

Population study of polymorphic microsatellite DNA in Taiwan

James Chun-I Lee,^{1,*} Ph.D.; Marie Lin,² M.D.; Li-Chin Tsai,¹ M.S.; Chien-Ming Hsu,¹ Ph.D.;
Hsing-Mei Hsieh,¹ Ph.D.; Nu-En Huang,³ M.A.; Rocky Tai-Ping Shih,³ M.D.;
Jin-Hui Wun,³ M.S.; Jan-Gowth Chang,⁴ M.D.; Ying-Chin Ko,⁵ M.D. Ph.D.;
Cheng-Hwai Tzeng,⁶ M.D.; Adrian Linacre,⁷ Ph.D.

¹ Department of Forensic Science, Central Police University, 56 Shu Jen Road, Taoyuan 333, Taiwan ROC.

² Mackay Memorial Hospital Taipei, Taiwan ROC.

³ Criminal Investigation Bureau, National Police Administration, Taiwan, ROC.

⁴ Department of Molecular Medicine, China Medical College Hospital, Taiwan ROC.

⁵ Department of Public Health, Kaohsiung Medical University, Kaohsiung, Taiwan ROC.

⁶ Department of Medicine, Taipei Veterans Hospital, Taiwan ROC.

⁷ Forensic Science Unit, Department of Pure & Applied Chemistry,
University of Strathclyde, Glasgow G1 1XW, UK.

Received: October 8, 2001/ Received in revised form: December 10, 2001 / Accepted: December 20, 2001

ABSTRACT

This study examined the population frequency and genetic diversity of 13 autosomal STR loci in 12 populations of Taiwanese. The populations included Bunun, Paiwan, Saisiat, Tao, Ami, Rukai, Atayal, Tsou, Pazeh, Puyumar, Hakka and Han. The STR loci were the CODIS including D3S1358, vWA, FGA, D8S1179, D21S11, D18S51, D5S818, D13S317, D7S820, D16S539, TH01, TPOX and CSF1PO. In this study no deviations from the Hardy-Weinberg equilibrium were observed for the 13 STR loci by the exact test. The range of matching probability for the 13 loci is from 1.44×10^{-13} in Tsou to 8.76×10^{-15} in Taiwanese Han population. Only one mutation event in D8S1179 locus, altering allele 16 to 15, was detected in 250 meiosis alleles in the mutation screening. The Neighbor-joining tree of cluster analysis was constructed with the genetic distance of 12 populations on 13 loci, which may be useful for phylogenetic study of populations in Taiwan. These data supported the Hanization hypothesis for populations in Taiwan.

Keywords: Short tandem repeat (STR), Population study, Taiwanese Han Population, Gene frequency

Introduction

The peoples of modern day Taiwan have proved to be an important subject for forensic anthropology and evolutionary studies. Much evidence from DNA studies has indicated that Taiwan was important in the colonization of SE Asia and Polynesia by modern man [1-5]. Previous genetic studies have centered on the mitochondrial DNA sequences [1-4] and Y-Chromosome markers [5]. The analysis of autosomal short tandem repeats for the modern Taiwan population has been examined [6] although the aboriginal population has yet to be studied.

Around 18,000 years ago when entering the end of glacial epoch, Taiwan was gradually separated from

mainland China by rising sea level. Finally about 10,000 years ago, Taiwan became an island in its present tobacco leaf shape. Archaeological and anthropological evidence indicates that the indigenous people inhabited Taiwan approximately 10,000 years ago, around the time of its isolation [7]. There were two distinct groups of these original settlers in Taiwan. One group lived on the rich plains of the northeast, west and southwest, called plain people (or "Pepo" people in Chinese). There were 11 tribes in the plain people. Since they lived on the plains, they almost certainly mixed with the later immigrating Chinese populations, the first of which was the Han population. The Han population was considered to be the first population to cross the China Sea to Taiwan. This population remained on the plains rather than inhab-

* Corresponding author, e-mail:jimlee@sun4.cpu.edu.tw

iting the inland mountainous regions. Although considered as only one population, the Han represented a large number of individuals compared to the indigenous population. It was during this period that most tribes have lost their language and culture, producing "Hanization" of the plain people. Only a few original languages can be identified [8]. The indigenous population that lived in the mountains, that run from North to South on the eastern seaboard, were called mountain people (or "Gau-san" people in Chinese). The Gausan population appears to have been separated from the Hanization that occurred on the plain region of Taiwan. The original Gausan people became divided into nine distinct linguistic groups (Bunun, Paiwan, Saisiat, Ami, Rukai, Atayal, Pazeh and Puyumar). The only other population in the Taiwanese region is the Tao (previously named Yami) tribe who lived in Orchid island, which is a small island off the southeastern coast of Taiwan. Tao tribe is believed to have migrated from southern Asian just before 500 years ago. These eight groups from the mountains and the Tao group have distinct languages that are mutually unintelligible, and have different cultures and social organizations.

Currently in Taiwan, the indigenous people constitute less than 2% of the total population [7]. The remaining population is made up of individuals who migrated from mainland China, the earliest waves of immigration, from approximately 400 years ago, were from the Fukien and Kwangtung provinces. These Chinese populations were from the Han area of China and increased the Hanization of the plain people. The descents of these people now constitute approximately 86% of the population [9]. In 1949, there was a large-scale migration from all 35 mainland provincial areas of China due to the Chinese civil war. The emigrants from this migration now constitute approximately 13% of the population. There is a large degree of admixture between these two non-indigenous groups, representing the second large-scale mixture of people in Taiwan. These people are now named Taiwanese Han population. In fact, Taiwan Han population includes indigenous plain people and two waves of Chinese immigrants. By a linguistic and culture study, there is Hakka subpopulation in the Taiwan Han population, which came from part of Kwangtung province. Hakka people constitute approximately 20% of the total Han population in Taiwan.

All the mountain aboriginal tribes have their own languages, custom, culture and social organizations, indicating that they were never part of the Hanization that occurred on the plains. These huge diversities presented in a small island, approximately 36,000 square kilometers, caused interesting discussion of the origin of those aborigines for decades. According to linguistic, culture and genetic studies, there are several theories about their origins. These include: (1) the descendents of proto-Malays who migrated to Taiwan from the Malay penin-

sula and Indonesian archipelago, and Polynesian expansion [3,4,10]; (2) descendents of Mongolia [11]; (3) descendents of Miao tribes in Kweichow province of mainland China [10]; and (4) the homeland of the Austronesians [12].

Polymorphic microsatellite DNA or short tandem repeat (STR) system has been widely used in forensic applications [13-17]. A large number of populations in the world have been screened using STR system [18-20]. These data will be another source to explore relationships among different races or populations in the world. This study analyzed 13 STR loci on Taiwanese Han population, Hakka, Pazeh (a tribe of plain people), and 9 tribes of mountain people.

Materials and methods

DNA samples were collected with informed consent from Mackay Memorial Hospital, Kaohsiung Medical University and the Central Police University, Taiwan. The samples came from unrelated healthy individuals and had been used previously in a phylogenetic study. Samples were collected from 12 populations including 101 Bununs, 99 Paiwans, 97 Saisiats, 117 Taos, 115 Amis, 65 Rukais, 83 Atayals, 92 Tsous, 61 Pazehs, 80 Puyumars, 100 Hakkas and 297 Hans. Additionally 106 family samples containing 250 parent/child pairs were collected from Taiwanese Han population. STR loci were amplified using the Profiler Plus™ and Cofiler™ kits (Applied Biosystems, CA, USA) to detect 13 STR loci (D3S1358, vWA, FGA, D8S1179, D21S11, D18S51, D5S818, D13S317, D7S820, D16S539, TH01, TPOX and CSF1PO). PCR amplification and allele detection were performed according to the manufacturer's instructions. Allele frequencies were calculated by computer program developed by the authors. Nei genetic distance [21] and Neighbor-joining tree were constructed with PHYLIP program [22].

Results and discussion

Samples from 1,307 unrelated individuals of 12 populations in Taiwan were screened using the 13 STR loci. Distributions of allele frequencies are given in Table 1. The Exact Test was used to detect deviation from Hardy-Weinberg equilibrium. In this study no deviations were observed for the 13 STR loci (with $p < 0.05$). Table 2 shows the p value of the exact test.

The range of matching probability for the 13 loci is from 1.44×10^{-13} in the Tsou population to 8.76×10^{-15} in Taiwanese Han population. Only one mutational event was detected in the 250 meiotic alleles that were screened. A mutation was detected in D8S1179 locus, which resulted in the loss of one repeat unit from allele 16 to 15.

Table 1 Gene frequencies (%) of 13 STR loci in 12 Taiwanese populations.

Loci & alleles	Populations											
	Bunun	Paiwan	Saisiat	Tao	Ami	Rukai	Atayal	Tsou	Pazeh	Puyumar	Hakka	Han
D3S1358												
12	0	0	0	0	0	0	0	0	0	0	0	0.2
13	0	0	0.5	0	0.4	0	0	0	0	0	0	0
14	0	0.5	8.2	2.1	1.3	2.2	1.8	0.5	5.7	1.3	4.5	4.4
15	23.8	34	25	23.9	34.8	34.6	39.8	30.3	33.6	35.6	35.5	38.4
16	28.7	25.5	39.3	26.9	30.8	41.9	31.9	36.8	29.5	28.1	31	30.5
17	45	30.5	26	32.1	28.3	16.2	18.1	27.1	25.5	26.2	22.5	19.5
18	2	9.5	1	13.7	3.5	5.1	7.8	3.7	5.7	6.9	6	6
19	0.5	0	0	1.3	0.9	0	0.6	1.6	0	1.9	0.5	1
vWA												
10	0	0.5	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0.2
13	0	0	0	0	0	0	0	0	0	0	0	0.5
14	23.8	12	8.2	11.1	15.2	7.4	20.5	23.4	26.2	7.5	29	26
15	1	3	8.2	6.8	4.3	3.7	0.6	1.1	4.9	3.1	1	3.1
16	11.8	8.5	19.4	17.9	15.2	5.1	22.3	3.7	10.7	11.2	12.5	15
17	24.3	23.5	31.1	12.4	17.5	27.2	15.7	25.5	24.6	29.4	27	24.1
18	24.3	32	25.5	25.2	28.3	36	28.9	20.2	20.5	26.2	20.5	19.5
19	14.3	20	5.6	21.9	14.3	16.2	12	21.8	11.5	18.8	8.5	10.4
20	0.5	0.5	2	4.7	4.8	3.7	0	4.3	1.6	2.5	1.5	1
21	0	0	0	0	0.4	0.7	0	0	0	1.3	0	0.2
FGA												
17	0	0	0	0	0	0	0	0	0	0	0	0.2
18	0	0	0	2.1	0.4	0	0.6	2.1	2.5	0	3	3.9
19	19.3	9	3.1	47.4	9.1	8.8	6.6	5.9	4.9	10.6	4	4.6
20	8.4	7.5	3.6	6.4	9.1	1.5	12.7	4.8	4.1	3.8	5	5.6
21	5	10	13.3	4.7	20	19.1	10.2	5.9	16.4	15	15	10.8
21.2	0	0	0	0	0	0	0	0	0	0	0	0.7
22	29.6	23.5	21.9	7.7	21.9	14	14.5	30.9	32	19.5	17	17.5
22.2	0	0	0	0.4	0	0	0	0	0	0	1	0.3
23	13.9	19	31.6	15.9	10.9	11	16.8	18.5	22.1	10.6	17.5	23.7
23.2	0	0.5	0	0	0	0	0	0	0	0	0	0.2
24	8.9	16.5	15.3	1.7	20.4	19.8	19.3	16.5	7.4	25.6	19.5	17.2
24.2	0	1.5	0	0	0.4	0	0	0	1.6	0	1	1.1
25	9.4	8.5	4.6	9.5	3.9	11.8	13.3	7.4	7.4	8.7	9	7
26	3	2	2	3.8	3.9	10.3	3.6	4.8	0.8	3.1	6	5.5
26.2	0	0	0	0	0	0	0	1.1	0.8	0	0.5	0
27	2	1.5	4.1	0.4	0	1.5	2.4	2.1	0	3.1	1	1.2
28	0	0.5	0.5	0	0	2.2	0	0	0	0	0.5	0.2
29	0.5	0	0	0	0	0	0	0	0	0	0	0.3
D8S1179												
8	0	0	0	0	0	0	0	0	0	0	0	0.2
9	0	0.5	0	0	0	0	0	0	0	0	0	0.2
10	37.6	13.5	24	12.8	16.5	13.2	34.9	17.6	13.9	10	17	11.3
11	7.4	17	7.7	9.4	8.3	14.8	9	12.8	12.3	23.1	11	10.4
12	5.9	16	8.7	2.1	15.2	16.9	1.8	16.5	12.3	9.4	13	11.4
13	23.4	27.5	16.8	56.4	22.6	14	22.3	30.9	25.4	21.9	20	20.6
14	16.3	17	32.6	9.8	23.9	29.4	21.2	6.9	22.2	26.9	15	20
15	6.4	3.5	8.2	7.3	12.2	10.3	8.4	13.8	12.3	5.6	17	18.8
16	3	3.5	2	1.3	1.3	0.7	1.8	1.6	1.6	3.1	6.5	5.6
17	0	1.5	0	0	0	0.7	0	0	0	0	0.5	1
18	0	0	0	0.9	0	0	0.6	0	0	0	0	0.5
D21S11												
23.2	0	0	0	0	0	0	0.6	0	0	0	0	0
27	0	0	0	0	0	0	0	0	0	0	0	0.3
28	0	10.5	3.1	0.9	3.5	6.6	1.8	4.8	2.5	2.5	6	6.3
28.2	0	0	0	0	0	0	0	0	0	0	0.5	0.2
29	6.9	14	30.9	29.5	12.2	19.1	24.7	26.1	23.8	30.6	29	28
29.2	0	0	0	0	0	0	0	2.7	0.8	0	0	0
30	19.8	28.5	9.2	13.2	12.2	32.4	23.5	26.1	28.6	19.4	25	29.5
30.2	0	1	0.5	0	0	1.5	0	0	0.8	0	1.5	1.4
31	10.9	17	14.8	3	21.3	4.4	16.9	8.5	13.1	3.1	10.5	8.7
31.2	26.7	5	8.7	8.1	7.4	2.9	16.9	12.2	6.6	8.7	6.5	5.6
32	2.5	7	3.6	0.9	8.3	16.9	4.2	2.7	7.4	3.1	2.5	3.1
32.2	26.7	9.5	17.9	20.5	22.5	9.6	10.8	7.4	14.8	24.5	13	13.3
33	0.5	0.5	8.7	0	0	0.7	0.6	0	0	0	0	0.5
33.2	4	4.5	2.6	23.9	11.3	4.4	0	9	1.6	8.1	5.5	2.4
34.2	2	2.5	0	0	1.3	1.5	0	0.5	0	0	0	0.7
D18S51												
10	0	0	0	0	0	0	0	0	0	0	0	0.2
11	3	0.5	0	0.4	0	0	0	1.1	0	0	1	0.3

Table 1 Gene frequencies (%) of 13 STR loci in 12 Taiwanese populations. (Continued)

Loci & alleles	Populations											
	Bunun	Paiwan	Saisiat	Tao	Ami	Rukai	Atayal	Tsou	Pazeh	Puyumar	Hakka	Han
12	7.4	0.5	6.6	10.3	3.9	2.2	3	4.3	2.5	2.5	3.5	3.9
13	1	7	5.6	10.7	2.2	5.9	2.4	3.2	15.6	15.6	17	21.8
14	32.1	32	24	25.6	14.8	29.4	31.9	25	14.8	21.3	17.5	20.5
15	24.8	20.5	28.1	10.7	38.3	28.8	28.3	25	23.6	31.2	18.5	20
16	16.8	20.5	21.9	9	7.4	20.6	25.4	20.1	16.4	13.1	13	11.5
17	8.9	9	9.7	21	11.3	5.1	8.4	13.3	10.7	5	7.5	7.8
18	1.5	2	3.6	6.8	3.5	0	0.6	2.7	4.9	0.6	9.5	4.3
19	4.5	3.5	0.5	4.7	14.3	2.2	0	1.6	4.1	6.3	4	3.2
20	0	3	0	0.4	2.6	0.7	0	2.7	2.5	0	1.5	1.9
21	0	1	0	0.4	0.9	3.7	0	0.5	3.3	0.6	3	1.7
22	0	0	0	0	0.4	0.7	0	0.5	1.6	2.5	3.5	1.7
23	0	0.5	0	0	0.4	0	0	0	0	0	0	0.5
24	0	0	0	0	0	0.7	0	0	0	1.3	0.5	0.5
25	0	0	0	0	0	0	0	0	0	0	0	0.2
D5S818												
6	0	0	0	0	0.4	0	0	0	0	0	0	0.2
7	1.5	1	3.1	0	2.2	0	4.2	1.1	2.5	0.6	3.5	2.4
8	0	0	0.5	0	0	0	0	0.5	0	0	0	0.3
9	8.4	2.5	0.5	0.4	7	4.4	1.2	16	0.8	3.1	7	7.7
10	26.3	37	32.7	39.3	33.5	39.7	45.2	28.2	29.5	40	24	21.3
11	46	22	21.9	14.6	27	16.9	30.8	22.8	27	36.3	30.5	31.9
12	9.9	25	32.1	22.2	24.3	32.4	7.8	19.1	27.9	17.5	17.5	22
13	7.9	12.5	7.7	23.5	5.2	6.6	10.8	11.2	11.5	2.5	17.5	12.8
14	0	0	1.5	0	0.4	0	0	1.1	0.8	0	0	1.4
D13S317												
7	0	0	0	0	0	0	0	0	0	0	0.5	0.5
8	23.7	24	17.9	12	17.8	30.9	15.1	24.5	25.5	23.7	32	25.3
9	3	10	7.1	15.8	8.3	11	6	7.4	8.2	12.5	11	14.2
10	12.9	22	16.8	11.5	25.7	4.4	21.1	8.5	9.8	9.4	13.5	15.7
11	50	25	44.9	50.9	36.1	30.1	45.8	51.6	31.1	30.6	24	26.4
12	10.4	15.5	10.7	9.4	11.7	19.9	6.6	5.9	19.7	22.5	14.5	14.8
13	0	2.5	2.6	0.4	0	3.7	5.4	2.1	4.1	0	4.5	3.1
14	0	1	0	0	0.4	0	0	0	1.6	1.3	0	0
D7S820												
7	5	0	0	0	0	0	0	1.1	0	0	0.5	0.3
8	4	19	11.2	27.4	10.9	7.4	7.8	16.5	17.2	25	16.5	14.5
9	1	0.5	1.6	6.8	3.9	2.9	4.2	10.6	4.9	5	4	7.5
10	30.2	19	26	6.4	22.6	33.8	25.9	10.1	18	12.5	11.5	16.9
11	33.6	37	43.9	41.9	38.7	30.1	39.8	43	31.1	34.3	38	32.9
12	19.8	17.5	11.2	15	16.1	9.6	15.1	8	25.5	19.4	25.5	24.3
13	6.4	5	6.1	2.1	4.8	14.7	7.2	9.6	3.3	1.9	2	3.4
14	0	2	0	0.4	3	1.5	0	1.1	0	1.9	2	0.2
D16S539												
7	0	0	0	0	0	0	0	0	3.3	0	0	0
8	0	0	0	0	0.4	0	0	0	0	0	0	0.3
9	31.7	30	29.6	23.5	20.4	16.2	40.4	27.7	25.4	18.8	22	24.9
10	29.2	11.5	12.2	7.7	6.5	16.2	8.4	28.2	12.3	18.8	12	13.7
11	15.8	20.5	27.6	28.2	38.3	11	15.1	15.4	20.5	21.3	29	28.8
12	13.9	13	22.5	23.9	23.6	27.2	16.9	19.7	24.6	18.8	25.5	21.8
13	9.4	20.5	7.1	16.7	10	27.9	12	8.5	13.9	19.2	10	9
14	0	4	0.5	0	0.4	1.5	6.6	0.5	0	3.1	1.5	1.2
15	0	0.5	0.5	0	0.4	0	0.6	0	0	0	0	0.3
TH01												
5	0	0	0	0	0	0	0	0	0.8	0	0	0
6	17.3	14.5	9.3	2.6	9.6	15.4	18.1	18.6	10.7	13.8	13	10.2
7	39.6	35	53.6	32.5	39.6	39.7	59	37.8	30.4	30	26	26.1
8	1.5	19.5	0	6.8	8.2	18.4	2.4	7.4	13.1	23.7	6.5	3.8
9	28.7	26	30.1	56.8	39.6	19.1	12.7	28.2	39.3	23.1	41	51.2
10*	12.9	5	7	1.3	3	7.4	7.8	8	5.7	9.4	13.5	8.7
TPOX												
7	0	0	0	0	0	0	0	0	0	0	0	0.2
8	18.8	60	48.5	55.6	54.8	52.2	30.1	35.6	49.2	56.3	57.5	58.2
9	14.9	8	17.3	15.4	10	5.9	4.2	13.9	4.1	8.7	8.5	11.9
10	0.5	0	0.5	8.5	0	0.7	2.4	0	1.6	0.6	3.5	1.7
11	64.8	29	33.7	20.5	33.9	41.2	63.3	50	42.6	32.5	28.5	24.9
12	0	3	0	0	0.9	0	0	0.5	2.5	1.9	1.5	3.1
13	1	0	0	0	0.4	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0.5	0
CSF1PO												
7	0	0	0	0	0	0	0	0	0.8	0	0	0.7
8	7.4	0	0	0	0.4	0	1.8	2.7	0.8	0.6	0	0.2

Table 1 Gene frequencies (%) of 13 STR loci in 12 Taiwanese populations. (Continued)

Loci & alleles	Populations											
	Bunun	Paiwan	Saisiat	Tao	Ami	Rukai	Atayal	Tsou	Pazeh	Puyumar	Hakka	Han
9	0	0.5	0.5	0.9	1.7	0	0.6	0.5	1.6	0	7.5	6.1
10	11.4	29	19.9	19.2	34.8	29.4	20.5	28.7	21.5	28.7	16	24.2
11	32.2	36.5	39.3	55.1	19.1	35.3	51.8	22.9	26.2	34.4	27	23.4
12	20.3	30	39.3	17.9	35.7	32.4	19.9	31.9	30.3	31.3	38	36.8
13	28.7	3.5	1	5.6	5.7	2.9	5.4	11.7	18	5	10.5	7.7
14	0	0.5	0	0.9	2.6	0	0	1.6	0.8	0	1	0.9
15	0	0	0	0.4	0	0	0	0	0	0	0	0

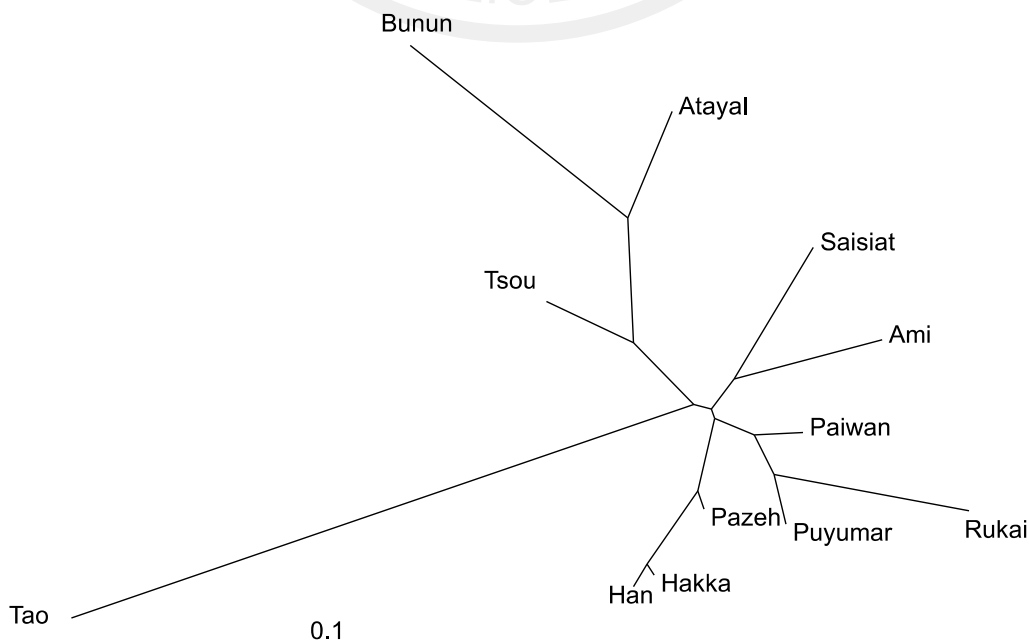
^a The frequency of alleles 9.3 and 10 were combined in this study.

Table 2 P value of the exact test for the 13 STR loci in 12 populations.

Loci	Populations											
	Bunun	Paiwan	Saisia	Tao	Ami	Rukai	Atayal	Tsou	Pazeh	Puyumar	Hakka	Han
D3S1358	0.421	0.179	0.400	0.832	0.838	0.455	0.654	1.000	0.152	1.000	0.662	0.537
vWA	0.589	1.000	0.233	1.000	0.489	0.435	0.791	0.587	1.000	0.282	0.467	0.172
FGA	0.635	0.579	0.637	0.095	0.273	0.285	0.289	0.809	0.081	0.376	0.352	0.747
D8S1179	0.832	0.800	0.649	0.038	0.078	0.079	0.809	0.329	1.001	0.252	0.215	0.722
D21S11	0.072	0.323	0.640	0.824	1.000	0.404	1.000	0.107	0.346	0.599	0.629	0.159
D18S51	0.497	0.817	0.453	0.142	0.430	0.385	1.000	0.785	0.727	1.000	1.003	0.494
D5S818	0.842	1.000	0.103	0.702	0.676	1.000	0.268	0.190	0.354	0.818	0.350	0.080
D13S317	0.072	0.287	0.309	0.141	0.688	0.779	0.057	1.000	0.553	0.291	0.249	0.548
D7S820	0.657	0.666	0.311	0.347	0.698	0.059	0.178	1.000	0.372	0.136	1.000	0.791
D16S539	0.176	1.000	0.627	0.493	0.430	0.763	0.822	0.214	1.001	0.107	0.629	0.051
TH01	0.535	0.381	1.000	0.451	0.439	1.000	0.652	0.661	1.000	0.298	0.013	0.815
TPOX	0.829	1.000	1.000	1.000	0.258	0.093	0.813	0.536	0.449	0.256	0.838	0.339
CSF1PO	0.170	0.188	1.000	0.134	1.000	0.288	1.000	0.467	0.377	0.621	0.291	0.707

Table 3 Genetic distance matrix among 12 populations on 13 STR loci generated using Nei's method [21].

	Bunun	Paiwan	Saisiat	Tao	Ami	Rukai	Atayal	Tsou	Pazeh	Puyumar	Hakka	Han
Bunun	0											
Paiwan	0.1380	0										
Saisiat	0.1313	0.0732	0									
Tao	0.2247	0.1450	0.1834	0								
Ami	0.1410	0.0606	0.0631	0.1761	0							
Rukai	0.1790	0.0489	0.0873	0.2345	0.0960	0						
Atayal	0.0807	0.0827	0.0696	0.2123	0.1155	0.1077	0					
Tsou	0.0873	0.0604	0.0853	0.1734	0.0949	0.0953	0.0844	0				
Pazeh	0.1030	0.0306	0.0548	0.1584	0.0529	0.0651	0.0946	0.0432	0			
Puyumar	0.1474	0.0410	0.0755	0.1552	0.0623	0.0516	0.1110	0.0777	0.0413	0		
Hakka	0.1500	0.0543	0.0753	0.1707	0.0676	0.0939	0.1203	0.0677	0.0138	0.0508	0	
Han	0.1696	0.0600	0.0786	0.1729	0.0727	0.1036	0.1394	0.0798	0.0177	0.0615	0.0027	0

**Fig. 1** Neighbor-joining tree constructed with the genetic distances in Table 3 using the PHYLIP program.

The degree of genetic relatedness between the 12 populations was examined. Genetic distances among the 12 populations, based on the 13 STR loci data, are shown in Table 3. A Neighbor-joining tree was constructed using the PHYLIP program and is shown in Fig. 1. The Neighbor-joining tree grouped the Pazeh, Hakka and Han populations in the same cluster, in agreement with the Hanizational hypothesis of Pepo population. The hypothesis that the Hakka and Han populations are closely related is supported by this study. The Tao population is the most isolated tribe, not only geographically but also genetically. This population has the least opportunity to mix with other tribes, confirmed by the longer branch in the Neighbor-joining tree than any other populations. The least similarity was observed between Tao population and any of the other populations, as shown in the genetic distances (Table 3 and Fig. 1). Genetic distances can be easily affected by sampling. There were greater chances of interbreeding among tribes in a small island like Taiwan.

In order to reveal the allele similarity within 26 alleles produced by the 13 loci, cross comparison was performed with the total number of 1,307 samples. All the samples produced 853,471 compared groups. The matching distribution of all compared groups in 26 alleles is shown in Table 3. No two individuals are identical in the 853,471 compared groups for all 26 alleles of the Taiwanese samples. The greatest similarity of compared group is the one matched pair of 24 alleles. The two individuals belong to the Atayal tribe. Three comparisons shared 21 of the 26 alleles. The six samples of the comparisons came from the tribes, Atayal, Bunun and Tao. For the 11 comparison groups that shared 20 alleles, there were 8 groups in which the samples came from individuals of the same tribe. The 8 groups were 2 from Amy, 1 from Atayal, 1 from Bunun, 1 from Hakka, 1 from Saisiat and 2 from Tao. Considering the population sizes for each of the 12 populations, it is inevitable that members of the same tribe show greater relatedness. Moreover the small population increases the possibility of sampling error. Since the highest matching probability is only 1.44×10^{-13} in this study, it is reasonable that no identical pair for 26 alleles was found in 853,471 possible combinations of sample pairs.

Acknowledgments

Financial support by No. NSC88-2314-B-015-001 from National Science Council and Ministry of Interior, Taiwan, Republic of China.

References

1. Sykes B, Leiboff A, Low-Beer J, Tetzner S, Richards M. The origins of the Polynesians: an interpretation

Table 4 Distributions of matched alleles on 13 STR loci in the cross comparisons of 1307 unrelated individuals.

Number of matched allele	Number of matched pair	Frequency	Cumulative frequency
0	9	0.0011	0.0011
1	146	0.0171	0.0182
2	956	0.1120	0.1302
3	4,597	0.5386	0.6688
4	15,301	1.7928	2.4616
5	37,927	4.4439	6.9054
6	73,530	8.6154	15.5209
7	114,281	13.3901	28.9110
8	144,120	16.8863	45.7973
9	148,754	17.4293	63.2266
10	126,654	14.8399	78.0665
11	90,098	10.5567	88.6232
12	53,701	6.2921	94.9152
13	26,753	3.1346	98.0498
14	11,140	1.3053	99.3551
15	3,883	0.4550	99.8101
16	1,208	0.1415	99.9516
17	294	0.0344	99.9861
18	75	0.0088	99.9948
19	29	0.0034	99.9982
20	11	0.0013	99.9995
21	3	0.0004	99.9999
22	0	0.0000	99.9999
23	0	0.0000	99.9999
24	1	0.0001	100.0000
25	0	0.0000	100.0000
26	0	0.0000	100.0000
Total	853,471	100.0000	100.0000

from mitochondrial linkage analysis. *Am J Hum Genet* 1995;57:1463-75.

2. Redd AJ, Takezaki N, Sherry ST, McGarvey ST, Sofro ASM, Stoenking M. Evolutionary history of the COII/tRNALYS intergenic 9 base pair deletion in the human mitochondrial DNAs from the Pacific. *Mol Biol Evol* 1995;12:604-15.
3. Melton T, Peterson R, Redd AJ, Saha N, Sofro ASM, Martinson J, Stoneking M. Polynesian genetic affinities with Southeast Asian populations as identified by mtDNA analysis. *Am J Hum Genet* 1995;57:403-14.
4. Melton T, Clifford S, Martinson J, Batzer M, Stoneking M. Genetic evidence for the proto-Austronesian homeland in Asia: mtDNA and nuclear DNA variation in Taiwanese aboriginal tribes. *Am J Hum Genet* 1998;63:1807-23.
5. Hagelberg E, Kayser M, Nagey M, Roewer L, Zimdahl H, Krawczak M, Lio P, Schiefenovel W. Molecular genetic evidence for the human settlement of the Pacific: analysis of mitochondrial DNA, Y Chromosome and HLA markers. *Philos Trans R Soc Lond B Biol Sci* 1999;354:141-52.
6. Lee JCI, Chen CH, Tsai LC, Linacre A, Chang JG. The screening of 13 short tandem repeat loci in the Chinese population. *Forensic Sci Int* 1997;87:137-44.
7. National Printing Office. Introduction of Taiwan: historical section. Taipei, ROC, 1997. (Chinese)
8. Li PJK. History and interaction of Pepo tribe in

- Taiwan. Formosa Folkways, Taipei, ROC, 1997. (Chinese)
9. Department of population Census. 1990 Census Report of the Republic of China in Taiwan. Taipei, ROC, 1992. (Chinese)
 10. Li PJK. Populations and movement of Austronesian in Taiwan. Formosa Folkways. Taipei, ROC, 1997. (Chinese)
 11. Lin M, Chu CC, Lee HL, Chang SL, Ohashi J, Tokunaga K, Akaza T, Juji T. Heterogeneity of Taiwan's indigenous population: possible relation to prehistoric Mongoloid dispersals. *Tissue Antigens* 2000;55:1-9.
 12. Bellwood P. The Austronesian dispersal and the origin of language. *Sci Am* 1991;7:70-3.
 13. Sparkes R, Kimpton C, Watson S, Oldroyd N, Clayton T, Barnett L, Arnold J, Thompson C, Hale R, Chapman J, Urquhart A, Gill P. The validation of a 7-locus multiplex STR test for use in forensic casework: (I) mixtures, ageing, degradation and species studies. *Int J Legal Med* 1996;109:186-94.
 14. Sparkes R, Kimpton C, Gilbard S, Carne P, Andersen J, Oldroyd N, Thomas D, Urquhart A, Gill P. The validation of a 7-locus multiplex STR test for use in forensic casework: (II) artefacts, casework studies and success rates. *Int J Legal Med* 1996;109:195-204.
 15. Pfeiffer H, Huhne J, Seitz B, Brinkmann B. Influence of soil and exposure period on DNA recovery from teeth. *Int J Legal Med* 1999;112:142-4.
 16. Sweet D, Hildebrand D. Saliva from cheese bite yields DNA profile of burglar: a case report. *Int J Legal Med* 1999;112:201-3.
 17. Larkin A, Harbison SA. An improved method for STR analysis of bloodstained denim. *Int J Legal Med* 1999;112:388-90.
 18. Hantschel M, Hausmann R, Lederer T, Martus P, Betz P. Population genetics of nine short tandem repeat (STR) loci - DNA typing using the AmpFISTR Profiler PCR amplification kit. *Int J Legal Med* 1999;112:393-5.
 19. Neuhuber F, Radacher M, Meisner N, Tutsch-Bauer. Nine STR markers plus amelogenin (AmpFISTR Profiler Plus): a forensic study in an Austrian population. *Int J Legal Med* 1999;113:60-2.
 20. Borys S, Vanstone H, Carmody G, Fournery R. Allele frequencies for the COFILER™ STR loci in the Canadian Caucasian and Canadian First Nations Populations. *J Forensic Sci* 2000;45:945-8.
 21. Nei M. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 1978;89:583-90.
 22. Felsenstein J. *PHYLIP* (Phylogeny Inference Package) Version 3.5c, University of Washington, 1986.