creative commons

COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

Master's Thesis of Biological Sciences

The genetic profiles and relatedness of the Three Kingdoms period individuals in Korea

삼국시대 한국인의 유전적 프로필과 혈연관계 분석

August 2022

Graduate School of Natural Sciences
Seoul National University
Biological Sciences Major

Don-Nyeong Lee

The genetic profiles and relatedness of the Three Kingdoms period individuals in Korea

Advised by Professor Choongwon Jeong

Submitting a master's thesis of
Biological Sciences

August 2022

Graduate School of Natural Sciences
Seoul National University
Biological Sciences Major

Don-Nyeong Lee

Confirming the master's thesis written by
Don-Nyeong Lee
August 2022

Chair _____(Seal)
Vice Chair _____(Seal)
Examiner _____(Seal)

Abstract

The Three Kingdoms period of Korea was the time when the ancient nations first appeared on the Korean peninsula. The Three Kingdoms period Koreans have been considered as direct ancestors of present-day Koreans. However, due to the lack of ancient genomes from the Three Kingdoms period, the genetic relationship between the Three Kingdoms period Koreans and present-day Koreans has remained largely unknown. In this study, to test hypothesis of the origins of present-day Koreans and investigate regional differences in genetic profile, genome-wide data of the Three Kingdoms period individuals from the Gunsan and Imdang sites are generated and analyzed. The jar coffin at the Dangbuk-ri site in Gunsan shows an unusual occasion of secondary multiple burial in a single jar coffin. Since the Ancient Tombs in Imdang-dong and Joyeong-dong present the evidence of sacrificial burials and are presumed as elite tombs of Abdokguk, genomes from this site are appropriate case not only for revealing genetic profiles, but also for understanding structure of the past society. In this study, genome-wide data of six individuals from the Gunsan jar coffin and 35 individuals from the Imdang site are generated. By using population genetic methods, genetic profiles are characterized and analyzed to test whether there are genetic differences between individuals from different regions, social statuses, or periods. Pedigrees of the Three Kingdoms period individuals are

reconstructed by using genetic relationship. In addition, evidence of relative marriage was investigated. The genetic profiles of the Three Kingdoms period individuals from Gunsan and Imdang are found to be similar to each other and that of present-day Koreans, suggesting long-term genetic stability in the broad region of the Korean peninsula at least for the last 1,500 years. Also, individuals from the Imdang site assigned to different social statuses have indistinguishable genetic profiles. As the archaeological records suggest, this study shows that the specific burial patterns of these sites reflect the genetic relationship between the individuals from each site. Individuals from the Gunsan jar coffin form an extended family consisting of a quartet family and both paternal and maternal relatives, showing little gender bias in the family structure. Individuals from the Imdang site form six distinct families that tend to correspond to their social statuses. In addition, close parental relatedness found in two Imdang individuals suggests that the relative marriage was practiced in Abdokguk at the Three Kingdoms period.

Keyword : Three Kingdoms period, ancient genome, population genetics, family burial, sacrificial burial, Gunsan Dangbuk-ri, Ancient Tombs in Imdang-dong and Joyeong-dong

Student Number : 2020-25735

Table of Contents

Abstract i

Table of Contents iii

List of Tables..... v

List of Figures vi

1. Introduction 1

1.1. Necessities for genomic studies of the Three Kingdoms period Koreans 1

1.2. A jar coffin from the Dangbuk-ri site 4

1.3. The Ancient Tombs in Imdang-dong and Joyeong-dong..... 6

1.4. Purpose of Research 7

2. Materials and Methods..... 12

2.1. Sampling of ancient human remains..... 12

2.2. Laboratory work and sequencing of ancient DNA 12

2.3. Data processing and authentication of ancient DNA sequencing data..... 13

2.4. Reprocessing of high-coverage whole genome sequencing data of present-day Koreans..... 14

2.5. Data set compilation for population genetic analysis.... 20

2.6. Principal component analysis 21

2.7. *F*-statistics and qpWave/qpAdm analysis 21

2.8. Genetic relationship analysis 22

2.9. Runs of Homozygosity analysis 23

3. Results 24

3.1. Ancient genome-wide data production 24

3.2. The genetic profiles of Three Kingdoms period Koreans 35

3.3. Genetic similarities between the Three Kingdoms period Koreans..... 46

3.4. A familial relationship of Three Kingdoms period individuals 50

4. Discussion..... 60

References..... 64

국문 초록 74

Appendix..... 76

List of Tables

Table 1. Close relative pairs detected from 104 present-day Koreans from Ulsan. 19

Table 2. A summary of sequencing and genetic information of the Three Kingdoms period individuals in this study. 26

Table 3. A summary of archaeological information and uniparental haplogroup of the Three Kingdoms period individuals in this study. 30

Table 4. A summary of deep shotgun sequencing information of 13 individuals from the Imdang site..... 34

Table 5. Admixture model of the Gunsan individuals and present-day Koreans using proximal sources. 42

Table 6. Admixture model of individuals from the Imdang site and present-day Koreans using proximal sources. 43

Table 7. Admixture model of the Three Kingdoms period and present-day Koreans and nearby ancient East Asian groups using distal sources. 45

Table 8. Genetic cladality test between the Three Kingdoms period individuals. 49

Table 9. Genetic relatedness between the six individuals from the Gunsan jar coffin. 55

Table 10. Genetic relatedness between the 35 individuals from the Imdang site. 57

List of Figures

Figure 1. Sampling sites of the key ancient and present-day East Asian populations used in this study..... 9

Figure 2. Excavated skeletons of multiple individuals in a single jar coffin, Dangbuk-ri site *in situ*..... 11

Figure 3. Genetic sex determination of present-day Ulsan Koreans..... 17

Figure 4. Relatives among present-day Ulsan Koreans..... 18

Figure 5. Principal component analysis from present-day Eurasian individuals..... 38

Figure 6. Principal component analysis from present-day East Asian individuals..... 39

Figure 7. Top 30 outgroup-f_3 statistics of the form f_3(Mbuti; Three Kingdoms period Korean group, world-wide) across ancient and present-day world-wide populations. 40

Figure 8. Genetic symmetry test of the Three Kingdoms period individuals and present-day Koreans. 41

Figure 9. QpAdm-based admixture modeling of Three Kingdoms period Koreans and other ancient and present-day mainland East Asian populations. 44

Figure 10. Genetic symmetry test between different social statuses of Imdang individuals..... 47

Figure 11. Genetic symmetry test of the Three Kingdoms period individuals from different sites. 48

Figure 12. Genetic relationship between the six individuals from the Gunsan jar coffin. 54

Figure 13. Estimated genetic relatedness of 35 Imdang individuals..... 56

Figure 14. Reconstructed pedigrees of individuals from the Imdang site..... 58

Figure 15. Inferred ROH from the Three Kingdoms period individuals. 59

1. Introduction

1.1. Necessities for genomic studies of the Three Kingdoms period Koreans

The Three Kingdoms period of Korea refers to the period from the 1st century BC to the 7th century AD, when the Three Kingdoms represented by Goguryeo, Baekje, and Silla occupied the Korean peninsula and parts of Manchuria (Peterson and Margulies, 2010). Historical, archaeological, and linguistic features support strong link between inhabitants of the Korean peninsula during the Three Kingdoms period (hereafter referred to as "Three Kingdoms period Koreans") and present-day Koreans (Lee and Ramsey, 2011). Geneticists have investigated whether this strong cultural connection between the Three Kingdoms period and present-day Koreans reflects their genetic continuity, that is whether present-day Koreans are direct descendants of Three Kingdoms period Koreans.

Previous genetic studies focused on mitochondrial DNA (mtDNA) (Kim et al., 2010; Lee et al., 2008), because the high copy number of mtDNA in cells makes it easier to extract mtDNA than nuclear DNA from ancient samples. However, mtDNA cannot provide sufficient information to test genetic continuity in a local context due to the broad distribution of mitochondrial haplogroups. In a previous study, mitochondrial DNA of 1st century BC individuals

from Nukdo island, at the southern end of the Korean peninsula belong to haplogroups common in present-day Koreans or Japanese (i.e. M, D, D4, and D4b) (Kim et al., 2010). Another study, analyzing mtDNA of a Three Kingdoms period Korean and two preceding Bronze Age Koreans, assigned their haplogroups as D4 which found in about 30% in present-day Koreans (Derenko et al., 2007; Lee et al., 2008; Lee et al., 2006). Although these studies claimed genetic similarity between the Three Kingdoms and present-day Koreans, information from mtDNA alone is insufficient to test the ancestor-descendant relationship.

Contrary to mtDNA, genome-wide data can provide sufficient statistical power to test population relationship in fine resolution by integrating information from hundreds of thousands even millions of markers across the genome. Therefore, genome-wide data of the Three Kingdoms period Koreans are necessary to clarify the relationship between the Three Kingdoms period and present-day Koreans. Indeed, many studies using ancient genomes have elucidated the genetic history of human populations over the world (Allentoft et al., 2015; Haak et al., 2015; Jeong et al., 2020; Ning et al., 2020). However, studies using ancient genomes from the Korean peninsula have only recently begun (Gelabert et al., 2022; Robbeets et al., 2021). A recent study reporting genetic profiles of individuals dated to 350-500 CE from the Gimhae, southern coastal region of the Korean peninsula, supports substantial genetic continuity in Korea since that time (Gelabert et al., 2022). This

study reported not only a present-day Korean-like genetic profile, but also a genetic profile with a significant amount of Jomon-related ancestry (about one-third). While human migrations between Korea and Japan are expected in this center of international trade in the Three Kingdoms period, it could not be confirmed whether genetic profiles found in this single region are representative of Three Kingdoms period Koreans. Therefore, Three Kingdoms period genomes from different sites are necessary to confirm the overall genetic homogeneity of the Three Kingdoms period Koreans and test genetic continuity between the Three Kingdoms period and present-day Koreans.

Since kinship has played an important role in the inheritance of power and wealth, inference of kinship between ancient individuals is one of the major interests in archaeology and anthropology. However, early kinship studies had depended on indirect information such as archaeological records or morphological traits, because there was little or no genetic information available (Alt and Vach, 1998). While multi-allelic autosomal markers harbor the key information to estimate genetic relatedness, early genetic studies used only mtDNA (Mooder et al., 2005; Rudbeck et al., 2005) or additionally used several to dozens of autosomal and Y chromosomal markers genotyped from PCR-based methods (Hawass et al., 2010; Keyser-Tracqui et al., 2003), due to the limitation of extraction and sequencing techniques using ancient DNA. Since poor preservation of ancient DNA causes missing data

and uncertain genotypes, small number of autosomal markers from ancient DNA usually cannot provide sufficient information to estimate accurate kinship, except for unusually well-preserved samples. Recently, Next-Generation Sequencing allowed production of genome-wide data of ancient samples, thus enabled estimation of accurate kinship between ancient individuals. Using such information, genomic studies have been performed to understand structure of past societies by incorporating genetic relatedness and archaeological records (Kennett et al., 2017; Mittnik et al., 2019).

In this study, I investigated the above explained two important topics of the genetic profiles and kinship structure of the Three Kingdoms period Koreans by analyzing genome-wide data of the two Three Kingdoms period sites about 180 km apart, one in the southwestern and the other in the southeastern region of Korea. Individuals from these two sites show burial practices that imply a close relationship between them. A jar coffin from the Gunsan shows an exceptional case of a secondary multiple burial in a single jar coffin. Therefore, it can be hypothesized that the individuals from the jar coffin are closely related genetically. Ancient Tombs in Imdang-dong and Joyeong-dong in Gyeongsan have been presumed to be the tombs of local ruling family groups with sacrificial burials (Kim, 2014).

1.2. A jar coffin from the Dangbuk-ri site

Jar coffin was a widespread burial type in ancient East Asia including the Korean peninsula in the Three Kingdoms period (Bacvarov, 2006; Boeyens et al., 2009; Kim, 1973; Naju national research institute of cultural heritage, 2020; Shewan et al., 2020). Jar coffin burials from the Neolithic to historical periods are found in all regions in the Korean peninsula. While large jar coffins comparable to human-height are found in Yeongsan River basin during 4th to 6th centuries AD (Naju national research institute of cultural heritage, 2020), most Korean jar coffins are much smaller. In most cases, these small jar coffins carry remains of a single individual that were secondarily buried after cremation. Therefore, most skeletal remains from jar coffins cannot provide morphological information to identify sex, age, and pathological history of the individual, as well as ancient DNA.

In this study, I investigated genome-wide data from human skeletal remains in a jar coffin from Dangbuk-ri site located in the Gunsan city, Jeollabuk-do province which is on southwest coastal region of the Korean peninsula (Figure 1). A total of 21 Three Kingdoms period burials were excavated from the Dangbuk-ri site during summer of 2016. Among the 21 burials, sixteen burials were the stone-cist type, four burials were the stone-chamber type, and the remaining one was a jar coffin. Based on burial types and artifacts, these burials are dated to the Woongjin period (475-538 AD) and the Sabi period (538-660 AD) of Baekje. Extraordinarily, this jar coffin harbored at least nine individuals, and their skeletal

remains are exceptionally well-preserved (Figure 2). In this study, genome-wide data of six individuals from the jar coffin are analyzed to test whether this unusual case of multiple burial reflects genetic relationship between the individuals.

1.3. The Ancient Tombs in Imdang-dong and Joyeong-dong

Human sacrificial burial is frequently considered as a landmark phenomenon that shows the establishment of a hierarchical society (National Research Institute of Cultural Heritage, 2009). Among the archaeological sites from the Three Kingdoms period, most sacrificial burials are found in tombs of Gaya and Silla before mid 6th century AD (National Research Institute of Cultural Heritage, 2009). One of the representative sacrificial burials of Gaya is found in Daeseong-dong site. A recent study reported genomes dated to 350-500 CE from the Daeseong-dong site and found no genetic differences by social status (Gelabert et al., 2022). In Silla, the clear appearance of sacrificial burial is confirmed in the 4th century from Joyeong-dong. Thereafter, sacrificial burial was frequently practiced in Silla until 6th century (Kim, 2002). Ancient Tombs in Imdang-dong and Joyeong-dong are one of the biggest complexes of tombs with sacrificial burials in Silla.

Ancient Tombs in Imdang-dong and Joyeong-dong are located in Gyeongsan, Gyeongsangbuk-do Province on the Korean

peninsula, and were excavated from 1982 (Figure 1). Archaeological records and historical documents indicate that there had been a small state called Abdokguk about from 1st century BC to 6th century AD at this site (Jang, 2007). Abdokguk was incorporated into Silla in about 4th century AD, but the ruling class of Abdokguk was maintained thereafter. (Jang, 2007). Ancient Tombs in Imdang-dong and Joyeong-dong are presumed as a series of Abdokguk elite tombs. High social status of the masters of the tombs is represented by the prestige goods such as crowns, earrings, belts, and bracelets made of precious materials such as gold, gilt-bronze, and silver (Kim, 2014). A previous study inferred social status and relationship between Imdang individuals based on excavated accessories and mtDNA sequences, but kinship could not be determined with such indirect information (Ha, 2011). In addition, it is not investigated whether there is genetic difference between individuals assigned to different social statuses. In this study, 46 individuals from 22 tombs that are likely related to human sacrificial burials are sampled to solve these problems.

1.4. Purpose of Research

This study aims to investigate the kinship of ancient individuals and genetic relationship between the Three kingdoms period and present-day Koreans by analyzing genome-wide data of the two Three Kingdoms period sites. Three hypotheses on genetic

relatedness and genetic profiles of the Three Kingdoms period individuals are tested. First, archaeology-driven hypothesis on the relationship of the individuals from the jar coffin in the Dangbuk-ri site and the Ancient Tombs in Imdang-dong and Joyeong-dong is tested by estimating genetic relatedness between the individuals. Second, genetic continuity between the Three Kingdoms period and present-day Koreans is tested to validate the hypothesis that claims the Three Kingdoms period Koreans as the direct ancestor of the present-day Koreans. Third, genetic differences between the individuals from different sites or social statuses are tested. As results, rich kinship information between the Three Kingdoms period individuals is obtained as expected from the archaeological records, and present-day Korean-like genetic profile is observed in both sites. This study provides valuable information about the genetic history and society of the Three Kingdoms period in Korea.

Figure 1. Sampling sites of the key ancient and present-day East Asian populations used in this study. Sampling sites of the Three Kingdoms period individuals, previously published ancient East Asian individuals, and present-day Koreans from Ulsan are presented. The base map is prepared using the data from the Natural Earth public domain map dataset (https://www.naturalearthdata.com/downloads/10m-raster-data/10m-cross-blend-hypso/). AR_EN = Amur River Early Neolithic individuals from the Zhalainuoer/Wuqi site; AR_Xianbei_IA = Amur River Iron Age individuals from the Xianbei context of the Mogushan site; Boshan = an Early Neolithic individual from Shandong region; DevilsCave_N = Early Neolithic individuals from the Russian Far East; HMMH_MN = Middle Neolithic individual from the Haminmangha site; Jomon_Ikawazu = a Jomon individual from the Ikawazu site; Liangdao2 = an Early Neolithic individual from the

Liangdao site; Miaozigou_MN = Middle Neolithic individuals from the Miaogizou site; Qihe = Early Neolithic individuals from the Qihe site; Shimao_LN = Late Neolithic individuals from the Shengedaliang site; Upper_YR_LN = Upper Yellow River Late Neolithic individuals from the Jinchankou and Lajia sites; WLR_MN = West Liao River Middle Neolithic individuals from the Banlashan site; WLR_LN = WLR Late Neolithic individuals from the Erdaojingzi site; WLR_BA/WLR_BA_o = WLR Bronze Age individuals from the Longtoushan site; Xitoucun = Late Neolithic individuals from the Xitoucun site; YR_MN = Yellow River Middle Neolithic individuals from the Xiaowu and Wanggou site; YR_LN = YR Late Neolithic individuals from the Haojiatai/Pingliangtai/Wadian site.

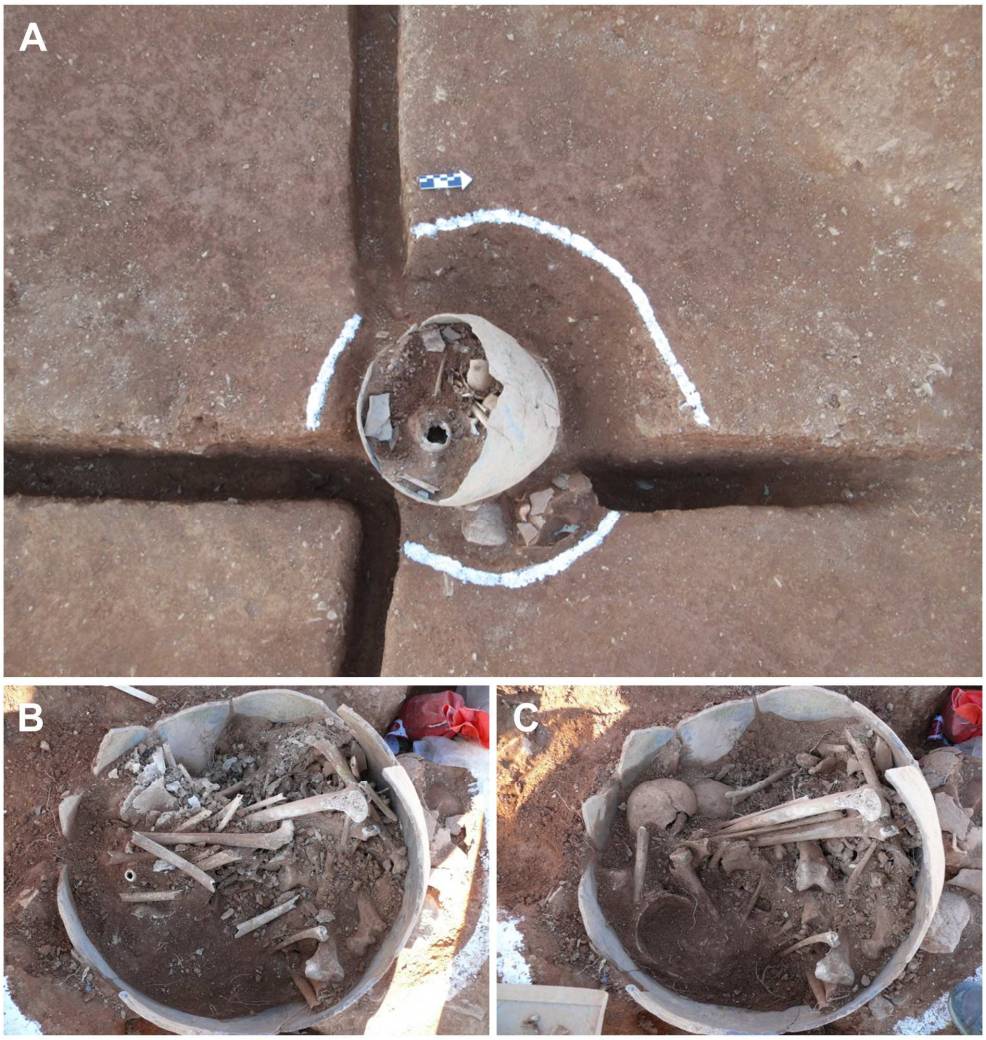


Figure 2. Excavated skeletons of multiple individuals in a single jar coffin, Dangbuk-ri site *in situ.* (A) A broad view of the Gunsan jar coffin. (B, C) The skeletons of the Gunsan jar coffin individuals.

2. Materials and Methods

2.1. Sampling of ancient human remains

Since compact bones are favorable for ancient DNA preservation, petrous parts of the temporal bones from seven Gunsan individuals and 33 Imdang individuals and teeth from 13 Imdang individuals were chosen. Two out of nine individuals from the Gunsan jar coffin were not sampled due to the poor preservation of temporal parts. The inner parts of the petrous bones and teeth were ground into fine powder.

2.2. Laboratory work and sequencing of ancient DNA

For each of the seven Gunsan individuals and 46 Imdang individuals, a double-strand double-indexed Illumina sequencing library was prepared from metagenomics DNA extracted from the tissue powder. DNA extraction and library preparation were processed following previously published protocols in a dedicated ancient DNA clean room facility of the Max Planck Institute for the science of Human History, Jena, Germany (Dabney et al., 2013). The uracil-DNA-glycosylase enzyme was partially treated to confine DNA damage to the ends of the reads. To enrich libraries, two rounds of in-solution capture for 1.24 million informative SNPs ("1240K") were performed for six Gunsan individuals and 35

Imdang individuals with sufficient levels of human DNA preservation (Mathieson et al., 2015). In addition, deep shotgun sequencing was performed for 13 Imdang individuals with especially large amount of human DNA. All libraries were sequenced on the Illumina HiSeq 4000 platform following the manufacturer's protocols using single-end 76 base pair sequencing design.

2.3. Data processing and authentication of ancient DNA sequencing data

The raw sequencing data from ancient samples were processed using the EAGER v1.92.50 wrapper (Peltzer et al., 2016). The Illumina adapter sequences were removed from raw reads using AdapterRemoval v2.3.0 (Schubert et al., 2016). Adapter-removed reads with 30bp or longer length were then mapped to the human reference genome (hs37d5; GRCh37 with decoy sequences) by using aln/samse modules in BWA 0.7.17 with "-n 0.01" option (Li and Durbin, 2009). Using DeDup v0.12.5, PCR duplicates were removed (Peltzer et al., 2016). Using samtools v1.9, Unique mapped reads with Phred-scaled mapping quality score 30 or higher were kept (Li et al., 2009). The first and last two bases of each read were soft-masked using the TrimBam function of bamUtils v1.0.14 to minimize the effect of postmortem DNA damage on genotyping (Jun et al., 2015). A single high-quality base with Phred-scaled base quality score 30 or higher was randomly

sampled using samtools mpileup and pileupCaller v.1.4.0.5 (downloaded from https://github.com/stschiff/sequenceTools) for SNPs in the 1240K capture panel to generate pseudo-haploid genotypes. End-masked BAM files were used for C/T and G/A SNPs. For remaining SNPs, unmasked BAM files were used. To confirm the authenticity of ancient DNA sequence data, multiple data quality measures were examined. First, chemical modifications specific to ancient DNA molecules (5' C>T and 3' G>A misincorporation) were tabulated using mapDamage v2.0.9 (Jonsson et al., 2013). Second, mitochondrial DNA contamination was measured for all individuals using Schmutzi v1.5.5.5 (Renaud et al., 2015). Third, for 20 male individuals, nuclear DNA contamination based on X chromosome was estimated using the contamination module of the ANGSD v0.929 program (Korneliussen et al., 2014). Mitochondrial haplogroups were determined by using the HaploGrep v2 program (Weissensteiner et al., 2016). Y haplogroups were determined by using a modified version of the yHaplo program with "--ancStopTresh 10" option to prevent halting in an internal branch due to missing data in the root-to-tip search (downloaded from https://github.com/alexhbnr/yhaplo) (Poznik, 2016).

2.4. Reprocessing of high-coverage whole genome sequencing data of present-day Koreans

High-coverage whole genome sequencing data of 104 present-day Koreans from the Ulsan city were downloaded from KoVariome

data (ftp://biodisk.org/Release/KPGP/KPGP_Data_2017_Release_Candidate/). FASTQ files with the Phred-64 scale quality scores were rescaled to the Phred-33 scale using seqtk v1.3-r106 with "seqtk seq -VQ64" option (https://github.com/lh3/seqtk). Using BWA mem program v0.7.17 with "-M" flag, reads were aligned to the human reference genome (hs37d5) (Li and Durbin, 2010). Properly aligned paired-end reads were kept using "-f 0x0003" filter in samtools view v1.9 (Li et al., 2009) and per-lane BAM files were merged into per-individual files using samtools merge. Duplicate reads were removed using Picard MarkDuplicates v2.20.0 (downloaded from https://broadinstitute.github.io/picard/) and reads with Phred-scaled mapping quality score 25 or higher were kept. Then, two genotype calls for the 1,233,013 SNPs in the 1240K panel were generated. First, using the UnifiedGenotyper module of the Genome Analysis ToolKit (GATK) v3.8.1.0 (McKenna et al., 2010) with the "--allSitesPLs" flag, genotype likelihoods for each individual were calculated. Posterior genotype probability was calculated by multiplying genotype likelihoods with GATK default prior [0.9985, 0.0010, 0.0005], and genotype calls with posterior probability 0.900 or higher were kept. Second, to create a pseudo-haploid call that imitates the genotype calling strategy of low-coverage ancient individuals, a base with the Phred-scaled base quality score 30 or higher from a read with mapping quality score 30 or higher per site per individual was randomly sampled using

samtools mpileup and pileupCaller v1.4.0.5. The two types of genotype calls of present-day Ulsan Koreans are then merged with the genome-wide data of world-wide populations genotyped on the Affymetrix Axiome® Genome-Wide Human Origins 1 array ("HumanOrigins") and the "1240K-Illumina" dataset for downstream analyses. Genetic sex of each individual was determined by comparing coverage of sex chromosomes with autosomes measured for the 1240K sites using samtools depth (Figure 3). Genetic outliers were observed by projecting Ulsan Koreans to the principal components calculated for 2,077 present-day Eurasian individuals using the smartpca program v16000 in the EIGENSOFT package v7.2.1 (Patterson et al., 2006). Close relative pairs from the remaining individuals were identified by calculating pairwise mismatch rate (PMR) for each pair using the random haploid calls. Kinship coefficients were measured using the -Z-genome module in PLINK v1.90b6.9 (Chang et al., 2015). For each relative pair up to the second-degree relatives, one individual was removed for the downstream group-based analyses (Figure 4; Table 1).

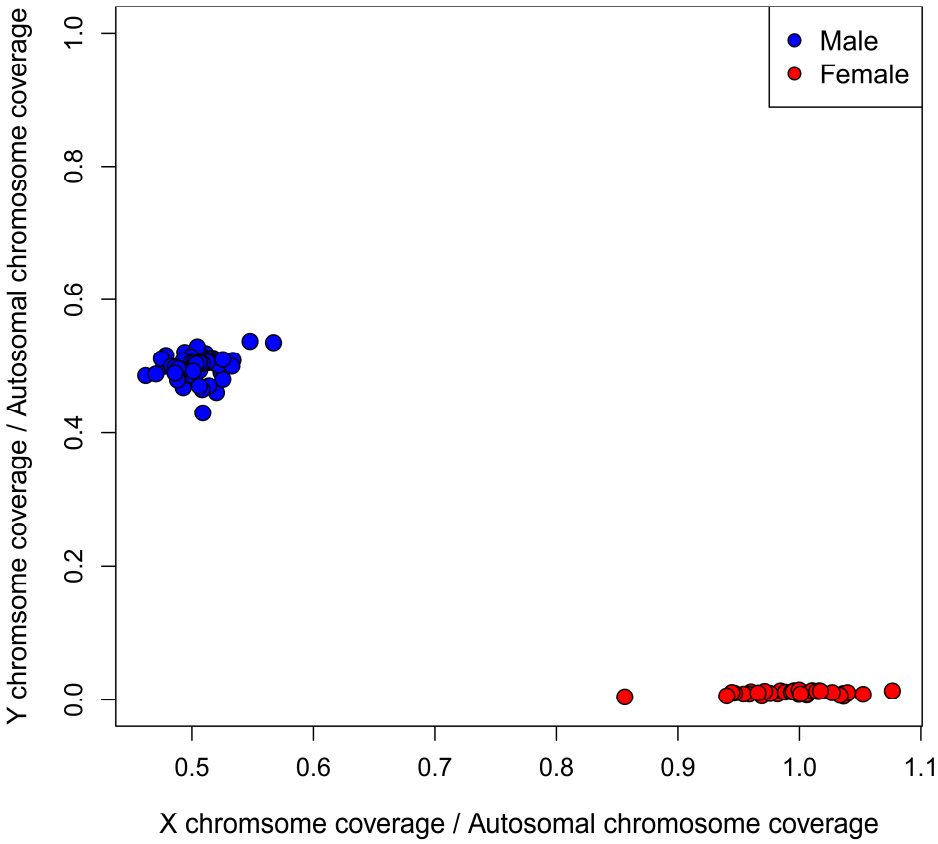


Figure 3. Genetic sex determination of present-day Ulsan Koreans. The ratio of X chromosome to autosomal chromosome coverage (x-axis) and Y chromosome to autosomal chromosome coverage (y-axis) are plotted. Genetic males and females are represented by blue and red circles, respectively.

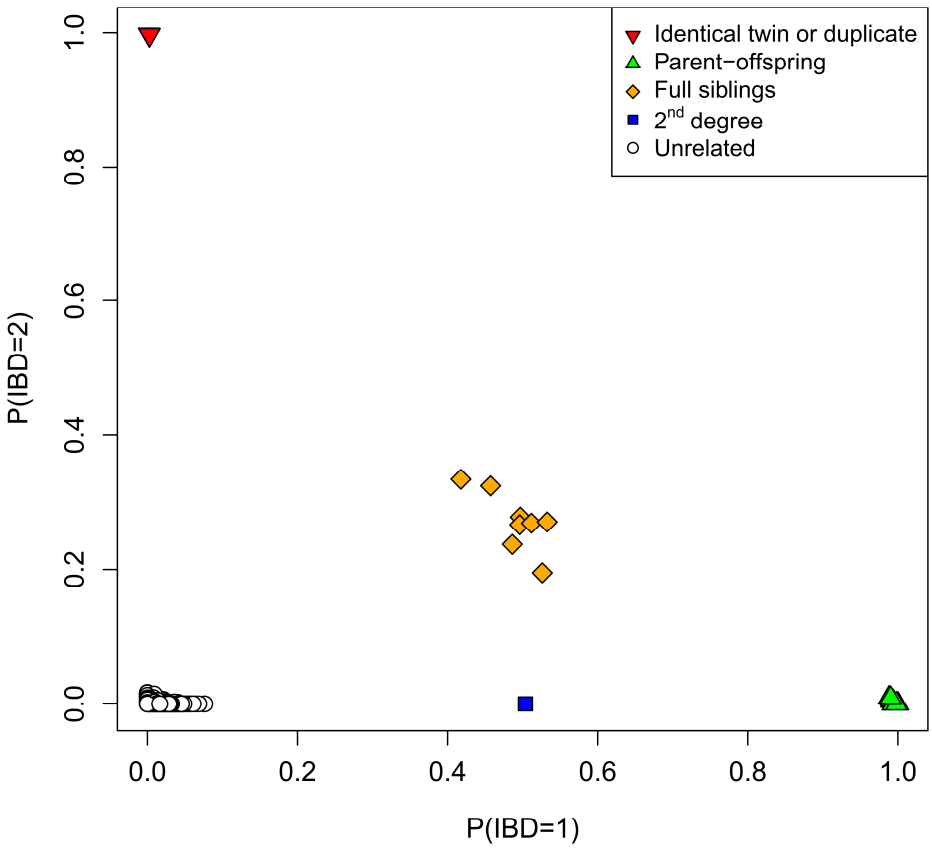


Figure 4. Relatives among present−day Ulsan Koreans. Probabilities of sharing one and two alleles per each SNP estimated by PLINK are plotted.

Table 1. Close relative pairs detected from 104 present-day Koreans from Ulsan. All relative pairs up to the 2nd degree relatives are shown. Z0, Z1, and Z2 indicate the probability of sharing 0, 1, and 2 alleles estimated by PLINK, respectively. PI_HAT indicates the estimate of the genetic relatedness estimated by PLINK. PMR indicates the pairwise mismatch rate of genotypes.

ID1	ID2	Z0	Z1	Z2	PI_HAT	PMR	Kinship
00088	00089	0.0001	0.0022	0.9976	0.9988	0.1197	identical
00090	00091	0.0002	0.0028	0.9970	0.9984	0.1216	identical
00231	00252	0.2183	0.4573	0.3243	0.5530	0.1728	full siblings
00252	00253	0.2482	0.4177	0.3341	0.5430	0.1742	full siblings
00337	00338	0.1966	0.5327	0.2707	0.5370	0.1749	full siblings
00003	00009	0.2253	0.4969	0.2779	0.5263	0.1767	full siblings
00219	00221	0.2194	0.5117	0.2689	0.5247	0.1772	full siblings
00005	00009	0.2374	0.4961	0.2665	0.5145	0.1775	full siblings
00231	00253	0.2765	0.4862	0.2373	0.4804	0.1819	full siblings
00003	00005	0.2789	0.5262	0.1950	0.4581	0.1843	full siblings
00353	00362	0.0012	0.9904	0.0084	0.5036	0.1786	parent-child
00002	00005	0.0023	0.9938	0.0039	0.5008	0.1787	parent-child
00337	00343	0.0015	0.9985	0.0000	0.4993	0.1788	parent-child
00001	00003	0.0044	0.9956	0.0000	0.4978	0.1795	parent-child
00342	00343	0.0019	0.9981	0.0000	0.4991	0.1795	parent-child
00232	00234	0.0028	0.9882	0.0090	0.5031	0.1796	parent-child
00002	00009	0.0000	1.0000	0.0000	0.5000	0.1797	parent-child
00001	00009	0.0049	0.9951	0.0000	0.4976	0.1798	parent-child
00228	00253	0.0067	0.9933	0.0000	0.4966	0.1799	parent-child
00234	00235	0.0000	1.0000	0.0000	0.5000	0.1799	parent-child
00230	00231	0.0032	0.9912	0.0056	0.5012	0.1800	parent-child
00205	00221	0.0000	1.0000	0.0000	0.5000	0.1800	parent-child
00002	00003	0.0000	1.0000	0.0000	0.5000	0.1800	parent-child
00228	00252	0.0000	1.0000	0.0000	0.5000	0.1800	parent-child
00230	00252	0.0065	0.9935	0.0000	0.4967	0.1800	parent-child
00228	00231	0.0000	1.0000	0.0000	0.5000	0.1802	parent-child
00001	00005	0.0000	1.0000	0.0000	0.5000	0.1803	parent-child
00205	00219	0.0000	1.0000	0.0000	0.5000	0.1806	parent-child
00230	00253	0.0000	1.0000	0.0000	0.5000	0.1810	parent-child
00220	00221	0.0044	0.9956	0.0000	0.4978	0.1812	parent-child
00219	00220	0.0043	0.9917	0.0040	0.4998	0.1813	parent-child
00231	00253	0.2765	0.4862	0.2373	0.4804	0.1819	full siblings
00003	00005	0.2789	0.5262	0.1950	0.4581	0.1843	full siblings
00338	00343	0.4963	0.5037	0.0000	0.2518	0.2081	2nd degree

2.5. Data set compilation for population genetic analysis

Genome-wide genotype data of 2,967 present-day individuals genotyped on the HumanOrigins array (Jeong et al., 2019; Lazaridis et al., 2016; Patterson et al., 2012) were merged with high-coverage genotype data of 104 present-day Ulsan Koreans (Kim et al., 2018), six Gunsan jar coffin individuals and 35 Imdang individuals from this study, and previously published ancient individuals (Allentoft et al., 2015; Damgaard et al., 2018a; Damgaard et al., 2018b; Fu et al., 2014; Fu et al., 2016; Haber et al., 2017; Harney et al., 2018; Jeong et al., 2016; Jeong et al., 2020; Jeong et al., 2018; Jones et al., 2015; Kanzawa-Kiriyama et al., 2019; Krzewińska et al., 2018; Lazaridis et al., 2017; Lazaridis et al., 2016; Lazaridis et al., 2014; Lipson et al., 2018; Mathieson et al., 2018; Mathieson et al., 2015; McColl et al., 2018; Moreno-Mayar et al., 2018; Narasimhan et al., 2019; Ning et al., 2020; Raghavan et al., 2014a; Raghavan et al., 2014b; Rasmussen et al., 2014; Rasmussen et al., 2010; Rasmussen et al., 2015; Sikora et al., 2019; Unterlander et al., 2017; Wang et al., 2020; Wang et al., 2021; Yang et al., 2020; Yang et al., 2017; Yu et al., 2020) for population genetic analysis. Also, 1240K dataset was produced by merging present-day world-wide populations genotyped on the 1240K sites (Mallick et al., 2016) with high-coverage genotype data of 104 present-day Ulsan Koreans, six Gunsan jar coffin individuals and

35 Imdang individuals from this study, and previously published ancient individuals (Appendix 1; Appendix 2).

2.6. Principal component analysis

Principal component analysis was implemented with 2,077 present-day Eurasian individuals from HumanOrigins dataset with the option "lsqproject: YES" using smartpca v16000 (Patterson et al., 2006). Gunsan jar coffin individuals, Imdang individuals, present-day Ulsan Koreans, and ancient East Asian individuals were projected onto the principal components (PCs). Also, PCA was ran with 455 present-day East Asian individuals including Ulsan Koreans with the option "lsqproject: YES" and "shrinkmode: YES" using smartpca v16000 for the East Asian-only PCA. Then ancient East Asian individuals including Gunsan jar coffin and Imdang individuals were projected onto the PCs.

2.7. *F*-statistics and qpWave/qpAdm analysis

F-statistics were calculated from 1240K dataset. Outgroup-f_3 statistics (Patterson et al., 2012; Raghavan et al., 2014b) of the form f_3(Mbuti; Three Kingdoms period Korean group, world-wide) were calculated using qp3Pop v650 to obtain a measurement of genetic affinity between two populations. F_4 statistics were calculated using qpDstat v970 (Patterson et al., 2012) with the

"f4mode: YES" option. 121 present-day world-wide populations and 167 ancient populations were used for these calculations (Appendix 1; Appendix 2). Admixture models for ancient and present-day East Asian populations were tested using qpWave v1200 and qpAdm v1201 program in the AdmixTools v7.0 (Lazaridis et al., 2016; Reich et al., 2012) with the "allsnps: NO" option on 1240K dataset. The outgroup set includes central African hunter-gatherers Mbuti (Mbuti.DG; n=5), Early Neolithic farmers from western Anatolia (Anatolia_N; n=23), Andamanese islanders Onge (Onge.DG; n=2), Neolithic Iranians from the Ganj Dareh site (Iran_N; n=8), an Epi-Paleolithic European hunter-gatherer from the Villabruna site (Villabruna; n=1), a Late Pleistocene Native American individual from the Upward Sun River site in Alaska (USR1; n=1), Early Neolithic hunter-gatherers from the western Baikal region (Baikal_EN; n=18), an Early Neolithic individual from Shandong region in China (Boshan; n=1), an Early Neolithic individual from the southern Chinese Liangdao site (Liangdao2; n=1), and Funadomari Jomon (Jomon_Funadomari; n=2).

2.8. Genetic relationship analysis

Genetic relatedness among the Three Kingdoms period individuals from the same site was estimated using PMR and lcMLkin. PMR between each individual was calculated using the autosomal SNPs in the 1240K panel. Probability of sharing 0, 1, and

2 alleles was estimated using lcMLkin v0.5.0 to distinguish between parent-offspring and full siblings (Lipatov et al., 2015). By removing one individual from each first-degree relative pair, three Gunsan individuals and 31 Imdang individuals were kept for the group-based analyses (Table 9; Table 10).

2.9. Runs of Homozygosity analysis

Runs of homozygosity (ROH) segments within the genome of each Three Kingdoms period Korean individuals were estimated to infer parental relatedness using hapROH (downloaded from https://pypi.org/project/hapROH/ v0.3a1) (Ringbauer et al., 2021).

3. Results[①]

3.1. Ancient genome-wide data production

Genetic investigation of Three Kingdoms period individuals from Gunsan jar coffin and Ancient Tombs in Imdang-dong and Joyeong-dong were performed. Based on archaeological contexts and artifacts, the Gunsan individuals and the Imdang individuals are dated to 5th to 7th century AD and 5th to 6th century AD, respectively. Six Gunsan individuals and 35 Imdang individuals were initially screened by shallow shotgun sequencing (Table 2; Table 3). Screened libraries were enriched for 1240K SNP panel (Mathieson et al., 2015) and genome-wide data with 1240K autosomal coverage of 0.1-2.9x and with 104-981K 1240K SNPs covered at least once are obtained. The authenticity of data from these individuals was verified by multiple measures. All of these individuals show post-mortem damages characteristic of ancient DNA. All males except for one individual with low X chromosome coverage show <4.5% nuclear contamination. All individuals with sufficient mitochondrial coverage except for one (JOY012; 10%) show <3% mitochondrial contamination. All six Gunsan individuals and 35 Imdang individuals including two individuals without contamination measure due to low X chromosome or mitochondrial

① The analysis of the Gunsan jar coffin data has been written into a manuscript and has been posted on bioRxiv (Lee et al., 2022). doi:https://doi.org/10.1101/2022.05.09.491093

coverage and 13 shotgun sequenced Imdang individuals (hereafter referred to as "Imdang.SG") (Table 4) were included into the downstream analyses. Haploid genotypes for 1240K panel SNPs were called by randomly sampling a single high-quality base. Previously published data of prsent-day and ancient world-wide individuals genotyped on the HumanOrigins and the 1240K panel SNPs were merged for downstream analyses (Appendix 1; Appendix 2).

Table 2. A summary of sequencing and genetic information of the Three Kingdoms period individuals in this study. Six out of the seven individuals from the Gunsan jar coffin and 35 out of the 46 individuals from the Imdang site yield sufficient levels of human DNA for genome-wide analysis. The number of covered target sites are estimated among 1,233,013 SNPs in the 1240K panel and 593,124 SNPs in the HumanOrigins panel. Contamination estimates based on X chromosome represent the point estimate ± 1 s.e.m.. Mitochondrial estimates represent the point estimate with the 95% credible interval. 13 Imdang individuals with large amount of human DNA are highlighted by bold face.

Lab ID	Genetic Sex	Pre-capture % human DNA	# of reads sequenced	# of uniquely mapped reads	Coverage				# of covered target sites		Post-mortem damage		Contamination estimates	
					Autosome	X	Y	MT	1240K	HumanOrigins	5' C>T	3' G>A	X	MT
GUC001	M	0.129	36,063,678	1,169,573	0.195	0.075	0.091	0.54	189,920	96,738	0.2928	0.2819	0.0184±0.0147	0.01 (0.00-0.02)
GUC002	M	1.219	25,871,919	5,363,973	1.097	0.420	0.518	3.11	639,758	327,372	0.2096	0.1952	0.0020±0.0017	0.01 (0.00-0.02)
GUC003	M	0.307	32,502,379	2,076,521	0.312	0.122	0.147	0.48	284,857	144,896	0.2623	0.2575	0.0106±0.0073	0.01 (0.00-0.02)
GUC004	F	0.350	36,179,211	2,592,575	0.397	0.304	0.004	0.56	342,704	174,677	0.2413	0.2376	N/A	N/A
GUC005	M	2.624	20,992,582	6,805,731	1.839	0.698	0.939	2.17	764,537	389,635	0.2140	0.2012	0.0058±0.0016	0.03 (0.01-0.05)
GUC007	F	1.656	24,249,836	5,438,347	1.087	0.812	0.015	2.22	643,530	330,307	0.2328	0.2119	N/A	0.02 (0.00-0.04)
GUC006	N/A	0.030	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
IMD002	F	11.432	59,364,005	17,530,951	2.283	1.850	0.019	15.53	958,963	493,582	0.0868	0.0836	N/A	0.01 (0.00-0.02)
IMD003	F	33.905	27,876,535	14,483,282	1.762	1.301	0.013	53.31	801,423	419,310	0.1334	0.1346	N/A	0.01 (0.00-0.02)
IMD004	M	10.823	48,892,689	15,798,683	2.505	0.989	1.213	14.49	970,748	496,640	0.0886	0.0766	0.0040±0.0011	0.01 (0.00-0.02)
IMD006	M	2.873	51,471,730	8,661,044	1.641	0.652	0.801	21.42	829,517	430,942	0.1072	0.1042	0.0105±0.0024	0.02 (0.01-0.03)

Table 2. Continued.

Lab ID	Genetic Sex	Pre-capture % human DNA	# of reads sequenced	# of uniquely mapped reads	Coverage				# of covered target sites		Post-mortem damage		Contamination estimates	
					Autosome	X	Y	MT	1240K	HumanOrigins	5' C>T	3' G>A	X	MT
IMD007	F	6.796	54,372,667	14,265,160	2.311	1.768	0.021	97.47	907,536	476,365	0.0927	0.0898	N/A	0.01 (0.00-0.02)
IMD008	F	0.251	48,074,409	869,190	0.094	0.075	0.002	0.40	104,362	54,079	0.1060	0.1009	N/A	0.01 (0.00-0.02)
IMD009	M	4.011	58,207,143	9,993,389	1.307	0.522	0.583	8.21	767,113	394,168	0.1527	0.1481	0.0064±0.0023	0.01 (0.00-0.02)
IMD010	F	39.769	42,051,240	23,584,400	2.687	2.098	0.030	21.22	962,538	497,304	0.1453	0.1319	N/A	0.01 (0.00-0.02)
IMD011	F	11.943	48,359,907	16,651,160	2.506	1.940	0.017	16.58	936,806	484,662	0.0709	0.0745	N/A	0.01 (0.00-0.02)
IMD016	F	28.265	52,306,876	20,634,335	1.771	1.456	0.013	20.81	884,299	456,750	0.0956	0.0950	N/A	0.01 (0.00-0.02)
IMD018	F	0.827	50,684,684	2,251,673	0.255	0.197	0.004	103.77	255,780	134,332	0.1217	0.1193	N/A	0.01 (0.00-0.02)
JOY001	M	2.613	46,760,280	6,353,475	0.889	0.350	0.421	2.09	614,278	314,757	0.1478	0.1412	0.0021±0.0023	0.01 (0.00-0.02)
JOY002	M	88.800	30,279,259	24,765,661	2.876	1.090	1.429	23.56	981,051	503,851	0.0578	0.0493	0.0016±0.0008	0.01 (0.00-0.02)
JOY003	F	84.261	25,298,776	20,081,692	2.640	2.013	0.033	25.29	950,311	494,221	0.0513	0.0470	N/A	0.01 (0.00-0.02)
JOY004	M	43.807	18,803,504	11,153,543	1.528	0.578	0.746	9.31	792,783	408,901	0.1159	0.0973	0.0044±0.0016	0.01 (0.00-0.02)
JOY005	F	33.068	22,540,591	12,211,177	1.905	1.442	0.023	11.84	844,085	439,567	0.1080	0.0926	N/A	0.01 (0.00-0.02)
JOY006	F	7.898	24,250,146	6,559,640	1.012	0.795	0.005	96.38	660,145	347,645	0.1056	0.1022	N/A	0.01 (0.00-0.02)
JOY007	M	33.293	38,894,327	20,411,241	2.728	1.032	1.350	20.67	965,108	495,900	0.0736	0.0578	0.0024±0.0008	0.01 (0.00-0.02)
JOY009	M	0.537	47,470,113	2,017,430	0.273	0.107	0.125	1.27	266,722	137,885	0.1205	0.1110	0.0062±0.0099	0.01 (0.00-0.02)
JOY010	M	4.000	53,350,309	10,583,225	1.845	0.714	0.891	8.33	852,461	439,931	0.1225	0.0991	0.0033±0.0014	0.01 (0.00-0.02)
JOY011	F	6.211	21,000,286	5,903,072	1.051	0.801	0.012	6.33	660,047	343,145	0.0780	0.0630	N/A	0.01 (0.00-0.02)
JOY012	M	0.670	56,549,278	2,801,429	0.373	0.148	0.167	2.30	346,302	179,540	0.1038	0.0906	0.0443±0.0143	0.10 (0.06-0.14)
JOY013	M	2.250	52,520,302	8,145,074	1.483	0.563	0.725	6.28	782,734	404,301	0.0895	0.0673	0.0051±0.0017	0.01 (0.00-0.02)

Table 2. Continued.

Lab ID	Genetic Sex	Pre-capture % human DNA	# of reads sequenced	# of uniquely mapped reads	Coverage				# of covered target sites		Post-mortem damage		Contamination estimates	
					Autosome	X	Y	MT	1240K	HumanOrigins	5' C>T	3' G>A	X	MT
JOY015	F	37.503	32,401,813	16,536,190	2.394	1.880	0.025	18.28	936,901	483,946	0.0956	0.0886	N/A	0.01 (0.00-0.02)
JOY016	F	29.829	19,072,705	9,357,195	1.399	1.065	0.009	13.37	746,413	388,521	0.0883	0.0783	N/A	0.01 (0.00-0.02)
JOY017	M	0.834	76,985,686	1,965,860	0.138	0.058	0.060	9.23	147,618	76,881	0.1360	0.1366	N/A	0.01 (0.00-0.02)
JOY018	M	19.472	38,378,884	15,632,041	2.610	1.036	1.320	8.57	965,974	493,226	0.0920	0.0885	0.0046±0.0012	0.01 (0.00-0.02)
JOY019	F	0.664	48,184,879	2,513,644	0.425	0.337	0.004	1.90	376,644	194,274	0.1083	0.0926	N/A	0.01 (0.00-0.02)
JOY020	M	38.808	39,708,634	21,764,928	2.895	1.097	1.363	12.62	981,658	504,309	0.0921	0.0789	0.0021±0.0008	0.01 (0.00-0.02)
JOY022	M	5.013	53,033,470	11,920,739	1.862	0.734	0.865	8.37	877,015	451,563	0.0933	0.0838	0.0105±0.0021	0.01 (0.00-0.02)
JOY023	F	5.188	71,112,119	15,487,797	1.779	1.366	0.017	20.84	869,168	454,946	0.0874	0.0681	N/A	0.01 (0.00-0.02)
JOY024	F	19.205	39,035,853	17,434,422	2.942	2.254	0.026	8.70	964,378	501,229	0.1018	0.0831	N/A	0.01 (0.00-0.02)
JOY025	M	3.113	57,133,223	10,605,469	1.849	0.711	0.895	9.99	855,314	441,664	0.1142	0.0961	0.0059±0.0016	0.01 (0.00-0.02)
JOY026	F	1.356	55,129,309	5,424,055	0.879	0.664	0.006	3.42	597,439	311,740	0.1139	0.1011	N/A	0.01 (0.00-0.02)
JOY028	F	3.063	59,750,783	7,142,515	0.866	0.697	0.007	5.56	625,751	324,592	0.1353	0.1287	N/A	0.01 (0.00-0.02)
IMD001	N/A	0.078	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
IMD005	N/A	0.018	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
IMD012	N/A	0.048	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
IMD013	N/A	0.008	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
IMD014	N/A	0.065	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
IMD015	N/A	0.046	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
IMD017	N/A	0.018	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Table 2. Continued.

Lab ID	Genetic Sex	Pre-capture % human DNA	# of reads sequenced	# of uniquely mapped reads	Coverage				# of covered target sites		Post-mortem damage		Contamination estimates	
					Autosome	X	Y	MT	1240K	HumanOrigins	5' C>T	3' G>A	X	MT
JOY008	N/A	0.015	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
JOY014	N/A	0.036	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
JOY021	N/A	0.089	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
JOY027	N/A	0.040	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Table 3. A summary of archaeological information and uniparental haplogroup of the Three Kingdoms period individuals in this study. Archaeological ID, date, social status, and age at death inferred from archaeological records, and uniparental haplogroup of each individual are shown. 13 Imdang individuals with large amount of human DNA are highlighted by bold face.

Lab ID	Archaeol.ID	period	Status	Age	MT haplogroup	Y haplogroup
GUC001	01; skull 1; Gunsan jar coffin	5th to 7th century AD	–	50+ years	D4	Q1a (Q-M1081, Q-L472)
GUC002	02; skull 2; Gunsan jar coffin	5th to 7th century AD	–	36–50 years	D4b2b1	Q1a1a1 (Q-M265, Q-M120)
GUC003	03; skull 3; Gunsan jar coffin	5th to 7th century AD	–	21–35 years	..	Q1a1 (Q-F1251, Q-F1096)
GUC004	04; skull 4; Gunsan jar coffin	5th to 7th century AD	–	Adult	..	–
GUC005	06; skull 6; Gunsan jar coffin	5th to 7th century AD	–	21–35 years	D4c (D4c1b1)	O1b2a2a1a (O-CTS7620)
GUC007	09; subadult 3; Gunsan jar coffin	5th to 7th century AD	–	15–18 years	D4	–
GUC006	08; subadult 2; Gunsan jar coffin	5th to 7th century AD	–	6–10 years	N/A	N/A
IMD002	011; Imdang #5B1-1	4th quarter of the 5th century AD	Unknown	36–50 years	C4a1a+195	–
IMD003	012; Imdang #5B1-2	4th quarter of the 5th century AD	Unknown	21–35 years	C4a1a+195	–
IMD004	013; Imdang #5B1-3	4th quarter of the 5th century AD	Sacrificed	15–18 years	M10a1a1a	O2a2b1a2a1 (O-F46)
IMD006	015; Imdang #5B1-5	4th quarter of the 5th century AD	Sacrificed	21–35 years	M10a1b	O1b2a1 (O-CTS10145, O-CTS713)

Table 3. Continued.

Lab ID	Archaeol.ID	period	Status	Age	MT haplogroup	Y haplogroup
IMD007	016; Imdang #5B1-6	4th quarter of the 5th century AD	Sacrificed	36-50 years	B5b1a	-
IMD008	021; Imdang #5D1-1	-	Unknown	21-35 years	M81	-
IMD009	024; Imdang #5D2-2	-	Unknown	36-50 years	D5a1	O1b2a1 (O-CTS10145; O-CTS713)
IMD010	027; Imdang #5D2-5	-	Unknown	Adult	A5a5	-
IMD011	034; Imdang #6A-6	1st quarter of the 6th century AD	Sacrificed	4-8 years	M7a1a1a	-
IMD016	043; Imdang #7B-1	2nd quarter of the 5th century AD	Unknown	8.5-13.5 years	D4b2a	-
IMD018	051; Imdang #7D-1	-	Unknown	21-35 years	D4e2	-
JOY001	052; Joyeong #CI-1-1	3rd quarter of the 5th century AD	Main	31-40 years	C4a1a	O1b2a1 (O-CTS10145; O-CTS713)
JOY002	053; Joyeong #CI-1-2	3rd quarter of the 5th century AD	Sacrificed	15-18 years	F2f	Q1a1a1 (Q-M265; Q-M120)
JOY003	054; Joyeong #CI-1-3	3rd quarter of the 5th century AD	Sacrificed	4-8 years	D4b2a2	-
JOY004	055; Joyeong #CI-1-4	3rd quarter of the 5th century AD	Sacrificed	41-60 years	G1a3	O1b2a1 (O-CTS10145; O-CTS713)
JOY005	056; Joyeong #CI-1-5	3rd quarter of the 5th century AD	Sacrificed	36-50 years	D4b2a2	-
JOY006	058; Joyeong #CI-2-2	3rd quarter of the 5th century AD	Sacrificed	21-35 years	G3a2a	-
JOY007	060; Joyeong #CI-2-4	3rd quarter of the 5th century AD	Sacrificed	15-18 years	G1a1a4	O2a1c1a5 (O-M5420)

Table 3. Continued.

Lab ID	Archaeol.ID	period	Status	Age	MT haplogroup	Y haplogroup
JOY009	064; Joyeong #CII-1-3	1st quarter of the 5th century AD	Sacrificed	About 10 years	B4a1b1	O2a2a1 (O-CTS3106; O-F2588)
JOY010	066; Joyeong #CII-1-5	1st quarter of the 5th century AD	Sacrificed	36-50 years	D4a	C2b (C-L1373)
JOY011	067; Joyeong #CII-1-6	1st quarter of the 5th century AD	Sacrificed	About 10 years	D5b1c	-
JOY012	068; Joyeong #CII-2-1	1st quarter of the 5th century AD	Main	Adult	D4c	O1b (O-L690; O-M268)
JOY013	072; Joyeong #CII-2-6	1st quarter of the 5th century AD	Sacrificed	18-20 years	Y2	O1b2a1 (O-CTS10145; O-CTS713)
JOY015	083; Joyeong #EII-2-1	4th quarter of the 5th century AD	Main	15-18 years	D4c1a	-
JOY016	085; Joyeong #EII-2-3	4th quarter of the 5th century AD	Sacrificed	7.5-12.5 years	N9a4a	-
JOY017	089; Joyeong #EII-3-1	1st quarter of the 6th century AD	Unknown	21-35 years	D5a2b	O1b1a1a1b (O-F3092, O-M1283)
JOY018	090; Joyeong #EII-3-2	1st quarter of the 6th century AD	Unknown	Subadult	N9a4a	O1b2a1 (O-CTS10145; O-CTS713)
JOY019	091; Joyeong #EII-3-3	1st quarter of the 6th century AD	Sacrificed	Adult	D5a2	-
JOY020	092; Joyeong #EII-3-4	1st quarter of the 6th century AD	Sacrificed	21-35 years	B4a1c1a1	O1b2a2a1 (O-Z24624; O-CTS723)
JOY022	097; Joyeong #EII-13-1	-	Unknown	21-35 years	D4a1	O1b2a2 (O-L682)
JOY023	100; Joyeong #EIII-2-2	1st quarter of the 5th century AD	Sacrificed	21-35 years	B4a2b1	-
JOY024	102; Joyeong #EIII-2-4	1st quarter of the 5th century AD	Sacrificed	21-35 years	G2a5	-

Table 3. Continued.

Lab ID	Archaeol.ID	period	Status	Age	MT haplogroup	Y haplogroup
JOY025	103; Joyeong #EIII-2-5	1st quarter of the 5th century AD	Sacrificed	21-35 years	B4b1a1b	N (N-M232; N-M231)
JOY026	108; Joyeong #EIII-4-2	2nd quarter of the 5th century AD	Sacrificed	31-50 years	D4a	-
JOY028	163; Joyeong #1A-6-1	-	Main	21-35 years	D4b2b2b	-
IMD001	005; Imdang #2 North Tomb 3	4th quarter of the 5th century AD	Unknown	21-35 years	N/A	N/A
IMD005	014; Imdang #5B1-4	4th quarter of the 5th century AD	Sacrificed	21-35 years	N/A	N/A
IMD012	036; Imdang #6B2-1	-	Unknown	21-35 years	N/A	N/A
IMD013	038; Imdang #7A-2	1st quarter of the 5th century AD	Sacrificed	8.5-13.5 years	N/A	N/A
IMD014	040; Imdang #7A2-1	1st quarter of the 5th century AD	Sacrificed	Adult	N/A	N/A
IMD015	041; Imdang #7A2-2	1st quarter of the 5th century AD	Sacrificed	Adult	N/A	N/A
IMD017	046; Imdang #7B-4	2nd quarter of the 5th century AD	Sacrificed	18-25 years	N/A	N/A
JOY008	063; Joyeong #CII-1-2	1st quarter of the 5th century AD	Sacrificed	16-18 years	N/A	N/A
JOY014	074; Joyeong #CII-2-8	1st quarter of the 5th century AD	Sacrificed	21-35 years	N/A	N/A
JOY021	096; Joyeong #EII-7-1	-	Unknown	21-40 years	N/A	N/A
JOY027	114; Joyeong #EIII-9-1	-	Main	Adult	N/A	N/A

Table 4. A summary of deep shotgun sequencing information of 13 individuals from the Imdang site. 13 Imdang individuals with especially large amount of human DNA are shotgun sequenced deeply and included into the downstream analyses. The number of covered target sites are estimated among 1,233,013 SNPs in the 1240K panel and 593,124 SNPs in the HumanOrigins panel. Contamination estimates based on X chromosome represent the point estimate ± 1 s.e.m.. Mitochondrial estimates represent the point estimate with the 95% credible interval.

Lab ID	# of reads sequenced	# of uniquely mapped reads	Coverage				# of covered target sites		Post-mortem damage		Contamination estimates	
			Autosome	X	Y	MT	1240K	HumanOrigins	5' C>T	3' G>A	X	MT
IMD003.A0101_SE	251,419,434	44,429,264	0.416	0.365	0.004	136.07	516,105	262,518	0.1387	0.1368	0.2172±0.0416	0.01(0.00-0.02)
IMD004.A0101_SE	270,220,007	18,354,716	0.262	0.117	0.123	14.97	356,255	184,842	0.0919	0.0906	N/A	0.01(0.00-0.02)
IMD010.A0101_SE	207,416,757	54,604,309	0.648	0.578	0.007	40.42	689,322	353,784	0.1488	0.1501	N/A	0.00(0.00-0.01)
IMD011.A0101_SE	273,171,759	20,437,164	0.248	0.224	0.002	17.16	343,086	177,044	0.0747	0.0727	0.1899±0.0102	0.01(0.00-0.02)
IMD016.A0101_SE	257,934,813	37,167,171	0.421	0.387	0.004	25.90	510,449	262,561	0.0971	0.0977	0.1880±0.0064	0.01(0.00-0.02)
JOY002.A0101_SE	137,139,244	91,832,233	1.511	0.649	0.675	88.39	1,023,662	517,401	0.0556	0.0585	0.0036±0.0008	0.01(0.00-0.02)
JOY004.A0101_SE	139,804,441	43,331,033	0.595	0.262	0.279	32.16	650,855	333,820	0.1149	0.1187	0.0039±0.0018	0.01(0.00-0.02)
JOY005.A0101_SE	235,248,102	53,929,366	0.721	0.625	0.008	44.50	726,383	374,301	0.1138	0.1153	0.2022±0.0042	0.01(0.00-0.02)
JOY007.A0101_SE	199,678,318	48,186,861	0.759	0.324	0.340	51.24	747,970	385,404	0.0730	0.0735	0.0038±0.0013	0.00(0.00-0.01)
JOY015.A0101_SE	166,995,815	36,685,618	0.485	0.430	0.005	32.86	572,038	293,644	0.0962	0.0981	0.1986±0.0055	0.01(0.00-0.02)
JOY016.A0101_SE	183,138,836	34,105,570	0.456	0.393	0.003	44.31	543,806	281,958	0.0902	0.0886	0.1902±0.0059	0.01(0.00-0.02)
JOY020.A0101_SE	167,456,912	46,353,782	0.672	0.281	0.286	23.65	696,471	362,663	0.0940	0.0936	0.0006±0.0010	0.01(0.00-0.02)
JOY024.A0101_SE	164,929,519	22,834,888	0.313	0.268	0.003	9.65	412,322	216,075	0.1024	0.1004	0.1713±0.0081	0.01(0.00-0.02)

3.2. The genetic profiles of Three Kingdoms period Koreans

To investigate the genetic profiles of the Three Kingdoms period individuals from two different sites, multiple methods for population genetic analysis were used. First, PCA was implemented with 2,077 present-day Eurasians. Then, the Three Kingdoms period Koreans were projected onto the top PCs (Figure 5). All of the Three Kingdoms period individuals except two form a tight cluster with each other and present-day Koreans. Two Imdang individuals slightly fall outside of the cluster along PC1, but this is likely an artificial effect due to their low coverage (IMD008, 0.094x; JOY017, 0.138x). Likewise, in East Asian-only PCA, the Three Kingdoms period Koreans also form a tight cluster with each other and present-day Koreans (Figure 6).

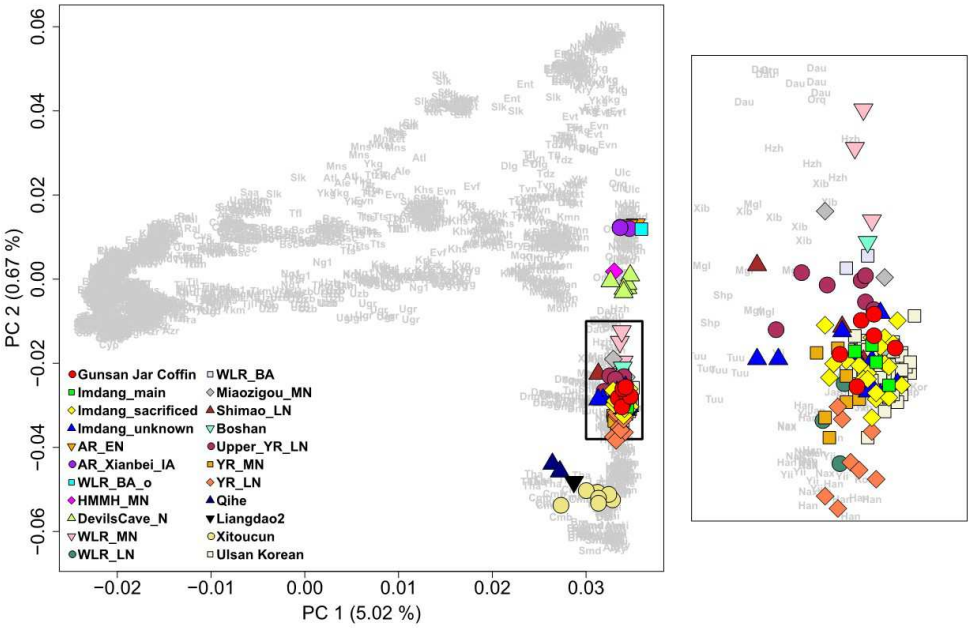
Second, outgroup-f_3 statistics of the form f_3(Mbuti; Gunsan/Imdang, world-wide) were calculated to measure genetic affinity between the Three Kingdoms period individuals and present-day and ancient world-wide populations (Figure 7). Gunsan individuals show highest genetic affinity with capture sequenced Imdang individuals followed by present-day Koreans. Imdang individuals show slightly different results according to sequencing methods, but they both share the most amount of genetic affinity with present-day Koreans.

Third, f_4 statistics of the form f_4(Mbuti, world-wide; present-

day Koreans, Gunsan/Imdang) were calculated to test the genetic symmetry of present-day Koreans and the Three Kingdoms period individuals in terms of their genetic affinity to present-day and ancient world-wide populations (Figure 8). Some groups break the symmetry between present-day Kroeans and the Three Kingdoms period individuals without clear geographic distribution.

Then by using qpAdm, a statistical tool for testing admixture model (Lazaridis et al., 2016; Reich et al., 2012), the genetic relationship between the Three Kingdoms period individuals and present-day Koreans was modeled (Table 5; Table 6). Present-day Ulsan Koreans are adequately modeled as a mixture of the Gunsan individuals or 13 shotgun sequenced Imdang individuals ("Imdang.SG") and a European source with a small negative coefficient. Conversely, Gunsan and Imdang.SG individuals are modeled as a mixture of Ulsan Koreans and a small proportion of a European source. These results of admixture model between the Three Kingdoms period individuals and Ulsan Koreans are likely due to a technical artifact from sample contamination or small reference bias rather than the genetic contribution of European ancestry, considering the small amount of nuclear contamination detected in the Three Kingdoms period individuals. Unlike admixture models using Imdang.SG, models using capture sequenced Imdang individuals ("Imdang") as a target or source do not fit the data properly, likely due to the technical artifact introduced from the sequencing methods.

In addition to proximal admixture model, the Three Kingdoms period individuals were modeled using more distal ancestries. Using Bronze Age individuals from the Longtoushan site (“WLR_BA”) as an intermediate northern proxy and Late Neolithic individuals from the Xitoucun site in southern China (“Xitoucun”) as a southern proxy, ancient and present-day mainland East Asian populations including the Three Kingdoms period individuals are properly positioned along the north-south genetic cline of East Asian gene pool (Figure 9; Table 7A). In three-way admixture model using Jomon_Ikawazu as the third component, no significant amount of Jomon contribution are detected in all target East Asian populations except for present-day Japanese. Another three-way admixture model using Middle Neolithic individuals from the Miaozigou site (“Miaozigou_MN”) as an intermediate northern proxy instead of WLR_BA shows small but significant amount of Jomon admixture in the Three Kingdoms period individuals and present-day Ulsan Koreans (Table 7B). However, given the location of the Longtoushan site distant to coast, distant spatiotemporal relationship between Miaozigou_MN and from the Three Kingdoms period to present-day Koreans, and genetic heterogeneity of Miaozigou_MN individuals, Jomon component in Koreans of this model is likely implausible.



Abz Abazin; Abk Abkhasian; Adg Adygei; Abn Albanian; Ale Aleut; Atl Aleut_Tlingit; Alt Altaian; Ack Altaian_Chelkans; Ami Ami; Arm Armenian; Ahm Armenian_Hemsheni; Aty Atayal; Avr Avar; Azr Azeri; Blk Balkar; Bsc Bashkir; Bsq Basque; Blr Belarusian; Bes Besermyan; Brn Borneo; Blg Bulgarian; Bry Buryat; Cmb Cambodian; Cch Chechen; Chk Chukchi; Cvs Chuvash; Ccs Circassian; Crt Croatian; Cyp Cypriot; Cze Czech; Dai Dai; Drg Darginian; Dau Daur; Dlg Dolgan; Dng Dungan; Ent Enets; Eng English; Ecs Eskimo_CS; Enk Eskimo_NK; Est Estonian; Evn Even; Evf Evenk_FarEast; Evt Evenk_Transbaikal; Ezd Ezid; Fin Finnish; Fre French; Ggz Gagauz; Grg Georgian; Ger German; Grk Greek; Han Han; Hzh Hezhen; Hun Hungarian; Ice Icelandic; Igs Ingushian; Iri Irish; Iru Irish_Ulster; Itn Italian_North; Its Italian_South; Ite Itelmen; Jap Japanese; Jas Jew_Ashkenazi; Jgr Jew_Georgian; Kbd Kabardinian; Ktg Kaitag; Klm Kalmyk; Krc Karachai; Krk Karakalpak; Krl Karelian; Kzk Kazakh; Ket Ket; Khs Khakass; Khk Khakass_Kachins; Kmn Khamnegan; Kin Kinh; Kor Korean; Kry Koryak; Kbc Kubachinian; Kmk Kumyk; Krd Kurd; Kyg Kyrgyz; Lah Lahu; Lak Lak; Lzg Lezgin; Lth Lithuanian; Mlt Maltese; Mns Mansi; Mia Miao; Mld Moldavian; Mon Mongol; Mgl Mongola; Mdv Mordovian; Nan Nanai; Nax Naxi; Ngd Negidal; Nga Nganasan; Nvh Nivh; Ng1 Nogai; Nwg Norwegian; Orc Orcadian; Orq Oroqen; Ost Ossetian; Pol Polish; Rom Romanian; Rus Russian; Rak Russian_Krasnoborsky; Ral Russian_Leshukonsky; Rap Russian_Pinezhsky; Saa Saami; Sar Sardinian; Sct Scottish; Slk Selkup; Smd Semende; She She; Shp Sherpa; Stl Shetlandic; Skh Shor_Khakassia; Smn Shor_Mountain; Scl Sicilian; Srb Sorb; Spa Spanish; Spn Spanish_North; Tbs Tabasaran; Tjl Tajik; Ttk Tatar_Kazan; Ttm Tatar_Mishar; Tts Tatar_Siberian; Ttz Tatar_Zabolotniye; Tha Thai; Tib Tibetan; Tdz Todzin; Tfl Tofalar; Tuu Tu; Tbl Tubalar; Tuj Tujia; Tra Turkish; Trb Turkish_Balikesir; Tkm Turkmen; Tvn Tuvinian; Udm Udmurt; Ukr Ukrainian; Ulc Ulchi; Ugr Uygur; Uzb Uzbek; Vep Veps; Xib Xibo; Ykt Yakut; Yil Yi; Ykg Yukagir

Figure 5. Principal component analysis from present-day Eurasian individuals. Principal components are constructed from 2,077 present-day Eurasian individuals. Ancient East Asian individuals including the Three Kingdoms period individuals and present-day Koreans from Ulsan are projected onto the top two PCs.

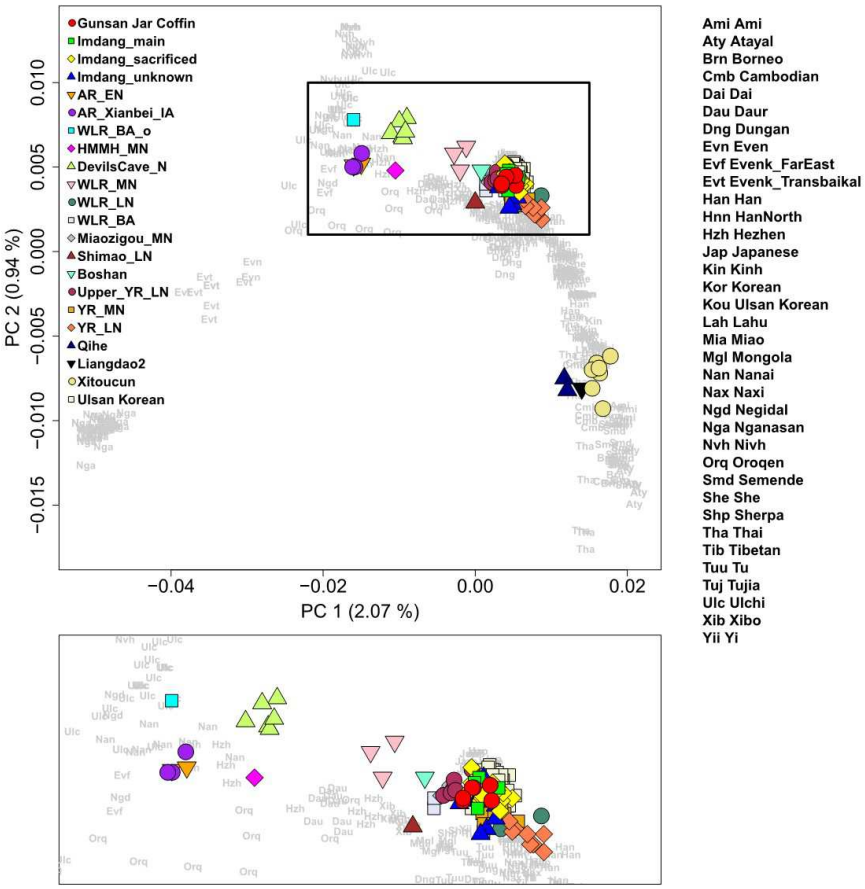


Figure 6. Principal component analysis from present-day East Asian individuals. Principal components are constructed from 455 present-day East Asian individuals including present-day Koreans from Ulsan. Ancient East Asian individuals including the Three Kingdoms period individuals are projected onto the top two PCs.

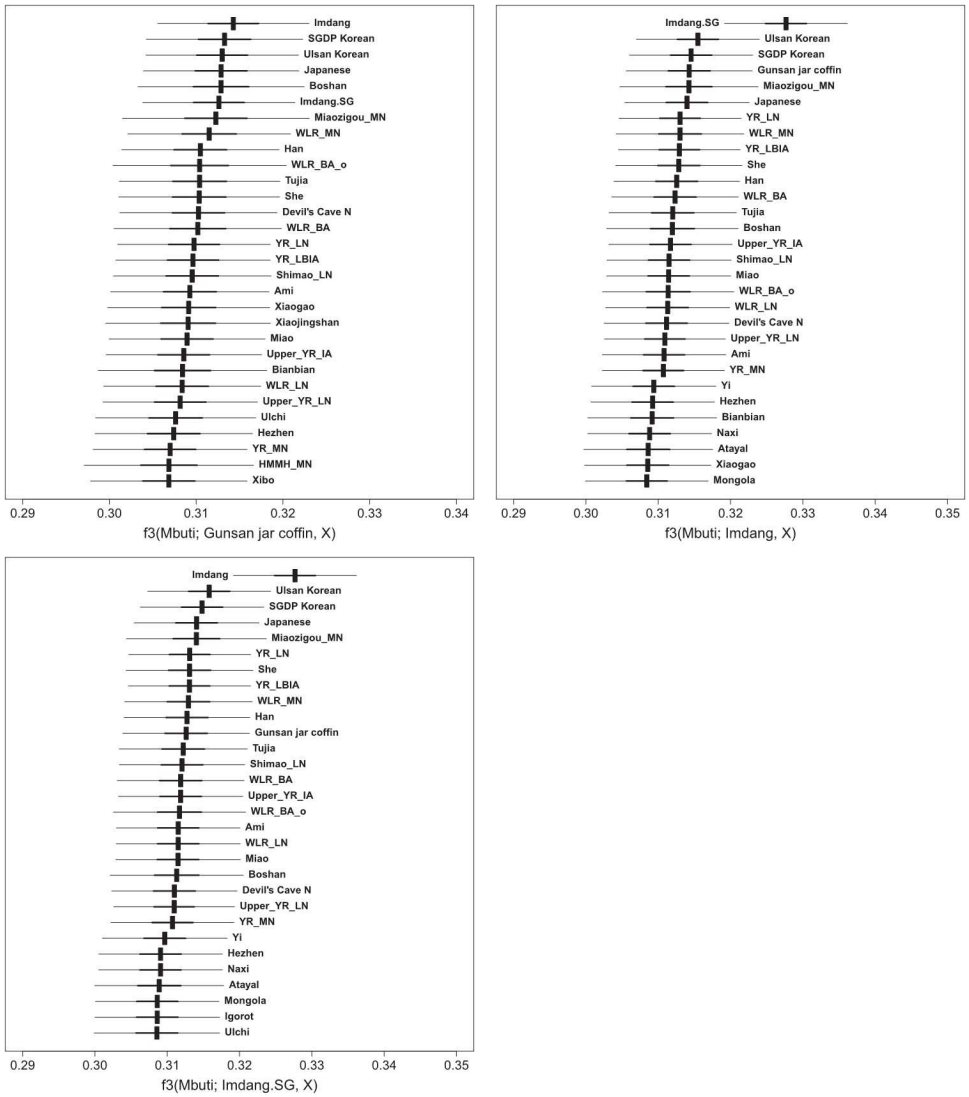


Figure 7. Top 30 outgroup−f_3 statistics of the form f_3(Mbuti; Three Kingdoms period Korean group, world−wide) across ancient and present−day world−wide populations. Gunsan individuals present the highest genetic affinity with Imdang. Imdang individuals present the highest genetic affinity with present−day Koreans. Horizontal bars show the point estimate ± 3 (thin) and ± 1 (thick) standard error measure (s.e.m.), respectively. S.e.m. are calculated by 5cM block jackknifing.

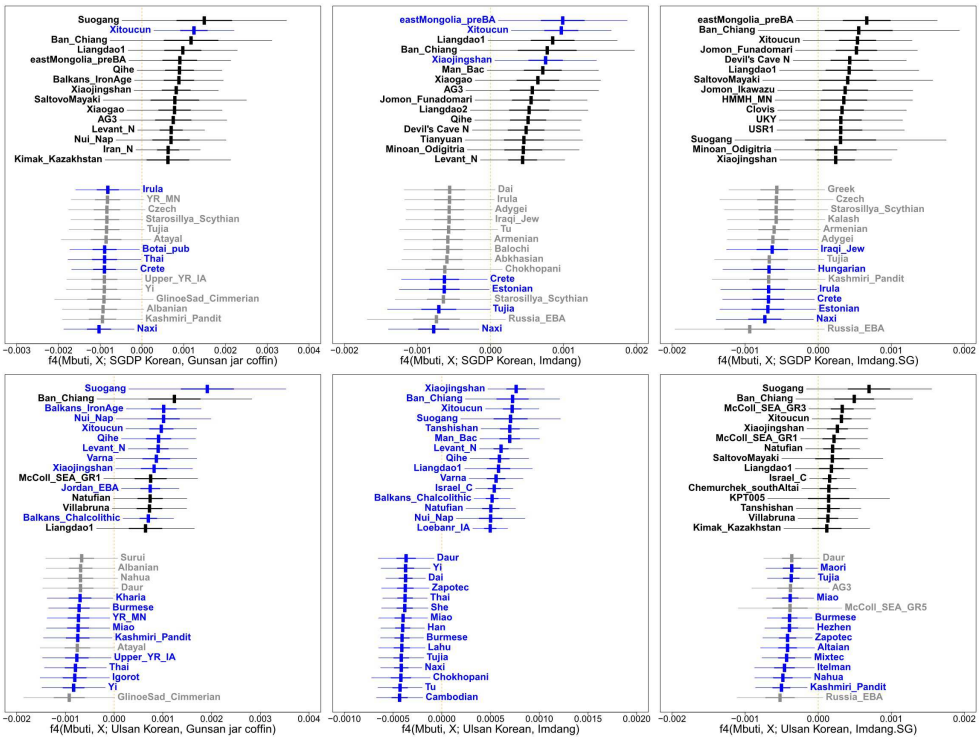


Figure 8. Genetic symmetry test of the Three Kingdoms period individuals and present-day Koreans. 15 most positive and negative f_4 statistics of the form f_4(Mbuti, world-wide; present-day Korean, Three Kingdoms period Korean group) across ancient and present-day world-wide populations are shown. F_4 statistics with $|Z| > 3$ are colored by blue. Horizontal bars represent the point estimate ± 3 (thin) and ± 1 (thick) standard error measure (s.e.m.), respectively. S.e.m. are calculated by 5cM block jackknifing.

Table 5. Admixture model of the Gunsan individuals and present-day Koreans using proximal sources. (A) The observed difference between the Gunsan individuals and the Ulsan Koreans is explained by using a small amount of European ancestry, here represented by the Early Neolithic Central Europeans ("LBK_EN") and the Middle-Late Bronze Age Russian Steppe population ("Sintashta_MLBA"), or AR_Xianbei_IA. (B) Non-significant contribution of AR_Xianbei_IA in a three-way model of Gunsan jar coffin+AR_Xianbei_IA+European suggests that the artifactual European ancestry is a more plausible explanation. *P*-values >0.05 are highlighted by bold face.

A. Two-way admixture models

Target	Ref_1	Ref_2	*P*-value	$Coef_1$	$Coef_2$	s.e.m.
Ulsan Korean	Gunsan jar coffin	AR_Xianbei_IA	**2.19×10^{-1}**	0.908	0.092	0.045
		Jomon_Ikawazu	**9.51×10^{-2}**	1.004	−0.004	0.007
		LBK_EN	**1.02×10^{-1}**	1.010	−0.010	0.005
		Sintashta_MLBA	**1.36×10^{-1}**	1.014	−0.014	0.006
Gunsan jar coffin	Ulsan Korean	AR_Xianbei_IA	**2.23×10^{-1}**	1.101	−0.101	0.054
		Jomon_Ikawazu	**9.62×10^{-2}**	0.996	0.004	0.007
		LBK_EN	**1.03×10^{-1}**	0.990	0.010	0.005
		Sintashta_MLBA	**1.38×10^{-1}**	0.986	0.014	0.006

B. Three-way admixture models

Target	Ref_1	Ref_2	Ref_3	*P*-value	$Coef_1$	$Coef_2$	$Coef_3$
Ulsan Korean	Gunsan jar coffin	AR_Xianbei_IA	LBK_EN	**5.46×10^{-1}**	0.943 (0.048)	0.072 (0.046)	−0.015 (0.006)
		AR_Xianbei_IA	Sintashta_MLBA	**6.72×10^{-1}**	0.945 (0.048)	0.074 (0.046)	−0.020 (0.007)
Gunsan jar coffin	Ulsan Korean	AR_Xianbei_IA	LBK_EN	**5.51×10^{-1}**	1.055 (0.054)	−0.072 (0.053)	0.017 (0.006)
		AR_Xianbei_IA	Sintashta_MLBA	**6.80×10^{-1}**	1.052 (0.054)	−0.074 (0.052)	0.022 (0.008)

Table 6. Admixture model of individuals from the Imdang site and present-day Koreans using proximal sources. Shotgun sequenced individuals from the Imdang site ("Imdang.SG") are adequately modeled as a mixture of present-day Koreans from Ulsan with a small amount of European ancestry, here represented by the Early Neolithic Central Europeans ("LBK_EN") and by the Middle-Late Bronze Age Russian Steppe population ("Sintashta_MLBA"). This small amount of European source is most likely an artifact due to a small amount of contamination and the reference bias. *P*-values >0.05 are highlighted by bold face.

Target	Ref_1	Ref_2	*P*-value	$Coef_1$	$Coef_2$	s.e.m.
Imdang	Ulsan Korean	AR_Xianbei_IA	7.29×10^{-9}	1.053	−0.053	0.023
Imdang	Ulsan Korean	Jomon_Ikawazu	1.69×10^{-12}	1.001	−0.001	0.002
Imdang	Ulsan Korean	LBK_EN	1.62×10^{-6}	0.987	0.013	0.002
Imdang	Ulsan Korean	Sintashta_MLBA	2.24×10^{-6}	0.984	0.016	0.002
Imdang.SG	Ulsan Korean	AR_Xianbei_IA	**2.90×10^{-1}**	1.010	−0.010	0.028
Imdang.SG	Ulsan Korean	Jomon_Ikawazu	**2.06×10^{-1}**	1.000	0.000	0.004
Imdang.SG	Ulsan Korean	LBK_EN	**6.79×10^{-1}**	0.994	0.006	0.003
Imdang.SG	Ulsan Korean	Sintashta_MLBA	**5.75×10^{-1}**	0.992	0.008	0.003
Ulsan Korean	Imdang	AR_Xianbei_IA	6.17×10^{-9}	0.950	0.050	0.021
Ulsan Korean	Imdang	Jomon_Ikawazu	1.62×10^{-12}	0.999	0.001	0.002
Ulsan Korean	Imdang	LBK_EN	1.69×10^{-6}	1.013	−0.013	0.002
Ulsan Korean	Imdang	Sintashta_MLBA	2.29×10^{-6}	1.016	−0.016	0.002
Ulsan Korean	Imdang.SG	AR_Xianbei_IA	**2.86×10^{-1}**	0.990	0.010	0.028
Ulsan Korean	Imdang.SG	Jomon_Ikawazu	**2.04×10^{-1}**	1.000	0.000	0.004
Ulsan Korean	Imdang.SG	LBK_EN	**6.78×10^{-1}**	1.006	−0.006	0.003
Ulsan Korean	Imdang.SG	Sintashta_MLBA	**5.73×10^{-1}**	1.007	−0.007	0.003

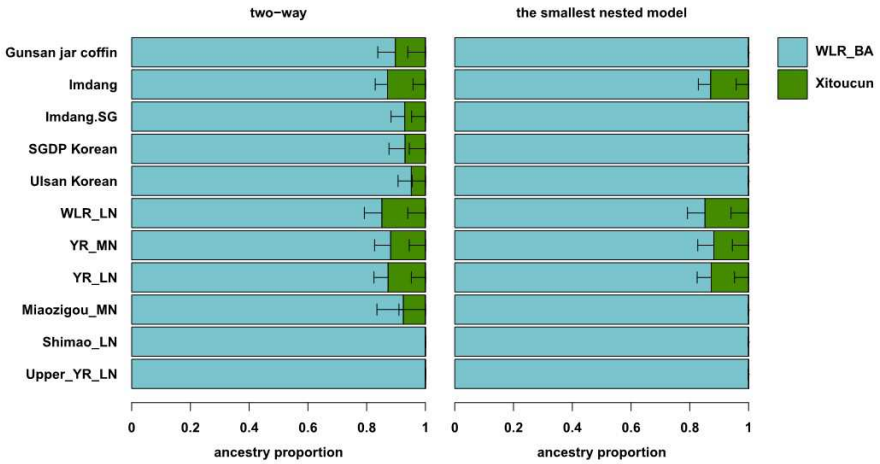


Figure 9. QpAdm-based admixture modeling of Three Kingdoms period Koreans and other ancient and present-day mainland East Asian populations. The results of a two-way admixture model composed of WLR_BA and Xitoucun are presented on the left side. The results of its smallest sub-model after removing all components that do not significantly increase model fit are represented on the right side. Horizontal bars represent the standard error measure (s.e.m.) of ancestry proportion estimates, calculated by the 5 cM block jackknifing procedure as implemented in the qpAdm program.

Table 7. Admixture model of the Three Kingdoms period and present-day Koreans and nearby ancient East Asian groups using distal sources. (A) Two-way and three-way admixture model of WLR_BA+Xitoucun and WLR_BA+Xitoucun+Jomon_Ikawazu are shown. (B) A three-way model of Miaozigou_MN+Xitoucun+Jomon_Ikawazu is shown. Each group name indicates the estimated ancestry proportion (± 1 s.e.m.). *P*-values >0.05 are highlighted by bold face.

A. qpAdm admixture modelling results using WLR_BA as a northern proxy

Target	*P*-value	WLR_BA	Xitoucun	Jomon_Ikawazu
Gunsan jar coffin	**4.59×10^{-1}**	0.898±0.060	0.102±0.060	
Gunsan jar coffin	**1.14×10^{-1}**	0.877±0.066	0.114±0.066	0.008±0.013
Imdang	**3.84×10^{-1}**	0.871±0.042	0.129±0.042	
Imdang	**5.46×10^{-2}**	0.836±0.048	0.165±0.049	−0.001±0.010
Imdang.SG	**3.50×10^{-1}**	0.930±0.047	0.070±0.047	
Imdang.SG	**7.16×10^{-2}**	0.894±0.052	0.104±0.054	0.002±0.011
SGDP Korean	**2.44×10^{-1}**	0.931±0.055	0.069±0.055	
SGDP Korean	**3.50×10^{-1}**	0.880±0.057	0.144±0.058	−0.024±0.012
Ulsan Korean	**2.20×10^{-1}**	0.952±0.045	0.048±0.045	
Ulsan Korean	2.07×10^{-2}	0.921±0.051	0.077±0.052	0.002±0.010
WLR_LN	4.89×10^{-2}	0.852±0.060	0.148±0.060	
WLR_LN	2.97×10^{-2}	0.758±0.070	0.274±0.072	−0.032±0.013
YR_MN	**1.08×10^{-1}**	0.882±0.055	0.118±0.055	
YR_MN	**6.71×10^{-2}**	0.822±0.059	0.197±0.061	−0.019±0.012
YR_LN	**1.13×10^{-1}**	0.873±0.048	0.127±0.048	
YR_LN	1.49×10^{-2}	0.847±0.055	0.162±0.057	−0.009±0.011
Miaozigou_MN	**1.57×10^{-1}**	0.925±0.090	0.075±0.090	
Miaozigou_MN	**1.56×10^{-1}**	0.913±0.104	0.118±0.107	−0.030±0.023
Shimao_LN	**1.75×10^{-1}**	1.056±0.064	−0.056±0.064	
Shimao_LN	**1.11×10^{-1}**	1.021±0.071	0.002±0.072	−0.023±0.015
Upper_YR_LN	**7.69×10^{-1}**	1.021±0.055	−0.021±0.055	
Upper_YR_LN	**8.91×10^{-1}**	0.983±0.058	0.036±0.059	−0.020±0.012
SGDP Japanese	4.37×10^{-18}	1.020±0.085	−0.020±0.085	
SGDP Japanese	**6.72×10^{-2}**	0.816±0.054	0.074±0.054	0.111±0.012

B. qpAdm admixture modelling results using Miaozigou_MN as a northern proxy

Target	*P*-value	Miaozigou_MN	Xitoucun	Jomon_Ikawazu
Gunsan jar coffin	**2.53×10^{-1}**	0.908±0.082	0.048±0.088	0.044±0.018
Imdang	**3.23×10^{-1}**	0.856±0.059	0.119±0.063	0.025±0.013
Imdang.SG	**1.82×10^{-1}**	0.911±0.068	0.060±0.072	0.029±0.015
SGDP Korean	**8.69×10^{-1}**	0.919±0.070	0.061±0.073	0.020±0.016
Ulsan Korean	**2.49×10^{-1}**	0.940±0.062	0.028±0.065	0.031±0.014
WLR_LN	**9.30×10^{-1}**	0.873±0.080	0.124±0.084	0.002±0.018
YR_MN	**2.01×10^{-1}**	0.784±0.076	0.214±0.080	0.002±0.016
YR_LN	**8.84×10^{-2}**	0.860±0.063	0.116±0.066	0.024±0.014
Shimao_LN	**5.72×10^{-1}**	1.014±0.082	−0.018±0.086	0.004±0.018
Upper_YR_LN	**3.55×10^{-1}**	1.003±0.075	−0.016±0.078	0.013±0.017
SGDP Japanese	**3.18×10^{-1}**	0.803±0.064	0.059±0.068	0.138±0.014

3.3. Genetic similarities between the Three Kingdoms period Koreans

Genetic cladality between the Three Kingdoms period individuals was formally tested. First, genetic cladality between different social statuses of Imdang individuals was tested. Imdang individuals were classified into three social statuses based on archaeological records. Masters of the tombs, sacrificed individuals, and ambiguous individuals are marked as Imdang_main, Imdang_sacrificed, and Imdang_unknown, respectively (Table 3). The f_4 statistics of the form f_4(Mbuti, world-wide; Imdang1, Imdang2) are nonsignificant ($|Z|<3$) for most ancient and present-day world-wide populations, and only a few groups break symmetry without clear geographic distribution (Figure 10). qpWave analysis cannot distinguish each pair of Imdang groups belong to different social statuses ($p>0.05$), suggesting symmetric relationship between the Imdang individuals from different social statuses in terms of their affinity to the outgroups (Table 8A). Second, genetic caldality between Gunsan individuals and Imdang individuals was tested. The f_4 statistics of the form f_4(Mbuti, world-wide; Gunsan jar coffin, Imdang) are nonsignificant ($|Z|<3$) and qpWave analysis cannot distinguish Imdang and Gunsan individuals, suggesting no substantial genetic differences between the Three Kingdoms period Koreans from Gunsan and Imdang (Figure 11; Table 8B).

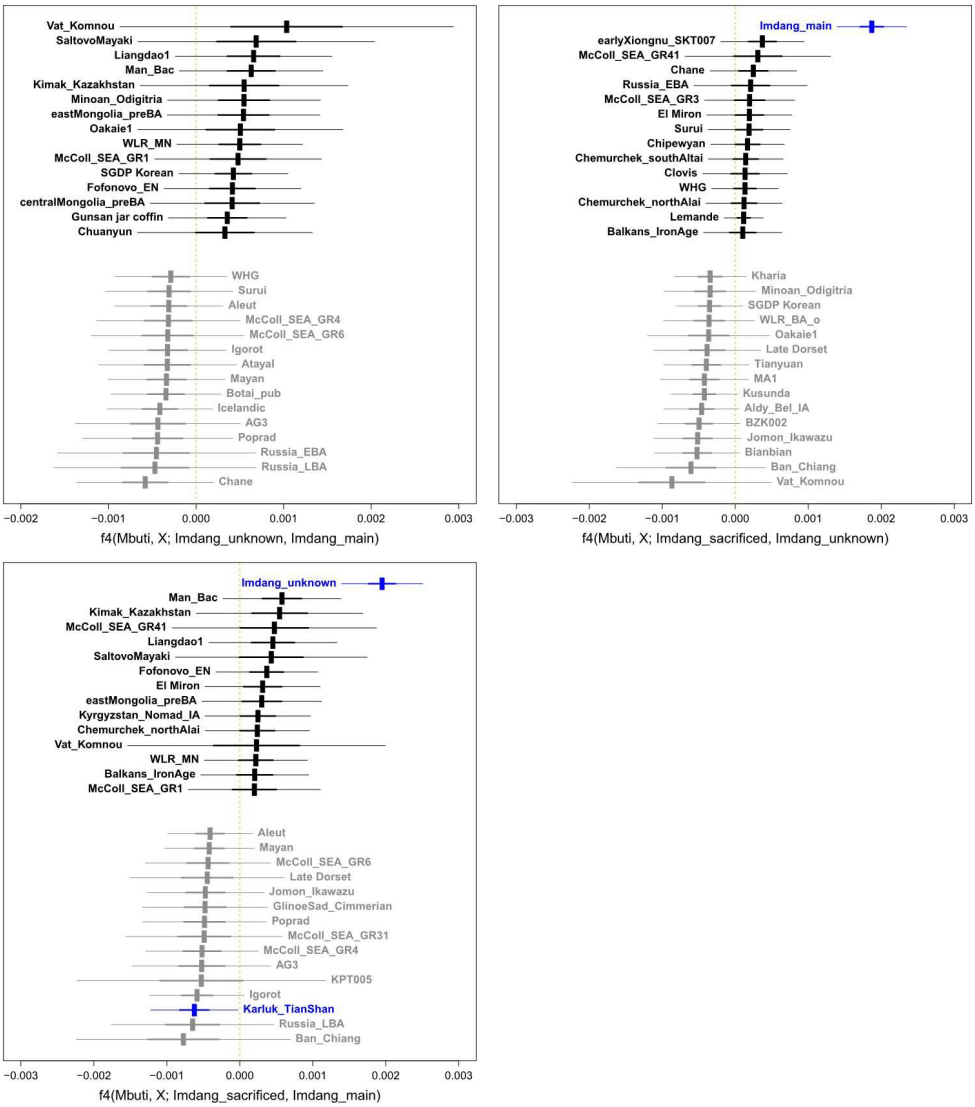


Figure 10. Genetic symmetry test between different social statuses of Imdang individuals. 15 most positive and negative f_4 statistics of the form f_4(Mbuti, world−wide; Imdang1, Imdang2) across ancient and present−day world−wide populations are shown. F_4 statistics with |Z| > 3 are colored by blue. Horizontal bars represent the point estimate ± 3 (thin) and ± 1 (thick) standard error measure (s.e.m.), respectively. S.e.m. are calculated by 5cM block jackknifing.

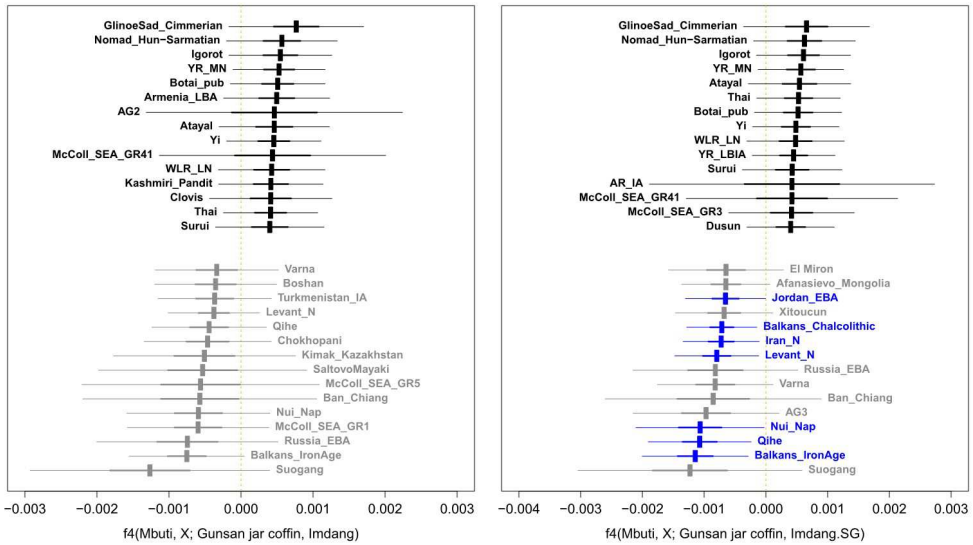


Figure 11. Genetic symmetry test of the Three Kingdoms period individuals from different sites. 15 most positive and negative f_4 statistics of the form f_4(Mbuti, world-wide; Gunsan jar coffin, Imdang) across ancient and present-day world-wide populations are shown. F_4 statistics with |Z| > 3 are colored by blue. Horizontal bars represent the point estimate ± 3 (thin) and ± 1 (thick) standard error measure (s.e.m.), respectively. S.e.m. are calculated by 5cM block jackknifing.

Table 8. Genetic cladality test between the Three Kingdoms period individuals. (A) Genetic cladality between the three social statuses of Imdang individuals is tested using the qpWave program. Non-significant *p*-values (p>0.05) support that different social statuses of the Imdang individuals are symmetrically related to the outgroups. (B) Genetic cladality between the Three Kingdoms period individuals from each site and present-day Koreans is tested using the qpWave program. Non-significant *p*-values (p>0.05) of the test between the Imdang individuals and Gunsan jar coffin individuals support that these two Three Kingdoms period groups are symmetrically related to the outgroups. *P*-values >0.05 are highlighted by bold face.

A. qpWave *p*-value for the test of cladality between different social statuses of Imdang individuals

Target1	Target2	Target3	*P*-value
Imdang_main	Imdang_unknown		**9.56×10^{-1}**
Imdang_main	Imdang_sacrificed		**7.16×10^{-1}**
Imdang_unknown	Imdang_sacrificed		**1.43×10^{-1}**
Imdang_main	Imdang_sacrificed	Imdang_unknown	**9.42×10^{-1}**

B. qpWave *p*-value for the test of cladality between Three Kingdoms period Korean individuals and present-day Korean individuals

Target1	Target2	*P*-value
Gunsan jar coffin	Imdang	**3.65×10^{-1}**
Gunsan jar coffin	Imdang.SG	**5.31×10^{-2}**
Gunsan jar coffin	SGDP Korean	2.93×10^{-2}
Gunsan jar coffin	Ulsan Korean	3.00×10^{-2}
Imdang	Imdang.SG	1.72×10^{-7}
Imdang	SGDP Korean	3.42×10^{-3}
Imdang	Ulsan Korean	1.93×10^{-17}
Imdang.SG	SGDP Korean	**1.30×10^{-1}**
Imdang.SG	Ulsan Korean	**1.88×10^{-1}**

3.4. A familial relationship of Three Kingdoms period individuals

Genetic relatedness between the Three Kingdoms period individuals was measured to test if the specific burial patterns of Gunsan jar coffin and Imdang site reflect close genetic relationship between the individuals. Indeed, estimated genetic relatedness showed closely related individuals among them from each site. Six individuals from the Gunsan jar coffin form a closely related extended family: six first-degree, two second-degree, and three third-degree or more distant relative pairs are observed from 15 pairs combined from six individuals (Figure 12A; Table 9). A most plausible pedigree was generated by incorporating the estimated genetic relatedness, uniparental haplogroup information and age at death (Fig 12B). The pedigree shows an extended family consisting of a quartet family and both paternal and maternal relatives. A couple and their two children organize the quartet family. One male (GUC005) is a first-degree relative of the quartet mother and a second-degree relative of two children, but unrelated to the quartet father. Relationship between GUC005 and GUC004 is more likely parent-offspring relationship than full sibling, because the probability of sharing both alleles at the same site is near-zero. Although both father-daughter and mother-son relationship are possible, a father-daughter relationship is suggested as a more probable relationship because GUC005 shares the mitochondrial

haplogroup with the quartet children. Therefore, GUC005 and (GUC001, GUC007) are likely half-siblings sharing the same mother. GUC003 and the quartet father are probably third-degree relatives and share the same Y haplogroup.

Then, genetic relatedness between the Imdang individuals was estimated. Six mutually exclusive groups including 17 individuals were detected among 595 pairs combined from 35 Imdang individuals, suggesting that 17 Imdang individuals form six distinct pedigrees. A genetically identical pair, four first-degree, eleven second-degree, about eight third-degree, and many more distant relative pairs are found from pairwise mismatch rate (Figure 13; Table 10). Up to third-degree relative pairs were considered primarily, because more distant relative pairs could not provide sufficient resolution to infer accurate kinship. By incorporating age at death, estimated time periods of tombs the individuals were buried, uniparental haplogroup information, and IBD probabilities, most plausible pedigrees are suggested (Figure 14). An identical pair inferred by PMR (IMD002 and IMD003) is likely a technical duplicate. Among the four first-degree relative pairs, a father-daughter relationship, a family trio, and a full-sibling are determined. The first-degree relationship between JOY010 and JOY011 is likely a father-daughter given different mitochondrial haplogroups and the near-zero sharing of both alleles at the same site (Figure 14A). Otherwise, JOY016 and JOY018 are likely full siblings given relatively high probability of sharing both alleles at

the same site and sharing of the same mitochondrial haplogroup (Figure 14E; Table 10). JOY003, JOY004, JOY005 likely form a family trio, because JOY004 and JOY005 are distantly related while JOY003 is first-degree related to the others (Figure 14B). The exact relationship of a second-degree relative pair (IMD016 and JOY009) and third-degree relative pairs could not be decided due to missing genetic data of individuals who link the pair of individuals. The second-degree relationship between IMD016 and JOY009 is likely half-sibling or uncle/aunt-nephew/niece, because they died too young to leave their offspring (Figure 14C). JOY019 is likely a first cousin of JOY016 and JOY018 given their genetic relatedness, age at death and the time period of the tomb they were buried (Figure 14E). JOY002 and JOY007 are likely first cousins, given that they are third-degree relatives and they were buried in the same tomb at similar age (Figure 14F). Four individuals (IMD003(=IMD002), IMD018, JOY001, JOY012) are second-degree relatives to each other. IMD009 is a third-degree relative of the three individuals (IMD003(=IMD002), IMD018, JOY001) and a second-degree relative of JOY012, suggesting these five individuals form a distinct pedigree including two masters (Figure 14D).

ROH segments within the genome of each Three Kingdoms period individuals are measured using hapROH. The results show long ROH segments from some Imdang individuals, suggesting a close genetic relationship between their parents (Figure 15).

JOY010 has a similar summed length and length distribution of ROH expected from offspring between first cousins. IMD003(=IMD002) shows similar ROH aspects expected from offspring between uncle/aunt−nephew/niece, suggesting that the mother of IMD003(=IMD002) is likely an aunt or niece of her father.

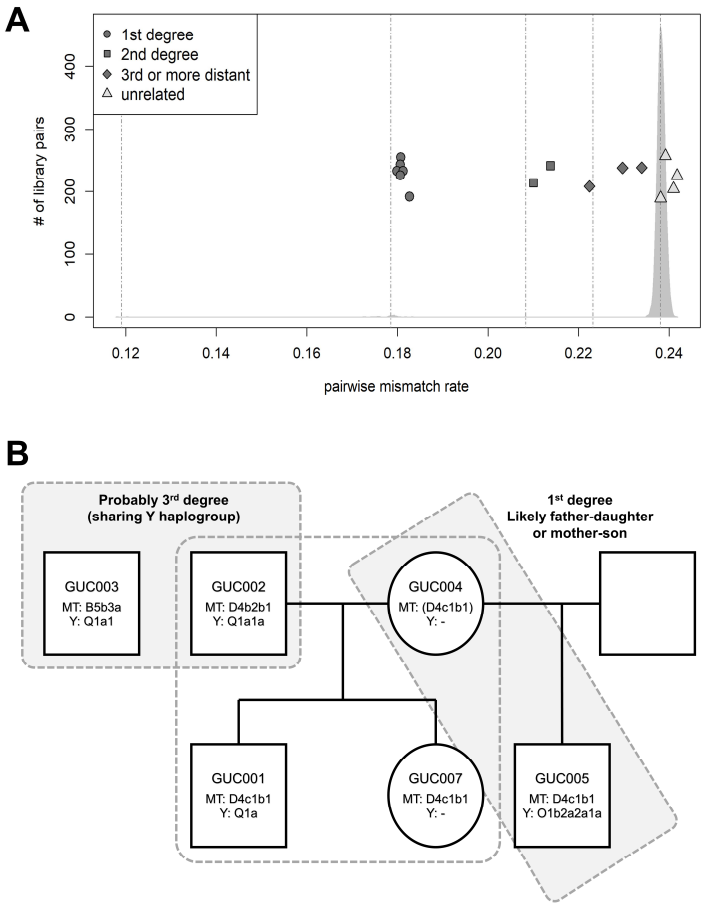


Figure 12. Genetic relationship between the six individuals from the Gunsan jar coffin. (A) Genetic relatedness of 15 pairs combined from six Gunsan individuals estimated with the PMR of genotypes is shown (color-filled shapes). On the background, density of the PMR values between 102 present-day Koreans from Ulsan are plotted. Dotted vertical lines show the expected PMR values of the identical, 1st degree, 2nd degree, 3rd degree relatives and the unrelated pairs of present-day Ulsan Koreans, from left to right, respectively. (B) Most plausible pedigree of the six individuals from the Gunsan jar coffin. MT and Y represent the mitochondrial haplogroup and Y haplogroup, respectively.

Table 9. Genetic relatedness between the six individuals from the Gunsan jar coffin. Six first-degree (five parent-offspring and full sibling), two second-degree, and three third-degree or more distant relative pairs are observed from 15 pairs combined from six individuals. Genetic relatedness is estimated based on the pairwise mismatch rate. Parent-offspring and full sibling are distinguished from first-degree relatives based on the probability of sharing both alleles estimated by the lcMLkin program. k0, k1 and k2 indicate probability of sharing 0, 1, and 2 alleles, respectively.

		Pairwise Mismatch Rate			lcMLkin			
ID1	ID2	Covered	Mismatch	Pr(mismatch)	k0	k1	k2	Kinship
GUC001	GUC002	135,698	24,525	0.1807	0.293	0.640	0.067	Son-Father
GUC001	GUC003	66,442	15,546	0.2340	0.992	0.008	0.000	> 3rd degree
GUC001	GUC004	77,168	14,096	0.1827	0.324	0.620	0.055	Son-Mother
GUC001	GUC005	154,884	33,116	0.2138	0.736	0.262	0.002	2nd degree
GUC001	GUC007	135,518	24,482	0.1807	0.453	0.314	0.233	Full sibling
GUC002	GUC003	196,448	43,698	0.2224	0.871	0.124	0.005	3rd degree
GUC002	GUC004	230,345	54,847	0.2381	0.996	0.004	0.000	Unrelated
GUC002	GUC005	487,550	116,613	0.2392	0.996	0.004	0.000	Unrelated
GUC002	GUC007	420,238	75,589	0.1799	0.172	0.812	0.016	Father-daughter
GUC003	GUC004	110,450	26,615	0.2410	0.995	0.005	0.000	Unrelated
GUC003	GUC005	225,119	54,425	0.2418	0.995	0.004	0.000	Unrelated
GUC003	GUC007	196,314	45,100	0.2297	0.978	0.022	0.000	> 3rd degree
GUC004	GUC005	265,920	48,192	0.1812	0.217	0.758	0.024	1st degree; likely mother-son or daughter-father
GUC004	GUC007	231,027	41,733	0.1806	0.254	0.722	0.024	Mother-daughter
GUC005	GUC007	489,161	102,749	0.2101	0.670	0.318	0.013	2nd degree

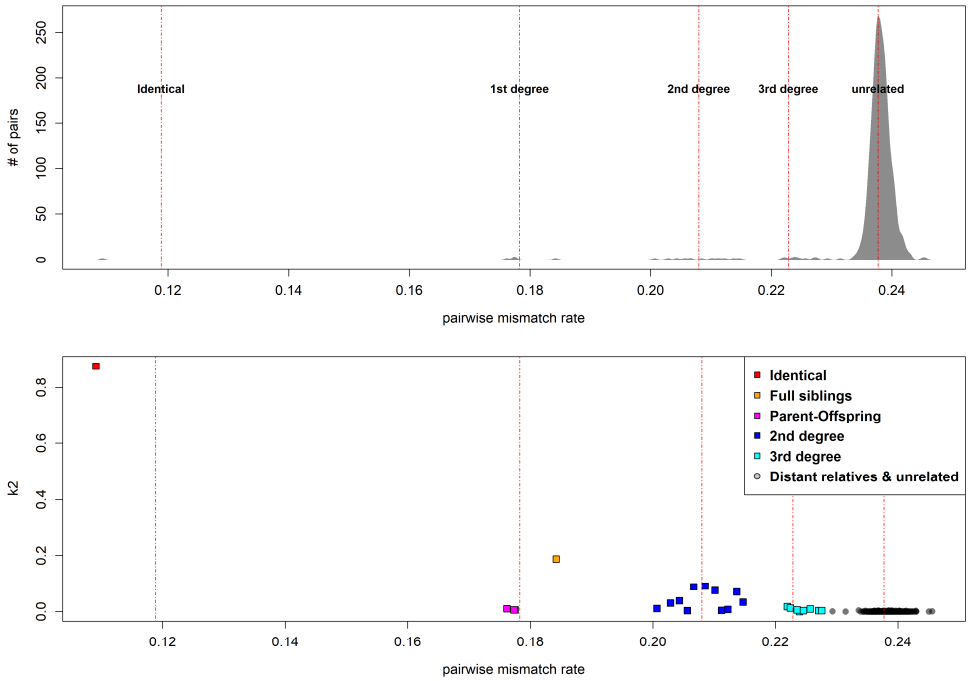


Figure 13. Estimated genetic relatedness of 35 Imdang individuals. Genetic relatedness of 595 pairs combined from 35 Imdang individuals estimated with the pairwise mismatch rate of genotypes is shown. Upper plot shows the density of the pairwise mismatch rate values between 35 Imdang individuals. Dotted vertical lines show the expected pairwise mismatch rate values of identical, 1st degree, 2nd degree, 3rd degree relatives and unrelated pairs, from left to right, respectively. Bottom plot shows the genetic relationship inferred from the pairwise mismatch rate of genotypes and estimated probability of sharing two alleles between 35 Imdang individuals.

Table 10. Genetic relatedness between the 35 individuals from the Imdang site. Genetic relatedness of relative pairs of Imdang individuals with bottom 30 pairwise mismatch rate values are shown. Genetic relatedness is estimated based on the pairwise mismatch rate. Parent-offspring and full sibling are distinguished from first-degree relatives based on the probability of sharing both alleles estimated by the lcMLkin program. k0, k1, and k2 indicate probability of sharing 0, 1, and 2 alleles, respectively.

		Pairwise Mismatch Rate			lcMLkin			
ID1	ID2	nSNPs	Mismatch	Pr(mismatch)	k0	k1	k2	Kinship
IMD002	IMD003	677,107	73,895	0.1091	0.014	0.110	0.876	Identical
JOY010	JOY011	521,135	91,814	0.1762	0.103	0.886	0.010	Father-daughter
JOY003	JOY005	710,503	125,991	0.1773	0.092	0.902	0.006	Mother-daughter
JOY003	JOY004	665,080	118,073	0.1775	0.083	0.911	0.006	Father-daughter
JOY016	JOY018	636,997	117,350	0.1842	0.412	0.400	0.188	Full-sibling
IMD002	JOY001	522,286	104,807	0.2007	0.608	0.381	0.011	Likely half-sibling
JOY001	IMD003	464,498	94,229	0.2029	0.659	0.311	0.030	Likely half-sibling
IMD002	JOY012	298,673	61,032	0.2043	0.739	0.222	0.038	2nd-degree
JOY012	IMD003	267,663	55,040	0.2056	0.728	0.269	0.003	2nd-degree
JOY001	JOY012	212,619	43,938	0.2067	0.786	0.125	0.089	2nd-degree
JOY001	IMD018	157,398	32,824	0.2085	0.824	0.085	0.091	Likely half-sibling
IMD009	JOY012	252,940	53,148	0.2101	0.815	0.110	0.075	2nd-degree
IMD016	JOY009	218,899	46,232	0.2112	0.683	0.313	0.004	2nd-degree
IMD002	IMD018	221,087	46,918	0.2122	0.830	0.164	0.007	Likely half-sibling
JOY012	IMD018	91,741	19,605	0.2137	0.913	0.016	0.070	2nd-degree
IMD003	IMD018	199,148	42,764	0.2147	0.898	0.068	0.033	Likely half-sibling
IMD002	IMD009	641,097	142,279	0.2219	0.949	0.034	0.017	3rd-degree
IMD009	JOY001	439,953	97,863	0.2224	0.925	0.063	0.012	3rd-degree
JOY002	JOY007	801,261	179,119	0.2236	0.868	0.126	0.006	3rd-degree
IMD009	IMD003	563,132	126,103	0.2239	0.941	0.058	0.000	3rd-degree
JOY016	JOY019	279,185	62,699	0.2246	0.922	0.075	0.003	3rd-degree
IMD009	IMD018	187,566	42,329	0.2257	0.956	0.035	0.009	3rd-degree
JOY002	JOY006	571,354	129,722	0.2270	0.917	0.080	0.003	3rd-degree
JOY018	JOY019	329,346	74,939	0.2275	0.943	0.054	0.003	3rd-degree
JOY015	JOY016	627,470	143,887	0.2293	0.965	0.034	0.001	≥4th-degree
JOY015	JOY019	324,922	75,199	0.2314	0.977	0.023	0.000	≥4th-degree
IMD008	JOY005	84,007	19,622	0.2336	0.990	0.006	0.004	≥4th-degree
JOY015	JOY018	776,514	181,649	0.2339	0.987	0.012	0.000	≥4th-degree
IMD002	JOY005	708,776	166,067	0.2343	0.997	0.003	0.000	≥4th-degree
JOY016	JOY020	646,075	151,468	0.2344	0.994	0.006	0.000	≥4th-degree

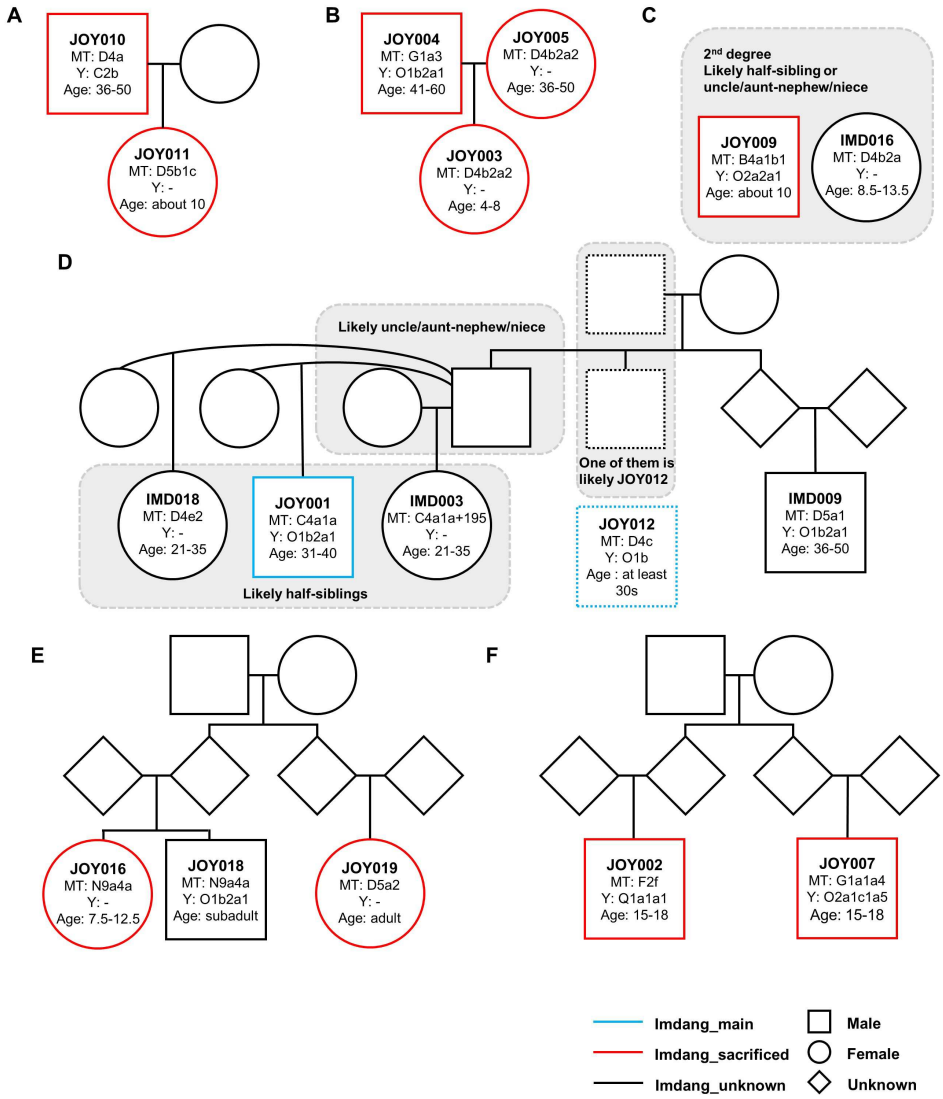


Figure 14. Reconstructed pedigrees of individuals from the Imdang site. Pedigrees of Imdang individuals are constructed by using relatedness inferred from PMR, age at death, time periods of tombs the individuals were buried, mitochondrial and Y haplogroup information, and IBD probabilities. Edge color of figures represents social status. Age, MT, and Y represent the age at death, mitochondrial haplogroup, and Y haplogroup, respectively.

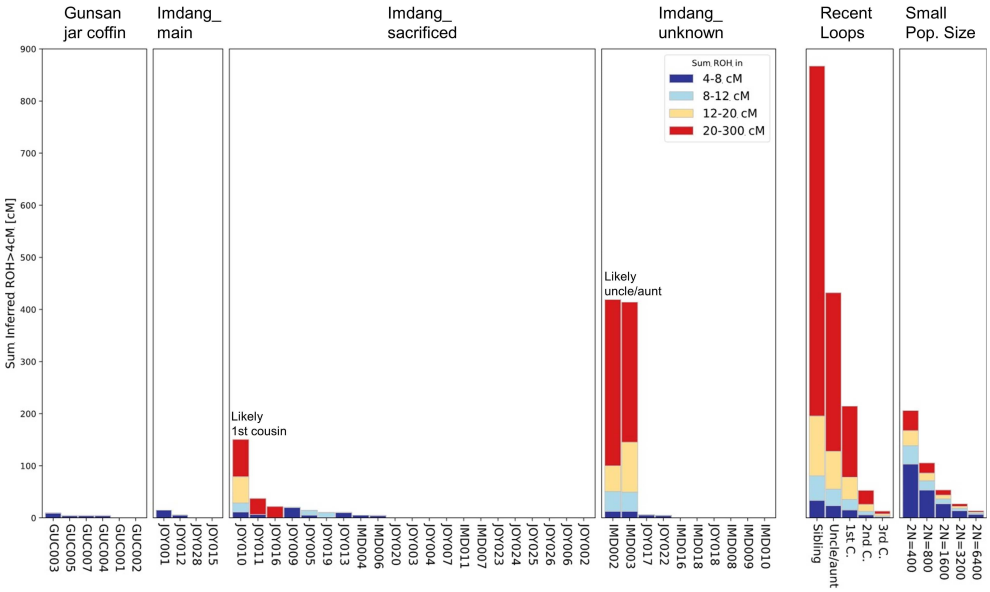


Figure 15. Inferred ROH from the Three Kingdoms period individuals. Sum of inferred ROH over 4cM from each Three Kingdoms period individuals are shown. Right boxes show expected sum of ROH over 4cM for offspring of close relatives (left) and in small population sizes (right).

4. Discussion

This study provides a greater insight into the genetic profiles and kinship structure of the Three Kingdoms period Koreans by analyzing genome-wide data of 41 individuals from two Three Kingdoms period sites. Three major hypotheses are tested by analyzing the Three Kingdoms period Korean genomes from both sites. First, as hypothesized by the archaeological records, rich kinship information is obtained from the Three Kingdoms period individuals from the jar coffin in the Dangbuk-ri site and the Ancient Tombs in Imdang-dong and Joyeong-dong. Second, genetic continuity between the Three Kingdoms period and present-day Koreans is confirmed. Third, no genetic differences are detected from the Three Kingdoms period individuals according to region and social status.

Genomic analysis of the Three Kingdoms period individuals from the Gunsan jar coffin shows a link between the unusual occasion of a secondary multiple burial and genetic relationship between the individuals. An extended family, consisting of a quartet family and both paternal and maternal relatives is found from the Gunsan individuals, showing little gender bias in the family structure. Since the Gunsan jar coffin is likely related to commoners rather than elites given the size and type of the jar coffin, the pedigree provides an example of family structure of Baekje commoners in Three Kingdoms period.

Ancient Tombs in Imdang-dong and Joyeong-dong are large-

scale cemetery of Abdokguk with sacrificial burial. These tombs include individuals assigned to various social statuses and ages. Family relationships between the individuals from the same social status are found. These intrastatus family relationships suggest the hereditary basis of social status. In the pedigree including two masters (JOY001, JOY012), the relationship between three individuals (JOY001, IMD003, IMD008) are likely half-siblings sharing the same father. This pedigree implies a probability of polygyny in the ruling class of the Abdokguk. In addition, family relationship between the individuals from Imdang-dong and Joyeong-dong confirms that the two sites are not assigned to distinct sociopolitical groups.

Close parental relatedness is detected in the two Imdang individuals. Consistent with prevalent relative marriage in the royal family of Silla (Ha, 1995), an offspring of close relatives about uncle/nephew or aunt/nephew is observed from the pedigree including masters. Close parental relatedness about first cousin is found in a sacrificed individual, suggesting that the relative marriage was practiced in both statuses.

There are some ambiguous relationships and missing individuals in the Imdang pedigrees. Further analysis on additional Imdang individuals will provide more precise relationship between the individuals, allowing reconstruction of higher-resolution pedigrees. Specifically, a pedigree including masters will be accurately inferred through sampling other close relatives.

This study shows the genetic similarity between the Three Kingdoms period Koreans from two sites and present-day Koreans, confirming the hypothesis that present-day Koreans are direct descendants of the Three Kingdoms period Koreans. Since the Gunsan jar coffin is likely related to commoners rather than elites, Gunsan individuals likely provide the genetic profile of the general population in the Three Kingdoms period lived in Gunsan. Although the Imdang individuals are assigned to distinct social statuses, there are no genetic differences between the different social statuses of Imdang individuals. Also, no genetic difference is found between the individuals from Gunsan and Imdang sites, which are about 180km apart. In consistent with the previous study (Gelabert et al., 2022), the Three Kingdoms period individuals and present-day Koreans share similar genetic profile. This genetic similarity support that the genetic profile of present-day Koreans had formed and been widespread already in the Three Kingdoms period. In addition, this result confirms that the previously reported Jomon-related genetic profile in Gimhae is likely caused by local human migration between Korea and Japan (Gelabert et al., 2022).

Although this study confirms genetic continuity between the Three Kingdoms period and present-day Koreans, it cannot be determined when the genetic profile of present-day Koreans is formed. Since the Bronze Age is an important period for elucidating the origins and formation of Koreans (Lee, 2002; Yu, 2010), Bronze Age genomes in the Korean peninsula are required to solve this

problem. The Korean peninsula harbors about 30,000 dolmens that are considered as representative Bronze Age tombs (Peterson and Margulies, 2010; Yun, 2017), but there are few Bronze Age human remains due to the unfavorable environment for preservation of human remains in Korea (Jee et al., 2008; Pang, 2016). A previous genomic study analyzed a Korean Bronze Age sample from below shell midden which locates on the west coast of South Korea, but it cannot be determined whether the genetic profile of this individual from coastal region can represent the genetic profile of general Bronze Age population lived in the Korean peninsula (Robbeets et al., 2021). In addition, low coverage of this Bronze Age genome did not allow high-resolution analysis (Robbeets et al., 2021). High-quality genomes from the Bronze Age individuals in the Korean peninsula will provide an important insight into the formation of present-day Korean genetic profile.

References

Allentoft, M.E., Sikora, M., Sjogren, K.G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P.B., Schroeder, H., Ahlstrom, T., Vinner, L.*, et al.* (2015). Population genomics of bronze age Eurasia. Nature *522*, 167–172.

Alt, K.W., and Vach, W. (1998). Kinship studies in skeletal remains—Concepts and examples. In Dental Anthropology (Wien: Springer), pp. 537–554.

Bacvarov, K. (2006). Early Neolithic jar burials in southeast Europe: a comparative approach. Documenta Praehistorica *33*, 101–106.

Boeyens, J., Van der Ryst, M., Coetzee, F., Steyn, M., and Loots, M. (2009). From uterus to jar: the significance of an infant pot burial from Melora Saddle, an early nineteenth-century African farmer site on the Waterberg Plateau. Southern African Humanities *21*, 213–238.

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.

Dabney, J., Knapp, M., Glocke, I., Gansauge, M.T., Weihmann, A., Nickel, B., Valdiosera, C., Garcia, N., Paabo, S., Arsuaga, J.L.*, et al.* (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. Proceedings of the National Academy of Sciences *110*, 15758–15763.

Damgaard, P.B., Marchi, N., Rasmussen, S., Peyrot, M., Renaud, G., Korneliussen, T., Moreno-Mayar, J.V., Pedersen, M.W., Goldberg, A., Usmanova, E.*, et al.* (2018a). 137 ancient human genomes from across the Eurasian steppes. Nature *557*, 369–374.

Damgaard, P.B., Martiniano, R., Kamm, J., Moreno-Mayar, J.V., Kroonen, G., Peyrot, M., Barjamovic, G., Rasmussen, S., Zacho, C., and Baimukhanov, N. (2018b). The first horse herders and the impact of early Bronze Age steppe expansions into Asia. Science *360*, eaar7711.

Derenko, M., Malyarchuk, B., Grzybowski, T., Denisova, G., Dambueva, I., Perkova, M., Dorzhu, C., Luzina, F., Lee, H.K., and Vanecek, T. (2007). Phylogeographic analysis of mitochondrial DNA in northern Asian populations. The American Journal of Human Genetics *81*, 1025–1041.

Fu, Q., Li, H., Moorjani, P., Jay, F., Slepchenko, S.M., Bondarev, A.A., Johnson, P.L., Aximu-Petri, A., Prufer, K., de Filippo, C.*, et al.* (2014). Genome sequence of a 45,000-year-old modern human from western Siberia. Nature *514*, 445–449.

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

Gelabert, P., Blazyte, A., Chang, Y., Fernandes, D.M., Jeon, S., Hong, J.G., Yoon, J., Ko, Y., Oberreiter, V., and Cheronet, O. (2022). Northeastern Asian and Jomon-related genetic structure in the Three Kingdoms period of Gimhae, Korea. Current Biology.

Ha, D. (2011). A study on the status of the victims of the Silla sacrificial burials at Imdang. Hanguk Kogo-Hakbo *79*, 175–204. Korean.

Ha, J.-Y. (1995). A study of the succession to the throne in the period of the Silla Sangdae – Centering around the consanguineous marriage in the royal family –. The Journal of the Research Institute for Silla Culture *12*, 173–199. Korean.

Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., and Stewardson, K. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. Nature *522*, 207–211.

Haber, M., Doumet-Serhal, C., Scheib, C., Xue, Y., Danecek, P., Mezzavilla, M., Youhanna, S., Martiniano, R., Prado-Martinez, J., Szpak, M.*, et al.* (2017). Continuity and admixture in the last five millennia of Levantine history from ancient Canaanite and present-day Lebanese genome sequences. The American Journal of Human Genetics *101*, 274–282.

Harney, É., May, H., Shalem, D., Rohland, N., Mallick, S., Lazaridis, I., Sarig, R., Stewardson, K., Nordenfelt, S., and Patterson, N. (2018). Ancient DNA from Chalcolithic Israel reveals the role of

population mixture in cultural transformation. Nature Communications *9*, 1–11.

Hawass, Z., Gad, Y.Z., Ismail, S., Khairat, R., Fathalla, D., Hasan, N., Ahmed, A., Elleithy, H., Ball, M., and Gaballah, F. (2010). Ancestry and pathology in King Tutankhamun's family. Jama *303*, 638–647.

Jang, Y.-s. (2007). Establishment and change of the ancient polity of Gyeongsan area researched through the Imdang site. The Journal of Korean Field Archaeology, 44–87. Korean.

Jee, S.H., Chung, Y.J., and Seo, M.S. (2008). Archaeogenetic research of excavated human bones from the ancient tombs. Mun Hwa Jae *41*, 99–108. Korean.

Jeong, C., Balanovsky, O., Lukianova, E., Kahbatkyzy, N., Flegontov, P., Zaporozhchenko, V., Immel, A., Wang, C.-C., Ixan, O., and Khussainova, E. (2019). The genetic history of admixture across inner Eurasia. Nature Ecology & Evolution *3*, 966–976.

Jeong, C., Ozga, A.T., Witonsky, D.B., Malmstrom, H., Edlund, H., Hofman, C.A., Hagan, R.W., Jakobsson, M., Lewis, C.M., Aldenderfer, M.S.*, et al.* (2016). Long-term genetic stability and a high-altitude East Asian origin for the peoples of the high valleys of the Himalayan arc. Proceedings of the National Academy of Sciences *113*, 7485–7490.

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.

Jeong, C., Wilkin, S., Amgalantugs, T., Bouwman, A.S., Taylor, W.T.T., Hagan, R.W., Bromage, S., Tsolmon, S., Trachsel, C., Grossmann, J.*, et al.* (2018). Bronze Age population dynamics and the rise of dairy pastoralism on the eastern Eurasian steppe. Proceedings of the National Academy of Sciences *115*, E11248–E11255.

Jones, E.R., Gonzalez-Fortes, G., Connell, S., Siska, V., Eriksson, A., Martiniano, R., McLaughlin, R.L., Gallego Llorente, M., Cassidy, L.M., Gamba, C.*, et al.* (2015). Upper Palaeolithic genomes reveal deep roots of modern Eurasians. Nature Communications *6*, 1–8.

Jonsson, H., Ginolhac, A., Schubert, M., Johnson, P.L., and Orlando,

L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics *29*, 1682-1684.

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 Research *25*, 918-925.

Kanzawa-Kiriyama, H., Jinam, T.A., Kawai, Y., Sato, T., Hosomichi, K., Tajima, A., Adachi, N., Matsumura, H., Kryukov, K., and Saitou, N. (2019). Late Jomon male and female genome sequences from the Funadomari site in Hokkaido, Japan. Anthropological Science *127*, 83-108.

Kennett, D.J., Plog, S., George, R.J., Culleton, B.J., Watson, A.S., Skoglund, P., Rohland, N., Mallick, S., Stewardson, K., and Kistler, L. (2017). Archaeogenomic evidence reveals prehistoric matrilineal dynasty. Nature communications *8*, 1-9.

Keyser-Tracqui, C., Crubezy, E., and Ludes, B. (2003). Nuclear and mitochondrial DNA analysis of a 2,000-year-old necropolis in the Egyin Gol Valley of Mongolia. The American Journal of Human Genetics *73*, 247-260.

Kim, A.-J., Kim, K.-J., Choi, J.-H., Choi, E.-H., Jung, Y.-J., Min, N.-Y., Lkhagvasuren, G., Rhee, S.-M., Kim, J.-H., and Noh, M.-S. (2010). Mitochondrial DNA analysis of ancient human bones excavated from Nukdo island, S. Korea. BMB reports *43*, 133-139.

Kim, D.-W. (2014). A study on the funeral system of ancient tombs at the Im-dang site (Gyeongsan: Yeungnam University). Korean.

Kim, J., Weber, J.A., Jho, S., Jang, J., Jun, J., Cho, Y.S., Kim, H.M., Kim, H., Kim, Y., Chung, O.*, et al.* (2018). KoVariome: Korean National Standard Reference Variome database of whole genomes with comprehensive SNV, indel, CNV, and SV analyses. Scientific Reports *8*, 5677.

Kim, W. (1973). Opening of Korean archaeology (Seoul: Ilchokak). Korean.

Kim, Y.-S. (2002). The burial of the living with the dead in Silla tomb. Komunhwa *55*, 64-87. Korean.

Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014).

ANGSD: analysis of next generation sequencing data. BMC Bioinformatics *15*, 1–13.

Krzewińska, M., Kılınç, G.M., Juras, A., Koptekin, D., Chyleński, M., Nikitin, A.G., Shcherbakov, N., Shuteleva, I., Leonova, T., and Kraeva, L. (2018). Ancient genomes suggest the eastern Pontic–Caspian steppe as the source of western Iron Age nomads. Science Advances *4*, eaat4457.

Lazaridis, I., Mittnik, A., Patterson, N., Mallick, S., Rohland, N., Pfrengle, S., Furtwangler, A., Peltzer, A., Posth, C., Vasilakis, A.*, et al.* (2017). Genetic origins of the Minoans and Mycenaeans. Nature *548*, 214–218.

Lazaridis, I., Nadel, D., Rollefson, G., Merrett, D.C., Rohland, N., Mallick, S., Fernandes, D., Novak, M., Gamarra, B., and Sirak, K. (2016). Genomic insights into the origin of farming in the ancient Near East. Nature *536*, 419–424.

Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., Sudmant, P.H., Schraiber, J.G., Castellano, S., Lipson, M.*, et al.* (2014). Ancient human genomes suggest three ancestral populations for present–day Europeans. Nature *513*, 409–413.

Lee, D.–N., Jeon, C.L., Kang, J., Burri, M., Krause, J., Woo, E.J., and Jeong, C. (2022). Genomic detection of a secondary family burial in a single jar coffin in early Medieval Korea. bioRxiv. doi: https://doi.org/10.1101/2022.05.09.491093

Lee, H.Y., Yoo, J.–E., Park, M.J., Kim, C.–Y., and Shin, K.–J. (2008). Genetic characterization and assessment of authenticity of ancient Korean skeletal remains. Human biology *80*, 239–250.

Lee, H.Y., Yoo, J.E., Park, M.J., Chung, U., Kim, C.Y., and 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*, 4408–4418.

Lee, K.–b. (2002). A new history of Korea (Seoul: Ilchokak). Korean.

Lee, K.–M., and Ramsey, S.R. (2011). A history of the Korean language (New York: Cambridge University Press).

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754-1760.

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics *26*, 589-595.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078-2079.

Lipatov, M., Sanjeev, K., Patro, R., and Veeramah, K.R. (2015). Maximum likelihood estimation of biological relatedness from low coverage sequencing data. bioRxiv. doi:https://doi.org/10.1101/023374

Lipson, M., Cheronet, O., Mallick, S., Rohland, N., Oxenham, M., Pietrusewsky, M., Pryce, T.O., Willis, A., Matsumura, H., Buckley, H.*, et al.* (2018). Ancient genomes document multiple waves of migration in Southeast Asian prehistory. Science *361*, 92-95.

Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., Zhao, M., Chennagiri, N., Nordenfelt, S., Tandon, A.*, et al.* (2016). The Simons genome diversity project: 300 genomes from 142 diverse populations. Nature *538*, 201-206.

Mathieson, I., Alpaslan-Roodenberg, S., Posth, C., Szecsenyi-Nagy, A., Rohland, N., Mallick, S., Olalde, I., Broomandkhoshbacht, N., Candilio, F., Cheronet, O.*, et al.* (2018). The genomic history of southeastern Europe. Nature *555*, 197-203.

Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S.A., Harney, E., Stewardson, K., Fernandes, D., Novak, M.*, et al.* (2015). Genome-wide patterns of selection in 230 ancient Eurasians. Nature *528*, 499-503.

McColl, H., Racimo, F., Vinner, L., Demeter, F., Gakuhari, T., Moreno-Mayar, J.V., van Driem, G., Gram Wilken, U., Seguin-Orlando, A., de la Fuente Castro, C.*, et al.* (2018). The prehistoric peopling of Southeast Asia. Science *361*, 88-92.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M.*, et*

al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res *20*, 1297-1303.

Mittnik, A., Massy, K., Knipper, C., Wittenborn, F., Friedrich, R., Pfrengle, S., Burri, M., Carlichi-Witjes, N., Deeg, H., and Furtwängler, A. (2019). Kinship-based social inequality in Bronze Age Europe. Science *366*, 731-734.

Mooder, K.P., Weber, A.W., Bamforth, F.J., Lieverse, A.R., Schurr, T.G., Bazaliiski, V.I., and Savel'ev, N.A. (2005). Matrilineal affinities and prehistoric Siberian mortuary practices: a case study from Neolithic Lake Baikal. Journal of Archaeological Science *32*, 619-634.

Moreno-Mayar, J.V., Potter, B.A., Vinner, L., Steinrucken, M., Rasmussen, S., Terhorst, J., Kamm, J.A., Albrechtsen, A., Malaspinas, A.S., Sikora, M.*, et al.* (2018). Terminal Pleistocene Alaskan genome reveals first founding population of Native Americans. Nature *553*, 203-207.

Naju national research institute of cultural heritage (2020). The rebirth of the big jar-coffin : Records on the production of the big jar-coffin (Jeollanam-do: Naju national research institute of cultural heritage).

Narasimhan, V.M., Patterson, N., Moorjani, P., Rohland, N., Bernardos, R., Mallick, S., Lazaridis, I., Nakatsuka, N., Olalde, I., Lipson, M.*, et al.* (2019). The formation of human populations in South and Central Asia. Science *365*, eaat7487.

National Research Institute of Cultural Heritage (2009). Specialised dictionary of Korean archaeology: burials (Daejeon: National Research Institute of Cultural Heritage). Korean.

Ning, C., Li, T., Wang, K., Zhang, F., Li, T., Wu, X., Gao, S., Zhang, Q., Zhang, H., and Hudson, M.J.J.N.C. (2020). Ancient genomes from northern China suggest links between subsistence changes and human migration. Nature Communications *11*, 1-9.

Pang, M.K. (2016). Physical anthropological characteristics of the people in the Korean peninsula of the Bronze Age. The Journal of Humanities and Social science *7*, 641-657. Korean.

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.

Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. PLOS Genetics *2*, e190.

Peltzer, A., Jager, G., Herbig, A., Seitz, A., Kniep, C., Krause, J., and Nieselt, K. (2016). EAGER: efficient ancient genome reconstruction. Genome Biology *17*, 1–14.

Peterson, M., and Margulies, P. (2010). A brief history of Korea (New York: Facts On File).

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

Raghavan, M., DeGiorgio, M., Albrechtsen, A., Moltke, I., Skoglund, P., Korneliussen, T.S., Gronnow, B., Appelt, M., Gullov, H.C., Friesen, T.M.*, et al.* (2014a). The genetic prehistory of the New World Arctic. Science *345*, 1255832.

Raghavan, M., Skoglund, P., Graf, K.E., Metspalu, M., Albrechtsen, A., Moltke, I., Rasmussen, S., Stafford, T.W., Jr., Orlando, L., Metspalu, E.*, et al.* (2014b). Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. Nature *505*, 87–91.

Rasmussen, M., Anzick, S.L., Waters, M.R., Skoglund, P., DeGiorgio, M., Stafford, T.W., Jr., Rasmussen, S., Moltke, I., Albrechtsen, A., Doyle, S.M.*, et al.* (2014). The genome of a Late Pleistocene human from a Clovis burial site in western Montana. Nature *506*, 225–229.

Rasmussen, M., Li, Y., Lindgreen, S., Pedersen, J.S., Albrechtsen, A., Moltke, I., Metspalu, M., Metspalu, E., Kivisild, T., Gupta, R.*, et al.* (2010). Ancient human genome sequence of an extinct Palaeo-Eskimo. Nature *463*, 757–762.

Rasmussen, M., Sikora, M., Albrechtsen, A., Korneliussen, T.S., Moreno-Mayar, J.V., Poznik, G.D., Zollikofer, C.P.E., de Leon, M.P., Allentoft, M.E., Moltke, I.*, et al.* (2015). The ancestry and affiliations of Kennewick Man. Nature *523*, 455–458.

Reich, D., Patterson, N., Campbell, D., Tandon, A., Mazieres, S., Ray, N., Parra, M.V., Rojas, W., Duque, C., Mesa, N.*, et al.* (2012). Reconstructing Native American population history. Nature *488*,

370–374.

Renaud, G., Slon, V., Duggan, A.T., and Kelso, J. (2015). Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. Genome Biology *16*, 1–18.

Ringbauer, H., Novembre, J., and Steinrücken, M. (2021). Parental relatedness through time revealed by runs of homozygosity in ancient DNA. Nature communications *12*, 1–11.

Robbeets, M., Bouckaert, R., Conte, M., Savelyev, A., Li, T., An, D.I., Shinoda, K.I., Cui, Y., Kawashima, T., Kim, G.*, et al.* (2021). Triangulation supports agricultural spread of the Transeurasian languages. Nature *599*, 616–621.

Rudbeck, L., Gilbert, M.T.P., Willerslev, E., Hansen, A.J., Lynnerup, N., Christensen, T., and Dissing, J. (2005). mtDNA analysis of human remains from an early Danish Christian cemetery. American Journal of Physical Anthropology *128*, 424–429.

Schubert, M., Lindgreen, S., and Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC research notes *9*, 1–7.

Shewan, L., Armstrong, R., O' Reilly, D., Halcrow, S., Beavan, N., and Sokha, T. (2020). Isotopic insights into the jar–and–coffin mortuary ritual of the Cardamom Mountains, Cambodia. Antiquity *94*, 1575–1591.

Sikora, M., Pitulko, V.V., Sousa, V.C., Allentoft, M.E., Vinner, L., Rasmussen, S., Margaryan, A., de Barros Damgaard, P., de la Fuente, C., and Renaud, G. (2019). The population history of northeastern Siberia since the Pleistocene. Nature *570*, 182–188.

Unterlander, M., Palstra, F., Lazaridis, I., Pilipenko, A., Hofmanova, Z., Gross, M., Sell, C., Blocher, J., Kirsanow, K., Rohland, N.*, et al.* (2017). Ancestry and demography and descendants of Iron Age nomads of the Eurasian Steppe. Nature Communications *8*, 1–10.

Wang, C.–C., Yeh, H.–Y., Popov, A.N., Zhang, H.–Q., Matsumura, H., Sirak, K., Cheronet, O., Kovalev, A., Rohland, N., and Kim, A.M. (2020). The genomic formation of human populations in East Asia. bioRxiv. doi:https://doi.org/10.1101/2020.03.25.004606

Wang, T., Wang, W., Xie, G., Li, Z., Fan, X., Yang, Q., Wu, X., Cao, P.,

Liu, Y., and Yang, R. (2021). Human population history at the crossroads of East and Southeast Asia since 11,000 years ago. Cell *184*, 3829–3841.

Weissensteiner, H., Pacher, D., Kloss-Brandstatter, A., Forer, L., Specht, G., Bandelt, H.J., Kronenberg, F., Salas, A., and Schonherr, S. (2016). HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. Nucleic Acids Research *44*, W58–63.

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.

Yang, M.A., Gao, X., Theunert, C., Tong, H., Aximu-Petri, A., Nickel, B., Slatkin, M., Meyer, M., Pääbo, S., and Kelso, J. (2017). 40,000-year-old individual from Asia provides insight into early population structure in Eurasia. Current Biology *27*, 3202–3208.

Yu, H., Spyrou, M.A., Karapetian, M., Shnaider, S., Radzeviciute, R., Nagele, K., Neumann, G.U., Penske, S., Zech, J., Lucas, M.*, et al.* (2020). Paleolithic to Bronze Age Siberians reveal connections with first Americans and across Eurasia. Cell *181*, 1232–1245.

Yu, T.-Y. (2010). A study of the Korean identity in the archaeological viewpoint. Journal of Ethnology *9*, 83–118. Korean.

Yun, H.-P. (2017). A study on the background behind the construction and symbolism of Bronze Age dolmens. Hanguk Cheongdonggi Hakbo, 80–106. Korean.

국문 초록

삼국시대 한국인의 유전적 프로필과 혈연관계 분석

삼국시대는 한반도에 고대 국가가 최초로 나타난 시기로, 삼국시대의 한반도 거주민은 현대 한국인의 직접 조상으로 여겨진다. 그러나 삼국시대 고유전체 정보의 부족으로 인해 이에 대한 증명은 불가능하였다. 본 연구에서는 이러한 가설을 검증하고 유전적 프로필의 지역 차를 알아보고자 전라북도 군산시와 경상북도 경산시 두 지역의 삼국시대 고유전체 자료를 생산해 비교, 분석하였다. 군산 당북리 유적의 옹관묘는 하나의 옹관에 복수의 시신이 동시에 안치된 이례적인 장법을 보인다. 경산 임당동과 조영동 고분군은 순장 풍습이 발견된 압독국 지배계층의 무덤이므로 유전적 프로필과 더불어 과거 사회 구조를 연구하기 적합한 사례이다. 본 연구에서는 군산 옹관묘 6개체, 임당 고분군 35개체의 전장유전체 자료를 생산하였다. 집단유전학 방법론을 이용한 전장유전체 자료 분석을 통해 두 지역의 삼국시대 한국인의 유전적 프로필을 파악하고 지역별, 계층별 유전적 차이와 현대 한국인과의 유전적 차이가 존재하는지 검증하였다. 또한, 개체 사이의 혈연관계를 밝혀 가계도를 제작하였으며, 근친혼의 흔적을 조사하였다. 군산과 임당 두 지역 개체의 유전적 프로필과 현대 한국인의 유전적 프로필은 유사하며, 이는 지난 1,500여 년 간 한반도의 넓은 지역에 걸친 유전적 안정성을 시사한다. 서로 다른 사회적 계층에 속하는 임당 고분군 개체 사이의 유전적 프로필 또한 유사한 것으로 나타났다. 본 연구에서는 군산 옹관묘에서 보이는 특이적인 매장 형태가 개체 사이의 혈연관계를 반영함을 밝혔다. 군산 옹관묘의 6개체는 4인 핵가족과

부계 친척, 모계친척 각 1명으로 구성된 확대가족을 이루며, 이는 가족구조의 성별 편차가 거의 존재하지 않음을 보인다. 임당 고분군 개체는 6개의 구분된 가계에 속하며, 동일한 신분의 피장자끼리 혈연관계를 갖는 경향을 보인다. 일부 임당 고분군 개체에서 발견한 부모 사이의 혈연관계는 삼국시대 압독국에서 근친혼이 이루어졌음을 시사한다.

주요어 : 삼국시대, 고유전체, 집단유전학, 가족묘, 순장, 군산 당북리 유적, 경산 임당동과 조영동 고분군

학번 : 2020-25735

Appendix

Appendix 1. A list of present-day world-wide populations used in this study. This list includes 150 present-day Eurasian populations genotyped on the HumanOrigins panel and 121 present-day world-wide populations genotyped on 1240K panel. "Symbol" column and "N" column represent distinct three-letter code assigned and the total number of individuals for each population, respectively. Populations used for each analysis method are shown.

Group name	Subgroup	Symbol	N	Eurasian PCA	East Asian PCA	f_3/ f_4	qpWave/ qpAdm	HO dataset	1240K dataset
Abazin		Abz	10	O				O	
Abkhasian		Abk	9	O				O	
Adygei		Adg	31	O				O	
Albanian		Abn	6	O				O	
Aleut		Ale	7	O				O	
Aleut_Tlingit		Atl	4	O				O	
Altaian		Alt	24	O				O	
Altaian_Chelkans		Ack	8	O				O	
Ami		Ami	10	O	O			O	
Armenian		Arm	10	O				O	
Armenian_Hemsheni		Ahm	8	O				O	
Atayal		Aty	9	O	O			O	
Avar		Avr	10	O				O	
Azeri		Azr	17	O				O	
Balkar		Blk	10	O				O	
Bashkir		Bsc	53	O				O	
Basque		Bsq	29	O				O	
Belarusian		Blr	10	O				O	
Besermyan		Bes	6	O				O	
Borneo		Brn	9	O	O			O	
Bulgarian		Blg	10	O				O	
Buryat		Bry	37	O				O	
Cambodian		Cmb	8	O	O			O	
Chechen		Cch	9	O				O	
Chukchi		Chk	20	O				O	
Chuvash		Cvs	16	O				O	
Circassian		Ccs	9	O				O	
Croatian		Crt	10	O				O	
Cypriot		Cyp	8	O				O	
Czech		Cze	10	O				O	
Dai		Dai	10	O	O			O	

Appendix 1. Continued.

Group name	Subgroup	Symbol	N	Eurasian PCA	East Asian PCA	f_3/ f_4	qpWave/ qpAdm	HO dataset	1240K dataset
Darginian		Drg	8	O				O	
Daur		Dau	9	O	O			O	
Dolgan		Dlg	4	O				O	
Dungan		Dng	13	O	O			O	
Enets		Ent	3	O				O	
English		Eng	10	O				O	
Eskimo_ChaplinSireniki		Ecs	7	O				O	
Eskimo_Naukan		Enk	12	O				O	
Estonian		Est	10	O				O	
Even		Evn	9	O				O	
	Even	Evn	2		O			O	
Evenk_FarEast		Evf	5	O				O	
	Evenk_FarEast	Evf	2		O			O	
Evenk_Transbaikal		Evt	8	O	O			O	
Ezid		Ezd	8	O				O	
Finnish		Fin	8	O				O	
French		Fre	61	O				O	
Gagauz		Ggz	9	O				O	
Georgian		Grg	23	O				O	
German		Ger	10	O				O	
Greek		Grk	20	O				O	
Han		Han	43	O				O	
	Han	Han	33		O			O	
	HanNorth	Hnn	10		O			O	
Hezhen		Hzh	8	O	O			O	
Hungarian		Hun	20	O				O	
Icelandic		Ice	12	O				O	
Ingushian		Igs	10	O				O	
Irish		Iri	24	O				O	
Irish_Ulster		Iru	6	O				O	
Italian_North		Itn	20	O				O	
Italian_South		Its	5	O				O	
Itelmen		Ite	6	O				O	
Japanese		Jap	29	O	O			O	
Jew_Ashkenazi		Jas	7	O				O	
Jew_Georgian		Jgr	7	O				O	
Kabardinian		Kbd	10	O				O	
Kaitag		Ktg	8	O				O	
Kalmyk		Klm	10	O				O	
Karachai		Krc	12	O				O	
Karakalpak		Krk	14	O				O	
Karelian		Krl	15	O				O	
Kazakh		Kzk	18	O				O	
Ket		Ket	22	O				O	
Khakass		Khs	16	O				O	

Appendix 1. Continued.

Group name	Subgroup	Symbol	N	Eurasian PCA	East Asian PCA	f_3/ f_4	qpWave/ qpAdm	HO dataset	1240K dataset
Khakass_Kachins		Khk	8	O				O	
Khamnegan		Kmn	10	O				O	
Kinh		Kin	8	O	O			O	
Korean		Kor	6	O				O	
	Korean	Kor	5		O			O	
Koryak		Kry	9	O				O	
Kubachinian		Kbc	6	O				O	
Kumyk		Kmk	8	O				O	
Kurd		Krd	8	O				O	
Kyrgyz		Kyg	9	O				O	
Lahu		Lah	8	O	O			O	
Lak		Lak	10	O				O	
Lezgin		Lzg	9	O				O	
Lithuanian		Lth	10	O				O	
Maltese		Mlt	8	O				O	
Mansi		Mns	8	O				O	
Miao		Mia	10	O	O			O	
Moldavian		Mld	10	O				O	
Mongol		Mon	34	O				O	
Mongola		Mgl	6	O	O			O	
Mordovian		Mdv	32	O				O	
Nanai		Nan	10	O	O			O	
Naxi		Nax	9	O	O			O	
Negidal		Ngd	3	O	O			O	
Nganasan		Nga	33	O	O			O	
Nivh		Nvh	10	O	O			O	
Nogai		Ng1	25	O				O	
Norwegian		Nwg	11	O				O	
Orcadian		Orc	13	O				O	
Oroqen		Orq	9	O	O			O	
Ossetian		Ost	16	O				O	
Polish		Pol	23	O				O	
Romanian		Rom	10	O				O	
Russian		Rus	71	O				O	
Russian_Archangelsk_Krasnoborsky		Rak	6	O				O	
Russian_Archangelsk_Leshukonsky		Ral	5	O				O	
Russian_Archangelsk_Pinezhsky		Rap	5	O				O	
Saami		Saa	3	O				O	
Sardinian		Sar	27	O				O	
Scottish		Sct	33	O				O	
Selkup		Slk	25	O				O	
Semende		Smd	10	O	O			O	
She		She	10	O	O			O	
Sherpa		Shp	2	O	O			O	

Appendix 1. Continued.

Group name	Subgroup	Symbol	N	Eurasian PCA	East Asian PCA	f_3/ f_4	qpWave/ qpAdm	HO dataset	1240K dataset
Shetlandic		Stl	10	O				O	
Shor_Khakassia		Skh	6	O				O	
Shor_Mountain		Smn	7	O				O	
Sicilian		Scl	11	O				O	
Sorb		Srb	9	O				O	
Spanish		Spa	53	O				O	
Spanish_North		Spn	5	O				O	
Tabasaran		Tbs	10	O				O	
Tajik		Tjl	31	O				O	
Tatar_Kazan		Ttk	13	O				O	
Tatar_Mishar		Ttm	10	O				O	
Tatar_Siberian		Tts	19	O				O	
Tatar_Siberian_Zabolotniye		Ttz	5	O				O	
Thai		Tha	10	O	O			O	
Tibetan		Tib	2	O	O			O	
Todzin		Tdz	3	O				O	
Tofalar		Tfl	13	O				O	
Tu		Tuu	10	O	O			O	
Tubalar		Tbl	27	O				O	
Tujia		Tuj	10	O	O			O	
Turkish		Tra	50	O				O	
Turkish_Balikesir		Trb	5	O				O	
Turkmen		Tkm	7	O				O	
Tuvinian		Tvn	20	O				O	
Udmurt		Udm	10	O				O	
Ukrainian		Ukr	21	O				O	
Ulchi		Ulc	25	O	O			O	
Ulsan Korean		Kou	88	O	O	O	O	O	O
Uygur		Ugr	10	O				O	
Uzbek		Uzb	27	O				O	
Veps		Vep	10	O				O	
Xibo		Xib	7	O	O			O	
Yakut		Ykt	20	O				O	
Yi		Yii	10	O	O			O	
Yukagir		Ykg	19	O				O	
Abkhasian.DG			2			O			O
Adygei.DG			2			O			O
Albanian.DG			1			O			O
Aleut.DG			2			O			O
Altaian.DG			1			O			O
Ami.DG			2			O			O
Armenian.DG			2			O			O
Atayal.DG			1			O			O

Appendix 1. Continued.

Group name	Subgroup	Symbol	N	Eurasian PCA	East Asian PCA	f_3/ f_4	qpWave/ qpAdm	HO dataset	1240K dataset
Australian.DG			1			O			O
Balochi.DG			2			O			O
Basque.DG			2			O			O
BedouinB.DG			2			O			O
Bengali.DG			2			O			O
Bergamo.DG			2			O			O
Bougainville.DG			2			O			O
Brahmin.DG			2			O			O
Brahui.DG			2			O			O
Bulgarian.DG			2			O			O
Burmese.DG			2			O			O
Burusho.DG			2			O			O
Cambodian.DG			2			O			O
Chane.DG			1			O			O
Chechen.DG			1			O			O
Chipewyan.DG			2			O			O
Chukchi.DG			1			O			O
Cree.DG			2			O			O
Crete.DG			2			O			O
Czech.DG			1			O			O
Dai.DG			5			O			O
Daur.DG			1			O			O
Druze.DG			2			O			O
Dusun.DG			2			O			O
English.DG			2			O			O
Eskimo_Chaplin.DG			1			O			O
Eskimo_Naukan.DG			2			O			O
Eskimo_Sireniki.DG			2			O			O
Estonian.DG			2			O			O
Even.DG			3			O			O
Finnish.DG			3			O			O
French.DG			4			O			O
Georgian.DG			2			O			O
Greek.DG			2			O			O
Han.DG			4			O			O
Hawaiian.DG			1			O			O
Hazara.DG			2			O			O
Hezhen.DG			2			O			O
Hungarian.DG			2			O			O
Icelandic.DG			2			O			O
Igbo.DG			2			O			O
Igorot.DG			2			O			O
Iranian.DG			2			O			O

Appendix 1. Continued.

Group name	Subgroup	Symbol	N	Eurasian PCA	East Asian PCA	f_3/ f_4	qpWave/ qpAdm	HO dataset	1240K dataset
Iraqi_Jew.DG			2			O			O
Irula.DG			2			O			O
Itelman.DG			1			O			O
Japanese.DG (SGDP Japanese)			3			O	O		O
Jordanian.DG			3			O			O
Kalash.DG			2			O			O
Kapu.DG			2			O			O
Karitiana.DG			4			O			O
Kashmiri_Pandit.DG			1			O			O
Kharia.DG			1			O			O
Khonda_Dora.DG			1			O			O
Kinh.DG			2			O			O
Kongo.DG			1			O			O
Korean.DG (SGDP Korean)			2			O	O		O
Kurumba.DG			1			O			O
Kusunda.DG			2			O			O
Kyrgyz.DG			2			O			O
Lahu.DG			2			O			O
Lemande.DG			2			O			O
Lezgin.DG			2			O			O
Madiga.DG			2			O			O
Makrani.DG			2			O			O
Mala.DG			3			O			O
Mansi.DG			2			O			O
Maori.DG			1			O			O
Mayan.DG			2			O			O
Mbuti.DG			5				O		O
Miao.DG			2			O			O
Mixe.DG			3			O			O
Mixtec.DG			2			O			O
Mongola.DG			2			O			O
Nahua.DG			1			O			O
Naxi.DG			3			O			O
North_Ossetian.DG			2			O			O
Onge.DG			2			O	O		O
Orcadian.DG			2			O			O
Oroqen.DG			2			O			O
Palestinian.DG			3			O			O
Papuan.DG			16			O			O
Pathan.DG			2			O			O
Piapoco.DG			2			O			O
Pima.DG			2			O			O
Polish.DG			1			O			O

Appendix 1. Continued.

Group name	Subgroup	Symbol	N	Eurasian PCA	East Asian PCA	f_3/ f_4	qpWave/ qpAdm	HO dataset	1240K dataset
unjabi.DG			4			O			O
Quechua.DG			3			O			O
Relli.DG			2			O			O
Russian.DG			2			O			O
Samaritan.DG			1			O			O
Sardinian.DG			4			O			O
She.DG			2			O			O
Sindhi.DG			2			O			O
Spanish.DG			2			O			O
Surui.DG			2			O			O
Tajik.DG			2			O			O
Thai.DG			2			O			O
Tlingit.DG			2			O			O
Tubalar.DG			2			O			O
Tu.DG			2			O			O
Tujia.DG			2			O			O
Turkish.DG			2			O			O
Tuscan.DG			2			O			O
Ulchi.DG			2			O			O
Uygur.DG			2			O			O
Xibo.DG			2			O			O
Yadava.DG			2			O			O
Yakut.DG			2			O			O
Yemenite_Jew.DG			2			O			O
Yi.DG			2			O			O
Zapotec.DG			2			O			O

Appendix 2. A list of ancient world-wide populations used in this study. This list includes 167 ancient world-wide populations genotyped on 1240K panel. "N" column represents the total number of individuals for each population. Populations used for each analysis method are shown. References that reported the samples and the individual IDs are represented.

Group name	Subgroup	N	Eurasian PCA	East Asian PCA	f_3/f_4	qpWave/ qpAdm	HO dataset	1240K dataset	Reference	Individual IDs
Afanasievo		24			O		O	O	Narasimhan et al. (2019) Science	I6712, I6711, I6713, I6715, I2069, I2071, I3388, I3950, I3954, I1829, I3952, I5269, I5270, I5271, I5272, I5273, I5277, I5278, I5279, I3387, I10565, I10564, I11112, I11752
Afanasievo_Mongolia		2			O		O	O	Jeong et al. (2020) Cell	SHT001.B0101, SHT002.A0101
AG2		1			O		O	O	Raghavan et al. (2014) Nature	AfontovaGora2.SG
AG3		1			O		O	O	Fu et al. (2016) Nature	AfontovaGora3_d
Alan		5			O		O	O	Damgaard et al. (2018) Nature	DA146, DA160, DA162, DA164, DA243
Aldy_Bel_IA		2			O		O	O	Unterländer et al. (2017) Nat. Commun.	I0576, I0577
Altai_MLBA		6			O		O	O	Jeong et al. (2020) Cell	ULI001.B0101, ULI003.A0101, SBG001.A0201, BER002.A0101, BIL001.A0101, ULZ001.A0101
Anatolia_EBA		3			O		O	O	Damgaard et al. (2018) Science	MA2213, MA2212, MA2210
Anatolia_MLBA		5			O		O	O	Damgaard et al. (2018) Science	MA2208, MA2206, MA2205, MA2203, MA2200
Anatolia_N		23			O	O	O	O	Mathieson et al. (2015) Nature	I1100, I1102, I1099, I1103, I1101, I1097, I0744, I1096, I1098, I0708, I0745, I0746, I0707, I0709, I0725, I0736, I0726, I0723, I0724, I0727, I1580, I1581, I1583
Andronovo.SG		5			O		O	O	Allentoft et al. (2015) Nature	RISE500.SG, RISE503.SG, RISE505.SG, RISE512.SG, DA231

Appendix 2. Continued.

Group name	Subgroup	N	Eurasian PCA	East Asian PCA	f_3/f_4	qpWave/ qpAdm	HO dataset	1240K dataset	Reference	Individual IDs
AR_EN		2	O	O	O		O	O	Ning et al. (2020) Nat. Commun.	WQM4, ZLNR−2
AR_IA		1			O		O	O	Ning et al. (2020) Nat. Commun.	ZLNR−1
Armenia_ChL		5			O		O	O	Lazaridis et al. (2016) Nature	I1634, I1632, I1631, I1409, I1407
Armenia_EBA		3			O		O	O	Lazaridis et al. (2016) Nature	I1635, I1633, I1658
Armenia_LBA.SG		5			O		O	O	Allentoft et al. (2015) Nature	RISE396.SG, RISE397.SG, RISE407.SG, RISE408.SG, RISE412.SG
AR_Xianbei_IA		3	O	O	O	O	O	O	Ning et al. (2020) Nat. Commun.	MGS−M6, MGS−M7L, MGS−M7R
Baikal_EBA		23			O		O	O	Damgaard et al. (2018) Science	DA354, DA358, DA360, DA334, DA335, DA336, DA337, DA338, DA339, DA343, DA353, DA356, DA361, KPT001, KPT002, KPT003, STB001, GLZ003, KPT004, ZPL001, ZPL002, KAG002, KPT006
Baikal_EN		18			O	O	O	O	Damgaard et al. (2018) Science	DA340, DA341, DA357, DA359, DA245, DA246, DA247, DA248, DA249, DA250, DA251, DA252, DA253, DA362, KAG001, STB002, IUO001, ANG001
Balkans_BronzeAge		12			O		O	O	Mathieson et al. (2018) Nature	Bul10, Bul6, Bul8, I2163, I2165, I3313, I2510, I2520, I4331, I4332, I2175, I2176
Balkans_Chalcolithic		14			O		O	O	Mathieson et al. (2018) Nature	I2427, I2431, I0785, I0781, I4088, I4089, I2509, I2519, I2423, I2424, I2425, I2426, I2430, I5079
Balkans_IronAge		1			O		O	O	Mathieson et al. (2018) Nature	I5769
Ban_Chiang		5			O		O	O	Lipson et al. (2018) Science	I8970, I8974, I8977, I8978, I4458_new
Bell_Beaker_Germany		8			O		O	O	Mathieson et al. (2015) Nature	I1546, I0806, I0805, I0113, I0112, I0111, I0108, I1549
Bianbian		1			O		O	O	Yang et al. (2020) Science	Bianbian
Boshan		1	O	O	O	O	O	O	Yang et al. (2020) Science	BS

Appendix 2. Continued.

Group name	Subgroup	N	Eurasian PCA	East Asian PCA	f_3/f_4	qpWave/ qpAdm	HO dataset	1240K dataset	Reference	Individual IDs
Botai_pub		3			O		O	O	Damgaard et al. (2018) Science	BOT2016, BOT15, BOT14
BZK002		1			O		O	O	Yu et al. (2020) Cell	BZK002
Canaanite_Lebanon.SG		5			O		O	O	Haber et al. (2017) Am. J. Hum. Genet.	ERS1790729.SG, ERS1790730.SG, ERS1790731.SG, ERS1790732.SG, ERS1790733.SG
centralMongolia_preBA		1			O		O	O	Jeong et al. (2020) Cell	ERM003.A0101
CentralSteppe_EMBA		2			O		O	O	Damgaard et al. (2018) Science	Kazakhstan_EBA2, Kazakhstan_EBA1
Chandman_IA		9			O		O	O	Jeong et al. (2020) Cell	CHN001.C0101, CHN003.A0101, CHN004.A0101, CHN006.A0101, CHN007.A0101, CHN008.A0101, CHN012.A0101, CHN015.A0101, CHN016.A0101
Chemurchek_northAlai		2			O		O	O	Jeong et al. (2020) Cell	KUR001.B0101, KUM001.B0101
Chemurchek_southAltai		2			O		O	O	Jeong et al. (2020) Cell	YAG001.B0101, IAG001.B0102
CHG		2			O		O	O	Jones et al. (2015) Nat. Commun.	KK1.SG, SATP.SG
Chokhopani.SG		1			O		O	O	Jeong et al. (2016) Proc. Natl. Acad. Sci. USA	C1.SG
Chuanyun		1			O		O	O	Yang et al. (2020) Science	L5694
Clovis.SG		1			O		O	O	Rasmussen et al. (2014) Nature	Anzick.SG
Corded_Ware_Estonia.SG		6			O		O	O	Allentoft et al. (2015) Nature	MA826.RISE00.SG, MA968.MA976.SG, MA969.SG, MA971.SG, MA973.SG, MA976.SG
Corded_Ware_Germany		12			O		O	O	Mathieson et al. (2015) Nature	I1542, I1536, I1544, I1538, I1539, I1534, I0106, I1540, I1532, I0049, I0103, I0104
DevilsCave_N		6	O	O	O		O	O	Sikora et al. (2019) Nature	NEO235, NEO236, NEO237, NEO238, NEO239, NEO240

Appendix 2. Continued.

Group name	Subgroup	N	Eurasian PCA	East Asian PCA	f_3/f_4	qpWave/ qpAdm	HO dataset	1240K dataset	Reference	Individual IDs
earlyMed_Turk		7			O		O	O	Jeong et al. (2020) Cell	ULI002.B0101, NOM001.B0101, TSB001.A0101, UGU001.A0101, ZAA002.A0101, ZAA004.A0101, ZAA007.A0101
earlyMed_Uigur		12			O		O	O	Jeong et al. (2020) Cell	OLN002.A0101, ZAA001.A0101, OLN003.B0101, OLN001.A0101, OLN001.B0101, OLN004.A0101, OLN005.A0101, OLN008.A0101, OLN009.A0101, OLN010.A0101, OLN011.A0101, OLN012.A0101
earlyXiongnu_rest		6			O		O	O	Jeong et al. (2020) Cell	JAG001.A0101, SKT002.A0101, SKT004.A0101, SKT005.A0101, SKT006.A0101, SKT012.A0101
earlyXiongnu_SKT007		1			O		O	O	Jeong et al. (2020) Cell	SKT007.A0101
earlyXiongnu_west		6			O		O	O	Jeong et al. (2020) Cell	AST001.A0101, SKT001.A0101, SKT003.A0101, SKT008.A0101, SKT009.A0101, SKT010.A0101
eastMongolia_preBA		1			O		O	O	Jeong et al. (2020) Cell	SOU001.A0101
EHG		4			O		O	O	Damgaard et al. (2018) Science; Mathieson et al. (2015) Nature	I0124, I0211, I0061, Sidelkino_EHG
ElMiron		1			O		O	O	Fu et al. (2016) Nature	ElMiron_d
Fofonovo_EN		4			O		O	O	Jeong et al. (2020) Cell	FNO001.A0101, FNO003.A0101, FNO006.B0101, FNO007.A0101
GlinoeSad_Cimmerian		2			O		O	O	Krzewińska et al. (2018) Sci. Adv.	cim357, cim358
Glinoe_Scythian		9			O		O	O	Krzewińska et al. (2018) Sci. Adv.	scy192, scy193, scy197, scy300, scy301, scy304, scy305, scy311, scy332
Globular_Amphora		9			O		O	O	Mathieson et al. (2018) Nature	I2405, I2433, I2434, I2440, I2441, I2403, ILK001, ILK002, ILK003
Gonur1_BA		20			O		O	O	Narasimhan et al. (2019) Science	I1781, I1782, I1787, I1788, I1790, I1793, I2085, I2087, I2116, I2121, I2125, I2128, I3374, I1784, I6119, I6124, I6217, I10411, I10410, I11042
Gunsan_Jar_Coffin		6	O	O	O	O	O	O	This study	GUC001.A0101, GUC002.A0101, GUC003.A0101, GUC004.A0101, GUC005.A0101, GUC007.A0101

Appendix 2. Continued.

Group name	Subgroup	N	Eurasian PCA	East Asian PCA	f_3/f_4	qpWave/ qpAdm	HO dataset	1240K dataset	Reference	Individual IDs
HMMH_MN		1	O	O	O		O	O	Ning et al. (2020) Nat. Commun.	HMF32
Hun_TianShan_1700BP		22			O		O	O	Damgaard et al. (2018) Nature	DA100, DA101, DA104, DA105, DA385, DA52, DA54, DA65, DA66, DA68, DA69, DA70, DA71, DA72, DA73, DA74, DA80, DA81, DA82, DA85, DA96, DA98
Iberia_ChL		12			O		O	O	Mathieson et al. (2015) Nature	I1282, I1276, I1284, I1280, I1314, I1277, I1272, I1281, I1300, I1271, I1303, I0581
Iberia_EN		4			O		O	O	Mathieson et al. (2015) Nature	I0409, I0412, I0410, I0413
Iberia_MN		4			O		O	O	Mathieson et al. (2015) Nature	I0405, I0407, I0408, I0406
Imdang		31			O	O	O	O	This study	IMD002.A0101, IMD004.A0101, IMD008.A0101, IMD009.A0101, IMD010.A0101, IMD011.A0101, IMD016.A0101, JOY001.A0101, JOY002.A0101, JOY004.A0101, JOY005.A0101, JOY007.A0101, JOY009.A0101, JOY010.A0101, JOY012.A0101, JOY013.A0101, JOY015.A0101, JOY018.A0101, JOY019.A0101, JOY020.A0101, JOY022.A0101, JOY023.A0101, JOY024.A0101, JOY025.A0101, JOY026.A0101, JOY028.A0101, IMD006.A0101, IMD007.A0101, IMD018.A0101, JOY006.A0101, JOY017.A0101
	Imdang_ main	4	O	O	O	O	O	O	This study	JOY001.A0101, JOY012.A0101, JOY015.A0101, JOY028.A0101
	Imdang_ sacrificed	21	O	O	O	O	O	O	This study	IMD004.A0101, IMD006.A0101, IMD007.A0101, IMD011.A0101, JOY002.A0101, JOY003.A0101, JOY004.A0101, JOY005.A0101, JOY006.A0101, JOY007.A0101, JOY009.A0101, JOY010.A0101, JOY011.A0101, JOY013.A0101, JOY016.A0101, JOY019.A0101, JOY020.A0101, JOY023.A0101, JOY024.A0101, JOY025.A0101, JOY026.A0101
	Imdang_ unknown	10	O	O	O	O	O	O	This study	IMD002.A0101, IMD003.A0101, IMD008.A0101, IMD009.A0101, IMD010.A0101, IMD016.A0101, IMD018.A0101, JOY017.A0101, JOY018.A0101, JOY022.A0101

Appendix 2. Continued.

Group name	Subgroup	N	Eurasian PCA	East Asian PCA	f_3/f_4	qpWave/ qpAdm	HO dataset	1240K dataset	Reference	Individual IDs
Imdang.SG		13			O	O	O	O	This study	IMD003.A0101_SE, IMD004.A0101_SE, IMD010.A0101_SE, IMD011.A0101_SE, IMD016.A0101_SE, JOY002.A0101_SE, JOY004.A0101_SE, JOY005.A0101_SE, JOY007.A0101_SE, JOY015.A0101_SE, JOY016.A0101_SE, JOY020.A0101_SE, JOY024.A0101_SE
Iran_ChL		5			O		O	O	Lazaridis et al. (2016) Nature	I1661, I1670, I1662, I1674, I1665
Iran_N		8			O	O	O	O	Lazaridis et al. (2016) Nature	I1290, I1944, I1945, I1947, I1949, I1954, I1951, I7527
Israel_C		19			O		O	O	Harney et al. (2018) Nat. Commun.	I0644, I1154, I1177, I1155, I1181, I1172, I1187, I1171, I1164, I1184, I1179, I1182, I1170, I1168, I1152, I1178, I1165, I1160, I1169
Jomon_Funadomari		2			O	O	O	O	Kanzawa-Kiriyama et al. (2019) Anthropol. Sci.	FUN5, FUN23
Jomon_Ikawazu		1			O	O	O	O	McColl et al. (2018) Science	IK002
Jordan_EBA		3			O		O	O	Lazaridis et al. (2016) Nature	I1705, I1706, I1730
Kangju_1800BP		6			O		O	O	Damgaard et al. (2018) Nature	DA121, DA123, DA125, DA206, DA226, DA229
Karakhanid_TianShan_1100BP		3			O		O	O	Damgaard et al. (2018) Nature	DA203, DA204, DA205
Karasuk.SG		7			O		O	O	Allentoft et al. (2015) Nature	RISE493.SG, RISE494.SG, RISE495.SG, RISE496.SG, RISE497.SG, RISE499.SG, RISE502.SG
Karluk_TianShan		2			O		O	O	Damgaard et al. (2018) Nature	DA222, DA230
Kazakhstan_Berel_IA		2			O		O	O	Unterländer et al. (2017) Nat. Commun.	I0563, I0562
Kennewick_WA.SG		1			O		O	O	Rasmussen et al. (2015) Nature	kennewick.SG

Appendix 2. Continued.

Group name	Subgroup	N	Eurasian PCA	East Asian PCA	f_3/f_4	qpWave/ qpAdm	HO dataset	1240K dataset	Reference	Individual IDs
Khovsgol_LBA		14			O		O	O	Jeong et al. (2018) Proc. Natl. Acad. Sci. USA	ARS001.A0101, ARS002.B0101, ARS003.B0101, ARS005.A0101, ARS006.A0101, ARS007.B0101, ARS008.A0101, ARS012.A0101, ARS013.A0101, ARS014.B0101, ARS016.A0101, ARS018.A0101, ARS024.A0101, ARS025.A0101
Kimak_Kazakhstan_1350BP		1			O		O	O	Damgaard et al. (2018) Nature	DA87
Kipchak		2			O		O	O	Damgaard et al. (2018) Nature	DA179, DA23
KPT005		1			O		O	O	Yu et al. (2020) Cell	KPT005
LateDorset.SG		1			O		O	O	Raghavan et al. (2014) Science	LateDorset-XIV-H_126.SG
lateMed_Khitan		3			O		O	O	Jeong et al. (2020) Cell	ULA001.A0101, ZAA003.A0101, ZAA005.A0101
lateMed_Mongol		61			O		O	O	Jeong et al. (2020) Cell	ARG002.B0101, GAN002.A0101, KRN001.A0101, UGU002.B0101, UGU003.A0101, SHA001.A0101, SHG001.A0102, ARG001.B0101, DAR002.B0101, ERD001.B0101, KHN001.A0101, KHO001.A0101, MRI001.A0101, RAH001.B0101, TSA001.B0101, ULN011.A0101, ZAM001.A0101, ARG003.A0101, BAU001.A0101, BAY001.A0101, BAZ001.A0101, BRG001.D0101, BRG002.B0101, BRG004.A0101, BRG005.B0101, BRL001.B0101, CHD001.B0101, CHN014.A0101, DAS001.A0101, DEE001.A0101, DUU002.A0101, GTO001.B0101, GUN002.A0101, IKU001.A0101, KGK001.A0101, KHL001.A0101, KHV002.A0101, KNN001.B0101, KNU001.A0101, KRN002.A0101, NRC001.A0101, SHG002.A0101, SHG003.A0101, TAH002.A0101, TAV001.A0101, TAV006.A0101, TAV011.A0101, TSA002.B0101, TSA003.A0101, TSA004.B0101, TSA005.A0101, TSA006.A0101, TSA007.A0101, UGO001.A0101, UGO002.A0101, UUS002.A0101, ZAM002.A0101, ZAR002.A0101, ZAY001.A0101, TAV005.A0101, SHR001.B0101

Appendix 2. Continued.

Group name	Subgroup	N	Eurasian PCA	East Asian PCA	f_3/f_4	qpWave/ qpAdm	HO dataset	1240K dataset	Reference	Individual IDs
lateXiongnu		25			O		O	O	Jeong et al. (2020) Cell	SAN001.A0101, UGU004.A0101, SOL001.A0101, DEL001.A0101, ULN004.B0101, BTO001.A0101, CHN010.A0101, DOL001.B0101, EME002.A0101, JAA001.A0101, KHO007.A0101, TAK001.A0101, TEV002.A0101, TUK003.A0101, UGU011.A0101, UVG001.A0101, TEV003.A0101, IMA001.B0101, IMA002.A0101, IMA003.C0101, IMA004.B0101, IMA005.C0101, IMA006.A0101, IMA007.A0101, IMA008.B0101
lateXiongnu_han		8			O		O	O	Jeong et al. (2020) Cell	ATS001.A0101, BAM001.A0101, BRU001.A0101, KHO006.A0101, SON001.A0101, TUH001.A0101, TUH002.A0101, TUK002.A0101
lateXiongnu_sarmatian		13			O		O	O	Jeong et al. (2020) Cell	TMI001.B0101, BUR001.A0101, NAI001.A0101, BUR002.B0101, BRL002.A0101, BUR003.A0101, BUR004.A0101, DUU001.B0101, HUD001.A0101, NAI002.A0101, UGU005.A0101, UGU006.A0101, UGU010.A0101
LBK_EN		6			O	O	O	O	Mathieson et al. (2015) Nature	I0797, I0795, I0022, I0026, I0025, I0054
LchashenMetsamor		2			O		O	O	Damgaard et al. (2018) Nature	DA31, DA35
Levant_N		12			O		O	O	Lazaridis et al. (2016) Nature	I1679, I1709, I1727, I1710, I1707, I1704, I1699, I0867, I1701, I1414, I1415, I1700
Liangdao1		1			O		O	O	Yang et al. (2020) Science	LD1
Liangdao2		1	O	O	O	O	O	O	Yang et al. (2020) Science	LD2
Loebanr_IA		33			O		O	O	Narasimhan et al. (2019) Science	I5400, I6553, I6555, I6556, I6292, I10000, I10001, I10974, I8998, I8999, I12134, I12137, I12980, I12988, I12982, I12981, I12987, I12983, I12984, I12979, I12456, I12455, I12457, I12458, I12459, I13224, I13222, I13223, I13227, I13226, I13221, I13220, I13228
MA1		1			O		O	O	Raghavan et al. (2014) Nature	MA1.SG

Appendix 2. Continued.

Group name	Subgroup	N	Eurasian PCA	East Asian PCA	f_3/f_4	qpWave/ qpAdm	HO dataset	1240K dataset	Reference	Individual IDs
Man_Bac		8			O		O	O	Lipson et al. (2018) Science	I0626, I0627, I1135, I1137, I1859, I2731, I10973, I2947_new
McColl_SEA_GR1		2			O		O	O	McColl et al. (2018) Science	La368, Ma911
McColl_SEA_GR2		6			O		O	O	McColl et al. (2018) Science	La364, La727, La898, Ma912, Vt833, Vt880
McColl_SEA_GR3		5			O		O	O	McColl et al. (2018) Science	Vt777, Vt779, Vt781, Vt796, Vt808
McColl_SEA_GR31		2			O		O	O	McColl et al. (2018) Science	Th531, Vt719
McColl_SEA_GR4		4			O		O	O	McColl et al. (2018) Science	Th519, Th521, Th530, Th703
McColl_SEA_GR41		1			O		O	O	McColl et al. (2018) Science	Vt778
McColl_SEA_GR5		2			O		O	O	McColl et al. (2018) Science	In661, In662
McColl_SEA_GR6		3			O		O	O	McColl et al. (2018) Science	Ma554, Ma555, Phl534
Mezhovskaya.SG		3			O		O	O	Allentoft et al. (2015) Nature	RISE523.SG, RISE524.SG, RISE525.SG
Miaozigou_MN		3	O	O	O	O	O	O	Ning et al. (2020) Nat. Commun.	MZGM10−1, MZGM16, MZGM25−2
Minoan_Lasithi		5			O		O	O	Lazaridis et al. (2017) Nature	I0070, I0071, I0073, I0074, I9005
Minoan_Odigitria		5			O		O	O	Lazaridis et al. (2017) Nature	I9127, I9128, I9129, I9130, I9131
Mokra_Cimmerian		1			O		O	O	Krzewińska et al. (2018) Sci. Adv.	cim359
Mycenaean		4			O		O	O	Lazaridis et al. (2017) Nature	I9010, I9006, I9033, I9041
Namazga_CA		4			O		O	O	Damgaard et al. (2018) Science	DA380, DA383, DA381, DA379
Natufian		6			O		O	O	Lazaridis et al. (2016) Nature	I0861, I1072, I1069, I1685, I1687, I1690

Appendix 2. Continued.

Group name	Subgroup	N	Eurasian PCA	East Asian PCA	f_3/f_4	qpWave/ qpAdm	HO dataset	1240K dataset	Reference	Individual IDs
Nomad_Hun−Sarmatian		2			O		O	O	Damgaard et al. (2018) Nature	DA20, DA27
Nomad_Kazakhstan_1700BP		2			O		O	O	Damgaard et al. (2018) Nature	DA177, DA95
Nomad_Kazakhstan_IA		2			O		O	O	Damgaard et al. (2018) Nature	DA18, DA92
Nomad_Kyrgyzstan_IA		2			O		O	O	Damgaard et al. (2018) Nature	DA129, DA221
Nomad_Med		10			O		O	O	Damgaard et al. (2018) Nature	DA106, DA116, DA118, DA124, DA126, DA128, DA142, DA93, DA94, DA99
Nui_Nap		2			O		O	O	Lipson et al. (2018) Science	I2497, I2948_new
Oakaie1		2			O		O	O	Lipson et al. (2018) Science	I7238_new, I4011_new
Okunevo		16			O		O	O	Damgaard et al. (2018) Science	RISE515, RISE516, RISE662, RISE664, RISE670, RISE671, RISE672, RISE673, RISE674, RISE675, RISE680, RISE683, RISE684, RISE685, RISE718, RISE719
Poltavka		8			O		O	O	Mathieson et al. (2015) Nature	I0126, I0440, I0374, I0371, I0442, I6294, I7671, I8745
Poprad		1			O		O	O	Damgaard et al. (2018) Nature	DA119
Potapovka		5			O		O	O	Mathieson et al. (2015) Nature	I0418, I0244, I0246, I0419, I7489
Qihe		2	O	O	O		O	O	Yang et al. (2020) Science; Wang et al. (2021) Cell	Qihe2_d, Qihe3
Russia_EBA.SG		1			O		O	O	Allentoft et al. (2015) Nature	RISE555.SG
Russia_IA.SG		5			O		O	O	Allentoft et al. (2015) Nature	RISE492.SG, RISE504.SG, RISE600.SG, RISE601.SG, RISE602.SG
Russia_LBA.SG		2			O		O	O	Allentoft et al. (2015) Nature	RISE553.SG, RISE554.SG
Saka_Kazakhstan_2500BP		6			O		O	O	Damgaard et al. (2018) Nature	DA10, DA11, DA13, DA14, DA15, DA16

Appendix 2. Continued.

Group name	Subgroup	N	Eurasian PCA	East Asian PCA	f_3/f_4	qpWave/ qpAdm	HO dataset	1240K dataset	Reference	Individual IDs
Saka_TianShan_2150BP		10			O		O	O	Damgaard et al. (2018) Nature	DA130, DA47, DA48, DA49, DA51, DA53, DA55, DA57, DA58, DA59
SaltovoMayaki		3			O		O	O	Damgaard et al. (2018) Nature	DA188, DA189, DA190
Saqqaq.SG		1			O		O	O	Rasmussen et al. (2010) Nature	Inuk.SG
Sarmatian		17			O		O	O	Damgaard et al. (2018) Nature; Krzewińska et al. (2018) Sci. Adv.; Mathieson et al. (2015) Nature; Unterländer et al. (2017) Nat. Commun.	I0574, I0575, I0247, DA134, DA139, DA141, DA143, DA144, DA145, DA202, DA26, DA30, chy001, chy002, tem001, tem002, tem003
Scythian_Hungary_2400BP		5			O		O	O	Damgaard et al. (2018) Nature	DA191, DA194, DA195, DA197, DA198
SHG		6			O		O	O	Mathieson et al. (2015) Nature	I0013, I0011, I0015, I0012, I0014, I0017
Shimao_LN		3	O	O	O	O	O	O	Ning et al. (2020) Nat. Commun.	SM-SGDLM27, SM-SGDLM6, SM-SGDLM7X
Sintashta_MLBA		38			O	O	O	O	Narasimhan et al. (2019) Science	I1060, I1062, I1065, I1022, I1088, I1084, I1082, I1055, I1018, I1019, I1086, I1090, I1024, I1063, I1029, I1064, I1089, I1053, I1027, I0982, I0937, I0938, I0939, I0942, I0943, I0984, I0986, I0989, I1003, I1006, I1008, I1011, I1012, I1013, I1061, I7480, I7670, I0987
Srubnaya		11			O		O	O	Mathieson et al. (2015) Nature	I0234, I0235, I0233, I0232, I0358, I0359, I0361, I0422, I0424, I0430, I0431
Starosillya_Scythian		3			O		O	O	Krzewińska et al. (2018) Sci. Adv.	scy006, scy009, scy010
Suogang		2			O		O	O	Yang et al. (2020) Science	SuogangB1_d, SuogangB3_d
Tagar		8			O		O	O	Damgaard et al. (2018) Nature	DA2, DA3, DA4, DA5, DA6, DA7, DA8, DA9

Appendix 2. Continued.

Group name	Subgroup	N	Eurasian PCA	East Asian PCA	f_3/f_4	qpWave/ qpAdm	HO dataset	1240K dataset	Reference	Individual IDs
Tanshishan		4			O		O	O	Yang et al. (2020) Science	L5696_d, L5698_d, L7417_d, L7415
Tianyuan		1			O		O	O	Yang et al. (2017) Curr. Biol.	Tianyuan
Turk		3			O		O	O	Damgaard et al. (2018) Nature	DA228, DA86, DA89
Turkmenistan_IA		1			O		O	O	Damgaard et al. (2018) Science	DA382
UKY		1			O		O	O	Yu et al. (2020) Cell	UKY001
Ulaanzuukh_SlabGrave		16			O		O	O	Jeong et al. (2020) Cell	BOR001.A0101, DAR001.B0101, ULN003.A0101, ULN005.A0102, BUL001.A0101, BUL002.A0101, ULN001.A0101, ULN002.A0101, ULN009.A0101, ULN010.A0101, MIT001.A0101, SHU001.A0101, ULN006.A0101, ULN007.A0101, ULN015.A0101, PTO001.A0101
Unetice_EBA		6			O		O	O	Mathieson et al. (2015) Nature	I0115, I0117, I0804, I0803, I0164, I0116
Upper_YR_IA		4			O		O	O	Ning et al. (2020) Nat. Commun.	DCZM17IV, DCZ-M21II, DCZ-M22IV, DCZ-M6
Upper_YR_LN		7	O	O	O	O	O	O	Ning et al. (2020) Nat. Commun.	JCKM1-1, LJM14, LJM25, LJM2, LJM3, LJM4, LJM5
USR1		1			O	O	O	O	Moreno-Mayar et al. (2018) Nature	USR1.SG
UstIda_LN		4			O		O	O	Damgaard et al. (2018) Science	DA342, DA344, DA345, DA355
Ust_Ishim		1			O		O	O	Fu et al. (2014) Nature	Ust_Ishim_published.DG
Varna		4			O		O	O	Mathieson et al. (2018) Nature	ANI152, ANI153, ANI160, ANI159-ANI181
Vat_Komnou		1			O		O	O	Lipson et al. (2018) Science	I1680_new

Appendix 2. Continued.

Group name	Subgroup	N	Eurasian PCA	East Asian PCA	f_3/f_4	qpWave/ qpAdm	HO dataset	1240K dataset	Reference	Individual IDs
Villabruna		1			O	O	O	O	Fu et al. (2016) Nature	Villabruna
WHG		10			O		O	O	Lazaridis et al. (2014) Nature	Loschbour_published.DG, I0585, I1507, Falkenstein, Iboussieres25−1, Iboussieres31−2, Rochedane, BerryAuBac, I2158, I1875
WLR_BA		2	O	O	O	O	O	O	Ning et al. (2020) Nat. Commun.	91KLH11, 91KLH18
WLR_BA_o		1	O	O	O		O	O	Ning et al. (2020) Nat. Commun.	91KLM2
WLR_LN		3	O	O	O	O	O	O	Ning et al. (2020) Nat. Commun.	EDM124, EDM139, EDM176
WLR_MN		3	O	O	O		O	O	Ning et al. (2020) Nat. Commun.	BLSM27S, BLSM41, BLSM45
Wusun_TianShan_2200BP		4			O		O	O	Damgaard et al. (2018) Nature	DA201, DA220, DA223, DA227
Xiaogao		1			O		O	O	Yang et al. (2020) Science	Xiaogao
Xiaojingshan		3			O		O	O	Yang et al. (2020) Science	XJS1309_M4, XJS1309_M7, XJS1311_M16
Xiongnu		3			O		O	O	Damgaard et al. (2018) Nature	DA38, DA39, DA41
Xitoucun		7	O	O	O		O	O	Yang et al. (2020) Science	L5692_d, L5706_d, L5701_d, L5703_d, L5704_d, L5700, L5705
Yamnaya_Kalmykia.SG		6			O		O	O	Allentoft et al. (2015) Nature	RISE240.SG, RISE546.SG, RISE547.SG, RISE548.SG, RISE550.SG, RISE552.SG
Yamnaya_Samara		9			O		O	O	Mathieson et al. (2015) Nature	I0370, I0441, I0444, I0439, I0357, I0429, I0438, I0443, I0231
YR_LBIA		6			O		O	O	Ning et al. (2020) Nat. Commun.	HJTM115, HJTW13, JXNTM23, JXNTM2, LGM41, LGM79

Appendix 2. Continued.

Group name	Subgroup	N	Eurasian PCA	East Asian PCA	f_3/f_4	qpWave/ qpAdm	HO dataset	1240K dataset	Reference	Individual IDs
YR_LN		8	O	O	O	O	O	O	Ning et al. (2020) Nat. Commun.	HJTM107, HJTM109, PLTM310, PLTM311, PLTM312, PLTM313, WD-WT1H16, WD-WT5M2
YR_MN		8	O	O	O	O	O	O	Ning et al. (2020) Nat. Commun.	WGH35-1, WGM20, WGM35, WGM43, WGM70, WGM76S, WGM94, XW-M1R18
Yumin		1			O		O	O	Yang et al. (2020) Science	Yumin
Zevakino_Chilikta_IA.SG		2			O		O	O	Unterländer et al. (2017) Nat. Commun.	IS2.SG, Ze6b.SG