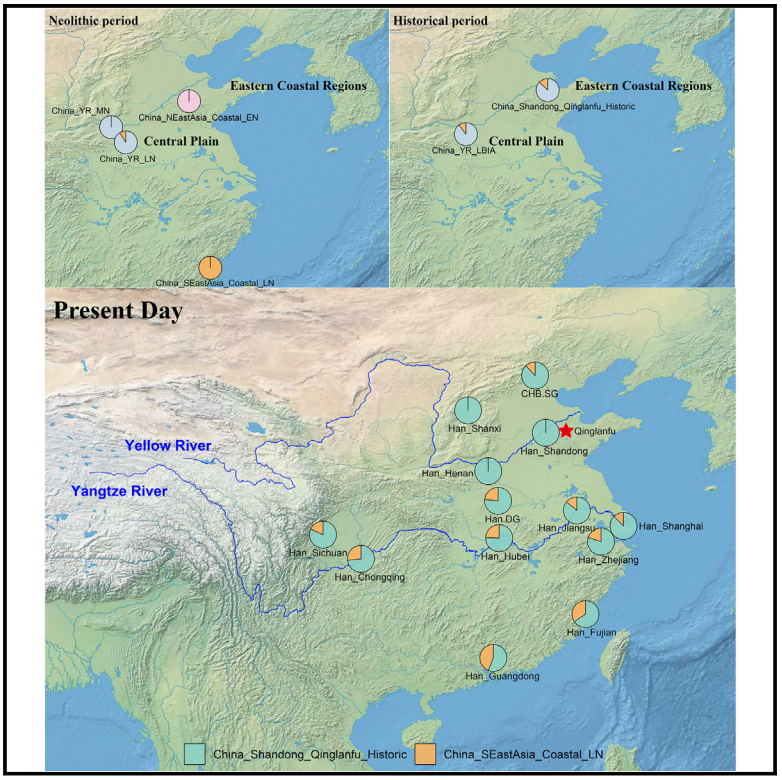
iScience

Article

Population expansion from central plain to northern coastal China inferred from ancient human genomes

Graphical abstract



Highlights

- The migration of Central Plain farmers promoted population turnover in Shandong
- Northern coastal China exhibited 2,000 years of genetic stability

Authors

Baitong Wang, Daohua Hao, Yu Xu, ..., Li Jin, Qun Zhang, Chuan-Chao Wang

Correspondence

51983964@qq.com (D.H.), stuyuxu@stu.xmu.edu.cn (Y.X.), zhangqun@whu.edu.cn (Q.Z.), chuanchaowang@fudan.edu.cn (C.-C.W.)

In brief

Human geography; Genomics; Evolutionary history; Paleogenetics

Wang et al., 2024, iScience 27, 111405
December 20, 2024
https://doi.org/10.1016/j.isci.2024.111405

CellPress

iScience

CellPress
OPEN ACCESS

Article

Population expansion from central plain to northern coastal China inferred from ancient human genomes

Baitong Wang,[1,9] Daohua Hao,[2,9,*] Yu Xu,[3,9,*] Kongyang Zhu,[3] Rui Wang,[3] Xiaomin Yang,[1] Qu Shen,[1] Mengting Xu,[3] Tianyou Bai,[3] Hao Ma,[3] Jiajing Zheng,[3] Xinyi Wang,[1] Xinyue Zou,[1] Hongming Zhou,[1] Xiaolu Mao,[1] Jiaxin Tang,[1] Yanying Peng,[1] Le Tao,[3] Haifeng He,[3] Haodong Chen,[3] Jianxin Guo,[4] Zhi Ji,[1] Yilan Liu,[3] Shaoqing Wen,[5] Li Jin,[6,7] Qun Zhang,[8,*] and Chuan-Chao Wang[1,3,7,10,*]

[1]Department of Anthropology and Ethnology, Institute of Anthropology, Fujian Provincial Key Laboratory of Philosophy and Social Sciences in Bioanthropology, School of Sociology and Anthropology, Xiamen University Xiamen 361005, China
[2]Shandong Provincial Institute of Cultural Relics and Archaeology, Jinan 250012, China
[3]State Key Laboratory of Cellular Stress Biology, School of Life Sciences, Xiamen University, Xiamen, China
[4]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650201, China
[5]Institute of Archaeological Science, Fudan University, Shanghai 200438, China
[6]State Key Laboratory of Genetic Engineering, Collaborative Innovation Center for Genetics and Development, School of Life Sciences and Human Phenome Institute, Fudan University, Shanghai 200433, China
[7]Ministry of Education Key Laboratory of Contemporary Anthropology, Department of Anthropology and Human Genetics, School of Life Sciences, Fudan University, Shanghai 200433, China
[8]Department of Archaeology, School of History, Wuhan University, Wuhan, China
[9]These authors contributed equally
[10]Lead contact
*Correspondence: 51983964@qq.com (D.H.), stuyuxu@stu.xmu.edu.cn (Y.X.), zhangqun@whu.edu.cn (Q.Z.), chuanchaowang@fudan.edu.cn (C.-C.W.)
https://doi.org/10.1016/j.isci.2024.111405

SUMMARY

The population history of the northern coastal Chinese is largely unknown due to the lack of ancient human genomes from the Neolithic to historical periods. In this study, we reported 14 newly generated ancient genomes from Linzi, one of China's densely populated and economically prosperous cities from the Zhou to Han Dynasties. The ancient samples in this study were dated to the Warring States period to the Eastern Han Dynasty (~2,000 BP). We found the samples derived all their ancestry from Late Bronze Age to Iron Age Middle Yellow River farmers rather than local Neolithic populations. They were genetically homogeneous with present-day Han Chinese of Shandong, suggesting 2,000 years of genetic stability. Our results highlight the role of the eastward migration of Yellow River farmers in the Central Plain to northern coastal China in forming the present-day genetic structure of Han Chinese.

INTRODUCTION

The Han Chinese population, with around 1.3 billion descendants, is the largest ethnic group in the world and could be roughly characterized by linguistic and genetic clines through different mixture proportions of ancient Yellow River farmers and ancient southern populations.[1–5] This suggests that the expansion of ancient Yellow River farmers, particularly the Middle Yellow River farmers, significantly shaped the modern Han Chinese population. Previous paleo-genomic studies have shown evidence of the spread of agricultural populations from the Middle Yellow River Basin, accompanied by millet farming, into various regions of China. This includes southern areas like the Yangtze River,[6] western areas, such as the Hexi Corridor,[7] and northern regions like the West Liao River Basin.[3] However, the extent of the contribution of Middle Yellow River farmers to the eastern coastal areas in China remains unknown.

Despite this, archaeological studies have provided valuable insights into the interaction between the Middle Yellow River Basin and eastern coastal regions or Lower Yellow River Basin. Lower Yellow River Basin is home to the Houli culture (8,500–7,500 BP), Beixin culture (7,000–6,100 BP), Dawenkou culture (6,000–4,000 BP), and Longshan culture (4,600–4,000 BP). Since the early Yangshao culture in the Middle Yellow River Basin, cultural exchanges between the Central Plains and the northern coastal region have gradually deepened with mutual influences.[8] The decorative style of faience of the early Dawenkou culture was influenced by Yangshao culture, and the artifacts and tooth extraction unearthed in the

late Yangshao tombs also clearly show the huge influence of Dawenkou culture.[8,9] Tooth extraction refers to the artificial removal of the alveolus of healthy permanent teeth in specific positions during growth. After extraction, the alveolus heals over time, resulting in slight shrinkage at the edges and a narrowing of the space between neighboring teeth. This custom of dental extraction originated in Beixin culture, flourished during the Dawenkou culture, and declined in Longshan culture, serving as an important symbol of the inhabitants of the Haidai area, particularly in Dawenkou culture.[10] In addition, archaeological evidence indicated interactions between the Yueshi culture in the Lower Yellow River Basin and the Erlitou cultures in the Middle Yellow River Basin in the Early Bronze Age.[11] However, due to the scarcity of ancient genomes from various periods in these regions, whether such cultural transmission was accompanied by population migration and turnover remains ambiguous.

Shandong, situated in the northern coastal area of China, holds significant importance as a territory within the Lower Yellow River Basin with a large population. Widely regarded as one of the birthplaces of ancient Chinese civilization,[12] Shandong has historically been a center of multiculturalism. Exploring the genetic structure of the people in Shandong is crucial for understanding the population history of ancient people in the northern coastal areas. Previous studies have revealed a close connection in maternal lineages between the populations in Shandong during the Longshan period and the Yangshao population in Henan.[13] Additionally, recently published genomes from the historical Shandong population, dating from the Warring States period to the Jin-Yuan Dynasties, are suggested to have descended from Late Bronze Age to Iron Age Middle Yellow River farmers.[14] However, further contemporary ancient Shandong samples are needed to confirm whether a broader movement from the central plain to northern coastal China occurred between the early Neolithic and the Historical Area or only outliers.

Here, we collected hundreds of human samples from ancient Linzi city of Shandong province from the Warring States period to the Eastern Han Dynasty (475 BC to AD 221) to fill this gap. As the capital of Qi, Linzi was densely populated and economically prosperous,[15] earning it the nickname ''the Ancient Rome of the East.'' Significant population movements have resulted in a high degree of demographic diversity in Linzi. We performed shotgun sequencing for 53 individuals, yielding 14 high-quality unrelated ancient genomes, increasing the temporal coverage in northern coastal China from around 9,700–7,800 BP in previous study to around 2,000 BP in our newly reported study.[16] These data, combined with other studies reporting Asian modern and ancient population data,[17] were used to achieve three objectives: (1) to construct the genetic structure of northern coastal Chinese in the Historical era; (2) to investigate when and which populations contribute to population turnover in northern coastal China; (3) to shed light on the evolutionary trajectory for northern coastal Chinese from Early Neolithic to Historical era and finally to present. Our results demonstrated that the genetic structure of modern northern coastal people was formed at least at 2,000 BP by the eastward migration of ancient Yellow River farmers who replaced the Early Neolithic Shandong people. After this structure was formed, it retained continuity and stability.

RESULTS

We performed shotgun sequencing for 53 individuals from Qinglanfu cemetery in Linzi, Shandong province (Data S1). We considered samples to fail filter thresholds: (1) if they did not have ancient DNA damage signatures (C to T substitution, less than 0.06); (2) if they had fewer than 30,000 targeted SNPs covered; (3) if they had ambiguous sex assignment; (4) if they showed evidence of contamination (the level of contamination greater than 0.03); (5) if they showed close relatives within second degrees among the samples. Of these 53 libraries, 14 unrelated samples passed our analysis filters (31,681–874,497 SNPs) (Data S1; Figure S1). The 14 samples from Qinglanfu cemetery showed genetic homogeneity with each other, deriving ancestry from a single source in pairwise-*qpWave* analysis (Figure S2). Therefore, we combined these samples into a single population for subsequent population genetic analysis, named China_Shandong_Qinglanfu_Historic.

Ancient genome profile for Shandong province in the historical era

The mitochondrial DNA (mtDNA) haplogroups of Qinglanfu samples were the ones prevalent in northern East Asia, including C, D4, D5, and Z, along with those dominant in East and Southeast Asia, such as M7, M10, and M11.[18] Notably, haplogroup D exhibited consistently high frequencies in Neolithic Shandong individuals dating back to before 4,600 BP, whereas haplogroups M7, C, and Z were evident in samples within 4,600 BP.[19] These findings suggest significant population movements and interactions within Shandong during historical epochs, underscoring the complex demographic dynamics of the region. The prevalent Y chromosome haplogroups of Qinglanfu samples were O1b1 and O2a2. O1b1 exhibits a higher frequency in the eastern coastal provinces and northern areas of China, whereas O2a2 is primarily found among the Han Chinese and certain populations in northern East Asia. These haplogroups have been identified in ancient populations of northern China, particularly those inhabiting the Middle and Upper Yellow River Basin,[3] emphasizing the paternal connections between the Shandong population and ancient populations of the Yellow River Basin.

To better understand the genetic structure of historical Shandong individuals, we then performed principal-component analysis (PCA) based on genome-wide SNP information that included ancient and present-day East Asian groups. All the Qinglanfu samples clustered tightly together and were projected into the genetic gradient of modern northern Han populations (including Shandong, Henan, and Shanxi) as well as overlapped with the ancient Middle Yellow River populations in the Central Plain (China_YR_LN and China_YR_LBIA) (Figures 1A and 1B). In addition, we found that Early Neolithic Shandong individuals were distant from the genetic gradient of present-day Han populations and more closely related to Northeast Asians in the PCA plot (Figure 1A) compared to historical Shandong individuals. Likewise, the results of the unsupervised ADMIXTURE analysis were consistent with those observations in the PCA plot. The Qinglanfu population was mainly composed of four ancestral components with a dark green component maximized in ancient

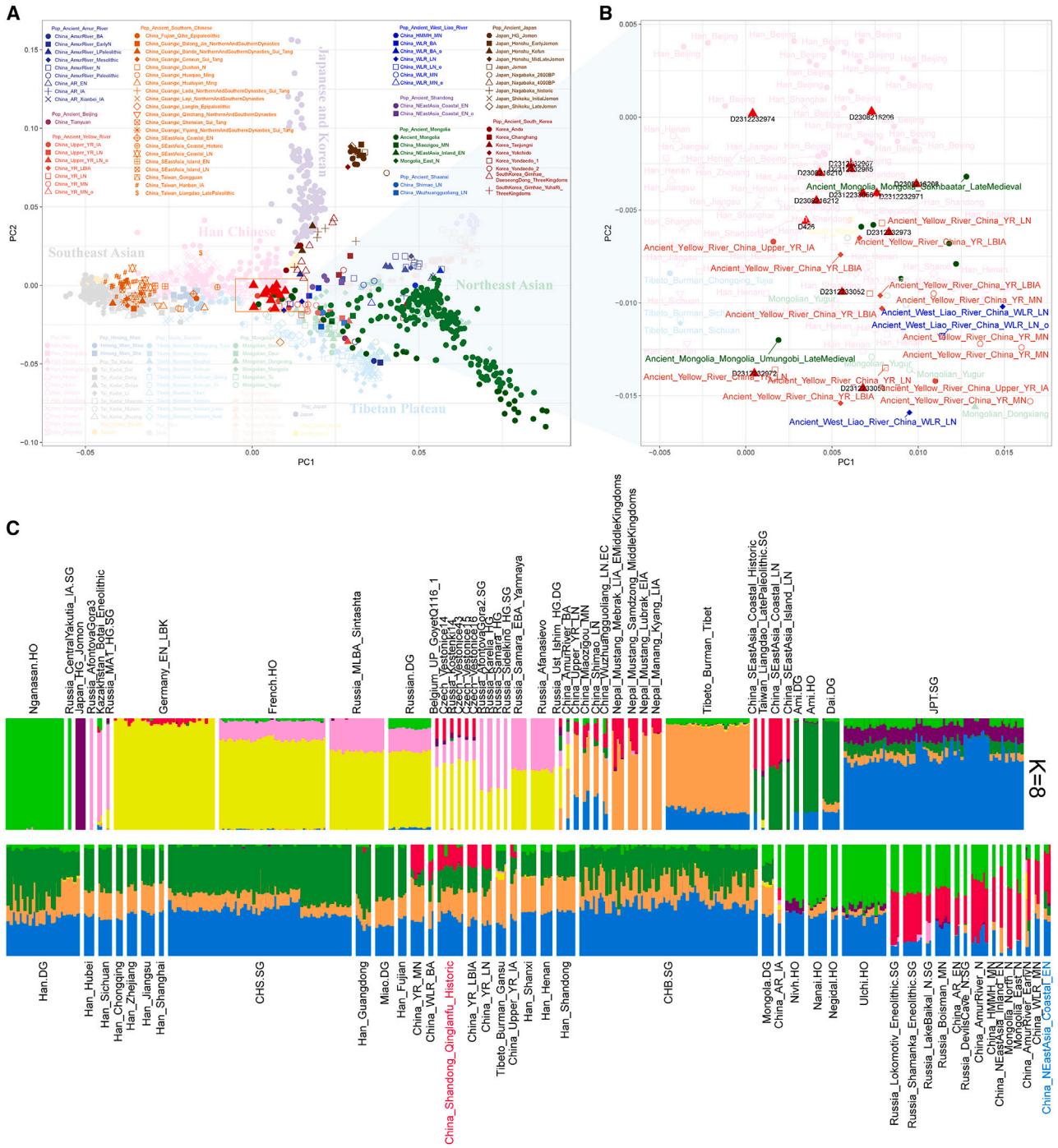


Figure 1. Genetic structure of the newly sampled Shandong individuals in the Historical era

(A) Principal-component analysis (PCA) of modern East Asian individuals and projection of ancient East Asian individuals onto the PCA plot. Dark colors indicate ancient populations, and light colors indicate modern populations. Red triangles indicate our historical Shandong samples.

(B) Interception of the detailed overlapped information among historical Shandong samples, modern Han populations, and Yellow River farmers related to Figure 1A.

(C) ADMIXTURE analysis for "HO" datasets among multiple populations, including East Asia, Southeast Asia, Northeast Asia, and Europe populations with K = 8 ancestral components (we performed this analysis for 1,933 individuals; in order to better visualize, we only present part populations here).

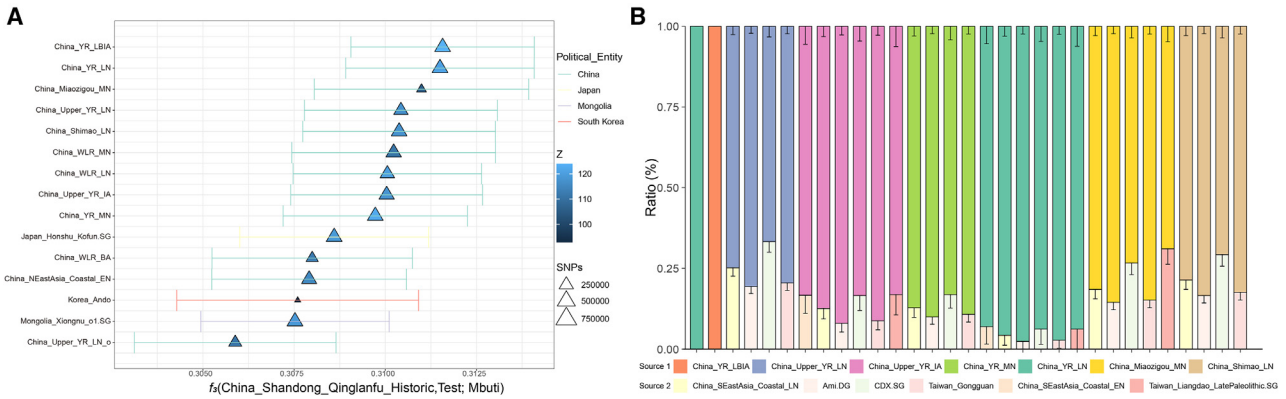


Figure 2. Qualitative relationship analysis between historical Shandong population and Yellow River populations

(A) Outgroup-f_3 test to calculate the relationship between China_Shandong_Qinglanfu_Historic and all ancient populations around the world from the "1240K" database in the form of f_3 (China_Shandong_Qinglanfu_Historic, all ancient populations from "1240K" database; Mbuti). The top 15 populations with the largest f_3 values are shown here. Data are represented as f_3 values +/− standard errors.

(B) *qpAdm* models to estimate the genetic composition of China_Shandong_Qinglanfu_Historic ancestry based on the "1240K" database. Data are represented as ratios +/− standard errors.

southern Chinese populations, an orange component maximized in highland Asian populations, a red component maximized in Northeast Asian populations, and a blue component maximized in Japanese populations (Figures 1C and S3). The composition proportions of the historical Shandong population were similar to those of ancient Central Plain populations and modern northern Han populations. In contrast, Early Neolithic Shandong individuals had a higher proportion of Northeast Asian components. The PCA and ADMIXTURE results suggested a population turnover in northern coastal China from the Early Neolithic to the Historical era.

Central plain farmers promote population turnover in Shandong

As previously mentioned, population turnover might have occurred in Shandong province. Next, we want to explore which populations contribute to this process. Outgroup-f_3 statistics, using all ancient populations worldwide, support the close genetic affinity among the historical Shandong and Middle Yellow River populations in the Central Plain, especially for China_YR_LN and China_YR_LBIA (Figure 2A). Furthermore, the results of f_4 (Mbuti, worldwide populations; Qinglanfu, Middle Yellow River populations) statistics also showed a closer relationship between the historical Shandong population and post-Late Neolithic Middle Yellow River populations, as reflected by nearly all populations contributing non-significant Z values ($-3< Z < 3$) (Data S2). However, negative results for f_4 (Mbuti, southern Chinese populations; Qinglanfu, China_YR_MN) indicated that the historical Shandong population shared more alleles with the southern East Asians than the Middle Neolithic Middle Yellow River population. Further, we also conducted f_4 (Mbuti, worldwide populations; Qinglanfu, China_NEastAsia_Coastal_EN) statistics to quantitatively test the relationship between these two populations. We found Early Neolithic Shandong individuals harbored significantly more Northeast Asian components as reflected by positive Z values ($Z > 3$), whereas historical Shandong individuals harbored significantly more southern Asian components as reflected by negative Z values ($Z < -3$) (Data S2).

We have two hypotheses to explain the formation of the observed genetic profiles in Qinglanfu samples: one is they were the descendants of a local Early Neolithic population with additional gene flow from southern populations or from Upper to Middle Yellow River populations and the other is that they were replaced by populations from the Central Plain.

We first used the pairwise-*qpWave* method to formally test whether the historical Shandong population and Yellow River farmers were genetically homogeneous and might be derived from one single genetic source (Figure S3). The results demonstrated nearly genetic homogeneity among Qinglanfu, China_YR_LN, and China_YR_LBIA (p value >0.05), but not between Qinglanfu and China_YR_MN nor between Qinglanfu and China_NEastAsia_Coastal_EN, which was consistent with previous f_4 results (Data S2). Moreover, we also found Qinglanfu, China_YR_LN, and China_YR_LBIA were consistent with being a clade according to the *qpWave* (Rank 0: p value >0.05, Data S3A). Therefore, we added southern East Asian populations as additional sources to build mixture models to account for the genetic composition of Qinglanfu samples. We found one-way models of using China_YR_LBIA or China_YR_LN as a single source that fit well to explain the genetic variations of Qinglanfu. Although, additional southern East Asian genetic components (~2.4%–7.0%) together with China_YR_LN as two sources could also be accepted in modeling the formation of Qinglanfu. However, the models with southern populations and China_YR_LBIA as two sources failed since Qinglanfu did not need more southern ancestry compared to China_YR_LBIA (Figure 2B; Data S3). As for the accepted models with China_YR_MN as one source, Qinglanfu samples were suggested to have more southern East Asian genetic components (~10%) (Data S3), which is similar to the composition ratios of modeling China_YR_LBIA and China_YR_LN using China_YR_MN and southern populations,[7] suggesting an increase of the southern ancestry through time in the Yellow River Basin.

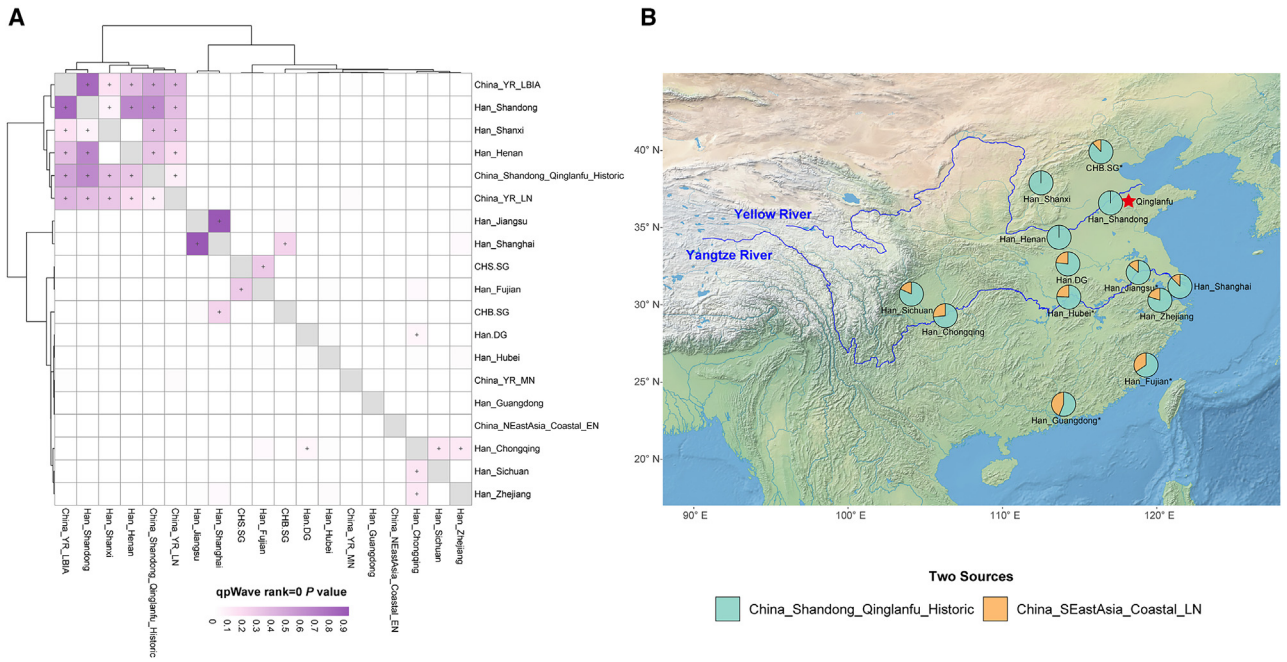


Figure 3. Qualitative relationship analysis between historical Shandong population and present-day Han populations

(A) Pairwise-*qpWave* analysis between China_Shandong_Qinglanfu_Historic and modern Han populations, using the "HO" database. "+" indicates not significant difference (*p* values >0.05).

(B) *qpAdm* modeling of ancestry related to China_Shandong_Qinglanfu_Historic and China SEastAsia_Coastal_LN in modern Han populations. The red asterisk indicates our newly reported ancient population. * indicates *p* values between 0.02 and 0.05. Made with Natural Earth. Free vector and raster map data @ naturalearthdata.com.

Furthermore, we co-analyzed the Qinglanfu population with Shandong_HE to determine if the genetic turnover in Shandong was a single outlier result for the Qinglanfu population. Our analyses showed that the two historical Shandong populations were genetically homogeneous, and both descended from the Late Bronze Age to the Iron Age of Middle Yellow River farmers (Figure S5 and Data S8A). In *qpAdm* analysis, one-way models using Shandong_HE as a single source fitted well to explain the genetic variations of Qinglanfu (Data S8B).

Finally, we tested if Qinglanfu samples could be modeled as an admixture of local Early Neolithic hunter-gatherers and southern populations or Upper to Middle Yellow River populations. However, we have not found any fitting model by trying all the available pairs of assigning local Early Neolithic hunter-gatherers and southern populations or other YR-related populations as sources (Data S5 and Data S6). These results suggested a higher probability (*p* values >0.05 for *qpAdm* and *qpWave* results; Figures 2B and 3A) that these historical Shandong people were migrants from the Middle Yellow River after the Middle Neolithic as opposed to descendants of the local Early Neolithic population with the additional gene flow from southern China and other YR-related populations.

Genetic relationship between Qinglanfu and modern Han Chinese populations

In the PCA plot, modern northern Han Chinese populations were located more closely to Qinglanfu than Middle Yellow River populations (Figure 1A). To quantitatively test the relationship between Qinglanfu and modern Han Chinese populations, we conducted f_4 statistics (Mbuti, 14 modern Han Chinese populations; Middle Yellow River populations, Qinglanfu) and found 14 Han Chinese populations could produce positive Z values compared to China_YR_MN (2.503–3.826, average: 3.00), China_YR_LN (1.044–2.255, average: 1.66), or China_YR_LBIA (0.312–2.047, average: 1.34) (Data S4A). These results suggested a higher genetic affinity between the historical Shandong samples and present-day Han Chinese populations. Therefore, we want to model modern Han populations using Qinglanfu as a proxy for ancestry sources.

We first performed pairwise-*qpWave* for all modern Han populations based on the "HO" database. We detected genetic homogeneity among historical Shandong populations (Qinglanfu and Shandong_HE) and Han Chinese in Shandong, Henan, and Shanxi (Rank 0: *p* value >0.05; Figures 3A and S5). Moreover, we also found these populations were consistent in deriving from a single source (Rank 0: *p* value >0.05; Data S4B and S8A). These results suggested genetic stability across northern Han populations since the Historical era despite experiencing frequent regime changes.[20,21] Finally, we formally modeled modern Han populations based on the *qpAdm*. Apart from the fact that Han populations from Shandong, Henan, and Shanxi could be adequately modeled as direct descendants of Qinglanfu, other Han populations could be modeled as deriving 56.1%–87.9% of ancestry from Qinglanfu samples, with the remaining from southern Chinese (Figure 3B; Data S4C).

Considering the robustness of our results, we also performed repeated analysis based on various outgroups (Data S3B and S3C) and based on the subset of bias-insensitive SNPs[22] (Data S7). All these new results were consistent with previous findings, including that Qinglanfu, China_YR_LN, and China_YR_LBIA were consistent with being a clade according to the *qpWave* (Rank 0: *p* value >0.05, Data S3A) and one-way models of using China_YR_LBIA or China_YR_LN as a single source fit well to explain the genetic variations of Qinglanfu (Data S3B).

DISCUSSION

The genetic composition of modern Han Chinese populations results from varying mixture proportions of ancient northern and southern populations, with the divide between these groups (ancient northern and southern populations) established by at least the end of the Last Glacial Maximum.[1–3,16,23] Although the genetic characteristics of Early Neolithic Shandong individuals were close to northern Chinese, more Northeast-Asian-related ancestry of these individuals cannot explain the genetic characteristics of present-day Han_Shandong individuals (Data S2). Previous studies found that maternal genetic structure increased diversity from the Neolithic to the Historical era based on the ancient mitochondrial genomes in Shandong, which suggested enhanced communication between Shandong and neighboring areas.[19] However, these studies could not solve the problem of which and when populations contributed to the formation of present-day northern coastal Chinese.

In this study, we reported 14 newly generated ancient genomes of northern coastal Chinese in the Historical era from Shandong Linzi. As the capital of Qi, Linzi experienced significant population movement and a high degree of demographic diversity, which could serve as a suitable region for studying population dynamics and interactions in cities.[15] We found these genomes were genetically homogeneous with modern Shandong individuals, deriving from a single source related to Central Plain farmers (Figure 3A; Data S4B), which suggested the present-day genetic profiles of northern coastal Chinese were formed at least at 2,000 BP and continued to maintain stability during the historical times.[20,21] We deduce that the genetic stability of northern Han populations might be associated with specific sociocultural contexts in China, such as a continuous cultural inheritance of thousands of years. Recent studies on the evolution of Chinese dialects have found that northern China shares an identical linguistic composition.[4,5] Although the limited samples from only one Qinglanfu population in our study do not rule out all potential circumstances—including its non-local nature or atypical Shandong ancestry—the additional ancient genomes from another Shandong cemetery[14] are genetically homogeneous with the Qinglanfu population, and both descended from Late Bronze Age to Iron Age Middle Yellow River farmers, suggesting a broader movement from the central plain to northern coastal China.

Previous ancient genomic studies have shown that Yellow River farmers spread to the west of China, the south of China, and the north of China to promote local population changes: from the Central Plains to the west to replace the Hexi Corridor populations in Historical era, as historical Hexi Corridor populations were genetically nearly identical to China_YR_LN and China_YR_LBIA[7]; from the Central Plains to the south to Yangtze River to form local Late Neolithic farmers[6]; from the Central Plains to the north to admix with Mongolian or promote farming in West Liao River basins.[3,20] Our results pointed to the eastward migration of ancient Yellow River farmers to northern coastal China to replace local Early Neolithic individuals. Taken together, we suggested Yellow River farmers spread rapidly from the Central Plains in all directions, replacing or admixing with local populations to form the main body of Han Chinese.

Limitations of the study

Although this study makes some progress in understanding the population history of Shandong province, the detailed population evolutionary trajectory in northern coastal China is far from clear. There were wide temporal and geographical gaps between available genomes, which currently limit our ability to describe the fine-scale genetic structure of the ancient populations in the Central Plain and northern coastal China. We can construct a more comprehensive picture of the population history of China as more ancient Chinese genomes are published in the future.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Chuan-Chao Wang (wang@xmu.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Alignment files (BAM format) have been deposited at the Genome Warehouse in the National Genomics Data Center, Beijing Institute of Genomics (China National Center for Bioinformation), Chinese Academy of Sciences and are publicly available as of the date of publication. Accession numbers are listed in the key resources table.
- This paper does not report the original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

This work was funded by the Major Project of the National Social Science Foundation of China granted to Chuan-Chao Wang (21&ZD285).

AUTHOR CONTRIBUTIONS

C.W. and Q.Z. designed the project and supervised the research and integrated resources. D.H. provided the archaeological materials. B.W., H.M., Q.S., X.W., X.Z., H.Z., X.M., J.T., Y.P., J.G., Z.J., and Y.L. assembled archaeological materials and performed dating. B.W., X.Y., L.T, H.H., and H.C. performed the ancient DNA extraction and library preparation. Y.X. performed all population genetics analyses under the supervision of K.Z., R.W., C.W., T.B., X.Y., and J.Z.; Y.X. prepared all figures and supplementary materials. S.W. and L.J. were involved in the discussion and project design. C.W., Y.X., B.W., and M.X. wrote the manuscript; C.W., Y.X., and B.W. revised the manuscript. All authors have read and approved the final version.

DECLARATION OF INTERESTS

The authors declare that they have no conflict of interest.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Archaeological information
  - Ethics statement
- METHOD DETAILS
  - Sampling, extraction of ancient DNA and library preparation
  - Radiocarbon dating
  - Ancient DNA sequence data processing
  - Contamination calculation, genetic sex determination, haplogroup assignment and kinship analysis
  - Data collection
  - Principal component analysis and ADMIXTURE analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.111405.

Received: April 8, 2024
Revised: August 6, 2024
Accepted: November 13, 2024
Published: November 18, 2024

REFERENCES

1. Xu, S., Yin, X., Li, S., Jin, W., Lou, H., Yang, L., Gong, X., Wang, H., Shen, Y., Pan, X., et al. (2009). Genomic Dissection of Population Substructure of Han Chinese and Its Implication in Association Studies. Am. J. Hum. Genet. *85*, 762–774.
2. Wang, C.-C., Yeh, H.-Y., Popov, A.N., Zhang, H.-Q., Matsumura, H., Sirak, K., Cheronet, O., Kovalev, A., Rohland, N., Kim, A.M., et al. (2021). Genomic insights into the formation of human populations in East Asia. Nature *591*, 413–419.
3. Ning, C., Li, T., Wang, K., Zhang, F., Li, T., Wu, X., Gao, S., Zhang, Q., Zhang, H., Hudson, M.J., et al. (2020). Ancient genomes from northern China suggest links between subsistence changes and human migration. Nat. Commun. *11*, 2700.
4. Xu, Y., and Wang, C.-C. (2024). Language evolution in China. Nat. Hum. Behav. *8*, 1008–1009.
5. Yang, C., Zhang, X., Yan, S., Yang, S., Wu, B., You, F., Cui, Y., Xie, N., Wang, Z., Jin, L., et al. (2024). Large-scale lexical and genetic alignment supports a hybrid model of Han Chinese demic and cultural diffusions. Nat. Hum. Behav. *8*, 1163–1176.
6. Tao, L., Yuan, H., Zhu, K., Liu, X., Guo, J., Min, R., He, H., Cao, D., Yang, X., Zhou, Z., et al. (2023). Ancient genomes reveal millet farming-related demic diffusion from the Yellow River into southwest China. Curr. Biol. *33*, 4995–5002.e7.
7. Xiong, J., Wang, R., Chen, G., Yang, Y., Du, P., Meng, H., Ma, M., Allen, E., Tao, L., Wang, H., et al. (2023). Inferring the demographic history of Hexi Corridor over the past two millennia from ancient genomes. Sci. Bull. *69*, 606–611.
8. Luan, F. (2017). Luan Fengshi Archaeological Anthology [in Chinese] (Cultural Relics Press).
9. Zhao, Y., Xiao, Y., and Zeng, W. (2019). The Westward Migration of the Residents of Dawenkou Culture Studied from Human Bones [in Chinese]. Southeast Cult., 56–65.
10. Zhao, Y., Zhang, X., Dong, W., Wang, F., and Zeng, W. (2022). Custom of dental extraction among the prehistoric residents in Haidai area. Acta Anthropol. Sin. *41*, 837–847.
11. Luan, F. (2006). The east cultural factors in Erlitou site [in Chinese]. Huaxia Archaeol., 46–53.
12. Chang, K.-C. (1987). The Archaeology of Ancient China (Yale University Press).
13. Miao, B., Liu, Y., Gu, W., Wei, Q., Wu, Q., Wang, W., Zhang, M., Ding, M., Wang, T., Liu, J., et al. (2021). Maternal genetic structure of a neolithic population of the Yangshao culture. J. Genet. Genom. *48*, 746–750.
14. Shen, Q., Wu, Z., Zan, J., Yang, X., Guo, J., Ji, Z., Wang, B., Liu, Y., Mao, X., Wang, X., et al. (2024). Ancient genomes illuminate the demographic history of Shandong over the past two millennia. J. Genet. Genom. https://doi.org/10.1016/j.jgg.2024.07.008.
15. S.P.I.o.C.R.a.A. (2013). Linzi City Site of Qi State [in Chinese] (Cultural Relics Press).
16. Yang, M.A., Fan, X., Sun, B., Chen, C., Lang, J., Ko, Y.-C., Tsang, C.-H., Chiu, H., Wang, T., Bao, Q., et al. (2020). Ancient DNA indicates human population shifts and admixture in northern and southern China. Science *369*, 282–288.
17. Mallick, S., Micco, A., Mah, M., Ringbauer, H., Lazaridis, I., Olalde, I., Patterson, N., and Reich, D. (2024). The Allen Ancient DNA Resource (AADR) a curated compendium of ancient human genomes. Sci. Data *11*, 182.
18. Tanaka, M., Cabrera, V.M., González, A.M., Larruga, J.M., Takeyasu, T., Fuku, N., Guo, L.-J., Hirose, R., Fujita, Y., Kurata, M., et al. (2004). Mitochondrial genome variation in eastern Asia and the peopling of Japan. Genome Res. *14*, 1832–1850.
19. Liu, J., Zeng, W., Sun, B., Mao, X., Zhao, Y., Wang, F., Li, Z., Luan, F., Guo, J., Zhu, C., et al. (2021). Maternal genetic structure in ancient Shandong between 9500 and 1800 years ago. Sci. Bull. *66*, 1129–1135.
20. Jeong, C., Wang, K., Wilkin, S., Taylor, W.T.T., Miller, B.K., Bemmann, J.H., Stahl, R., Chiovelli, C., Knolle, F., Ulziibayar, S., et al. (2020). A Dynamic 6,000-Year Genetic History of Eurasia's Eastern Steppe. Cell *183*, 890–904.e29.
21. Yang, X.-M., Meng, H.-L., Zhang, J.-L., Yu, Y., Allen, E., Xia, Z.-Y., Zhu, K.-Y., Du, P.-X., Ren, X.-Y., Xiong, J.-X., et al. (2023). Ancient genome of Empress Ashina reveals the Northeast Asian origin of Göktürk Khanate. J. Syst. Evol. *61*, 1056–1064.
22. Rohland, N., Mallick, S., Mah, M., Maier, R., Patterson, N., and Reich, D. (2022). Three assays for in-solution enrichment of ancient human DNA at more than a million SNPs. Genome Res. *32*, 2068–2078.
23. Mao, X., Zhang, H., Qiao, S., Liu, Y., Chang, F., Xie, P., Zhang, M., Wang, T., Li, M., Cao, P., et al. (2021). The deep population history of northern East Asia from the Late Pleistocene to the Holocene. Cell *184*, 3256–3266.e13.
24. Schubert, M., Lindgreen, S., and Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res. Notes *9*, 88.
25. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics *25*, 1754–1760.
26. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.
27. Jun, G., Wing, M.K., Abecasis, G.R., and Kang, H.M. (2015). An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data. Genome Res. *25*, 918–925.
28. Peltzer, A., Jäger, G., Herbig, A., Seitz, A., Kniep, C., Krause, J., and Nieselt, K. (2016). EAGER: efficient ancient genome reconstruction. Genome Biol. *17*, 60.

29. Skoglund, P., Northoff, B.H., Shunkov, M.V., Derevianko, A.P., Pääbo, S., Krause, J., and Jakobsson, M. (2014). Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. Proc. Natl. Acad. Sci. USA *111*, 2229–2234.

30. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L.F., and Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics *29*, 1682–1684.

31. Renaud, G., Slon, V., Duggan, A.T., and Kelso, J. (2015). Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. Genome Biol. *16*, 224.

32. Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinf. *15*, 356.

33. Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.J., Kronenberg, F., Salas, A., and Schönherr, S. (2016). HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. Nucleic Acids Res. *44*, W58–W63.

34. Poznik, G.D. (2016). Identifying Y-chromosome haplogroups in arbitrarily large samples of sequenced or genotyped men. Preprint at bioRxiv. https://doi.org/10.1101/088716.

35. Patterson, N., Price, A.L., and Reich, D. (2006). Population Structure and Eigenanalysis. PLoS Genet. *2*, e190.

36. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. *19*, 1655–1664.

37. Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient Admixture in Human History. Genetics *192*, 1065–1093.

38. Knapp, M., Clarke, A.C., Horsburgh, K.A., and Matisoo-Smith, E.A. (2012). Setting the stage – Building and working in an ancient DNA laboratory. Ann. Anatomy *194*, 3–6.

39. Meyer, M., and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb. Protoc. *2010*, t5448.

40. Ramsey, C.B., and Lee, S. (2013). Recent and planned developments of the program OxCal. Radiocarbon *55*, 720–730.

41. Reimer, P.J., Austin, W.E.N., Bard, E., Bayliss, A., Blackwell, P.G., Bronk Ramsey, C., Butzin, M., Cheng, H., Edwards, R.L., Friedrich, M., et al. (2020). The IntCal20 Northern Hemisphere radiocarbon age calibration curve (0–55 cal kBP). Radiocarbon *62*, 725–757.

42. Mittnik, A., Wang, C.-C., Svoboda, J., and Krause, J. (2016). A molecular approach to the sexing of the triple burial at the Upper Paleolithic Site of Dolní Věstonice. PLoS One *11*, e0163019.

43. Skoglund, P., Storå, J., Götherström, A., and Jakobsson, M. (2013). Accurate sex identification of ancient human remains using DNA shotgun sequencing. J. Archaeol. Sci. *40*, 4477–4482.

44. Fu, Q., Posth, C., Hajdinjak, M., Petr, M., Mallick, S., Fernandes, D., Furtwängler, A., Haak, W., Meyer, M., Mittnik, A., et al. (2016). The genetic history of Ice Age Europe. Nature *534*, 200–205.

45. Monroy Kuhn, J.M., Jakobsson, M., and Günther, T. (2018). Estimating genetic kin relationships in prehistoric populations. PLoS One *13*, e0195491.

46. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience *4*, 7.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Ancient human remains	This paper	M286_D2312233052
Ancient human remains	This paper	M289_D2312233053
Ancient human remains	This paper	M323_D2312233056
Ancient human remains	This paper	M357N_D2312232965
Ancient human remains	This paper	M417_D2312232967
Ancient human remains	This paper	M426_D426
Ancient human remains	This paper	M443_D2308216206
Ancient human remains	This paper	M515_D2308216208
Ancient human remains	This paper	M516_D2312232971
Ancient human remains	This paper	M527_D2312232972
Ancient human remains	This paper	M531_D2312232973
Ancient human remains	This paper	M548_D2308216210
Ancient human remains	This paper	M596_D2312232974
Ancient human remains	This paper	M720_D2308216212
Chemicals, peptides, and recombinant proteins		
Ethanol	Sinopharm	100092008
NaClO	Sinopharm	80010428
0.5 M EDTA, pH 8.0	Thermo Fisher Scientific	AM9262
Proteinase K	Beyotime	ST533
Guanidine hydrochloride	Sigma	G3272
Isopropanol	Sigma	34863
Acetic acid	Sigma	695092
Sodium acetate	Sigma	S7899
Tween 20	Sigma	P7947
Isothermal amplification buffer	NEB	B0537S
Deoxynucleotide (dNTP) solution mix	NEB	N0447L
Bst 2.0 DNA polymerase	NEB	M0537L
AMPure XP Beads	Beckman	A63881
Agarose	Biowest	111860
Tris-EDTA buffer solution (1003)	Sigma	T9285
Critical commercial assays		
MinElute PCR Purification Kit	QIAGEN	28006
NEBNext Ultra II DNA Library Prep Kit	NEB	E7645
Deposited data		
BAM files reported in this paper have been deposited in the GSA-Human (https://ngdc.cncb.ac.cn/gsa-human/) and are publicly available as of the date of publication.	This paper	GSA-Human: HRA008852
Genotype data	Reich Lab website	https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data.
Software and algorithms		
AdapterRemoval v2.3.3	Schubert et al.[24]	https://github.com/MikkelSchubert/adapterremoval; RRID:SCR_011834

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
BWA v0.7.17	Li et al.[25]	https://bio-bwa.sourceforge.net/; RRID:SCR_010910
SAMtools v1.18	Li et al.[26]	http://samtools.sourceforge.net; RRID:SCR_002105
bamUtil v1.0.15	Jun et al.[27]	https://github.com/statgen/bamUtil
DeDup v0.12.8	Peltzer et al.[28]	https://github.com/apeltzer/DeDup
pileupCaller	https://github.com/stschiff/sequenceTools	https://github.com/stschiff/sequenceTools
PMDtools v0.60	Skoglund et al.[29]	https://github.com/pontussk/PMDtools
mapDamage v2.2.1	Jónsson et al.[30]	https://ginolhac.github.io/mapDamage/; RRID:SCR_001240
Schmutzi	Renaud et al.[31]	https://github.com/grenaud/schmutzi
ANGSD v0.940	Korneliussen et al.[32]	http://www.popgen.dk/angsd/index.php/ANGSD; RRID:SCR_021865
HaploGrep2 v2.4.028	Weissensteiner et al.[33]	https://haplogrep.uibk.ac.at/index.html
Yhaplo v1.1.2	Poznik et al.[34]	https://github.com/23andMe/yhaplo
EIGENSOFT	Patterson et al.[35]	https://github.com/DReichLab/EIG; RRID:SCR_004965
ADMIXTURE v1.3.0	Alexander et al.[36]	https://dalexander.github.io/admixture/download.html; RRID: SCR_001263
ADMIXTOOLS (*qp3Pop*, *qpDstat*, *qpWave*, *qpAdm*)	Patterson et al.[37]	https://github.com/DReichLab/AdmixTools/; RRID:SCR_018495

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Archaeological information

Located in the middle of Shandong province, Linzi was established around 859 BC and was known as the capital city of the Qi State for more than 600 years. Since the late Western Zhou Dynasty, Linzi has become the largest craft and commercial center in China and has become one of the most prosperous Chinese cities by the Warring States period. During the Han Dynasty, along with population growth, the city's economy developed rapidly, leaving abundant cultural relics.[15]

Since the 1970s, over 30,000 tombs of the Han Dynasty have been excavated in Linzi. The large and medium-sized aristocratic tombs have intricate structures, but the majority of tombs are smaller in size and are spread out over a wide area. These small tombs, which display typical features of northern Shandong, are crucial to studying the Han Dynasty population in Linzi and surrounding areas. Numerous human skeletons excavated from Han tombs provide valuable historical insights into the Linzi population during the Han Dynasty. Nevertheless, there has been limited scholarly investigation into the human skeletal remains from these cemeteries in Linzi. The human skeletons in this study were excavated from the Qinglanfu Cemetery in Qidu Town, the eastern region of Linzi and the southern area of the ancient Linzi city of Qi. Since 2021, Shandong Provincial Institution of Cultural Relics and Archaeology has carried out an archaeological excavation in the cemetery, with the main period of the skeletons extending from the Warring States period to the Eastern Han Dynasty. Most tombs during the Warring States period were small, while the tombs in the Han Dynasty contained both small and medium-sized structures. The excavation area in 2022 covered approximately 25,000 square meters and is still ongoing, with a total of about 500 tombs to be excavated. The cemetery excavation is of great significance in exploring the genetic history of the population in Shandong during the Historical era.

Ethics statement

The research protocol was approved by the Medical Ethics Committee of Xiamen University. The provincial archaeology institute that managed the samples also approved the use of archaeological materials.

METHOD DETAILS

Sampling, extraction of ancient DNA and library preparation

This study collected human skeletal remains from the Qinglanfu site. All samples were processed in the dedicated ancient DNA clean room at the Institute of Anthropology, Xiamen University, following established precautions for working with ancient human DNA.[38] The human skeletal remains were first cleaned with 75% ethanol, and the surface was cleaned using a drill bit. Next, the

samples were washed with 10% NaClO and exposed to ultraviolet light for 30 min. We collected 80–200 mg of bone powder by drilling deep into the petrous part of the temporal bone. After adding 1 mL of 0.5 mM EDTA and 0.25 mg/mL Proteinase K to the bone powder, the mixture was agitated at 300 rpm and incubated overnight at 37°C for lysis. After centrifugation, the precipitate was discarded, and 12.5 mL of binding buffer was added to the supernatant. The binding buffer contains 5 M guanidine hydrochloride, 40% isopropanol, 25 mM sodium acetate, and 0.05% Tween-20 (Sigma Aldrich, Germany) at pH 5.5. Then, we purified samples using the MinElute kit (Qiagen, Germany) and eluted DNA extract by 0.1'TE. We used the NEBNext Ultra II DNA Library Prep Kit with an adaptor from blunt-ended ligation-based approaches[39] instead of a circular NEBNext Adaptor to prepare double-stranded libraries. Next, we purified the DNA library with the AMPure XP beads (Beckman Coulter, USA) and employed the conventional agarose gel electrophoresis method to inspect the library strips. Finally, sequencing was performed on the Illumina NovaSeq platform. All libraries were sequenced using the DNBSEQ-T7 platform to produce pair-end reads (2x150 bp for nuclear DNA).

Radiocarbon dating

One human bone sample was selected and analyzed using accelerator mass spectrometry (AMS) at Beta Analytic Testing Laboratory in Xiamen, China. The resulting 14C date was calibrated using OxCal v4.4[40] and IntCal20[41] calibration curves. The individual was dated in the Han Dynasty period (50 BC - 84 AD), consistent with the spanning time estimated from archaeological stratigraphic layers and excavated grave goods.

Ancient DNA sequence data processing

We first performed AdapterRemoval v2.3.3[24] to trim adapters, filter low-quality reads and merge the pair-end sequence reads. Merged reads were mapped to the human reference genome hs37d5 (GRCh37 with decoy sequences) using BWA v0.7.17,[25] with "aln -l 1024 -n 0.01" parameters. PCR duplicates were removed using DeDup v0.12.8.[28] Low mapping quality reads (MAPQ < 30) were removed by SAMtools v1.18.[26] PMDtools v0.60[29] and mapDamage v2.2.1[30] were performed to calculate the ancient DNA damage patterns, and BamUtil v1.0.15[27] was used to cut the ends of each mapped read from bam files according to the DNA damage patterns to ensure that the transition rates at the ends of reads <2% (Data S1). These trimmed "bam" files were used to generate pseudo-haploid calls within the "1240k" panel and "HO" panel by pileupCaller (https://github.com/stschiff/sequenceTools), with "–randomHaploid" parameter.

Contamination calculation, genetic sex determination, haplogroup assignment and kinship analysis

The contamination rate for each sample was evaluated by Schmutzi[31] and ANGSD v0.940.[32] Three different methods, Rx,[42] Ry[43] and comparison of genome coverage between X and Y chromosomes,[44] were performed to determine genetic sex for each sample. Sex determination is only assumed to be accurate if it is consistent across methods. The mitochondrial haplotype for each sample was identified by Haplogrep2 v2.4.028, and the Y-chromosome DNA haplogroup for each sample was determined by Yhaplo v1.1.2.[34] The genetic kinship between ancient individuals was detected by READ.[45]

Data collection

Our data were merged with the latest version of Allen Ancient DNA Resource v54.1.p1 datasets ("Human Origin" dataset and "1240k" dataset)[17] using mergeit implemented in EIGENSOFT. The filtered thresholds for these published datasets are as follows: (1) we only considered the data in column 34 of the ".anno" file where the assessment was "PASS"; (2) we filtered the low coverage data with SNP sites below 15,000; (3) we only considered the data that did not have the relatives or duplicates detected in these datasets; (4) the outlier samples for modern populations were excluded from further analysis. The subset of bias-insensitive SNPs obtained from the last column of the annotation in Data S1 (PassFilterForMetaAnalysisBias = 1) of the research work for Nadin Rohland et al.[22]

Principal component analysis and ADMIXTURE analysis

Smartpca v18140[35] was performed to do Principal Component Analysis (PCA), using the "HO" dataset as more modern populations in this dataset, with "lsqproject: YES altnormstyle: NO numoutlieriter: 0" parameters. Modern populations were used to calculate the principal components (PCs), and ancient individuals were projected onto the PCs. We pruned the linkage disequilibrium for the "HO" dataset by plink v1.90,[46] with "–indep-pairwise 200 25 0.4 –allow-no-sex" parameters. The filtered "HO" dataset was used for unsupervised admixture analysis by ADMIXTURE v1.3.0[36] with K ranging from 2 to 11.

QUANTIFICATION AND STATISTICAL ANALYSIS

We performed outgroup-f_3 statistics by qp3Pop v651, using the "1240k" dataset as more SNP sites in this dataset, with "inbreed: YES" parameters. We performed f_4 statistics by qpDstat v980, using the "1240k" dataset, with "f4mode: YES printsd: YES inbreed: YES" parameters. For f_3 statistics, we only considered overlapping SNP sites > 40,000 and individuals within a population > 1. Both software programs are implemented in ADMIXTOOLS.[37]

We performed *qpWave* analysis by *qpWave* v1520 and *qpAdm* analysis by *qpAdm* v810, using both the "HO" and "1240k" datasets. For *qpWave* analysis, the outgroup was "Mbuti.DG, Russia_MA1_HG.SG, China_Tianyuan, Mongolia_North_N, China_Guangxi_Dushan_N, Taiwan_Hanben_IA, Tibeto_Burman_Tibet, Nganasan.HO" for "HO" dataset related analysis, and the outgroup was "Mbuti.DG, Russia_MA1_HG.SG, China_Tianyuan, Mongolia_North_N, Russia_CentralYakutia_LN.SG, China_Guangxi_Dushan_N, Taiwan_Hanben_IA, Nepal_Mustang_Lubrak_EIA" for "1240k" dataset related analysis. For *qpAdm* analysis, the outgroup was "Mbuti.DG, Russia_MA1_HG.SG, China_Tianyuan, Mongolia_North_N, Russia_CentralYakutia_LN.SG, China_Guangxi_Dushan_N, Taiwan_Hanben_IA, Nepal_Mustang_Lubrak_EIA". Both software programs are implemented in ADMIXTOOLS.[37]