LUPINE PUBLISHERS

Journal Of Anthropological And Archaeological Sciences

DOI: 10.32474/JAAS.2026.12.000396

ISSN: 2690-5752

Research Article

Genetic Austronesian Heritage of the Siraya People of Taiwan

Jin Yuan Huang[1], Zhong Sian Chen[2], Marie Lin[3], Jean Alain Trejaut[4]*

[1,2,3,4]Molecular Anthropology and Transfusion Medicine Research Laboratory, Mackay Memorial Hospital, Taipei, Taiwan

***Corresponding author:** Jean Alain Trejaut, Molecular Anthropology and Transfusion Medicine Research Laboratory, Mackay Memorial Hospital, Taipei, Taiwan

Received: April 01, 2026 **Published:** June 30, 2026

Abstract

The Siraya people have a rich history, having settled in Taiwan approximately 6,000 years ago while speaking Austronesian languages. Today, they exist in various groups across the island. Unfortunately, the past four centuries have seen substantial migrations of non-Austronesian (NAN) speaking groups from East Asia, which, coupled with significant Sinicization, have contributed to the extinction of the Siraya language. Despite this, many elements of their unique culture and customs continue to thrive.

Efforts to revive their heritage and secure official recognition as an indigenous group highlight the ongoing struggle to honor their identity. The origins of the Siraya remain contentious, but increasing research into their genetic variation holds promise for illuminating various theories emerging from anthropology, linguistics, and governmental narratives. Analysis of their genetic structures, including Human leukocyte antigen (HLA), mitochondrial DNA (mtDNA), and non-recombining Y chromosome (NRY), shows that their maternal lineage and, to a lesser extent, their autosomal and paternal lineages strongly reflect Austronesian characteristics and underscore the need to recognize their profound historical legacy.

Keywords:

Population Genetics; Siraya; Taiwan Aborigines; Y Chromosome; Bayesian Analysis of Population Structure; Mtdna; HLA

Introduction

Despite migration, diffusion, and introgression, island populations that have been isolated for millennia are more likely to generate genetic specificity that allows geneticists to retrace their ancestral history. For thousands of years, the Indigenous villages of Taiwan operated as independent entities and often engaged in conflicts with one another. The Siraya people, located near Tainan, were the first group of "Plain indigenous people" (or Pingpu) encountered by the Dutch who referred to them as the Sideia. In the mid-seventeenth century, the Dutch reorganized these villages into groups based on parish divisions, native trade systems, and the arrangement of hunting grounds. This reorganization encouraged the Indigenous people to seek external resources and fostered cooperation among neighboring villages [1].

Over the past 250 years, numerous migrants from mainland China arrived in Taiwan, leading to the assimilation of the Siraya people into the ethnic Chinese community through intermarriages. After World War II, most Pingpu groups, including the Siraya, lost their indigenous status [2]. In 2022, after more than a decade of seeking recognition, the Constitutional Court of Taiwan rejected Siraya's claim to be recognized as Indigenous people [3]. Despite this setback, the Siraya possess many distinct cultural traits. Unfortunately, their original language, a branch of the Austronesian language family, is now extinct, although it is well-documented in literature [4]. Over time, it is believed that the Siraya language (or proto-Siraya) evolved into three subgroups: Siraya, Taivuan, and Makatau. Today, most Siraya people reside in the southwestern plains of Taiwan, primarily in Tainan County. However, some groups, particularly the Taivuan, were displaced to Pingtung County during the Cheng Dynasty (1661-1683) and again from 1722 to 1736. Simultaneously, the Makatau people in Pingtung County were moved to the eastern foothills of the Central Mountain Range [5]. Additionally, records indicate that in 1804, the Siraya and several other plain tribes faced further displacements to the east coast of Taiwan [6].

The name "Siraya" is a transliteration of the word for "people" in their language. This name was assigned by Japanese scholars during the period from 1895 to 1945 5. It is believed that the Niaosong culture, which existed from 1600 to 500 B.P. and is characterized by a significant number of bird-head-shaped artifacts uncovered at the Niaosong archaeological site in Tainan, serves as the ancestral culture of the Siraya [7]. However, further research is needed to fully understand the complex origins of the Siraya people.

Molecular anthropology explores the genetic diversity in various parts of the human genome, including mitochondrial DNA (mtDNA), Y-chromosome DNA, the Human Leukocyte Antigen (HLA)

system, and the complete human genome. For instance, autosomal inheritance, which in this study involves the highly polymorphic HLA system, represents genetic contribution from both parents. The HLA system has a slow mutation rate, it is influenced by selective pressure and shows significant recombination, especially in mixed groups. In contrast, uniparental inheritance systems, such as mtDNA and Y-chromosome DNA, are passed down through one parent only. Mitochondrial DNA is inherited strictly from mother to daughter without genetic recombination. Its mutation rate is moderate, which makes mtDNA particularly valuable for studying human populations and tracing maternal migration events throughout history. Similarly, the Y-chromosome is inherited from father to son and is used for research on paternal evolution [8]. Although the Y-chromosome has a faster rate of variation, its study enables researchers to estimate the coalescence age of specific traits and link them to historical migration events [9].

In this study, we analyze the genetic variation of the Siraya groups by examining the HLA, mtDNA, and Y-Chromosome gene systems. These genetic systems will help estimate the genetic contributions—both paternal and maternal—of the Siraya from other Taiwanese indigenous groups and relevant reference populations from abroad, including regions such as Northeast China, South China, East Coast China, Mainland Southeast Asia, and Island Southeast Asia. Last, we employ Bayesian Analysis of Population Structure (BAPS) to explore gene flow and immigration patterns among the Siraya groups and their neighboring populations.

Materials and Methods

Samples

Unrelated healthy Siraya individuals (n=541) (Figure 1, Table 1) were recruited for this study. Group 1 comprised Siraya individuals from the Tainan County (n= 159), group 2, 3 and 4 relocated to the Khaohsiung County (n=57), the Pingtung County (n=147), the Hualien County (n= 66) respectively. Except for Makatao (n = 50) from the Pingtung County, all individuals were analyzed for Mitochondrial DNA (mtDNA), Non-recombining Y chromosome (NRY) and Human Leukocyte Antigen (HLA) gene systems [10,11]. Their profiles were compared to other groups in Taiwan (Table 1) previously characterized for the same gene systems [12-14] and neighboring populations of North China (n = 252), the East China coast (Fujian and Matsu; n= 198) [15,16] South China (n = 1163), the Philippines (n = 260), Western and Eastern Indonesia (n= 326 + 72) [12,14,17-19] and Mainland Southeast Asia (MSEA): Vietnam (n = 603) [20], Thailand (n = 560) [21] and Malaysia (n = 86) [22,23] (Table 1).

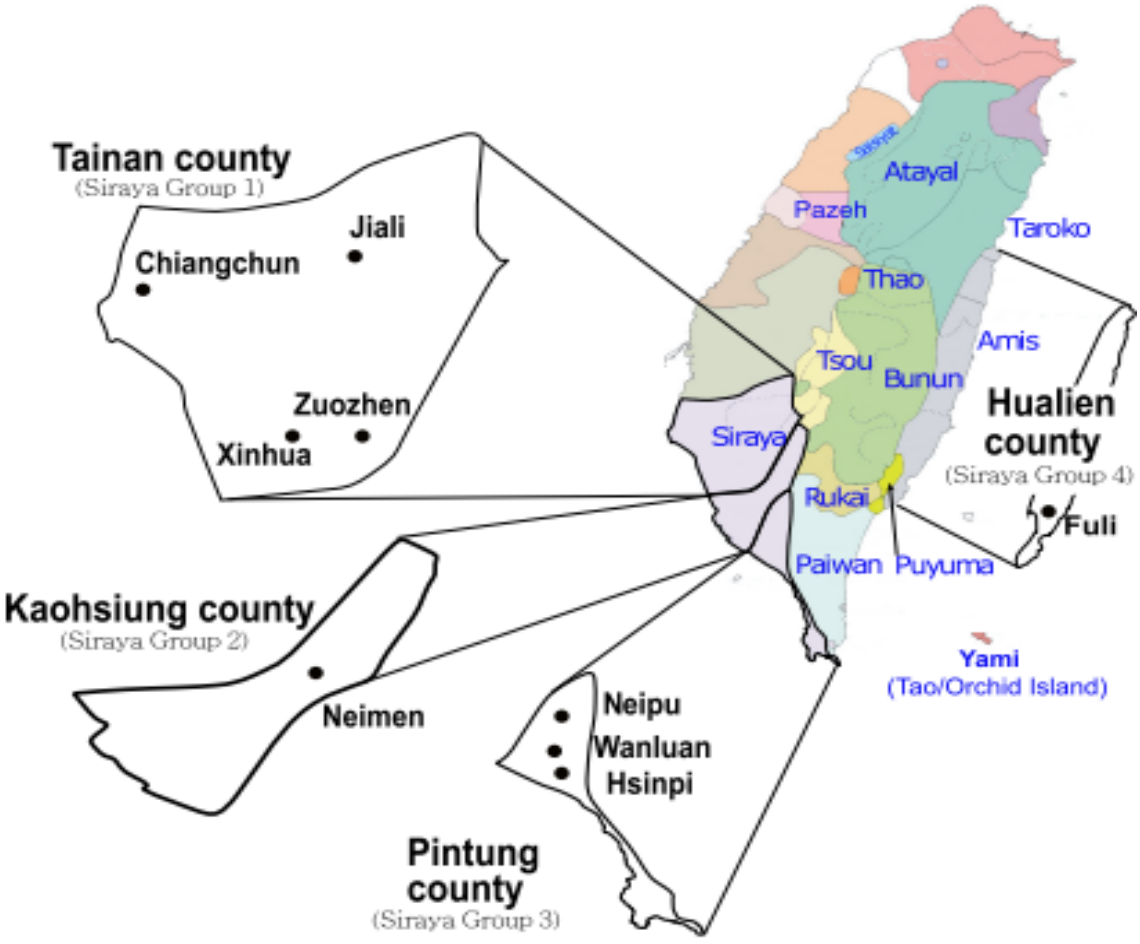


Figure 1: Location of Siraya groups.

Table 1: Population samples.

Population			This study sampling location	Code	Linguistic affiliation	mtDNA (n)	Y Chr (n)	HLA (n)	mtDNA References	Y Chromosome	HLA
Island Southeast Asia (ISEA)			Batan	Bat	Malayo-Polynesian	29	59	50	Delfin et al. 2011	Loo et al 2011	AFND 2020
			Philippines	Phi	Malayo-Polynesian	231	353	55	Delfin et al. 2011	Delfin et al. 2010	AFND 2020
			West Indonesia	WInd	Malayo-Polynesian	284	253	236	Kim et al. 2011	Kim et al. 2011	AFND 2020
			East Indonesia	EInd	Malayo-Polynesian	416	119	NA	Karafet 2005	Karafet 2005, Trejaut et al. 2014	AFND 2020
Taiwan	Taiwanese Plain tribes	Siraya Group 1	Tainan County, Xinhua Township	SirTX	sinitic	58	30	51	This study	This study	This study
			Tainan County, Jiali Township	SirTJ	sinitic	18	17	26	This study	This study	This study
			Tainan County, Zuozhen Township	SirTZ	sinitic	26	na	18	This study	NA	This study
			Tainan County, Chiangchun Township	SirTC	sinitic	57	38	na	This study	This study	NA
		Siraya Group 2	Kaohsiung County, Neimen Township	SirKN	sinitic	57	39	29	This study	This study	This study
			Pintung County, Wanluan Township	SirPW	sinitic	69	44	36	This study	This study	This study
		Siraya Group 3	Pintung County, Hsinpi Township	SirPH	sinitic	43	25	47	This study	This study	This study
			Pintung County, Neipu Township	SirPN	sinitic	35	16	16	This study	This study	This study
		Siraya Group 4	Hualien County, Fuli Township	SirHF	sinitic	66	37	51	This study	This study	This study
		Other plain tribes	Makatao	Mak	sinitic	50	NA	NA	Ko et al. 2014	NA	NA
			Pazeh	Paz	sinitic	65	38	105	This study	Trejaut et al. 2014	This study
	Taiwan Indigenous Peoples (Austronesian Speaking Taiwanese; AN_Tw)		Atayal	Ata	Formosan	50	52	105	Ko et al. 2014	Trejaut et al. 2014	AFND 2020
			Saisiyat	Sai	Formosan	24	24	51	Ko et al.2014	Trejaut et al. 2014	AFND 2020
			Tsou	Tso	Formosan	48	41	51	Ko et al.2014	Trejaut et al. 2014	AFND 2020
			Thao	Thao	Formosan	30	16	30	Trejaut et al. 2019	Trejaut et al. 2014	AFND 2020
			Bunun	Bun	Formosan	50	56	101	Ko et al.2014	Trejaut et al. 2014	AFND 2020
			Puyuma	Puy	Formosan	40	23	50	Ko et al.2014	Trejaut et al. 2014	AFND 2020
			Rukai	Ruk	Formosan	50	29	50	Ko et al.2014	Trejaut et al. 2014	AFND 2020
			Paiwan	Pai	Formosan	49	25	51	Ko et al.2014	Trejaut et al. 2014	AFND 2020
			Amis	Ami	Formosan	50	38	98	Ko et al.2014	Trejaut et al. 2014	AFND 2020
			Yami (Tao / Orchid Island)	Yam	Formosan	44	30	50	Ko et al.2014	Trejaut et al. 2014	AFND 2020
	Taiwanese Han (Non-AN_Tw; NAN_Tw)		Urban Taiwanese (Minnan & Hakka Taipei)	UrTw	sinitic	123	155	324	Lin et al. 2023	Trejaut et al. 2014	Lin et al. 2001
			Hakka Pintung	Hak	sinitic	46	32	55	Ko et al. 2014	Trejaut et al. 2014	NA
			Minnan Kaohsiung	Min	sinitic	49	60	101	Ko et al. 2014	Trejaut et al. 2014	NA
China	Han		North China	Nch	sinitic	252	157	105	Kong et al. 2003, Peng et al. 2011, Li	Kim et al. 2011	AFND 2020
			East coast China	Ech	sinitic	148	53	60	Loo et al. 2011	Trejaut et al. 2014	AFND 2020
			South China	Sch	sinitic	1163	100	101	Kim et al. 2011	Kim et al. 2011	AFND 2020
			Matsu	Matsu	sinitic	50	NA	NA	Chen et al. 2021	NA	NA
Mainland Southeast Asia (MSEA)			Malaysia	Mal	Malayo-Polynesian	86	14	173	Jinam et al. 2012	Tofanelli 2009	AFND 2020
			Thailand	Thai	Tai kadai	559	40	142	Kim et al. 2011	Kim et al. 2011	AFND 2020
			Vietnam	Vie	Khmon Mien	609	48	103	Kim et al. 2011	Kim et al. 2011	AFND 2020
Total						5024	2061	2521			

NA: not applicable

*: Pooled data

AFND 2020: Allele frequency Net Databank: http://www.allelefrequencies.net. (accessed on the 24th of August 2022)

Data analysis

DNA was first extracted from 500 µl of buffy-coat from each Chick swab or blood sample using the QIAmp DNA kit (QIAmp® DNA Blood Mini kit, Qiagen inc. Taiwan). For mtDNA typing, the control region HVS-1, nucleotide positions (nps) of coding region nps 8000 to 9000 and 9800 to 10900 were sequenced as previously described 12. Additionally, complete mtDNA sequencing of 39 selected samples (Supplementary Data 1) were performed on both strands using the Perkin-Elmer/Applied Biosystems Division (ABI Taiwan) DyeDeoxy Terminator Cycle Sequencing Kit (Foster City, California, United States) according to the recommendations of the manufacturer on an automated DNA sequencer (ABI Model 377). Mitochondrial haplogroups were assigned according to Build 17 of Phylotree [24] using Haplogrep-2 available online: https://haplogrep.i-med.ac.at/app [25].

Genotyping of Non-recombining Y chromosome (NRY) was obtained 14 from 81 Y-SNPs (Table 1) using direct sequencing of amplicons obtained from specific primer pairs as described in the Y Chromosome Consortium 2002 [26,27], and furthere analyzed using 17 Y-chromosome short tandem repeats (Y-STRs): DYS19, DYS385I, DYS385II, DYS389I, DYS389II, DYSS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635, and Y GATA-H4 [14].

Statistical analysis

Multidimensional scaling (MDS): MDS plots were constructed using haplogroup frequencies with the Paleontological statistics version 3.25 (PAST) [28].

Phylogenetic analysis:

To infer the mtDNA phylogenetic relationships between Siraya and other groups of Taiwan, East Asia, Peninsular Asia, and ISEA we adjusted all mtDNA haplogroup assignments using the same control and coding region fragments used for our local data set (control region HVS-1, coding region fragments nps 8000 to 9000 and 9800 to 10900).

Table 2: Gene Diversity and Mixture estimate with Han and NAN_Tw.

Genetic Systems	Haplogroups	All China (Putative parental Han)	All Tw_Han	Taiwan Plain Tribes			All AN_Tw (No Yami)	All ISEA	All MSEA
					Other plain Tribes				
				All Siraya tribes	Pazeh	Makatao #			
Human Leukocyte Antigen	Sample size (n)	48.00	480	**274**	105	na	637	341	418
	Mean Allele diversity	0.88	0.88	**0.83**	0.87	na	0.80	0.85	0.89
	Standard error	0.02	0.02	**0.03**	0.03	na	0.05	0.04	0.03
	Total number of alleles	108	87	**70**	48	na	45	86	228
	% alleles shared with Han	52.33%	34.67%	**36.39%**	31.38%	na	27.03%	17.68%	17.68%
	% alleles shared with AN_Tw	47.67%	50.39%	**56.47%**	64.45%	na	72.97%	20.92%	20.92%
	% of undefined origin	0.00%	14.94%	**7.14%**	4.17%	na	0.00%	61.40%	61.40%
Mitochondrial DNA	Sample Size (n)	516.00	218	**429**	65	50	489	544	1245.00
	Haplotype Diversity	0.996	0.993	**0.983**	0.975	0.973	0.956	0.970	0.99
	Standard Error	0.000	0.000	**0.001**	0.002	0.002	0.002	0.002	0.001
	Total number of haplogroups	241	124	**145**	41	32	74	109	218
	% mtDNA haplogroups shared with Han	80.43%	37.3%	**17.3%**	19.8%	15.4%	6.9%	11.0%	29.5%
	% mtDNA haplogroups shared with AN_Tw	19.57%	25.6%	**36.5%**	36.3%	62.7%	93.1%	25.7%	18.6%
	% of undefined origin	0.00%	37.1%	**46.2%**	43.9%	21.9%	0.0%	63.3%	51.8%
Calculation using 7 Y STRs haplotypes	Sample Size (n)	310	247	**221**	38	25	354	665	102
	SNP Haplogroup diversity	0.90	0.84	**0.82**	0.67	0.73	0.30	0.89	0.36
	Standard Error	0.02	0.02	**0.03**	0.06	0.04	0.07	0.02	0.07
	Total number of SNP Haplogroups	24	19	**17**	7	8	12	26	16
	% Y STRs haplotypes shared with Han	90.99%	7.58%	**8.10%**	8.64%	7.13%	4.12%	4.81%	12.01%
	% Y STRs haplotypes shared with AN_Tw	9.01%	7.36%	**21.36%**	19.93%	32.87%	95.88%	7.65%	4.26%
	% of undefined origin	0.00%	85.06%	**70.54%**	71.43%	60.00%	0.00%	87.54%	83.72%

Similarly, for the Y-SNP-STR phylogenetic relationships, only seven STRs were retained to match other Y chromosome determination protocols (DYS19, DYS389I, DYS389II, DYSS390, DYS391, DYS392, DYS393).

Population gene flow and ancestry sharing:

We inferred mixture processes in the Siraya groups from lineage sharing of the three gene systems, autosomal HLA, mitochondrial and NRY polymorphisms. Mixtures were estimated assuming that a population can be regarded as the result of a hybridization or mixture process among two putative parental populations, namely the Han and the Austronesian speakers of Taiwan [29] (Table 2). The putative parental Han comprised Northeast, East and South china, and the putative Austronesian group comprised all AN_Tw haplogroups (Table 2 and *Supplementary Table S1*).

Bayesian population mixture analysis was estimated with BAPS v. 6.1 (Bayesian Analysis of Population Structure) [30] along with the Evanno statistic [31] to detect the uppermost hierarchical level of structure (or best K) based on the rate of change in the log marginal probability of data between successive K values. Finally, a mixture analysis with the BAPS option "Admixture of individuals based on mixture clustering" was used to produce a heat map showing gene flow between all Source and Target groups and maximum K ≤ 30 (Figure 3 and Supplementary Figure S3).

Results and Discussions

Genetic structure

Data on genetic diversity for HLA, mtDNA, and NRY were obtained from 5,024 unrelated individuals across Asia. This included 476 Siraya individuals for HLA analysis, 266 for mtDNA, and 274 for NRY (Table 1). We categorized the Siraya samples into four groups based on their geographic locations. The analysis identified 70 distinct HLA alleles (HLA-A, -B, and -DRB1), 145 mtDNA haplogroups, and 17 NRY SNP haplogroups among the Siraya samples.

We calculated gene diversity indices based on the frequencies of haplogroups and alleles across all genetic systems (Table 2 and Supplementary Table S1). Overall, the Siraya population showed higher gene diversity compared to the AN_Tw group, although it was slightly lower than that of mainland China. This difference is likely attributed to reductions caused by genetic drift and founder effects.

We used the HLA gene system to evaluate deviations from Hardy-Weinberg (HW) equilibrium. Our analysis, which included four-digit allele assignments, showed that the HLA-A, -B, and -DRB1 loci did not significantly deviate from the expected HW equilibrium ($P > 0.05$) (data not shown). However, we found that Siraya group 1 displayed a significant deviation for the HLA-DRB1 locus. This deviation is likely attributed to the low diversity of HLA-DRB1 in this group, as well as the sensitivity of HW equilibrium to factors such as excess homozygosity, rare alleles, or potential genotyping errors [32].

Mixture proportion analysis

The HLA gene system is diploid and has a very low mutation rate. Consequently, its alleles are preserved for thousands of years. Over the past 6,000 years, the genetic diversity of these alleles observed between the Han and AN_Tw groups has been minimal. As a result, the variation between these groups is mainly attributed to genetic drift rather than point mutations, gene conversions, or recombination events.

Admixture proportions in Siraya were estimated using the haplogroup sharing method developed by Maca-Meyer 29. This method, along with profile sharing, highlights the diversity of undefined origins. Table 2 presents these proportions based on

a surogate Han parental population, which includes groups from Northern, East Coast, and Southern China, as well as an AN_Tw parental population that encompasses all indigenous groups of Taiwan.

It was not surprising to find that the Han HLA alleles present in the AN_Tw group exceeded 47% (see Table 2 and Supplementary Table S1). As a result, only HLA alleles found in the Siraya and other non-Austronesian speaking (NAN) groups that share more than 47% similarity with AN_Tw (as noted in the first column of Table 2) can be reliably attributed to Austronesian admixture. It is important to mention that we utilized a four-digit generic level for HLA genotyping, in contrast to the six-digit classification used in more advanced sequence-based typing methods. Therefore, the HLA data presented in Table 2 and Supplementary Table S1 represent only generic alleles and may underestimate the genetic mixing between AN_Tw and Siraya. Furthermore, with the exception of Siraya Group 4 (data not shown), Groups 1 to 3 exhibited only a few HLA alleles, with frequencies ranging from 2.27% to 5.56% (A*29:01, B*15:12, B*35:02, B*56:04, and B*81:01) that were not found in our Han control dataset. This discrepancy may be due to point mutations or recombination that occurred after a long period of isolation in Taiwan, or more likely, due to migration from other regions of East Asia. A search for allele frequencies in the worldwide population database indicates that, although these Siraya alleles are distributed globally (with frequencies ranging from 2% to 10% in Malaysia), they are rare among the Han population.

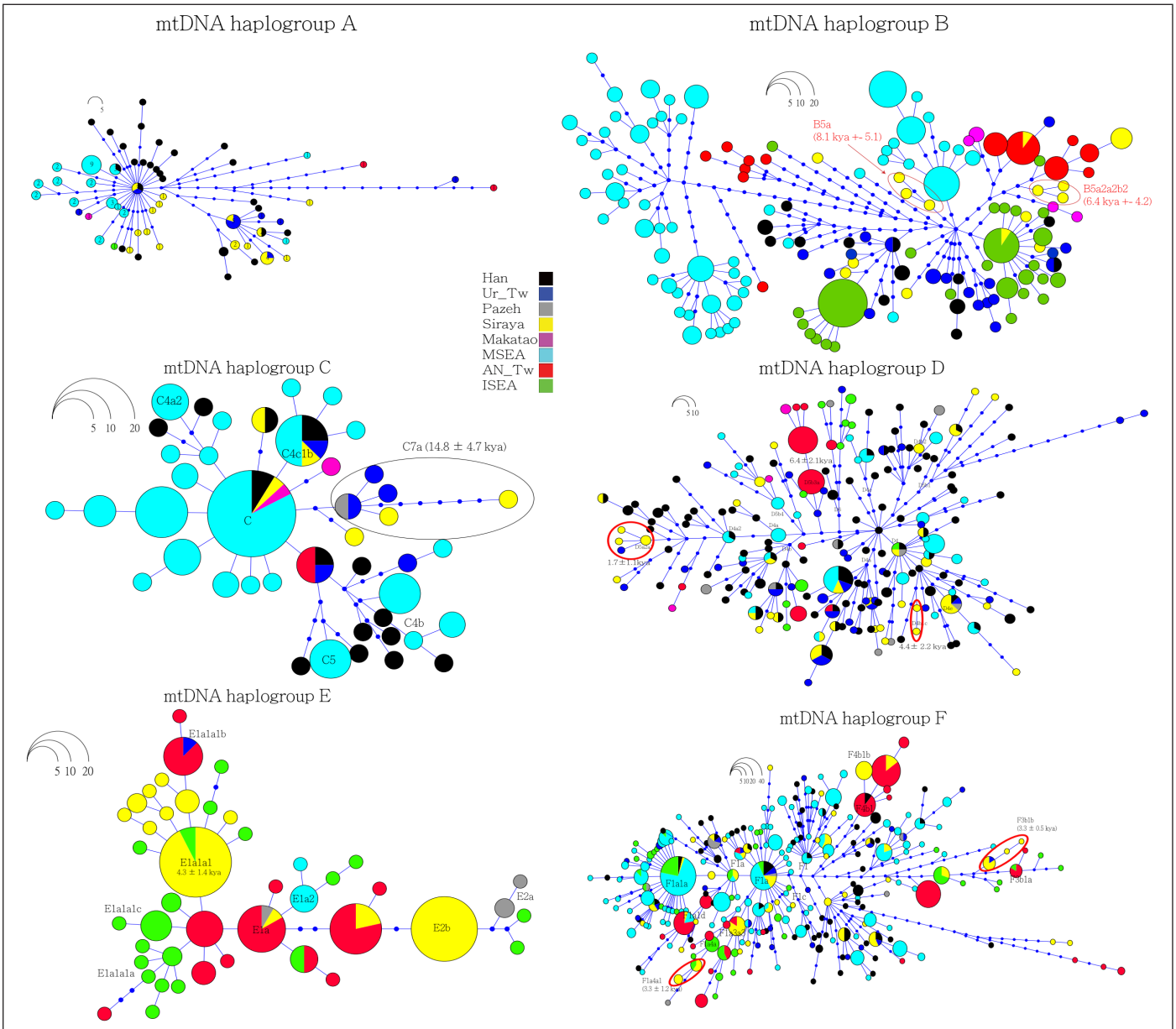
The mtDNA and NRY gene systems are haploid, and their mutation rates are faster than those of the HLA system [33,34]. This characteristic allows researchers to estimate the age of the most recent common ancestor (MRCA) and, in the case of unique findings in Taiwan, to determine the time of settlement or first migration to the island. As expected, the diversity of mtDNA and NRY shared between Han and indigenous Taiwanese (AN_Tw) populations was significantly lower (19.6% for mtDNA and 9% for NRY) compared to the diversity seen in the HLA system and between AN_Tw and the Siraya population, which showed 36.5% and 21.4% respectively. Overall, with the exception of Siraya Group 2, which primarily consists of Han males, we can reliably estimate that the AN_Tw component contributes approximately 16.9% of mtDNA and 12.4% of NRY in the other Siraya groups. To compare the genetics of the Siraya with published data (Table 1), we utilized a generic definition for mtDNA haplogroups, encompassing the HVR-I and 2000 nucleotides from the coding region. For NRY, we employed a limited number of single nucleotide polymorphisms (SNPs) along with seven Y-STRs. Additionally, some private mtDNA and NRY diversity were identified and used to estimate the settlement date of their carriers (Supplementary Figure S1 and S2).

We conducted a mixture analysis involving four potential genetic contributors: Han, ISEA (which combines the Philippines and Western Indonesia), MSEA (which includes Thailand, Malaysia, and Vietnam), and a putative ancestral parental Austronesian group. The latter was formed by merging all AN_Tw groups and removing haplotypes (for NRY) or haplogroups (for mtDNA) that were identical to those of Han (Table 2 and Supplementary Table S2). As indicated in Table 2, the depth of the HLA typing assignment suggested an overestimation of the mixture between AN_Tw and Siraya. However, for mtDNA and NRY, Siraya showed a slightly greater mixture with AN_Tw compared to Han.

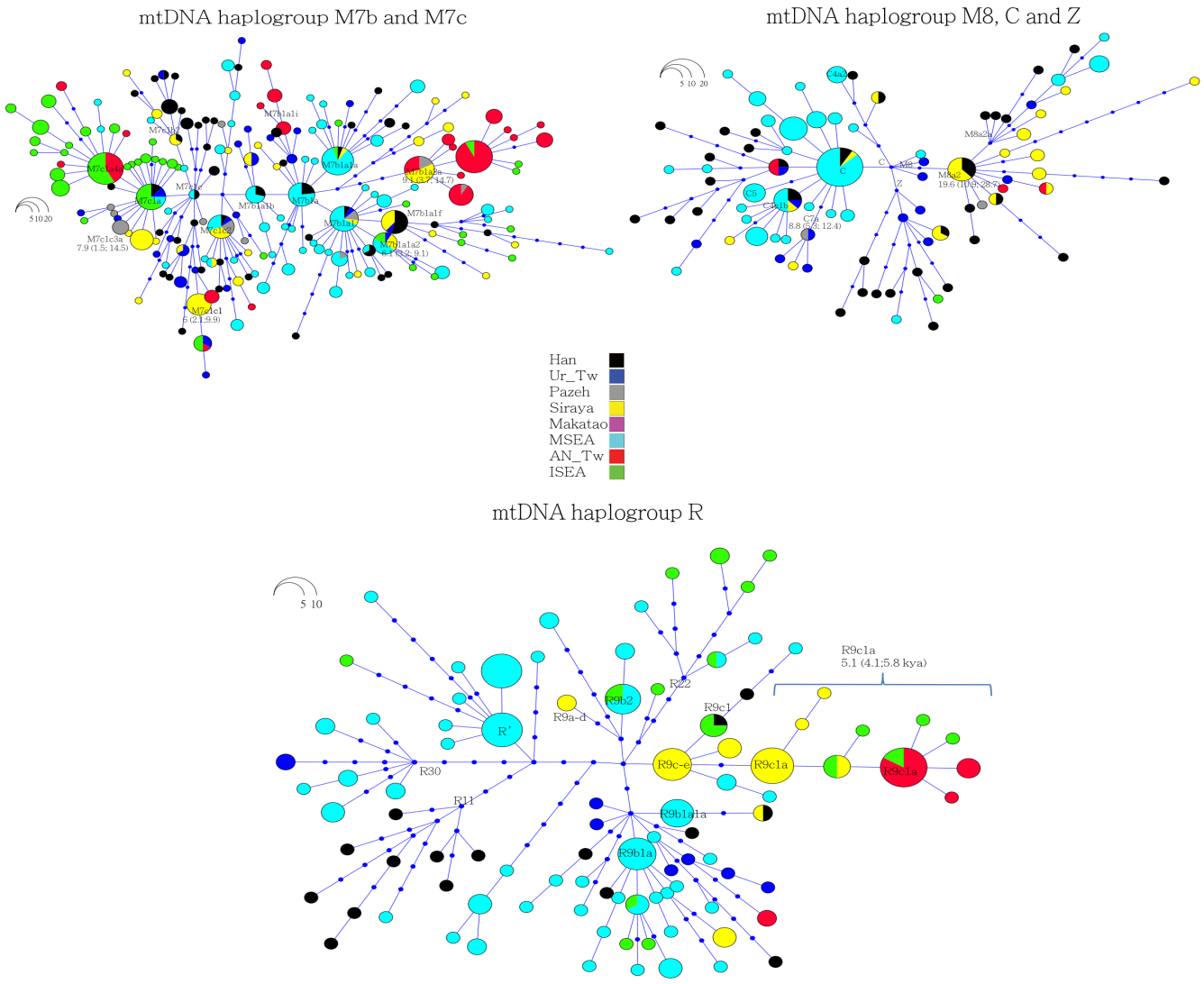
A significant portion, over 38%, of the mitochondrial DNA (mtDNA) and Y-chromosomal DNA (NRY) polymorphisms observed in the Siraya population (as detailed in Supplementary Table S2) were not identified in our control dataset. Unlike HLA, this considerable amount of undefined genetic origin indicates a collection of new lineages, predominantly reflecting unique variations within the Siraya for the NRY, alongside potential laboratory errors. This genetic diversity can be attributed to the cultural and geographical isolation of these groups, coupled with the rapid mutation rates of the NRY and mtDNA systems, which inversely correlates with the extent of undefined shared origins. Further analysis using Phylotree 17 24 enabled us to estimate that 20% of the undefined mtDNA lineages correspond to MSEA origins, mainly from Thailand and Malaysia. Additionally, 51% are linked to other Han groups in China, while the remaining 29% trace back to Austronesian (AN) groups in Taiwan and Island Southeast Asia (ISEA), including two lineages from the Philippines (F3b1b and R9c) that appear to have Negrito ancestry [35].

Principal Component (PC) Analysis

The principal component (PC) plots were created using the Past software [28], utilizing allele and haplogroup frequencies from HLA, mtDNA, and NRY populations in China, Taiwan, Island Southeast Asia, and Mainland Southeast Asia (Figure 2). The analysis accounted for 48.4%, 36.0%, and 69.8% of the total genetic variation, respectively (Figure 2). The HLA PC plot distinctly separated the Siraya from other groups, indicating that prolonged isolation has allowed genetic drift to create a specific HLA profile for the Siraya. In contrast, the Ur_Tw and Han_Tw groups (which include the Hakka and Minnan) are located within the HLA Han and MSEA cluster. This likely suggests that these groups have recently migrated from mainland China to Taiwan. This result is further supported by the mtDNA and NRY PC analyses, which align with the findings presented in our mixture analysis (see Supplementary Table S2). The mtDNA PC analysis indicates a mixture between the Tainan Siraya Group 1 and AN_Tw (Amis and Tsou). Interestingly, the positioning of the Siraya in the PC plots of both mtDNA and the Y chromosome (Figure 2) reveals a connection to the Han and a closer relationship to the Southern AN_Tw groups (including connections with some populations in the Philippines) compared to the Northern AN_Tw. This suggests that the pre-Han settlement Siraya people may have shared a similar migratory history with other Southern AN_Tw groups. However, their settlement in the western plains of Taiwan resulted in genetic sinicization due to the more recent influx of Han groups from China over the past 400 years.



Supplementary Figure S1: mtDNA Haplogroups A, B, C, D, E, F.
Network of mtDNA haplogroups A, B, C, D and F using the concatenated regions of HVS-I, nps 16051 to 16400, and coding regions nps 8001 to 9000 and 9801 to 10900. Network were first generated with the minimum-reduced algorithm, then using the median-joining algorithm for estimating. Coalescence time was calculated using 8,940 years per site. Final graphic plots were processed with the Haploview package version 4.2.



Supplemenary Figure S2: MtDNA haplogroup M and R.
Network of mtDNA haplogroups M7b, M7c, M8, Z and R using the concatenated regions of HVS-I, nps 16051 to 16400, and coding regions nps 8001 to 9000 and 9801 to 10900. Network were first generated with the minimum-reduced algorithm, then using the median-joining algorithm for estimating. Coalescence time was calculated using 8,940 years per site. Final graphic plots were processed with the Haploview package version 4.2.

Supplementary Table S2: Sharing of pooled Siraya groups with control groups.

	Putative parents	HLA	mtDNA	NRY
Sharing of pooled Siraya groups with	All China	86.18%	11.01%	6.09%
	All AN_Tw (No Yami)	0.00%	12.95%	11.09%
	All ISEA	10.75%	11.23%	0.83%
	All MSEA	3.07%	3.27%	1.30%
	Not seen in control data set	0.00%	61.54%	80.68%

A significant portion, over 38%, of the mitochondrial DNA (mtDNA) and Y-chromosomal DNA (NRY) polymorphisms observed in the Siraya population (as detailed in Supplementary Table S2) were not identified in our control dataset. Unlike HLA, this considerable amount of undefined genetic origin indicates a collection of new lineages, predominantly reflecting unique variations within the Siraya for the NRY, alongside potential laboratory errors. This genetic diversity can be attributed to the cultural and geographical isolation of these groups, coupled with the rapid mutation rates of the NRY and mtDNA systems, which inversely correlates with the extent of undefined shared origins. Further analysis using Phylotree 17 24 enabled us to estimate that 20% of the undefined mtDNA lineages correspond to MSEA origins, mainly from Thailand and Malaysia. Additionally, 51% are linked to other Han groups in China, while the remaining 29% trace back to Austronesian (AN) groups in Taiwan and Island Southeast Asia (ISEA), including two lineages from the Philippines (F3b1b and R9c) that appear to have Negrito ancestry [35].

Principal Component (PC) Analysis

The principal component (PC) plots were created using the Past software [28], utilizing allele and haplogroup frequencies from HLA, mtDNA, and NRY populations in China, Taiwan, Island Southeast Asia, and Mainland Southeast Asia (Figure 2). The analysis accounted for 48.4%, 36.0%, and 69.8% of the total genetic variation, respectively (Figure 2). The HLA PC plot distinctly separated the Siraya from other groups, indicating that prolonged isolation has allowed genetic drift to create a specific HLA profile for the Siraya. In contrast, the Ur_Tw and Han_Tw groups (which include the Hakka and Minnan) are located within the HLA Han and MSEA cluster. This likely suggests that these groups have recently migrated from mainland China to Taiwan. This result is further supported by the mtDNA and NRY PC analyses, which align with the findings presented in our mixture analysis (see Supplementary Table S2). The mtDNA PC analysis indicates a mixture between the Tainan Siraya Group 1 and AN_Tw (Amis and Tsou). Interestingly, the positioning of the Siraya in the PC plots of both mtDNA and the Y chromosome (Figure 2) reveals a connection to the Han and a closer relationship to the Southern AN_Tw groups (including connections with some populations in the Philippines) compared to the Northern AN_Tw. This suggests that the pre-Han settlement Siraya people may have shared a similar migratory history with other Southern AN_Tw groups. However, their settlement in the western plains of Taiwan resulted in genetic sinicization due to the more recent influx of Han groups from China over the past 400 years.

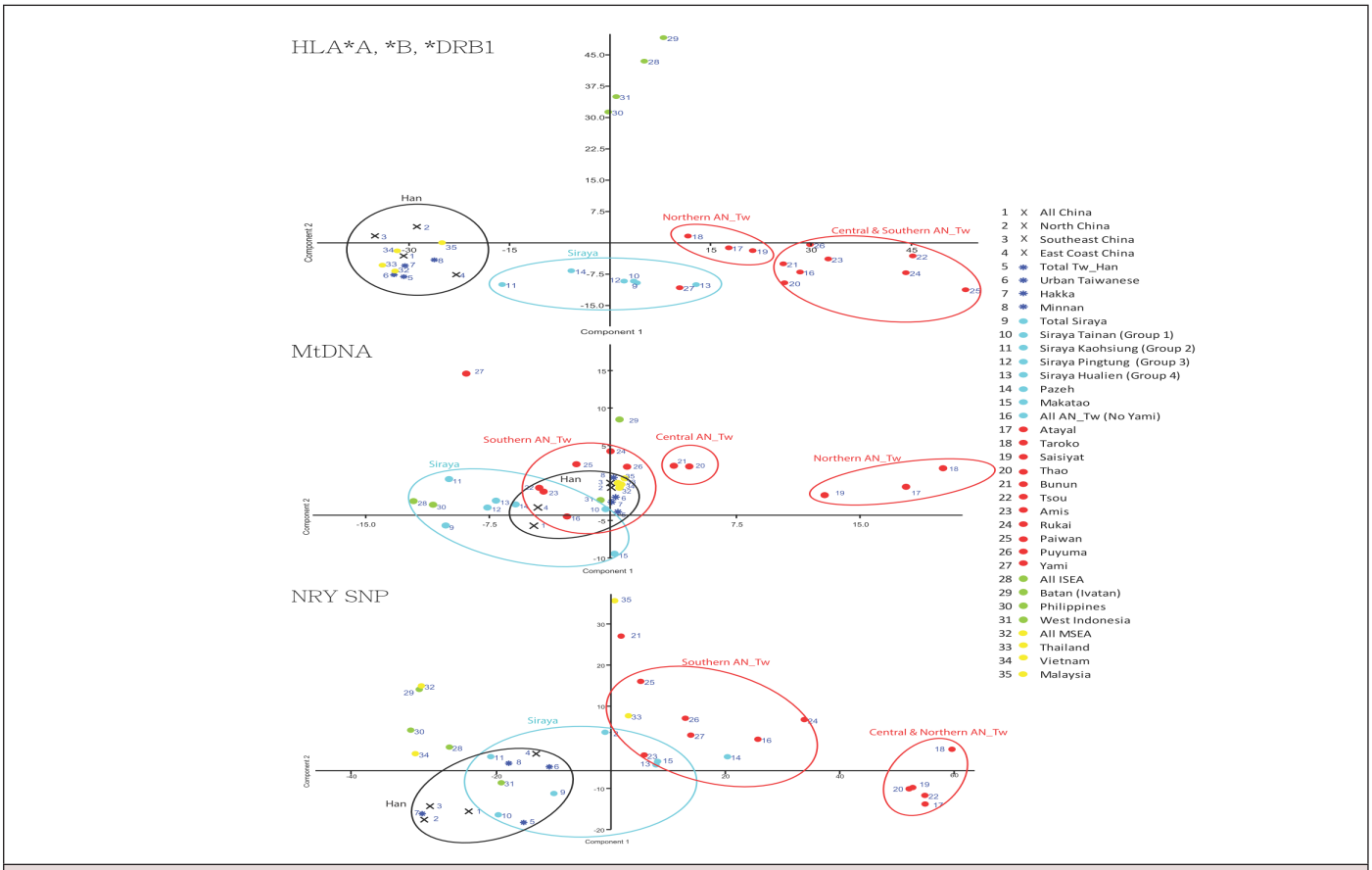


Figure 2: Principal component for HLA *A, *B, *DRB1, mtDNA and NRY.
The first two PCs of the HLA plot explain 48.4% variation, with PC1 (36.2%) principally influenced by alleles A*24:02 and DRB1*09:01, and PC2 (12.1%) influenced by alleles A*11:01, A*23:01, B*15:01, DRB1*12:01 and DRB1*15:01. The mtDNA plot explains 36.4% variation with PC1 mediated by haplogroup M7b1a2a, and PC2 by B4c1b2a, F1a1d1, B4a2a1 and E2b. The NRY plot accounts for 69.8% variation with PC1 affected by NRY haplogroup O1a1 and PC2 by O1a2, O2a1a and O3a2c. Overall, the Siraya (mtDNA and NRY) show affinities with the Han and southern AN_Tw.

Age estimate of settlement

mtDNA

Haplogroup A: Haplogroup A did not exhibit any clusters specific to the Siraya population (Supplementary Figure S1).

Haplogroup B: Haplogroup B displayed three clusters unique to the Siraya (Supplementary Figure S1). One of these clusters belongs to B5a, with an estimated age of 8.1 ± 5.1 Kya and originating from MSEA. The other two clusters are B5a2a2b2 (6.3 ± 4.2 Kya) and B4a2a (3.3 ± 2.8 Kya), both of which were observed only among the AN [36].

Haplogroup C: Haplogroup C (Supplementary Figure S1) revealed a cluster that includes C7a lineages, estimated to be 14.8 ± 4.7 Kya. This cluster is found exclusively in NAN Taiwanese individuals and the Siraya, suggesting a possible expansion from MSEA or South China [21], as well as a recent migration to Taiwan. Notably, there were no terminal clusters specific to the Siraya observed.

Haplogroup A: Haplogroup A did not show any cluster specific to Siraya (Supplementary Figure S1).

Haplogroup B: Haplogroup B showed three clusters unique to Siraya (Supplementary Figure S1). One belonging to B5a (8.1 kya ± 5.1) of MSEA origin, and two others, B5a2a2b2 (6.3 kya ± 4.2) and B4a2a (3.3 ± 2.8 Kya) seen only among AN_Tw [36].

Haplogroup C: Haplogroup C (Supplementary Figure S1) showed a cluster comprising C7a lineages (14.8 ± 4.7 kya) seen only in NAN Taiwanese and Siraya, likely suggesting expansion in MSEA or South China 21 and a recent migrantion to Taiwan since no terminal clusters specific to Siraya were seen.

Haplogroup D: Currently, haplogroup D is primarily found in China. It first arrived in Taiwan with the Austronesian agriculturalists as D5b3a, approximately 6.4 ± 2.1 thousand years ago (kya), and has been identified among the southern Austronesian groups, including the Paiwan, Rukai, and Puyuma. Notably, the haplogroup D found in the Siraya people is not shared with the Austronesian groups in Taiwan and is situated among the twigs of the network plot (Supplementary Figure S1). While all known types of haplogroup D present in the Siraya have Chinese ancestry, only two clusters demonstrate expansion in Taiwan: D5a2a1 (1.7 ± 1.1 kya) and D4h1c (4.4 ± 2.2 kya).

Haplogroup E: Haplogroup E is believed to be a descendant of M9 from southeastern China and is thought to have reached Taiwan around six thousand years ago (BP) with the first Neolithic settlers [37]. The most common subtypes of haplogroup E found among the Austronesian groups in Taiwan include E1a1a1 and E2b, particularly prevalent in the Puyuma and the Siraya Groups 3 (Pingtung) and 4 (Hualian). E1a1a1 is widespread throughout the Austronesian groups, with the highest frequency observed in the Saisiyat and Siraya populations, where its estimated presence dates back to approximately 4.3 ± 1.4 kya. Interestingly, descendants of E1a1a1 are evident in other Austronesian groups, indicating that this haplogroup came along with the first Neolithic settlers. Although E2b is common among the Siraya people, it exhibits low diversity. Its estimated presence in Taiwan ranges from 4.3 to 5.7 ± 2.3 kya.

Haplogroup F is found throughout continental East Asia, Taiwan, and Island Southeast Asia (ISEA) [21]. However, certain subtypes among the Taiwanese Indigenous people (Tw_AN) specifically F1a3a, F1a4a, and F3b1a2, are unique to this group. Notably, the subtypes F1a3a3 and F1a4a1a are exclusive to the Siraya people and fall under the F1a haplogroup, which is also specific to Tw_AN. Additionally, F3b1b in the Siraya represents a sister clade to F3b1a2 in the Ancestral Taiwanese (AN_Tw) lineage (Supplementary Figure S1). The diversity of haplogroup F among the Siraya is the highest observed in East Asia, following China. The terminal clusters F1a4a1a (estimated to have existed around 3.3 ± 1.2 Kya), and F3b1b (3.3 ± 0.5 Kya) provide insight into the time of isolation of the Siraya people. While this supports the notion of an Austronesian primary origin for the Siraya, the absence of the NAN F cluster among them suggests that the bearers of F1 and F3 haplogroups in the Siraya represent a recently mixed group.

Haplogroups M7b and M7c: Haplogroups M7b and M7c (Supplementary Figure S2) represent two significant subclasses that are present throughout East Asia, Taiwan, and Island Southeast Asia (ISEA). Notably, in the Siraya population, M7b1a2a (with an age estimate of 9.1 kya, ranging from 3.7 to 14.7 kya) appears to be ancestral to other subtypes of this clade, specifically in the AN_Tw populations [33,38]. The M7b1a1a2 subtype of M7b1 (estimated at 11.5 kya, ranging from 5.5 to 17.6 kya) is found in the Siraya and the NAN groups of Taiwan and is thought to result from a mixture with populations from Mainland Southeast Asia (MSEA). Additionally, subtypes of M7c1c1 (approximately 6 kya, with a range of 2.1 to 9.9 kya) are primarily found across Taiwan and ISEA, although their rare occurrence among the Hakka suggests that Austronesian speakers may have been introduced in Hakka communities. Finally, the closest ancestors of M7c1c3a (about 7.9 kya, ranging from 1.5 to 14.5 kya) have been reported in MSEA and Indonesia, likely arriving in Taiwan along with the first Neolithic settlers.

Haplogroups M8, C, and Z: The C and Z clades (see Supplementary Figure S2) are descendants of M8, which likely originated in Thailand and South China [39]. Haplogroups M8 and Z have not produced any terminal clusters that allow for significant estimates of coalescence within the Siraya population. However, a C7a cluster, which includes Ur_Tw, a single Siraya sample, and Makatao samples, has produced an estimate of local expansion in Taiwan at approximately 8.8 kya (ranging from 5.3 to 12.4 kya). The coalescence of this clade in Thailand, estimated at 13.6 kya (ranging from 14.9 to 18.4 kya), suggests early migration from MSEA to Taiwan.

Haplogroup R: Subtypes of haplogroup R are widespread across East Asia, with R11 found in China, R30 in India, R9b in MSEA, and R9c predominating among the Negrito population of Palawan (52%, Philippines) [40]. R9c is also common in MSEA [39], Indonesia [22], Austronesian populations in Taiwan (AN_Tw) [41], and the Siraya population (over 2.3%). Notably, R9c1a appears to be ancestral to the cluster observed among AN_Tw, indicating that R9c bearers were likely part of the Neolithic settlement in Taiwan.

NRY

The non-recombining Y chromosome (NRY) O haplogroup is the dominant type in East Asia. It includes three major subgroups: O1a, O1b_P31, and O2_M122, which are primarily found in Northeast Asia, Southeast Asia, and the Mainland Southeast Asia (MSEA), respectively. Among the Austronesian groups, the AN_Tw and Siraya populations are descendants of haplogroup O, and represents the major Neolithic migration towards Taiwan and Island Southeast Asia (ISEA). In contrast to the NAN_Tw group, 93.6% of the O haplogroup profile observed in Siraya is shared with either NAN_Tw (O2_M122) or Southern AN_Tw (O1a1_P203 or O1a2_M110). Haplogroup O1a1_P203, along with the less common O1a_M119, is prevalent among Southern AN_Tw, ISEA, and in the Fujian Province of China, where the O1a_M119 subtypes likely originated [42]. O1a2_M110 is also dominant among the Southern AN_Tw but shows a decline toward ISEA, except in the Nias Islands south of Sumatra [43], where it has been observed in some Daic-speaking groups in Southeast China [44]. Haplogroup O1b_P31 in the Siraya population is mainly represented by the subtype O1b1a1a1a1a1_M88. This subtype is found throughout MSEA and is particularly prominent in Malaysia [45]. In a haplotype network (data not shown), the branches of O1b_P31 in Siraya mostly occupy the tips of the clade indicating recent diversity; however, some deeper nodes suggest an origin in MSEA [45]. Haplogroup O2a1c_JST002611 is commonly found in the Fujian Province of China but is rare among Austronesian groups, with the exception of the Bunun. Its presence in the Siraya population suggests early dispersal, likely during the Taiwan Metal Age from Southeast Asia [46]. Haplogroup O2a1b_P164 is also distributed across Fujian, Taiwan, and ISEA. Unlike O1a1, O1b_P31, and O2a1c_JST002611, the network associated with O2a1b_P164 indicates a possible early expansion among the Siraya, dating back to approximately 3.9 kya BP ± 2.5 kya [36].

Bayesian Analysis of Population Structure (BAPS)

Gene Flow and the Human Leukocyte Antigen (HLA):

The population structure, genetic diversity, and admixture, along with the relative average of the HLA ancestry of the Siraya groups with other groups from East Asia using BAPS 30 are illustrated in Figure 3. We determined the best spatial matching using the average log-likelihood of K (Delta K), which showed an optimal value at K = 10 when evaluated from K = 2 to K = 20 31. The Siraya cluster consists of two groups: one includes all Siraya subgroups, and the other consists of the Pazeh group, a heavily Sinicized group from the western plains. This cluster exhibits an 18% genetic affinity with groups from East China (Fujian) and urban Taiwanese populations (Ur_Tw), as shown in Figure 3B. Although we applied a 10% pruning of the gene flow to enhance the visual representation, we estimate that the residual relationship between the Siraya and other groups primarily reflects AN_Tw ancestry.

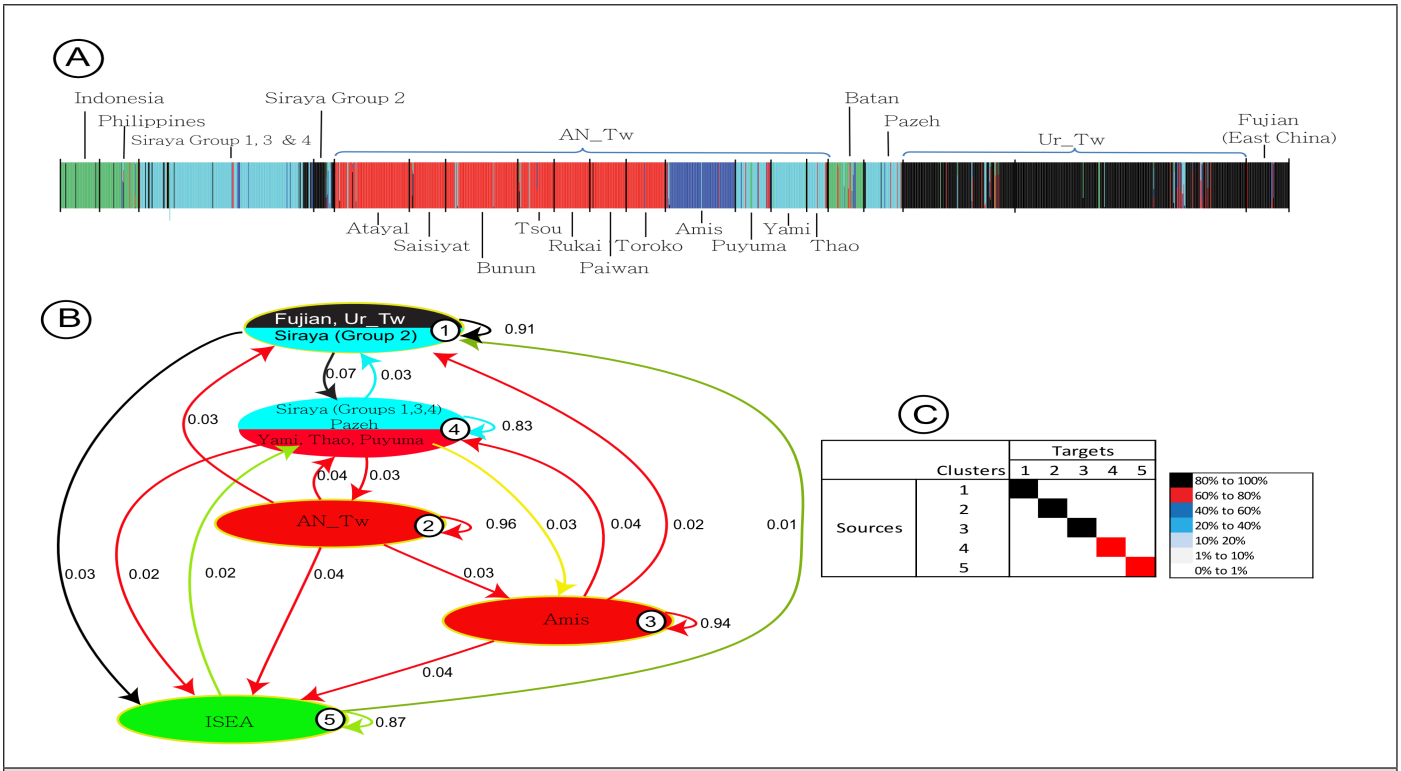
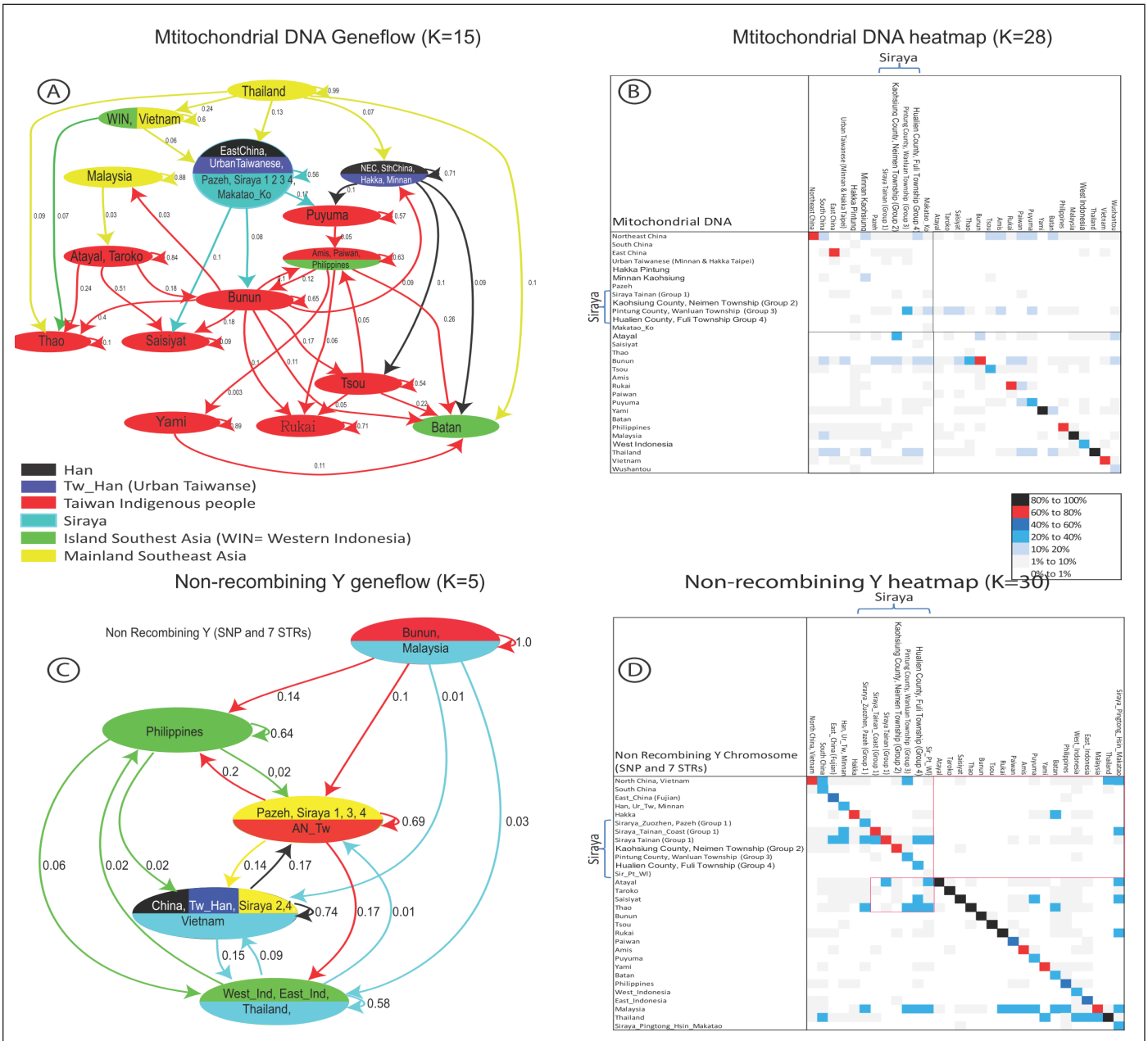


Figure 3: Bayesian Analysis of Population Structure (BAPS) using HLA *A, *B and *DRB1 gene systems for K=10. Admixture analysis of the Siraya population. (A) Each column represents a single individual and is colored according to the proportion of genetic variation assigned to each BAPS cluster. Each cluster represents the upper hierarchical level of the structure at K = 10 and p < 0.01. (B) Gene flow network identified between clusters of populations. Arrows show the average fraction of genetic variation from the source cluster to the target cluster. Note: we pruned gene flow under 10% to improve the graph visual. Back loops arrows show the within-cluster original profile (C) Heatmap summarizing back loops and gene flows. (D) Phylogenetic tree based on Nei's distances (averaged over loci).

The heat map depicting gene flow for all HLA groups at K = 10 (Figure 3C) reveals a slightly higher gene source within clusters for most groups. However, this gene source is significantly lower for the Siraya cluster, which shows only 15% (illustrated by the looping arrows in Figure 3B and the diagonal in Figure 3C). Due to the slow mutation rate of the HLA gene system, the gene source within clusters primarily reflects a genetic profile and distribution shaped by genetic drift, admixture, and haplotype recombinations, rather than the structural gene diversity and mutations accumulated over millennia. Although the Siraya cluster appears to be a mixture of Han and Austronesian (AN) groups (as shown in Figures 3A and 3B), the phylogenetic analysis (Figure 3D) indicates that the Siraya cluster has shifted closer to the AN clade. Additionally, the diploid multi-locus sequence typing analysis of the HLA system did not provide any insights regarding sex linearity. Therefore, we further examined the relationship of the Siraya cluster with other regions using single-sex gene systems, including mitochondrial DNA (mtDNA) and Y-chromosome (NRY) analysis (Supplementary Figure S3).



Supplementary Figure S3: A and C: Bayesian analysis of gene flow was conducted for mitochondrial DNA (MtDNA) with K=15 and nuclear Y chromosome (NRY) with K=5, at a significance level of $p < 0.01$.
B and D display heatmaps generated using the BAPS option "Admixture of individuals based on mixture clustering," for K=28 and K=30, respectively, also at $p < 0.01$. The diagonal of the heatmaps represents the proportion of gene sources originating from within the BAPS clusters, while the other cells illustrate the gene flow between source and target populations.

MtDNA and NRY Gene Flow:

MtDNA: The best spatial match for mtDNA, measured using the average log-likelihood 31 was achieved at K=6. To better illustrate the influences of North and South China from East China (represented by Fujian), we extended the gene flow plot to K=15 (Supplementary Figure S3 A) while maintaining the original six colors that represent K=6. In addition, all Siraya groups showed a strong affinity for East China, likely due to the initial settlement location of the Siraya in Tainan County and the fact that the first NAN settlers predominantly came from Southeast China and Fujian.

NRY: In contrast, the NRY gene flow (Supplementary Figure S3 C) reveals considerable within-cluster diversity and significant bidirectional gene flow between the two clusters containing the Siraya groups. It also indicates that all groups in Taiwan have been influenced by the presence of NRY haplogroups.

MtDNA and NRY Heatmaps:

MtDNA: The corresponding mtDNA Heatmap was run for K=28 (Supplementary Figure S3 B). Contrary to the AN_Tw groups, and except for Siraya group 3 (Pingtong), Siraya group 1, 2, and 4 showed low within group diversity. As indicated in Table 1, these groups are scattered throughout Taiwan and likely have been strongly influenced by local mixture. Furthermore, the top right quadrant shows mtDNA mixture between Northeast China and AN_Tw suggesting the majority of settlers in the last 400 years were probably males who intermarried with AN_Tw.

NRY: To the contrary of the mtDNA heatmap, NRY heatmap (Supplementary Figure S3 D) shows strong within cluster diversity for all groups indicating the group's isolation after initial settlement. However, the Siraya have more NRY mixtures with NAN_Tw than mtDNA with AN_Tw.

Conclusion

The Siraya people are among the original inhabitants of Taiwan, with a rich cultural heritage that spans several thousand years. Today's population reflects the influence of Han Chinese settlers over the past four hundred years. Despite this demographic change, the Siraya have successfully maintained a distinct cultural identity, including their customs and traditions, and have been actively working to revive their original Austronesian language. When using genetic systems with a low mutation rate, such as the polymorphic multi-loci HLA diploid system, classifying mixed ethnic groups like the Siraya can be challenging due to significant genetic recombination. However, our analyses of the Y-chromosome (NRY), mitochondrial DNA (mtDNA), and HLA indicate that the Siraya people, apart from their current language use, have preserved their original Austronesian genetic heritage. We advocate for the Siraya to be recognized as an officially acknowledged indigenous group in Taiwan. It is important to conduct further research to confirm this, and, as with any genetic studies, we must approach the investigation of the Siraya with respect for their cultural heritage and rights.

Acknowledgements

The project was supported by a grant from the Ministry of Science and Technology of

Taiwan #106-2320-B-195-001-MY2. We acknowledge with gratitude the skilled assistance of Ying-Hui Lai, and Lan-Rong Chen from the Mackay Memorial Hospital in Taipei (Taiwan). We also greatly appreciate the kindness of the Siraya peoples who donated their bio-samples for this investigation.

Author contributions

J.A.T. and M.L conceived the project, M.L. collected samples and contributed reagents. J.Y.H and J.A.T. designed the experiments, J.Y.H and Z.S.C did the bench work, J.Y.H and J.A.T. did the analysis, J.A.T. wrote the manuscript.

Data availability

Raw data for mtDNA partial sequencing of all Siraya samples and NRY data for SNP and seven STRs are reported in Supplementary Table S3.

MtDNA complete sequencing of Siraya samples is reported in Supplementary data 1. Allele and haplogroup frequencies, Supplementary Table S1, Supplementary Table 3, Supplementary Data 1 may be requested from the corresponding author.

Supplementary information: Supplementary material is available at the journal

website: https://lupinepublishers.com/anthropological-and-archaeological-sciences/

References

1. Li Ruiyuan (2015) From Single Society to Community: The Formation of Sideia in the Seventeenth Century. (National Cheng Kung University (Taiwan).
2. Jacobs J B (2014) Taiwans Colonial Experiences and the Development of Ethnic Identities: Some Hypotheses. Asian Languages and Studies, Monash University. Taiwan Comp. Perspect 5: 47-59.
3. Chiang M (2022) (Radio T. I. Indigenous group secures major achievement towards recognition. Taiwan News 14 November.
4. Li P J K (2001) The Dispersal of The Formosan Aborigines in Taiwan. Lang. Linguist 2(1): 271-278.
5. Ferrell R (1971) Aboriginal Peoples of the Southwestern Taiwan Plains. Bull. Inst. Ethnol 217–235.
6. Nakamura T (National T. U. Household Registration of Indigenous Villages During the Dutch Occupation. in Taiwan History During the Dutch Occupation (in Chinese); National Library of Taiwan (eds. Wu, T.-M., Wong, K.-I. &Shu, S.-Y.) 1 1-38.
7. Liu Y C (2002) Prehistory and Austronesians in Taiwan: an archeological perspective. (The Phoebe Museum of anthropology).
8. Underhill P A, Kivisild T (2007) Use of y chromosome and mitochondrial DNA population structure in tracing human migrations. Annu. Rev. Genet 41: 539-564.
9. Mitchell R J, Hammer M F (1996) Human evolution and the Y chromosome. Curr. Opin. Genet. Dev 6(6): 737-742.
10. Lin M, C C Chu, S L Chang, H L Lee, J H Loo, et al (2001) The origin of Minnan and Hakka, the so-called 'Taiwanese', inferred by HLA study. Tissue Antigens 57(3): 192-199.
11. Gonzalez-Galarza F F, Antony McCabe, Eduardo J Melo Dos Santos, James Jones, Louise Takeshita, et al (2020) Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools. Nucleic Acids Res 48(D1): 783-788.
12. Trejaut J A, Toomas Kivisild, Jun Hun Loo, Chien Liang Lee, Chun Lin He, et al (2005) Traces of archaic mitochondrial lineages persist in Austronesian-speaking Formosan populations. PLoS Biol 3(8): 247.
13. Chu C C, Lee H L, Trejaut J, Chang H L, Lin M et al (2004) HLA-A, -B, -Cw and -DRB1 allele frequencies in Ami, Atayal, Bunun. Hakka, Paiwan, Pazeh, Puyuma, Rukai, Saisiat, Tsou, Taroko, Thao and Tao populations from Taiwan. Hum. Immunol. Spec. Issue HLA alleles other Immunogenet. Polymorph. Freq. from worldwide Popul. Guest Ed. Derek Middelton, John Sanil Manavalan, Marcelo A. Fernandes-Vina. ASHI 65, 1102-1181.
14. Trejaut J A, Estella S Poloni, Ju-Chen Yen, Ying-Hui Lai, Jun-Hun Loo, et al (2014) Taiwan Y-chromosomal DNA variation and its relationship with Island Southeast Asia. BMC Genet 15: 77.
15. Chen L-R, Jean Alain Trejaut, Ying-Hui Lai, Zong-Sian Chen, Jin-Yuan Huang, et al (2019) Mitochondrial DNA Polymorphisms of the Saisiyat Indigenous Group of Taiwan. Search for a Negrito Signature. Edelweiss J. Biomed. Res. Rev. Rev 1: 12-18.
16. Luo X Q, Pan-Xin Du, Ling-Xiang Wang, Bo-Yan Zhou, Yu-Chun Li, et al (2020) Uniparental Genetic Analyses Reveal the Major Origin of Fujian Tanka from Ancient Indigenous Daic Populations. Hum Biol 91(4): 257-277.
17. Tabbada K A, Jean Trejaut, Jun-Hun Loo, Yao-Ming Chen, Marie Lin, et al (2010) Philippine mitochondrial DNA diversity: a populated viaduct between Taiwan and Indonesia? Mol Biol Evol 27(1): 21-31.
18. Trejaut J, Lee C L, Yen J C, Loo J H, Lin M, et al (2011) Ancient migration routes of Austronesian-speaking populations in oceanic Southeast Asia and Melanesia might mimic the spread of nasopharyngeal carcinoma. Chin J Cancer 30(2): 96-105.
19. Loo J H, Jean A Trejaut, Ju-Chen Yen, Zong-Sian Chen, Chien-Liang Lee, et al (2011) Genetic affinities between the Yami tribe people of Orchid Island and the Philippine Islanders of the Batanes archipelago. BMC Genet 12: 21.
20. Duong N T, Enrico Macholdt, Nguyen Dang Ton, Leonardo Arias, Roland Schröder, et al (2018) Complete human mtDNA genome sequences from Vietnam and the phylogeography of Mainland Southeast Asia. Sci Rep 8(1): 11651.
21. Kutanan W, Jatupol Kampuansai, Metawee Srikummool, Daoroong Kangwanpong, Silvia Ghirotto, et al (2017) Complete mitochondrial genomes of Thai and Lao populations indicate an ancient origin of Austroasiatic groups and demic diffusion in the spread of Tai-Kadai languages. Hum Genet 136(1): 85-98.
22. Hill C, Pedro Soares, Maru Mormina, Vincent Macaulay, Dougie Clarke, et al (2007) A mitochondrial stratigraphy for island southeast Asia. Am J Hum Genet 80(1): 29-43.
23. Jinam T A, Lih-Chun Hong, Maude E Phipps, Mark Stoneking, Mahmood Ameen, et al (2012) Evolutionary history of continental southeast asians: Early train hypothesis based on genetic analysis of mitochondrial and autosomal DNA data. Mol. Biol. Evol 29(11): 3513-3527.
24. van Oven M, Kayser M (2009) Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Hum Mutat 30(2): 386-394.
25. Weissensteiner H, Dominic Pacher, Anita Kloss-Brandstätter, Lukas Forer, Günther Specht, et al (2016) HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. Nucl. Acids. Res 44(W1): W 58-63.
26. Karafet T M, Fernando L Mendez, Monica B Meilerman, Peter A Underhill, Stephen L Zegura, et al (2008) New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. Genome Res 18(5): 830-838.
27. ISSOG. International Society of Genetic Genealogy. Y-DNA Haplogroup Tree 2019, Version: 15.46 Date: 12 March 2020, (2019).
28. Hammer O, Harper D A T, Ryan P D Package (2001) PAST: Paleontological Statistics Software. Paleontological Statistics Software Package for Education and Data Analysis. Palaeontol. Electron 4(1): 9.
29. Maca-Meyer N, Matilde Arnay, Juan Carlos Rando, Carlos Flores, Ana M González, et al (2004) Ancient mtDNA analysis and the origin of the Guanches. Eur J Hum Genet 12(2): 155-162.
30. Corander J, Marttinen Pekka, Mäntyniemi Samu (2006) A Bayesian Method for Identification of Stock Mixtures from Molecular Marker Data. Fish. Bull 104(4): 550-558.
31. Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. Mol. Ecol 14(18): 2611-2620.
32. Morin P A, Richard G Leduc, Frederick I Archer, Karen K Martien, Ryan Huebinger, et al (2009) Significant deviations from Hardy-Weinberg equilibrium caused by low levels of microsatellite genotyping errors. Mol. Ecol. Resour 9(2): 498-504.
33. Soares P, Luca Ermini, Noel Thomson, Maru Mormina, Teresa Rito, et al (2009) Correcting for purifying selection: an improved human mitochondrial molecular clock. Am J Hum Genet 84(6): 740-759.

34. Zhivotovsky L A, Peter A Underhill, Cengiz Cinnioğlu, Manfred Kayser, Bharti Morar, et al (2004) The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. Am J Hum Genet 74(1): 50-61.

35. Delfin F, Albert Min-Shan Ko, Mingkun Li, Ellen D Gunnarsdóttir, Kristina A Tabbada, et al (2014) Complete mtDNA genomes of Filipino ethnolinguistic groups: a melting pot of recent and ancient lineages in the Asia-Pacific region. Eur J Hum Genet 22(2): 228-237.

36. Trejaut J A, Frank Muyard, Ying-Hui Lai, Lan-Rong Chen, Zong-Sian Chen, et al (2019) Genetic diversity of the Thao people of Taiwan using Y-chromosome, mitochondrial DNA and HLA gene systems. BMC Evol Biol 19(1): 64.

37. Soares P, Jean Alain Trejaut, Jun-Hun Loo, Catherine Hill, Maru Mormina, et al (2008) Climate change and postglacial human dispersals in southeast Asia. Mol Biol Evol 25(6): 1209-1218.

38. Soares P A, Jean A Trejaut, Teresa Rito, Bruno Cavadas, Catherine Hill, et al (2016) Resolving the ancestry of Austronesian-speaking populations. Hum Genet 135: 309-326.

39. Kutanan W, Kampuansai J, Brunelli A, Silvia Ghirotto, Pittayawat Pittayaporn, et al (2018) New insights from Thailand into the maternal genetic history of Mainland Southeast Asia. Eur J Hum Genet 26(6): 898-911.

40. Scholes C, Katherine Siddle, Axel Ducourneau, Federica Crivellaro, Mari Järve, et al (2011) Genetic diversity and evidence for population admixture in Batak Negritos from Palawan. Am J Phys Anthr 146(1): 62-72.

41. Ko A M, Chung-Yu Chen, Qiaomei Fu, Frederick Delfin, Mingkun Li, et al (2014) Early Austronesians: into and out of Taiwan. Am J Hum Genet 94(3): 426-436.

42. Wang C C, Li H (2013) Inferring human history in East Asia from Y chromosomes. Investig Genet 4: 11.

43. Van Oven M, Johannes M Hämmerle, Marja van Schoor, Geoff Kushnick, Petra Pennekamp, et al (2011) Unexpected Island effects at an extreme: reduced Y chromosome and mitochondrial DNA diversity in Nias. Mol Biol Evol 28(4): 1349-1361.

44. Li P J K (2008) Time perspective of Formosan Aborigines. In Past Human Migrations in East Asia: Matching Archaeology, Linguistics and Genetics R.B. A Sanchez- Mazas, M Ross, I Peiros, M Lin. ed 211-218.

45. Brunelli A, Jatupol Kampuansai, Mark Seielstad, Khemika Lomthaisong, Daoroong Kangwanpong, et al (2017) Y chromosomal evidence on the origin of northern Thai people. PLoS One 12(7): e0181935.

46. Lin M, Trejaut J A (2023) Diversity and distribution of mitochondrial DNA in non-Austronesian-speaking Taiwanese individuals. Hum. Genome Var 10(2): 2.