Population Genetic Study in Hainan Island, China II. Genetic Affinity Analyses

Naruya Saitou¹, Keiichi Omoto², Chuanshu Du³, and Ruofu Du⁴

- 1 Laboratory of Evolutionary Genetics, National Institute of Genetics, Mishima, Japan
- 2 International Research Center for Japanese Studies, Kyoto, and Department of Anthropology, Graduate School of Science, University of Tokyo, Tokyo, Japan
- 3 Department of Medical Genetics, Sun Yat-sen University of Medical Sciences, Guangzhou, China
- 4 Department of Human Genetics, Institute of Genetics, Academica Sinica, Beijing, China

Received May 13, 1993

Abstract A population genetic study was carried out on six populations (two Li populations, two Miao populations, one Hui population, and one Han population) in Hainan Island, Southern China. Allele frequency data for these six populations and those for some other Asian populations were used to estimate genetic distances between populations, and genetic affinity dendrograms and networks were constructed by using UPGMA and the neighbor-joining method, respectively. The two Li populations showed a close relationship, while the two Miao populations were relatively distant from each other. These six populations of Hainan Island showed a close affinity with Zhuang of Southern China, Thailanders, and Javanese, but they are distant from populations around northern China (Japanese, Mongolian, and Korean). Divergence time between the Li-Miao group and Japanese was estimated to be roughly 19,000–26,000 years.

Key Words: ethnic minorities in China, Hainan Island, genetic polymorphism, genetic affinity network, population genetics

INTRODUCTION

Several ethnic groups are living in Hainan Island of China, including three ethnic minorities, Li, Miao, and Hui. In order to understand the relationships among these populations in Hainan Island and other human populations in Asia, a joint Japanese-Chinese field survey was carried out in 1985 and 1987, and a total of 792 blood samples were obtained from six human populations (two Li and Miao populations, one Hui and Han populations; see Omoto *et al.*, 1993). A total of 23 genetic loci were examined for those six populations and Omoto *et al.* (1993) presented allele frequency data for those loci.

The purpose of this paper is to present the result of genetic affinity analyses for these Hainan Island populations and for surrounding Asian populations. A possible history of these Hainan Island populations will be discussed based on the results on genetic affinity.

MATERIAL AND METHODS

Allele frequency data

Omoto *et al.* (1993) studied six human populations of Hainan Island. These populations were named Li-1, Li-2, Miao-1, Miao-2, Hui, and Han. Blood samples for Li-1, Miao-1, and Hui were collected in 1985, and six serum protein loci (Hp, Tf, Gc, C6, C7, and Bf), seven red blood cell enzyme loci (ACP, ESD, 6PGD, PGM1, GPT, ADA, and AK), and six blood group loci (ABO, MN, Rh, Duffy, Kell, and Lutheran) were examined, except Rh for Hui. Blood samples for Li-2, Miao-2, and Han were collected in 1987, and five serum protein loci (Hp, Tf, Gc, C6, and Pi), four red blood cell enzyme loci (ACP, ESD, 6PGD, and PGM1), and eight blood group loci (ABO, MNSs, Rh, Duffy, Kell, Lutheran, P, and Diego) were examined. Gc and PGM1 loci were subtyped, and "subtype" will be added when subtyped data are used in the following.

Different sets of genetic loci were examined in the two studies held in 1985 and 1987, and data for the following 13 loci are available for all the six Hainan Island populations: Hp, Tf, Gc-subtype, C6, ACP, ESD, 6PGD, PGM1-subtype, ABO, MN, Duffy, Kell, and Lutheran. Allele frequency data for these shared loci were used for genetic affinity analysis of the six populations. Because only Hui population lacked the allele frequency data for the Rh blood group locus, data for 14 loci (Rh and the above 13 loci) were also used for estimating the genetic affinity of the remaining five populations (Li-1, Li-2, Miao-1, Miao-2, and Han). Japanese and English populations were also included in the analysis. Allele frequency data for Japanese were taken from Omoto (1978) except for Gc-subtype, C6, PGM1-subtype, Kell, and Lutheran. Allele frequency data of the Japanese population for Gc-subtype, C6, and PGM1-subtype were taken from Yuasa *et al.* (1984), Tokunaga *et al.* (1986), and Yuasa *et al.* (1973). Data for the English population were taken from Roychoudhury and Nei (1988).

We also compared the six Hainan Island populations with 12 other human populations mostly living in Asia: Mongolian, Zhuang, Japanese, Ainu, Korean, Southern Chinese, Malays, Javanese, Thailanders, Filipinos, Negritos (Aetas in the Philippines), and English. Because C6, Gc, and Lutheran were not examined in some of those populations, allele frequency data for the following 10 loci were compared: Hp, Tf, ACP, ESD, 6PGD, PGM1, ABO, MN, Duffy, and Kell. In this case, conventional electrophoresis data for PGM1 were used, and subtypes were

combined for populations in which only subtyped data were available.

Data for Japanese and English were as above. Allele frequency data for Ainu were taken from Omoto (1978) except those for the Kell locus. Data for Negritos were taken from Omoto *et al.* (1978) except those for the four blood group loci. Data for Korean, Southern Chinese (Fujien Province), Malays, and Javanese were taken from Saha and Tay (1992). Those for Mongolian and Zhuang were taken from Goedde *et al.* (1984) and Yuan *et al.* (1984) except for Hp. Remaining allele frequency data were taken from Roychoudhury and Nei (1988).

Lastly, Li-1, Miao-1, Hui, Japanese, and English were compared for the following 18 commonly examined loci: Hp, Tf, Gc-subtype, C6, C7, Bf, ACP, ESD, 6PGD, PGM1-subtype, GPT, ADA, AK, ABO, MN, Duffy, Kell, and Lutheran. Allele frequency data for Japanese were taken as above, with the additional 5 loci. Data for C7 and Bf were from Nakamura *et al.* (1984) and Horai (1976), respectively, while those for GPT, ADA, and AK were taken from Omoto (1978). All the data for English were taken from Roychoudhury and Nei (1988).

Genetic distances and genetic affinity relationship construction

Two kinds of genetic distances were used in the present study. One is Nei's (1972) standard genetic distance (D_{std}). D_{std} between populations X and Y is defined by

$$D_{std} = -\log (J_{XY}/\sqrt{J_X}J_Y). \tag{1}$$

 J_{XY} , J_X , and J_Y are defined by

$$J_{XY} = \sum_{k=1}^{n} \sum_{i=1}^{m_k} x_{ik} y_{ik} / n,$$
 (2a)

$$J_{X} = \sum_{k=1}^{n} \sum_{i=1}^{m_{k}} x_{ik}^{2} / n,$$
(2b)

$$J_{Y} = \sum_{k=1}^{n} \sum_{i=1}^{m_{k}} y_{ik}^{2} / n, \qquad (2c)$$

where x_{ik} and y_{ik} are frequencies of allele *i* of locus *k* for populations *X* and *Y*, respectively, *n* is the number of loci compared, and m_k is the number of alleles at locus *k*. J_X and J_Y are homozygosities for populations *X* and *Y*, respectively, and heterozygosities or gene diversities for these populations are defined by

SAITOU et al.

$$H_X = 1 - J_X, \tag{3a}$$

$$H_Y = 1 - J_Y. \tag{3b}$$

 D_{std} has been shown to be proportional with divergence time of populations when many genetic loci are randomly sampled from the genome (Nei, 1987).

The other genetic distance measure used is D_A proposed by Nei *et al.* (1983), which is defined by

$$D_{A} = \sum_{k=1}^{n} (1 - \sum_{i=1}^{m_{k}} \sqrt{x_{ik} y_{ik}}) / n.$$
(4)

 D_A distance was shown to recover the true branching pattern of populations more efficiently than other genetic distances by computer simulation (Nei *et al.*, 1983). It should be mentioned that D_A is closely related to the chord distance (Dc) of Cavalli-Sforza and Edwards (1967), which is defined by

$$D_{C} = 2\sqrt{2}D_{A} / \pi.$$
⁽⁵⁾

Genetic affinity dendrograms were constructed by using UPGMA (Sokal and Sneath, 1963), which is often called "cluster analysis" in morphological studies. Genetic affinity networks were constructed by using the neighbor-joining (NJ) method of Saitou and Nei (1987). In the case of UPGMA, a pair of populations with the smallest genetic distance are first clustered and this process continues until all the populations are clustered. Therefore, the similarity between populations is the dominant factor for clustering in UPGMA. As for the NJ method, however, the principle of minimum evolution is used, and the population pair with the smallest distance may not be clustered first. This is because all the distance data are simultaneously considered to choose the clustering pair (neighbors) in the case of the NJ method. Since the least square method is used to estimate branch lengths for the NJ method, some branches may become negative when the distance matrix used is far from additivity. It should also be noted that a UPGMA tree is rooted (dendrogram), while an NJ tree is unrooted (network).

RESULTS

 Comparison of Li-1, Miao-1, Hui, Li-2, Miao-2, Han, Japanese, and English using allele frequency data for 13 polymorphic loci Genetic distances between the above eight populations using allele frequency data

132

	Miao-1	Li-1	Hui	Miao-2	Li-2	Han	Japanese	English
Miao-1	0.375	0.009	0.014	0.010	0.009	0.014	0.014	0.044
Li-1	0.012	0.320	0.008	0.007	0.002	0.005	0.012	0.045
Hui	0.013	0.006	0.323	0.008	0.010	0.007	0.010	0.052
Miao-2	0.016	0.010	0.008	0.344	0.007	0.006	0.011	0.042
Li-2	0.013	0.003	0.011	0.009	0.318	0.005	0.015	0.045
Han	0.018	0.008	0.008	0.007	0.009	0.308	0.010	0.045
Japanese	0.018	0.016	0.008	0.015	0.023	0.009	0.344	0.039
English	0.067	0.075	0.074	0.072	0.075	0.069	0.061	0.335

Table 1. Genetic distances among Li-1, Miao-1, Hui, Li-2, Miao-2, and Han populations

Figures above and below the diagonal are D_A and D_{std} , respectively, and those on the diagonal are gene diversity. Allele frequency data for 13 polymorphic loci (Hp, Tf, Gc-subtype, C6, ACP, ESD, 6PGD, PGM1-subtype, ABO, MN, Duffy, Kell, and Lutheran) were used.



Fig. 1. A UPGMA dendrogram for Li-1, Miao-1, Hui, Li-2, Miao-2, Han, Japanese, and English. *D*_{std} distances of Table 1 are used.



Fig. 2. NJ networks for Li-1, Miao-1, Hui, Li-2, Miao-2, and Han populations. Each line of the network is proportional to branch lengths (distances). Negative lengths are shown in broken lines. (A) D_{std} distances of Table 1 were used. (B) D_A distances of Table 1 are used.

for 13 polymorphic loci are given below (D_{std}) and above (D_A) the diagonal in Table 1. Figures on the diagonal are gene diversities (average heterozygosities) for each population. D_A values are often smaller than D_{std} values, and both genetic distances are much smaller than gene diversities. Clearly, most of the genetic variation exists within populations.

UPGMA dendrograms constructed by using D_{std} and D_A distances are shown in Fig. 1. The clustering pattern of the UPGMA tree (Fig. 1B) based on D_A distances is somewhat different from that (Fig. 1A) based on D_{std} , in terms of the clustering of Han, Miao-2, and Hui. Japanese are slightly distant from those populations, and English are far apart from the remaining populations.

NJ networks constructed by using D_{std} and D_A distances are shown in Fig. 2. Two Li populations (Li-1 and Li-2) have the smallest genetic distance and they are always clustered in all four trees in Figs. 1 and 2. The genetic distance between two Miao populations (Miao-1 and Miao-2) is much larger than that between the two Li populations, and they do not form a cluster in any tree. Miao-1 is always located at the position of the outgroup to all the remaining five Hainan populations in all four trees of Figs. 1 and 2.

(2) Comparison of Li-1, Miao-1, Li-2, Miao-2, Han, Japanese, and English using allele frequency data for 14 polymorphic loci

Genetic distances for the above seven populations using allele frequency data for 14 polymorphic loci are given in Table 2. Allele frequency data for the Rh blood group locus were added by excluding the Hui population from the previous comparison. Compared to genetic distances of Table 1 in which 13 polymorphic loci were used, those of Table 2 are slightly larger for many pairs of populations. This is probably because there are eight alleles in the Rh locus and the contribution of



Fig. 3. UPGMA dendrograms for Li-1, Miao-1, Li-2, Miao-2, Han, Japanese, and English. (A) D_{std} distances of Table 2 were used. (B) D_A distances of Table 2 are used.

the locus to genetic distances is larger than the average of the 13 polymorphic loci used in the previous section.

UPGMA dendrograms were produced based on D_{std} and D_A distances (Fig. 3). Japanese is now clearly distant from the five Hainan Island populations, and English is far apart from the remaining populations. The two Li populations were clustered

	Miao-1	Li-1	Miao-2	Li-2	Han	Japanese	English
Miao-1	0.364	0.011	0.011	0.008	0.019	0.020	0.060
Li-1	0.008	0.327	0.010	0.006	0.007	0.019	0.056
Miao-2	0.015	0.009	0.336	0.007	0.009	0.019	0.060
Li-2	0.011	0.002	0.009	0.318	0.011	0.026	0.066
Han	0.018	0.008	0.008	0.011	0.316	0.014	0.052
Japanese	0.027	0.026	0.023	0.036	0.013	0.347	0.045
English	0.088	0.094	0.092	0.098	0.080	0.069	0.365

 Table 2. Genetic distances among Li-1, Miao-1, Li-2, Miao-2, Han, Japanese, and English populations

Figures above and below the diagonal are D_A and D_{std} , respectively, and those on the diagonal are gene diversity. Allele frequency data for 14 polymorphic loci (Hp, Tf, Gc-subtype, C6, ACP, ESD, 6PGD, PGM1-subtype, ABO, MN, Duffy, Kell, Lutheran, and Rh) were used.



Fig. 4. NJ networks for Li-1, Miao-1, Li-2, Miao-2, and Han populations. Each line of the network is proportional to branch lengths (distances). Negative lengths are shown in broken lines. (A) D_{std} distances of Table 2 are used. (B) D_A distances of Table 2 are used.

as in Fig. 1, and the location of Miao-1 is the same for Figs. 1A, 1B, 3A, and 3B. If we ignore Hui in Fig. 1A, the branching pattern becomes identical with that of Fig. 3A. It should be noted that D_{std} distances were used both for Figs. 1A and 3A. As for the trees in which D_A distances were used, however, there is an inconsistency on the clustering of Han, Miao-2, and the Li cluster (see Figs. 1B and 3B). This inconsistency may be explained by small lengths of some internal branches.

NJ networks based on genetic distances of Table 2 are shown in Fig. 4. Now Han becomes the outgroup to the remaining four Hainan populations in both Figs. 4A and 4B. The two Li populations clustered in Figs. 2A, 2B, and 4A, whereas they did not cluster in Fig. 4B, where D_A distances were used. Because the branching pattern among the five Hainan Island populations differs from tree to tree, the relationship among Han, Miao-1, Miao-2, and the Li cluster is not clear.

It seems that D_{std} distances are more stable than D_A distances when allele frequency data for different set of loci are used. Thus only D_{std} distances are used in the following analyses.

(3) Comparison of six populations in Hainan Island with 12 other human populations using allele frequency data for 10 polymorphic loci

Genetic distances (*D*_{std}) among 18 human populations including the six Hainan Island populations are given in Table 3, and a UPGMA dendrogram was constructed based on the genetic distances of Table 3 (see Fig. 5). All the Asian populations except Negritos (Aetas) of the Philippines are closely related to each other. All the six populations of the Hainan Island (those with asterisks in Fig. 5) belong to cluster A, in which Thailanders, Javanese, and Zhuang of southern China are also included. Two Li populations again make a cluster, and they were closest to Zhuang. On the other hand, populations of East Asia (Japanese, Mongolian, Korean, and Ainu) form a different cluster, and they show more affinity with Filipinos, Malays, and Southern Chinese than to cluster A.

An NJ network was constructed based on the genetic distances of Table 3. When all the 18 populations were compared, English and Malays were clustered, and a large negative branch length was observed for the branch going to Malays (tree not shown). The genetic distance (0.026) between Malays and English is unusually smaller than those between the remaining populations and English (the average = 0.063), while distances between Malays and Southern Chinese, Javanese, Thailanders, Mongolian, and Japanese are smaller than 0.01. This situation is quite different from the additivity of branch lengths, which is the ideal situation for the NJ method (Saitou and Nei, 1987), and apparently this created a large negative branch for Malays. Therefore, English were excluded from the comparison, and the NJ network for the remaining 17 populations was constructed (see Fig. 6).

Now Malays are clustered with Southern Chinese as in the UPGMA dendrogram

		Ξ	6	(3)	(4)	(5)	(9)	6	(8)	(6)	(10)	([])	(12)	(13)	(14)	(15)	(16)	(12)	(18)
(1)	Miao-1	0.346																	
(2)	Li-1	0.010	0.284																
(3)	Hui	0.010	0.005	0.282															
(4)	Miao-2	0.017	0.008	0.005	0.298														
(5)	Li-2	0.014	0.002	0.010	0.008	0.277													
(9)	Han	0.019	0.007	0.008	0.008	0.009	0.262												
(7)	Japanese	0.016	0.015	0.009	0.013	0.022	0.008	0.296											
(8)	English	0.067	0.076	0.073	0.075	0.083	0.070	0.052	0.316										
(6)	Korean	0.020	0.019	0.011	0.017	0.027	0.009	0.002	0.066	0.277									
(10)	S. Chinese	0.016	0.012	0.011	0.015	0.017	0.006	0.005	0.037	0.009	0.302								
(11)	Malays	0.019	0.018	0.018	0.018	0.021	0.013	0.009	0.026	0.015	0.003	0.315							
(12)	Javanese	0.017	0.007	0.008	0.009	0.009	0.005	0.011	0.054	0.012	0.005	0.008	0.285						
(13)	Thailanders	0.015	0.006	0.008	0.009	0.008	0.002	0.007	0.056	0.010	0.003	0.008	0.003	0.289					
(14)	Filipinos	0.015	0.020	0.015	0.029	0.028	0.015	0.008	0.052	0.009	0.007	0.013	0.014	0.013	0.286				
(15)	Negritos	0.049	0.053	0.043	0.044	0.060	0.030	0.020	0.057	0.021	0.022	0.022	0.038	0.030	0.025	0.281			
(16)	Ainu	0.028	0.025	0.017	0.016	0.032	0.015	0.004	0.061	0.005	0.014	0.015	0.017	0.014	0.022	0.024	0.283		
(17)	Mongolian	0.012	0.015	0.009	0.014	0.020	0.008	0.002	0.055	0.002	0.005	0.009	0.009	0.007	0.007	0.021	0.006	0.303	
(18)	Zhuang	0.009	0.003	0.004	0.009	0.005	0.005	0.012	0.071	0.013	0.008	0.014	0.006	0.005	0.012	0.038	0.023	0.010	0.292
Figur 6PGI	es below the), PGM1, AB	diagona O, MN,	I are $D_{\rm s}$ Duffy,	and I and I and I	those or ell) wer	n the dia e used.	agonal	are gen	e diver	sity. Al	lele fre	quency	data fo	r 10 po	lymorpł	nic loci	(Hp, T	f, ACP	, ESD,

Table 3. Genetic distances among 18 human populations

SAITOU et al.

of Fig. 5. It is noteworthy that all the six populations of Hainan Island (those with asterisks in Fig. 6) are included in the cluster above the branch A, in which Thailanders, Javanese, and Zhuang are also included. This cluster corresponds to cluster A of Fig. 5. There are three neighboring pairs of populations in this cluster; [Li-1*, Li-2*], [Miao-2*, Hui*], and [Han*, Thailanders]. The same three pairs of populations are clustered in the UPGMA dendrogram (see Fig. 5). There are two groups inside this cluster separated by branch B; one including four ethnic minorities in southern China (Miao, Li, Hui, and Zhuang) and the other including Thailanders, Javanese, and Han in Hainan Island. The latter group has some affinity with a group consisting of Malays, Southern Chinese, and Filipinos, though these two groups are separated by the relatively long branch A (see Fig. 6).



Fig. 5. A UPGMA dendrogram of 18 human populations based on genetic distance data of Table 3. Populations with asterisks are those living in Hainan Island. Cluster A includes all those six Hainan Island populations.



Fig. 6. An NJ network of 17 human populations based on genetic distance data of Table 3. Each line of the network is proportional to branch lengths (distances). Negative lengths are shown in broken lines. Populations with asterisks are those living in Hainan Island. The cluster above the branch A includes all those six Hainan Island populations.

The four populations of East Asia (Japanese, Mongolian, Korean, and Ainu) do not make a single cluster, but they form a group located between the three populations (Malays, Southern Chinese, and Filipinos) and Negritos, which is distant from all the remaining populations.

(4) Comparison of Li-1, Miao-1, Hui, Japanese, and English using allele frequency data for 18 polymorphic loci and 18 monomorphic loci

Genetic distances (D_{std}) among Li-1, Miao-1, Hui, Japanese, and English are given in Table 4, with their standard errors in parentheses. When genetic distances were computed, we used the method of Omoto (1982), who considered a hypothetical situation of the existence of some monomorphic loci to obtain a realistic divergence time between populations. In the present case, existence of 18 monomorphic loci as well as 18 polymorphic loci was assumed so as to bring the distance between English and Japanese and gene diversities for these populations close to the estimates obtained from the comparison of many loci. Nei and Roychoudhury (1982) estimated gene diversities (average heterozygosities) of English and Japanese to be 0.14 and 0.16 for protein loci and 0.11 and 0.20 for blood group loci, respectively, and D_{std} between Mongoloids (data were mostly from Japanese) and English was estimated to be 0.019 based on allele frequency data for 85 loci. In Table 4, gene diversities (figures on diagonal) for Japanese and English are estimated to be both 0.152 and the genetic distance (D_{std}) between Japanese and English is 0.0167 ± 0.0087. Both gene diversities and the genetic distance are more or less similar to the corresponding estimates of Nei and Roychoudhury (1982). Needless to say, those estimates in Table 4 are smaller than corresponding estimates in Tables 1 and 2 because of inclusion of 18 monomorphic loci.

Figure 7 shows a UPGMA dendrogram (Fig. 7A) and an NJ network (Fig. 7B), based on the genetic distances of Table 4. Open boxes in the UPGMA dendrogram represent standard errors of each branching point. Standard errors (SEs) of branching points were estimated by assuming the covariances between distances to be zero, hence slightly overestimated SEs may be obtained. In any case, it is clear that SEs of branching points are quite large.

Although Hui and Japanese make a cluster in the UPGMA dendrogram, Hui has an affinity with the other two Hainan populations, as seen in the NJ network. This reflects smaller genetic distances between Hui and the Li-Miao group than those between Japanese and the Li-Miao group (see Table 4). It is possible that Hui experienced a certain admixture. Therefore Hui was excluded from the further analysis.

	Miao-1	Li-1	Hui	Japanese	English
Miao-1	0.1637 (0.0390)				
Li-1	0.0021 (0.0011)	0.1440 (0.0365)			
Hui	0.0034 (0.0012)	0.0047 (0.0020)	0.1569 (0.0390)		
Japanese	0.0069 (0.0030)	0.0087 (0.0040)	0.0027 (0.0017)	0.1521 (0.0371)	
English	0.0197 (0.0082)	0.0241 (0.0107)	0.0203 (0.0121)	0.0167 (0.0087)	0.1518 (0.0361)

Table 4. Genetic distances among Li-1, Miao-1, Hui, Japanese, and English

Figures below the diagonal are D_{std} and those on the diagonal are gene diversities. Figures in parentheses are standard errors. Allele frequency data for 18 polymorphic loci (Hp, Tf, Gc-subtype, C6, C7, Bf, ACP, ESD, 6PGD, PGM1-subtype, GPT, ADA, AK, ABO, MN, Duffy, Kell, and Lutheran) were used, and data for 18 monomorphic loci were assumed.



Fig. 7. UPGMA dendrogram (A) and NJ network (B) for Li-1, Miao-1, Hui, Japanese, and English. D_{std} distances of Table 4 are used. Open boxes represent 1SE for each branching point of the UPGMA dendrogram. Each line of the NJ network is proportional to branch lengths (distances).

If we compare the genetic distance (0.0069) between Japanese and Miao-1 and that (0.0087) between Japanese and Li-1, they are more or less similar with each other, the average being 0.0078. Thus the divergence time (T_{J-ML}) between Japanese and the ancestral population for Miao-1 and Li-1 is estimated in the following way. Nei and Roychoudhury (1982) estimated the divergence time between Mongoloids (data were mostly taken from Japanese) and English to be 41,000 years, and a more recent estimate for that is 55,000 years (Nei and Ota, 1991). If the genetic distance of 0.0167 between Japanese and English (Table 4) corresponds to these divergence time estimates, that (0.0078) between Japanese and the Miao-Li cluster should correspond to T_{J-ML}. Hence the estimate for T_{J-ML}, under the assumption of the constant evolutionary rate, becomes 19,000–26,000 years, depending on the estimates for divergence time between Mongoloids and English.

DISCUSSION

A possible history of Li, Miao, and Hui populations of Hainan Island

We analyzed the genetic affinity of six human populations living in Hainan Island in the present study. Two Li populations (Li-1 and Li-2) are always clustered in most of the trees, suggesting a close genetic relationship of these two populations. This supports the common origin hypothesis for these two Li populations (Omoto *et al.*, 1993). A preliminary study on the genetic affinity of Hainan Island as well as some ethnic minorities in northern China (Omoto *et al.*, 1994) also supported the close affinity of the two Li populations. Although we studied only two Li subgroups (Qi-Li = Li-1, and Benti-Li = Li-2), it is possible that all the five Li subgroups started to differentiate after their common ancestral population migrated into Hainan Island. According to an archaeological evidence, the earliest record of human occupation in this island is about 3,000 years ago (Yu, 1985). We would like to suggest that the common ancestral population of all the five present day Li subgroups migrated into Hainan Island from the mainland China around that time. This hypothesis can also explain the close genetic affinity of Li with Miao and Zhuang, who probably have lived in southern part of China for a long time.

Miao tribe in Hainan Island are said to be descendants of migrants from Guangxi Province of mainland China as early as in the time of Ming dynasty (Omoto *et al.*, 1993). Miao-1 and Miao-2 populations show a relatively large genetic distance and they do not make a cluster in any tree. Therefore, it is possible that there was more than one migration of Miao populations to Hainan Island as Omoto *et al.* (1993) suggested, and that the original stock of different migrations may be genetically different from each other. However, an alternative possibility can also explain the pattern observed for Miao populations. If the population size of these two Miao populations have been kept rather small, the random genetic drift may have caused a large genetic differentiation within a relatively short time even if they initially originated from the common ancestral population. In fact, the two Miao populations were clustered in an NJ network of Omoto *et al.* (1994) in which data for slightly different set of polymorphic loci were used.

Hui, a moslem group, has a close genetic affinity with Li, Miao, and Zhuang ethnic minorities of southern China (see Figs. 5 and 6). This relationship is different from that of Omoto *et al.* (1994), in which Hui was more closely related with populations of East Asia (Japanese, Mongolian, and Korean). When Hui was compared with Miao-1, Li-1, Japanese, and English for 18 polymorphic loci, however, Hui was shown to have some affinity with Japanese (Table 4 and Fig. 7). Therefore, it is possible that the ancestral population of Hui migrated from a northern part of China to Hainan Island, and gene flow from the surrounding populations of the island took place after the migration. In fact, a majority of Hui populations is distributed in the northwestern part of China. This northern origin

SAITOU et al.

hypothesis for Hui of Hainan Island is consistent with Omoto *et al.*'s (1993) conjecture that malaria was not endemic in Hui's original homeland. However, it is still not clear why Hui of Hainan Island speaks a Chamic language which is spoken by some northern Vietnamese populations.

Relationship of Hainan Island populations with other Asian populations

When six populations of Hainan Island were compared with 12 other populations, those Hainan Island populations showed affinity with Thailanders, Javanese, and Zhuang in both UPGMA and NJ trees (Figs. 5 and 6, respectively). This clustering pattern is consistent with the geographical proximity of the Island with southern part of the mainland China and with Southeast Asia. It should be mentioned, however, that some Southeast Asian populations (Malays and Filipinos) were somewhat distant from the Hainan Island populations and were more closely related with northern East Asian populations (Japanese, Mongolian, and Korean).

There are two studies on the genetic affinity of human populations including some Hainan Island populations. Saitou *et al.* (1992) presented an NJ network for 18 East Asian populations based on the allele frequency data of HLA-A, B, and C loci. A close affinity among Li, Miao (in the mainland China), Zhuang, and Han (in Hainan Island) were observed, and it is consistent with the result of the present study. Imanishi *et al.* (1992) analyzed allele frequency data of HLA-A and B loci for 77 populations over the world, and constructed a UPGMA dendrogram and an NJ network. Li, Miao (in southern mainland China), Thailanders, Vietnamese, Buyi (in southern mainland China), and southern Chinese showed a close affinity in both trees. This is also consistent with our present results.

It can be summarized that Hainan Island populations seem to have a closer genetic affinity with some surrounding populations of southern China and Southeast Asia than those of northern East Asia.

Genetic differentiation between northern and southern Asian populations

Saha and Tay (1992) showed a UPGMA dendrogram of 9 Asian populations based on data for 15 polymorphic loci. There was a clear dichotomy of southern and northern East Asian populations in the dendrogram; Korean, Mongolian, and Japanese were included in the northern cluster, while the southern cluster consisted of Southern Chinese, Zhuang, Malays, and Javanese.

A distinct dichotomy of southern and northern populations of East Asia was also recognized in Saitou *et al.*'s (1992) NJ network. Li and Han of Hainan Island, Miao, Yiao, Zhuang, and Dong of southern mainland China as well as Han of Guangxi and Guangzhou belonged to the southern cluster, whereas Han, Mongolian, Tibetan, Hazakh, and Hui of northern China as well as Japanese and Korean belonged to the northern cluster. Imanishi *et al.*'s (1992) UPGMA dendrogram also showed a

clear distinction between northern and southern populations of East Asia, though this distinction was not clearly observed in their NJ network. Matsumoto (1987) reviewed a distribution of allele frequencies for the Gm locus, and there is a clear difference between northern and southern populations of Asia.

Four northern populations of East Asia (Japanese, Mongolian, Korean, and Ainu) formed a closely related group in the present study, but the variation among the remaining southern populations was much larger than that for those four northern populations (see Figs. 5 and 6). It should be remembered that data for only 10 polymorphic loci were used for constructing those trees, and the reliability of the branching pattern of those trees may not be high.

Cavalli-Sforza *et al.* (1988) constructed a UPGMA dendrogram for 42 populations over the world, and showed that populations of northern East Asia such as Japanese, Ainu, Koreans, Mongolian, and Tibetans are more closely related to Caucasoid populations than to those of Southeast Asian and southern Chinese. Recently, Nei and Roychoudhury (1993) analyzed a different set of allele frequency data for 29 polymorphic loci for 26 human populations, and found that northern and southern Asian populations are more closely related with each other than to Caucasoid both in UPGMA and NJ trees. The reason for this discrepancy between the two studies may be that Cavalli-Sforza *et al.* (1988) used many unshared loci for comparison and only 5 loci were commonly examined for all 42 populations (Nei and Roychoudhury, 1993). The results of the present study consistently showed a distant relationship between English (a Caucasoid) and the remaining Asian populations, which supports the result of Nei and Roychoudhury (1993).

If Japanese and the Li-Miao group can be considered as the representatives of northern and southern East Asian populations, estimate of T_{J-ML} (19,000–26,000 years) corresponds to the divergence time between northern and southern populations. Although it is a rough estimate, it seems reasonable that at least 10,000 years passed after the ancestral populations of the present northern and southern populations of East Asia diverged.

At last, we would like to note that a unique history of Negritos among the Asian populations (Omoto, 1984) is also shown in Figs. 5 and 6, where Negritos was distantly related with all the other East and Southeast Asian populations.

Conclusion

Allele frequency data only for a dozen or so genetic loci were used in the present study, and a definite conclusion is difficult to extract. Because the amount of genetic differentiation is so small among the populations of interest, we need a lot more genetic polymorphism data to elucidate conclusive statements on the genetic affinity of Hainan Island populations. DNA polymorphism such as VNTR may be the tool for a detailed analysis of genetic affinity between closely related human populations in the future.

SAITOU et al.

ACKNOWLEDGEMENTS

The authors thank Drs. K. Aoki and T. Imanishi for their helpful comments. This study was supported by Grants-in-Aid for Scientific Research (International Scientific Research Program) and for Scientific Research on Priority Areas (Prehistoric Mongoloid Dispersals) of Ministry of Education, Science and Culture, Japan.

REFERENCES

- Cavalli-Sforza, L.L., Piazza, A., Menozzi, P., and Mountain, J. (1988) Reconstruction of human evolution: Bringing together genetic, archaeological, and linguistic data. *Proc. Natl. Acad. Sci. USA* 85, 6002–6006.
- Cavalli-Sforza, L.L., and Edwards, A.W.F. (1967) Phylogenetic analysis: Models and estimation procedures. Am. J. Hum. Genet. 19, 233–257.
- Goedde, H.W., Benkmann, H.-G., Kriese, L., Bogdanski, P., Agarwal, D.P., Du, R., Chen, L., Cui, M., Yuan, Y., Xu, J., Li, S., and Wang, Y. (1984) Population genetic studies in three Chinese minorities. Am. J. Phys. Anthropol. 64, 277–284.
- Horai. S. (1976) Genetic polymorphism of human serum factor B (Bf) in Japanese. Jpn. J. Human Genet. 21, 177–186.
- Imanishi, T., Wakisaka, A., and Gojobori, T. (1992) Genetic relationships among various human populations indicated by MHC polymorphisms. In *HLA1991* Vol. 1 (Tsuji, K., Aizawa, M., and Sasazuki, T., eds.), Oxford Univ. Press, Oxford, pp. 627–632.
- Matsumoto, H. (1987) Characteristics of the Mongoloid and neighboring populations on the basis of the genetic markers of immunoglobulins. J. Anthrop. Soc. Nippon 95, 291–304. (In Japanese)
- Nakajima, H. (1973) Unpublished data cited in Akaishi, S., and Kudo, T. (1975) Blood groups. In JIBP Synthesis Vol. 2, Anthropological and Genetic Studies on the Japanese (Watanabe, S., Kondo, S., and Matsunaga, E., eds.), Univ. of Tokyo Press, Tokyo, pp. 77–107.
- Nakamura, S., Ooue, O., and Abe, K. (1984) Genetic polymorphism of the seventh component of complement in a Japanese population. *Hum. Genet.* 66, 279–281.
- Nei, M. (1972) Genetic distance between populations. Am. Nat. 106, 283-292.
- Nei, M. (1987) Molecular Evolutionary Genetics, Columbia Univ. Press, New York.
- Nei, M., and Ota, T. (1991) Evolutionary relationships of human populations at the molecular level. In *Evolution of Life* (Osawa, S., and Honjo, T., eds.), Springer-Verlag, Tokyo, pp. 415–428.
- Nei, M., and Roychoudhury, A.K. (1982) Genetic relationship and evolution of human races. *Evol. Biol.* 14, 1–59.
- Nei, M., and Roychoudhury, A.K. (1993) Evolutionary relationships of human populations on a global scale. *Mol. Biol. Evol.* 10, 927–943.
- Nei, M., Tajima, F., and Tateno, Y. (1983) Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. J. Mol. Evol. 19, 153–170.
- Omoto, K. (1978) Genetic polymorphism of Japanese. In *Anthropology* Vol. 6, *The Japanese II* (Ikeda, J., ed.). Yuzankaku, Tokyo, pp. 217–263. (In Japanese)
- Omoto, K. (1982) Allelic diversity of human populations in the Asian-Pacific area. In *Molecular Evolution, Protein Polymorphism and the Neutral Theory* (Kimura, M., ed.), Japan Scientific Societies Press, Tokyo, pp. 193–214.
- Omoto, K. (1984) The Negritos: Genetic origins and microevolution. *Acta Anthropogenetica* **8**, 137–147.
- Omoto, K., Hirai, M., Harihara, S., Jin, F., Misawa, S., Washio, K., Tokunaga, K., Saitou, N.,

Yamazaki, K., Du, R., Hao, L., Yuan, Y., Luo, Z., Xu, J., Hu, J., Wei, X., Li, S., Zhao, H., Zhang, Z., Niu, K., Du, C., and Lu, B. (1994) Population genetic studies on national minorities in China. In *Prehistoric Mongoloid Dispersals* (Akazawa, T., and Szathmary, E., eds.), Oxford Univ. Press, Oxford (in press).

- Omoto, K., Hirai, M., Jin, F., Misawa, S., Washio, K., Tokunaga, K., Yamazaki, K., Saitou, N., Du, C., Lu, B., Luo, Z., Zeng, Z., Xu, J., Zao, H., Zhang, Z., Wei, X., Niu, K., Hao, L., and Du, R. (1993) Population genetic study in Hainan Island, China. I. Distribution of blood genetic markers. *Anthropol. Sci.* 101, 1–24.
- Omoto, K., Misawa, S., Harada, S., Sumpaico, J., Medado, P. M., and Ogonuki, H. (1978) Population genetic studies of the Philippine Negritos. I. A pilot survey of red cell enzyme and serum protein groups. Am. J. Hum. Genet. 30, 190–201.
- Roychoudhury, A. K., and Nei, M. (1988) Human Polymorphic Genes: A World Distribution, Oxford Univ. Press, Oxford.
- Saha, N., and Tay, J.S.H. (1992) Origin of the Koreans: A population genetic study. Am. J. Phys. Anthropol. 88, 27–36.
- Saitou, N., and Nei, M. (1987) The neighbor-joining method: A new method for constructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425.
- Saitou, N., Tokunaga, K., and Omoto, K. (1992) Genetic affinities of human populations. In Soc. Study Hum. Biol. Symp. 33: Isolation, Migration, and Health (Roberts, D.F., Fujiki, N., and Torizuka, K. eds.), Cambridge Univ. Press, Cambridge, pp. 118–129.
- Sokal, R., and Sneath, P.A.P. (1963) *Principles of Numerical Taxonomy*, W.H. Freedman, San Francisco.
- Tokunaga, K., Dewald, G., Omoto, K., and Juji, T. (1986) Family study on the polymorphisms of the sixth and seventh components (C6 and C7) of human complement: Linkage and haplotype analyses. Amer. J. Hum. Genet. 39, 414–419.
- Yu, G.Y. (1985) Hainan Island, Geodesic Publishing Co. of Guandong Province. (In Chinese)
- Yuan, Y., Du, R., Chen, L., Xu, J., Cui, M., Wang, Y., Li, S., Goedde, H.W., Benkmann, H.-G., Kriese, L., and Bogdanski, P. (1984) Distribution of eight blood group systems and ABH secretion in Mongolian, Korean, and Zhuang nationalities in China. *Ann. Hum. Biol.* 11, 377–388.
- Yuasa, I., Ikebuchi, J., Suenaga, K., and Ito, K. (1986) Phosphoglucomutase-1 subtypes: Polymorphic occurrence of PGM1*7+ and geographical variation in Japan. *Hum. Hered.* 36, 233–237.
- Yuasa, I., Suenaga, K., Gotoh, Y., Ito, K., and Yokoyama, N. (1984) Gc types in western Japan: Report of a new variant Gc 1C35. *Hum. Hered.* 34, 174–177.