AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 133:1128–1136 (2007)

Tracing the Genetic History of the Chinese People: Mitochondrial DNA Analysis of a Neolithic Population from the Lajia Site

Shi-Zhu Gao,[1,2] Yi-Dai Yang,[1] Yue Xu,[3] Quan-Chao Zhang,[1] Hong Zhu,[1] and Hui Zhou[1,3]*

[1]*Laboratory of Ancient DNA, Research Center for Chinese Frontier Archaeology of Jilin University, Changchun 130012, China*
[2]*College of Pharmacia Sciences, Jilin University, Changchun 130021, China*
[3]*College of Life Science, Jilin University, Changchun 130023, China*

KEY WORDS ancient mtDNA; Lajia Site; Chinese people

ABSTRACT Ancient DNA analysis was conducted on the dental remains of specimens from the Lajia site, dating back 3,800–4,000 years. The Lajia site is located in Minhe county, Qinghai province, in northwestern China. Archaeological studies link Lajia to the late period of the Qijia culture, one of the most important Neolithic civilizations of the upper Yellow River region, the cradle of Chinese civilization. Excavations at the site revealed that the inhabitants died in their houses as the result of a sudden flood. The Lajia site provides a rare chance to study the putative families, all of whom died at the same instant. Possible maternal familial relationships were investigated through mitochondrial DNA (mtDNA) sequence analysis. Twelve sequences from individuals found in one house were assigned to only five haplotypes, consistent with a possible close kinship. Results from analyses of RFLP typing and HVI motifs suggest that the Lajia people belonged to the haplogroups B, C, D, M*, and M10. This study, combined with archaeological and anthropological investigations, provides a better understanding of the genetic history of the Chinese people. Am J Phys Anthropol 133:1128–1136, 2007.

The Lajia site is located in the Guanting basin of the upper Yellow River, in Minhe county, Qinghai province, northwestern China (Fig. 1). According to radiocarbon dating, the Lajia site is 3,800–4,000 years old (Xia et al., 2003). Archaeological studies have associated the Lajia site with the late period of the Qijia culture, a major culture that flourished during the late Neolithic Age to early Bronze Age (Ren et al., 2002) in the Hexi Corridor, a major route leading to the central regions of the Yellow River. The Qijia culture belonged to the cultures of the Di-Qiang, an ancient tribe confederation in northwestern China (Liu, 2003). According to ethnological studies, the Di-Qiang population contributed to the development of the current Han and Tibeto-Burman (TB) speaking populations (Yang and Ding, 2003). In the late Neolithic Age to early Bronze Age period, a branch of the Di-Qiang people migrated eastward and merged with tribes in the middle and lower Yellow River valley, bringing into being one constituent of the Huaxia ethnic group (Yang and Ding, 2003). The Huaxia ethnic group later absorbed many other cultures, including some from the Yangtze River region, to form the famous Huaxia civilization. After a long process of integration and expansion, the Huaxia civilization developed into the Han, which forms the largest ethnic group in present-day China (Tian, 2001; Xu, 2003). In addition, a wave of the Di-Qiang tribes migrated to the southwest of China starting 4,000–5,000 years ago and mixed with several indigenous tribes. The southwest migrants of the Di-Qiang people developed into part of the Tibeto-Burman (TB) speaking populations that are now primarily distributed in the Qinghai, Tibet, Sichuan, Yunnan, and Hunan regions of China (Yang and Ding, 2003).

In 2000, archaeologists working in the northeast part of the Lajia site discovered 16 human remains in two ruined houses, designated F3 and F4. Excavations at the site revealed that the cause of the death of the inhabitants was a sudden earthquake followed by a flood (Xia et al., 2003). The Lajia civilization ceased to exist after this natural disaster, so that the Lajia site is sometimes termed "the eastern Pompeii" by archaeologists (Zhang et al., 2004).

Two of the 16 human remains were found in F3. The skeletons of these two individuals were in close apposition, against the eastern wall. Anthropomorphic analyses suggested that these remains belonged to an approximately 35-year-old female and 3–4-year-old child. It was supposed that the female was the mother of the child and was embracing her child tightly to protect him/her from danger during the disaster.

The 14 skeletons found in F4 were placed into six groups (Fig. 2). One group (Group III) found at the eastern wall also consists of a presumed mother–child pair.

Grant sponsor: Natural Science Foundation of China, NSFC; Grant number: J0530184.

*Correspondence to: Hui Zhou, Laboratory of Ancient DNA, Research Center for Chinese Frontier Archaeology, Jilin University, Jiefang Road 115, Changchun, 130023 China.
E-mail: zhouhui@mail.jlu.edu.cn

Received 10 April 2006; accepted 28 February 2007

DOI 10.1002/ajpa.20623
Published online 16 May 2007 in Wiley InterScience (www.interscience.wiley.com).

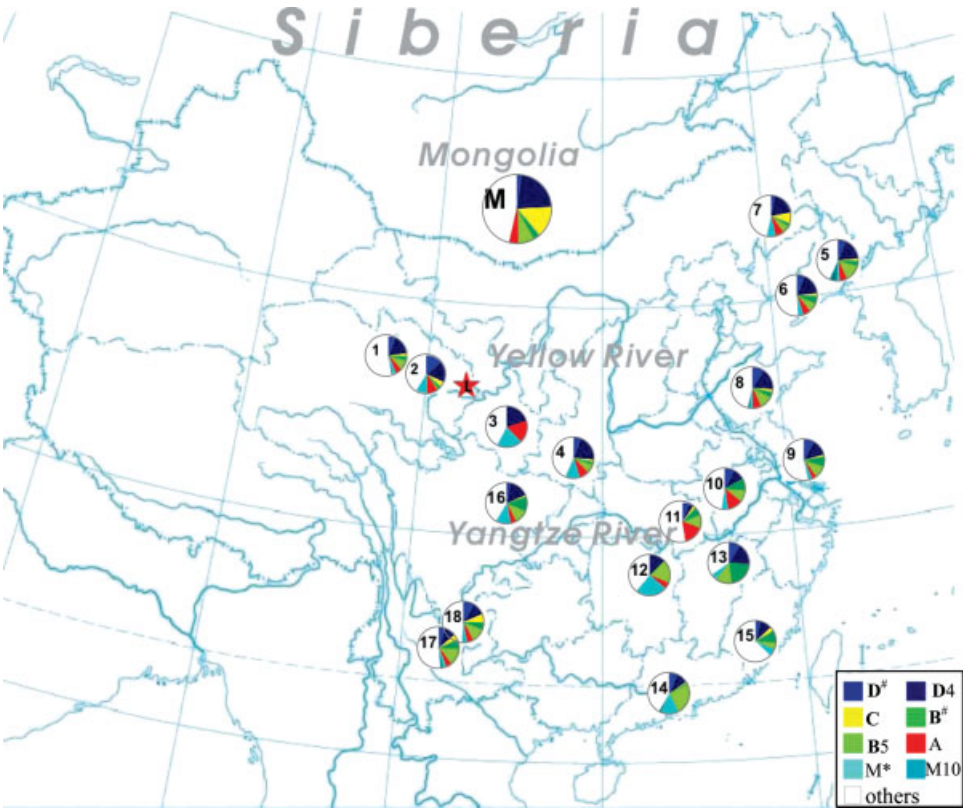

Fig. 1. Geographic location of the Lajia population (red star) and the modern Han populations. Population codes are as reported in table 6. Pie charts show the distribution of the A, B, C, D, and M*/M10 in the Han populations. D# represents the sub-haplogroups of D except D4. B# represents the sub-haplogroups of B except B5. M* includes some unidentified M10 haplotypes. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

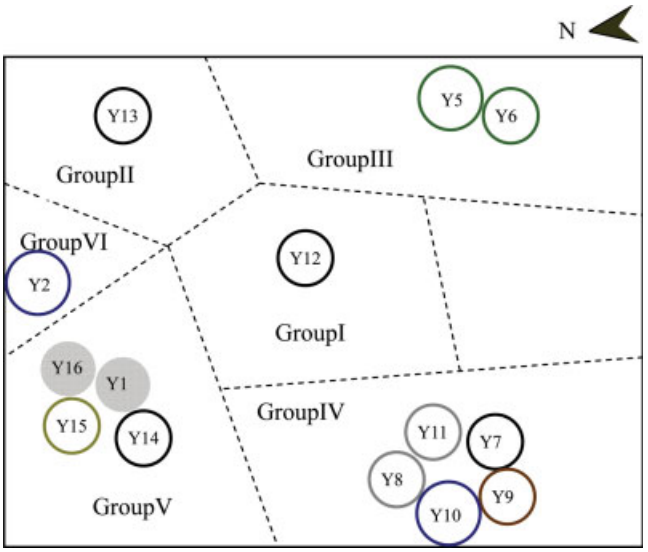

Fig. 2. Spatial distribution of skeletons and assignment into groups for F4. Circles having the same color represent individuals sharing the same mtDNA haplotypes. Greyed-out circles represent samples that could not be amplified. Adult individuals are indicated by larger circles. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Group V and VI are comprised of five skeletons in the north or northwestern corner, and Group IV is comprised of five skeletons, mostly children, at the west wall a southwestern corner. A youth's remains (Group I) were found near the fireplace at the center of the house and a juvenile's skeleton (Group II) was found northeast of the fireplace. All are suspected to be close relatives who died in the flood. The familial relationships within these 16 specimens are of great interest. Unlike most archaeological burial sites that have hundreds of individuals and were occupied often for thousands of years, the Lajia site provides a rare opportunity to study a group of people that lived together and died suddenly at a single point in time.

In the present study, we used mtDNA analysis to investigate the maternal relationships of the ancient inhabitants whose remains were found in F3 and F4 and their relationship to modern populations. The results provide valuable new insights into the familial relationships of the Lajia specimens as well as their relationship to modern Han and TB populations.

MATERIALS AND METHODS

Sample collection

Samples for DNA analysis were collected from the teeth of the 16 human remains found at the Lajia site. The use of teeth as a source of DNA has the advantage that multiple, independent samples can be obtained per individual (O'Rourke et al., 2000). All teeth were sampled from well-preserved molars without caries and cracks to minimize the possibility of modern DNA contamination.

The estimated age and sex of the skeletons is shown in Table 1.

Ancient DNA extraction, amplification, and sequencing

Teeth were first soaked in 10% bleach for 20 min, rinsed with ethanol and distilled water, and then UV-irradiated for 30 min on each side (Ovchinnikov et al., 1999). Each tooth was powdered in a 6750 Freezer Mill (SPEX, USA). DNA was carefully extracted from 500 mg of the tooth powder using the Geneclean for Ancient DNA kit (BIO 101, Vista, CA) according to the manufacturer's protocol. Amplification of the mtDNA was performed on four subregions (a, b, c, and d) of the first hypervariable segment of the control region (HVI) using the primers listed in Table 2.

TABLE 1. Estimated age and sex of the Lajia individuals

Specimen numbers	Sex[a]	Age
House (F3)		
Y3	Female	About 35 years old
Y4	?	3–4 years old
House (F4)		
Y1	?	4–5 years old
Y2	Male	40–45 years old
Y5	Female	28–30 years old
Y6	?	1–2 years old
Y7	Male	14–18 years old
Y8	Male?	11–14 years old
Y9	Male?	10–13 years old
Y10	Female	30–35 years old
Y11	?	7–9 years old
Y12	Male	15–17 years old
Y13	Male?	8–10 years old
Y14	Male?	10–13 years old
Y15	?	6–8 years old
Y16	?	3–4 years old

[a] Male? represents presumed male and ? represents sex unknown.

PCR amplification was carried out in 25 μl of a reaction mixture containing 67 mM Tris–HCl (pH 8.8), 2 mM $MgCl_2$, 50 mM KCl, 500 mM each of dNTPs, 0.5 μM of each primer, 1 U of *Taq* polymerase (Promega) and 1.3 mg/ml BSA. Cycling parameters were 94°C for 10 min, followed by 33 cycles with 94°C for 1 min, 45 s at 52°C for HVIa and HVIb or 45 s at 55°C for HVIc and HVId, elongation for 1 min at 72°C, extension 10 min at 72°C and storage at 4°C. To verify the reliability of the experiments, negative controls containing all of the reagents but without the ancient DNA sample were included in each DNA extraction/PCR run. Amplification products were checked on a 2% agarose gel and purified with QIAquick Gel Extraction Kit (Qiagen). PCR products were sequenced using the ABI 310 Terminator Sequencing kit (PE Applied Biosystems) according to the manufacturer's instructions and were analyzed on the ABI PRISM 3100 automatic sequencer (PE Applied Biosystems). DNA sequences were analyzed using CLUSTAL X 1.83.

RFLP typing

Because control regions of the mtDNA exhibit a relatively high evolutionary rate, unique haplogroups should not be assigned to specimens using only these regions (Richards et al., 1996; Macaulay et al., 1999). Thus, various diagnostic genetic markers were chosen to aid haplogroup determination. Ancient DNA was analyzed by restriction fragment length polymorphism (RFLP) for the microhaplogroup M and its constituent haplogroups C, D, G, and M10. The B haplogroup was identified based on electrophoresis with known size standards. The primers and marker positions used are provided in Table 2. PCR amplifications were performed as described above using an annealing temperature of 52°C. Restriction digestions were carried out following the manufacturer's recommendations (MBI). Fragment length was analyzed using an ABI PRISM 3100 automatic sequencer (PE Applied Biosystems) in accordance with the manufacturer's instructions.

TABLE 2. Primers for sequencing and RFLP

Haplogroup	Primer	RFLP	Length
HVI-a	L16055 5′-GGAAGCAGATTTGGGTAC-3′ H16142 5′-ATGTACTACAGGTGGTCAAG-3′	Sequencing	124
HVI-b	L16131 5′-CACCATGAATATTGTACGGT-3′ H16218 5′-TGTGTGATAGTTGAGGGTTG-3′	Sequencing	126
HVI-c	L16185 5′-ACCCAATCCACATCAAAACC-3′ H16286 5′-TGTACTGTTAAGGGTGGGTAGG-3′	Sequencing	142
HVI-d	L16281 5′-CCTCACCCACTAGGATACCAA-3′ H16379 5′-CAAGGGACCCCTATCTGAGG-3′	Sequencing	138
M	L10286 5′-ATGAGCCCTACAAACAACT-3′ H10441 5′-GAGTCGAAATCATTCGTTT-3′	+ *Alu*I 10397	156
G	L4687 5′-CTATCCTCTTCAACAATATACTCT-3′ H4866 5′-ATGTGAGAAGAAGCAGGC-3′	+ *Hha*I 4831	179
C	L4687 5′-CTATCCTCTTCAACAATATACTCT-3′ H4866 5′-ATGTGAGAAGAAGCAGGC-3′	− *Bse*AI 4715	179
B	L8215 5′-ACAGTTTCATGCCCATCGTC-3′ H8297 5′-ATGCTAAGTTAGCTTTACAG-3′	CoII/$tRNA^{lys}$ 9-bp deletion	121/112
D	L5140 5′-GCACCACGACCCTACTACTA-3′ H5305 5′-GGGATGATGAGGCTATTGT-3′	–*Alu*I 5176	166
M10	L10600 5′-CTACTCTCATAACCCTCAAC-3′ H10762 5′-CATTGGAGTAGGTTTAGG-3′	+*Rsa*I 10646	163
M7	L9787 5′-CTCAACATTTTTTGTAGCCAC-3′ L9947 5′-CCACATCTACAAAATGCCAG-3′	–*Hinf*I 9820	161

TABLE 3. mtDNA HVI sequences of the Lajia haplotypes

		16140	16189	16197	16223	16256	16260	16267	16269	16311	16319	16327	16362	16365
H[a]	CRS[b]	t	t	c	c	c	c	c	t	t	g	c	a	t
1	Y2	·	·	·	t	·	·	·	·	·	·	·	·	c
	Y10	·	·	·	t	·	·	·	·	·	·	·	·	c
2	Y5	·	·	·	t	t	t	·	·	·	·	·	·	·
	Y6	·	·	·	t	t	t	·	·	·	·	·	·	·
3	Y7	·	·	·	t	·	·	·	c	·	·	t	·	·
	Y12	·	·	·	t	·	·	·	c	·	·	t	·	·
	Y13	·	·	·	t	·	·	·	c	·	·	t	·	·
	Y14	·	·	·	t	·	·	·	c	·	·	t	·	·
4	Y8	·	·	·	t	·	·	·	·	c	·	·	·	·
	Y11	·	·	·	t	·	·	·		c	·	·	·	·
5	Y4	c	·	·	t	·	·	·	·	·	·	·	g	·
6	Y15	·	·	·	t	t	·	·	·	c	·	·	g	·
7	Y3	c	c	t	·	·	·	·	·	·	a	·	·	·
8	Y9	·	c	·	·	·	·	t	·	·	·	·	·	·

[a] H indicates the haplotype numbers.
[b] Polymorphic nucleotide sites are numbered according to the report by Andrews et al. (1999) and dots (·) indicate identity with the reference sequence.

Precautions against contamination

Standard contamination precautions cited by O'Rourke et al. (2000) were followed as closely as possible. Pre- and post-PCR experimental areas were separated. Facemasks, gloves, lab coats, and pipettes with aerosol-resistant tips were used. Negative extraction controls and negative PCR controls were employed. The equipment and benches were treated with DNA contamination removal solution (DNA-QUANT OFF^{TM} Q-BIO gene) and ultraviolet irradiation at 254 nm. In order to detect possible exogenous sources of contamination, the mtDNA of the individuals involved in processing the samples were typed and compared to the results. Two individuals were working with the samples, Shi-Zhu Gao and Yi-Dai Yang. Their mtDNA types were 126-294-296-304 (Gao) and 129-223-257A-260-261 (Yang), which belonged to the haplogroup T and N9a, respectively. These haplotypes matched none of the sequences of the ancient samples and only two hypervariable sites were shared between the sequences of these individuals and the Lajia population. In order to limit cross-contamination, the ancient samples were kept out of contact with all other DNA samples during the experimentation. At least three extractions and three amplifications of each extract were carried out for each sample to assess the reproducibility and the authenticity of the results.

DNA analysis

A neighbor-joining tree was constructed using MEGA version 3.1 (Kumar et al., 2004), based on Kimura2-parameter model. In order to show the relationships between the Lajia people and the modern populations, some present populations were included in the phylogenetic analysis. Han (Wen et al., 2004a; Yao et al., 2002a) and TB (Wen et al., 2004b) populations from China, Mongolians (Kolman et al., 1996) from Mongolia, as well as Buryats and Yakuts (Pakendorf et al., 2003) from Siberia were selected for analysis. European population (German) (Richards et al., 1996) was also included to show the East Asian specific haplotypes of the Lajia people.

RESULTS

mtDNA analysis

Reproducible sequences were obtained for 14 of the 16 individual remains. Two DNA samples, Y1 and Y16, could not be successfully amplified. The 325-bp fragments of the mtDNA HVI control region were compared to the revised Cambridge Reference Sequence (Andrews et al., 1999). From the 14 individuals, a total of 8 different sequences defined by 13 variable nucleotide positions were obtained (Table 3). Differences between sequences could be attributed to transitional substitutions and involved mainly the pyrimidines.

The most frequent mtDNA type was scored in four of the fourteen individuals (Y7, Y12, Y13, and Y14). These individuals were found in Groups I, II, IV, and V, respectively (see Fig. 2). Although these individuals may be brothers based on the shared haplotype, Y chromosome data is needed to confirm this point. Three other mtDNA types were shared by individuals Y2 and Y10, Y5 and Y6, and Y8 and Y11, respectively. The two adults (Y2 and Y10) from group IV and group VI may be a brother–sister pair, whereas Y5, a 28–30-year-old women, may be the mother of Y6, whose age is estimated at 1–2 years. The children Y8 and Y11 (from Group IV) may be siblings. The remaining four mtDNA types are each represented by a single individual. The two mtDNA sequences from the individuals (Y3 and Y4) of F3 differ from each other at five nucleotide positions. Although these two subjects may be genetically linked by a mother/son (or daughter) relationship, this result unambiguously excludes kinship through the maternal lineage for Y3 and Y4.

In the neighbor-joining tree, the Lajia sequences were widely scattered among the Han and TB sequences (Fig. 3). Haplotypes 2 and 5, haplotypes 4 and 6, and haplotypes 7 and 8 clustered closely with each other.

Haplogroup assignments

To infer the potential haplogroup status from the short segment of the control region that was sequenced, hap-

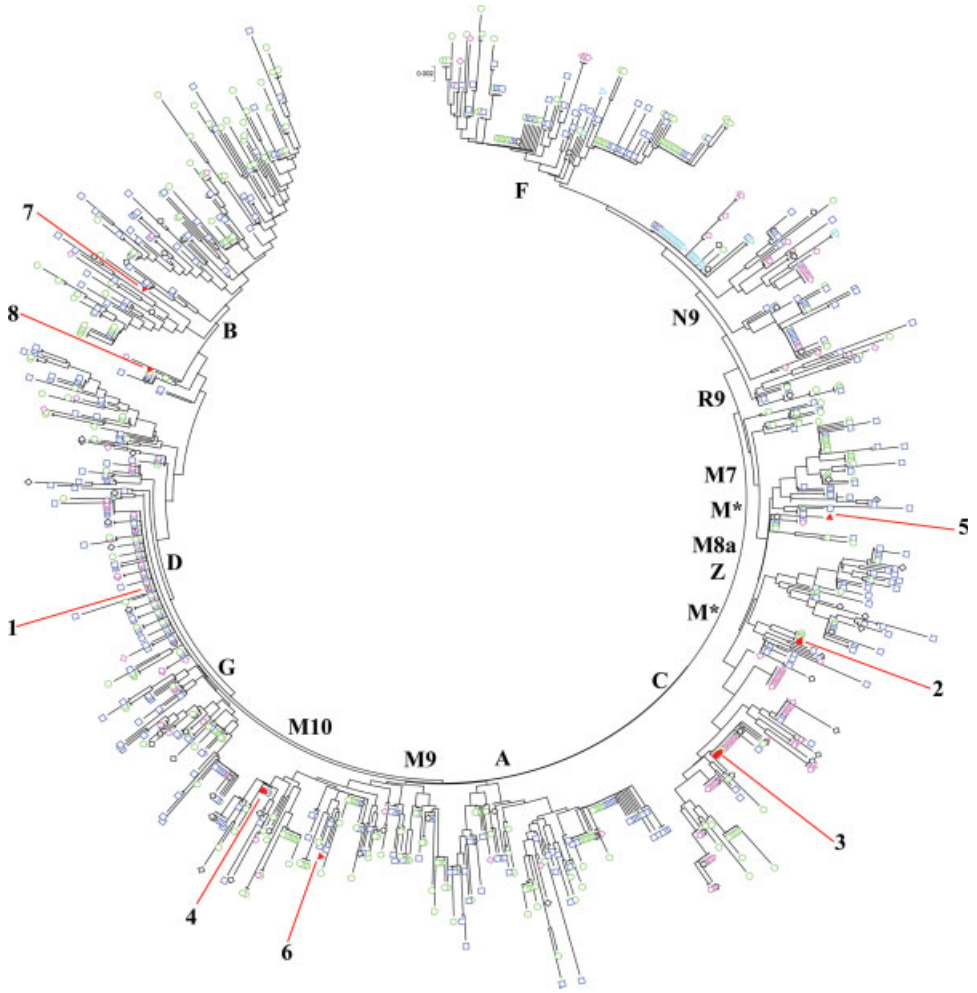

Fig. 3. Neighbor-joining tree of HVI mtDNA sequences showing the Lajia individuals and modern populations. Lajia individuals are indicated by solid triangles with a number corresponding to the haplotype number. The TB populations is represented by circles, the Han population by the squares, the Siberian populations by the light diamond, the Mongolian population by the black diamond and the European population by the light triangles. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

logroup-specific HVI motif searches were performed against mtDNA haplotypes in GenBank (Table 4).

By using the East Asian mtDNA classification tree (Kivisild et al., 2002; Kong et al., 2003; Yao et al., 2002a, 2003), haplogroup assignment was attempted based on HVI sequences (from 16055 to 16379) in combination with RFLP data (Table 5). Haplotype 1 (16223–16362) bears the full motif 16223–16362 for the subclade of haplogroup D and G. It was assigned to super-haplogroup M by 10397*Alu*I+, which was consistent with its HVI motif. The 5140–5305 fragment could not be amplified successfully, so it was not possible to identify the characteristic site (5176*Alu*I−) of haplogroup D. However, the potential haplogroup status of this sequence can be inferred. The *Hha*I− site at 4831 rules out haplogroup G. Haplotype 1 is linked to haplogroup D. According to the diagnostic sites of the subhaplogroups of D, haplotype one can be classified as D4. In D4, haplotype 1 has a high frequency and is widely distributed in East Asian populations (Table 4).

Haplotype 3 shows substitutions at sites 16223 (C/T), 16298 (T/C), and 16327 (C/T), which assign the sequence to haplogroup C. This assignment is also supported by the RFLP results of 10397*Alu*I+ and 4715*Bse*AI−. Haplotype 3 is widely distributed in east and north East Asian, south Siberian, and American Indian populations. Interestingly, the recently identified sub-haplogroup C5 (16189–16298–16327) is a major branch of C in southern East Asian populations such as Zhuang, Dai, Lahu, Miao, and Yao but is absent in northern East Asian populations (Wen et al., 2005). In contrast, haplotype 3 is dominant in the northern East Asians, which is consistent with the distribution of haplogroup C.

Haplotype 7 (16140–16189–16197–16319) and haplotype 8 (16189–16267) are both classified as haplogroup B due to the 9-bp deletion in the CoII/tRNA^{lys} intergenic region. Haplotype 7 might be assigned to B5 by the transition at 16140. Compared with B5b, haplotype 7 also matches a characteristic site of 16319. Haplotype 8 matches neither B5 nor B4 diagnostic sites, and might be an ancient haplotype of B.

Haplotype 4 (16223–16311) and haplotype 6 (16223–16256–16311–16343) are assigned to super-haplogroup M by 10397*Alu*I+. They both have the transition at 16311(T/C) and 10646 (+10646 *Rsa*I). On the phylogenic tree of East Asian mtDNA lineages, they can be located

on the limb of haplogroup M10. Hapoltype 4 is found in the North Han population (including Liaoning, Shandong). Some other Han (including Zhejiang, inner-Mongolian, and Fujian) and TB populations (including Tibet and Tujia) have the same HVI sequence, but information concerning 10646 is required to know whether haplotype 4 exists in those populations. It should be pointed out that the HVI sequences of 16223–16311, but no M10 type are found in South Asian populations (India, Indian Andaman land, Thailand) (Sun et al., 2006; Thangaraj et al., 2005).

TABLE 4. Distribution of HVI sequences in GenBank

Haplotype	Sample	Sharing Population
1	Y2 Y10	Chinese Han
		TB (Tibet, Tujia, Pumi, Naxi, Lahu, Bai, Aini)
		Other Chinese north ethnic populations (Mongolian, Hui, Uzbek, Daur, Korean)
		Aboriginal Tanwanese
		Japanese Shizuoka
		Chinese Han
2	Y5 Y6	India*
		Japanese Tottori*
3	Y7Y12Y13Y14	Chinese Han
		TB (Tujia, Yi, Bai)
		Other Chinese north ethnic populations (Daur, Ewenki, Mongolian, kazak)
		Chinese south ethnic populations (Miao, Yao, Dai)
		Russian (Uralic region, Buryat, Chukchee, Kamchatka)
		Indian
		Ancient Neosib1 (Neolithic)
4	Y8 Y11	Chinese Han (Liaoning Shandong)
		Japanese
		Tibetan
		Thailand
		Ethiopian
5	Y4	Indonesian*
6	Y15	Taiwanese *
		South Korea*
		Japanese *
7	Y3	Unique
8	Y9	Unique

* Represents a one-step neighbor.

Haplotype 2 (16223–16256–16260) and haplotype 5 (16140–16223–16343) can be assigned to haplogroup M by 10397*Alu*I+, but their HVI motifs cannot be linked to any of the haplogroup M branches. They have been temporarily classified to haplogroup M*. Sequences identical to haplotype 2 and 5 have not previously been reported, but sequences with one step difference from these two types have been found (Table 4).

The above analyses demonstrate that the mtDNA haplogroups B, C, D, M10, and M* could be identified in the ancient Lajia population. These mitochondrial lineages all belong to major limbs of the East Asian mtDNA tree.

DISCUSSION

Maternal relationship

The ancient inhabitants of the Lajia site died simultaneously in a flood following a sudden earthquake. Due to the unexpected nature of this catastrophe, it is reasonable to assume that these people were part of an extended family that lived together. Thus, a detailed analysis of the DNA of this group of people may provide valuable information about the familial and social structures that existed in the prehistoric upper Yellow River region. In addition, the molecular data may provide a profile of the genetic structure of the Lajia population.

The 14 ancient inhabitants of the Lajia site, who died in two adjoining houses (F3 and F4), belong to eight different haplotypes. The two individuals located in F3 (Y3 and Y4) did not share the same maternal line (mother–child pair) and their mtDNA haplotypes were different from the individuals in F4. A patrilineal relationship remains a possibility since their skeletal remains were found close together. The three adults, two youths, and seven children (except Y1 and Y16) found in F4 belonged to six maternal lines. The 14 skeletal remains could be placed into six groups according to their proximity. Group I, a youth (Y12), shared the same maternal line with three other teenagers (Y7, Y13, Y14) of Group IV, II, and V. These individuals might be four descendants of one family. However, neither of the two females in the house was their mother. Group III, a mother–child pair, did not have the matrilineal affinities with any individuals of the other groups. Group IV consisted of one adult female (Y10), one youth (Y7), and three children (Y8, Y9, Y11). At first it was thought that this group might represent a mother and her children, but our analysis showed that the female was not the mother of these children. Moreover, mtDNA sequencing showed that only Y8 and Y11 shared an identical maternal line. The female (Y10) might be the sister of Y2 (Group VI). Y2 was the only

TABLE 5. Haplogroup attribution and HVI motifs for each sample

H[a]	Sample	HVI haplotype[b]	Restriction site	
1	Y2 Y10	**16223 16362**	10397*Alu*I+, 4831 *Hha*I−	D
2	Y5 Y6	**16223** 16256 16260	10397*Alu*I+, 4831*Hha*I−, 5176 *Alu*I+, 4715 *Bse*AI−, 9820*Hinf*I−	M*
3	Y7 Y12 Y13 Y14	**16223 16298 16327**	10397*Alu*I+, 4715 *Bse*AI−	C
4	Y8 Y11	**16223 16311**	10397*Alu*I+, 10646 *Rsa*I+	M10
5	Y4	16140 **16223** 16343	10397*Alu*I+, 4831*Hha*I−, 5176 *Alu*I+, 4715 *Bse*AI−, 9820*Hinf*I−	M*
6	Y15	**16223** 16256 **16311** 16343	10397*Alu*I+, 10646 *Rsa*I+	M10
7	Y3	16140 **16189** 16197 16319	10397*Alu*I−, 9bp deletion	B
8	Y9	**16189** 16267	10397*Alu*I−, 9bp deletion	B

[a] H indicates the haplotype numbers.
[b] Diagnostic polymorphisms are highlighted in bold.

TABLE 6. mtDNA haplogroup frequency of Chinese and Asian populations

Population	Location	Sample number	Haplogroup frequency								References
			D	D4	C	B	B5	M* [11]	M10[12,13]	A	
HAN											
1. Qinghai1	Northwest of China	78	23.1	17.9	2.6	12.8	3.8	1.3–3.9	1.3	5.1	1
2. Qinghai2	Northwest of China	44	31.8	18.2	4.5	4.5	2.3	9.0	ND	9.0	2
3. Gansu	Northwest of China	45	20.0	20.0	0.0	0.0	0.0	20.0	ND	17.8	2
4. Shaanxi	Northwest of China	53	26.4	20.8	1.9	9.4	1.9	11.3	ND	7.5	2
5. Liaoning1	Northeast of China	51	23.5	17.6	2.0	17.6	3.9	7.8	5.9	5.9	1
6. Liaoning2	Northeast of China	51	23.5	17.6	2.0	13.7	6.1	3.9	ND	5.9	2
7. Inner Mongolia	North of China	45	22.2	20.0	8.9	8.9	2.2	6.7	ND	6.7	2
8. Shandong	East of China	126	26.2	17.4	2.4	15.1	4.0	4.0	2.0	6.3	1,3
9. Jiangsu	East of China	67	20.9	13.4	1.5	17.9	7.5	3.0	ND	3.3	2
10. Anhui	East of China	42	16.7	9.5	0.0	19.1	9.5	4.8	ND	11.9	2
11. Hubei	Middle of China	42	9.5	4.8	2.4	19.0	7.1	0.0	0.0	16.7	1
12. Hunan	Middle of China	16	12.5	12.5	0.0	18.8	0.0	25.0	ND	4.8	2
13. Jiangxi	Southeast of China	23	26.1	17.4	0.0	34.8	21.7	4.4	ND	0.0	2
14. Guangdong	South of China	99	15.2	9.1	1.0	26.3	1.0	15.9	2.0	0.0	1
15. Fujian	Southeast of China	54	13.0	9.3	3.8	14.8	9.2	5.7	ND	0.0	2
16. Sichuan	Southwest of China	70	18.6	15.7	1.4	22.9	11.4	11.4	ND	4.3	2
17. Yunan1	Southwest of China	43	14.0	9.3	4.7	20.9	7.0	4.7	2.3	4.7	1
18. Yunan2	Southwest of China	59	18.6	10.2	7.0	17.0	5.1	3.4	ND	5.0	2
TB Populations											
1. Tibetan	Qinghai	56	7.1	ND	5.3	1.8	0.0	35.7	7.1	21.4	4
2. Tibetan	Yunnan	99	25.3	ND	2.0	7.1	4.0	33.3	2.9	12.1	4,5,6
3. Pumi	Yunnan	36	16.7	ND	22.2	2.8	0.0	19.4	11.1	13.9	4
4. Aini	Yunnan	50	24.0	ND	2.0	16.0	2.0	14.0	6.3	2	4
5. Hani	Yunnan	33	27.3	ND	3.0	18.2	9.1	12.1	3.0	12.1	4
6. Lahu	Yunnan	82	14.6	ND	1.2	14.6	0.0	8.5	0	0	4,6
7. Tujia	Hunan	94	20.2	ND	9.6	17.0	7.4	4.3	3.1	9.6	4
8. Naxi	Yunnan	45	4.4	ND	9.1	24.4	6.7	15.6	11.1	8.9	4
9. Yi	Yunnan	87	21.8	ND	3.4	14.9	2.3	13.8	0	11.4	4
10. Bai	Yunnan	77	20.8	ND	2.6	9.1	3.9	23.4	5.3	6.5	4,5
Siberia Populations											
1. Northern Siberia	Ket, Mansi, Nganasan,	160	15.6	0.6	19.4	0	0	0.6	ND	3.8	7,8,9
2. Southern Siberia	Tuvan, Buryat, Tofalar	166	10.2	0	47.6	7.8	1.2	0.9	ND	1.2	7
3. Southestern Siberia	Negidal, Ulchi, Nivkhi	176	24.4	8	9.6	2.3	2.3	1.1	ND	0	7
Mongolian	Mongolia	103	24.3	22.3	14.6	10.7	2.9	0	ND	3.9	10

[1] Yao et al., 2002a, [2]Wen et al., 2004a, [3]Yao et al., 2003, [4]Wen et al., 2004b, [5]Yao et al., 2002b, [6]Qian et al., 2001, [7]Starikovskaya et al., 2005, [8]Derbeneva et al., 2002a, [9]Derbeneva et al., 2002b, [10]Kolman et al., 1996, [11]The frequencies of M* includes the M10 type, [12]ND represents not determined, [13]M10 frequencies in the TB populations are the frequencies of M* haplotypes which has a substitution at position 16311.

adult male in the house and he shared an identical mtDNA haplotype with Y10. Y7 was from the same family as Y12, Y13, and Y14. Y9 had no maternal links with any of the individuals in F4. Of the four children that comprised Group V (Y1, Y14, Y15, and Y16), mtDNA sequences were obtained only for Y14 and Y15. Y15, similar to Y9, had no maternal links with the other individuals in the house. The finding of multiple maternal lines indicates that Lajia community did not belong to a matrilineal clan and the inhabitants of F4 were not from one or two single families. The community may represent an extended family comprised of several small families that shared some parental relationships, and that came together as a way to survive severe circumstances. Y2 might be the father of some of the children (although this cannot be determined from mtDNA); Y5 and Y6 represent a mother–child pair; and Y7, Y12, Y13, and Y14 are likely to be offspring of the same mother.

Haplogroup comparison

The genetic relationship between the ancient Lajia individuals examined in this report and the modern Chinese populations, especially the Han and TB populations, is of considerable interest. The Lajia people had a culture background that suggests they may be connected to the ancient Di-Qiang tribe of northwest China that figure in ethnologic histories concerning the origin of the Han and TB populations.

The ancient Lajia individuals belonged to haplogroups D4, B, C, M10, and M*. A compilation of haplogroup frequencies within the modern Han and TB population is provided in Table 6. The major haplogroup within the modern Han population is D, with most of the D lineages belonging to sub-haplogroup D4. D4 predominates for the northern Han (17.4–20.8%) but has a relatively lower frequency in the south of China (9.1–17.4%). The Mongolian population also has high frequency of D4 (22.3%), although the frequency of D4 in Siberia declines northward and westward. The distribution of haplogroup D4 indicates that there was an expansion of D4 outward from the center of East Asia. In Qinghai province, where the Lajia site is located, D4 is the most frequent haplogroup in the Han population (17.9%, 18.2%). The most common haplogroup in the population of central and southern Siberia is C, with the highest frequency (47.6%) being found in the population of southern Siberia. The frequency of the C haplogroup declines south-

ward, although it is still quite common (14.6%) in the population of Mongolia. In the Han population, the northern East Asia-dominating haplogroup C generally has relatively higher frequencies in the northern populations. For example, 8.9% of the Han in Inner Mongolia have the C haplotype, as do 2.6–4.5% of the Han in Qinghai province. The detection of the D4 and C haplotypes in individuals from the Lajia site is consistent with the modern geographical distribution of these haplopgroups.

In contrast to haplogroups D4 and C, haplogroup B predominates in Southeast Asia and declines northward. In southern China 14.8–34.8% of the population has haplogroup B whereas in northern China the frequency ranges from 0% to 17.6%. The frequency of haplogroup B in the Han population of Qinghai province ranges from 4.5% to 12.8%. Haplogroup B, like R9, is one of the most ancient haplogroups in the East Asian mtDNA tree (ages >50,000 years) (Yao et al., 2002a). Haplogroups B and R9 may have had their earliest diversification in Southeast Asia and later were distributed widely throughout most of East Asia.

A higher degree of polymorphism is observed in Southeast Asia compared to Northeast Asia. In the TB populations, A, B, D, F, and M* are the predominant haplogroups (Wen et al., 2004b). The D haplogroup is highly prevalent in most TB populations, whereas the other haplogroups show differential distributions across geographic regions (Table 6). The frequencies of A and M* decline from north to south, while the frequencies of B and F increase from north to south. The trends for the A, B, and F haplogroups in the TB populations are the same as for the Han population; however, the trend for the M* haplogroup is the opposite. Major haplotypes of M* that are present in Southeast Asia are absent or rare in Northeast Asia. These haplotypes possibly represent the most ancient, basal branches of M. This result indicates that the M* haplogroups in the TB and Han populations are derived from different sources, with that for the TB populations perhaps originating in the north (Wen et al., 2004b). M* lineages may thus be one of the characteristic traces remaining from the ancient northwestern tribes in China. Among East Asian populations, the Tibetan population in Qinghai province exhibits the highest frequencies of M* (35.7%). Four haplotypes (including M10 type) of M* were detected in the individuals from the ancient Lajia site.

Recently, M10, a new sub-haplogroup of M*, was defined based on substitution of positions 10646 and 16311 (Yao et al., 2002a). M10 might be a candidate for the unidentified M* lineages. In the Han population, the frequency of M10 ranges from 0% to 2.3%. The frequency of M10 in the TB populations cannot be calculated for the absence of data from the 10646 site. It should be noted, though, that the frequency of M* haplotypes which has a substitution at position 16311 is relatively high in the TB populations (0–11.1%), with a frequency of 7.1% in the Tibetan population of Qinghai province. Of the Lajia individuals, two types of M10 were also found.

The A haplogroup is present with a relatively high frequency in northern TB populations (Table 6). The frequency of the A haplogroup in the Tibetan population of Qinghai province is 21.4%. The reason why no A haplotype was detected in the Lajia individuals may be due to the small number of individuals that were examined. Alternative explanations could be that the TB populations received the A haplogroup from another tribe of migrants from the north or that the frequency of the A haplogroup has increased with time due to genetic drift. We anticipate that additional DNA sequence data obtained from members of the ancient Qijia culture may help resolve these and similar questions in the future.

CONCLUSIONS

The present study reveals that the 14 subjects found in the Lajia site shared some close maternal kinships. However, the detection of different haplotypes in individuals within the same house excludes the possibility of a matrilineal social structure. All of the ancient sequences were identified as Asian haplogroups based on the sequence of the mtDNA HVI region (16055–16379) and corresponding RFLP sites within mtDNA coding region. The mtDNA haplogroups B, C, D, M*, and M10 that were detected in the ancient individuals demonstrate the specificity and continuity of part of the mitochondrial gene pool over several millennia in this location. The geographic distributions of these detected haplogroups in the modern Han and TB populations indicate that the ancient people that lived in Northwest China may contribute to the maternal gene pool of these modern populations, which is consistent with their ethnic history.

ACKNOWLEDGMENTS

We thank Maolin Ye and Ming-Hui Wang for sample collection and Li Jin for providing the mtDNA sequences of Tibeto-Burman speaking populations. We are also grateful to Graham P. Côté for helpful comments on the manuscript.

LITERATURE CITED

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 23:147.

Derbeneva OA, Starikovskaya EB, Wallace DC, Sukernik RI. 2002a. Traces of early Eurasians in the Mansi of northwest Siberia revealed by mitochondrial DNA analysis. Am J Hum Genet 70:1009–1014.

Derbeneva OA, Starikovskaya EB, Volodko NV, Wallace DC, Sukernik RI. 2002b. Mitochondrial DNA variation in Kets and Nganasans and the early peopling of Northern Eurasia. Russ J Genet 38:1316–1321.

Kivisild T, Tolk HV, Parik J, Wang Y, Papiha SS, Bandelt HJ, Villems R. 2002. The emerging limbs and twigs of the East Asian mtDNA tree. Mol Biol Evol 19:1737–1751.

Kolman CJ, Sambuughin N, Bermingham E. 1996. Mitochondrial DNA analysis of Mongolian populations and implications for the Origin of New World Founders. Genetics 142:1321–1334.

Kong QP, Yao YG, Sun C, Bandelt HJ, Zhu CL, Zhang YP. 2003. Phylogeny of East Asian mitochondrial DNA lineages inferred from complete sequences. Am J Hum Genet 73:671–676.

Kumar S, Tamura K Nei M. 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Briefings Bioinformatics 5:150–163.

Liu XG. 2003. The cultures of Bronze Age in Qinghai region. Stor Relics 6:29–32.

Macaulay V, Richards M, Hickey E, Vega E, Cruciani F, Guida V, Scozzari R, Bonne-Tamir B, Sykes B, Torroni A. 1999. The emerging tree of west Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. Am J Hum Genet 64:232–249.

O'Rourke DH, Hayes MG, Carlyle SW. 2000. Ancient DNA studies in physical anthropology. Annu Rev Anthropol 29:217–42.

Ovchinnikov I, Buzhilova A, Mednikova M, Goodwin W, Curry G. 1999. Ethnic affinities of the ancient human Jety-Asar population by mitochondrial DNA analysis. Electrophoresis 20:1729–1732.

Pakendorf B, Wiebe V, Tarskaia LA, Spitsyn VA, Soodyall H, Rodewald A, Stoneking M. 2003. Mitochondrial DNA evidence for admixed origins of central Siberian populations. Am J Phys Anthropol 120:211–224.

Qian YP, Chu ZT, Dai Q, Wei CD, Chu JY, Tajima A, Horai S. 2001. Mitochondrial DNA polymorphisms in Yunnan nationalities in China. J Hum Genet 46:211–220.

Ren XY, Wang DG, Cai LH, He KZ, Ye ML. 2002. The expedition of Lajia site in Minhe county of Qinghai province in 2000. Archaeology 12:12–25.

Richards M, Corte-Real H, Forster P, Macaulay V, Wilkinson-Herbots H, Demaine A, Papiha S, Hedges R, Bandelt HJ, Sykes B. 1996. Paleolithic and Neolithic lineages in the European mitochondrial gene pool. Am J Hum Genet 59:185–203.

Starikovskaya EB, Sukernik RI, Derbeneva OA, Voldko NV, Ruiz-Pesini E, Torroni A, Brown MD, Lott MT, Hosseini SH, Huoponen K, Wallace DC. 2005. Mitochondrial DNA diversity in indigenous populations of the southern extent of Siberia, and the origins of Native American haplogroups. Ann Hum Genet 69:67–89.

Sun C, Kong QP, Palanichamy MG, Agrawal S, Bandelt HJ, Yao YG, Khan F, Zhu CL, Chaudhuri TK, Zhang YP. 2006. The dazzling array of basal branches in the mtDNA macrohaplogroup M from India as inferred from complete genomes. Mol Biol Evol 23:683–690.

Thangaraj K, Chaubey G, Kivisild T, Reddy AG, Singh VK, Rasalkar AA, Singh L. 2005. Reconstructing the origin of Andaman Islanders. Science 308:965–996.

Tian XY. 2001. The history of Chinese nationalities. Beijing: Huaxia Press. p 25–178.

Wen B, Li H, Lu D, Song X, Zhang F, He Y, Li F, Gao Y, Mao X, Zhang L, Qian J, Tan J, Jin J, Huang W, Deka R, Su B, Chakraborty R, Jin L. 2004a. Genetic evidence supports demic diffusion of Han culture. Nature 431:302–305.

Wen B, Xie XH, Gao S, Li H, Shi H, Song XF, Qian TZ, Xiao CJ, Jin JZ, Su B, Lu DR, Chakraborty R, Jin L. 2004b. Analyses of genetic structure of Tibeto-Burman populations reveals sex-biased admixture in Southern Tibeto-Burmans. Am J Hum Genet 74:856–865.

Wen B, Li H, Gao S, Mao XY, Gao Y, Li F, Zhang F, He YG, Dong YL, Zhang YJ, Huang WJ, Jin JZ, Xiao CJ, Lu DR, Chakraborty R, Su B, Deka R, Jin L. 2005. Genetic structure of Hmong-Mien speaking populations in East Asia as revealed by mtDNA lineages. Mol Biol Evol 22: 725–734.

Xia ZK, Yang XY, Ye ML. 2003. The prehistoric disaster of the Lajia site. Chin Sci Bull 48:1200–1204.

Xu YT. 2003. A brief study on the origin of Han nationality. J Central Univ Nationalities 3:59–64.

Yang SM, Ding H. 2003. The ethnography of China. Beijing: Central University for Nationalities Press.

Yao YG, Kong QP, Bandelt HJ, Kivisild T, Zhang YP. 2002a. Phylogeographic differentiation of mitochondrial DNA in Han Chinese. Am J Hum Genet 70:635–651.

Yao YG, Nie L, Harpending H, Fu YX, Yuan ZG, Zhang YP. 2002b. Genetic relationship of Chinese ethnic populations revealed by mtDNA sequence diversity. Am J Phys Anthropol 118:63–76.

Yao YG, Kong QP, Man XY, Bandelt HJ, Zhang YP. 2003. Reconstructing the evolutionary history of China: a caveat about inferences drawn from ancient DNA. Mol Biol Evol 20:214–219.

Zhang Y, Yu LJ, Qian R. 2004. "The eastern Pompeii"–the Lajia site. Silk Road 4:11–14.