Bull. Natl. Mus. Nat. Sci., Ser. D, 40, pp. 1–12, November 22, 2014

Mitochondrial genetic diversity of Pingpu tribes in Taiwan

Ken-ichi Shinoda[1], Tsuneo Kakuda[2], Hideaki Kanzawa-Kiriyama[1], Noboru Adachi[2], Pei-Ying Tsai[3,4], Hsi Kuei Tsai[5]

[1] Department of Anthropology, National Museum of Nature and Science, 4–1–1, Amakubo, Tsukua City, Ibaraki prefecture 305–0005, Japan
[2] Department of Legal Medicine, Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi, 1110 Shimo-kato, Chuo 409–3898, Yamanashi, Japan
[3] Center of Faculty Development, National Taiwan University college of Medicine
[4] Museum of Medical Humanities, National Taiwan University
[5] Department of Anatomy, Faculty of Medicine, National Taiwan University, No. 1 Jen Ai Road Section 1, Taipei 10051, Taiwan R.O.C.

Abstract The origin and demographic history of the aboriginal populations of Taiwan have not been clearly resolved. To investigate the genetic characteristics of the Plains tribes or the so-called Pingpu population, thought to be originally Austronesian-speaking inhabitants of Taiwan, the mitochondrial DNA (mtDNA) hypervariable segment 1 (HVS1) region and diagnostic variants in coding regions in 35 individuals sampled from historical skeletal remains obtained from National Taiwan University were analyzed. The Pingpu population showed high haplotype diversity and nucleotide diversity compared with the populations of other aboriginal Taiwanese and southern Chinese populations, which probably reflects their complex history. The analysis showed that the haplogroups that are predominant in southern East Asia, including F, B, and M7, account for 64% of mtDNA in Pingpu populations. Genetic characteristics of the Pingpu population differ from those of other mountain aboriginal people of Taiwan. Principal component analysis of haplogroup frequencies demonstrated closer relatedness between Pingpu and other southern China lineages, especially from Guangdong, Hainan Island, and North Vietnam, suggesting a general south China origin of maternal lineages in the Pingpu populations. In spite of the recent gene flow from the Han Chinese living in Taiwan, maternal lineages of the Pingpu population could be traced back to southern China during the Neolithic period.

Key words: Mitochondrial DNA, Haplogroup, Taiwan, Ethnic group, Pingpu tribe, Population genetics.

Introduction

Analyzing patterns of genetic diversity from living groups to study population history has been a burgeoning field of population genetics over the past three decades. Moreover, ancient DNA analysis has become an increasingly important tool in elucidating the geographical origins of ancient populations and their relationships. Because of the complex history and prehistory of Taiwan and the associated expansion of the Austronesian language family (Bellwood, 2000; Gray and Jordan, 2003), it is not surprising that indigenous tribes of Taiwan have often attracted the interest of population geneticists (e.g. Su *et al.*, 1999; Lin *et al.*, 2000, 2005; Chu *et al.*, 2001; Kayser *et al.*, 2000; Tajima *et al.*, 2003). Moreover, a large number of mtDNA analyses have revealed the genetic characteristics of living indigenous mountain people (Trejaut *et al.*, 2005). Although the molecular analyses of ancient bones has become one of the more frequently used methods in recent years, few studies have focused on the population genetics of archaeological sites (Shinoda, 2008, Ko *et al.* 2014). Because of the lack of the DNA data from historical samples, we still know very little about the genetic structure of ancient populations.

During the last glaciation (about 18,000–15,000 years before present [BP]), Taiwan was connected to the Eurasian continent. Although sporadic human skeletal remains have been found from that time, there is a lack of adequate evidence. Certain skeletal evidence dating back to the Neolithic period has been found. The oldest culture in Taiwan—the Tapenkeng culture—had spread to the coastal area of Taiwan after 4,500 years BC. This culture retained continuity with the Neolithic culture of the opposite shore of the Eurasian continent. It is obvious that the Tapenkeng culture was brought to Taiwan by immigrants of southern China. These people were thought to be Austronesian language speakers. They had occupied all of Taiwan until the 17th century and continue to do so to date.

To clarify the population history of the aboriginal tribes in Taiwan, ancient DNA analysis was conducted on human skeletal remains housed in the National Taiwan University Faculty of Medicine. The samples analyzed belonged to the Pingpu tribes, also called the Plains tribes, which mainly settled in the western plains of Taiwan and are distinct from aboriginal mountain people. However, the discrimination is tentative. Both mountain and plains tribes are of Austronesian descent.

The Pingpu people had a primitive agricultural lifestyle before the arrival of Han immigrants. Since then, they faced strong forces from the Dutch, Spanish, and Han Chinese and were deeply influenced by these foreign cultures, especially Han culture. As a result, their native languages and cultures diminished and got lost. The Pingpu people originally lived in western Taiwan. When faced with oppression by the Han Chinese 400 years ago, some Pingpu tribes chose to depart from their homeland and move to other plains of Taiwan.

Although some anthropological investigations and research on the Plains tribes have been performed, mainly 60 years ago (e.g., Hsu, 1947; Wang, 1949; Yuan, 1960), no genetic analyses have been performed on them. Moreover, the genetic composition of aboriginal Taiwanese has been altered in the face of the recent industrialized society, and the current population may not reflect the original composition. Thus, the original genetic characteristics of the Plains tribes are not clear. The aims of this study were to characterize the genetic composition of the ancient Pingpu tribes and to address their genetic relationship with other populations of Taiwan and Southeast Asia. This is the first systematic study on the genetic structure of the Plains populations.

Materials and Methods

Archaeological sites

In this analysis, we used human skeletal remains excavated from two Pingpu archaeological sites (Ogulan and Xiluo) in Taiwan. The approximate locations of these sites are shown in Figure 1. Skeletal remains had been excavated by Dr. Kanaseki of the Imperial University of Taihoku before World War II and these tombs were used until 19th centuries. The remains are now housed at the department of Anatomy, Faculty of Medicine of National Taiwan University.

The Ogulan site is located in the Puli Basin, which is one of the largest basins located around the mountainous area of Nantou in central Taiwan. The Pingpu people who settled in the Ogulan site belonged to the "Pazeh (Pazih) tribe," which was originally located in Fengyuan of Taichung County. The Xiluo site is located in an urban township in Yunlin County. The earliest inhabitants here were members of the "Babuza tribe," which is one of the Pingpu tribes of central Taiwan.

Archeological specimens

Tooth enamel forms a natural barrier to exogenous DNA contamination. Furthermore, the DNA recovered from teeth appears to lack most of the inhibitors of the enzymatic amplification of ancient DNA (Woodward *et al.* 1994). Therefore, tooth samples were used in the present analysis. A total of 35 well-preserved tooth samples (22 from Ogulan and 13 from Xiluo) were selected for DNA analysis. A list of all samples used in this study is presented in Table 1.

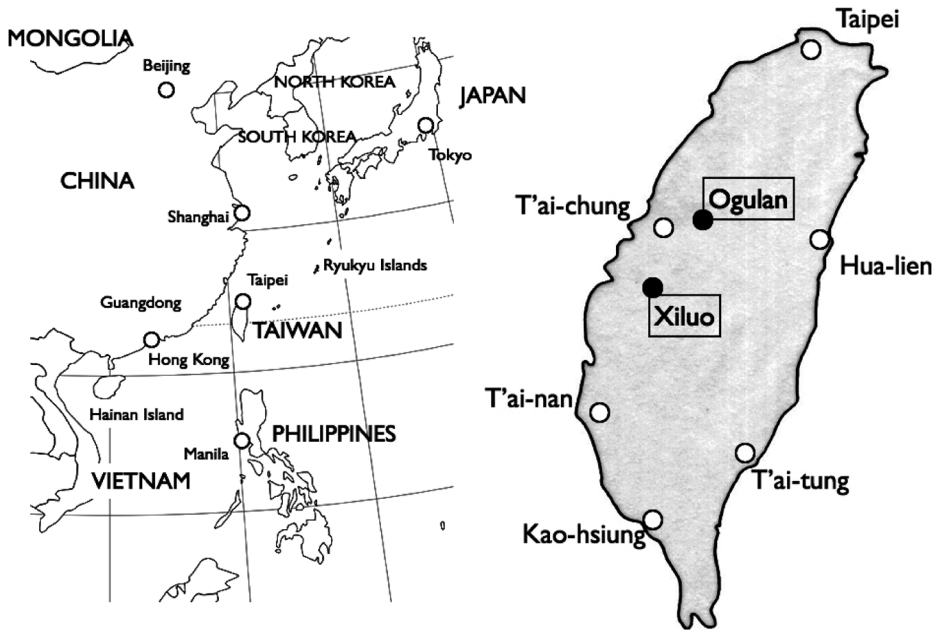

Fig. 1. The map of East and Southeast Asia (left) and location of the Xilou and Ogulan sites (right).

Table 1. Sample used for DNA exreaction

No.	Code Name	Sample					No.	Code Name	Sample				
		portion	side	type	sex	age			portion	side	type	sex	age
	Ogulan site							*Xilou site*					
1	53	Mandible	Right	M3	M	Mature	1	23	Mandible	Left	M3	M	Senior
2	113	Mandible	Right	M2	F	Mature	2	40	Mandible	Right	M2	M	Mature
3	561	Mandible	Right	M3	F	Mature	3	117	Mandible	Right	M2	M	Mature
4	562	Mandible	Left	M3	F	Mature	4	118	Maxilla	Right	M3	M	Mature
5	565	Mandible	Right	M2	F	Adult	5	119	Maxilla	Left	M2	M	Mature
6	566	Maxilla	Right	M3	F	Mature	6	529	Maxilla	Right	M3	F	Senior
7	568	Mandible	Right	M3	F	Mature	7	533	Mandible	Left	M3	F	Mature
8	569	Mandible	Right	M3	F	Adult	8	536	Mandible	Right	M2	F	Mature
9	570	Maxilla	Left	M3	F	Mature	9	539	Mandible	Left	M2	F	Senior
10	572	Mandible	Left	M3	F	Mature	10	541	Maxilla	Right	M2	M	Mature
11	605	Maxilla	Right	M2	M	Senior	11	543	Maxilla	Right	M2	M	Senior
12	608	Mandible	Left	M3	M	Senior	12	546	Maxilla	Right	M2	M	Mature
13	610	Mandible	Left	M3	M	Mature	13	548	Maxilla	Left	M2	M	Senior
14	613	Mandible	Right	M3	M	Mature							
15	615	Mandible	Left	M3	M	Senior							
16	617	Mandible	Left	M2	M	Mature							
17	618	Maxilla	Left	M3	M	Senior							
18	621	Maxilla	Right	M3	F	Mature							
19	622	Mandible	Left	M3	M	Mature							
20	624	Mandible	Right	M3	F	Senior							
21	626	Maxilla	Right	M3	M	Mature							
22	635	Maxilla	Right	M3	M	Senior							

* Sex and age were judged from morphologocal feature.

Authentication methods and extraction of DNA

DNA analyses were performed at the National Museum of Nature and Science and Yamanashi University, which are dedicated to ancient DNA analysis. We employed standard precautions to avoid contamination, e.g., separation of pre- and post-PCR experimental areas, use of disposable laboratory ware and filter-plugged pipette tips, treatment with DNA contamination removal solution (DNA Away; Molecular Bio Products, San Diego, CA, USA), UV irradiation of equipment and benches, and negative extraction and PCR controls (Shinoda *et al.*, 2006).

First, exact replicas of each tooth were prepared for the other morphological studies. To prevent contamination from post-excavation handling, the tooth samples were rinsed with DNA-decontamination agents and then washed thoroughly with distilled water before drying. Next, the samples were encased in silicone rubber (Provil novo Heraeus Kulzer GmbH, Hanau, Germany). The tip of the root of the tooth sample was removed by a horizontal cut using a cutting disk, and the dentin around cavities and dental pulp were powdered and removed through the root tip by using a dental drill as described by Gilbert *et al.* (2003). The powdered samples were decalcified with 0.5 M EDTA (pH 8.0) (Invitrogen, Carlsbad, CA, U.S.A.) at room temperature overnight, and then, the EDTA buffer was replaced by fresh buffer, and samples were decalcified for a further 48 h. Decalcified samples were lysed in 500 μl of Fast Lyse (Genetic ID, Fairfield, IA, U.S.A.) with 30 μl of 20 mg/ml Proteinase K (Invitrogen) at 60℃ for 4 h. DNA was extracted from the lysate using a FAST ID DNA Extraction Kit (Genetic ID) in accordance with the technical manual.

Amplification and sequencing of HVR1

Segments of hypervariable region 1 (HVR1; nucleotide positions 16121–16238, HVR1-1; 16209–16291, HVR1-2; and 160289–16366; HVR1-3) of mtDNAs, as per the revised Cambridge reference sequence (rCRSs; Andrews *et al.*, 1999), were sequenced in all samples. Primers used to amplify the hypervariable regions were the same as those used by Shinoda *et al.* 2013.

Aliquots (2 μl) of the extracts were used as templates for PCR. Amplifications were carried out in 15-μl reaction mixtures containing 1 unit of Taq DNA polymerase (HotStarTaq™ DNA polymerase; Qiagen, Duesseldorf, Germany), 0.1 M of each primer, and 100 mM of dNTP in 1× PCR buffer provided by the manufacturer. The PCR conditions were as follows: incubation at 95℃ for 15 min; 40 cycles of 94℃ for 20 s; 50–56℃ for 20 s, and 72℃ for 15 s; and final extension at 72℃ for 1 min.

The PCR products were subjected to agarose gel electrophoresis and were recovered by using a QIAEX II agarose gel extraction kit (Qiagen). Aliquots of the samples were prepared for sequencing on a BigDye cycle sequencing kit Ver.3.1 (Applied Biosystems, Foster City, CA, USA) The primers used in the PCR amplification were also used in the sequencing reaction. Bidirectional sequencing was performed to enable identification of polymorphisms or ambiguous bases by using a single primer. The sequencing reactions were performed on a DNA Sequencer (ABI model 3130).

Data analysis and genotyping of mtDNA

The 246-base pair (bp) hypervariable segment 1 (HVS1) sequence was edited and aligned against the rCRSs using the Genetyx-Mac software (version 15). The nucleotide diversity and mean number of pairwise differences between the mitochondrial D-loop sequences were computed using the Arlequin software package (version 3.1; Excoffier *et al.*, 2005) based on Tamura–Nei distances and a gamma parameter value of 0.26 (Mayer *et al.*, 1999).

To confidently assign mtDNAs to the relevant haplogroups, 56 haplogroup-diagnostic single nucleotide polymorphisms (SNPs) that define major haplogroups found in Southeast Asian and East Asian populations were analyzed by PCR-Luminex methods. SNPs that define major hap-

Table 2. Primer sequence of PCR Luminex methods

Foward primer	primer oligo sequence 5′>3′	Reverse primer	primer oligo sequence 5′>3′	PCR position		size	Target marker	Haplogroup
00663-F	TCACCCCATAAACAAATAGGTTTGGTC	00663-R	CTTGCATGTGTAATCTTACTAAGAGCTAATAGAAAG	634	701	68 bp	00663	A
00827-F	ACCCCCACGGGAAACAGCAGTG	00827-R	GGGTTAGTATAGCTTAGTTAAACTTTCGTTTATTGCTA	801	869	69 bp	00827	B4b.d.e
02766-F	TATGGAGCTTTAATTTATTAATGCAAACAGTAC	02772-R	CGAAATTTTTAATGCAGGTTTGGTAGTTTAGGAC	2729	2810	82 bp	02772/02766	M7a/D4d1
03010-F	GACCTCGATGTTGGATCAGGACATC	03010-R	CGAACCTTTAATAGCGGCTGCAC	2983	3036	54 bp	03010	D4
03497-F	CCATAAAACTCTTCACCAAAGAGCC	03497-R	GGTGATGTAGAGGGTGATGGTAGATGTG	3461	3525	65 bp	03497	B4c1
03644-F	CTATTTATTCTAGCCACCTCTAGCCTAGC	03644-R	CTCACCCTGATCAGAGGATTGAG	3613	3670	58 bp	03644	D4h/M12
03970-F	TACGCCGCAGGCCCCTTCGC	03970-R	TGTTTATTATAATAATGTTTGTGTATTCGGCTATGA	3949	4013	65 bp	03970	F
04048-F	TCACCACTACAATCTTCCTAGGAACA	04071-R	TAGAAGTAGGGTCTTGGTGACAAAATATGTTG	4016	4104	89 bp	04071/04048	M7b.c/M7b
04386-F	AGAATCCAAAATTCTCCGTGCCAC	04386-R	CCGATAGCTTATTTAGCTGACCTTACTTTAG	4359	4428	70 bp	04386	M7a
04491-F	CGTACTAATTAATCCCCTGGCCCAAC	04491-R	GAGTGTGCCTGCAAAGATGGTAGAGTAG	4463	4523	61 bp	04491	M9
04715-F	CCTTCTAATAGCTATCCTCTTCAACAATATACTCTC	04715-R	TGAGTATTGATTGGTAGTATTGGTTATGGTTC	4676	4751	76 bp	04715	M8
04833-F	CACTTCTGAGTCCCAGAGGTTACCCAAG	04833-R	GAAGAAGCAGGCCGGATGTCAG	4803	4860	58 bp	04833	G
04883-F	TGCTTCTTCTCACATGACAAAAACTAGC	04883-R	GTTTAGTGAGGGAGAGATTTGGTATATGATTG	4852	4919	68 bp	04883	D
05108-F	CCTAACATAACCATTCTTAATTTAACTATTTATATTATCCTA	05127-R	GATAGTAGTAGGGTCGTGGTGCTGGAGTTTA	5061	5160	100 bp	05108/05127	G/B4f
05178-F	CGACCCTACTACTATCTCGCACCTGA	05178-R	GGTGGATGGAATTAAGGGTGTTAGTC	5146	5210	65 bp	05178	D
05231-F	CATCCACCCTCCTCTCCCTAG	05231-R	CATTTGGGCAAAAAGCCGGTTAG	5203	5260	58 bp	05231	N9a
05417-F	TCCCCATATCTAACAACGTAAAAATAAAATGAC	05417-R	GAATGGGGTGGGTTTTGTATGTTCA	5383	5444	62 bp	05417	N9
06179-F	CTAGTTCCCCTAATAATCGGTGCC	06179-R	AGTCAGAAGCTTATGTTGTTTATGCGG	6147	6214	68 bp	06179	M8a
06752-F	TGGTCTGAGCTATGATATCAATTGGCTTC	06752-R	TCTATTCCTACTGTAAATATATGGTGTGCTCAC	6721	6796	76 bp	06752	Z
08020-F	TCCTTGACGTTGACAATCGAGTAGTACTC	08020-R	ATGTAATTATTATACGAATGGGGGCTTCA	7989	8051	63 bp	08020	D4b
08200-F	GTCAATGCTCTGAAATCTGTGGAGCA	08200-R	TTTTAGGGGAATTAATTCTAGGACGATG	8166	8236	71 bp	08200	G1
08272-F	CTTTGAAATAGGGCCCGTATTTACCCTATAG	08272-R	TGCTAAGTTAGCTTTACAGTGGGCTCTAG	8239	8315	77 bp	08272	9del(B)
08392-F	ACCTCTTTACAGTGAAATGCCCCAACTA	08392-R	GTGTAAGGAGTATGGGGGTAATTATGGTG	8350	8423	74 bp	08392/8383	Y/D4d2
08473-F	ACCCAACTAAAAATATTAAACACAAACTACCACCTAC	08473-R	TATAATTTTTTATTTTTATGGGCTTTGGTGAG	8435	8507	73 bp	08473	D4a
08701-F	CAACAATGACTAATCAAACTAACCTCAAAACAA	08701-R	GAGATCAGGTTCGTCCTTTAGTGTTGTG	8662	8734	73 bp	08701	M
08794-F	TCATTTTTATTGCCACAACTAACCTCCTCG	08794-R	AGTTGGGTGGTTGGTGTAAATGAG	8753	8818	66 bp	08794/08793	A/M10
09123-F	CATTAACCTTCCCTCTACACTTATCATCTTCAC	09123-R	GGATTAAGGCGACAGCGATTTCTAG	9075	9154	80 bp	09123	B4a
09180-F	GCCTTAATCCAAGCCTACGTTTTCAC	09180-R	TCATTATGTGTTGTCGTGCAGGTAG	9145	9210	66 bp	09180	D5a.b
09950-F	CCGCCTGATACTGGCATTTTGTAG	09950-R	GAGTAAGACCCTCATCAATAGATGGAGAC	9919	9994	76 bp	09950	B5
10310-F	CCTACCATGAGCCCTACAAACAACTAAC	10310-R	TGATGATTAATAAGAGGGATGACATAACTATTAGTG	10280	10342	63 bp	10310	F
10400-F	TGGCCTATGAGTGACTACAAAAAGGATTAG	10400-R	TGAGTCGAAATCATTCGTTTTGTTTAAAC	10363	10442	80 bp	10400/10398	N/M
10873-F	ACCACCCACAGCCTAATTATTAGCATC	10873-R	GCTAAATAGGTTGTTGTTGATTTGGTTAAAAAATAG	10841	10912	72 bp	10873	N
11215-F	GAACGCCTGAACGCAGGCAC	11215-R	TGAGTAGGGGAAGGGAGCCTACTAG	11180	11243	64 bp	11215	D4e
12091-F	CACGAGAAAACACCCTCATGTTCATAC	12091-R	CGGTAATGATGTCGGGGTTGAG	12051	12125	75 bp	12091/12085-92	M7c/D4c.m,..
12705-F	CCCAAACATTAATCAGTTCTTCAAATATCTACTC	12705-R	GGTTGTTAGCGGTAACTAAGATTAGTATGGTAATTAG	12669	12745	77 bp	12705	A.N9.M
12811-F	CGTAGGAATTATATCCTTCTTGCTCATC	12811-R	TGCTGTGTTGGCATCTGCTCG	12774	12837	64 bp	12811	M7b
13183-F	CCACTAATCCAAACTCTAACACTATGCTTAG	13183-R	GCGCAGACTGCTGCGAACAG	13147	13211	65 bp	13183	N9b
13263-F	AAAAAATCGTAGCCTTCTCCACTTCA	13263-R	GGTTGATGCCGATTGTAACTATTATGAGTC	13232	13297	66 bp	13263	C
13563-F	AACATATCATACACAAACGCCTGAGC	13563-R	TCAGGGAGGTAGCGATGAGAGTAATAG	13534	13591	58 bp	13563	G2
15346-F	GCCCTAGCAGCACTCCACCTCCTA	15346-R	GTGATTCCTAGGGGGTTGTTTGATC	15317	15381	65 bp	15346	B4c
15487-F	TCTCTCCTTAATGACATTAACACTATTCTCAC	15497-R	GCTAGGGTATAATTGTCTGGGTCGC	15454	15522	69 bp	15487/15497	M8/G1
15535-F	GACAATTATACCCTAGCCAACCCCTTAA	15535-R	TAGGAAATATCATTCGGGCTTGATGTG	15506	15571	66 bp	15535	B4b.d.e

logroups were detected by using suspension-array technology (Luminex 100; Luminex, Austin, Texas, U.S.A.) at the laboratory of G&G Science, Fukushima. The methodology used for genotyping has been described in detail elsewhere (Itoh *et al.*, 2005, Shinoda *et al.*, 2012). Forty-one portions were amplified; Table 2 lists the primer sequences of this analysis.

In some cases, mtDNA SNPs were detected using the amplified product length polymorphism (APLP) method (Umetsu *et al.*, 2001, 2005). This method has been applied to ancient DNA analyses and has yielded convincing results (Adachi *et al.*, 2004; Shinoda *et al.*, 2006). In this study, 26 SNPs in the coding region and a 9-bp pair repeat variation in the non-coding cytochrome oxidase II/tRNALys intergenic region were analyzed by the multiplex APLP method using the primer sets described by Adachi *et al.* (2011). The polymorphic sites examined in this study have been proved to cover most of the haplogroup-defining mutations in East Asian mtDNAs. The constitution of the PCR reaction mixture, thermal conditions, and the method for separating and detecting PCR products were the same as described by Adachi *et al.* (2009).

Results and Discussion

Table 3 shows the results of PCR amplification of HVR1 and genotyping of coding regions by PCR-Luminex and APLP methods. Ancient DNA data were successfully extracted from all 35 samples included in this study. We were able to successfully retrieve different variants of the HVR1 (16121–16366) sequence. However, some parts of HVR1 from 11 samples failed to yield a product on amplification or contained several ambiguous nucleotide positions during the sequencing reaction, possibly due to misincorporations during PCR.

To analyze mtDNA variations at a finer level, 246-bp sequences of HVS1 from 24 samples (Ogulan, 18; Xilou, 6) were used to gain information on the internal population diversity (Table 4). Twenty-two distinct types of sequences were defined by 44 polymorphic sites in total. In general, populations of the same ethnic origin shared a relatively large number of sequences with one another. However, mainly due to the small sample size of the Xilou site, we could not find shared haplotypes between the Ogulan and Xilou sites.

Nucleotide diversity among the individuals was estimated to be 0.028, which is higher than the corresponding value estimated for the Taiwanese, East Asian, and Oceanian populations (ranging from 0.014 to 0.019). Haplotype diversity (0.98) is also higher than the corresponding value estimated for the aboriginal Taiwanese (ranging from 0.83 to 0.91) (Trejaut *et al.*, 2005). The Pingpu population has intermarried and integrated into the Minnan and Hakka ethnic groups, which belonged to the Han Chinese population since the 17th century. It seems reasonable to suppose that the genetic characteristics of the Pingpu population has been affected by these people and has therefore come to have considerable genetic diversity.

Phylogenetic analysis based on coding region information clustered the observed haplotypes into 14 distinct major haplogroups and subhaplogroups. These results indicate that all mtDNA lineages of Pingpu individuals were assigned to the previously defined haplogroups in Southeast Asians and East Asians.

Figure 2 shows the haplogroup frequencies of each site and modern Han Taiwanese. As shown the figure, haplogroup frequencies of aboriginal Taiwanese and Han Taiwanese was quite different. Although haplogroup composition among these Pingpu groups was slightly different, haplogroup F was dominant in both groups. Haplogroup F1a, which is common and widespread in Southeast Asia (Kivisild *et al.* 2002), is found largely in both groups, especially in Xilou. Haplogroup B, particularly haplogroup B4, has been observed at high frequencies in the islands of Southeast Asia. Thus, the frequency of this haplogroup reflects the genetic affinity to the islands in Southeast Asia. However, haplogroup B is only present at low frequencies in the Ogulan site

Table 3. Result of the analysis

No.	Code Name	HV1-1 (16121–16238) (16,000+)	HV1-2 (16209–16291) (16,000+)	HV1-3 (16289–16366) (16,000+)	APLP	SNP PCR-Luminex	Haplogroup
Xilou site							
1	23	129, 172, 190, 218	N.D.	N.D.	F1a		F1a
2	40	N.D.	223	304		N/F,B	F*
3	117	223	223, 290	319		A	A
4	118	N.D.	278	304		F*	F*
5	119	*CRS*	N.D.	304		N.D.	F*
6	529	129, 172	*CRS*	304	F1		F1a1
7	533	129, 172	*CRS*	304	F1a1	N	F1a1
8	536	223	N.D.	N.D.	D4		D4
9	539	185, 223	223	294, 362		M9	M9
10	541	185, 223	223	294, 362		M9	M9
11	543	186, 192	N.D.	295, 319		M7c	M7c
12	546	145, 172, 223	223, 245, 256, 261	362		N9a	N9a
13	548	N.D.	N.D.	N.D.		M7b2	M7b2
Ogulan site							
1	53	182, 183, 189	*CRS*	311, 355, 358, 366	R		N.D.
2	113	185, 223	223, 260	298	M8		M8
3	561	170, 207	207	N.D.	F4		F4a
4	562	129, 162, 172	*CRS*	304	F1a		F1a1a
5	565	223	223	362	E		E
6	566	126	231	*CRS*		N.D.	Y2
7	568	148, 182, 183, 189, 223	223	362	D*		D5
8	569	N.D.	N.D.	294, 304		N.D.	F1c
9	570	189	294	304		F*	F1c
10	572	172, 210, 220	210, 220, 265	298, 311, 362	F3		F3b1a
11	605	N.D.	261	*CRS*		B4a	B4a
12	608	*CRS*	*CRS*	*CRS*	M7c		M7c
13	610	223	223	362		D4e	D4e
14	613	142, 188, 223	223	295, 362		M7c	M7c
15	615	223	223	295, 328, 362		N.D.	M7c
16	617	155, 182, 183, 189	266A	*CRS*	B*		B5
17	618	150, 168, 223	223	295, 362		M7c	M7c
18	621	223	223, 227	*CRS*	M7c		M7c
19	622	223	223, 227, 272	319, 362, 366	G/M12		G2a4
20	624	129, 223	223, 274	311, 317, 362		D	D4
21	626	129, 162, 172	259	304	F*		F1a1a
22	635	182, 183, 189, 217	217, 261	356	B4a	N*	B4a

All polymorphic sites are numbered according to the revised Cambridge reference sequence (Andrews *et al.*, 1999).
CRS indicates that the sequence of the segment is identical to the revised Cambridge reference sequence.
N.D. denotes "not determined".
The asterisk denotes that the haplogroup status cannot be identified further.

(13.6%) and absent in the Xilou site. Haplogroups E and Y2 are common indigenous haplogroups in the islands of Southeast Asia. They are also common in Taiwanese aboriginals of mountainous areas but are almost absent in China (Hill *et al.*, 2007). Only two samples examined in the present study belongs to these haplogroup. This suggests only some genetic connection between the populations of the island areas and the Pingpu population. In the present study, the results of the mtDNA analyses demonstrate that the populations of the islands in Southeast Asia have a different maternal ancestry from most of the Pingpu population.

Although haplogroups N9a and M9 are widespread in mainland Southeast Asia and may have a southern origin, they are absent in the Ogulan site and three samples from the Xilou site. Haplogroups M7b and M7c are also widespread in Southeast Asia. Although M7c was found at both sites, M7b was only found in the Xilou site. Haplogroups A, D, G, and M8 had high frequencies

Table 4. Divesity Indices of the the Pingpu populations calculated from HVS1 region

Sites	Size	Number of haplotype	Number of loci	Number of polymorphic sites	Number of unique haplotype	Haplotype diversity	Nucleotide diversoty	Gene diversity	Mean number of pairwise differences
Ogulan	18	18	246	40	18 (100%)	1	0.030+/−0.016	1.00+/−0.019	7.314+/−3.591
Xilou	6	4	246	13	4 (66.7%)	0.867	0.024+/−0.015	1.00+/−0.096	5.867+/−3.265
Total	24	22	246	44	22 (91.7%)	0.989	0.028+/−0.015	1.00+/−0.012	6.996+/−3.401

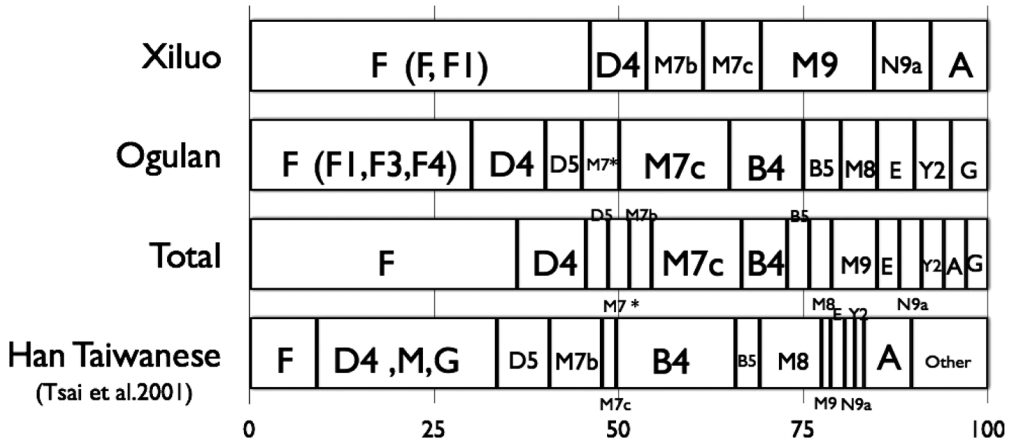

Fig. 2. Estimated frequencies of mtDNA haplogroups in Pingpu populations and Han Taiwanese.

in north China (Yao *et al.* 2002a; Kong *et al.* 2003) but had low frequencies in the Pingpu samples. To clarify the genetic characteristics of the regional population, the distribution of haplogroup frequencies of the Pingpu population was compared with information available from other Asian populations in nearby regions (Table 5).

Principal components analysis based on haplogroup frequencies was performed. This statistical technique aims to reveal the internal structure of multivariate data in a way that best explains variance; it involves the summarization of correlating parameters to explain overall changes on the basis of a small number of variables.

The results of the data calculations using this technique are graphically represented in Figures 3 and 4. Of the overall information, about 28.6% is integrated along the axis of the first principal component (PC1), about 22.1% along the axis of the second principal component (PC2), and 15% along the axis of the third principal component (PC3). Thus, approximately 66% of all information is two-dimensionally accounted for in Figures 3 and 4. Judging from the value of the eigenvector (Table 6), axis 1 indicates the percentage of haplogroups A and D/M, which represent the East Asian haplogroups, relative to haplogroups B4, M7c, and E, which represent the Southeast Asian haplogroup. As the value of axis 1 increases (the groups located on the right), the percentage of haplogroups A and D/M increases. Thus, the arrangement of groups along axis 1 reflects the north-to-south geographic relationship among groups in East to Southeast Asia. The Pingpu people are positioned in the middle of axis 1, similar to the position of the present-day Hainan islanders and the North Vietnamese, Yunnan, and Guangdong people, who live somewhat farther north of Southeast Asia.

Axis 2 indicates the percentage of haplogroups Y2, M7b, E, and B5 relative to haplogroups M7c. As the value of axis 2 increases (groups located in the upper part), the percentage of haplogroups Y2, M7b, E, and B5 becomes higher than that of haplogroup M7c. This axis segre-

Table 5. The Haplogroup Distribution Frequencies (%) in the 18 Populations Considered in the Present Study

	Aboriginal Taiwanese									
Haplogroup	Pingpu	Atayal	Saisiat	Tsou	Bunun	Paiwan	Rukai	Puyuma	Amis	Tao
A	3.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
B4	6.1	10.9	4.8	18.3	34.8	23.7	14.0	21.2	49.0	50.0
B5	3.0	0.9	12.7	13.3	6.7	9.1	10.0	0.0	0.0	0.0
F	36.3	28.4	14.3	21.6	30.3	38.2	42.0	40.4	6.1	18.8
N9a	3.0	1.8	0.0	0.0	1.1	1.8	0.0	1.9	7.1	0.0
Y2	3.0	2.8	9.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0
D/M	12.2	6.4	1.6	1.7	0.0	16.3	20.0	1.9	8.2	0.0
M7b	3.0	39.4	23.8	0.0	3.4	0.0	8.0	0.0	6.1	0.0
M7c	15.2	0.0	0.0	1.7	0.0	7.3	6.0	28.9	10.2	26.6
M9	6.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
E	3.0	6.4	31.7	13.3	23.6	3.6	0.0	5.8	13.3	4.7
Other	6.0	2.9	1.3	30.0	0.0	0.0	0.0	0.0	0.0	0.0
n	33	109	63	60	89	55	50	52	98	64
References	Present study				Trejaut *et al.* 2005					

Haplogroup	Han Taiwanese	Philippines	North Vietnam	Guangdong	Hong Kong	Hainan Island	Yunnan	Norh China
A	6.5	0.0	0.9	2.6	4.0	0.7	0.6	9.0
B4	12.9	35.6	13.2	15.6	10.1	15.2	12.0	9.0
B5	3.2	5.1	12.3	6.1	5.6	9.0	5.7	4.2
F	9.0	6.8	21.5	23.5	19.4	18.7	20.9	11.7
N9a	1.3	3.4	3.4	3.8	2.1	1.8	0.0	3.7
Y2	1.3	3.4	0.0	0.4	0.5	0.0	0.0	1.6
D/M	31.6	5.1	19.0	26.0	30.0	26.6	32.9	30.2
M7b	7.1	6.8	12.6	8.6	12.5	10.3	15.2	4.2
M7c	1.9	6.8	4.0	4.4	2.9	4.7	0.0	2.7
M9	1.3	1.7	1.5	0.7	1.6	0.9	1.3	3.2
E	1.9	20.3	0.0	0.2	0.0	0.0	0.0	0.0
Other	21.9	5.1	11.7	8.2	11.4	12.2	11.4	20.6
n	155	59	326	546	337	447	158	189
References	Tsai *et al.* 2001	Tajima *et al.* 2004	Peng *et al.* 2011		Irwin *et al.* 2009	Peng *et al.* 2011		Yao *et al.* 2002

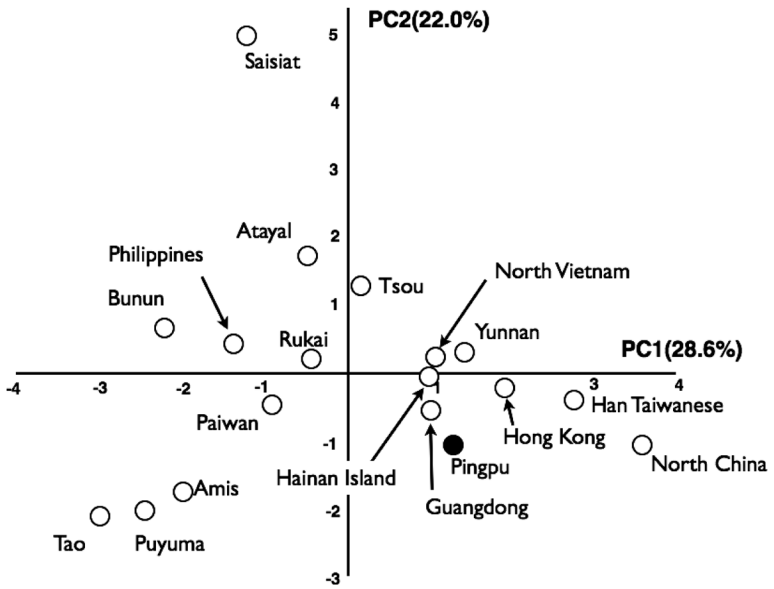

Fig. 3. Principal components (PC) analysis of populations in Taiwan, southern China, and Southeast Asia. Plot of PC1 and PC2 scores.

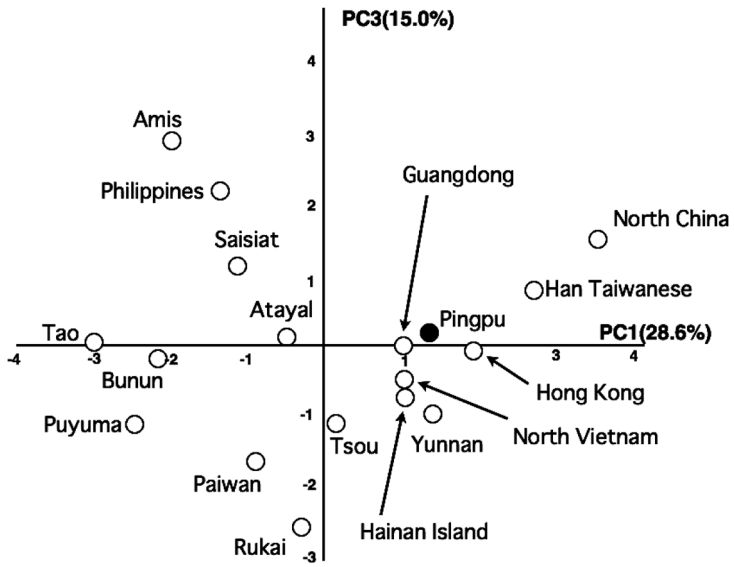

Fig. 4. Plot of PC1 and PC3 scores of Southeast Asian and East Asian populations.

Table 6. Result of principal component analysis of eighteen Asian populations

	PC1	PC2	PC3
Eigenvalue	3.4333	2.649	1.8023
Contribution	28.6111	22.0749	15.019
Cumurlative	28.6111	50.686	65.705
Eigenvector			
A	0.43351	−0.15024	0.21685
B4	−0.38361	−0.26943	0.23839
B5	0.10349	0.38467	−0.29278
F	−0.1192	−0.07006	−0.60074
N9a	0.04461	−0.2617	0.47357
Y2	−0.02305	0.44986	0.29817
D/M	0.4649	−0.12354	−0.08813
M7b	0.06068	0.38893	0.08568
M7c	−0.28353	−0.39571	−0.0535
M9	0.30835	−0.14135	0.15618
E	−0.30132	0.3701	0.30614
Other	0.38907	0.03186	0.01675

gated the aboriginal Taiwanese populations from one another and reflects the north-to-south geographic relationship among aboriginal groups in Taiwan. Judging from the value of the eigenvector, axis 3 indicates the percentage of haplogroup F and B5 relative to haplogroups N9a and Y2. Similar to PC2, this component also separates the Pingpu from the aboriginal Taiwanese mountain population.

PC analysis reveals that the mtDNA haplogroup profiles of the Pingpu population are similar to the profiles of the populations from mainland southern China and North Vietnam and distinct from aboriginal and Han Taiwanese. This is in agreement with the hypothesis that the Pingpu population has descended from immigrants from southern China, particularly Guangxi, and has a common origin with the populations from this region. It is said that the influence of the Pingpu by the Han Chinese immigrants began in the 17th century. However, our analysis indicates that their maternal genetic components continue to retain their original form.

In general, the mtDNA phylogeny in the Pingpu population was represented as a subset in the context of East Asia and Southeast Asia. The ancestors of the Pingpu people might have been from the populations in mainland southern China in the Neolithic period. This agrees well with the archaeological evidence. The recent gene flow from the Han Chinese living in Taiwan might play important roles in shaping the gene pool of the Pingpu population. Although they lost their original language, maternal lineages of Pingpu population could be traced back to southern China during the Neolithic period. However, it

should be kept in mind that mtDNA data yields no information on the genetic connections along the paternal line. In the absence of additional independent lines of evidence, it is not possible to specify the biological relationships among excavated individuals beyond those described in this paper.

Acknowledgement

We thank Dr. Lu KS of the Department of Anatomy, National Taiwan University, for helpful advice and for permitting us to use the skeletal materials. This research was supported in part by a Grant-in-Aid for Scientific Research (No. 20370096) from the Japan Society for the Promotion of Science.

References

Adachi N., Shinoda K., Umetsu K., and Matsumura H., (2009) Mitochondrial DNA analysis of Jomon skeletons from the Funadomari site, Hokkaido, and its implication for the origin of Native Americans. *American Journal of Physical Anthropology*, 138: 255–265.

Adachi N., Umetsu K., Takigawa W., and Sakaue K., (2004) Phylogebetic analysis of the human ancient mitochondrial DNA. *Journal of Archaeological Science*, 31: 1339–1348.

Adachi N., Shinoda K., Umetsu K., Kitano T., Matsumura H., Fujiyama R., Sawada J., and Tanaka M., (2011) Mitochondrial DNA analysis of Hokkaido Jomon skeletons: Remnants of archaic maternal lineages at the southwestern edge of former Beringia. *American Journal of Physical Anthropology*, 146: 346–360.

Andrews R. M., Kubacka I., Chinnery P. F., Lightowlers R. N., Turnbull D. M., and Howell N., (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nature Genetics*, 23:147.

Bellwood P., (2001) Early agriculturalist population diasporas? Farming, languages, and genes. *Annual Review of Anthropology*, 30: 181–207.

Excoffier L., Laval G., and Schneider S., (2005) Arlequin (version 3.1): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, 1: 47–50.

Chu C. C., Lin M., Nakajima F., Lee H. L., Chang S. L., Juji T., and Tokunaga K., (2001) Diversity of HLA among Taiwan's indigenous tribes and the Ivatans in the Philippines. *Tissue Antigens*, 58: 9–18.

Gilbert M. T. P., Willerslev E., Hansen A. J., Barnes I., Rudbeck L., Lynnerup N., and Cooper A., (2003) Distribution patterns of postmortem damage in human mitochondrial DNA. *American Journal of Human Genetics*, 72: 32–47.

Gray R. D., and Jordan F. M., (2000) Language trees support the express train sequence of Austronesian expansion. *Nature*, 405: 1052–5.

Hill C., Soares P., Mormina M., Macaulay V., Clarke D., Blumbach P. B., Vizuete-Forster M., Forster P., Bulbeck D., Oppenheimer S., and Richards M., (2007) A mitochondrial stratigraphy for island Southeast Asia. *American Journal of Human Genetics*, 80: 29–43.

Hsu Y-C., (1947) Skeletal anthropometric study of Ogulan pingpu tribe. *Bulletins of Anatomical Department of the National Taiwan University*, *Formosa. Fascicle* 1: 121–198. (In Japanese).

Irwin J. A., Saunier J. L., Beh P., Strouss K. M., Paintner C. D., and Parsons T. J., (2009) Mitochondrial DNA control region variation in a population sample from Hong Kong, China. *Forensic Science International Genetics*, 3: e119–125.

Itoh Y., Mizuki N., Shimada T., Azuma F., Itakura M., Kashiwase K., Kikkawa E., Kulski J. K., Satake M., and Inoko H., (2005) High throughput DNA typing of HLA-A, -B, -C, and -DRB1 loci by a PCR-SSOP-Luminex method in the Japanese population. *Immunogenetics*, 57: 717–729.

Kayser M., Brauer S., Weiss G., Underhill P. A., Roewer L, Schiefenh?vel W., and Stoneking M., (2000) Melanesian origin of Polynesian Y chromosomes. *Current Biology*, 10: 1237–1246.

Kivisild T., Tolk H-V., Parik J., Wang Y., Papiha S. S., Bandelt H. J., and Villems R., (2002) The emerging limbs and twigs of the East Asian mtDNA tree. *Molecular Biology and Evolution*, 19: 1737–1751.

Ko A.M-S., Chen C-Y., Fu Q., Delfin F., Li M., Chui H-L., Stoneking M., and Ko Y-C., (2014) Early Austronesians: Into and out of Taiwan. *American Journal of Human Genetics*, 94: 426–436.

Kong Q. P., Yao Y. G., Sun C., Bandelt H. J., Zhu C. L. and Zhang P., (2003) Phylogeny of East Asian mitochondrial DNA lineages inferred from complete sequences. *American Journal of Human Genetics*, 73: 674–676.

Lin M., Chu C. C., Lee H. L., Chang S. L., Ohashi J., Tokunaga K., Akaza T., and Juji T., (2000) Heterogeneity of Taiwan's indigenous population: possible relation to prehistoric Mongoloid dispersals. *Tissue Antigens*, 55: 1–9.

Lin L., Chu C. C., Broadberry R. E., Yu L-C., Joo J-H., and Trejaout A., (2005) Genetic diversity of Taiwan's indigenous peoples. In The peopling of East Asia. (ed) Sagart P., Blench R., and Sanchez-Mazas A., Routledge Curzon, 230–347.

Mayer S., Weiss G., and von Haeseler A., (1999) Pattern of nucleotide substitution and rate of heterogeneity in the hypervariable regions I and II of human mtDNA. *Genetics*, 152: 1103–1110.

Peng M-S., He J-D, Liu H-X, and Zhang Y-P., (2011) Tracing the legacy of the early Hainan Islanders -a perspective from mitochondrial DNA. *BMC Evolutionary Biology*, 11: 46.

Shinoda, K., Adachi N., Guillen S., and Shimada I., (2006) Mitochondrial DNA analysis of ancient Peruvian highlanders. *American Journal of Physical Anthropology*, 131: 98–107.

Shinoda K., (2008) MtDNA analysis of Bunun remains stored in the National Taiwan University. *Anthropological Science* (Japanese Series), 116(2): 154–160. Japanese with English summary.

Shinoda K., Kakuda, T., and Doi, N., (2012) Mitochondrial DNA polymorphisms in late Shell midden period skeletal remains excavated from two archaeological sites in Okinawa. *Bulletin of the National Museum of Nature and Science Series D*., 38: 51–61.

Shinoda K., Kakuda, T., and Doi, N., (2013) Ancient DNA analysis of skeletal remains from the Gusuku period excavated from two archaeological sites in the Ryukyu Islands, Japan. *Bulletin of National Museum of Nature and Science*, *Series D*, 39: 1–8.

Su B., Xiao J., Underhill P., Deka R., Zhang W., Akey J., Huang W., Shen D., Lu D., Luo J., Chu J., Tan J., Shen P., Davis R., Cavalli-Sforza L., Chakraborty R., Xiong M., Du R., Oefner P., Chen Z., and Jin L., (1999) Y-Chromosome evidence for a northward migration of modern humans into Eastern Asia during the last Ice Age. *American Journal of Human Genetics*, 65: 1718–1724.

Tajima A., Sun C. S., Pan I. H., Ishida T., Saitou N., and Horai S., (2003) Mitochondrial DNA polymorphisms in nine aboriginal groups of Taiwan: Implications for the population history of aboriginal Taiwanese. *Human Genetics*, 113: 24–33.

Tsai L. C., Lin C. Y., Lee J. C., Chang J. G., Linacre A., and Goodwin W., (2001) Sequence polymorphism of mitochondrial D-loop DNA in the Taiwanese Han population. *Forensic Science International*, 119: 239–247.

Tajima A., Hayami M., Tokunaga K., Juji T., Matsuo M., Marzuki S., Omoto K., and Horai S., (2004) Genetic origins of the Ainu inferred from combined DNA analyses of maternal and paternal lineages. *Journal of Human Genetics*, 49: 187–193.

Trejaut J. A., Kivisild T., Loo J. H., Lee C. L., He C. L., Hsu C. J., Li Z. Y., and Lin M., (2005) Traces of archaic mitochondrial lineages persist in Austronesian-speaking Formosan populations. *PLoS Biology*, 3: 1362–1372.

Umetsu K., Tanaka M., Yuasa I., Saitou N., Takeyasu T., Fuku N., Naito E., Ago K., Nakayashiki N., Miyoshi A., Kashimura S., Watanabe G., and Osawa M., (2001) Multiplex amplified product-length polymorphism analysis for rapid detection of human mitochondrial DNA variations. *Electrophoresis*, 22: 3533–3538.

Umetsu K., Tanaka M., Yuasa I., Adachi N., Miyoshi A., Kashimura S., Park K. S., Wei Y. H., Watanabe G., and Osawa M., (2005) Multiplex amplified product-length polymorphism analysis of 36 mitochondrial single-nucleotide polymorphisms for haplogrouping of East Asian populations. *Electrophoresis*, 26: 91–98.

Wang C-M., (1947) Somatometric study of Ogulan Pingpu tribe. *Bulletins of Anatomical Department of the National Taiwan University*, *Formosa. Fascicle* 7: 1–90. (In Japanese)

Wen B., Li H., Gao S., Mao X., Gao Y., Li F., Zhang F., He Y., Dong Y., Zhang Y., Huang W., Jin J., Xiao C., Lu D., Chakraborty R., Su B., Deka R., and Jin L., (2005) Genetic structure of Hmong-Mien speaking populations in East Asia as revealed by mtDNA lineages. *Molecular Biology and Evolution*, 22: 725–734.

Woodward S. R., King M. J., Chiu N. M., Kuchar L. M., and Griggs C. W., (1994) Amplification of ancient nuclear DNA from teeth and soft tissues. *PCR Methods and Applications*, 3: 244–247.

Yao Y. G., Nie L., Harpending H., Fu Y. X., Yuan Z. G., and Zhang Y. P., (2002a) Genetic relationship of Chinese ethnic populations revealed by mtDNA sequence diversity. *American Journal of Physical Anthropology*, 118: 63–76.

Yao Y. G., Kong Q. P., Bandelt H. J., Kivisild T., and Zhang Y. P., (2002b) Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *American Journal of Human Genetics*, 70: 635–651.

Yuan T-M., (1960) Anthropometric study of skeletal remains excavated from Xilou site of Pingpu tribe. *Quarterly Journal of Anthropology*, Vol.VII, No.3–4: 9–51. (In Japanese)