

## ORIGINAL ARTICLE

# Genetic diversity of two Neolithic populations provides evidence of farming expansions in North China

Ye Zhang<sup>1</sup>, Jiawei Li<sup>1</sup>, Yongbin Zhao<sup>1,2</sup>, Xiyang Wu<sup>1</sup>, Hongjie Li<sup>1,3</sup>, Lu Yao<sup>4</sup>, Hong Zhu<sup>3</sup> and Hui Zhou<sup>1,3</sup>

The West Liao River Valley and the Yellow River Valley are recognized Neolithic farming centers in North China. The population dynamics between these two centers have significantly contributed to the present-day genetic patterns and the agricultural advances of North China. To understand the Neolithic farming expansions between the West Liao River Valley and the Yellow River Valley, we analyzed mitochondrial DNA (mtDNA) and the Y chromosome of 48 individuals from two archeological sites, Jiangjialiang (>3000 BC) and Sanguan (~1500 BC). These two sites are situated between the two farming centers and experienced a subsistence shift from hunting to farming. We did not find a significant difference in the mtDNA, but their genetic variations in the Y chromosome were different. Individuals from the Jiangjialiang belonged to two Y haplogroups, N1 (not N1a or N1c) and N1c. The individuals from the Sanguan are Y haplogroup O3. Two stages of migration are supported. Populations from the West Liao River Valley spread south at about 3000 BC, and a second northward expansion from the Yellow River Valley occurred later (3000–1500 BC).

*Journal of Human Genetics* advance online publication, 1 September 2016; doi:10.1038/jhg.2016.107

## INTRODUCTION

Archeological evidence from prehistoric sites in North China points to the west Liao River and the Yellow River Valley as separately developed Neolithic centers in North China (Figures 1 and 2).<sup>1</sup> Ancient communities in North China transitioned from hunting and gathering to domestication as their predominant subsistence strategy 10 000 years ago.<sup>1–3</sup> The process of population movements and interaction during the Neolithic expansions between the west Liao River Valley and the Yellow River Valley are not well understood, but have a major role in the modern distribution of haplogroups across North China.

The northern farming center is in the West Liao River Valley.<sup>4</sup> The earliest evidence of crop domestication dates to 6500 BC at the Xinglongwa site.<sup>2</sup> The common millet became a staple diet during the Hongshan culture period (4500–3000 BC, characterized by their jade artifacts and dragon-shaped relics),<sup>5,6</sup> indicating that agriculture was a chief way of life. Agriculture provides high yields and allows the support of an increasing population and flourishing culture,<sup>6</sup> such as the Xiaohewan culture (3000–2200 BC) and the Lower Xiajiadian culture (2200–1600 BC).<sup>7,8</sup> Following the Lower Xiajiadian culture, paleoclimate records showed a regional shift toward a cold and steppe-like environment, which encouraged nomadic subsistence strategy ~1000 BC.<sup>9–11</sup> Consequently, the widespread agricultural practices disappeared from the archeological records in this region, but some characteristics of this farming center remained in North China, such as the widespread use of jade.<sup>12</sup>

To the south, the farming center traces to the Yellow River Valley.<sup>2</sup> The earliest evidence of crop domestication in this region dates to 8000 BC at the Cishan site.<sup>3</sup> The majority of archeological sites located at the Yellow River Valley sites belong to the Yangshao culture (5000–3000 BC), which is known for its distinctive geometric or animal prints patterned pottery in red or black.<sup>13</sup> The common millet and foxtail millet were grown throughout this region at that time period.<sup>2,14,15</sup> Yangshao culture holds significance in Chinese history as scholars, arguing that it gives rise to the present-day Han Chinese culture.<sup>16</sup>

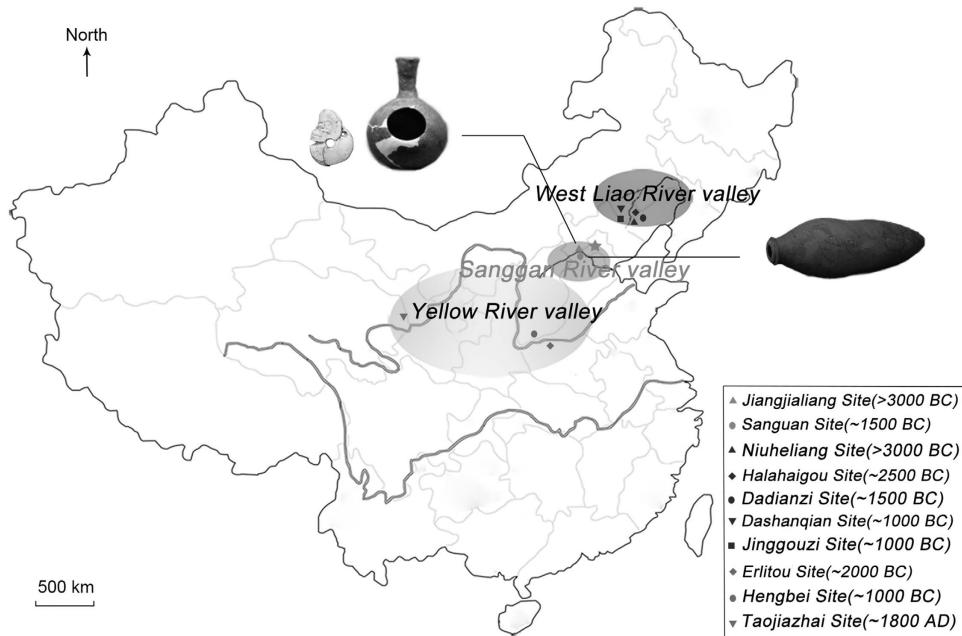
The Sanggan River Valley is centrally located and accessible between the West Liao River Valley and the Yellow River Valley (Figure 1). Until ~4300 BC, hunting-and-gathering was the main way of life for the ancient community, since local climate aridity is not favorable for intensified agriculture development.<sup>17</sup> After this period, archeological evidence indicates a shift toward a farming strategy.<sup>17–20</sup> Regional data from the Sanggan River Valley suggest that technologies from the west Liao River Valley and the Yellow River Valley were influential on the development and adaptations of agricultural practices.<sup>21</sup> Hence, the genetic changes of populations in this region during their lifestyle shifted toward an agriculturally focused subsistence strategy can provide evidence of Neolithic farming expansions in North China.

The Jiangjialiang (JL) and Sanguan (SG) sites are located in the Sanggan River Valley (Figure 1), yet display different subsistence economies. The JL site was one of the largest Neolithic sites in the Sanggan River Valley. It is the oldest Neolithic Sanggan River Valley site with human remains.<sup>19</sup> A previous excavation uncovered a total of

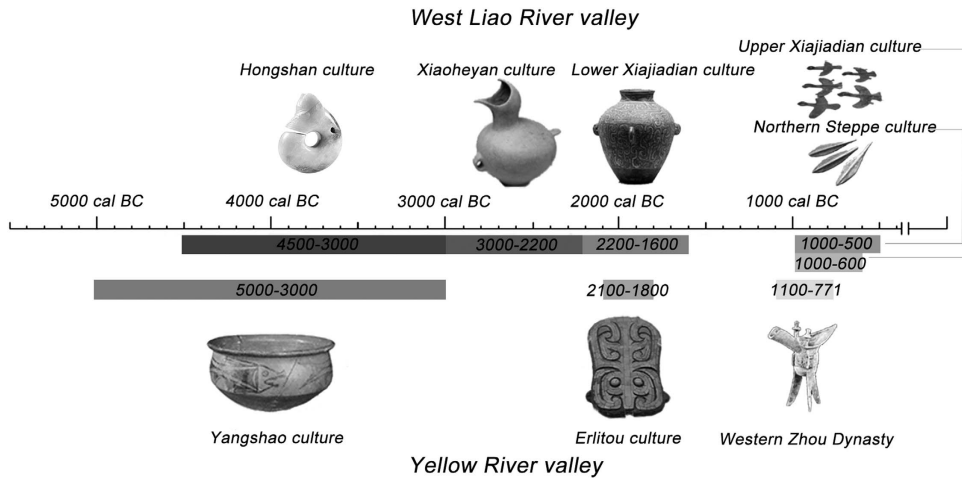
<sup>1</sup>Laboratory of Ancient DNA, School of Life Science, Jilin University, Changchun, China; <sup>2</sup>Laboratory of Ancient DNA, College of Life Science, Jilin Normal University, Siping, China; <sup>3</sup>Laboratory of Anthropology, Research Center for Chinese Frontier Archaeology, Jilin University, Changchun, China and <sup>4</sup>Department of Anthropology, Committee on Evolutionary Biology, University of Chicago, Chicago, USA

Correspondence: Dr H Zhou, Laboratory of Ancient DNA, School of Life Science, Jilin University, Qianjin Street No. 2699, Changchun, Jilin 130021, China.  
E-mail: zhouhui@jlu.edu.cn

Received 30 May 2016; revised 23 July 2016; accepted 26 July 2016



**Figure 1** Geographic location of the valleys and archeological sites used in this study. The study sites in this research: Jiangjialiang site (Xueshan culture) and Sangan site (Lower Xiajiadian culture) in the Sanggan River Valley. The genetic data were previously published for 11 sites: Niuheliang site (Hongshan culture); Halahaigou site (Xiaohayan culture); Dadianzi site (Lower Xiajiadian culture); Dashanqian site (upper Xiajiadian culture); and Jinggouzi site (Northern Stepe culture) in the West Liao River Valley; and Erlitou site (Erlitou culture), Hengbei site (Western Zhou Dynasty) and Taojiazhai site (Han Dynasty) in the Yellow River Valley. The red star represents Beijing, China. A full color version of this figure is available at the *Journal of Human Genetics* journal online.



**Figure 2** Cultures studied here from the West Liao River Valley and the Yellow River Valley. Characteristic artifacts are shown, along with approximate dates for each culture. A full color version of this figure is available at the *Journal of Human Genetics* journal online.

nine buildings and 78 graves (Supplementary Figure S1).<sup>19</sup> Radiocarbon dating of the nine buildings in the lowest stratigraphic layer indicated that they were constructed in  $4850 \pm 80$  BC.<sup>19,22</sup> Artifacts were found in the buildings and graves, including pottery, millstones, stone axes and flake tools. The non-decorative sand inclusion pottery is stylistically similar to the early Xueshan culture (3600–2900 BC).<sup>23</sup> A large number of millstones were excavated in buildings and graves at this site, which demonstrated that by ~7000 years ago people had already begun to practice agricultural techniques. However, stone axes and flake tools are also found in the buildings and graves, suggesting that hunting still had a crucial role.<sup>22</sup>

The SG site ( $1435 \pm 170$  BC) is more recent than the JJL site by 1500 years. It is approximately 30 km from the JJL site.<sup>24</sup> The SG site is

not geographically isolated from the JJL site. Excavation from the SG site and the JJL site yield bowls with similar characteristics. This suggested a close cultural affinity between these two sites.<sup>19</sup> The SG population may extensively practice agricultural subsistence strategy, and share aspects of the Lower Xiajiadian culture (2500–1500 BC).<sup>8</sup> For the SG population, agriculture was the primary form of subsistence, replacing the earlier hunting and gathering lifestyle.

JJL people and SG people represent populations in the same region whose subsistence changed from hunting to farming. In this study, mtDNA and the Y chromosome from ancient humans were sampled and analyzed to determine the genetic structure of the JJL and SG populations. By comparing the genetic diversity of the JJL and SG populations with ancient populations from the West Liao River Valley

and the Yellow River Valley in North China (Figure 1), we aim to understand the Neolithic farming expansions in North China.

## MATERIALS AND METHODS

### Sample selection

Forty-one individuals were randomly selected from 74 human skeletal remains from the JIL site (Supplementary Figure S1). The remains were kept at the State Museum of Research Center for Chinese Frontier Archeology at Jilin University. Anthropological sex identification had been previously performed on the majority of these samples (Supplementary Table S1).<sup>22</sup> From the SG site, we selected seven samples whose mtDNA had already been observed<sup>25</sup> for a Y-chromosome single-nucleotide polymorphism (SNP) analysis. The genetic samples were exclusively taken from teeth. All human remains used in this study were excavated by the Cultural Relics Institute of Hebei Province, with permission from the State Administration of Cultural Heritage of China.

### Sample decontamination, grinding and DNA extraction

Proper precautionary processes were taken to minimize contamination. Soil and tartar were removed from the surface of the teeth from the 41 individuals of JIL population. A liquid treatment was administered with a 5% sodium hypochlorite solution, rinsed once with ddH<sub>2</sub>O and 95% ethanol. These samples were dried on each side in a UV-radiation box for 30 min. For each sample, 200 mg of fine enamel powder was collected using a dental drill (STRONG 90, Taegu, Korea) into a 15 ml tube. The enamel powder was stored at -20 °C. For the seven SG individuals, we used enamel powder collected from the previous study.<sup>25</sup> Each sample was digested in 5 ml of EDTA (0.5 mol, PH 8.0) and 70 µl of proteinase K (100 mg ml<sup>-1</sup>) for 12–16 h in a rotating hybridization oven at 55 °C. The supernatant was transferred to an ultrafiltration tube (centricon YM-10, Merck Millipore, Darmstadt, Germany) and condensed to ~100 µl at 6300 r.p.m. DNA was then extracted and eluted to 70 µl using the QIAquick PCR Purification Kit (Qiagen, Düsseldorf, Germany), using the manufacturer's protocol.

### Mitochondrial DNA sequencing and SNP typing

Within the mitochondrial genome, the hypervariable region 1 (HVR1) and coding region SNPs are the focus of this study. The HVR1 was amplified with two overlapping primer pairs spanning 375 bp (nucleotide position 16035–16409). Twelve haplogroup-defining SNPs in the coding region of the mitochondrial genome were typed to determine mtDNA lineages. The primer sequences are shown in Supplementary Table S2. The HVR1 and SNP PCR products were amplified using Takara Ex Taq Hot-Start DNA polymerase. The amplified products were further sequenced directly using forward and reverse primers and an ABI 310 Terminator Sequencing Kit on an ABI PRISM 310 automated DNA sequencer (Applied Biosystems, Foster City, CA, USA) without additional purification. These mitochondrial sequences were then analyzed for mtDNA variability using the program Sequencher 5.0 according to the Cambridge Reference Sequence.<sup>26</sup> MitoTool 1.1.2 were used to yield a list of the variants and determine the haplogroup status of that lineage according to Phylotree (www.phylotree.org).<sup>27,28</sup> Haplogroups M/N, B, C, D, D4 and F were detected using product-length polymorphism analysis. Haplogroups A, D5, G, M10, R and Z were identified by sequencing using amplification products. The PCR amplification was performed as described by Li *et al.*<sup>39</sup> with modified primer annealing time to 30s. To reveal maternal genetic differences between the ancient populations of North China,  $F_{st}$  genetic distances were calculated using Arlequin 3.11<sup>29</sup> by the HVR1 data (nucleotide positions 16051–16400). These populations include JIL, SG, Erlitou (ELT), Hengbei (HB), Niheliang (NHL), Halahaigou (HLHG) and Dadianzi (DDZ).

### Sex identification, Y-chromosome SNP and Y-STR

The amelogenin gene and SRY locus were used for sex identification of individuals from the JIL and SG sites.<sup>30,31</sup> The male samples were then selected for Y-chromosome analyses. The most prevalent Eastern Asian Y-chromosomal haplogroup F-M89 was detected in this study. Subsequent analyses were restricted to seven sub-haplogroups (K-M9, NO-M214, N-M231, N1a-M128, N1c-Tat, O-M175 and O3-M122). All PCR amplifications were performed as

described by Karafet *et al.*<sup>32</sup> The primer sequences used are shown in Supplementary Table S2.

Short tandem repeats (STRs) on the Y chromosome (Y-STRs) from the JIL samples were analyzed on 18 loci (DYS19, DYS385a/b, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS447, DYS448, DYS456, DYS458, DYS635 and Y GATA H4) using the AGCU Y18 STR Kit (AGCU ScienTech, Wuxi, China). Y-STR data are not available to the SG population because DNA extracts from the SG individuals ran out. PCR amplification was performed as described by Cui *et al.*<sup>38</sup> The products from the Y-STR reactions were analyzed using GeneMapper software 4.2 (Applied Biosystems) and Network 5.0 (Fluxus Technology Ltd, Kiel, Germany).

### Precautions against ancient DNA contamination

To prevent modern sources of contamination in ancient DNA, researchers in the pre-PCR lab wore full-body protective clothing, facemasks and several layers of gloves. A strict workflow protocols were followed when doing the laboratory work. The post and pre-PCR labs are located in separate buildings. Sample preparation, DNA extraction, purification and PCR set-up were completed in the pre-PCR laboratory. PCR and sequencing were carried out in the post-PCR laboratory. Every step in the pre-PCR laboratory was executed in hoods that were irradiated with UV for at least 30 min and cleaned with DNA-Off before, between, and after each use. The rooms and other equipment were also treated as described above. Extraction and PCR negative controls yielded negative results, as expected. We selected teeth specimens *in situ* to avoid sampling duplication. Twenty individuals have performed at least two duplicated extractions. All the results were verified by performing two duplicated amplifications. HVR1 sequences from the SG samples from this study yield the same results as previous study conducted Zhao *et al.*<sup>25</sup> HVR1 and Y-STR of all researchers who worked in the lab were sequenced to check for contamination (Supplementary Table S3).

## RESULTS

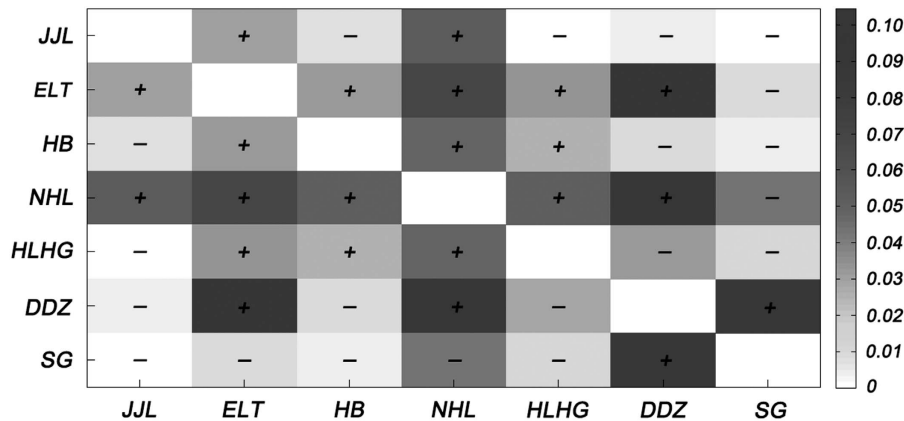
The resulting mtDNA HVR1 sequences from JIL specimens were compared with the revised Cambridge Reference Sequence.<sup>26</sup> We identified 23 polymorphic sites from 41 JIL individuals. A total of 36 haplotypes from 10 haplogroups (D, A, G, B, M10, C, Z, F, R and M\*) were observed (Supplementary Table S1). Two samples marked M\* represent a failure in the typing process. Haplogroups D (34%), A (15%), G (12%), B (10%) and M10 (7%) are present in the JIL population at a higher frequency than the other haplogroups (C, Z, F and R). Haplogroups C (5%), R (5%), Z (5%) and F (2%) were observed at a low frequency in the JIL population (Table 1). To better understand the regional genetic relationship in the Sanggan River Valley, West Liao River Valley and Yellow River Valley,  $F_{st}$  genetic distances between seven ancient populations were calculated (Figure 3 and Supplementary Table S4).<sup>33–37</sup> The result indicated that the smallest  $F_{st}$  value was between the JIL and SG populations ( $F_{st}=0.00009$ ). Samples from the TJZ site are not included in the  $F_{st}$  analysis, as specimens are excavated from a family cemetery with similar mtDNA variation.

Using the HVR1 data from Zhao *et al.*<sup>25</sup> (Supplementary Table S5), we determined that the haplogroups belonging to the SG population include B, D, G, A, M10 and M\*, all of which are also present in the JIL population at a higher frequency. With the small sample size, the haplogroup frequency of the SG population was not calculated.

In total, 17 out of 41 individuals from the JIL site were classified as male through sex identification. Twenty individuals were classified as female, including two individuals (95JILM38 and 95JILM65x) who were identified as females in molecular identification but were considered males through morphological identification.<sup>22</sup> The results for three individuals (95JILM43z, 95JILM52 and 95JILM53) were unclear using the molecular identification method (Supplementary Table S1). All JIL males expressed the Y haplogroup N-M231 (Supplementary Table S5), the most common haplogroup in North

**Table 1** Distribution of mtDNA haplogroups (%) in the eight ancient populations and the modern North China population

Geographical region	Population	Number	D	A	G	B	M10	R	C	Z	F	M8a	M9a	N9a	M7	Other	Reference
West Liao River Valley	JJL	41	34.1	14.6	12.2	9.8	7.3	4.9	4.9	4.9	2.4	0	0	0	0	4.9	This article
	NHL	28	28.6	10.7	0	10.7	0	0	0	3.6	3.6	3.6	3.6	35.6	0	0	Zhao <sup>33</sup>
	HLHG	36	47.2	8.3	13.9	11.1	0	2.8	5.6	0	2.8	0	0	8.3	0	0	Zhao <sup>33</sup>
	DDZ	14	42.9	7.1	7.1	0	7.1	0	0	7.14	7.14	0	7.14	0	14.3	0	Li <i>et al.</i> <sup>39</sup>
Yellow River Valley	DSQ	14	15.4	0	38.5	7.7	0	7.7	0	0	23.1	0	0	7.7	0	0	Zhao <sup>33</sup>
	ELT	26	48.2	3.7	0	18.5	0	0	0	0	7.41	0	0	3.7	7.4	11.1	Liu <sup>34</sup>
	HB	52	26.9	9.6	1.9	13.5	5.8	0	1.9	1.9	11.5	3.9	7.7	3.9	1.9	9.6	Zhao <i>et al.</i> <sup>35</sup>
North China	TJZ	29	27.6	13.8	0	20.7	3.5	0	0	6.9	17.2	0	0	3.5	0	6.9	Zhao <i>et al.</i> <sup>36</sup>
	Modern	1100	24.7	8.2	5.9	5.5	0	1.1	6.5	3.5	9.6	12.4	2.3	2.6	5.8	11.5	Xue <i>et al.</i> <sup>37</sup>



**Figure 3**  $F_{st}$  values between ancient populations of North China, based on HVR1 data. The '+' indicates the  $F_{st}$  value is significant, with a  $P$ -value less than 0.05; the '-' indicates the  $F_{st}$  was not significant, with a  $P$ -value greater than 0.05. A full color version of this figure is available at the *Journal of Human Genetics* journal online.

China 3000 years ago.<sup>38</sup> Haplogroup N1 ( $\times$ N1a,  $\times$ N1c) is present in 10 out of the 17 males while the haplogroup N1c-Tat was observed in the other seven male individuals (Table 2).

For the SG samples, four of seven individuals were classified as male, and they all belonged to the Y haplogroup O3-M122 (Supplementary Tables S2 and S5), common in the ancient people of the Yellow River Valley.<sup>38,39</sup> Haplogroup O3-M122 is the main haplogroup observed in the Han Chinese people today,<sup>40-42</sup> however was not identified in the JJL population.

Y-STR data were available from the JJL samples (Supplementary Table S5). DYS447, DYS439 and DYS392 were not successfully amplified in almost all the samples. This is likely because the lengths of these amplicons are too long for ancient DNA. From the Median-joining network analysis, we found the Y-STRs in the N1c population were extremely variable. While the N1 ( $\times$ N1a,  $\times$ N1c) population can be divided into three paternal groups (Supplementary Figure S2). JJL3, JJL6, JJL25, JJL40 and JJL63z individuals share one paternal relative, JJL9, JJL10 and JJL11 individuals share another paternal relative, and JJL16 and JJL29 share a third paternal relative.

## DISCUSSION

The mtDNA haplogroups D, A, G, B and M10 were present at a high frequency in the JJL population. These are the most common haplogroups found in the ancient and present-day northeastern Chinese populations.<sup>33,36,37,39</sup> MtDNA haplogroups C, Z, F and R were present at a low frequency in the JJL population. This is expected because low frequencies of these haplogroups are also observed in

most other northeastern Chinese ancient populations (Table 1).<sup>33,39</sup> All haplogroups observed in the JJL population were exclusive to North China, indicating that the JJL population was an ancient local population of North China. No evidence of in-migration is found based on haplogroups present in regions outside of Northeast China, such as the West Eurasian maternal lineages U and W in the Tianshanbeilu site of Xinjiang in northwestern China.<sup>43</sup> Unfortunately, only seven samples were excavated from the SG site, so the data were too small to the mtDNA structure analysis.

Archaeological evidence clearly suggested that the ancient JJL and SG communities utilized different subsistence strategies.<sup>22,44</sup> The subsistence strategy was changed during this 1500-year period from both hunting and farming at the JJL site to extensive farming at the SG site. The mtDNA data from this study, however, did not reveal a significant difference of each population. In fact, all the haplogroups (B, D, G, A and M10) in the SG population were present at a high frequency in the JJL population. This is further supported by the small  $F_{st}$  distance between the JJL and SG populations (Figure 3). The mtDNA results indicate that there might be a close maternal relationship between the two populations. More mtDNA data from relevant spatial and temporal ranges are needed to more accurately describe the maternal model of population dynamics during the subsistence shift observed in the archeological records.

The resulting Y-chromosome analysis indicated the presence of a rare haplogroup N1c-Tat in the JJL population. This is an unexpected result for ancient populations in North China. Searching through previously reported Y-chromosome variations for 119 known ancient

**Table 2** Geographic locations and the Y-chromosome haplogroup distribution of ancient populations in this study

Geographical region	Site	Culture	Size	N1 ( $\times$ N1a, $\times$ N1c) (%)	N1c (%)	O3 (%)	O2 (%)	O* (%)	Q (%)	C (%)	Reference
Sanggan River Valley	JJL	Xueshan culture	17	58.8	41.2	0	0	0	0	0	This article
	SG	Lower Xiajiadian culture	4	0	0	100	0	0	0	0	This article
West Liao River Valley	NHL	Hongshan culture	6	66.7	0	13.7	0	0	0	13.7	Cui <i>et al.</i> <sup>38</sup>
	HLHG	Xiaohayan culture	12	100	0	0	0	0	0	0	Cui <i>et al.</i> <sup>38</sup>
	DDZ	Lower Xiajiadian culture	5	60	0	40	0	0	0	0	Li <i>et al.</i> <sup>39</sup>
	DSQ	Upper Xiajiadian culture	9	11.1	33.3	44.4	0	0	0	11.1	Cui <i>et al.</i> <sup>38</sup>
	JGZ	Northern Steppe culture	12	0	0	0	0	0	0	100	Cui <i>et al.</i> <sup>38</sup>
Yellow River Valley	HB	Western Zhou dynasty	22	4.5	0	27.3	9.1	18.2	40.9	0	Zhao <i>et al.</i> <sup>48</sup>
	TJZ	Han dynasty	12	0	0	100	0	0	0	0	Zhao <i>et al.</i> <sup>36</sup>

individuals from 13 archeological sites (approximately dated to 6000–2000 years ago) in North China, the N1c-Tat was found in only three individuals from the Bronze Age Dashanqian site (~1000 BC) in the West Liao River Valley.<sup>38,39,45</sup>

Considering the rarity of the N1c-Tat haplogroup distribution in the ancient populations throughout the late Neolithic and Bronze Ages, it is possible that N1c-Tat represented the local population of the Sanggan River Valley. These people may be the original settlers who used a hunter-gatherer subsistence strategy in the Sanggan River Valley. During the shift from hunting to farming, they likely embraced new techniques and then began to use farming technology.

The haplogroup N1 ( $\times$ N1a,  $\times$ N1c) was widely distributed at a high density in ancient populations of North China but at a much lower frequency (<5.96%) in modern East Asian populations.<sup>42</sup> Previous studies supported that haplogroup N1 expanded into North China 12–18 kya.<sup>46,47</sup> High-frequency distributions of N1 ( $\times$ N1a,  $\times$ N1c) (Table 2) were observed in populations from the West Liao River Valley during the Neolithic and early Bronze Age,<sup>38</sup> showing that it became the dominant Y-chromosome lineage of North China at that time. Through the West Liao River Valley, researchers argued that the N1 ( $\times$ N1a,  $\times$ N1c) haplogroup migrated to northwestern China (for example, the Bronze Age Tianshanbeilu site in Xinjiang)<sup>43</sup> and the Yellow River Valley (for example, the Hengbei site).<sup>48</sup> The presence of N1 ( $\times$ N1a,  $\times$ N1c) haplogroup in the ancient JJL population suggested that this population maybe in close affinity with the West Liao River Valley populations. This suggests possible genetic exchanges between the West Liao River Valley and the Sanggan River Valley populations.

Haplogroup O3-M122 is the main lineage of the Yellow River Valley.<sup>49</sup> It is, however, not present in the JJL population by this study. Artifacts excavated from the JJL site also showed more stylistic features of the Hongshan culture of the West Liao River Valley.<sup>19</sup> Genetic and archeological evidence indicates that people from the Yellow River Valley probably did not reach the Sanggan River Valley before 3000 BC. Therefore, people from the West Liao River Valley, not the Yellow River Valley, most likely first influenced the agricultural practices in the Sanggan River Valley.

Haplogroup O3-M122 was observed in all four males from the SG site. Considering the SG site is more recent than the JJL site, the presence of haplogroup O3-M122 suggests that people carrying this haplogroup from the Yellow River Valley may have migrated to the Sanggan River Valley sometime between 3000 and 1500 BC. Haplogroups N1 ( $\times$ N1a,  $\times$ N1c) and N1c were not present in the SG samples in this research. Their absence may be due to the small sample size of the SG population. Another hypothesis is that the

Yellow River Valley population carrying haplogroup O3-M122 completely replaced the local people carrying N1 ( $\times$ N1a,  $\times$ N1c) and N1c. This hypothesis is not very compelling because haplogroup N1 ( $\times$ N1a,  $\times$ N1c) and N1c were still found in the later populations in North China. The more likely scenario is that the Yellow River Valley populations admixed with the local populations during this expansion. For example, the Dadianzi population, a Lower Xiajiadian cultural group of West Liao River Valley, carried N1 ( $\times$ N1a,  $\times$ N1c) and O3 haplogroups.<sup>39,45</sup> Similarly the later Dashanqian population, characterized to the Upper Xiajiadian culture, had evidence of N1 ( $\times$ N1a,  $\times$ N1c), N1c and O3 haplogroups.<sup>38,45</sup>

Combining the Y-chromosome analyses with the archeological evidence,<sup>8,19,23</sup> we find Neolithic farming expansions from the Yellow River Valley and the West Liao River Valley. The West Liao River Valley population expanded earlier, with agriculturalists spreading south and reaching the Sanggan River Valley before 3000 BC. This migration was combined with the appearance of agricultural techniques in the Sanggan River Valley.<sup>19,22</sup> The culture also shifted into a unique transitional culture, the Xueshan culture, which was characterized by both hunting and farming subsistence strategies.<sup>23</sup>

The later expansion from the Yellow River Valley may happen during 3000–1500 BC. This expansion also greatly influenced people in the Sanggan River Valley genetically and culturally.<sup>8</sup> During this expansion, people from the Yellow River Valley migrated to the Sanggan River Valley and finally settled there. They also contributed to the change from the Xueshan culture to the Lower Xiajiadian culture, a highly developed agricultural culture. Another consequence of the expansion of the Yellow River Valley population is that a portion of local Sanggan River Valley people might have migrated to the West Liao River Valley. This hypothesis is supported by the appearance of the haplogroup N1c in individuals from the Dashanqian site.<sup>45</sup>

More ancient data, especially for the Y chromosome, during adjacent time periods and from the Yellow River Valley and the Sanggan River Valley, are needed for further research regarding farming expansion in North China.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ACKNOWLEDGEMENTS

This study was supported by grants from the National Natural Science Foundation of China (Grant numbers 31371266, J1210007 and 31301025) and the National Social Science Foundation of China (Grant No.11& ZD182). We are grateful to Fei Xie of Hebei Cultural Relics and Jun Li and Xiaozheng Zhang of Hebei Archeology Institute for providing access to the human remains.

- 1 He, H. *A Study on the Production of Setaria Italica Beauv. Ancient Times of China* (Nanjing Agricultural University, 2010).
- 2 An, Z. The Neolithic archaeology of China. *Curr. Anthropol.* **29**, 753–759 (1988).
- 3 Lu, H., Zhang, J., Liu, K. B., Wu, N., Li, Y., Zhou, K. *et al*. Earliest domestication of common millet (*Panicum miliaceum*) in East Asia extended to 10 000 years ago. *Proc. Natl Acad. Sci. USA* **106**, 7367–7372 (2009).
- 4 Liu, G. The mode and characteristics of the origin of civilizations of Hongshan culture and West Liaohé River basin. *Cultural Relics Archaeol. Inner Mongolia* **1**, 43–51 (2010).
- 5 Tian, G. *The Origin of Culture in West Liao River Valley, North China* (Northeast Normal University, 2003).
- 6 Peterson, C. E., Lu, X., Drennan, R. D. & Zhu, D. Hongshan chiefly communities in neolithic northeastern China. *Proc. Natl Acad. Sci. USA* **107**, 5756–5761 (2010).
- 7 Suo, X.-F. & Li, S.-B. The time, fountainhead and direction of Xiaohéyan Culture. *Res. Chin. Front. Archaeol.* **1**, 57–63 (2008).
- 8 Ping, C. Review of lower Xiajiadian culture. *Beijing Cultural Relics Archaeol.* **1**, 99–123 (2002).
- 9 Li, Y.-h., Yin, H.-n., Zhang, X.-Y. & Chen, Z.-J. The environment disaster events and the evolution of man-land relation in the West Liaoning during 5000 a BP. *J. Glaciol. Geocryol.* **25**, 19–26 (2003).
- 10 Han, M. A study on settlements and environment of prehistoric times in the West Liao River Valley. *Acta Archaeol. Sin.* **1**, 1–20 (2010).
- 11 Shao, H., Yang, J. A review of the Steppe metal route by analysing the bronze remains from upper Xiajiadian culture. *Archaeology* **10**, 85–99 (2015).
- 12 Zhou, X. *Study on Hongshan Culture Jade* (Jilin University, 2014).
- 13 Wu, S. The neolithic decorative patterns of Yellow River, Yangzi River and Southern China. *Cultural Relics* **5**, 59–72 (1975).
- 14 Wang, R. The origin of Yangshao culture. *Archaeology* **6**, 70–82 (2003).
- 15 Zohary, D. & Hopf, M. *Domestication of Plants in the Old World* (Oxford University Press, New York, 2000).
- 16 Liu, L. State emergence in early China. *Annu. Rev. Anthropol.* **38**, 217–232 (2009).
- 17 Zhang, Z. Geological environment since the late Pleistocene and the future environment evolution trend for existence in Northern China. *Quat. Sci. Rev.* **21**, 208–217 (2001).
- 18 Bar-Yosef, O. & Wang, Y. Paleolithic archaeology in China. *Annu. Rev. Anthropol.* **41**, 319–335 (2012).
- 19 Cultural Relics Institute Hebei Province. The excavation report of neolithic Jiangjialiang site, Hebei province. *Archaeology* **2**, 13–29 (2001).
- 20 Ji, D.-X. *Environmental Archaeological Perspective Dispersal of Anatomically Modern Human and Origin of Agriculture in Northern China-A Case Study in Gansu and Ningxia* (Lanzhou University, 2007).
- 21 Xue, L. The 'dragon' of Hongshan culture and the 'flower' of Yangshao culture. *J. Chifeng Univ. (Soc. Sci.)* **29**, 5–6 (2008).
- 22 Li, F. *A Research of the Neolithic Human Skeletons from Jiangjialiang, Yangyuan in Hebei Province, China* (Jilin University, 2004).
- 23 Han, J. Review of Xueshan first stage culture. *Huaxia Archaeol.* **4**, 46–54 (2003).
- 24 Institute of Archaeology, Chinese Academy of Social Sciences. Radiocarbon dating report(15). *Archaeology* **7**, 658–662 (1988).
- 25 Zhao, X., Ge, B., Zhang, Q., Cai, D., Zhou, H. & Zhu, H. Molecular biological analysis of genetic structure of human remains from Sanguan site in Hebei province. *Cultural Relics Age* **1**, 3–33 (2009).
- 26 Andrews, R. M., Kubacka, I., Chinnery, P. F., Lightowlers, R. N., Turnbull, D. M. & Howell, N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* **23**, 147 (1999).
- 27 Fan, L. & Yao, Y. G. MitoTool: a web server for the analysis and retrieval of human mitochondrial DNA sequence variations. *Mitochondrion* **11**, 351–356 (2011).
- 28 Fan, L. & Yao, Y. G. An update to MitoTool: using a new scoring system for faster mtDNA haplogroup determination. *Mitochondrion* **13**, 360–363 (2013).
- 29 Excoffier, L., Laval, G. & Schneider, S. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol. Bioinform.* **1**, 47–50 (2005).
- 30 Mannucci, A., Sullivan, K. M., Ivanov, P. L. & Gill, P. Forensic application of a rapid and quantitative DNA sex test by amplification of the X-Y homologous gene amelogenin. *Int. J. Legal Med.* **106**, 190–193 (1994).
- 31 Drobnic, K. A new primer set in a SRY gene for sex identification. *Int. Congr. Ser.* **268–270** (2006).
- 32 Karafet, T. M., Mendez, F. L., Meilerman, M. B., Underhill, P. A., Zegura, S. L. & Hammer, M. F. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res.* **18**, 830–838 (2008).
- 33 Zhao, X. *Physical Anthropological and Molecular Archaeological Research on Ancient Populations in Western Liaoning before Qin Dynasty* (Jilin University, 2009).
- 34 Liu, H. *Molecular Archaeology Study on People of Xia Dynasty Excavated from Erlitou Site, Henan province* (Chinese Academy of Sciences, 2011).
- 35 Zhao, Y.-B., Zhang, Y., Zhang, Q.-C., Li, H.-J., Cui, Y.-Q., Zhi, Xu, L. J. *et al*. Ancient DNA reveals that the genetic structure of the northern Han Chinese was shaped prior to 3 000 years ago. *PLoS ONE* **10**, e0125676 (2015).
- 36 Zhao, Y. B., Li, H. J., Li, S. N., Yu, C. C., Gao, S. Z., Xu, Z. *et al*. Ancient DNA evidence supports the contribution of Di-Qiang people to the Han Chinese gene pool. *Am. J. Phys. Anthropol.* **144**, 258–268 (2011).
- 37 Xue, F. H., Wang, Y., Xu, S. H., Zhang, F., Wen, B., Wu, X. S. *et al*. A spatial analysis of genetic structure of human populations in China reveals distinct difference between maternal and paternal lineages. *Eur. J. Hum. Genet.* **16**, 705–717 (2008).
- 38 Cui, Y. Q., Li, H. J., Ning, C., Zhang, Y., Chen, L., Zhao, X. *et al*. Y chromosome analysis of prehistoric human populations in the West Liao River Valley, Northeast China. *BMC Evol. Biol.* **13**, 216 (2013).
- 39 Li, H. J., Zhao, X., Zhao, Y. B., Li, C. X., Si, D. Y., Zhou, H. *et al*. Genetic characteristics and migration history of a bronze culture population in the West Liao-River valley revealed by ancient DNA. *J. Hum. Genet.* **56**, 815–822 (2011).
- 40 Kang, L. L., Lu, Y., Wang, C. C., Hu, K., Chen, F., Liu, K. *et al*. Y-chromosome O3 haplogroup diversity in Sino-Tibetan populations reveals two migration routes into the Eastern Himalayas. *Ann. Hum. Genet.* **76**, 92–99 (2012).
- 41 Shi, H., Dong, Y. L., Wen, B., Xiao, C. J., Underhill, P. A., Shen, P. D. *et al*. Y-chromosome evidence of southern origin of the East Asian-specific haplogroup O3-M122. *Am. J. Hum. Genet.* **77**, 408–419 (2005).
- 42 Zhong, H., Shi, H., Qi, X. B., Duan, Z. Y., Tan, P. P., Jin, L. *et al*. Extended Y chromosome investigation suggests Postglacial migrations of modern Humans into east Asia via the northern route. *Mol. Biol. Evol.* **28**, 717–727 (2011).
- 43 Gao, S. Z., Zhang, Y., Wei, D., Li, H. J., Zhao, Y. B., Cui, Y. Q. *et al*. Ancient DNA reveals a migration of the ancient Di-qiang populations into Xinjiang as early as the early Bronze Age. *Am. J. Phys. Anthropol.* **157**, 71–80 (2015).
- 44 Zheng, J. *A Study of Late Neolithic Remains in Zones Along the Yanshan Mountains* (Jilin University, 2012).
- 45 Li, H. *Y-Chromosome Genetic Diversity of the Ancient Northern Chinese Populations* (Jilin University, 2012).
- 46 Shi, H., Qi, X., Zhong, H., Peng, Y., Zhang, X., Ma, R. Z. *et al*. Genetic evidence of an East Asian origin and paleolithic northward migration of Y-chromosome haplogroup N. *PLoS ONE* **8**, e66102 (2013).
- 47 Shi, H., Zhong, H., Peng, Y., Dong, Y. L., Qi, X. B., Zhang, F. *et al*. Y chromosome evidence of earliest modern human settlement in East Asia and multiple origins of Tibetan and Japanese populations. *BMC Biol.* **6**, 45 (2008).
- 48 Zhao, Y. B., Zhang, Y., Li, H. J., Cui, Y. Q., Zhu, H. & Zhou, H. Ancient DNA evidence reveals that the Y chromosome haplogroup Q1a1 admixed into the Han Chinese 3 000 years ago. *Am. J. Hum. Biol.* **26**, 813–821 (2014).
- 49 Li, H., Huang, Y., Mustavich, L. F., Zhang, F., Tan, J. Z., Wang, L. E. *et al*. Y chromosomes of prehistoric people along the Yangtze River. *Hum. Genet.* **122**, 383–388 (2007).

Supplementary Information accompanies the paper on Journal of Human Genetics website (<http://www.nature.com/jhg>)