Isolation, Migration and Health

33rd Symposium Volume of the Society for the Study of Human Biology

EDITED BY

D. F. ROBERTS
Department of Human Genetics
University of Newcastle upon Tyne

N. FUJIKI
Department of Internal Medicine and Medical Genetics
Fukui, Medical School,
Fukui, Japan

AND

K. TORIZUKA
President, Fukui Medical School,
Fukui, Japan

CAMBRIDGE
UNIVERSITY PRESS 1992

12 *Genetic affinities of human populations*

NARUYA SAITOU, KATSUSHI TOKUNAGA AND KEIICHI OMOTO

Introduction

Differentiation of human populations does not necessarily follow the predictions from a simple model of population fission, because gene flow between them may occur after relatively long isolation. This situation is quite different from the phylogenetic tree of different species, where no gene migration is assumed after speciation. Thus for describing the genetic relationship of populations, instead of a dendrogram (rooted tree), which would describe the phylogenetic tree of populations under the assumption of no migration after fission, a network (unrooted tree) of genetic affinity seems to be more appropriate (Figure 12.1).

The neighbour-joining (NJ) method (Saitou & Nei, 1987), in which the principle of minimum evolution is used, may be suitable for constructing genetic affinity networks of populations. The NJ method does not assume constancy of the evolutionary rate, and it has been shown by computer simulation that it is efficient in

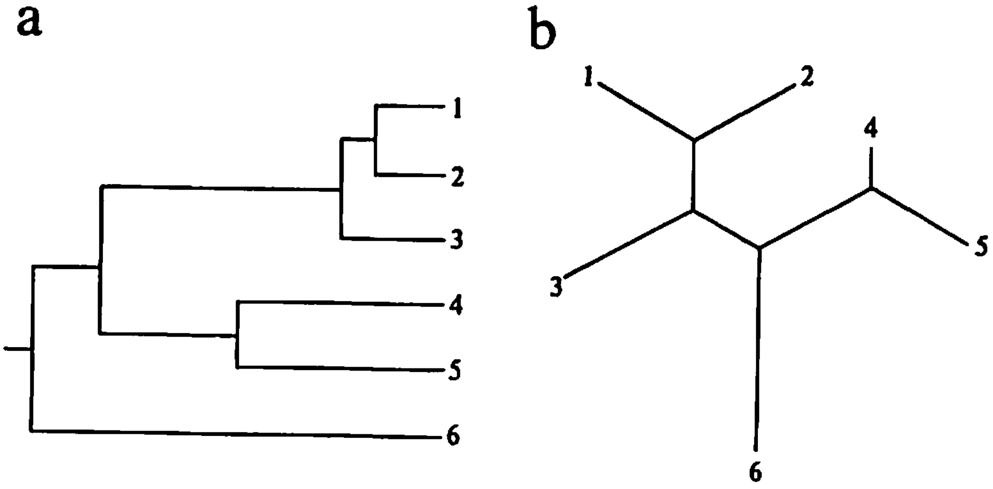

Figure 12.1. A rooted tree (a) and an unrooted tree (b) of 6 populations.

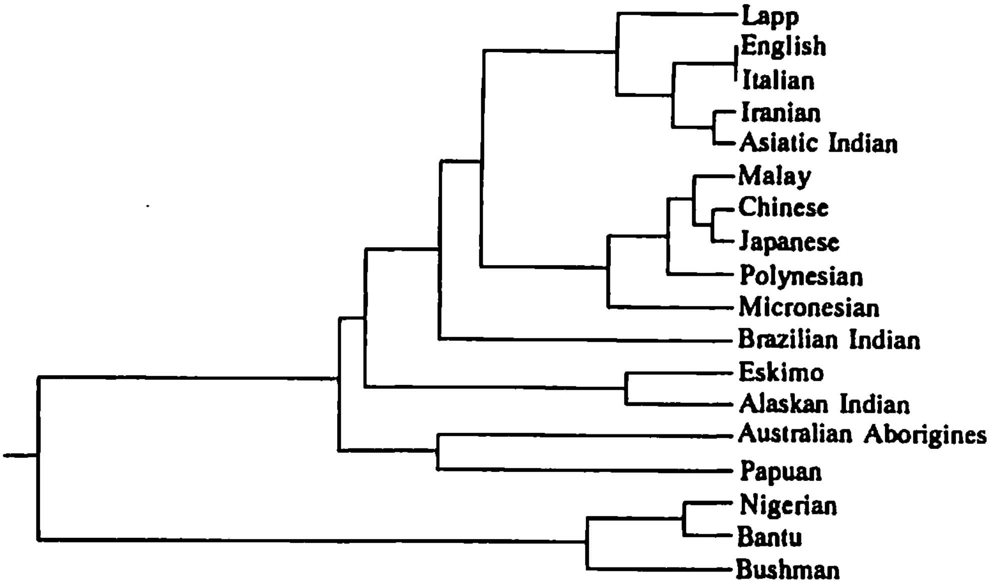

Figure 12.2. A dendrogram for 18 human populations constructed using UPGMA (modified from Nei & Roychoudhury, 1982).

reconstructing the true phylogenetic trees (Saitou & Nei, 1987; Sourdis & Nei, 1988; Saitou & Imanishi, 1989). While in these studies the evolution of nucleotide sequences was simulated, it is likely that the NJ method is also efficient in reconstructing trees from genetic distance matrices based on allele frequency data. In the present study this method is applied to three sets of genetic distance data and the resulting affinity networks are compared with those obtained by other tree-making methods.

Genetic distance data on 18 human populations

Nei and Roychoudhury (1982) compiled allele frequency data on world populations and computed Nei's (1972) genetic distances among 18 human populations based on 23 genetic loci (their Table XI). They constructed a dendrogram of these populations (Figure 12.2) by using UPGMA (Sokal & Sneath, 1963). In this dendrogram, three subSaharan African populations (Nigerian, Bantu and Bushman) stand apart from the remaining populations, and Caucasoid (Lapp, English, Italian, Iranian and Asiatic Indian) and Asian Mongoloid populations (Malay, Chinese, Japanese, Polynesian and Micronesian) each constitute a monophyletic group. However, Amerind populations (Eskimo, Alaskan Indian and Brazilian Indian), who are considered to be genetically close to Asian Mongoloids, are located outside the Caucasoid-Asian Mongoloid cluster. Nei and Roychoudhury (1982) attributed this

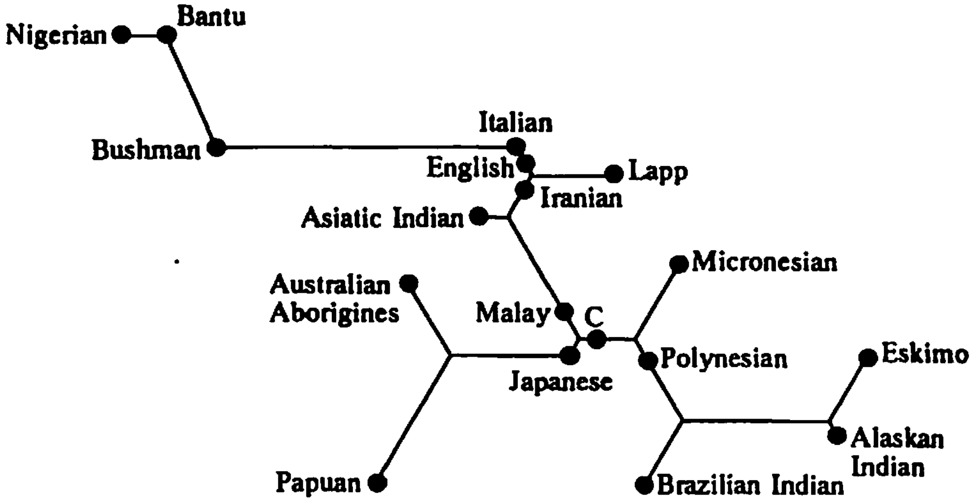

Figure 12.3. An affinity network for 18 human populations constructed using the Wagner distance method (modified from Nei & Saitou, 1986). C = Chinese.

unusual Amerind location to the inbreeding of Amerind populations. Australoid populations (Aborigines and Papuan) were also shown to be genetically similar to Asian Mongoloids (Omoto, 1982). An essentially similar dendrogram was obtained when allele frequency data for the HLA-A and B loci were added (Ryman *et al.*, 1983).

Nei and Saitou (1986) applied the Wagner distance method (Farris, 1972) to the genetic distance matrix of Nei and Roychoudhury (1982). The affinity network they obtained is modified in Figure 12.3, in that negative branch lengths are converted to positive ones as follows: branches of negative length are omitted, and those with positive lengths are drawn proportional to their lengths (=genetic distances). This rule also applies to the following figures. There are marked differences in clustering of populations between the dendrogram (Figure 12.2) and the network (Figure 12.3). Although the three African populations remain monophyletic, neither Caucasoid nor Asian Mongoloid populations are monophyletic. On the other hand, Amerind populations are now monophyletic, and this cluster as well as the Australoid cluster are closer to Asian Mongoloid than to Caucasoid, which is located between African and the other populations. In the following, Asian Mongoloid, Amerind and Australoid as a whole will be called 'Pan-Mongoloid'.

The NJ method was applied to the genetic distance matrix of Nei and Roychoudhury (1982), and the network obtained is shown

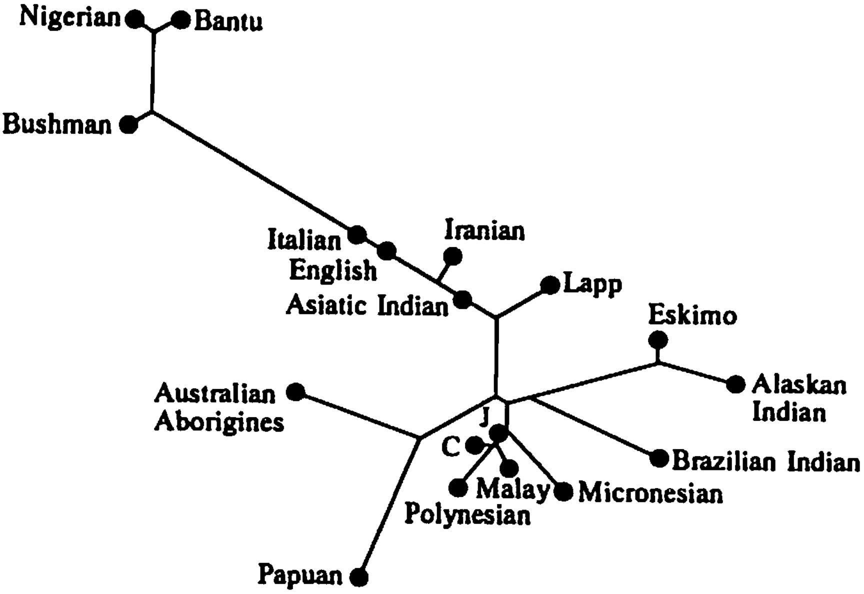

Figure 12.4. An affinity network for 18 human populations constructed using the NJ method (distance matrix data from Nei & Roychoudhury, 1982). C = Chinese, J = Japanese.

in Figure 12.4. In its general features the clustering pattern is similar to that of Figure 12.3; African populations are monophyletic and are far from the remainder, Caucasoid populations are located between African and Pan-Mongoloid, and Amerind populations are monophyletic. Because both the Wagner distance and the neighbour-joining (NJ) methods are intended to produce minimal evolution networks, these similarities are expected. However, there are some important differences. Asian Mongoloid populations are tightly clustered to become a monophyletic group in the NJ network, and the Lapp, instead of the Asiatic Indian, is located between the other Caucasoid populations and the Pan-Mongoloid cluster. Some branch lengths also show differences between Figures 12.3 and 12.4.

Li's (1981) method and modified Farris method (Tateno *et al.*, 1982) were also applied to the same distance matrix data (Figures 12.5 and 12.6, respectively). A UPGMA dendrogram was first constructed in Li's (1981) method, and the dendrogram was then modified by distance transformation. Therefore, the position of the root of Figure 12.5 is identical with that of the UPGMA dendrogram, but Figure 12.5 was drawn as if there were no root. Tateno *et al.*'s (1982) method is a modification of Farris' (1972) Wagner distance method. Although these two trees (Figures 12.5 and 12.6) seem generally to resemble Figures 12.3 and 12.4, there

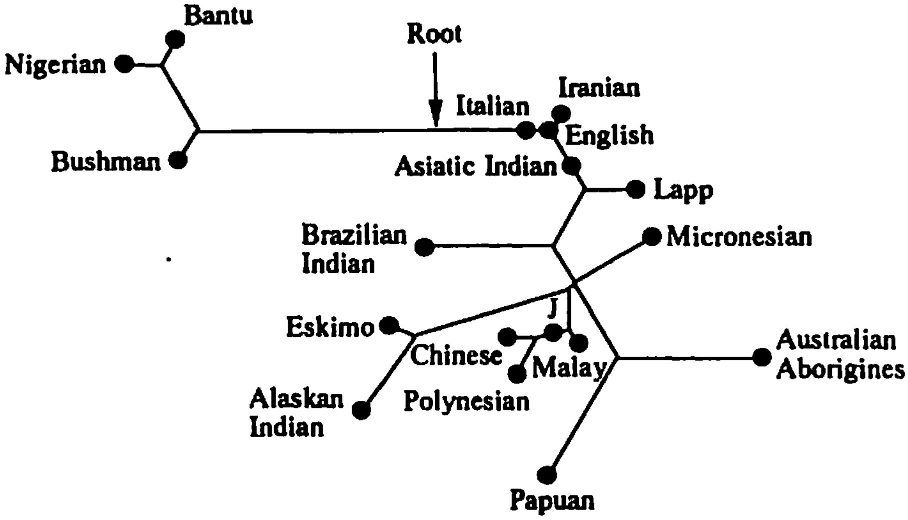

Figure 12.5. An affinity network for 18 human populations constructed using Li's method (distance matrix data from Nei & Roychoudhury, 1982). J = Japanese.

are some conspicuous differences in the Pan-Mongoloid cluster, especially in the location of the Brazilian Indian population. This is located between Caucasoid and the remaining Pan-Mongoloid populations in Figure 12.5, whereas it becomes a part of the Asian Mongoloid cluster in Figure 12.6. Monophyletic grouping of Amerind poulations (Brazilian Indian, Eskimo, and Alaskan Indian) observed in Figures 12.3 and 12.4 seems to be more reasonable if the geographical distribution of these populations is considered.

Some of the clustering features are shared by all of the five figures (one dendrogram, Figure 12.2, and four networks, Figures 12.3-6) produced from the same genetic distance matrix. These are the African cluster (Bantu, Bushman, and Nigerian), the Australoid cluster (Australian Aborigines and Papuan), and the North Amerind cluster (Eskimo and Alaskan Indian). These concordant clusterings seem reasonable in view of the geographical distribution of the populations. Because the amount of migration between a pair of populations is expected to be inversely related to the geographical distance between them, the geographical proximity of populations seems to be a good indicator of their genetic affinity.

If the geographical proximity of populations is used as a criterion for comparing different networks, the Amerind cluster in Figure 12.3 (Wagner distance network) and Figure 12.4 (NJ network), and the Asian-Mongoloid cluster observed in Figure 12.2

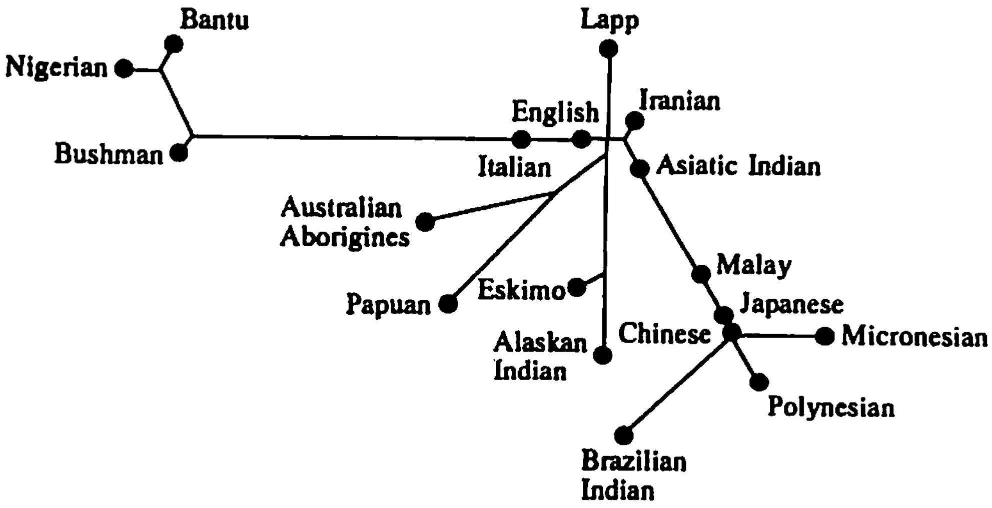

Figure 12.6. An affinity network for 18 human populations constructed using modified Farris method (distance matrix data from Nei & Roychoudhury, 1982).

(UPGMA dendrogram) and Figure 12.4, may represent the true picture. The Caucasoid populations constitute a monophyletic cluster in the UPGMA dendrograms, and they are located between the African cluster and the Pan-Mongoloid cluster in all four networks (Figures 12.3-6). This suggests that Caucasoid populations have experienced gene migrations from Negroid and Mongoloid populations. This possibility was also suggested by Nei and Livshits (1989) who studied the relationship between three major groups of humans.

Genetic distance data on 30 populations from HLA loci

The second set of data analysed is taken from Wakisaka *et al.* (1986) who estimated Nei's (1972) genetic distances for 30 populations based on the HLA allele frequencies, using at locus A 16 antigens, B 32 antigens, C 6 antigens, DR 10 antigens, DRw52/53 2 antigens, and DQ 3 antigens. Wakisaka *et al.* (1986) constructed a UPGMA dendrogram and a modified Farris network based on the distance matrix. A network constructed using the NJ method from the same distance matrix is shown in Figure 12.7. As in the case of the previous analysis, there are marked differences between the UPGMA dendrogram and the two networks. For example, the American Black population is known to be derived from admixture between African Black and European; its intermediate position is realised in the two networks

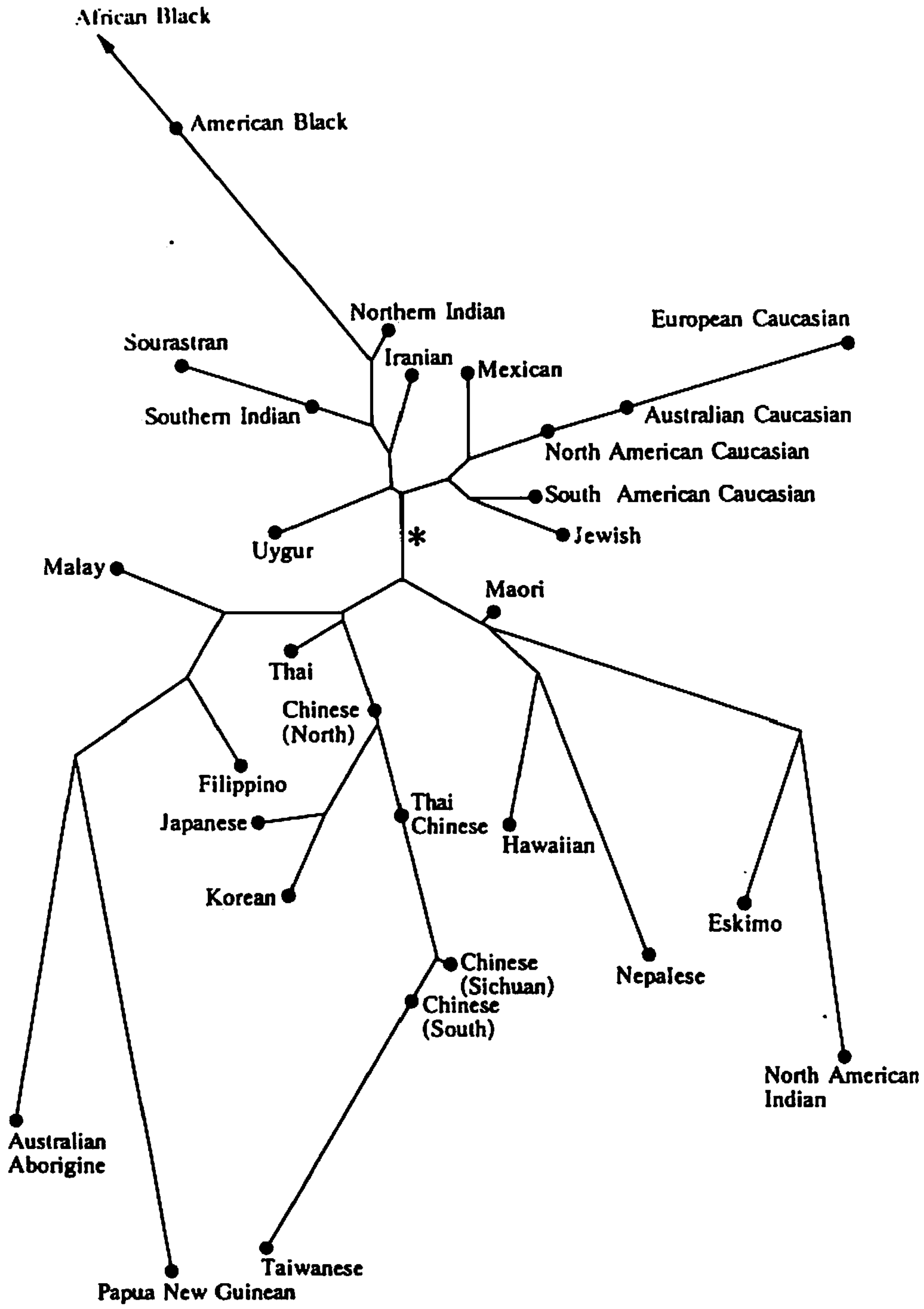

Figure 12.7. An affinity network for 30 populations constructed using the NJ method (distance matrix data from Wakisaka *et al.*, 1986).

(e.g. Figure 12.7), while it clusters with Caucasian populations in the UPGMA dendrogram (Figure 2 of Wakisaka *et al.*, 1986). The position of the Amerind cluster (Eskimo and North American Indian) is also problematic; they are outside the cluster containing Caucasian, Asian Mongoloids, and Polynesian. The situation is similar to the UPGMA dendrogram (Figure 12.2) based on the 23 genetic loci, as discussed above.

Although the two networks are more similar to each other than to the UPGMA dendrogram, there are many differences between the network constructed by the modified Farris method and that constructed by the NJ method. In the NJ network of Figure 12.7, Pan-Mongoloid populations are monophyletic, being separated from the remaining populations by the internal branch denoted by an asterisk. On the other hand, Caucasian populations are located between the Pan-Mongoloid and the African clusters. This is consistent with the NJ network of Figure 12.4 for a different set of data. A similar analysis for mitochondrial DNA data (Saitou & Harihara, 1992) also shows this tendency. However, the Australoid cluster (Australian Aborigine and Papua New Guinean), which is a subcluster of the Pan-Mongoloid cluster in the NJ network, is outside the Caucasian-Mongoloid cluster in the modified Farris network. The position of Japanese also differs in the two networks. In the NJ network (Figure 12.7), Japanese and Korean group together, as a subcluster of one containing various Chinese populations, while in the modified Farris network the Japanese are located with the Amerind cluster. In summary, the clustering of populations in the NJ network seems to reflect the geographical locations of these populations more closely than those in the modified Farris network or in the UPGMA dendrogram. Interestingly, this conclusion is the same as that of the previous analysis.

There is, however, a noteworthy inconsistency between the NJ network and geographical location in the clustering of populations. Hawaii and Nepal are geographically far apart, but the populations living there form a cluster in the NJ network (Figure 12.7). This genetic proximity is also observed in the modified Farris network, and the Polynesian group containing Hawaiian and Maori clusters with Nepalese in the UPGMA tree (Wakisaka *et al.*, 1986). The genetic distance matrix shows Hawaiian and Malay to be equally close, indeed the closest, to Nepalese. Because these genetic distances were computed from allele frequency data at six HLA loci that are tightly linked on chromosome 6 and not independent of each other, the effect of random genetic drift, or any other differentiating process, may be

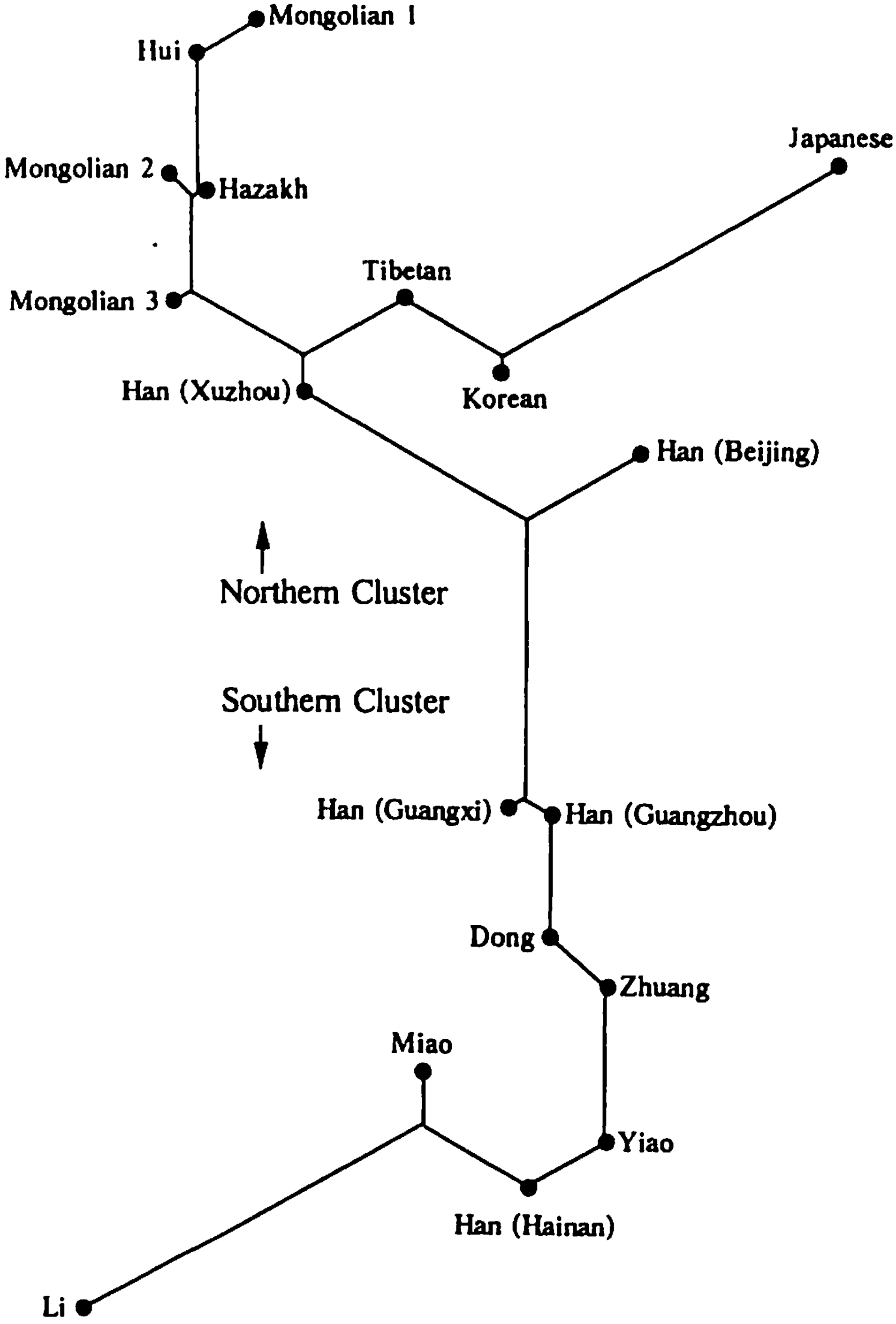

Figure 12.8. An affinity network for 18 east Asian populations constructed by using the NJ method (data from Sun *et al.*, 1986; Aizawa *et al.*, 1986).

exaggerated and apparently larger than that for a set of six unlinked loci. If this is the case, peoples who diverged long ago may show quite similar allele frequencies only by chance. The seemingly close affinity between Hawaiian and Nepalese populations may be so caused.

Genetic distance data on 18 East Asian populations from HLA loci

Allele frequency data were also analysed for the HLA-A, B and C loci of Sun *et al.* (1986) for 16 Chinese populations as well as those of Japanese and Korean. The data for these last two populations were taken from Aizawa *et al.* (1986). Nei's (1972) genetic distances were computed based on the combined allele frequency data (not shown), and an affinity network was constructed using the NJ method (Figure 12.8).

Two clusters can be recognised corresponding to southern and northern populations of East Asia. In the southern cluster, Li and Miao populations are ethnic minorities of Hainan Island, and Dong, Zhuang, and Yiao populations are also ethnic minorities of the southern part of mainland China. Han populations living in the southern provinces of Hainan, Guangxi, and Guangzhou also belong to the southern cluster, while Han of Beijing and Xuzhou belong to the northern cluster. The northern cluster also includes three Mongolian populations, Hui, Hazakh, and Tibetan. Interestingly, the Korean population is closest to the Japanese, and they also belong to the northern cluster. A similar pattern was reported by Omoto *et al.* (1989) who constructed an NJ network based on the allele frequency data of blood group, serum protein and red cell enzyme loci for several Chinese populations and Japanese. Extensive data on Gm haplotype frequencies (Matsumoto, 1987) also show marked differences between northern and southern Chinese populations, and Japanese and Koreans were closer to the northern Chinese populations. These results suggest that significant proportions of the present populations of Japan and Korea are the descendants of migrants from northern Chinese or their related populations. The good correlation between geographical proximity and the genetic affinity of populations is also observed in this set of data.

Conclusion

For describing the genetic relationship of populations, a network (unrooted tree) of genetic affinity seems to be more appropriate than a dendrogram (rooted tree). The neighbour-joining method, in which the principle of minimum evolution is used for constructing networks, appears suitable for this purpose. This method was applied to three sets of genetic distance data and the resulting affinity networks were compared with those obtained by other methods. It was shown that the genetic affinity networks generally reflect the geographical location, isolation and migration of human populations.

References

Aizawa, M., Natori, T., Wakisaka, A. & Konoeda, Y. (1986). *HLA in Asia-Oceania 1986*. Sapporo: Hokkaido University Press.

Farris, J. S. (1972). Estimating phylogenetic trees from distance matrices. *American Naturalist*, **106**, 645–668.

Li, W. -H. (1981). A simple method for constructing phylogenetic trees from distance matrices. *Proceedings of the National Academy of Sciences*, USA, **78**, 1085–1089.

Matsumoto, H. (1987). Characteristics of the Mongoloid and neighboring populations on the basis of the genetic markers of immunoglobulins (in Japanese). *Journal of the Anthropological Society of Nippon*, **95**, 305–324.

Nei, M. (1972). Genetic distance between populations. *American Naturalist*, **106**, 283–292.

Nei, M. & Livshits, G. (1989). Genetic relationships of Europeans, Asians and Africans and the origin of modern *Homo sapiens*. *Human Heredity*, **39**, 276–281.

Nei, M. & Roychoudhury, A. K. (1982). Genetic relationship and evolution of human races. *Evolutionary Biology*, **14**, 1–59.

Nei, M. & Saitou, N. (1986). Genetic relationship of human populations and ethnic differences in relation to drugs and food. In *Ethnic differences in reaction to drugs and xenobiotics*, ed. W. Kalow, H. W. Goedde & D. P. Agarwal, pp. 21–37. New York: Alan R. Liss, inc.

Omoto, K. (1982). Allelic diversity of human populations in the Asia-Pacific area. In *Molecular evolution, protein polymorphism, and the neutral theory*, ed. M. Kimura, pp. 193–214. Tokyo: Japan Scientific Societies Press.

Omoto, K., Saitou, N., Misawa, S., Yamazaki, K., Du, R., Zu, J., Zhao, H., Zhang, Z., Wei, X., Niu, K., Hao, L., Du, C. & He, X. (1989). Japanese-Chinese joint study on national minorities in Hainan Island (second study). I. Blood genetic markers. *Journal of the Anthropological Society of Nippon*, (Abstract), **97**, 292.

Ryman, N., Chakraborty, R. & Nei, M. (1983). Differences in the relative distribution of human gene diversity between electrophoretic and red and white cell antigen loci. *Human Heredity*, **33**, 93–102.

Saitou, N. & Harihara, S. (1992). Gene phylogeny and population phylogeny reconstructed from human mitochondrial DNA data. In *Human evolution in the pacific region*, ed. C. K. Ho, E. Kranz & M. Stoneking. Washington; Washington State University Press.

Saitou, N. & Imanishi, T. (1989). Relative efficiences of the Fitch-Margoliash, maximum-parsimony, maximum-likelihood, minimum-evolution, and neighbor-joining methods of phylogenetic tree construction in obtaining the correct tree. *Molecular Biological Evolution*, **6**, 514–525.

Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biological Evolution*, **4**, 406–425.

Sokal, R. & Sneath, P. H. P. (1963). *Principles of numerical taxonomy*. San Francisco: W -H Freeman, Co.

Sourdis, J. & Nei, M. (1988). Relative efficiencies of the maximum parsimony and distance-matrix methods in obtaining the correct phylogenetic tree. *Molecular Biological Evolution*, **5**, 298–311.

Sun, Y., Lee, J., Gao, X., An, J., Li, S., Song, C., Shi, Q. & Li, Z. (1986). HLA antigens in Chinese populations. In *HLA in Asia-Oceania 1986*, ed. M. Aizawa, T. Natori, A. Wakisaka & Y. Konoeda, pp. 502–510. Sapporo: Hokkaido University Press.

Tateno, Y., Nei, M. & Tajima, F. (1982). Accuracy of estimated phylogenetic trees from molecular data. I. Distantly related species. *Journal of Molecular Evolution*, 18, 387–404.

Wakisaka, A., Hawkin, S., Konoeda, Y., Takada, A. & Aizawa, M. (1986). Anthropological study using HLA antigen frequencies as a genetic marker. In *HLA in Asia-Oceania 1986* ed. M. Aizawa, T. Natori, W. Wakisoka & Konoeda, pp. 197–208. Sapporo: Hokkaido University Press.