

The genetic structure and admixture of Manchus and Koreans in northeast China

Na Sun, Le Tao, Rui Wang, Kongyang Zhu, Xiangjun Hai & Chuan-Chao Wang

To cite this article: Na Sun, Le Tao, Rui Wang, Kongyang Zhu, Xiangjun Hai & Chuan-Chao Wang (2023) The genetic structure and admixture of Manchus and Koreans in northeast China, *Annals of Human Biology*, 50:1, 161-171, DOI: [10.1080/03014460.2023.2182912](https://doi.org/10.1080/03014460.2023.2182912)

To link to this article: <https://doi.org/10.1080/03014460.2023.2182912>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



[View supplementary material](#)



Published online: 18 May 2023.



[Submit your article to this journal](#)



Article views: 3584



[View related articles](#)



[View Crossmark data](#)

RESEARCH PAPER

OPEN ACCESS



The genetic structure and admixture of Manchus and Koreans in northeast China

Na Sun^a, Le Tao^b, Rui Wang^b, Kongyang Zhu^b, Xiangjun Hai^c and Chuan-Chao Wang^{b,d,e,f,g}

^aCollege of Foreign Languages, Huaqiao University, Quanzhou, China; ^bState Key Laboratory of Cellular Stress Biology, School of Life Sciences, Xiamen University, Xiamen, China; ^cKey Laboratory of Environmental Ecology and Population Health in Northwest Minority Areas, Northwest Minzu University, Lanzhou, China; ^dDepartment of Anthropology and Ethnology, Institute of Anthropology, School of Sociology and Anthropology, Xiamen University, Xiamen, China; ^eState Key Laboratory of Marine Environmental Science, Xiamen University, Xiamen, China; ^fInstitute of Artificial Intelligence, Xiamen University, Xiamen, China; ^gMinistry of Education Key Laboratory of Contemporary Anthropology, Department of Anthropology and Human Genetics, School of Life Sciences, Fudan University, Shanghai, China

ABSTRACT

Background: The fine-scale genetic profiles and population history of Manchus and Koreans remain unclear.

Aim: To infer a fine-scale genetic structure and admixture of Manchu and Korean populations.

Subjects and methods: We collected and genotyped 16 Manchus from Liaoning and 18 Koreans from Jilin province with about 700K genome-wide SNPs. We analysed the data using principal component analysis (PCA), ADMIXTURE, Fst, TreeMix, *f*-statistics, *qpWave*, and *qpAdm*.

Results: Manchus and Koreans showed a genetic affinity with northern East Asians. Chinese Koreans showed a long-term genetic continuity with Bronze Age populations from the West Liao River and had a strong affinity with Koreans in South Korea and Japan. Manchus had a different genetic profile compared with other Tungusic populations since the Manchus received additional genetic influence from the southern Chinese but didn't have West Eurasian-related admixture.

Conclusions: The genetic formation of Manchus involving southern Chinese was consistent with the extensive interactions between Manchus and populations from central and southern China. The large-scale genetic continuity between ancient West Liao River farmers and Koreans highlighted the role farming expansion played in the peopling of the Korean Peninsula.

ARTICLE HISTORY

Received 17 July 2022

Revised 26 December 2022

Accepted 27 January 2023

KEYWORDS

Genetic structure; population history; Manchu, Korean

Introduction

Manchu is the third largest ethnic minority group in China. Manchus were mainly distributed in the northeast of China, especially in Liaoning and Jilin Province. In addition, Manchus were scattered all over China, including Gansu, Guizhou, Hunan, and other provinces. Historically, the origins of the Manchu can be traced back to the Sushen in the Zhou Dynasty, Yinglou in the Han Dynasty, Wuji in the Southern and Northern Dynasties, Mohe in the Sui and Tang Dynasties, and Jurchen in the Jin Dynasty. During the Ming Dynasty, Jianzhou Jurchen developed into today's Manchu. The ancestral language of the Manchu is the Jurchen language. The present-day Manchu language belongs to the Manchuric branch of the Tungusic family, and they have the largest population in the Tungusic family. Tungusic languages are spoken in Eastern Siberia and Manchuria by Tungusic peoples, including Manchu, Xibo, Oroqen, Hezhen,

Nanai, Ulchi, Evenk, Negidal, and so on. Archaeolinguistic study suggested that the homeland of proto-Tungusic was the region around Lake Khanka, the far east part of the border between Russia and China (Wang and Robbeets 2020).

Previous genetics studies of Manchus, including those looking at autosomal short tandem repeats (STRs), Y-haplotype, and X-STRs (Xue et al. 2005; Liu et al. 2013; Xing et al. 2019), suggested there were small genetic distances between Northern Hans and Manchus. The mtDNA suggested Manchus had the admixture signal from south and north East Asia and showed a close genetic relationship with the neighbouring Mongolians, Koreans, and the Liaoning Han Chinese (Zhao et al. 2011). From the paternal Y-chromosome side, the dominant type in Manchus is O2-M122 with a proportion of 42.6% (Katoh et al. 2005). The Rst values calculated by Y-STRs show the Manchus were significantly different from some Chinese populations like Tibetans and Uyghurs, but have a close affinity with Northern Hans and

CONTACT Le Tao ✉ taole@stu.xmu.edu.cn State Key Laboratory of Cellular Stress Biology, School of Life Sciences, Xiamen University, Xiamen, China; Xiangjun Hai ✉ yxhxj@xbmu.edu.cn Key Laboratory of Environmental Ecology and Population Health in Northwest Minority Areas, Northwest Minzu University, Lanzhou, China; Chuan-Chao Wang ✉ wang@xmu.edu.cn Department of Anthropology and Ethnology, Institute of Anthropology, School of Sociology and Anthropology, Xiamen University, Xiamen, China

Na Sun and Le Tao contributed equally to this work and should be considered the co-first authors.

Supplemental data for this article can be accessed online at <https://doi.org/10.1080/03014460.2023.2182912>.

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

Mongolians (Kato et al. 2005; Xue et al. 2006; He and Guo 2013; Bai et al. 2016; Atif et al. 2019). The qualitative and quantitative analyses of the high-density genome-wide SNP data of 93 Manchus collected from Xinbin, Liaoning, showed a large-scale admixture with northern Han Chinese (Zhang et al. 2021). The Manchu can be modelled as a mixture of the northern populations represented by the ancient Mohe and Xianbei people and the southern populations represented by the Iron Age Taiwan samples (Zhang et al. 2021).

Unlike Manchus, Koreans constitute a cross-border ethnic group and comprise the main nationality of the Korean Peninsula. The largest inhabited area of the Chinese Koreans is the Yanbian Korean Autonomous Prefecture in Jilin Province. From a linguistic view, Korean is considered similar to Japanese as an isolated language of unknown phylogeny (Song 2005). Korean has similar factors to Japonic, Tungusic, and Mongolic languages in terms of grammatical features. Just like Japanese and Manchu, Korean has many Chinese loanwords. However, the subject-object-verb structure of Korean is the same as Manchu, which is different from the subject-verb-object structure of Chinese.

From an mtDNA perspective, Koreans display a typical East Asian profile (Jin et al. 2009). Y-chromosome haplotype O2b-SRY465 suggested the origin of proto-Korean can be traced back to northeastern China during the Neolithic (9,900-10,000 years BP) and Bronze Ages (3,450-2,350 years BP) (Kim et al. 2011). Kim et al. reported 88 modern Korean genomes and found two major genetic components of East Siberia and Southeast Asia (Kim et al. 2020). Modern Koreans can be best described as an admixture of Neolithic Northeast Asians and the Iron Age Southeast Asians. Gelabert et al. reported the first paleogenomics data from Korea, which traced back to the 4th-7th CE of Korea's Three Kingdoms period. These ancient genomes can be modelled as an admixture between a Bronze Age northern Chinese genetic source (Yellow_River LBIA or Liao_River BA) and a Jomon-related ancestry (Gelabert et al. 2022). However, there is no genomic analysis reported on Korean ethnicity in China. We note that the formation of populations is not subjected to

national borders, but it's interesting to investigate the genetic diversity of Chinese Koreans.

Here, we reported the genome-wide SNP data of 16 Manchu individuals living in Jinzhou, Liaoning, and 18 Korean individuals living in Antu County of Yanbian Korean Autonomous Prefecture. Zhang et al. (2021) collected and genotyped Manchu individuals in Xinbin County, eastern Liaoning Province, but in this study, we collected and genotyped the Manchu samples from Jinzhou city in central-west Liaoning Province. Since Manchu has a large population in Liaoning Province, it is better to genotype more samples to infer the possible genetic substructure within Manchu. In addition, Jinzhou is located east of the Liaodong Corridor, which has historically connected most of the land transport between north China and northeast China. The ethnic memory of native Manchus in Beizhen, Jinzhou, recorded their homeland as Changbai Mountain (Supplementary Figure S1). Antu County is an ethnic Korean autonomous region at the foot of the Changbai Mountains. We sampled these two places in the hope of providing more detailed information on migration and admixture events of ethnic minorities in northeast China.

Subject and methods

Sampling and genotyping

We collected saliva samples from 16 Manchus in Heishan and Beizhen (two adjacent sites) County in Jinzhou city, Liaoning province (group label "Manchu_Jinzhou" in the following analysis), and 18 Koreans in Erdaobaihe Town, Antu County, Yanbian Korean Autonomous Prefecture, and Jilin Province (group label "Korean_Antu" in the following analysis) (Figure 1). All samples were taken with informed consent. The studies involving human participants were reviewed and approved by the Medical Ethics Committee of Xiamen University (Approval Number: XDYX2019009). Genomic DNA was extracted by QIAamp DNA Blood Mini kit (QIAGEN, Hilden, Germany) and quantified by the Nanodrop-2000 spectrophotometer (Thermo Scientific, Wilmington, DE,

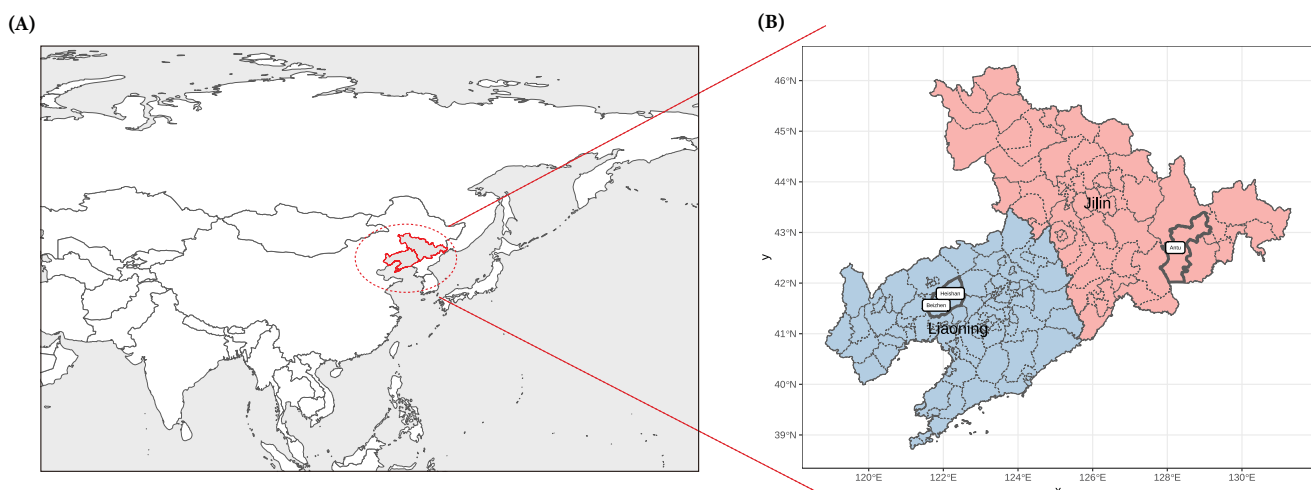


Figure 1. Overview of sampling sites in this study. (A) Geographical view of the sampling location in East Asia. (B) Geographical details of the sampling sites of Liaoning and Jilin Provinces.

United States). We genotyped DNA samples with the Illumina WeGene Arrays containing 717,228 single nucleotide polymorphisms (SNPs). We used KING software (Manichaikul et al. 2010) to remove individuals with third degrees with other samples to guarantee that the samples in the population analysis were unrelated. Finally, 15 Manchus and 18 Koreans were reserved.

Merging data

We merged our newly collected samples with genome-wide SNP data from previously published modern and ancient populations (Patterson et al. 2012; Liu et al. 2020; Ning et al. 2020; Mao et al. 2021; Robbeets et al. 2021; Wang et al. 2021; Gelabert et al. 2022). Data merging was conducted by the *mergeit* program in EIGENSOFT.

Here we generated three reference datasets for our later genetic analysis:

1. The 1240K dataset contains both ancient and modern human individuals from either shotgun sequencing data or in-solution target capture at 1,233,013 sites from the David Reich laboratory (<https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data>). 206,105 SNP sites were reserved after merging. This “1240k-merged” dataset was used in *qpWave* and *qpAdm* analysis.
2. The above 1240K dataset merged with present-day individuals typed on the Human Origins arrays (HO) with 597,573 sites. This “HO-merged” dataset with 69,659 SNP sites contains more modern individuals and is used in *f*-statistics, and *Treemix* analysis.
3. We merged previously published genome-wide data from Guizhou and Liaoning (Chen et al. 2021; Zhang et al. 2021) with the above 1240K and HO dataset. After data merging, 42,912 sites remained. This dataset was used in PCA and ADMIXTURE to calculate the *F*_{st} values.

Principal component analysis (PCA)

We conducted PCA using the *smartpca* package implemented in EIGENSOFT software (Patterson et al. 2006). We calculated the principal components (PC1 and PC2) using modern reference populations and then projected ancient individuals onto PC1 and PC2 using the *lsqproject*: YES option.

ADMIXTURE analysis

We first pruned the linkage disequilibrium SNPs on our dataset using PLINK 1.9 (Purcell et al. 2007) by parameters “-indep-pairwise 200 25 0.4.” Finally, 4611 of 42,192 variants were removed. We ran ADMIXTURE from *K*=2 to *K*=8 and checked the cross-validation error (CV error) of each one. We chose the best-fitting model with the minimum CV error (*K*=4) and visualised the proportion of each individual by an in-house script.

*F*_{st}

We merged previously published Manchu data to calculate the *F*_{st} values, which measure population differentiation and genetic distance. We used the *smartpca* program in EIGENSOFT (Patterson et al. 2006) with default parameters and option *f*stonly: YES.

Treemix

Treemix v1.13 (Pickrell and Pritchard 2012) was used to construct a maximum likelihood tree for the set of populations. An African population Mbuti was considered as the outgroup and set as the root population *via* option: -root Mbuti.

Three-population test (*f*₃ statistics)

Two different forms of *f*₃ statistics were used in our subsequent analysis by the *qp3Pop* program in the AdmixTools (Patterson et al. 2012).

First, we conducted outgroup-*f*₃-statistics in the form of *f*₃ (Studied population, reference population; outgroup). The outgroup-*f*₃-statistics calculated the shared genetic drift between our studied population and chosen reference population since their divergence from the outgroup. The African population Mbuti was used as the outgroup.

Following this, the admixture-*f*₃-statistics in the form of *f*₃ (Source1, Source2; Studied population) were performed to test whether our studied population was an admixture for the selected reference populations Source1 and Source2. The significant negative *f*₃ values with a Z-score < -3 indicated that two source populations could be related to the genetic ancestors of the studied population.

Four-population test (*f*₄ statistics)

The *f*₄ statistics in the form of *f*₄ (Mbuti, B; C, D) were calculated by *qpDstat* program in AdmixTools (Patterson et al. 2012) with parameter *f*4mode: YES. We used Mbuti as the outgroup. The significant positive *f*₄ statistic with Z-score greater than 3 indicated a possible gene flow between the populations B and D. On the contrary, a significant negative statistic *f*₄ (Z-score < -3) indicated a possible gene flow between population C and D. The *f*₄ value that is close to 0 with absolute value Z-score less than 3 indicated B shared a similar number of alleles with both C and D.

Admixture coefficient modelling by qpAdm

We modelled our studied groups using the *qpWave* and *qpAdm* programs in AdmixTools (Patterson et al. 2012) with option *allsnps*: YES. The “1240k-merged” dataset mentioned above was used in this analysis. The following ten populations in the 1240k dataset were used as outgroups: an African population (Mbuti, *n*=4), an indigenous Taiwan group (Atayal, *n*=1), an Oceania population (Papuan), a Bronze Age Steppe population (Russia_Afanasievo, *n*=23), Andamanese islanders (Onge, *n*=15), a late Neolithic

population in southeast Asia (Malaysia, $n=1$), a late Palaeolithic Chinese ancient individual (Tianyuan, $n=1$), hunter-gathers from Japanese archipelago (Japan_Jomon, $n=6$), and Neolithic populations from the Amur River Basin (AR_EN, $n=2$) and Yellow River (Miaozigou_MN, $n=3$).

We first used ancient farmers from Northeast Asia (represented by WLR_BA) or the Yellow River (represented by YR_MN) and some ancient Koreans as one single source in the modelling. Based on the result of ADMIXTURE and f statistics, we then added Iron Age southern East Asians (represented by Taiwan_Hanben) as sources for the two-way admixture model. Considering historical factors, we also explored if ancient populations Heishui_Mohe and Mongolia_XiongNu had contributed to our studied populations in a three-way admixture model.

Results

Population genetic structure

To qualitatively infer the population genetic structure, we first performed PCA (Figure 2). We observed two genetic clines corresponding well with geography and linguistics in

the first two principal components (dotted wire circle). The north-south gradient cline of East Asia included Tai-Kadai, Austronesian, Austroasiatic, Hmong-Mien, Han Chinese and some Tibetan-Burman. The other cline mainly had Tungstic and Mongolic populations in Northeast Asia. Our studied Manchu and Korean samples were located between these two clines. Manchus clustered closely with north Han Chinese and some Yugurs, while Koreans in Antu (Korean_Antu) clustered closely with Koreans from South Korea, Japanese, and the Bronze Age populations from the West Liao River.

We further performed an unsupervised model-based ADMIXTURE clustering analysis and observed the lowest cross-validation error at $K=4$ (Figure 3 and Supplementary Figure S2). Both the studied groups, Manchus and Koreans, consisted of four ancestral components, which were enriched in Tibetans or Hans (pink and yellow), Southeast Asians (yellow) and Northeast Asians (orange). The Manchus showed a similar genetic profile with northern Han Chinese, while the Korean_Antu samples were genetically similar to Koreans from South Korea, which is consistent with the PCA results. The studied Korean_Antu contained more orange

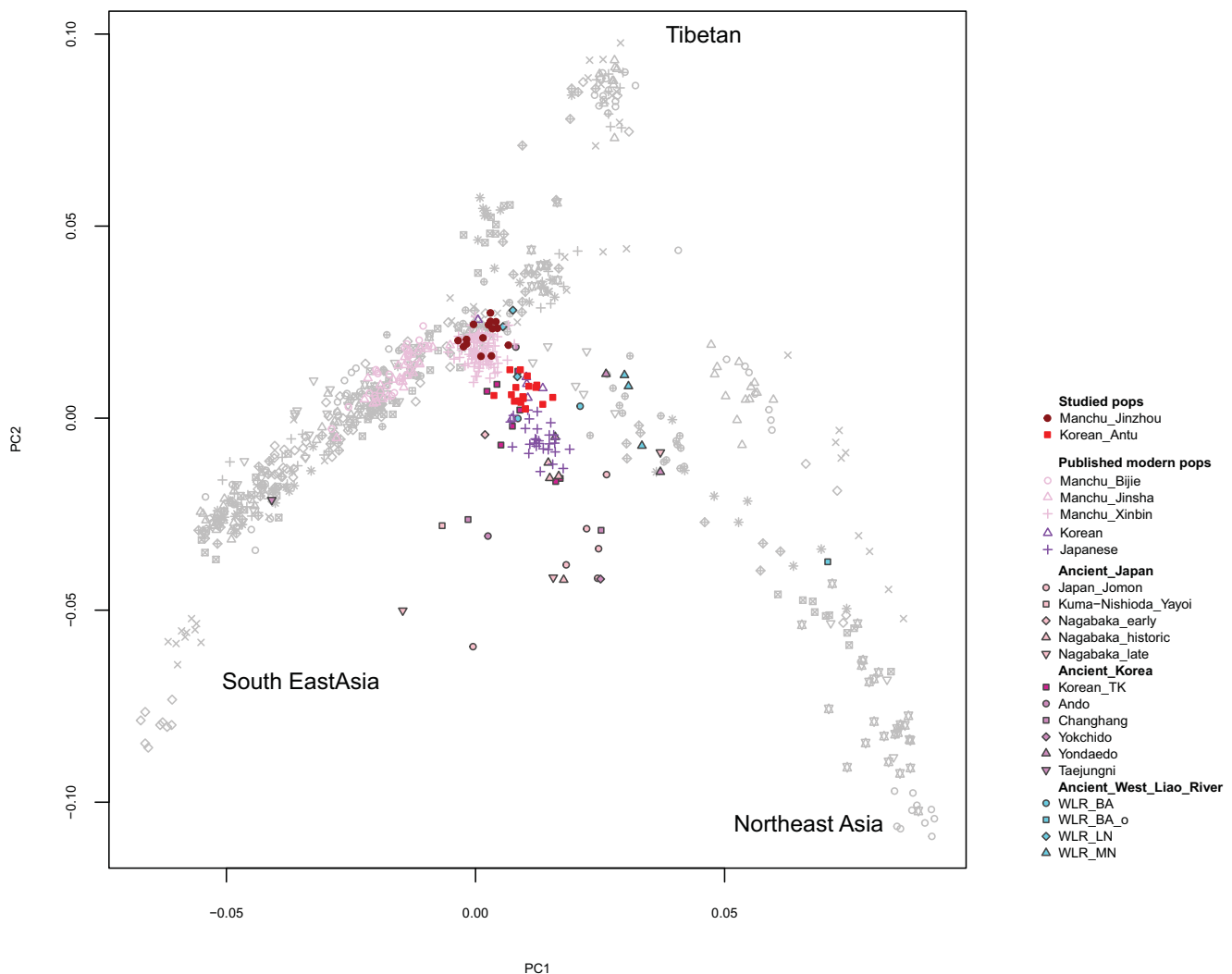


Figure 2. Genetic structure of ancient and present-day populations included in this study. Principal component analysis (PCA) of ancient individuals projected onto modern East Asia. More details are shown in Figure S3.

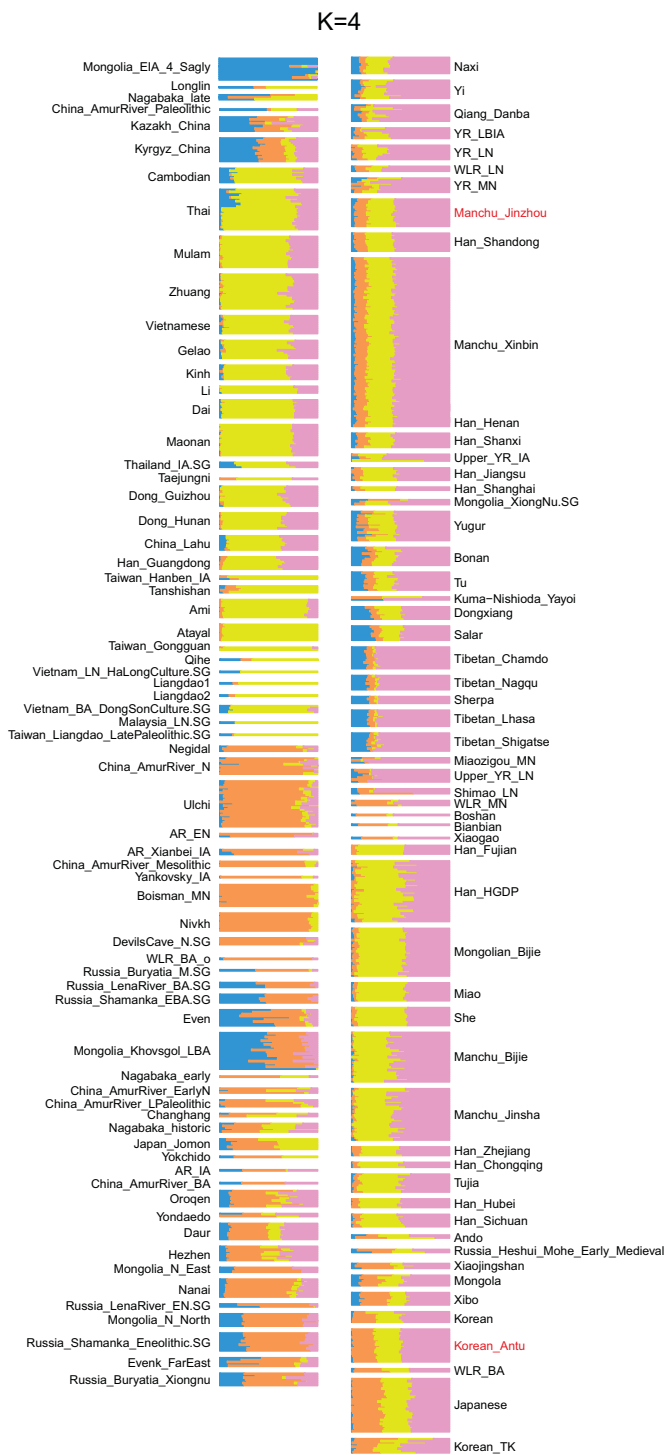


Figure 3. Results of ADMIXTURE plot for ancients and modern East Asians for $K=4$. The cross validation (CV) is lowest when $K=4$ (Supplementary Fig. S2B), the lowest CV error correlates to the best unsupervised model. Other models ($K=2$ to $K=8$ are shown in S Supplementary Figure S2A).

components than Manchus in Jinzhou, which indicated Koreans might have more Northeast Asia-related ancestry.

Population relationships between reference modern East Asians

Manchu is a geographically widespread minority. We observed a genetic substructure in Manchus in northern and

southern China. Manchus in Liaoning in northeast China (Manchu_Jinzhou and Manchu_Xinbin) and Manchus in Guizhou in southwest China (Manchu_Bijie and Manchu_Jinsha) were genetically heterogeneous (Figure 4). In the PCA plot (Figure 2), our studied Manchus clustered with Manchu_Xinbin but clearly split from Manchus in Guizhou. Guizhou Manchus are closer to the southern Han, Hmong-Mien and local Guizhou Mongolians (Mongolian_Bijie).

Consistent with the PCA and ADMIXTURE results, both Koreans and Japanese show high outgroup- f_3 values and low F_{st} values with studied Korean_Antu, indicating a genetic affinity of Korean_Antu with Japanese and Koreans. Both the outgroup- f_3 value and F_{st} value indicated a close relationship between studied Manchus and Koreans. In addition, Koreans, Japanese, and studied Korean_Antu cluster together on the Maximum Likelihood (ML) tree inferred by Treemix. We also detected a possible gene flow from Tungusic-speaking populations in northeast Asia to the Korean-Japanese cluster (Figure 5).

Compared to some Turkic and Tungusic speaking groups in northern East Asia (such as Kazakh_China, Kyrgyz_China, Even) (Figure 6(A,B), Supplementary Table S2), Manchus and Korean_Antu share more genetic drift with Han Chinese, Hmong-Mien, Tai-Kadai, Japanese, and Koreans as shown in outgroup- f_3 and pairwise f_4 statistics in the form of f_4 (Mbuti, Manchu/Korean; EastAsia1, EastAsia2). Manchus were genetically different from other Tungusic-speaking populations as shown in the values of f_4 (Outgroup, X; Manchu_Jinzhou, other Tungusic) where X represents an East Asian population (Figure 7(A)). Similar patterns can be seen in Korean_Antu (Supplementary Figure S3). The f_4 value shows a closer relationship between Southeast or East Asia and Manchus compared to other Tungusic groups, which we suspect was caused by the gene flow from West Eurasians into other Tungusic groups and the genetic influence on Manchus from southern China. We used Bronze Age Russia_Afanasievo pastoralists from the Eurasian Steppe as a western source to infer the possible influence of West Eurasians in studied Manchus and three Tungusic populations Hezhen, Oroqen, and Xibo (Supplementary Table S4). We have not detected significant evidence of West Eurasian-related influence in Manchus but found small proportions of western ancestry (4%~6%) in the other three Tungusic populations.

Population relationships between ancient populations

Manchus and Koreans have the closest affinity with ancients in the Yellow River (YR) and Yankovsky_IA (ancients from the Iron age far east) as shown in outgroup f_3 (Studied population, Ancient, Mbuti) (Figure 6(C,D)). The studied populations share more genetic drift with the ancient Yellow River and West Liao River basins than with Tibetan, Hmong-Mien, Tai-Kadai, and other southern and northeastern populations (Figure 7(B,C)). However, Manchus and Koreans shared similar amounts of genetic drift with the ancient populations of the Yellow River and West Liao River basins compared to the Han Chinese.

Koreans are spread across the Korean Peninsula and north-east China. We found Koreans in South Korea are genetically

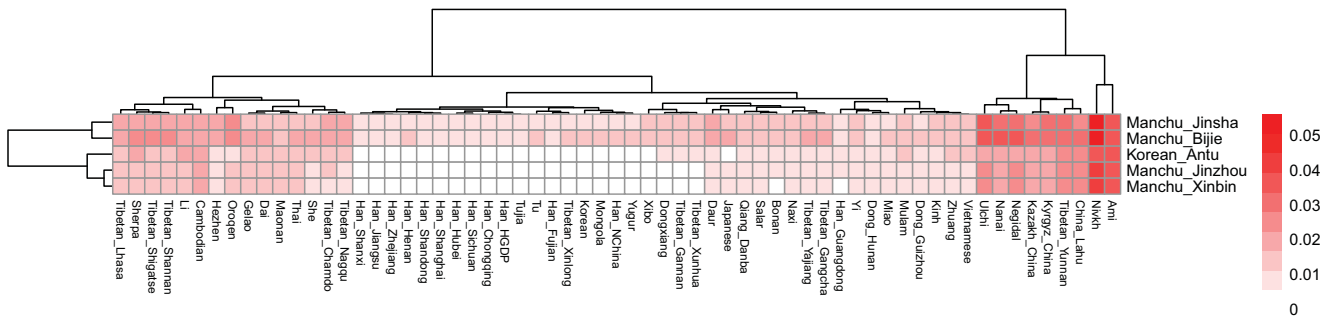


Figure 4. Fst value between Manchus and studied Korean with modern East Asia. The smaller the diversity between groups, the smaller the Fst value.

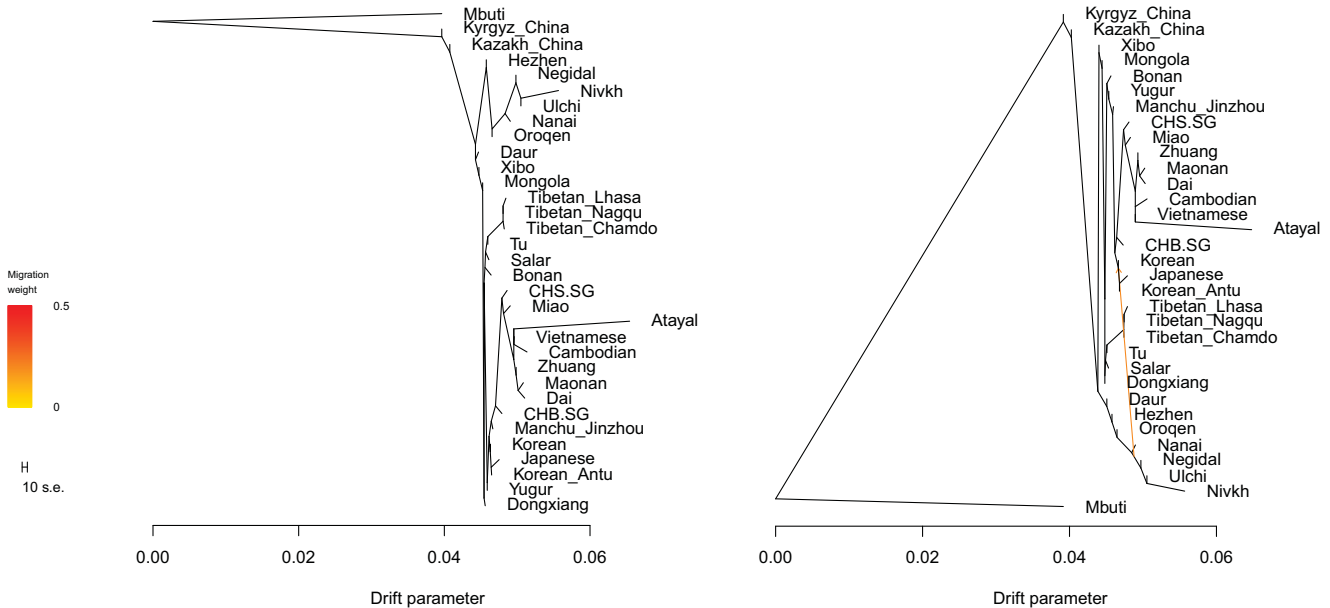


Figure 5. Treemix results for 0 and 1 migration event.

homogeneous with our studied Koreans. On the other hand, historical Koreans (Korean_TK) share more genetic drifts with our studied Koreans. However, other ancient Koreans across the Neolithic have no significant difference in shared genetic drifts compared to other East Asians (Figure 7(C)).

Considering the observed close relationship with Han Chinese, we focus on the Han population in f_4 (Mbuti, ancients; Manchu/Korean, East Asia) (Figure 7(B,C)). Ancient populations of East Asia shared a similar amount of alleles with Manchus and northern Han from Henan, Shanxi, and Shandong provinces. But ancient populations from Southeast Asia, such as Malaysia_LN, shared more alleles with southern Han from Sichuan and Fujian provinces than Manchu. The f_4 statistics gave significant positive Z-scores when we put Koreans in place of Manchus and used ancient groups of northern East Asia and southern Siberia as “ancients”, showing that Koreans harboured more northern East Asian or southern Siberian-related ancestry than Han Chinese.

Admixture model of studied Manchus and Koreans

We calculated the admixture f_3 statistics using Manchus and Koreans as target populations and all modern populations across East and Southeast Asia as potential sources

(Supplementary Table S1). Tibetan_Chamdo, Ulchi, or Nanai as northern sources, with Ami or Tai-Kadai speaking groups as southern sources, can generate the top negative f_3 values in modelling the admixture of Manchu. While Northeast Asians such as Nanai or Ulchi with Tai-Kadai people such as Mulam or Maonan generated the top negative f_3 values using Koreans as the target. The negative f_3 values indicated the genetic formation of both Manchus and Koreans had probably involved an admixture between a northern and a southern source.

We next used *qpAdm* to model the genetic formation of Manchus and Koreans (Figure 8). We found Koreans can be fitted as deriving from a single source with Bronze Age farmers from the West Liao River (WLR_BA). The Neolithic or Bronze ancients in Korea (Taejungni and Ando) can be modelled as the single source of both studied Manchus and Koreans, showing the long-term genetic continuity in the West Liao River region and Korea. However, Manchus were suggested to have received additional genetic influence from southern China. In a two-way model, we can model Manchus as an admixture of approximately 17% of Iron Age Taiwan-related ancestry and 83% of WLR-related ancestry. We note that the two-way model still held when the northern

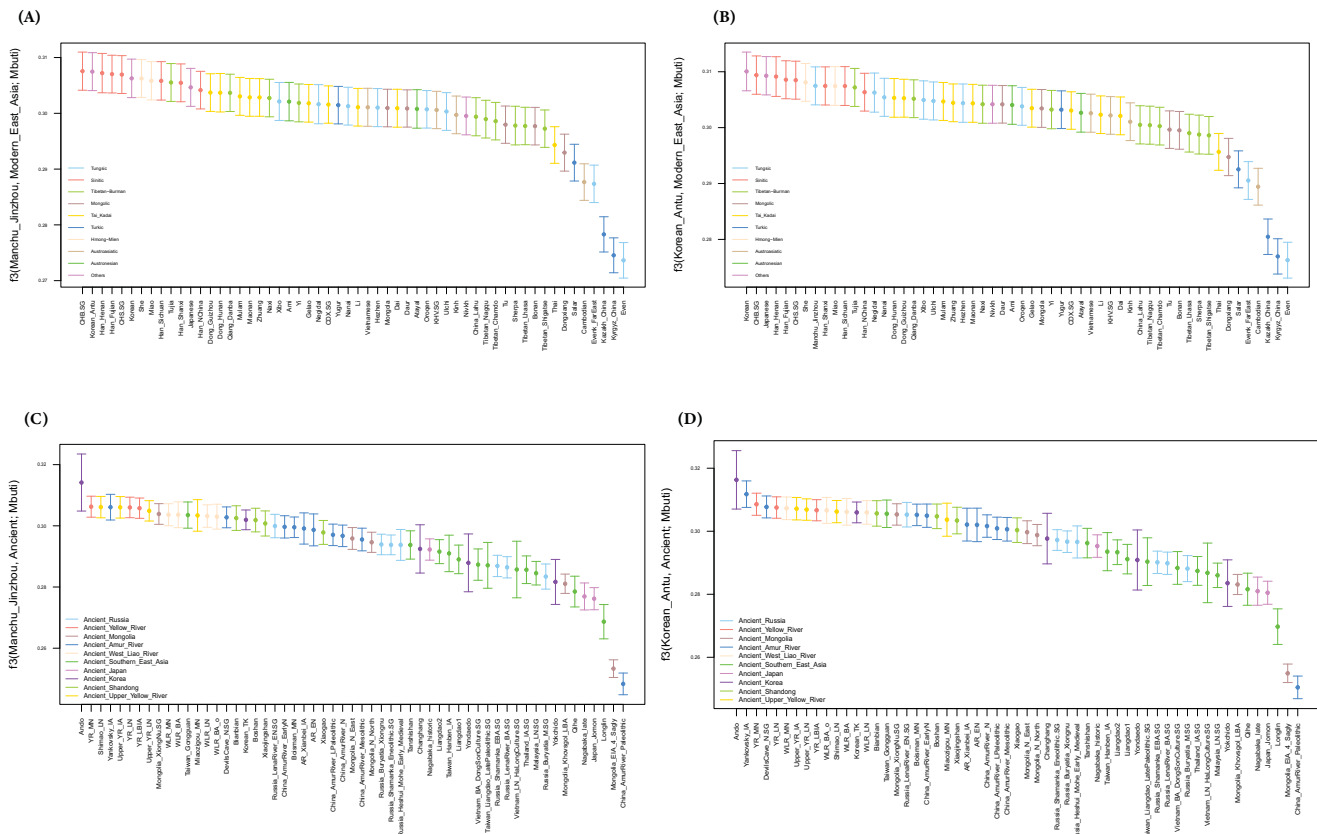


Figure 6. Outgroup-f3 values are shown in error bar plot. The error bar is marked as the standard deviation. Each group is classified in different colours by linguistic. Higher f3 value represented more shared genetic drifts. (A,B) Outgroup-f3 (Studied pops, Modern East Asia; Mbuti). (C,D) Outgroup-f3 (Studied pops, ancient reference East Asians; Mbuti).

ancestral source was replaced by the Iron Age upper Yellow River group (Upper_YR_IA), indicating the limited resolution of the currently available data in distinguishing Upper_YR_IA and WLR_BA. When we added Heishui_Mohe or Mongolia_Xiongnu as the third source in the 3-way models (Supplementary Table S3), Manchus can be modelled as 7.5% of Taiwan_Hanben, 84% of YR_LN and the remaining 8.5% of Heishui_Mohe related ancestry. To compare the difference between our target and closely related groups, we next modelled the formation of neighbouring populations, including Koreans and Mongolians (Supplementary Table S3). With the same sources and outgroups, modern Koreans can also be modelled as an admixture of WLR_BA(85%) and Taiwan_Hanben(15%). We observed Mongolians can be modelled with less southern ancestry from Taiwan_Hanben related populations but more ancestry from Northeast Asians, such as Xianbei or Mohe.

Paternal and maternal haplogroup genotyping of each individual

In the Supplementary Table S5, we show mtDNA and Y-chromosome haplogroups of the studied Manchus and Koreans. The dominant paternal Y-chromosome haplogroup in Manchus is O2a1c1, similar to northern Han Chinese. On the other hand, we identified a variety of mtDNA haplogroups, including B4, C4, D4, F1, F2, M8, and N9 in our studied Manchus, and most of these maternal lineages are

also prevalent in Han Chinese. Surprisingly, we found a U5 haplotype in the studied Manchus. Haplogroup U is broadly distributed among West Eurasians (Malyarchuk et al. 2010). The presence of U5 in the Manchus indicated the West Eurasian genetic influence in the formation of Manchu people.

The studied Koreans in Antu have a similar maternal genetic profile to Koreans in South Korea (Lee et al. 2006), which have the dominant D4 lineages but also have A5, B4, F1, G1, M7, M9, N9, and Y1 as other East Asians. The Y chromosomal haplogroups of six male samples belonged to O1, O2, and C2, all commonly found lineages in Han Chinese and Koreans (Cai et al. 2009).

Discussion

As a minority that once established a regime with great influence, the Manchus have integrated and interacted with many ethnic groups over the centuries. In the Qing Dynasty, established by the Manchus, they had large-scale intermarriages with Han Chinese. In the last centuries, large-scale migration from northeast China to the central plain through Liaodong Corridor and the Brave journey to the northeast (known as “Chuang Guandong” in Chinese) brought extensive interactions between Manchus and populations from central and southern China. Those historical reasons may explain the genetic affinity between Manchus and other populations all over China, including northern Sino-Tibetan and

(C) f4(Mbuti, ancient; Korean_Antu, East_Asia)

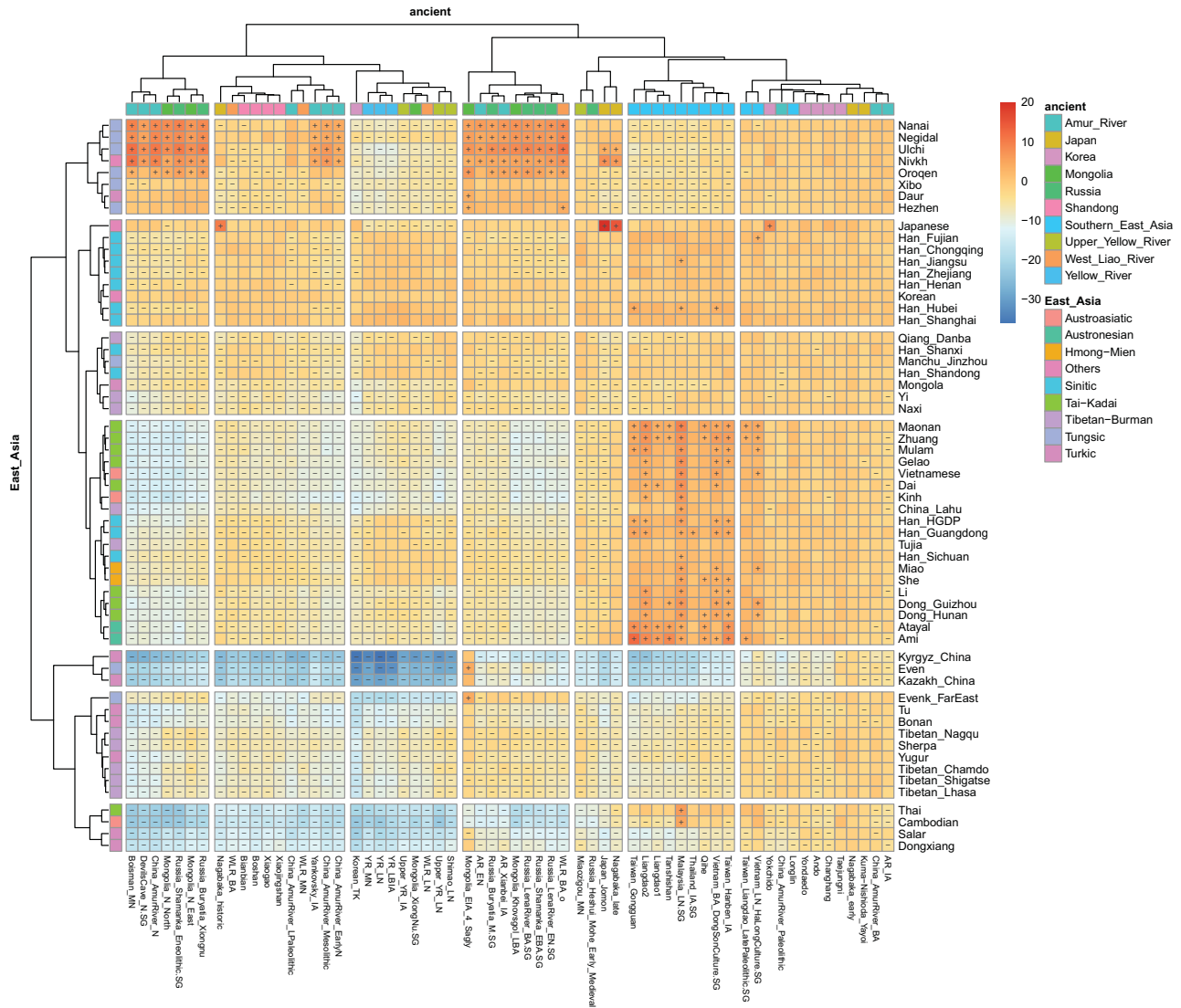


Figure 7. -F4 Continued.

southern Tai-Kadai or Hmong-Mien groups in our study. Moreover, the close relationship between northern Hans and Manchus has been proven in many previous studies, and is confirmed in this study once again (Tian 2004; Zhao 2007).

Although the Manchu have the largest population in the Tungusic language family, the genetic profile of the Manchu is different from other Tungusic groups such as the Oroqen, Hezhen, and Xibo. We found small proportions of West Eurasian-related components in other Tungusic groups but not in Manchus. Manchus have a close affinity with ancient populations from the West Liao River and Yellow River Basin. Manchus can be modelled as an admixture of Bronze Age West Liao River farmers and Iron Age Taiwan_Hanben, indicating the north-south admixture in the formation of Manchu. A previous study found Manchus can be modelled as deriving 32.4% ancestry from ancient Mohe people and the remaining ancestry from the farming-related ancient populations in the Yellow River Basin (Zhang et al. 2021). However we failed in the two-way model with Heishui_Mohe as the source. The reason may be that we have

approximately 200,000 SNP sites after merging with the 1240k dataset, which is 80,000 sites more than the previous study (almost 120,000 sites). We can have a better resolution when we have more SNPs, and more SNPs may detect the subtle genetic difference leading to a failure in the modelling, which could not be found with a smaller number of SNPs. Another reason may be our smaller sample size and the different sampling site of the Manchu: our samples were collected from Jinzhou, but the samples of Zhang et al. (2021) were from Xinbin.

Gelabert et al. show that the 4th–5th-century South Korean populations had a varied proportion of indigenous Jomon-related ancestry, which does not survive in present-day Koreans (Gelabert et al. 2022). Consistent with the previous study, we have not detected Jomon-related ancestry in present-day Chinese Koreans. Our studied Koreans in Jilin showed a strong connection with present-day people from South Korea, which supported that the origin of Chinese Koreans can be traced back to the recent migration of the Korean Peninsula. Chinese Koreans can also be modelled as

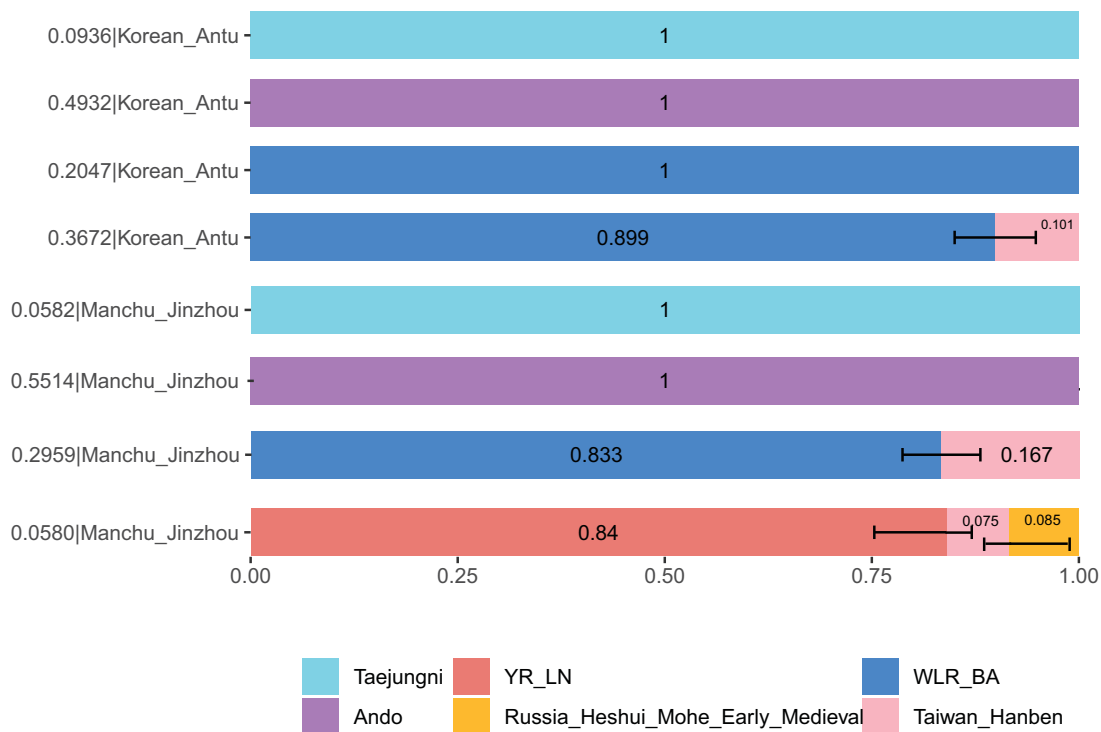


Figure 8. Admixture model of qpAdm. Each bar represents a success model ($p > .05$) of our chosen outgroups. Each label of the bar contains target population and p value of this model. The error bar denoted the standard error estimated using jack-knife. Only successful models of the studied population are shown in this graph, other unsuccessful models or models that use other target populations are shown in Table S3.

deriving ancestry from a single source related to WLR_BA, consisting of the transmission route of farming from the northeast to the Korean Peninsula and even the Japanese islands (Kwak et al. 2017; Kim and Park 2020).

Previous studies have shown that Manchus and Koreans correspond in language, history and culture. Manchus and Koreans have a linguistic connection because Koreans have many Manchu loanwords and the same grammar structure. Chinese Manchus and Koreans are the main ethnic minorities in northeast China and have had many communications in history. In addition, Chinese Manchus and Koreans are deeply influenced by the Han culture and have genetic exchanges with Han (Kim and Park 2020). The historical memory of native Manchus in Beizhen, Jinzhou, recorded their original place as Changbai Mountain (Figure S1). We here found Manchus were genetically similar to Chinese Koreans from Changbai Mountain, which is consistent with the linguistic connection between Manchus and Koreans.

To further investigate the genetic profile in northeast China, more subgroups of Manchus and Koreans in this area should be collected in the future. In addition, the reference dataset of modern populations was mainly generated *via* the Affymetrix Human Origins arrays. We have only about 70,000 sites left in the analysis after merging with the Human Origin dataset, which is the limitation of our research.

Acknowledgements

S. Fang and Z. Xu from the Information and Network Center of Xiamen University are acknowledged for their help with high-performance computing.

Ethics statement

The studies involving human participants were reviewed and approved by the Ethical Committee of Xiamen University. The participants provided their written informed consent to participate in this study.

Author contributions

C-C.W. and X.H. designed this study, N.S. collected the samples and conducted the experiment, L.T. analysed the data, L.T., and C-C.W. wrote the manuscript, X.H. modified the manuscript. All authors reviewed the manuscript.

Disclosure statement

The authors declare that the research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

This work was funded by the National Social Science Fund of China (20BY195) "Language evolution and its correlation with population mixing history in Northeast Asia" to Na Sun, National Natural Science Foundation of China (32270667), the "Double First Class University Plan" key construction project of Xiamen University (0310/X2106027), Nanqiang Outstanding Young Talents Program of Xiamen University (X2123302), the Major Project of National Social Science Foundation of China granted to C-CW (21&ZD285), the Major Project of National Social Science Foundation of China (20&ZD248), the Major Project of National Social Science Foundation of China (2021MZD014), and the European Research Council (ERC) grant (ERC-2019-ADG-883700-TRAM).

Data availability statement

The datasets presented in this study can be found in online repositories. The repository's name and accession number can be found below: <https://zenodo.org/record/6847790>.

References

- Atif A, Kasim K, Rakha A, Noor A, Cheema AS, Hadi S, Xing J. 2019. Population data of 23 Y STRs from Manchu population of Liaoning Province, Northeast China. *Int J Legal Med.* 133(3):785–788.
- Bai R, Liu Y, Lv X, Shi M, Ma S. 2016. Genetic polymorphisms of 17 Y chromosomal STRs in She and Manchu ethnic populations from China. *Forensic Sci Int Genet.* 22:e12–e14.
- Cai X, Wang X, Li S, Qian J, Qian D, Chen F, Yang Y, Yuan Z, Xu J, Bai Y, et al. 2009. Association of mitochondrial DNA haplogroups with exceptional longevity in a Chinese population. *PLoS One.* 4(7):e6423.
- Chen J, He G, Ren Z, Wang Q, Liu Y, Zhang H, Yang M, Zhang H, Ji J, Zhao J, et al. 2021. Genomic insights into the admixture history of Mongolic- and Tungusic-speaking populations from Southwestern East Asia. *Front Genet.* 12:685285.
- Gelabert P, Blazyte A, Chang Y, Fernandes DM, Jeon S, Hong JG, Yoon J, Ko Y, Oberreiter V, Cheronet O, et al. 2022. Northeastern Asian and Jomon-related genetic structure in the Three Kingdoms period of Gimhae, Korea. *Curr Biol.* 32(15):3232–3244.e6.
- He J, Guo F. 2013. Population genetics of 17 Y-STR loci in Chinese Manchu population from Liaoning Province, Northeast China. *Forensic Sci Int Genet.* 7(3):e84–e85.
- Jin H-J, Tyler-Smith C, Kim W. 2009. The peopling of Korea revealed by analyses of mitochondrial DNA and Y-chromosomal markers. *PLoS One.* 4:e4210.
- Katoh T, Munkhbat B, Tounai K, Mano S, Ando H, Oyungereel G, Chae G-T, Han H, Jia G-J, Tokunaga K, et al. 2005. Genetic features of Mongolian ethnic groups revealed by Y-chromosomal analysis. *Gene.* 346:63–70.
- Kim J, Jeon S, Choi J-P, Blazyte A, Jeon Y, Kim J-I, Ohashi J, Tokunaga K, Sugano S, Fucharoen S, et al. 2020. The origin and composition of Korean ethnicity analyzed by ancient and present-day genome sequences. *Genome Biol Evol.* 12(5):553–565.
- Kim J, Park J. 2020. Millet vs rice: an evaluation of the farming/language dispersal hypothesis in the Korean context. *Evolut Hum Sci.* 2:e12.
- Kim S-H, Kim K-C, Shin D-J, Jin H-J, Kwak K-D, Han M-S, Song J-M, Kim W, Kim W. 2011. High frequencies of Y-chromosome haplogroup O2b-SRY465 lineages in Korea: a genetic perspective on the peopling of Korea. *Investig Genet.* 2(1):10.
- Kwak S, Kim G, Lee G-A. 2017. Beyond rice farming: evidence from central Korea reveals wide resource utilization in the Songgukri culture during the late-Holocene. *Holocene.* 27(8):1092–1102.
- Lee HY, Yoo J-E, Park MJ, Chung U, Kim C-Y, Shin K-J. 2006. East Asian mtDNA haplogroup determination in Koreans: haplogroup-level coding region SNP analysis and subhaplogroup-level control region sequence analysis. *Electrophoresis.* 27(22):4408–4418.
- Liu D, Duong NT, Ton ND, Van Phong N, Pakendorf B, Van Hai N, Stoneking M. 2020. Extensive ethnolinguistic diversity in Vietnam reflects multiple sources of genetic diversity. *Mol Biol Evol.* 37(9): 2503–2519.
- Liu J, Guo L, Qi R, Li S, Yin J, Zhang W, Sun Z, Tian X, Gao B. 2013. Allele frequencies of 19 autosomal STR loci in Manchu population of China with phylogenetic structure among worldwide populations. *Gene.* 529(2):282–287.
- Malyarchuk B, Derenko M, Grzybowski T, Perkova M, Rogalla U, Vanacek T, Tsybovsky I. 2010. The peopling of Europe from the mitochondrial haplogroup U5 perspective. *PLoS One.* 5(4):e10285.
- Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen W-M. 2010. Robust relationship inference in genome-wide association studies. *Bioinformatics.* 26(22):2867–2873.
- 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(12):3256–3266.e13.
- Ning C, Li T, Wang K, Zhang F, Li T, Wu X, Gao S, Zhang Q, Zhang H, Hudson MJ, et al. 2020. Ancient genomes from northern China suggest links between subsistence changes and human migration. *Nat Commun.* 11(1):2700.
- Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D. 2012. Ancient admixture in human history. *Genetics.* 192(3):1065–1093.
- Patterson N, Price AL, Reich D. 2006. Population structure and Eigenanalysis. *PLoS Genet.* 2(12):e190.
- Pickrell JK, Pritchard JK. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* 8(11):e1002967.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 81(3):559–575.
- Robbeets M, Bouckaert R, Conte M, Saveliev A, Li T, An D-I, Shinoda K, Cui Y, Kawashima T, Kim G, et al. 2021. Triangulation supports agricultural spread of the Transeurasian languages. *Nature.* 599(7886):616–621.
- Song JJ. 2005. *The Korean language: structure, use and context.* London: Routledge.
- Tian Z. 2004. The origin and historical contribution of “Chuang Guandong.” *Hist Geogr Northeast China.* 7:21–28.
- Wang C-C, Robbeets M. 2020. The homeland of Proto-Tungusic inferred from contemporary words and ancient genomes. *Evolut Hum Sci.* 2: e8.
- Wang C-C, Yeh H-Y, Popov AN, Zhang H-Q, Matsumura H, Sirak K, Cheronet O, Kovalev A, Rohland N, Kim AM, et al. 2021. Genomic insights into the formation of human populations in East Asia. *Nature.* 591:413–419.
- Xing J, Adnan A, Rakha A, Kasim K, Noor A, Xuan J, Zhang X, Yao J, McNeven D, Wang B. 2019. Genetic analysis of 12 X-STRs for forensic purposes in Liaoning Manchu population from China. *Gene.* 683:153–158.
- Xue Y, Zerjal T, Bao W, Zhu S, Lim S-K, Shu Q, Xu J, Du R, Fu S, Li P, et al. 2005. Recent spread of a Y-chromosomal lineage in Northern China and Mongolia. *Am J Hum Genet.* 77(6):1112–1116.
- Xue Y, Zerjal T, Bao W, Zhu S, Shu Q, Xu J, Du R, Fu S, Li P, Hurles ME, et al. 2006. Male demography in East Asia: a North–South contrast in human population expansion times. *Genetics.* 172(4):2431–2439.
- Zhang X, He G, Li W, Wang Y, Li X, Chen Y, Qu Q, Wang Y, Xi H, Wang C-C, et al. 2021. Genomic insight into the population admixture history of Tungusic-speaking Manchu people in Northeast China. *Front Genet.* 12:754492.
- Zhao Y-B, Sun W-Y, Zhan Y, Wang D, Yu C-C. 2011. Mitochondrial DNA evidence of southward migration of Manchus in China. *Mol Biol.* 45(5):759–765.
- Zhao Y-I. 2007. From Manchulization, Chinesization to the ethnic unity of multicultures—evolution of relations among ethnic groups in Northeast Qing Dynasty. *Northeast Asia Forum.* 16:116–120.