Genetic Structure of a 2,500-Year-Old Human Population in China and Its Spatiotemporal Changes

Li Wang,† Hiroki Oota,* Naruya Saitou,‡ Feng Jin,§ Takayuki Matsushita,‖ and Shintaroh Ueda**

*Department of Biological Sciences, Graduate School of Science, University of Tokyo, Tokyo, Japan; †Department of Medical and Animal Genetics, Institute of Genetics, Chinese Academy of Sciences, Beijing, China; ‡Laboratory of Evolutionary Genetics, National Institute of Genetics, Mishima, Japan; §Laboratory of Human Genetics, Institute of Genetics, Chinese Academy of Sciences, Beijing, China; and ‖Doigahama Site Anthropological Museum, Houhoku, Japan

To examine temporal changes in population genetic structure, we compared the mitochondrial DNA (mtDNA) sequences of three populations that lived in the same location, Linzi, China, in different periods: 2,500 years ago (the Spring–Autumn era), 2,000 years ago (the Han era), and the present day. Two indices were used to compare the genetic differences: the frequency distributions of the radiating haplotype groups and the genetic distances among the populations. The results indicate that the genetic backgrounds of the three populations are distinct from each other. Inconsistent with the geographical distribution, the 2,500-year-old Linzi population showed greater genetic similarity to present-day European populations than to present-day east Asian populations. The 2,000-year-old Linzi population had features that were intermediate between the present-day European/2,500-year-old Linzi populations and the present-day east Asian populations. These relationships suggest the occurrence of drastic spatiotemporal changes in the genetic structure of Chinese people during the past 2,500 years.

Introduction

Molecular studies have identified genetic relationships among the present-day human populations. However, population history is still in controversy; migration and admixture in the past make it difficult to interpret these relations in a straightforward manner when examining the origin of modern humans and their expansion across the earth. Genetic research on ancient populations is therefore absolutely requisite for disclosing past events and investigating population history. Direct analysis of the Neanderthal DNA has supported the dispersal of modern humans out of Africa, showing its mtDNA sequence to be outside the variation of present-day human populations (Krings et al. 1997).

Here we investigated temporal changes in genetic structure of human populations during the past 2,500 years in China using mtDNA sequences. All samples were collected in Linzi at the lower reaches of the Yellow River in China. In a previous study (Oota et al. 1999*b*), we examined DNA from the human remains excavated from the 2,000-year-old site of Linzi. In the present study, we extracted DNAs from the human remains excavated from the 2,500-year-old site of Linzi and the present-day Han Chinese living in Linzi. We determined nucleotide sequences of their mitochondrial D-loop regions and compared genetic structures of the three populations that lived in the same location in different periods by evaluating the frequency distributions of the radiating haplotype groups and the genetic distances among the populations, including the present-day Eurasian populations.

Key words: ancient DNA, mitochondrial DNA, Chinese genetic diversity population structure.

Address for correspondence and reprints: Shintaroh Ueda, Department of Biological Sciences, Graduate School of Science, University of Tokyo, Tokyo 113-0033, Japan. E-mail: sueda@biol.s.u-tokyo.ac.jp.

Mol. Biol. Evol. 17(9):1396–1400. 2000

Materials and Methods

Sample Collection

All three populations were sampled in Linzi, which is now part of the city of Zibo in Shandong Province of China. In a previous study (Oota et al. 1999*b*), we examined DNA from human remains found at the 2,000-year-old Yixi site of Linzi. In the present study, DNA was extracted from bone samples of 63 individuals collected at the 2,500-year-old Liangchun site of Linzi, after receiving authorization from the Cultural Relics Bureau of the People's Republic of China. These sites have been dated to the Spring–Autumn (770 B.C.–403 B.C.) and the Han era (206 B.C.–A.D. 220) in China, respectively, on the basis of the archaeological finds (earthenwares) excavated. To examine the present-day population, we collected blood samples from 50 Han Chinese individuals living in Linzi whose parents were both born there.

Ancient DNA Extraction and Amplification

The bone samples were exposed to ultraviolet radiation to destroy possible contaminating DNA on their outer surfaces. Then, the DNA was extracted from the spongy layer of the bone as previously described (Kurosaki, Matsushita, and Ueda 1993; Oota et al. 1995, 1999*a*, 1999*b*). PCR amplification was carried out in 40 μl of a reaction mixture containing 10 mM Tris-HCl (pH 8.3), 2 mM $MgCl_2$, 50 mM KCl, 200 μM each of dNTP, 20 pmol of each primer, and 2 U of *Taq* polymerase (Ampli-Taq Gold, PE Biosystems). PCR began with at 95°C for 9 min, followed by 40 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 1 min. To verify the reliability of the experiments, negative controls containing all of the reagents but without remains/DNA were included in each DNA extraction/PCR run. PCR products purified on spin columns (S-300, Pharmacia) were directly sequenced for both strands using a commercial

kit (FS *Taq* DyeDeoxy Terminator Cycle Sequencing Kit, PE Biosystems).

Phylogenetic Analysis

The phylogenetic network of haplotypes and the population tree were constructed using the network construction (Bandelt et al. 1995) and the neighbor-joining (Saitou and Nei 1987) methods, respectively. In addition to the nucleotide sequences determined in this study, we used mtDNA sequence data from four central Asian (Comas et al. 1998) and five European (Richards et al. 1996) present-day populations and the data from our previous paper (Oota et al. 1999*b*). Nucleotide diversity (the mean of pairwise nucleotide differences per site) and evolutionary distances (net values of nucleotide substitutions) between populations (Nei 1987) were calculated using the program dnapopdist, which was newly developed for this study. The neighbor-joining tree was drawn using the program Dendromaker, version 4.1 (Imanishi 1998), and the midpoint rooting method.

Results and Discussion

Ancient DNA samples are usually damaged and fragmented, and sequences of less than a few hundred nucleotides should be amplified by PCR. Some unknown molecules are difficult to remove and can inhibit PCR when extremely small amounts of DNA extracted from ancient remains are being amplified. Overlapping DNA regions must be sequenced to assess the reliability of the results. Therefore, we analyzed hypervariable region I of the mtDNA by PCR using two sets of primers that amplify the overlapping MT1 (positions 16190–16422) and MT4 (positions 16135–16366) regions.

We successfully amplified and sequenced both of these DNA sequences for 40 of the 63 2,500-year-old remains from the Liangchun site. The nucleotide sequences of both strands in each DNA region were completely complementary in all of them. However, the nucleotide sequences of the overlapping regions of MT1 and MT4 were not identical in 6 of the 40 samples; those six samples were excluded from the subsequent analysis. Since heteroplasmy due to length variation has been reported for mtDNA sequences (Bendall and Sykes 1995), a cytosine tract (positions 16184–16193) was excluded from the comparison, as it was in our previous analysis of DNA from 2,000-year-old human remains (Oota et al. 1999*b*). The nucleotide sequence data reported in this paper will appear in the DDBJ/EMBL/GenBank nucleotide sequence databases with the accession numbers AB031108–AB031191.

In ancient DNA analyses, both the sequence length that can be amplified and the number of samples available for analysis are limited, problems that are not encountered in studies of present-day DNA samples. Therefore, we first constructed a phylogenetic reference network using the 185-bp MT1 and MT4 nucleotide sequences (positions 16194–16378) of 1,298 present-day Asian and circum-Pacific individuals. Then, we traced the genealogy of the mtDNA sequences of the ancient human remains. Based on the phylogenetic network, we identified six radiation groups. These six groups can be characterized by the five nucleotide sites 16217, 16223, 16304, 16319, and 16362 (see Oota et al. 1999*b*; more details will be published elsewhere). Groups I–IV have sequences TTTGT, TTTGC, TTTAC, TCTGT, CCTGT, and TCCGT, respectively. We would like to note that haplotype 1 of figure 1 is the same as the Cambridge reference sequence (CRS) (Anderson et al. 1981). It should also be noted that haplotype 7 and its offspring are not clearly classified into group II according to those five nucleotide sites because of a reticulation (see fig. 1). We did find the same phylogenetic relationship for the longer sequence data as for the 185-bp sequence data.

Based on this present-day reference network, we constructed a network of the mtDNA sequences of the two ancient populations and the present-day people of Linzi (fig. 1). Ten individuals of the 2,500-year-old Linzi population had mtDNA type with 16274A; this mtDNA type was not found in either the 2,000-year-old or the present-day Linzi populations. Sixty-five percent (22 of 34) of the individuals of the 2,500-year-old Linzi population belong to group IV, whereas none of the 2,000-year-old population and only 8% of the present-day Linzi population belong to that group. In contrast, 38% (5 of 13) of the 2,000-year-old Linzi population belong to group VI, compared with only 9% and 10% of the 2,500-year-old and the present-day Linzi populations, respectively. The 2,000-year-old and present-day Linzi populations showed high frequencies for group I (23% and 30%, respectively) and for group II (31% and 36%, respectively). Other present-day east Asian populations, including Mongols, Koreans, and mainland Japanese, also have high frequencies for groups I and II (fig. 2).

Heterogeneity of substitution rate for the human mtDNA D-loop region has been known (e.g., Excoffier and Yang 1999; Meyer, Weiss, and von Haeseler 1999), and sites with high rates, such as 16223, 16311, and 16362, in fact caused reticulations, observed in the phylogenetic network of figure 1. Backward mutations occurring at those high-rate sites may cause erroneous classification of the six groups defined above. However, relative substitution rates of those sites are about four times as high as the average rate (Excoffier and Yang 1999), and the effect of ''contamination'' caused by backward mutations may be relatively small. In any case, frequency estimates of the six groups should be considered as a rough measure for comparing different human populations.

We also estimated the nucleotide diversity within each population and the genetic distances between the populations. We used the 172-bp mtDNA sequence data for comparison due to the lack of 13-bp data (positions 16366–16378) in the present-day European populations. The nucleotide diversities were 0.020, 0.022, and 0.027 for the 2,500-year-old, the 2,000-year-old, and the present-day Linzi populations, respectively. Nucleotide diversities within the other present-day populations were similar (range 0.011–0.027). The smallest genetic distance for the present-day Linzi population was that from the Mongols, followed by those from mainland Japanese

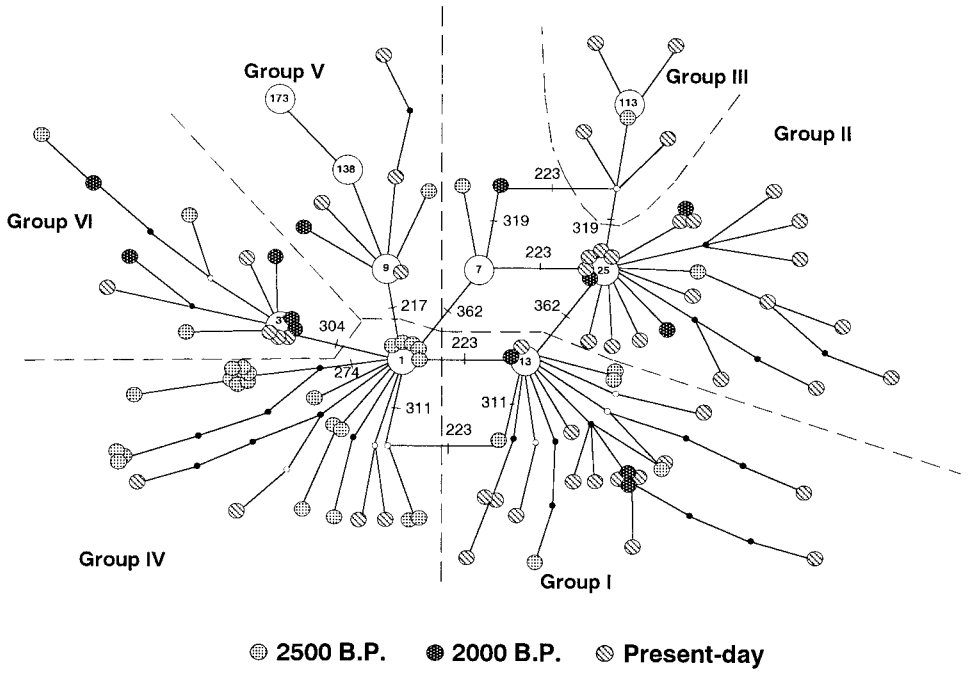

FIG. 1.—A phylogenetic network. Open circles represent distinct mtDNA types found in present-day Asian and circum-Pacific populations, and full circles represent mtDNA types that are absent from those populations. Shaded circles represent mtDNA types seen in the 2,500-year-old to present-day Linzi people in China. Larger circles represent backbone types 1, 3, 7, 9, 13, 25, 113, 138, and 173. The numbers on the branches are the nucleotide positions at which substitutions have occurred (add 16,000 to these numbers to obtain the CRS positions). Reticulations in the network indicate the existence of incompatible nucleotide configurations.

and Koreans. Surprisingly, the three smallest genetic distances for the 2,000-year-old Linzi population were from the present-day central Asian populations: the Kirghiz (Sary-Tash), followed by the Kazakh and the Uighurs. Even more surprisingly, the three smallest genetic distances for the 2,500-year-old Linzi population were from the Turkish, Icelander, and Finnish, rather than from the east Asian populations. The results indicate that the genetic backgrounds of the three populations in Linzi are distinct from each other. Figure 3 shows the phylogenetic tree based on those genetic distances; present-day populations from east Asia, including the present-day Linzi population, form a cluster, which is consistent with their geographical distribution. However, the 2,000-year-old Linzi population lies outside the present-day east Asian cluster, and the 2,500-year-old Linzi popu-

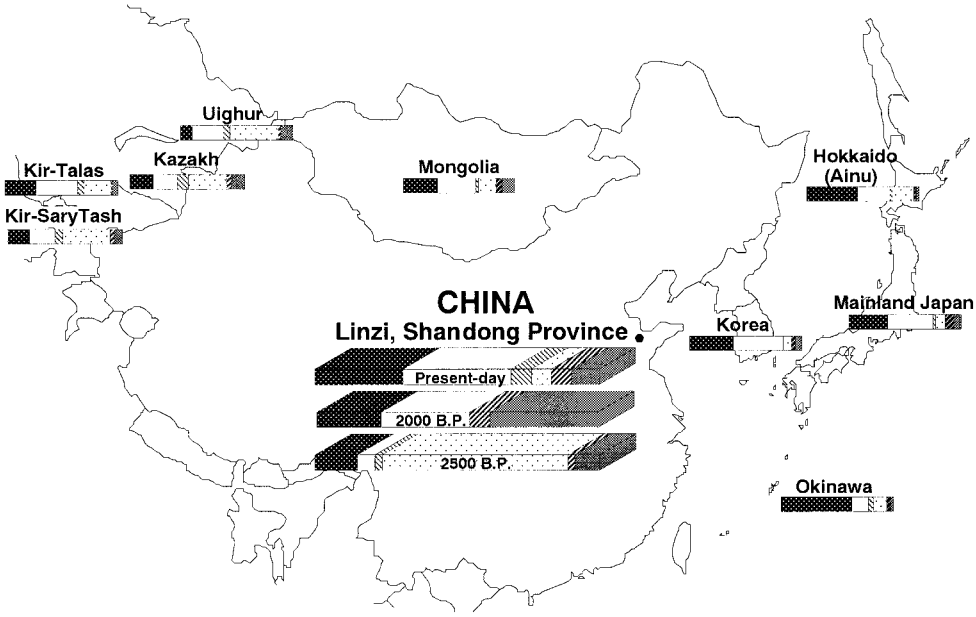

FIG. 2.—Geographic distribution of six radiation groups in the east-central Eurasian continent. Radiation groups I–VI are shown from left to right.

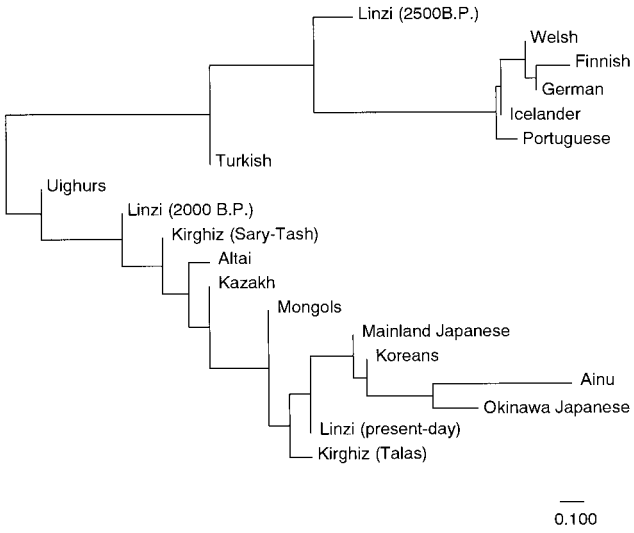

FIG. 3.—A neighbor-joining tree for 19 human populations. Branch lengths are proportional to genetic distances.

lation clusters with the present-day European populations.

The multidimensional scaling method was also applied to the genetic distance matrix data (see fig. 4). The overall constellation of the 19 populations is quite similar to that of the neighbor-joining tree of figure 3. This two-dimensional scattergram explains 96.2% of the variation, and the significant portion is already explained by horizontal axis 1. Because the present-day European populations are on the left side and the present-day east Asian populations are on the right side, this representation can be equated to the geographical map of populations, except for the two ancient populations in China. This strengthens the very odd location of the 2,500-year-old Linzi population.

It should be noted that the population distance analysis is based only on sequence differences within and between populations. Therefore, ambiguity of the phylogenetic relationship of mtDNA haplotypes as represented by a network with reticulations (see fig. 1) does not cause serious problems in the population analyses.

We compared the genetic structures of three populations that lived in the same location during three separate historical periods and found that the genetic structure of the inhabitants of Linzi has changed greatly over time. The period of Chinese history that dates to 2,500 years ago corresponds to the transition period from the Spring–Autumn era to the Warring States era, and the period around 2,000 years ago was in the middle of the Han era. Linzi, our sampling location, was the capital of the feudal state Qi in the Spring–Autumn and the Warring States eras. Qin, one of the feudal states during those periods, conquered other states, including Qi, and established the first unified nation in China. Subsequently, the Han dynasty followed Qin after great disturbances of war. Therefore, our finding that the population structure of Linzi changed drastically during those periods can be concordant with these historical events.

The similarity between the genetic structures of the 2,500-year-old Linzi population and the present-day European populations indicates that there was a genetic shift in the Linzi area from a European-like population to a population more like those found in present-day east Asia, probably caused by migration. This is in accord with the existence of the Eurasiatic superfamily languages, which surround a linguistically unique Sino-Tibetan language, the present-day Chinese language (Ruhlen 1987, 1994; Cavalli-Sforza, Menozzi, and Piazza 1994). Future molecular studies of ancient populations will help us discover the places and times of human diversification and the migration routes of ancient populations.

Acknowledgments

We gratefully acknowledge the Institute of Archaeology, Shandong Province, China, and the Cultural Relics Bureau of the People's Republic of China. We also thank Drs. Kangxin Han, Xunzhang Luo, Y. Manabe and J. Oyamada for encouragement, Dr. O. Kondo for multidimensional scaling analysis, and Drs. J. Bertranpetit and F. Calafell for mtDNA sequences of the central Asians. This study was supported by grants from the Ministry of Education, Science, Sports, and Culture of Japan and the NIG Cooperative Research Program. L.W. was supported by The JSPS Postdoctoral Fellowship for Foreign Researchers in Japan.

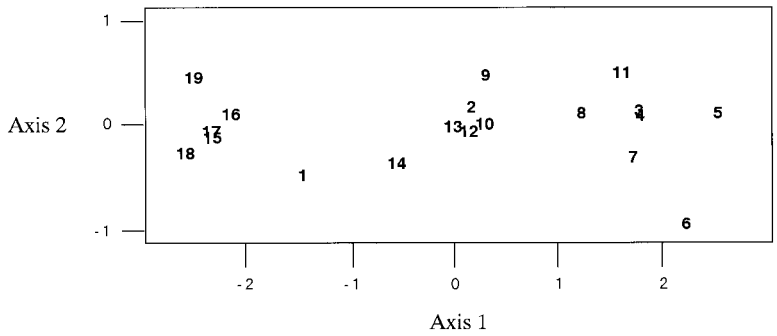

FIG. 4.—The multidimensional scaling of 19 human populations. Populations are designated as follows. 1: the 2,500-year-old Linzi population; 2: the 2,000-year-old Linzi population; 3: the present-day Linzi population; 4: mainland Japanese; 5: Koreans; 6: Ainu; 7: Okinawa Japanese; 8: Mongols; 9: Altai; 10: Kazakh; 11: Kirghiz (Talas); 12: Kirghiz (Sary-Tash); 13: Uighurs; 14: Turkish; 15: Portuguese; 16: Icelander; 17: German; 18: Finnish; 19: Welsh.

LITERATURE CITED

ANDERSON, S., A. T. BANKIER, B. G. BARRELL et al. (14 co-authors). 1981. Sequence and organization of the human mitochondrial genome. Nature **290**:457–465.

BANDELT, H. J., P. FORSTER, B. C. SYKES, and M. B. RICHARDS. 1995. Mitochondrial portraits of human populations using median networks. Genetics **141**:743–753.

BENDALL, K. E., and B. C. SYKES. 1995. Length heteroplasmy in the first hypervariable segment of the human mtDNA control region. Am. J. Hum. Genet. **57**:248–256.

CAVALLI-SFORZA, L. L., P. MENOZZI, and A. PIAZZA. 1994. The history and geography of human genes, Princeton University Press, Princeton, N.J.

COMAS, D., F. CALAFELL, E. MATEU et al. (12 co-authors). 1998. Trading genes along the silk road: mtDNA sequences and the origin of central Asian populations. Am. J. Hum. Genet. **63**:1824–1838.

EXCOFFIER, L., and Z. YANG. 1999. Substitution rate variation among sites in mitochondrial hypervariable region I of humans and chimpanzees. Mol. Biol. Evol. **16**:1357–1368.

IMANISHI, T. 1998. http://www.cib.nig.ac.jp/dda/timanish/dendromaker/home.html.

KRINGS, M., A. STONE, R. W. SCHMITZ, H. KRAINITZKI, M. STONEKING, and S. PÄÄBO. 1997. Neanderthal DNA sequences and the origin of modern humans. Cell **90**:19–30.

KUROSAKI, K., T. MATSUSHITA, and S. UEDA. 1993. Individual DNA identification from ancient human remains. Am. J. Hum. Genet. **53**:638–643.

MEYER, S. G., G. WEISS, and A. VON HAESELER. 1999. Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. Genetics **152**:1103–1110.

NEI, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

OOTA, H., N. SAITOU, K. KUROSAKI, S. POOKAJORN, T. ISHIDA, T. MATSUSHITA, and S. UEDA. 1999*a*. Ancient DNA: a new strategy for studying population history. Pp. 25–41 *in* K. OMOTO, ed. Interdisciplinary perspectives on the origins of the Japanese. International Research Center for Japanese Studies.

OOTA, H., N. SAITOU, T. MATSUSHITA, and S. UEDA. 1995. A genetic study of 2,000-year-old human remains from Japan using mitochondrial DNA sequences. Am. J. Phys. Anthropol. **96**:133–145.

———. 1999*b*. Molecular genetic analysis of remains of a 2,000-year-old human population in China—and its relevance for the origin of the modern Japanese population. Am. J. Hum. Genet. **64**:250–258.

RICHARDS, M., H. CÔRTE-REAL, P. FORSTER, V. MACAULAY, H. WILKINSON-HERBOTS, A. DEMAINE, S. PAPIHA, R. HEDGE, H. BANDELT, and B. SYKES. 1996. Paleolithic and neolithic lineages in the European mitochondrial gene pool. Am. J. Hum. Genet. **59**:185–203.

RUHLEN, M. 1987. A guide to the world's languages, Vol. 1. Classification. Stanford University Press, Stanford, Calif.

———. 1994. The origin of language. John Wiley and Sons, New York.

SAITOU, N., and M. NEI. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.

MASAMI HASEGAWA, reviewing editor

Accepted May 29, 2000