NRY chromosome variation in Northern Island Melanesia. *Molecular Biology and Evolution*, 23(8), 1628–1641. <https://doi.org/10.1093/molbev/msl028>
- Scholes, C., Siddle, K., Ducourneau, A., Crivellaro, F., Jarve, M., Rootsi, S., ... Migliano, A. B. (2011). Genetic diversity and evidence for population admixture in Batak Negritos from Palawan. *American Journal of Physical Anthropology*, 146(1), 62–72. <https://doi.org/10.1002/ajpa.21544>
- Sengupta, S., Zhivotovskiy, L. A., King, R., Mehdi, S. Q., Edmonds, C. A., Chow, C. E., ... Underhill, P. A. (2006). Polarity and temporality of high-resolution Y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists. *American Journal of Human Genetics*, 78(2), 202–221. <https://doi.org/10.1086/499411>
- Soares, P., Trejaut, J. A., Loo, J. H., Hill, C., Mormina, M., Lee, C. L., ... Richards, M. B. (2008). Climate change and post-glacial human dispersals in Southeast Asia. *Molecular Biology and Evolution*, 25(6), 1209–1218. <https://doi.org/10.1093/molbev/msn068>
- Soares, P. A., Trejaut, J. A., Rito, T., Cavadas, B., Hill, C., Eng, K. K., ... Richards, M. B. (2016). Resolving the ancestry of Austronesian-speaking populations. *Human Genetics*, 135(3), 309–326. <https://doi.org/10.1007/s00439-015-1620-z>
- Su, B., Jin, L., Underhill, P., Martinson, J., Saha, N., McGarvey, S. T., ... Deka, R. (2000). Polynesian origins: Insights from the Y chromosome. *Proceedings of the National Academy of Sciences of the United States of America*, 97(15), 8225–8228.
- Tabbada, K. A., Trejaut, J., Loo, J. H., Chen, Y. M., Lin, M., Mirazon-Lahr, M., ... de Ungria, M. C. (2010). Philippine mitochondrial DNA diversity: A populated viaduct between Taiwan and Indonesia? *Molecular Biology and Evolution*, 27(1), 21–31. <https://doi.org/10.1093/molbev/msp215>
- The HUGO Pan-Asian SNP Consortium, Abdulla, M. A., Ahmed, I., Assawamakin, A., Bhak, J., Brahmachari, S. K., ... Indian Genome Variation Consortium. (2009). Mapping human genetic diversity in Asia. *Science*, 326(5959), 1541–1545. <https://doi.org/10.1126/science.1177074>
- Thomson, R., Pritchard, J. K., Shen, P., Oefner, P. J., & Feldman, M. W. (2000). Recent common ancestry of human Y chromosomes: Evidence from DNA sequence data. *Proceedings of the National Academy of Sciences of the United States of America*, 97(13), 7360–7365.
- Trejaut, J. A., Kivisild, T., Loo, J. H., Lee, C. L., He, C. L., Hsu, C. J., ... Lin, M. (2005). Traces of archaic mitochondrial lineages persist in Austronesian-speaking Formosan populations. *PLoS Biology*, 3(8), e247. <https://doi.org/10.1371/journal.pbio.0030247>
- Trejaut, J. A., Muiyard, F., Lai, Y. H., Chen, L. R., Chen, Z. S., Loo, J. H., ... Lin, M. (2019). Genetic diversity of the Thao people of Taiwan using Y-chromosome, mitochondrial DNA and HLA gene systems. *BMC Evolutionary Biology*, 19(1), 64. <https://doi.org/10.1186/s12862-019-1389-0>
- Trejaut, J. A., Poloni, E. S., Yen, J. C., Lai, Y. H., Loo, J. H., Lee, C. L., ... Lin, M. (2014). Taiwan Y-chromosomal DNA variation and its relationship with Island Southeast Asia. *BMC Genetics*, 15, 77. <https://doi.org/10.1186/1471-2156-15-77>
- Vallée, F., Luciani, A., & Cox, M. P. (2016). Reconstructing demography and social behavior during the Neolithic expansion from genomic diversity across Island Southeast Asia. *Genetics*, 204(4), 1495–1506. <https://doi.org/10.1534/genetics.116.191379>
- Wang, C. C., Jin, L., & Li, H. (2014). Natural selection on human Y chromosomes. *Journal of Genetics and Genomics*, 41(2), 47–52. <https://doi.org/10.1016/j.jgg.2014.01.006>
- Wang, Y., Song, F., Zhu, J., Zhang, S., Yang, Y., Chen, T., ... Zhao, W. (2017). GSA: Genome sequence archive. *Genomics, Proteomics & Bioinformatics*, 15(1), 14–18. <https://doi.org/10.1016/j.gpb.2017.01.001>
- Wei, L. H., Yan, S., Teo, Y. Y., Huang, Y. Z., Wang, L. X., Yu, G., ... Li, H. (2017). Phylogeography of Y-chromosome haplogroup O3a2b2-N6 reveals patrilineal traces of Austronesian populations on the eastern coastal regions of Asia. *PLoS One*, 12(4), e0175080. <https://doi.org/10.1371/journal.pone.0175080>
- Wilson Sayres, M. A., Lohmueller, K. E., & Nielsen, R. (2014). Natural selection reduced diversity on human y chromosomes. *PLoS Genetics*, 10(1), e1004064. <https://doi.org/10.1371/journal.pgen.1004064>
- Wong, L. P., Ong, R. T., Poh, W. T., Liu, X., Chen, P., Li, R., ... Teo, Y. Y. (2013). Deep whole-genome sequencing of 100 southeast Asian Malays. *American Journal of Human Genetics*, 92(1), 52–66. <https://doi.org/10.1016/j.ajhg.2012.12.005>
- Xue, Y., Wang, Q., Long, Q., Ng, B. L., Swerdlow, H., Burton, J., ... Tyler-Smith, C. (2009). Human Y chromosome base-substitution mutation rate measured by direct sequencing in a deep-rooting pedigree. *Current Biology*, 19(17), 1453–1457. <https://doi.org/10.1016/j.cub.2009.07.032>
- Y Chromosome Consortium. (2002). A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Research*, 12(2), 339–348. <https://doi.org/10.1101/gr.217602>
- Yan, S., Wang, C. C., Zheng, H. X., Wang, W., Qin, Z. D., Wei, L. H., ... Jin, L. (2014). Y chromosomes of 40% Chinese descend from three Neolithic super-grandfathers. *PLoS One*, 9(8), e105691. <https://doi.org/10.1371/journal.pone.0105691>
- Yang, M. A., Fan, X., Sun, B., Chen, C., Lang, J., Ko, Y. C., ... Fu, Q. (2020). Ancient DNA indicates human population shifts and admixture in northern and southern China. *Science*, 369(6501), 282–288. eaba0909. <https://doi.org/10.1126/science.aba0909>
- Yang, X., Chen, Q., Ma, Y., Li, Z., Hung, H.-C., Zhang, Q., ... Fu, X. (2018). New radiocarbon and archaeobotanical evidence reveal the timing and route of southward dispersal of rice farming in south China. *Science Bulletin*, 63(22), 1495–1501. <https://doi.org/10.1016/j.scib.2018.10.011>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Sun J, Wei L-H, Wang L-X, et al. Paternal gene pool of Malays in Southeast Asia and its applications for the early expansion of Austronesians. *Am J Hum Biol.* 2020; e23486. <https://doi.org/10.1002/ajhb.23486>