Ancient genomic DNA analysis of Jomon people

Kanzawa-Kiriyama Hideaki

DOCTOR OF PHILOSOPHY

Department of Genetics

School of Life Science

The Graduate University for Advanced Studies (SOKENDAI)

2013

Acknowledgements

First and foremost I would like to thank my academic supervisor, Professor Naruya Saitou for his expert advice and support throughout my study. I would also like to thank members of my current and previous progress committee: Drs. Ituro Inoue, Takashi Gojobori, Toshihiko Shiroishi, Kazuho Ikeo, Tishiyuki Takano, Naoki Osada, and Minori Shinya, for their useful advice and constructive criticisms. I am also grateful to Dr. Kryukov Kirill at University of Tokai, Drs. Timothy Jinam and Kazuyoshi Hosomichi at National Institute of Genetics, Dr. Ken-ichi Shinoda at National Science Museum, Dr. Noboru Adachi at University of Yamanashi, Drs. Minoru Yoneda, Shintaroh Ueda, Gen Suwa, and Aiko Saso at University of Tokyo for their assistance and collaboration on my research.

I would also like to take this opportunity to thank Dr. Kenta Sumiyama, former assistant professor at Saitou Laboratory, and Mr. Tsuneo Kakuda at National Science Museum, for their supports on many experiments, Drs. Tomoya Baba, Katsuhiko Yanagihara, and Yasuo Shirakibara of the National Institute of Genetics, for kindly providing me access to their experimental rooms, Ms Nadeeka Hettiarachchi Nilmini and Mr. Isaac Babarinde of Saitou Laboratory, for support on my data analysis.

Thanks also to all the members of Saitou Laboratory, past and present, for all their valuable support and friendship. Lastly, I wish to thank my family for their endless support and encouragement.

List of Figures

List of Tables

Abbreviation

YBP: year before present

mtDNA: mitochondrial DNA

C.I.: confidence interval

HGDP: Human Genome Diversity Project

PASNP: Pan-Asian SNP Consortium

JAHPGC: Japanese Archipelago Human Population Genetics Consortium

rCRS: revised Cambridge Reference Sequence

YRI: Yoruba in Ibadan, Nigeria

CEU: Utah residents with Northern and Western European ancestry

GIH: Gujarati Indian in Houston,TX

CHB: Han Chinese, Beijing

CHS: Southern Han Chinese, China

KHV: Kinh in Ho Chi Minh City, Vietnam

GWD: Gambian in Western Division, The Gambia

ESN: Esan in Nigeria

MSL: Mende in Sierra Leone

CDX: Chinese Dai in Xishuangbanna, China

JPT: Japanese in Tokyo, Japan

DNK: Dinka from Africa

CLM: Colombian in Medellin, Colombia

Abstract

Clarifying the genetic relationship between Neolithic East Asian Hunter-Gatherers and modern human populations is one of the keystones to understand the complex history of modern East Eurasian populations. Neolithic hunter-gatherers, Jomon people, inhabited the Japanese Archipelago from 16,000 years ago, and while their origin and relationship with modern humans have been long debated, details remain unclear. To solve these questions, I investigated mitochondrial DNA and nuclear DNA of Jomon individuals excavated from Sanganji Shell Mound, Fukushima Prefecture, Tohoku region, Japan. I decided mitochondrial DNA haplotypes from four Sanganji Jomon individuals, and conducted statistical analyses. I found that their mitochondrial DNA haplotypes, M7a and N9b, were unique to modern Japanese. When I compared frequencies of mitochondrial DNA haplotypes of modern East Eurasians and other Jomon populations which were previously reported, I observed that Northern Jomon people were genetically closer to Udegey of Southern Siberia than other continental East Eurasians. This implies that Northern Jomon people originated from Northeast Asia. Microblade culture with Araya type burin was introduced from Northeast Asia into Japanese Archipelago via Sakhalin in the Upper Paleolithic period. The genetic similarity among Northern Jomon people and modern Southern Siberian is therefore plausible. I also found that Jomon people were genetically heterogeneous in terms of their geographical regions. It is not clear whether this heterogeneity indicates the multi-origin or sub-structure of Jomon people. For investigating their origin and relationship with modern humans based on genomic DNAs, I sequenced 60 million base pairs of nuclear Jomon genomes with next generation sequencer, GAIIx and Hiseq 2000. The Jomon genomic DNA clearly evidenced that Neolithic Jomon people were genetically isolated from modern East Eurasians for a long time, and they diverged from the ancestors of both northern and southern East Eurasians but postdated the divergence of East Eurasian and Melanesians. It implies that the genetic structure of modern East Eurasians was formed after the divergence of Jomon people ancestors and other East Eurasians. Another conclusion is that Jomon genomic DNAs were highly inherited to Ainu people and Okinawa people, who live in

Northern and Southern part of the Japanese Archipelago, respectively, compared to Mainland people. The finding supports the dual structure model proposed by Hanihara Kazuro and others based on craniofacial data analyses. Another controversial issues are the genetic relationship between archaic humans and modern humans. Using phylogenetic network analysis and *D*-statistic analysis, I observed that Jomon people and other non-Africans shared more derived alleles with Vindija Neanderthal than Africans. The sharing confirms the existence of archaic gene flow from Vindija Neanderthal into the ancestors of modern non-Africans. Furthermore, I also detected Denisovan gene flow into not only Melanesians but also to Sanganji Jomon and Dai ethnic minority group of Southern China, while Ainu people as well as modern East Asians did not show a sign of the gene flow. This indicates that first migrants from Sundaland into East Asia had genetic components of Denisovan, but the remnants were diluted or disappeared in modern East Asians. The Denisovan DNA in East Eurasians and Melanesians is a footprint of early modern human migration into East Eurasia.

The heterogeneity of Jomon people in mitochondrial DNA implies the complexity of the Jomon history. To investigate whether Jomon people from different regions and period were genetically heterogeneous, I sequenced more Jomon genomes from four samples; Shikkariabe Jomon from Aomori Prefecture, Yugura Jomon from Nagano Prefecture, Daizenno-Minami Jomon from Chiba Prefecture, and Odake Jomon from Toyama Prefecture, respectively. I successfully sequenced about 10% and 80% of Jomon nuclear genomes from Yugura Jomon and Shikkariabe Jomon DNA samples, respectively. Using principal component analysis, I found that Jomon people were homogeneous at least in northern part of the Japanese Archipelago. The small pairwise distance between Shikkariabe Jomon and Yugura Jomon also supports their homogeneity. I also investigated the genetic relationship between Shikkariabe Jomon and archaic humans to examine whether Jomon people really have more Denisovan DNA materials than other East Eurasians. Using principal component analysis, phylogenetic network, and *D*-statistic analysis, I confirmed the evidence of Denisovan gene flow into the ancestors of Shikkariabe Jomon people.

I also sequenced the genome of Upper Paleolithic Ryukyuan, who lived ca. 20,000 years

ago, to investigate the earlier stages of human migration into the Japanese Archipelago, and the genetic relationship between Upper Paleolithic humans and Jomon people. I obtained 0.03% of the whole genome, and informative SNPs were limited. Despite of this small dataset, I observed a genetic similarity between Jomon and Paleolithic Ryukyuan. I speculate that both Jomon people and Paleolithic Ryukyuan share same ancestors, who migrated from Southeast Asia into East Asia.

In summary, I observed that Jomon people were genetically quite unique in East Eurasia, and it suggests that Jomon people were the descendants of early stage of humans who migrated into East Asia. The finding indicates that Jomon people are very important populations to clarify how early modern humans migrated into East Asia. A draft sequence of the Shikkariabe Jomon genome made me possible to compare the Jomon genome with other ancient and modern human genomes. This will certainly deepen the understanding the history of not only modern Japanese and Jomon people but also East Eurasians and Native Americans. Part of the Jomon genome from several individuals approached to the truth of their population structure (homogeneous or heterogeneous). In addition, genome sequences from 20,000 year-old Upper Paleolithic Ryukyuan will accelerate the elucidation of genetic relationship between Upper Paleolithic Japanese and Jomon/present-day individuals. Finally, one of the most important things is the genetic relationship between Jomon people and archaic humans. I observed the genetic relationship not only between Jomon and Vindija Neanderthal but also between Jomon and Denisovan. Since the conclusion of this study is based on partial sequences of the Jomon genome, further analyses with more genome sequence data and other statistical methods are necessary in future studies. The existence of Denisovan DNA material in Jomon people will be one of the main debates connected to not only the history of Japanese but also the origin and population structure of East Eurasians.

Chapter 1

General Introduction

1.1 Out of Africa and the history of East Eurasians

The history and origin of modern humans were studied for a long time in many fields such as archeology, anthropology, linguistics, and genetics. The analysis of mitochondrial and nuclear DNA revealed that *Homo sapiens* originated in Africa about 200,000 years before present (YBP), and dispersed from Africa by 60,000 YBP. The questions of how and when did they migrate from Africa are still under debates. Some part of the Paleolithic hunter-gatherers, who were in an early stage of modern humans, migrated northward and entered into the present-day Europe area, and the others migrated eastward and reached into East Eurasia (Figure 1.1). Southeast Asia was occupied by Paleolithic humans at least 50,000 YBP, and was a corridor through which the humans traveled not only into Sahul land (present-day Australia, Papua New Guinea, and Tasmania) but perhaps also into East Asia. Archeological evidences suggest that Paleolithic hunter-gatherers had reached Australia and East Asia by 50,000 YBP and 40,000 YBP, respectively. The stream of this series of human dispersal reached the Japanese Archipelago by 30,000 YBP.

Rice, *Oryza sativa japonica*, was first domesticated from a specific population of *Oriza rufipogon* around the middle area of the Pearl River in southern China (Huang et al., 2012), and the timing of rice domestication was between 9000 and 4500 YBP (Fuller et al., 2009). The rice agriculture dispersed to many regions of East Eurasia by 3,000 YBP. The disperse caused unquantifiable but a greater or lesser degree of genetic exchanges between the original indigenous populations and immigrants from mainland East Asia (Jacob, 1967). The rice agriculture entered into Japanese Archipelago via Korean peninsula about 3,000 YBP, and it seems that the genetic exchange between hunter-gatherers and agricultural people, so called Yayoi immigrants, is no exception. Recent advances in research methods have demonstrated a rapid population transition and associated

large-scale genetic exchange between newly immigrated agriculturists and pre-existing hunter-gatherers.

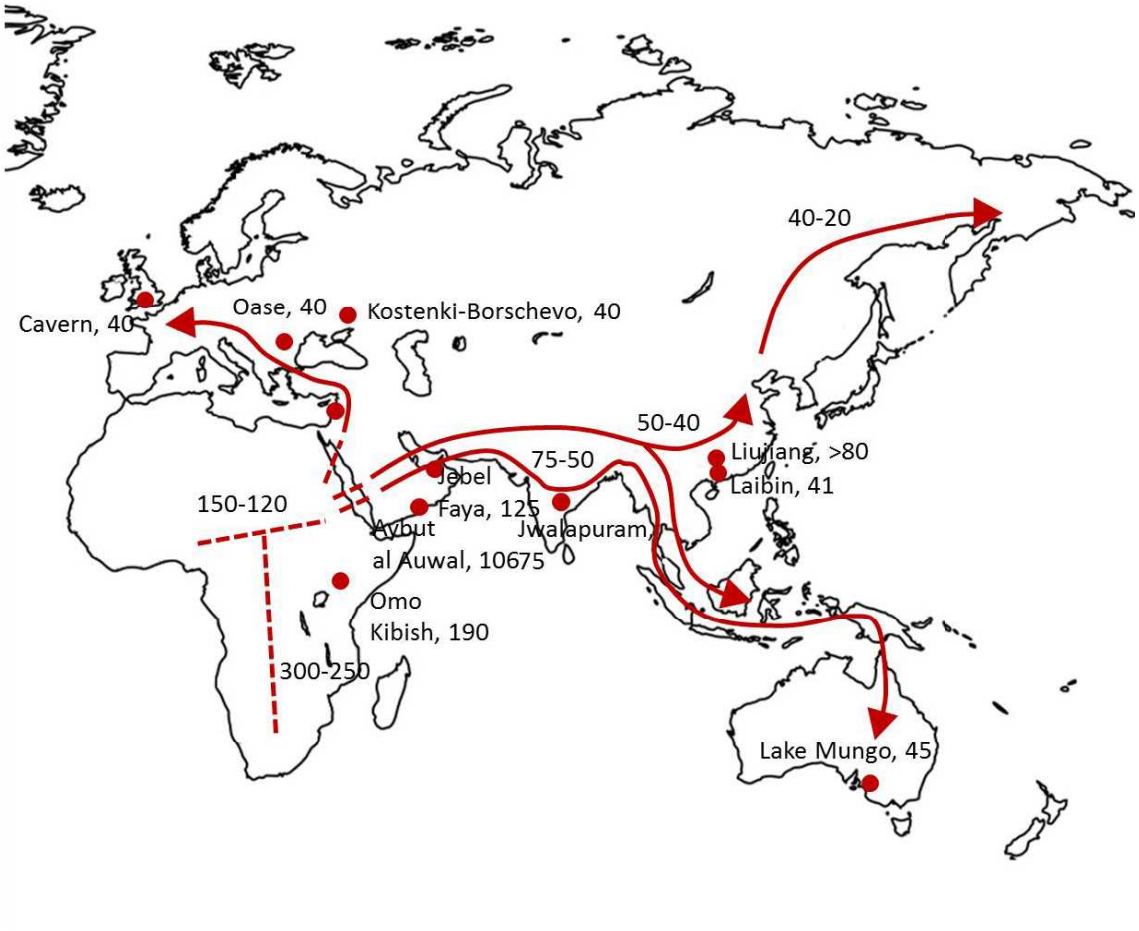


Figure 1.1: Populations and timescales involved in the origin of modern humans
(From Scally and Durbin [2012] with modification)

1.2 History of human populations on the Japanese Archipelago

In the last glacial maximum (between 26,500 and 19,000 YBP), temperatures were much lower than present days (Clark et al., 2009). Accordingly, the sea level was more than 100 meters lower than the present level at the peak (Lambeck et al., 2002). At that period, the Japanese Archipelago connected to the Eurasian continent, and continental humans could easily migrate into the Japanese Archipelago. Paleolithic stone tools evidence the existence of humans. Hand-axes, one of Paleolithic stone tools, were present before 30,000 YBP in the Japanese Archipelago (Imamura, 1996). This skill came from somewhere of continental East Eurasia. Knife-shape tools, which were widely used during the Upper Paleolithic period of the Japanese Archipelago, are rarely found in continental Eurasia. During the last stage of the Upper Paleolithic period, microblade tools prevailed throughout much of the Japanese Archipelago. The stone tool industry dispersed from continental Eurasia. The microblade differentiated into some types. Araya type burin, which widely spread over Northeast Asia, was introduced into Northern part of the Japanese Archipelago via Sakhalin, and non-Arayan type burin, which was distributed over Southern East Asia, was introduced into southern part of the Japanese Archipelago (Figure 1.2).

After the Upper Paleolithic period, Jomon culture began in the Japanese Archipelago. Cord marked ("jomon" in Japanese) pottery is the critical feature of Jomon culture, which ranged geographically from Hokkaido to Okinawa, stretching over 4,000 km from north to south. The Jomon period ranged from 16000 to 3000 YBP (~2,300 YBP in the northern of Japan) (Kobayashi, 2008, Harunari and Imamura, 2004) (Figure 1.3), and the Jomon period is typically divided into the following six periods: Incipient (16,000~11,000 cal BP), Earliest (10,500~7,000 cal BP), Early (7,000~5,470 cal BP), Middle (5,400~4,420 cal BP), Late (4,420~3,220 cal BP), and Final (3,220~2,300 cal BP). The Japanese Archipelago was apart from the East Eurasian continent after the end of the last glacial maximum, at 10,000-9,000 BP. Jomon people were not only geographically but perhaps also genetically isolated from continental Eurasia after that period.

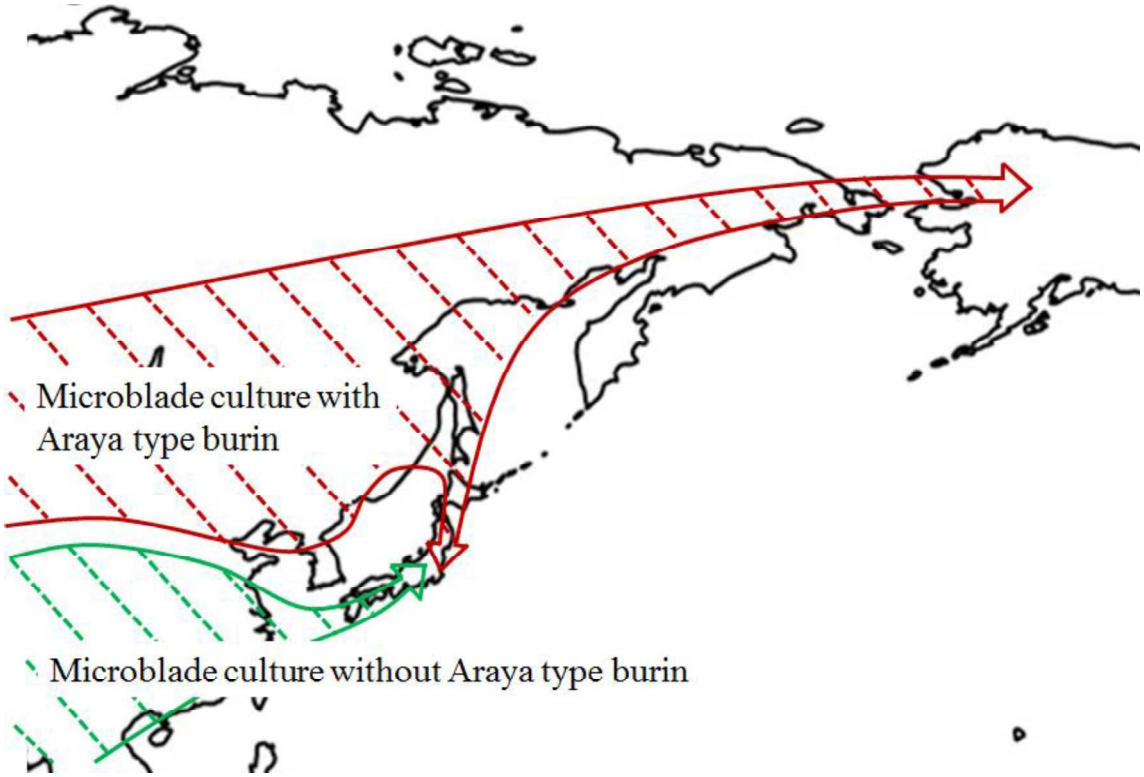


Figure 1.2: Introduction of microblade stone tools into Japanese Archipelago
(Based on Yoshizaki [1986])

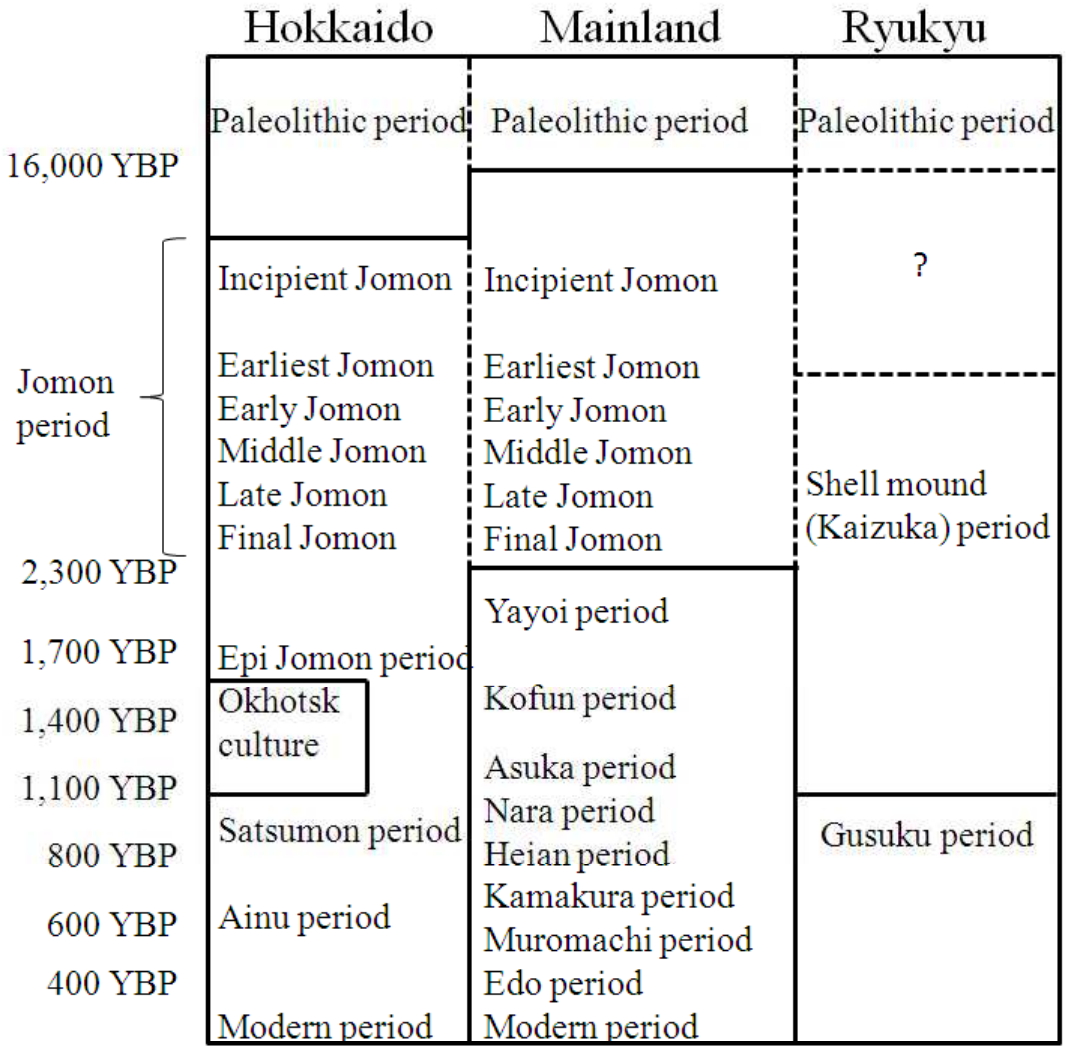


Figure 1.3: Brief chronological table of the cultural traditions in the three areas of the Japanese Archipelago
(from Nakahashi, 2005 with some modifications)

During the last stage of the Final Jomon period, agricultural people migrated into Japanese Archipelago via Korean Peninsula. It is currently believed that those agricultural migrants admixed with indigenous Jomon people, and became the genetic basis of modern Japanese. Within modern Japanese, Ainu and Ryukyuan, who live in Northern and Southern part of the Japanese Archipelago, respectively, are morphologically closer to Jomon people. This suggests that the two populations were less admixed with agricultural migrants and their descendants.

The origin of Jomon people has been debated for long time. Craniofacial data suggest that Jomon people came from somewhere in Southeast Asia (Turner, 1987, 1990; Hanihara, 1991; Matsumura, 2007; Matsumura et al., 2009, Yamaguchi, 1999), while archaeology and recent studies of cranial morphology suggest that the Jomon people were of northern origin (Imamura, 1996; Hanihara and Ishida, 2009; Nakashima et al., 2010). Molecular biology is a strong discipline to investigate the history of human migration. For the last 30 years, nuclear classic genetic markers (e.g. blood type, HLA, and Y-chromosome) and mitochondrial DNA data were available for the study of human history. Studies based on classic markers proposed that Jomon people were northern origin (Nei, 1995; Omoto and Saitou, 1997). Recently, genome wide SNP (single nucleotide polymorphism) data became available to investigate the history of human migrations. Jinam et al. (2012) presented the genetic uniqueness of Ainu people, but did not discuss about the origin of Jomon people. Modern Japanese are probably the admixture of Jomon and the agricultural people, and it is difficult to estimate the origin of Jomon people from the genetic information of modern humans, which are “indirect” evidences. As direct evidence, DNA sequences of Jomon people have been reported for almost a quarter century (Horai et al., 1989, 1991; Shinoda and Kanai, 1999; Shinoda, 2003; Adachi et al., 2008, 2009a, 2009b, 2011). The direct evidence may clarify the origin of Jomon people.

1.3 Sanganji Shell Mound

To clarify the Jomon origin, I attempted to sequence mitochondrial DNA and nuclear DNA from Jomon samples as direct evidence. I focused on samples taken from skeletal remains of the late to final Jomon period Sanganji Shell mound site in Tohoku Region, Japan (Figure 1.4). Sanganji Shell Mound is located in the coastal area of Fukushima Prefecture (Figure 1.5). The chronological age of the burials, belonging to the late to final Jomon period, is considered to lie within the range from 4,000 to 2,500 YBP, based on Jomon pottery chronology. The 1952-1954 excavations of the Sanganji Shell Mound were carried out by the Special Committee for Jomon Chronology of the Japanese Archaeological Association (1952 excavation), and by the Department of Anthropology, The University of Tokyo (1954 excavation). The remains have been stored in The University Museum, The University of Tokyo, for more than 50 years. Figure 1.6 shows the excavation trench of Sanganji Shell Mound in 1954, and skeletons of at least 40 Jomon individuals were excavated from the trench. The details of the excavation trench in 1952 were not reported, and we do not know the relationship among the individuals from 1952 excavation and those from 1954 excavation.

Ancient mitochondrial DNA analyses on Sanganji Jomon people are described in Chapter 2 of this Thesis, and this content was already published by Kanzawa-Kiriyama et al. (2013).

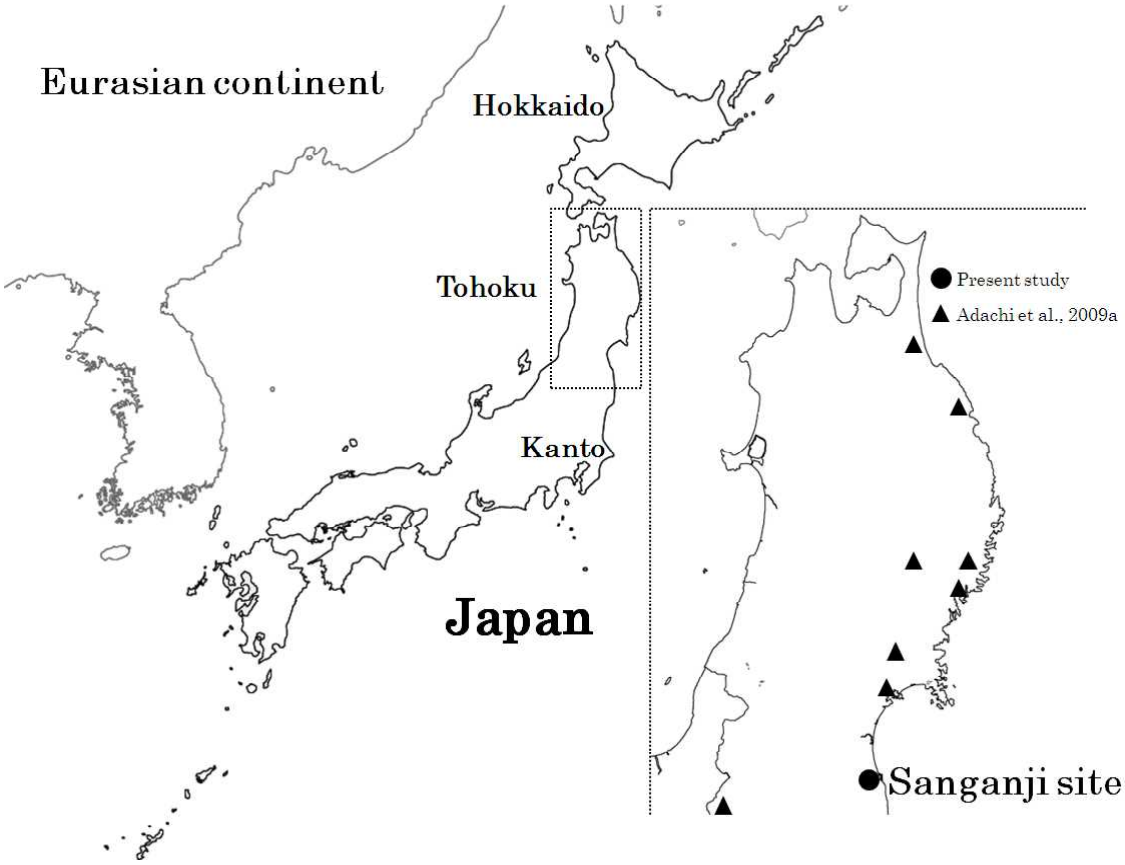


Figure 1.4: **Geographical location of the Sanganji site and other sites of Tohoku Jomon people**

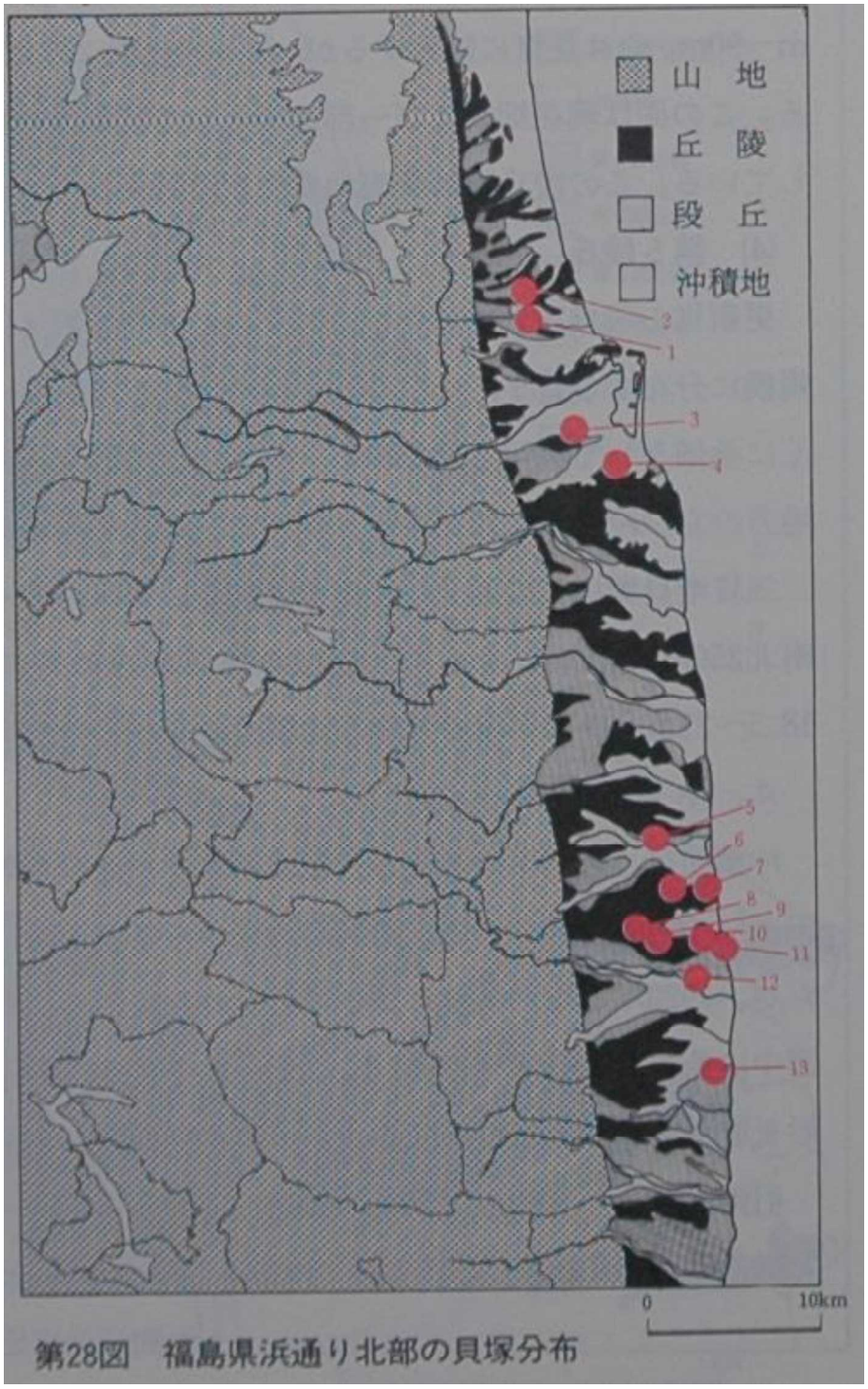


Figure 1.5: Geographical location of the Sanganji Shell Mound and other Shell Mounds located in coastal aria of Fukushima prefecture (from Fukushima Museum, 1988)

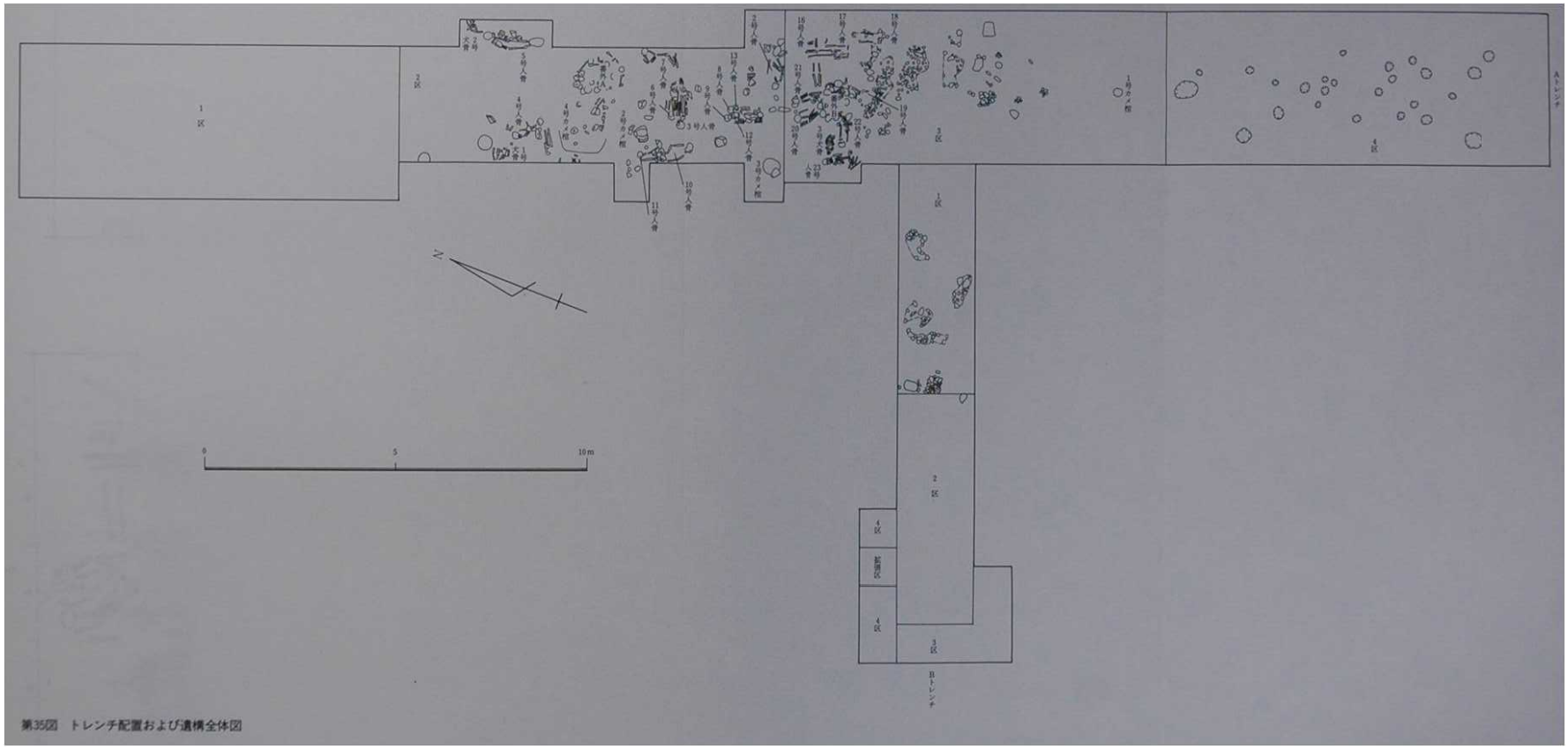


Figure 1.6: Excavation trench of Sanganji Shell Mound in 1954 (from Fukushima Museum, 1988**)**

Chapter 2

Mitochondrial DNA analysis of four Sanganji Jomon individuals

2.1 Introduction

Archeological, morphological, and genetic methods have been used to elucidate the history of human migration for a long time, and mitochondrial DNA (mtDNA) data of modern humans have also been used to elucidate our history from 1980s. However, those genetic data of modern humans are the summation of all ancient migrations. Therefore, genetic origin of Jomon people, and genetic relationship between Jomon people and modern East Eurasians, is largely unknown. To directly investigate the origin of the Jomon from a genetic standpoint, in the last few decades, mtDNAs from ancient skeletal remains have been analyzed (Horai et al., 1989, 1991; Shinoda and Kanai, 1999; Shinoda, 2003; Adachi et al., 2008, 2009a, 2009b, 2011). Analyses of ancient materials are useful, because they directly clarify maternally inherited mtDNA into haplogroups. So far, Jomon materials from three regions, the Hokkaido, Tohoku, and Kanto regions (Figure 2.1) have been analyzed. Comparisons of mtDNA haplogroup frequencies with other East and Northeast Asian populations indicated that the genetic structure of the Hokkaido Jomon (the northernmost of the Jomon populations) exhibited some similarities with indigenous modern Siberian populations inhabiting the lower Amur riverine system, particularly the Ulchi and Udegey people (Adachi et al., 2011). However, the genetic affinity of the other Jomon people, especially of the Tohoku Jomon was not sufficiently classified in Adachi et al. (2009a). Furthermore, the mtDNA haplogroups, N9b and M7a, characteristic to the Tohoku Jomon are scarce in the other modern populations except the modern Japanese (including Ainu and Ryukyuan people) and Udegey. In order to further understand the origin of the Jomon people based on ancient mitochondrial DNA, additional analyses of the Tohoku Jomon is essential.

Figure 2.1: A map of East Eurasia and geographic locations of the East Asian and Siberian populations compared in the present study

2.2 Materials and Methods

2.2.1 Sample collection

The Sanganji shell mound is located in the northern part of Fukushima Prefecture, Tohoku region (Figure 1.4). The chronological age of the burials, belonging to the late to final Jomon period, is considered to lie within the range from 4,000 to 2,500 years BP. To analyze mtDNA haplogroups, I used teeth from four Sanganji Jomon individuals housed in The University Museum, The University of Tokyo (Table 2.1). Three of these came from skeletal remains catalogued as UMUT-131421 (Endo and Endo, 1979), which includes six to eight individuals from what is known as secondary burial C′ (Fukushima Museum, 1988). In the following analysis, I designate these three individuals as 131421-1, 131421-2, and 131421-3. The fourth individual that I examined, UMUT-131464 (Endo and Endo, 1979), comes from secondary burial B, which includes approximately 10 individuals (Fukushima Museum, 1988). A single, well-preserved molar tooth was removed from the mandible of each individual (isolated teeth were excluded to avoid the possibility of sampling from the same individual). Mandibles which had both right and left side molars were chosen, so that the antimere of the analyzed tooth would remain available for general morphological studies. Ablation patterns of the Jomon people have been discussed in relation with population movement (Harunari, 2002; Funahashi, 2010). Tooth ablation type of individuals 131421-1 and 131421-2 (no mandibular teeth extracted) was probably the 0-type (only the upper canine extracted, no mandibular tooth extraction). According to Harunari's (2002) interpretations, the ablation type would indicate in-group status of the individuals and not immigration through marriage. If so, these two individuals may have been born at Sanganji, and it would be possible that they are relatives.

Table 2.1: Information of analyzed Sanganji Jomon teeth samples

Sample ID	Sex	Tooth ablation type	Date of excavation	Analyzed tooth part
131421-1	Female?	0 [1]	1954	Right M3
131421-2	Male	0 [1]	1954	Right M3
131421-3	Male	? [2]	1954	Right M2
131464	Female	2C [3]	1952	Right M2

1 There are no tooth extractions in the mandibular anterior teeth (all alveoli present), inferring ablation of only the upper canines (or a rare case at Sanganji of no ablation).
2 The left I1 alveolus is closed after tooth loss, while the left I2 is intact. The right side mandible is not preserved. This suggests either individual tooth ablation interpreted as mourning ablation (Harunari, 2002), a rare case at Sanganji of transitional stage of the 4I pattern (all four lower incisors extracted), or type 0 with natural loss of the left I1.
3 All lower anterior teeth (canine and incisors) were extracted, implying the 2C ablation type (all lower anterior teeth and upper canines).

2.2.2 DNA extraction

Tooth samples were soaked in a 13% sodium hypochlorite solution for 3min, rinsed several times with HPLC-grade water (Wako, Japan), then treated with ultraviolet radiation at least overnight, and allowed to air-dry with ultraviolet radiation for 60 min. Moreover, using a heat-treated drill, the outer surface of the samples was removed. Next, the samples were again rinsed with WPLC-grade water and allowed to be air-dried under ultraviolet radiation for 60 min. The tip of the tooth root was cut horizontally using a cutting disk; the dental pulp was reduced to powder by using a drill and collected into a tube.

DNA extraction from the powdered samples was carried out based on the method of Rohland and Hofreiter (2007) with some modifications. The powdered samples were dissolved with 0.5 M EDTA (pH 8.0) (Wako, Japan) and 0.25 mg/ml proteinaseK (TaKaRa, Japan) overnight at room temperature. The dissolved solution was centrifuged, and the supernatant was collected. The collected solution was concentrated using VivaSpin6 (Sartorius, Japan) to a volume less than 300 µl. The concentrated solution was added to 1.5 ml binding buffer (5 M GuSCN, 25 mM NaCl and 50 mM Tris (pH 8.0)) plus 100 µl silica suspension using silica (Sigma) and incubated under agitation for three hours at room temperature. The silica pellet was collected via centrifugation, the supernatant was removed, and the pellet was washed twice with binding buffer. Subsequently, the pellet was dried at room temperature for 15 min. DNA was eluted with 50 µl 1xTE buffer at room temperature for 10 min, and after centrifugation, the aqueous solution was transferred to a new tube. DNA was eluted twice, and 100 µl of DNA extract was finally obtained from each sample.

2.2.3 PCR amplification and sequencing

Ancient DNA is invariably of shorter length, and the length of the DNA sequences that can be amplified by PCR is limited (Pääbo et al., 2004; Haak et al., 2010). Moreover, as heavy handling and storage for many years influence the ability to obtain authentic endogenic DNA (Melchior et al., 2010), I anticipated that the analysis of Sanganji Jomon samples, excavated in

1952 and 1954, would be difficult and the success rate would be low. To overcome this problem, in addition to the primers that target long DNA segments (158-195 bp), I used short lengths of target DNA (58-90 bp) containing haplogroup specific SNPs, because shorter DNA regions are more readily amplified by PCR than longer DNA regions (Pääbo et al., 2004; Haak et al., 2010).

To determine mtDNA haplogroups, hypervariable region (HVR) 1, HVR-2, 28 haplogroup/sub-haplogroup specific single nucleotide polymorphisms (SNPs), and a 9-bp repeat variation in the noncoding cytochrome oxidase II/tRNA ^{Lys} intergenic region, were analyzed following previous publications (Umetsu et al., 2005; Van Oven and Kayser, 2008; Adachi et al., 2009a, 2009b, 2011; Haak et al., 2010). Primers used to amplify HVR-1, HVR-2, and haplogroup-determining regions are listed in Tables 2.2 and 2.3. Analyzed haplogroups were compared with those of modern Japanese and Jomon people (Horai et al., 1996; Shinoda and Kanai, 1999; Shinoda, 2003; Maruyama et al., 2003; Tajima et al., 2004; Tanaka et al., 2004; Adachi et al., 2008, 2009a, 2009b, 2011; Matsukusa et al., 2010). To further characterize the sub-haplogroups within N9 and M7a, I used additional primers listed in Table 2.4.

A total of 1 µl of the extracted DNA was used as the template for PCR. Amplifications were carried out in a total reaction volume of 20 µl containing of 0.5 µM each primer, and the Multiplex PCR assay kit (TaKaRa) reagents. The PCR conditions were incubation at 94°C for 30 sec, followed by 42 cycles of 94°C for 30 sec, 54°C for 20 sec, 72°C for 15 sec, and 1cycle of 72°C for 10 min. An aliquot (2.5 µl) of the PCR product was checked on a 2% agarose gel, and remaining (17.5 µl) was then purified with Monofas (GL Sciences). The PCR products of SNPs analysis were used as template for direct nucleotide sequencing, but I performed cloning of the PCR products of HVR1 and HVR2 because of these small quantities. Sequencing reactions were performed with BigDye Terminators v1.1 Cycle Sequencing kit (Applied Biosystems, Foster City, CA), and -21M13 and M13RVN primers were used in direct sequencing. All sequencing reactions were carried out using 3130 DNA Sequencer (Applied Biosystems), and the sequence of each region was compared with the revised CRS (Cambridge reference sequence) (Andrews et al., 1999).

Table 2.2: Primers used for PCR amplifications of mitochondrial DNA hypervariable regions 1 and 2

Primer		Nucleotide positions [2]	Product size
ID	Sequence [1]		
L127	AGCACCCTATGTCGCAGTAT	128-256	169
H257	TCTGTGTGGAAAGCGGCTGT		
L15998	CCATTAGCACCCAAAGCTA	15999-16141	182
H16142	ATGTACTACAGGTGGTCAAG		
L16120	TTACTGCCAGCCACCATGAA	16121-16238	158
H16239	TGGCTTTGGAGTTGCAGTTG		
L16208	CCCCATGCTTACAAGCAAG	16209-16366	195
H16367	CTGAGGGGGGTCATCCAT		

[1] Primer of Adachi et al. (2009b) were referred.

[2] Numbers of nucleotide positions are relative to the revised Cambridge reference sequence (rCRS) (Andrews et al., 1999).

Table 2.3: Primers used for PCR amplifications of mitochondrial DNA haplogroup specific SNPs

Haplogroup	Haplogroup-defining mutations	Primer ID	Primer Sequence (5'-3')	Nucleotide positions [1]	Product size
M/D5	10400/10397	M13-L10382 M13R-H10421	(-21M13)[2]AAGTCTGGCCTATGAGTGACTACAA (M13RVN)[3]TGAGTCGAAATCATTCGTTTTG	10383-10420	85
N	10873	M13-L10870 M13R-H10888	(-21M13)CCACAGCCTAATTATTAGCATCATC (M13RVN)GCTAAATAGGTTGTTGTTGATTTGG	10871-10887	67
A/G	4824/4833	M13-L4812 M13R-H4842	(-21M13)TAGCCCCCTTTCACTTCTGA (M13RVN)AAGAAGCAGGCCGGATGT	4813-4841	68
B	8281-8289 deletion	M13-L8268 M13R-H8295	(-21M13)AATAGGGCCCGTATTTACCCTATA (M13RVN)AGGTTAATGCTAAGTTAGCTTTACAGTG	8269-8294	78
D	5178	M13-L5171 M13R-H5204	(-21M13)ACCCTACTACTATCTCGCACCTGA (M13RVN)CTAGGGAGAGGAGGGTGGAT	5172-5203	76
D4/E	3010/3027	M13-L3005 M13R-H3029	(-21M13)CTCGATGTTGGATCAGGACA (M13RVN)TCGTTGAACAAACGAACCTT	3006-3028	60
F	10310	M13-L10298 M13R-H10326	(-21M13)CCCTACCATGAGCCCTACAA (M13RVN)GGATGATGATTAATAAGAGGGATGA	10299-10325	72
M7	9824	M13-L9816 M13R-H9851	(-21M13)TTTTGTAGCCACAGGCTTCC (M13RVN)GGCGGATGAAGCAGATAGTG	9817-9850	74
M7a	2772	M13-L2742 M13R-H2778	(-21M13)GAGAAGACCCTATGGAGCTTTAAT (M13RVN)TAATGCAGGTTTGGTAGTTTAGGA	2743-2777	83
M7c	5442	M13-L5434 M13R-H5471	(-21M13)AAATGACAGTTTGAACATACAAAACC (M13RVN)AAAGGGGAGATAGGTAGGAGTAGC	5435-5470	86
M8	15487T	M13-L15450 M13R-H15489	(-21M13)GCCCTCGGCTTACTTCTCTT (M13RVN)CTGGGTCGCCTAGGAGGT	15451-15488	76
M9	3394	M13-L3376 M13R-H3397	M13-GGCATTCCTAATGCTTACCG (M13RVN)GGGCCTTTGCGTAGTTGTAT	3377-3396	60
M10	15071	M13-L15057 M13R-H15077	(-21M13)ATCGGGCGAGGCCTATATTA (M13RVN)ATGCCGATGTTTCAGGTTTC	15058-15076	59
M12/M7b	4170/4164	M13-L4151 M13R-H4177	(-21M13)GAACAGCATACCCCCGATT (M13RVN)TGCTAGGGTGAGTGGTAGGAA	4152-4176	77
N9	5417	M13-L5385 M13R-H5422	(-21M13)TCCACCTCAATCACACTACTCC (M13RVN)TGGGGTGGGTTTTGTATGTT	5386-5421	78

[1] Numbers of nucleotide positions are relative to the revised Cambridge reference sequence (rCRS) (Andrews et al., 1999).
[2] (-21M13) sequence represents 5'-TGTAAAACGACGGCCAGT-3'.
[3] (M13RVN) sequence represents 5'-TGTGGAATTGTGAGCGG-3'.

Table 2.4: Primers used for PCR amplifications of mitochondrial DNA sub-haplogroup specific SNPs

Haplogroup	Haplogroup-defining mutations	Primer ID	Primer Sequence (5'-3')	Nucleotide positions[3]	Product size
M7a1	14364	M13-L14357	(-21M13)[1]CCACCCCATCATACTCTTTCA	14358-14398	83
		M13R-H14399	(M13RVN)[2]GGTTGAGGTCTTGGTGAGTGT		
	16324	M13-L16302	(-21M13)CAAACCTACCCACCCTTAACA	16303-16345	84
		M13R-H16346	(M13RVN)GGGACGAGAAGGGATTTGAC		
M7a2	15422	M13-L15401	(-21M13)CCTCCCATTCCGATAAAATCA	15402-15426	65
		M13R-H15427	(M13RVN)AAGTAAGCCGAGGGCGTCT		
	16140	M13-L16119	(-21M13)ATTACTGCCAGCCACCATGA	16120-16169	90
		M13R-H16170	(M13RVN)AGGGGGTTTTGATGTGGATT		
N9a	5231	M13-L5217	(-21M13)AATTCCATCCACCCTCCTCT	5218-5237	58
		M13R-H5238	(M13RVN)GGGCAAAAAGCCGGTTAG		
N9b	13183	M13-L13147	(-21M13)CCCCCTAGCAGAAAATAGCC	13148-13187	80
		M13R-H13188	(M13RVN)AGACTGCTGCGAACAGAGTG		
N9b1	12501	M13-L12492	(-21M13)TCAGTCTCTTCCCCACAACA	12493-12523	71
		M13R-H12524	(M13RVN)GGCTCAGTGTCAGTTCGAGA		
N9b2	16294	M13-L16292	(-21M13)CACTAGGATACCAACAAACCTACCC	16293-16326	84
		M13R-H16327	(M13RVN)TTTGACTGTAATGTGCTATGTACGG		
N9b3	14996	M13-L14965	(-21M13)CCCACATCACTCGAGACGTA	14966-14997	72
		M13R-H14998	(M13RVN)AGAATATTGAGGCGCCATTG		
Y	8392	M13-L8380	(-21M13)CAGTGAAATGCCCCAACTAAAT	8381-8394	59
		M13R-H8395	(M13RVN)GGAGTATGGGGGTAATTATGGTG		

[1] (-21M13) sequence represents 5'-TGTAAAACGACGGCCAGT-3'.

[2] (M13RVN) sequence represents 5'-TGTGGAATTGTGAGCGG-3'.

[3] Numbers of nucleotide positions are relative to the revised Cambridge reference sequence (rCRS) (Andrews et al., 1999).

2.2.4 Contamination precautions

During each step of sample handling, the following precautions were taken to minimize the risk of contamination.

1) The equipment used (drill, pipettes, centrifuge, vortex machine and PCR rack) were treated with a DNA contamination removal solution (DNA-AWAY, Molecular BioProducts, San Diego, CA) and with ultraviolet irradiation.

2) Some heat-resistant implements (tube rack made of stainless steel, drill tips, cutting disk, and glass bottle) were treated with dry heat sterilization at 200°C for 12 hours.

3) Disposable equipments (tubes and filtered pipette tips) were irradiated with ultraviolet rays.

4) The working place was divided into pre-PCR and post-PCR areas.

5) Clean-bench was treated with overnight ultraviolet radiation.

6) Glove, face mask, and clean-room items (coats, caps, shoes and socks) were also used when ancient DNA samples were handled.

7) Clean-room items were treated with ultraviolet irradiation overnight.

8) To exclude any possible contamination with modern human DNAs, simultaneous negative controls during DNA extraction and PCR amplification steps were added.

2.2.5 Data analysis

To investigate the relationship between the Tohoku Jomon people and the other ancient and modern East Asians, I compared mtDNA haplogroup frequencies (Table 2.5) using Arlequin version 3.11 (Excoffier et al., 2005). For the analysis, I used four ancient populations and 14 modern East Eurasians (Figure 2.1). For pairwise population comparisons, the *F*st formula used in Arlequin is identical to the weighted average F-statistic over loci, θw, defined by Weir and Cockerham (1984) (Michalakis and Excoffier, 1996). Equation to calculate *F*st value is as follows:

$$\hat{\theta}_W = \frac{\sum_u a_u}{\sum_u (a_u + b_u + c_u)} \quad \textbf{(2.1)}$$

$$a_u = \frac{\bar{n}}{n_c}\left\{s^2 - \frac{1}{\bar{n}-1}\left[\bar{p}(1-\bar{p}) - \frac{r-1}{r}s^2 - \frac{1}{4}\bar{h}\right]\right\} \quad \textbf{(2.2)}$$

$$b_u = \frac{\bar{n}}{\bar{n}-1}\left[\bar{p}(1-\bar{p}) - \frac{r-1}{r}s^2 - \frac{2\bar{n}-1}{4\bar{n}}\bar{h}\right] \quad \textbf{(2.3)}$$

$$c_u = \frac{1}{2}\bar{h} \quad \textbf{(2.4)}$$

where: a_u, b_u, and c_u, observed components of variance (a_u for between populations; b_u for between individuals within populations; and c_u for between gametes within individuals); n bar, the average sample size; n_c, the squared coefficient of variation of sample sizes; p bar, the average sample frequency of alleles; s^2, the sample variance of allele frequencies over populations; h bar, the average heterozygote frequency for alleles. For further details, please see Weir and Cockerham (1984).

Three ancient populations, the Hokkaido (Adachi et al., 2011), Tohoku (present study; Adachi et al., 2009a), and Kanto Jomon (Shinoda and Kanai, 1999; Shinoda, 2003), the Okhotsk people (Sato et al., 2009), and the following 14 modern populations were compared: Ainu (Tajima et al., 2003), Mainland Japanese (Maruyama et al., 2003), Ryukyuan (Umetsu et al., 2005), Korean (Lee et al., 2006), Northern Chinese (Yao et al., 2002), Yangtze river region Chinese (Yao et al., 2002), Southern Chinese (Yao et al., 2002), Aboriginal Taiwanese (Trejaut et al., 2005), Northern Siberians (Starikovskaya et al., 1998), Kamchatkans (Schurr et al., 1999), Ulchi (Starikovskaya et al., 2005), Udegey (Starikovskaya et al., 2005), Nivkhi (Starikovskaya et al., 2005), and Negidal (Starikovskaya et al., 2005). Principle component analysis (PCA) scatterplots and phylogenetic networks were constructed with R (R Development Core Team, 2010) and splitstree4 (Huson and Bryant, 2006), respectively, based on pairwise *F*st values.

Table 2.5: Frequencies (%) of mitochondrial DNA haplogroups in ancient and modern East Eurasian populations

	Population																	
	Ancient East Asians				Modern East Asians								Siberian					
	Tohoku Jomon	Hokkaido Jomon	Kanto Jomon	Okhotsk	Ai	MJ	Ry	Ko	NC	YC	SC	TA	NS	Ka	Ul	Ud	Ni	Ne
n	23	54	54	37	51	211	326	591	151	162	119	640	145	202	87	46	56	33
M7a	34.8	7.4	3.7	5.4	15.7	10.0	23.3	1.5								19.6		
M7b			1.9		3.9	2.8	4.0	3.9	3.3	5.6	6.7	11.1						
M7c						0.5	0.6	2.7	2.6	0.6	2.5	7.8						
M7d															1.1			
M7*[1]										0.6	0.8				1.1			
N9a						1.9	0.3	7.3	4.6	1.9	2.5	1.6						
N9b	60.9	64.8	5.6	10.8	7.8	1.9	4.6	0.3							6.9	30.4		
Y				43.2	19.6	0.5	0.6	1.9	2.0	1.9		1.4		8.4	37.9	8.7	66.1	21.2
A			7.4	8.1	3.9	9.0	6.7	8.5	5.3	12.3	0.8		73.1	5.4				
B			9.3	2.7	2.0	10.9	14.1	15.7	13.9	17.3	26.1	32.0						12.1
C				5.4		0.9	0.3	2.7	3.3	6.2	1.7		6.2	31.2	13.8	17.4		15.2
D4[2]	4.3	16.7	18.5		13.7	36.0	36.2	25.9	23.2	19.8	7.6	1.3	16.6	1.0	17.2		5.4	6.1
D5					3.9	5.2	2.1	5.2	7.3	7.4	5.0	4.1						
D6											0.8							
D*[3]								1.2			0.8				4.6		23.2	18.2
E												12.0						
F			1.9		2.0	5.2	2.5	7.6	9.3	13.6	23.5	25.2			1.1			
G1		11.1		24.3	21.6	6.6	0.6	2.4					4.1	48.0	10.3		5.4	27.3
G*[4]			1.9		3.9	3.8	0.9	5.9	6.6	0.6	0.8	0.2			1.1			
M8a			9.3			0.5	0.6	1.2	6.0	2.5	1.7				4.6	15.2		
M9					2.0	0.9	1.8	1.7	3.3		0.8					8.7		
M10			33.3			1.4		1.5	3.3		2.5	0.6						
M11								0.7	0.7		0.8							
M12											1.7							
R						0.5		0.3	2.6	3.1	4.2	2.5						
Z						1.4		0.8	1.3	3.7				5.9				
Others[5]			7.4				0.6	1.0	1.3	3.1	8.4	0.3						

[1] A total of haplogroup MJ except for haplogroup MJa, MJb, MJc and MJd. [2] Haplogroup D4 includes haplogroup D1, D2 and D3. [3] A total of haplogroup D except for haplogroup D4, D5 and D6. [4] A total of haplogroup G except for haplogroup G1. [5] It is difficult to estimate specific haplogroup.

Tohoku Jomons (present study; Adachi et al., 2009a), Hokkaido Jomons (Adachi et al., 2011), Kanto Jomons (Shinoda and Kanai, 1999; Shinoda, 2003), Okhotsk people (Sato et al., 2009), Ainu (Ai) (Tajima et al.), Mainland Japanese (MJ) (Maruyama et al., 2003), Ryukyuan (Ry) (Umetsu et al., 2005), Korean (Ko) (Lee et al., 2006), Northern Chinese (NC), Yangtze river region Chinese (YC), Southern Chinese (SC) (Yao et al., 2002), Taiwanese Aboriginals (TA) (Trejaut et al., 2005), Northern Siberian (NC) (Starikovskaya et al., 1998), Kamchatkan (Schurr et al., 1999), Ulchi (Ul), Udegey (Ud), Nivkhi (Ni) and Negidal (Ne) (Starikovskaya et al., 2005).

Northern Chinese, Liaoning, Qingdao and Zibo; Yangtze river region Chinese, Wuhan and Shanghai; Southern Chinese, Gongdong, Guangdong, Guangzhou and Hong Kong; Northern Siberian, Chukchi and Eskimos; Kamchatkan, Koryak and Itel'men.

2.3 Results

2.3.1 mitochondrial DNA haplotyping

I successfully determined the mtDNA haplogroup of all four individuals that I examined (Table 2.6), using the primer pairs targeting short DNA segments (58-90 bp). Two individuals were classified into haplotype M7a2, and the others were classified into haplotype N9b. Both of the haplotypes were typical in Northern Jomon people, but not in modern Japanese. In each individual, all SNP sites were consistent with previously observed haplogroups, and this thus reduces the chance of contamination. When I used the primer pairs targeting long DNA segments (158-195 bp) in the analysis of HVR, amplification efficiency and quantity of PCR products decreased dramatically possibly due to the sample conditions. The result corresponds to the observations of Pääbo et al. (2004) and Haak et al. (2010). To estimate the level of modern DNA contamination, long PCR products were sequenced with cloning. In these analysis, some PCR products contained sequences that were inconsistent with the SNP analysis. This indicates that, to reduce the effect of contamination in the analysis of ancient DNA, focusing on short DNA regions is better, and I need to reduce total number of PCR cycles.

2.3.2 Principal component analysis and phylogenetic network

In a PCA plot, modern East Asian clustered, and Kanto Jomon was within the cluster (Figure 2.2). However, Northern Jomon as well as southern and northern Siberians was located in the outside of the cluster. This indicates the uniqueness of Northern Jomon people.

In phylogenetic network analysis, I observed that northern Jomon was apart from modern East Asians, while there genetic relationship between northern Jomon and Udegey from Southern Siberia (Figure 2.3).

Table 2.6: Observed nucleotide changes in four Sanganji Jomon individuals

Haplogroup	Nucleotide position[1]	Sample	Sanganji131421-1	Sanganji131421-2	Sanganji131421-3	Sanganji131464
		sex	f?	m	m	f
		Tooth ablation type	0	0	4I?	2C
		rCRS[2]				
M	10400	C	**T**	**T**	**C**	**C**
N	10873	T	**C**	**C**	**T**	**T**
A	4824	A	A	A	A	A
B	8281-8289 del	2	2	2	2	2
D	5178	C	C	C	C	C
D4	3010	G	ND	ND	ND	ND
D5	10397	A	A	A	A	A
E	3027	T	ND	ND	ND	ND
F	10310	G	G	G	G	G
G	4833	A	A	A	A	A
M7	9824	T	**C**	**C**	T	T
M7a	2772	C	**T**	**T**	-	-
M7a1	14364	G	G	G	-	-
	16324	T	T	T	-	-
M7a2	15422	A	**G**	**G**	-	-
	16140	T	**C**	**C**	-	-
M7b	4164	C	C	C	C	C
M7c	5442	T	ND	ND	-	-
M8	15487T	A	A	A	A	A
M9	3394	T	ND	ND	ND	ND
M10	15071	T	T	T	T	T
M12	4170	A	A	A	A	A
N9	5417	G	G	G	**A**	**A**
N9a	5231	G	-	-	G	G
N9b	13183	A	-	-	**G**	**G**
N9b1	12501	G	-	-	G	G
N9b2	16294	C	-	-	C	-
N9b3	14996	G	-	-	G	G
Y	8392	G	-	-	ND	ND
HVR	128-256		ND	**146**	ND	CRS
	15999-16141		ND	ND	ND	ND
	16121-16238		ND	ND	ND	ND
	16209-16366		ND	ND	ND	16256, **16294**
Haplogroup			**M7a2**	**M7a2**	**N9b***	**N9b2**

[1] Numbers of nucleotide positions are relative to the revised Cambridge reference sequence (rCRS) (Andrews et al., 1999).

[2] The sequences of rCRS

"ND" denotes "not determined", "-" denotes "not analyzed", and diagnostic polymorphisms are emphasized by bold type. Haplogroup B have 9bp (CCCCCTCTA) deletion. "1" denotes the presence of the 9bp deletion, and "2" denotes nondeletion (i.e., two repeats of the 9bp fragment).

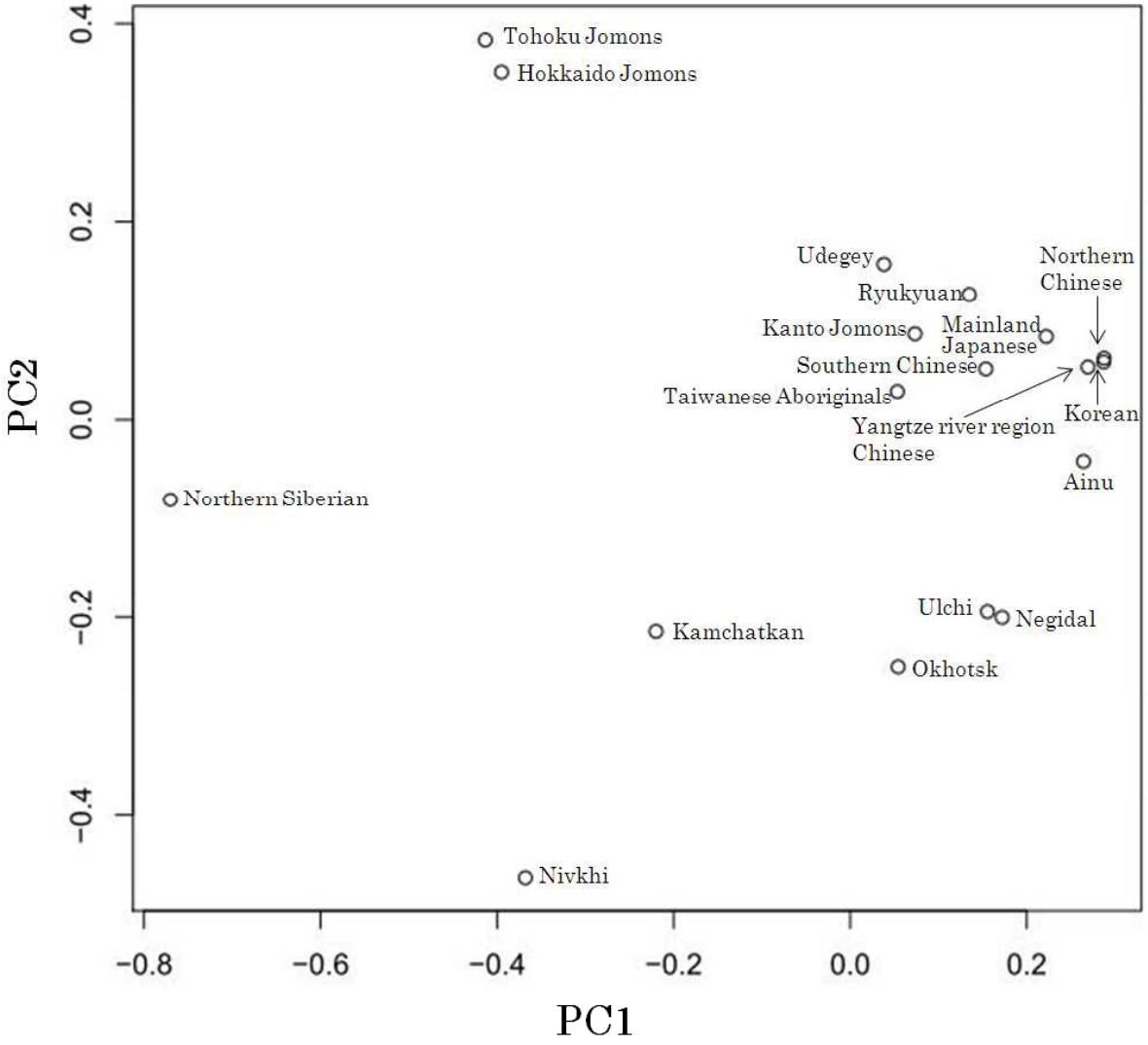


Figure 2.2. Principle Component Analysis of 18 East Asian populations based on *F*st values.

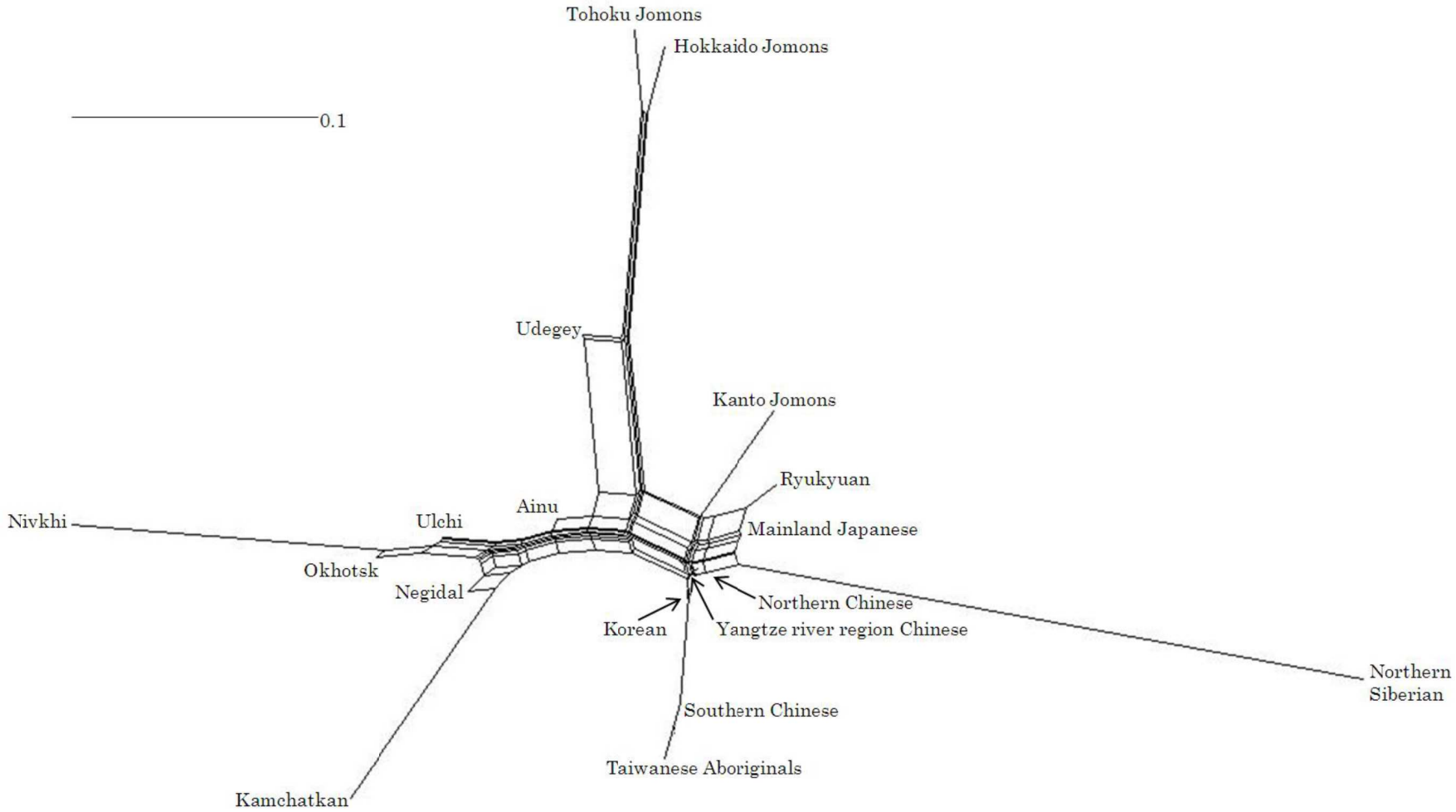


Figure 2.3. A phylogenetic network of 18 East Asian populations based on *F*st values

2.4 Discussion

Two of four Sanganji Jomon individuals belonged to haplogroup N9b (Tables 2.6 and 2.7). Haplogroup N9b has been observed in the northern Jomon populations at high frequencies (Hokkaido Jomon, 64.8% (Adachi et al., 2011), Tohoku Jomon, 63.2% (Adachi et al., 2009a)), and I found this haplogroup in high frequency (50.0%) in the Sanganji Jomon, although the sample size was small. On the contrary, the frequency of haplogroup N9b was low in the Kanto Jomon at 5.6% (Shinoda and Kanai, 1999, Shinoda, 2003). In modern populations, haplogroup N9b is present in the Japanese Archipelago at low frequencies (<10%) (Maruyama et al., 2003; Tajima et al., 2004; Umetsu et al., 2005), but at a high frequency (30.4%) in the Udegey from Southern Siberia. It seems that haplogroup N9b was one of the main haplogroups in ancient and modern Northeast Asian populations.

I further subdivided the haplogroups N9b observed in the two Sanganji Jomon individuals using specific primers into four sub-haplogroups (N9b1, N9b2, N9b3, and N9b*) (Table 2.4). Samples which could not be assigned to sub-haplogroups N9b1, N9b2, and N9b3 were classified as sub-haplogroup N9b*. Of the two Sanganji Jomon samples, one was classified as sub-haplogroup N9b2 and another was classified as sub-haplogroup N9b*.

Haplogroup M7a is classified into three sub-haplogroups M7a1, M7a2, and M7a*; individuals who were not classified into sub-haplogroups M7a1 and M7a2 were classified into sub-haplogroup M7a*. I found that two Sanganji individuals belonged to haplogroup M7a2 (Tables 2.6 and 2.8). In modern populations, haplogroup M7a1 was observed in modern Japanese populations at a high frequency: Ainu, 15.7% (Tajima et al., 2004); mainland Japanese, 9.5% (Maruyama et al., 2003); Ryukyuan, 26.6% (Horai et al., 1996, Matsukusa et al., 2010). However, haplogroup M7a2 was scarcely observed in the mainland Japanese (0.5%), and this haplogroup was not observed in the Ryukyuan. On the contrary, haplogroup M7a2 was observed in the Udegey from southern Siberia at high frequency (19.6%, Starikovskaya et al., 2005).

Table 2.7: Frequencies of N9b sub-haplogroups (%) in ancient and modern East Eurasian populations

	Populations										
	Ancient East Asians					Southern Siberian		Modern East Asians			
	Sangarji Jomon	Tohoku Jomon	Hokkaido Jomon	Kanto Jomon[1]	Okhotsk	Ulchi	Udegey	Ainu[1]	Mainland Japanese	Ryukyuan	Korean
References	This study	Adachi et al., 2009a	Adachi et al., 2011	Shinoda and Kanai, 1999; Shinoda, 2003	Sato et al., 2009	Starikovskaya et al., 2005	Starikovskaya et al., 2005	Tajima et al., 2004	Tanaka et al., 2004	Umetsu et al., 2005	Lee et al., 2006
n	4	19	54	54	37	87	46	51	672	326	591
N9b1			**55.6**						**1.3**		
N9b2	**25.0**	(63.2)[2]		(5.6)[2]	(10.8)[2]	(6.9)[2]	(30.4)[2]	(7.8)[2]	0.6	(4.6)[2]	(0.3)[2]
N9b3									0.3		
N9b*	**25.0**		9.3						0.4		

[1] For these populations, haplogroups were inferred from the sequences of hypervariable regions (Adachi et al., 2009a, 2009b).
The haplogroups observed at highest frequency in each population were emphasized by bold type.
[2] The mitochondrial DNA sequence data of these populations were insufficient to classify them into each sub-haplogroup.

Table 2.8: Frequencies (%) of M7a sub-haplogroups in ancient and modern East Eurasian populations

	Populations									
	Ancient East Asians						Modern populations			
	Sanganji Jomon	Tohoku Jomon	Hokkaido Jomon	Kanto Jomon[1]	Okhotsk	Udegey	Ainu[1]	Mainland Japanese	Ryukyuan[1]	Korean
References	This study	Adachi et al., 2009a	Adachi et al., 2011	Shinoda and Kanai, 1999; Shinoda, 2003	Sato et al., 2009	Starikovskaya et al., 2005	Tajima et al., 2004	Maruyama et al., 2003	Horai et al., 1996; Matsukusa et al., 2010	Lee et al., 2006
n	4	19	54	54	37	46	51	211	274	591
M7a1				**3.7**			**15.7**	**9.5**	**26.6**	**1.3**
M7a2	**50.0**	(31.6) [2]	1.9		**5.4**	**19.6**		0.5		
M7a* [3]			**5.6**						1.1	0.2

[1] For these populations, haplogroups were inferred from the sequences of hypervariable regions (Adachi et al., 2009a, 2009b).

[2] The mitochondrial DNA sequence data of Tohoku Jomon people are insufficient to classify them into each sub-haplogroups, and the frequency of Tohoku Jomon people is the total of haplogroup M7a1, M7a2, and M7a*.

[3] Individuals that are not classified into Sub-haplogroups denote M7a*.

The haplogroups observed at highest frequency in each population were emphasized by bold type.

In ancient East Asian populations, haplogroup M7a (which includes M7a1, M7a2, and M7a*) is widely observed in the Jomon people (Hokkaido Jomon, 7.4%; Tohoku Jomon, 31.6%; Kanto Jomon, 3.7%). That the Sanganji Jomon (present study) also had this haplogroup apparently at high frequency (50.0%) supports the suggestion of Adachi et al. (2009b) that haplogroup M7a is a putative ‘‘Jomon genotype’’. However, I have now identified occurrence of sub-haplogroup M7a2 for the first time in the Tohoku Jomon, which enables some further considerations.

Although the currently available data is still limited, interestingly, the frequencies of haplogroups M7a1 and M7a2 appear quite different among the Jomon people (Table 2.8). The finding may indicate that during the later Jomon period, M7a1 was mainly distributed in the middle to southern part of the Japanese Archipelago, and haplogroup M7a2 (and also M7a*) was distributed further north, from the Tohoku and Hokkaido regions of the Japanese Archipelago and southern Siberia. Adachi et al. (2011) suggested that haplogroup M7a was of southern origin, and that this haplogroup was introduced to Japan around the last glacial maximum (LGM). If I accept this hypothesis, haplogroup M7a2, which is observed in the northern Jomon people, might have originated after introduction of M7a into the Japanese Archipelago. The sub-haplotype subsequently became distributed further northward by migration, and must have reached southern Siberia.

Tooth ablation types of both Sanganji individuals 13142-1 and 13142-2 were probably the 0-type (only the upper canine extracted, no mandibular tooth extraction), which suggests an in-group status of the individuals according to Harunari’s (2002) interpretation of ablation patterns. On the other hand, individual 13142-3 did not exhibit a clear pattern and 131464 was the 2C type, which suggests an out-group status. Therefore, individuals 13142-1 and 13142-2, both possibly born at Sanganji and with the same haplotype, could have been relatives. However, since sub-haplogroup M7a2 might have been the major type in that region, it is also possible that they shared the same haplotype without genealogical relationship.

The genetic transition from the Jomon people to modern Japanese is explained by “the dual structure model”, where substantial number of Yayoi people migrated into the Japanese

Archipelago via the Korean Peninsula about 3,000 years ago, and admixed with Jomon people (Hanihara, 1991). This model also suggests that the effect of the Yayoi immigrants is small in the Ainu and Ryukyuan peoples, who live at the margins of the Japanese Archipelago and retain more Jomon components than the mainland Japanese. Therefore, high frequencies of the "Jomon genotype" (haplogroup N9b and M7a) in the Ainu and Ryukyuan peoples have been considered to partially support this model (Adachi et al., 2009b). Then, it is expected that the Ainu people who live in Hokkaido, the northernmost Japanese Archipelago, would have haplogroups M7a* and M7a2 in high frequencies, because the northern Jomon populations have been so far shown to have these haplogroups. However, the Ainu people have only haplogroup M7a1 (Table 2.8). This may indicate either sampling bias, or that the haplogroup M7a of the Ainu was not derived from the indigenous northern Jomon genotype, and that this haplogroup was introduced into the Ainu population after the Jomon period. The latter interpretation implies that the genetic effect from mainland Japan after the Jomon period was stronger than previously considered. Jinam et al. (2012) showed that modern Ainu people recently admixed with mainland Japanese. The suggestion raises the possibility that the haplotype M7a1 was introduced into Ainu people within several generations.

To investigate the relationships between the Jomon people from the Tohoku region (Sanganji Jomon, present study; the Tohoku Jomon, Adachi et al., 2011) and other East and Northeast Asian ancient and modern populations, statistical analysis based on haplogroup frequencies (Table 2.5) was performed. In this analysis, I merged sub-haplogroups N9b1, N9b2, N9b3, and N9b* into haplogroup N9b, and sub-haplogroups M7a1, M7a2, and M7a* into haplogroup M7a, because sub-haplogroup frequency data were missing in some populations used in the statistical analysis (Tables 2.7 and 2.8). The approximate geographical locations of these populations are shown in Figure 2.1 and Figure 2.2.

Within the Japanese Archipelago, contrasting opinions on the population structure of the Jomon have been proposed; the Jomon people were relatively homogeneous (Dodo, 1982; Yamaguchi, 1982; Kondo, 1994; Ossenberg et al., 2006; Matsumura, 2007) or heterogeneous (Adachi et al., 2009a, 2009b; Hanihara and Ishida, 2009; Nakashima et al., 2010). When the

ancient populations of the Japanese Archipelago (Kanto, Tohoku, and Hokkaido Jomons and Okhotsk) were compared, the population differentiation test demonstrated that they were statistically different from each other (highest p-value was 0.007 between the Tohoku and Hokkaido Jomon). The difference indicates inter-regional heterogeneity of the ancient Japanese Archipelago populations, consistent with results of recent genetic (Adachi et al., 2009a, 2009b) and morphological (Hanihara and Ishida, 2009; Nakashima et al., 2010) research.

In order to obtain, for the first time, an indication of overall genetic similarities of the Jomon regional populations as seen from the mtDNA haplotypes, I computed *F*st values between population pairs. The *p*-values between the Tohoku Jomon and the other ancient or modern comparative populations were less than 5%, based on 100 permutation of the Arlequin software. The results showed a close genetic similarity between the Tohoku and Hokkaido Jomon people (*F*st = 0.061, *P* =0.036±0.015) (Table 2.9), and supports previous interpretations that the Jomon ancestors of the northern part of Japan (Hokkaido) might have expanded southward to Honshu Island with a series of bottlenecks (Hanihara and Ishida, 2009; Nakashima et al., 2010). However, genetic similarity between the Tohoku Jomon and the geographically adjacent Kanto Jomon was not observed (*F*st = 0.268); the differences are shown in the PCA plot and the phylogenetic network (Figures 2.2 and 2.3). These results are an indication of genetic differences between the Kanto and northern Jomon populations.

Table 2.9: Pairwise *F*st values between each pair of populations

	n[1]	TJ	HJ	KJ	Ok	Ai	Ul	Ud	Ni	Ne	Ka	NS	MJ	Ry	Ko	NC	YC	SC
Tohoku Jomon (TJ)	23																	
Hokkaido Jomon (HJ)	54	0.061																
Kanto Jomon (KJ)	50	0.268	0.267															
Okhotsk (Ok)	37	0.292	0.287	0.203														
Ainu (Ai)	51	0.194	0.204	0.115	0.034													
Ulchi (Ul)	87	0.276	0.258	0.159	0.028	0.044												
Udegey (Ud)	46	0.088	0.138	0.149	0.137	0.086	0.118											
Nivkhi (Ni)	56	0.479	0.463	0.333	0.105	0.187	0.091	0.299										
Negidal (Ne)	33	0.308	0.295	0.157	0.049	0.039	0.045	0.137	0.171									
Kamchatkan (Ka)	202	0.382	0.349	0.271	0.152	0.135	0.171	0.224	0.344	0.079								
Northern Siberian (NS)	145	0.533	0.507	0.350	0.404	0.343	0.376	0.409	0.530	0.401	0.395							
Mainland Japanese (MJ)	211	0.239	0.224	0.082	0.174	0.060	0.120	0.150	0.282	0.120	0.219	0.262						
Ryukyuan (Ry)	324	0.216	0.231	0.107	0.205	0.079	0.151	0.146	0.304	0.160	0.264	0.292	0.013					
Korean (Ko)	585	0.238	0.218	0.072	0.153	0.066	0.105	0.136	0.243	0.096	0.196	0.228	0.012	0.040				
Northern Chinese (NC)	149	0.241	0.229	0.061	0.153	0.066	0.100	0.122	0.252	0.096	0.210	0.275	0.019	0.050	0.000			
Yangtze river region Chinese (YC)	157	0.251	0.241	0.082	0.156	0.078	0.113	0.134	0.262	0.100	0.207	0.241	0.029	0.060	0.008	0.005		
Southern Chinese (SC)	109	0.280	0.283	0.116	0.193	0.119	0.164	0.165	0.302	0.126	0.251	0.359	0.090	0.117	0.047	0.038	0.026	
Taiwanese Aboriginals (TA)	638	0.295	0.294	0.158	0.209	0.150	0.194	0.193	0.300	0.152	0.260	0.342	0.136	0.156	0.085	0.079	0.063	0.012

[1] The individuals that were not classified into specific haplogroups ("others" in Table 5) were removed in population comparison analysis.

*F*st p-value of all comparisons except for Ko-NC and NC-YC are less than 0.05.

Adachi et al. (2009a) mentioned that three regional Jomon populations shared some haplogroups (M7a and N9b), and that genetic similarity decreased gradually with increased geographical distance. But since sub-haplogroups M7a2 and N9b2 have so far not been shown to be shared between the Tohoku and Kanto Jomon people, it may be that, in terms of the maternal lineage, gene flow between geographically close Tohoku and Kanto region were limited in the Jomon period. In addition, in the population comparison analysis, although I found genetic similarity between the Hokkaido and Tohoku Jomon, so far they lack shared haplogroups at the sub-haplogroup level (M7a*, D4h2, and G1b of the Hokkaido Jomon were not seen in the Tohoku Jomon, and N9b2 and D4b of the Tohoku Jomon was not seen in the Hokkaido Jomon). Again, this may be indicating relatively limited gene flow in the Jomon period. This interpretation is consistent with the results of the population differentiation test, and the observation of inter-regional heterogeneity among the ancient Japanese Archipelago populations. However, the above interpretations need to be confirmed and refined by larger samples of sub-haplogroup determinations and better temporal control of the Jomon materials.

Compared to modern East Asian populations, the Tohoku and Hokkaido Jomon people were genetically close to the Udegey of southern Siberia (Udegey and Tohoku Jomon, *F*st = 0.088, *P* =0.01; Udegey and Hokkaido Jomon, *F*st = 0.138, *P* =0.00) (Table 2.9). The Udegey is also geographically closer to the Jomon populations than are the other southern Siberian populations (Figure 2.1). In the phylogenetic network shown in Figure 2.3, based on shared mtDNA haplogroups M7a and N9b (Table 2.5), it seems possible that the Udegey represents admixture of southern Siberian populations and the northern (Hokkaido and Tohoku) Jomon people. This implies some degree of gene flow between the Udegey people ancestors and the northern Jomon. Moreover, as I mentioned earlier, haplogroup N9b and M7a2 are hardly observed in the other East Asian populations except in the Japanese Archipelago (Table 2.7 and 2.8). One interpretation would be that a northern population with haplogroups N9b and M7a2 migrated into the Tohoku region via Hokkaido, although a southern origin has been considered for the M7a haplotype (see above for discussion of the southern haplogroup hypothesis). The interpretation would be compatible with the conclusion of previous studies that the Jomon people

were (largely) of northern origin (Nei, 1995; Omoto and Saitou, 1997; Hanihara and Ishida, 2009; Nakashima et al., 2010; Adachi et al., 2011).

Chapter 3

Nuclear DNA analysis of three Sanganji Jomon individuals

3.1 Introduction

In the previous chapter, I discussed the origin and history of Jomon people based on Jomon mtDNA haplotypes. Clarifying their origin is highly important to understand the history of not only modern Japanese but also modern East Eurasians. Previous studies and my mtDNA haplotyping of Sanganji Jomons show the genetic similarity between Udegey people and Northern Jomon people, and it implies their Northern origin (Adachi et al., 2010; Chapter 2 of this Thesis). In addition, their Jomon genotypes hint the uniqueness and the complexity on the history of Jomon people. However, the mtDNA is inherited in maternal lineages, and is considered just a single locus. Therefore, its genetic information is limited even if its complete genomic sequence is obtained. Furthermore, the Jomon genotypes, haplotype M7a and N9b, rarely observed in modern continental East Eurasians, which made it difficult to understand more detail of the Jomon origin and the genetic relationship between Jomon people and modern East Eurasians. Therefore, the analysis of the different genetic markers, namely the markers existing on nuclear DNA, is essential.

Next generation sequencing technologies makes us possible to analyze tiny amounts of ancient DNA (e.g., Miller et al., 2008; Green et al., 2010; Rasmussen et al., 2010; Reich et al., 2011), and it is quickly becoming practicable to analyze not only Jomon mtDNA but also nuclear DNA, which contain much more genetic information than mtDNA. Since the success rate of mitochondrial DNA haplotyping of four Sanganji Jomon samples was 100% (Table 2.6), skeletal remain materials from Sangnaji shell mound may also contain nuclear DNA. With this prospect of an expanded range of materials suitable for ancient nuclear DNA analysis with the emergence of next generation sequencing, further studies on the Sanganji Jomons may enable a better resolution to the origin of Jomon people.

3.2 Material and Methods

3.2.1 DNA extraction, library preparation, and sequencing

DNA extraction from four molar teeth was carried out in a previous study at National Institute of Genetics, Mishima. For DNA library preparation, I chose the DNA extract of Sanganji 131421-3 because the remnant of the DNA solution was sufficient for next generation sequencing. All the workflow of library preparation was carried out in a clean room at room G108 of National Institute of Genetics. Sixteen microliters of the remaining fraction was used for library preparation. Libraries were built from the extracted DNA solution using the GS Titanium Rapid Library Preparation kit (454 Life Science Corporation), with the following modifications to the protocol of Rasmussen et al. (2011). In the adapter ligation step, I used 1µl of Illumina adapter mix instead of 2 µl to minimize the amount of adapter dimer in following PCR step. The libraries were amplified in two rounds of PCR. First, a PCR was set up with Multiplex PCR kit (Qiagen) as follows: 5 µl DNA library, 25 µl Mix, 20-100 nM Multiplexing PCR primer 1.0, 20-100 nM Multiplexing PCR primer 2.0, and H_2O to 50 µl. Cycling conditions were 15 min at 95°C, 12 cycles of 30 sec at 95°C, 30 sec at 60°C, and 30 sec at 72°C, with a final extension at 72˚C for 10 min. PCR products were purified on a QiaQuick PCR purification kit and eluted in 50µl with TE buffer. All the steps of adapter ligation and first round PCR amplification were carried out in the ancient DNA laboratory of National Science Museum, Tsukuba through collaboration with Dr. Ken-ichi Shinoda. I used totally 20 to 30µl of purified PCR products for second round PCR amplification.

To minimize the contamination of PCR products into the experimental room, the second round PCR amplification and later experiments were carried out in room A202 of National Institute of Genetics, Mishima. A second round PCR was set up as follows: 5µl product from first PCR, 25µl Multiplex PCR mix, 500-2,000 nM Multiplexing PCR primer 1.0, 500-2,000 nM Multiplexing PCR primer 2.0, and H_2O to 50 µl, cycling conditions for the second PCR were, 95°C 15 min, 15 cycles of 95°C for 30 sec, 60°C for 30 sec, 72°C for 30 sec, with a final extension 72°C for 10 min. The PCR products were merged into one tube, and purified on a

QiaQuick PCR purification kit. The PCR products were finally eluted in 30-50 µl with TE buffer. The half of second round PCR products was treated with two restriction enzymes, FastDigest® Bsh1236I (Thermo Scientific) and FastDigest® TauI (Thermo Scientific), to enrich human derived sequence reads. Bsh1236I and TauI recognize CGCG and GCSGC sites, respectively. After the digestion, I denatured the enzymes with a phenol-chloroform treatment, and performed purification using a QiaQuick PCR purification kit. The remnants were amplified in third rounds of PCR. The PCR products were purified on a QiaQuick PCR purification kit and eluted in 30-50 µl with TE buffer. Amplification success was determined using Agilent 2100 Bioanalyzer DNA High-Sensitivity chip. Since the second and third PCR products included adapter dimer and other artificial short fragments, I purified those PCR products with 2% agarose gel (Certified low-range Ultra Agarose, BIO-RAD). DNA was separated for 30 min at 100 volts. A gel slice containing library molecules with inserts longer than 50 bp was excised to exclude short fragments and primer dimer, and the DNA libraries were extracted from the gel on a Qiagen Gel Extraction kit (Qiagen) and eluted to 30 µl volume with TE buffer. The libraries were quantified with the Agilent 2100 Bioanalyzer DNA High-Sensitivity chip.

Four lanes of Illumina GAIIx platforms were used, and 120-bp paired-end reads were generated for the library following the manufacturer's protocol. Sequencing was done at University of Tokai with the kind arrangement of Professor Ituro Inoue.

To obtain more sequence data from Jomon samples, I additionally prepared two libraries from two Sanganji Jomon individuals, Sanganji 131421-2 and 131464. I failed prepare library from Sanganji 131421-1 because the remnants of extracted DNA solution were limited. Furthermore, I newly extracted DNA from two teeth samples of Sanganji 131421-3 and 131464, kindly provided by Professor Gen Suwa at University of Tokyo Museum, to construct more DNA libraries. Before the library preparation, I typed mitochondrial DNA haplotype with APLP method (Adachi et al., 2011) to confirm that the DNA extracts include Jomon DNA. The later DNA extraction was carried out in National Science Museum, Tsukuba, to confirm whether the two sequence output of same individual extracted independently in two institutes were concordant with each other, and to weed out the possibility that the DNA derives from modern

human DNA contamination. For the DNA extraction, I used the protocol of Adachi et al. (2013) with some modifications (Adachi et al., 2013). During the preparation of sequencing library, I used Illumina index adapter, index 5, 6, 12, and 16, to reduce the effect of cross contamination of human DNA from the previous sequencing run. Additionally, since ~0.1% of sequence data is from previous sequencing run in Illlumina sequencing machines (Illumina, tracking number: KK-005001), to remove the cross contamination of human DNA, I regulated that the DNA libraries of previous sequencing run was non-human. Hiseq 2000 was used for sequencing, and 100-bp paired-end reads were generated for the library following the manufacturer's protocol. Since I sometimes prepared two libraries from the same Jomon individual, I need to characterize them to discriminate each library. I add (1) and (2) behind the sample ID for first and second library, respectively (e.g. Sanganji 131421-3(1) and Sanganji 131421-3(2), respectively). Sequencing was done by Hokkaido System Science Co., Ltd.

3.2.2 Carbon 14 dating

The dates of four samples were identified with carbon 14 dating. The dating was made by Professor Minoru Yoneda at University Museum, University of Tokyo. I listed the summary of dates in Table 3.1.

Table 3.1: Carbon 14 dating of four Sanganji Jomon individuals

Sample ID	Parts	Conventional ^{14}C age (BP ± 1σ)	δ 13 C (‰)
131421-1	Root of the tooth	3156 ± 19	-16.87 ± 0.20
131421-2	Root of the tooth	2923 ± 19	-19.41 ± 0.14
131421-3	Root of the tooth	2994 ± 19	-19.60 ± 0.15
131464	Root of the tooth	3061 ± 19	-17.64 ± 0.16

3.2.3 Mapping and damage estimates

Metagenomic analysis was performed using MEGABLAST homology search with the help of Dr. Kirill Kryukov of Saitou Laboratory at National Institute of Genetics, Mishima (Imanishi Laboratory of Tokai University Medical School from August 2013). A total of 4,000 sequence reads of the Sanganji 131421-3(1) were searched; since four lanes of GAIIx were used for sequencing, every 1,000 sequence reads were randomly chosen from each sequence lane.

Most of the sequence reads included adaptor sequence in the 3' end of the reads, because the original DNA fragments, which were caught with adapter, were shorter than 120-bp. Those adapter sequences were trimmed from output data with the help of Dr. Kazuyoshi Hosomichi of Inoue Laboratory at National Institute of Genetics, Mishima. Trimming was also made for 3' end of sequences having low base quality, less than 20. Trimmed sequence reads were mapped to human reference genome (hg19) with BWA software (Li and Durbin, 2009). I examined the frequencies of sequence reads mapped to hg19 with samtools, flagstat option (Li et al., 2009). To clarify the authenticity of mapped sequence reads, with the help of Dr. Kirill Kryukov I checked the effect of post-mortem misincorporation and depurination, which are characteristic of ancient DNA. Mapped reads were compared with human reference genome to estimate the frequency of the misincorporation and depurination. Examination of depurination was kindly suggested by Professor Shintaroh Ueda at Department of Biological Sciences, University of Tokyo. In the misincorporation frequency estimation, we used forward reads in the first +1 to +25 bases of mapped reads, and used reverse reads in the last -25 to -1 bases. For estimation of depurination, nucleotide frequencies of human reference genome in -10 to +30 bases of mapped reads were analyzed. To minimize the effect of those misincorporations in the following statistical analysis, the first and last five bases of mapped sequence reads were removed with the help of Dr. Kirill Kryukov.

3.2.4 Mitochondrial DNA haplotyping and investigating authenticity

To classify each Jomon individuals into specific mitochondrial DNA haplotypes, I used "PhyloTree.org" http://www.phylotree.org/ (van Oven and Kayser, 2008). I also confirmed

whether the mitochondrial DNA haplotypes decided from shotgun sequencing data were consistent with previous haplotyping based on Sanger sequencing method.

To explore the authenticity of mapped sequence reads, I estimated the frequencies of heterogeneity, $F_{inconsistent}$, in haplotype specific mutation sites with the equation as follows;

$$F_{inconsistent} = (n_{inconsistent}/n_{consistent})*100 \quad (3.1)$$

$n_{inconsistent}$ indicates the number of reads inconsistent with mitochondrial DNA haplotype of each individual, and $n_{consistent}$ indicates the number of reads consistent with mitochondrial DNA haplotype of each individual. Since the number of total reads mapped to mtDNA is few, I used t-test with 1 degree of freedom. 95% confidence interval (C.I.) was 6.314.

3.2.5 Sex determination and contamination estimates with sex chromosomes

For sex determination in ancient individual, I examined the ratio of the numbers of sequence reads mapped to X chromosome and Y chromosome, respectively. I used only mate-mapped forward reads, which are forward sequences and have mapping pair (not singleton, locating same chromosomes, and having CIGAR larger than 50). The ratio of bases covering X chromosome and Y chromosome was approximately 10:1 in ancient Aboriginal Australians (71,371,413 bases and 7,130,997 bases, respectively) according to the ratio of genome size (Rasmussen et al., 2011). Therefore, I estimated the sex as male if the ratio were close to 10:1, and I estimated the sex as female if the ratio were quite different from 10:1.

I recognized the sequence reads mapped to chromosome Y in female samples as modern human DNA contamination. I estimated the frequency of contamination, $F_{inconsistent}$, in ancient females by using the number of sequence reads mapped to chromosome X and Y. The equation to calculate $F_{inconsistent}$ is as follows;

$$F_{inconsistent} = [n_y*10/(n_x/2)]*100 \quad (3.2)$$

n_x and n_y indicate the number of reads mapped to X chromosome and Y chromosome, respectively. I assumed binominal distribution to calculate 95% C.I.

3.2.6 Estimating the frequency of post-mortem changes in Sanganji Jomon samples

To infer the degree of post-mortem change in Sanganji Jomon, I calculated the number of differences among Sanganji Jomon, 14 present-day humans, Vindija Neanderthal, Denisovan, and Chimpanzee (Table 3.2). I downloaded human genome data (BAM and BAM.bai format) of GWD (Gambian in Western Division, The Gambia), ESN (Esan in Nigeria), MSL (Mende in Sierra Leone), CDX (Chinese Dai in Xishuangbanna, China), and JPT (Japanese in Tokyo, Japan) from ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp/ (The 1000 Genomes project consortium, 2012), and of San, Yoruba, DNK (Dinka), Mbuti, French, Sardinian, Papuan, Han, Karitiana, and Denisovan from http://cdna.eva.mpg.de/denisova/ (Meyer et al., 2013), and of Vindija Neanderthal from http://genome.ucsc.edu/Neandertal/ftp://hgdownload.cse.ucsc.edu/gbdb/hg19/neandertal/seqAlis/ (Green et al., 2010), and also downloaded Chimpanzee genome (PanTro2) from http://hgdownload.soe.ucsc.edu/goldenPath/hg19/vsPanTro2/axtNet/ (The Chimpanzee Sequencing and Analysis Consortium, 2005). With the help of Dr. Kirill Kryukov, I chose the sequence reads having mapping quality greater than 30 to minimize the effect of sequencing error.

As Denisovan is archaic human inhabited Denisova Cave, Altai Mountains, Siberia (Meyer et al., 2012), in addition to Sanganji Jomon, I also removed the first and last five bases of mapped sequence reads in Denisovan genome to minimize the effect of post-mortem changes.

I merged genome data of two Jomon individuals, Sanganji 131421-3(1) and 131464(1) (the data was mainly used in later statistical analysis), and chose the sequence reads having mapping quality greater than 30. In Sanganji Jomons, most sequence data were derived from a single template because of low coverage of the Jomon genome. For making the data condition even, I randomly chose one nucleotide having base quality greater than 30, and prepared haploid data in all Sanganji Jomon, present-day humans, and Denisovan.

Table 3.2: List of individuals for estimating the frequency of post-mortem changes in ancient DNA and for *D*-statistic analysis

Sample ID	Sample name	Dataset	Population
PanTro2	PanTro2	UCSC	Chimpanzee
Denisovan (high coverage)	Denisovan	Meyer et al.,	Archaic human
Vindija Neanderthal	Vindija Neanderthal	Green et al.,	Archaic human
HGDP01029	San	Meyer et al.,	African
HGDP00927	Yoruba	Meyer et al.,	African
HGDP0456	Mbuti	Meyer et al.,	African
HG02922	ESN	1000 Genome Project	African
HG02571	GWD	1000 Genome Project	African
HG03052	MSL	1000 Genome Project	African
DNK02	DNK	1000 Genome Project	African
HGDP00521	French	Meyer et al.,	European
HGDP00665	Sardinian	Meyer et al.,	European
HGDP00542	Papuan	Meyer et al.,	Melanesian
131421-3(1) 131464(2)	Sanganji Jomon	This study	Ancient Japanese
HG02398	CDX (Southern Chinese Dai)	1000 Genome Project	East Asian
NA18960	JPT (Japanese Tokyo)	1000 Genome Project	East Asian
HGDP00778	Han	Meyer et al.,	East Asian
HGDP00998	Karitiana	Meyer et al.,	Native American

In Chimpanzee genome, PanTro2, original data (reference) were haploid sequences. Since the genome sequence was produced with utilizing a combination of whole genome plasmid reads as well as fosmid and BAC end sequences, I assumed that no-sequencing error in the genome sequence data in PanTro2.

I chose the nucleotide positions that all 16 individuals overlap with Sanganji Jomons. I then removed the nucleotide positions where three or four types of base were observed. To infer the pattern of post-mortem change, I counted the number of differences between seven Africans and non-African individuals at the nucleotide positions that the bases are common in all seven Africans. I estimated the frequency of post-mortem change by using the numbers. The number of differences between seven Africans and present-day humans include both real differences (N_d) and sequencing and PCR error (N_e). The number of differences between seven Africans and Sanganji Jomons additionally includes post-mortem changes (N_p). I calculated the frequency of post-mortem change in Sanganji Jomons as follows, based on my own reasoning;

$$N_{o,am} = N_{d,am} + N_{e,am} \quad (3.1)$$

$$N_{o,as} = N_{d,as} + N_{e,as} + N_p \quad (3.2)$$

$$F_p = (N_p/L)*100 = \{(N_{o,as} - N_{o,am})/L\text{-}APD\}*100 \quad (3.3)$$

am is the pair of African and present-day humans, *as* is the pair of African and Sanganji Jomons, N_o is the number of observed differences, N_d is the number of real differences, N_e is the number of errors, F_p is the frequency of post-mortem change in Sanganji Jomons, L is the length of overlapping sequences among all 17 individuals, and *APD* is average pairwise distance between Chimpanzee and all African and non-African individuals. There are seven different pairs of *am*, and $\hat{F}_p$ is finally obtained by taking average of seven different F_p values.

I also conducted different approach to infer the frequency of post-mortem changes in Sanganji Jomons. To infer the degree, I calculated the number of differences among all 17 individuals with MEGA 5.0 (Tamura et al., 2011). I focused on only transversion-type substitution. I assumed that there are no sequencing errors in Chimpanzee genome, and that the

genetic distance from Chimpanzee to all present-day humans are equal due to a constant rate molecular clock. I also assumed that the genetic distance between African individual and non-Africans are equal. I averaged the number of differences between Chimpanzee and present-day humans, and between African individual and Non-Africans. I subtracted each numbers from the number of differences between Chimpanzee and Sanganji Jomons, and between African individual and Sanganji Jomons, respectively. From the observed pair-wise distances, I calculated the frequency of post-mortem changes as follows;

$$F_p = [(O_A - APD)/(L - O_A)]*100 \qquad (3.4)$$

O_A indicates is observed pair-wise distances between ancient sample and Chimpanzee or non-African. $\hat{F}_p$ is finally obtained by taking average of seven different F_p values.

3.2.7 Principal component analysis

3.2.7.1 Data preparations and data analysis

I observed that genomic sequences of Jomon individuals include many post-mortem changes, which can induce false results in principal component analysis (PCA). To weed out the errors in the analysis, I focused on single nucleotide polymorphisms (SNPs). For the PCA, I used three Sanganji Jomon genome data, Sanganji 131421-3(1), 131421-3(2), and 131464(1). Because of low coverage of the Jomon genome, most sequence data were derived from a single template. I therefore prepared haploid data, and then duplicated it to produce diploid data in all three Jomon genomes. Thus all SNP sites became homozygous on them. To maximize the number of SNPs, I also prepared a different dataset by merging the sequence data of Sanganji 131421-3(1) and 131464(1). I didn't include Sanganji 131421-3(2), because I estimated that the sequence data includes relatively higher modern human DNA contamination than other two Sanganji genome. To allow comparison between Sanganji Jomon and modern human individuals, I prepared homozygous diploid datasets for modern human individuals. There are two types of data sets available: genome wide SNP data (e.g. HapMap project data), and whole genome sequence data

(SAM/BAM format). For the genome wide SNP data, I randomly chose one allele in each SNP site and duplicated it to prepare homozygous diploid datasets in each modern human sample with the help of Dr. Kirill Kryukov. For the genome sequence data, I collected nucleotide positions validated by dbSNP (http://www.ncbi.nlm.nih.gov/SNP/) (Database of Single Nucleotide Polymorphisms (dbSNP)). I randomly chose one base from multiple sequence reads in each nucleotide position and duplicated it to prepare homozygous diploid datasets. Nucleotide positions with three or more alleles were removed.

I filtered out SNP sites that had call rates were less than 95% and 90% in genome wide SNP and whole genome sequence data, respectively. To minimize the effect of PCR and sequence error, SNP with minor allele frequencies of less than 1% were removed. I used PLINK software for the SNP filtering (Purcell et al., 2007). After filtering, I extracted SNP sites that overlapped with merged Sanganji Jomon dataset. I carried out PCA based on the overlapping SNPs by using the smartpca program in the EIGENSOFT software package (Patterson et al., 2006).

3.2.7.2 Japanese Archipelago Human Population Genetics Consortium data & HapMap CHB

I obtained SNP data of Ainu, mainland Japanese, and Ryukyuan from Japanese Archipelago Human Population Genetics Consortium data (JAHPGC) (Jinam et al., 2012), and HapMap CHB (Han Chinese, Beijing) (The International HapMap Consortium, 2005). Figure 3.1 shows geographic locations of each population. Probably due to DNA degradation, only 13 of 36 Ainu individuals passed cQC (contrast quality control) threshold of 0.04 (Jinam et al., 2012), which measures the differences in contrast distributions for homozygote and heterozygote genotypes, and chose high quality SNP sites. Since 6 of 13 individuals seem to admix with mainland Japanese, and three pairs were close relatives (parents and offsprings), only five Ainu individuals were used in PCA. To maximize the number of individuals in the Ainu, I used the datasets prepared in Jinam et al. (2012), who did further filtering based on confidence scores (less than 0.008) generated with Affymetrix Birdseed Ver2 algorithm, and retained the maximum number of Ainu individuals. After the removal of close relatives, 32 individuals were available.

Dr. Timothy Jinam at Inoue Laboratory of National Institute of Genetics helped me for these procedures.

3.2.7.3 Human Genome Diversity Project

I used Human Genome Diversity Project (HGDP-CEPH) data to investigate the genetic relationship between Sanganji Jomon individual and worldwide populations (Li et al., 2008). Figure 3.1 shows geographic locations of each population.

3.2.7.4 1000 genome project

I downloaded human genome data of The 1000 Genomes Project Consortium (2012) from ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp/. To use large numbers of SNPs, I chose the individuals having at least 18GB-sized BAM file. I further chose the individuals having Illumina-based sequence data. Figure 3.1 shows geographic locations of each population.

3.2.7.5 Investigating the genetic relationship between HGDP humans and archaic humans

I examined the genetic relationship between HGDP humans and archaic humans with PCA. Protocol of the PCA is different from a standard PCA. With the help of Dr. Kirill Kryukov, I randomly chose one allele in each SNP sites and duplicated it to prepare homozygous diploid datasets in each present-day human. I first carried out a PCA on chimpanzee, Neanderthal and Denisovan, without using data from Jomon and present-day humans at all. Using the SNP weights from the PCA, I can then project the present-day humans onto the plane of PC1 and PC2. This allows us to explore the relationship of diverse present-day humans relative to archaic humans and chimpanzee, and to test if the genetic differences among present-day humans are correlated to the differences among these non-modern humans.

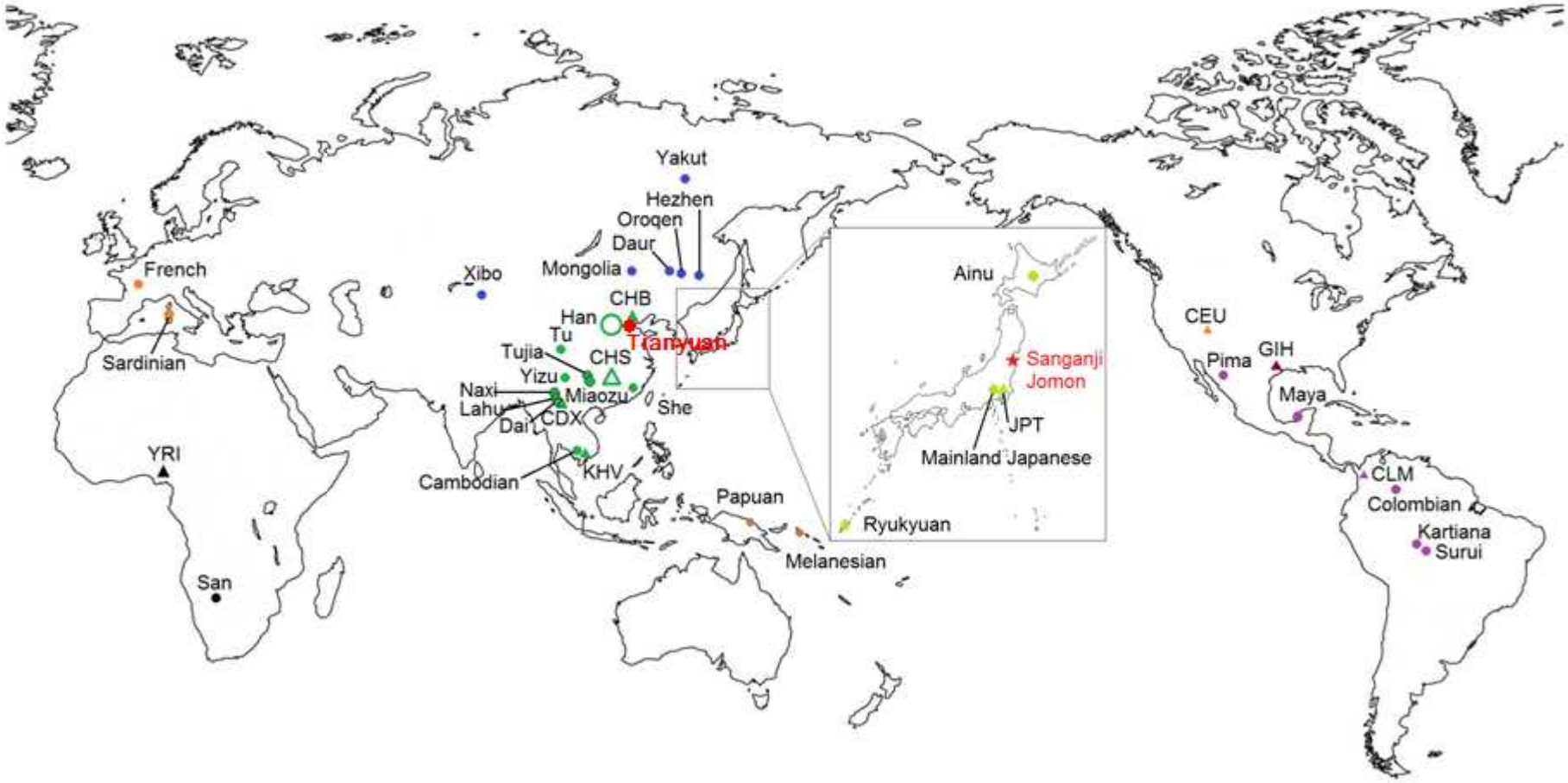


Figure 3.1: Geographical locations of JAHPGC, HapMap, HGDP-CEPH, 1000 genome project populations, and Tianyuan

3.2.8 Allele sharing analysis

I examined the frequency of shared SNPs between the Sanganji Jomon individual, and modern human populations. For genome wide SNP data, I used original genotype, not homogeneous diploid datasets I used in PCA. For human genome data, I instead used homogeneous diploid datasets, which I prepared in chapter 3.2.6, because the coverage of human genome was low in the datasets of present-day human individuals, and it was difficult to call both haplotypes at one SNP locus correctly. I used PLINK software to calculate the frequency of shared alleles with Sanganji Jomon in every SNP sites. Then, I took average of the frequencies of all SNP sites.

In Ainu people, I chose individuals with fewer admixture with mainland Japanese to minimize the effect of recent admixture. In other populations, all the individuals were used for the analysis.

3.2.9 Phylogenetic analysis

3.2.9.1 Data preparation and data analysis

To compare Sanganji Jomon individuals and worldwide populations, I prepared a set of allele frequencies in each population. The datasets were same with the dataset used in PCA. The frequencies of all nucleotide positions were 0.0 or 1.0 in Sanganji Jomon individual, which artificially makes his blanch length much longer than other blanches. I computed Nei's genetic distances (Nei 1972) in each pairs of populations with 1,000 bootstrap replicates, and constructed neighbor-joining tree (Saitou and Nei 1987) by using Phylip software (Felsenstein, 2005). I used Splitstree4 as a viewer of the neighbor-joining tree (Bryant & Moulton, 2004; Huson & Bryant, 2006). For phylogenetic network analysis, I directly used the data of Nei's genetic distances and constructed the network with Splitstree4. Since the branch length of Sanganji Jomon individual was very long, I truncated the length in figures of phylogenetic tree and networks.

For the comparison with modern human individuals, Vindija Neanderthal, Denisovan, and Chimpanzee, I used sequence data instead of allele frequencies. The genome data of modern humans, Denisovan, and Chimpanzee were already described in Section 3.2.5. I additionally

downloaded Genome data of Vindija Neanderthal from UCSC database. I remapped the genome data to human reference genome, hg19. I did same data filtering as described in Section 3.2.5 for the Vindija Neanderthal genome. I also removed the first and last five bases of mapped sequence reads in Vindija Neanderthal genome to minimize the effect of post-mortem change. I collected the nucleotide positions that all the 14 present-day humans, Vindija Neanderthal, Denisovan, and Chimp overlapped with Sanganji Jomons. To minimize the effects of post-mortem change and PCR and sequencing error, I additionally filtered out transition substitution sites. I computed pairwise distances in each pair with 100 bootstrap replicates, and constructed phylogenetic tree using MEGA 5.0 (Tamura et al., 2011). I also constructed phylogenetic network with Splitstree4 based on pairwise distances (Bryant & Moulton, 2004; Huson & Bryant, 2006). Since post-mortem changes and errors might occur randomly in the sequence reads, the topology of phylogenetic network within modern humans would be intact. For the relationship between archaic humans and modern humans, I need to consider the effect of post-mortem change in not only archaic humans but also Sanganji Jomons. When I constructed phylogenetic networks, I adjusted the number of pair-wise distance to take the effect of post-mortem changes into account. I adjusted the distance with the equations as follows;

$$E_{pd} = O_{pd} - (L - E_{pd})* P_i - (L - E_{pd})*P_j \quad (3.5)$$

namely,

$$E_{pd} = (O_{pd} - (P_i + P_j)*L)/(1- P_i - P_j) \quad (3.6)$$

E_{pd} is Expected number of pairwise distance, O_{pd} is observed pair-wise distances, L is the total length of sequence reads overlapped in all 17 individuals, P_i and P_j are the frequency of post-mortem change in ancient samples. If one of two individuals was modern humans, P_i or P_j was zero.

3.2.10 *D*-statistic analysis to infer the genetic relationship between Sanganji Jomon and other modern and ancient individuals

To further investigate the genetic relationship between Sanganji Jomon and present-day humans as well as archaic humans, I carried out *D*-statistic analysis (Reich et al., 2010). *D*-statistic analysis is a non-model-based analysis to detect gene flow among population or individual. I constructed input file with EIGENSOFT, convertf option (Patterson et al., 2006), and calculated *D* with ADMIXtools (Patterson et al., 2012). The equation is as follows;

$$D = (n_{ABAB} - n_{ABBA})/(n_{ABAB} + n_{ABBA}) \quad (3.7)$$

$$Z = D/SE \quad (3.8)$$

A is ancestral allele observed in Chimpanzee and *B* is derived allele (Figure 3.2). I used Chimpanzee or Africans as outgroup, and chose three individuals, hominin_1 (H1), hominin_2 (H2), and hominin_3 (H3), from other individuals. H1 is an outgroup of H2 and H3. The order of four alleles corresponds to Outgroup-H1-H2-H3. n_{ABAB} and n_{ABBA} indicates the number of ABAB and ABBA, respectively. "*SE*" is standard error, and I set block jackknife as 0.05 to calculate *SE*. If there was gene flow between H1 and H3, the value of *D* and *Z*-score becomes positive, and if there was gene flow between H1 and H2, the value of *D* and *Z*-score becomes negative. I considered that if the *Z*-score is larger than +4 or smaller than -4 (the criteria is strict than previous studies (Reich et al., 2011; Skoglund and Jakobsson, 2011), there was gene flow between them. When I included Vindija Neanderthal for the *D*-statistic analysis, the number of total nucleotide sites was limited, which gives high SE-values and non-significant *Z*-scores. Therefore, when I included Vindija Neanderthal for the *D*-statistic analysis, I used +2 and -2 instead of +4 and -4.

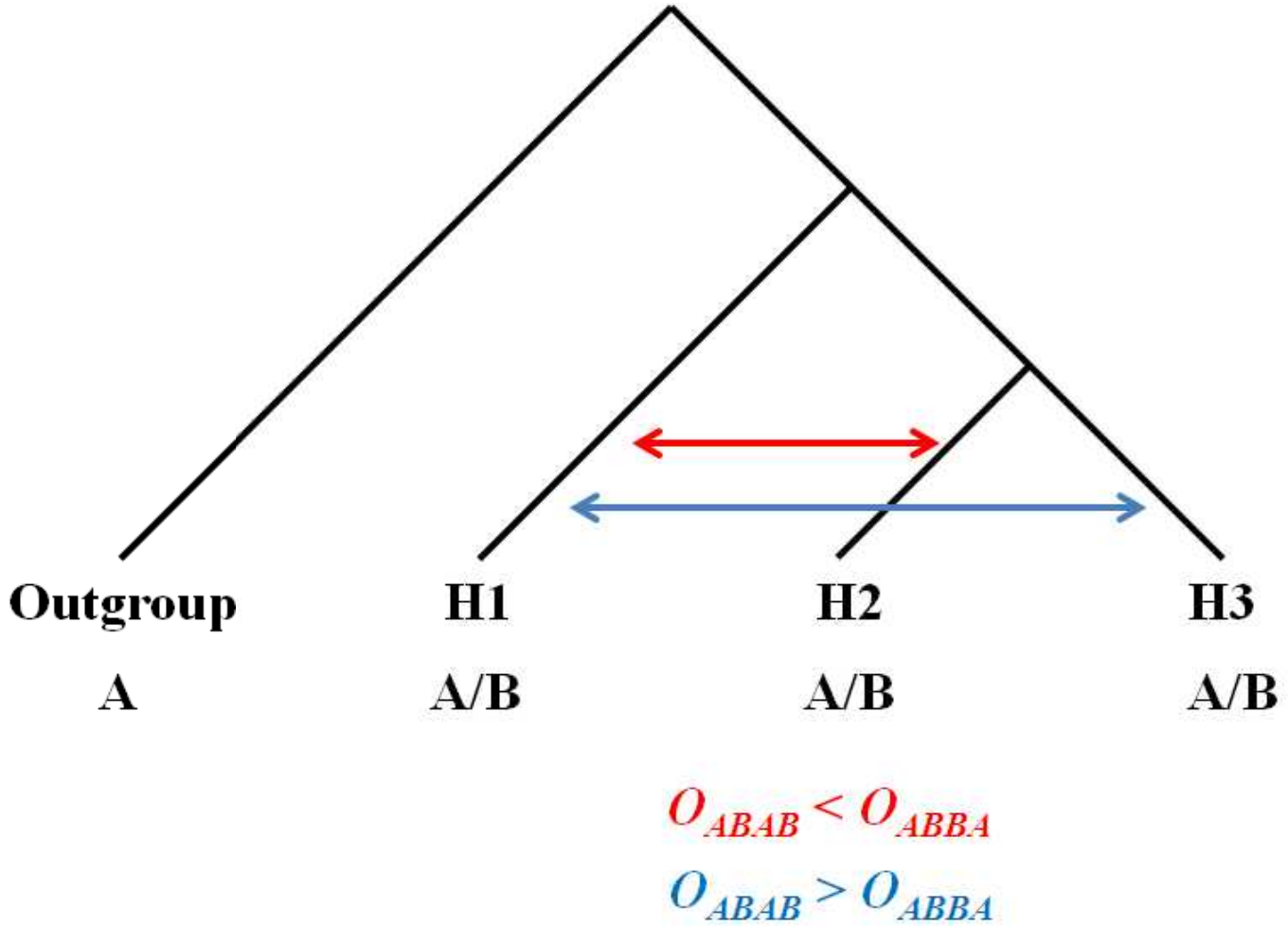


Figure 3.2: Model of *D*-statistic analysis. Capital A and B indicate ancestral allele and derived allele, respectively
If there were no-gene flow among H1, H2, and H3, O_{ABAB} and O_{ABBA} should be equal. If there were gene flow between H1 and H2 or between H1 and H3, [$O_{ABAB} < O_{ABBA}$] or [$O_{ABAB} > O_{ABBA}$] is expected, respectively.

Ancient samples include post-mortem change, and the change skews the *D*-value and *Z*-score. To weed out the effect of post-mortem change in *D*-statistical analysis, I adjusted the *D*-value and *Z*-score in the pairs, which include ancient samples. I consider the pair of (Chimpanzee, archaic humans; present-day humans, Sanganji Jomons), and used the frequency of post-mortem change estimated in chapter 3.2.5 for the adjustment. *Pi* and *Pj* are the frequency of post-mortem change in archaic humans and Sanganji Jomons, respectively. For example, ABBA can change into ABBB at frequency *Pi*, and into AABA at frequency *Pj*, respectively. I assumed that all the three individuals, archaic humans, present-day humans, and Sanganji Jomons, equally include PCR error and sequencing error, and neglected its effect. So, the observed n_{ABAB} (O_{ABAB}) and n_{ABBA} (O_{ABBA}) can be described with the expected n_{ABAB} (E_{ABAB}) and n_{ABBA} (E_{ABBA}) as follows;

$$O_{ABAB} = E_{ABAB} + PiE_{AAAB} + PjE_{ABAA} - PiE_{ABAB} - PjE_{ABAB} \quad (3.9)$$

$$O_{ABAA} = E_{ABAA} + PiE_{AAAA} + PjE_{ABAB} - PiE_{ABAA} - PjE_{ABAA} \quad (3.10)$$

$$O_{AAAB} = E_{AAAB} + PiE_{ABAB} + PjE_{AAAA} - PiE_{AAAB} - PjE_{AAAB} \quad (3.11)$$

$$O_{ABBA} = E_{ABBA} + PiE_{AABA} + PjE_{ABBB} - PiE_{ABBA} - PjE_{ABBA} \quad (3.12)$$

$$O_{ABBB} = E_{ABBB} + PiE_{AABB} + PjE_{ABBA} - PiE_{ABBB} - PjE_{ABBB} \quad (3.13)$$

$$O_{AAAB} = E_{AAAB} + PiE_{ABBA} + PjE_{AABB} - PiE_{AABA} - PjE_{AABA} \quad (3.14)$$

$$O_{AAAA} = E_{AAAA} + PiE_{ABAA} + PjE_{AAAB} - PiE_{AAAA} - PjE_{AAAA} \quad (3.15)$$

In equation (3.15), I assumed E_{ABAA} as (O_{ABAA}-PiO_{AAAA}), and E_{AAAB} as (O_{AAAB}-PjO_{AAAA}). From (3.9), (3.10), (3.11) and (3.15), E_{ABAB} can be calculated as follows;

$$E_{ABAB} \fallingdotseq [O_{ABAB} - PjO_{ABAA}/(1-Pi-Pj) - PiO_{AAAB}/(1-Pi-Pj) + 2PiPj[\{1+(Pi)^2+(Pj)^2\}O_{AAAA} - PiO_{ABAA} - PjO_{AAAB}]/(1-Pi-Pj)^2]/(1-Pi-Pj) \quad (3.16)$$

While, from (3.12), (3.13), and (3.14), E_{ABBA} can be calculated as follows;

$$E_{ABBA} \fallingdotseq \{O_{ABBA} - (Pi+Pj)O_{ABBB}/(1 - Pi - Pj)\}/(1 - Pi - Pj) \quad (3.17)$$

For example, in (Chimpanzee, Denisovan; San, Sanganji Jomons), observed data are as follows;

(O_{ABAB}, O_{ABBA}, O_{ABBB}, O_{ABAA}, O_{AAAB}, O_{AABA}, O_{AAAA}, O_{AABB}) = (1864, 1896, 2013193, 39091, 111899, 14256, 50725773, 4606)

From (3.9), D_{standard} and Z_{standard} becomes -0.0085 and -0.480, respectively, and *Z*-score was non-significant. However, after the modifications with (3.16) and (3.17), the D_{revision} and Z_{revision} becomes 0.1269 and 7.160, respectively, and *Z*-score became significant. If I use present-day humans instead of archaic humans, *Pi* is zero.

3.3 Results

3.3.1 Mapping and Contents of sequence reads

A total of 1,178,246,254 reads were sequenced from five Sanganji Jomon DNA libraries (Table 3.3). Ancient DNA includes many non-human sequence reads (e.g. Green et al., 2010). In the sequence data of Sanganji 131421-3(1), about half of the read were not assigned uniquely to the specific species in MEGABLAST (Figure 3.3). About 40% of the read were assigned to bacteria. The tendency is consistent with previous studies (e.g., Green et al., 2010). The sequence reads classified into Primates were only 2.6% of total reads.

Table 3.3: Summary of genomic sequence data from three Sanganji Jomon individuals

Sample	Sanganji 131421-2(1)	Sanganji 131421-3(1)	Sanganji 131421-3(2)	Sanganji 131464(1)	Sanganji 131464(2)
Illumina index	5	-	16	12	6
Total reads	41,688,616	198,645,228	472,784,726	107,931,912	357,195,772
Hits to hg19 (%)	379,850 (0.91%)	1,365,861 (0.68%)	835,495 (0.18%)	1,906,607 (1.77%)	110,502 (0.03%)
Non-restriction enzyme treatment	-	501,837 (0.46%)	-	-	-
restriction enzyme treatment	-	864,024 (0.97%)	-	-	-
Total reads (>= Mapq30) (%)	8610 (0.02%)	1,062,498 (0.53%)	674,416 (0.14%)	1,377,955 (1.28%)	2,195 (0.0006%)
Properly paired (%)	1,069 (12.42%)	210,109 (19.78%)	346,007 (51.30%)	284,963 (20.68%)	748 (34.08%)
With intself and mate mapped	1,981	973,078	646,208	1,226,518	984
Singletons (%)	6,629 (76.99%)	89,420 (8.42%)	28,207 (4.18%)	151,437 (10.99%)	1,210 (55.13%)
With mapped to a different chr	5	1,309	1,092	1,971	1
Hits after rmdup (%)	8,598 (0.02%)	553,061 (0.28%)	586,788 (0.12%)	1,157,557 (1.07%)	2,189 (0.0006%)
Bases covering hg19 (only autosome) (%)	-	18,133,262 (0.63%)	30,865,776 (1.07%)	46,018,723 (1.59%)	-
Coverage of mtDNA (%)	-	11,140 (67.23%)	14,748 (89.01%)	14,012 (84.57%)	-
Sequence reads mapped to X-chromosome *1	-	3,098	6,348	16,665	-
Sequence reads mapped to Y-chromosome *1	-	303	413	23	-
Relative frequency of base pair mapped to X and Y (X : Y)	-	10.2 : 1	15.4 : 1	724.6 : 1	-
Sex identification based on X : Y	-	Male	Male	Female	-
Sex identification based on morphological study	-	Male	Male	Female	-

*1 For sex identification and contamination estimates, only the numbers of pairwise-forward reads were counted.

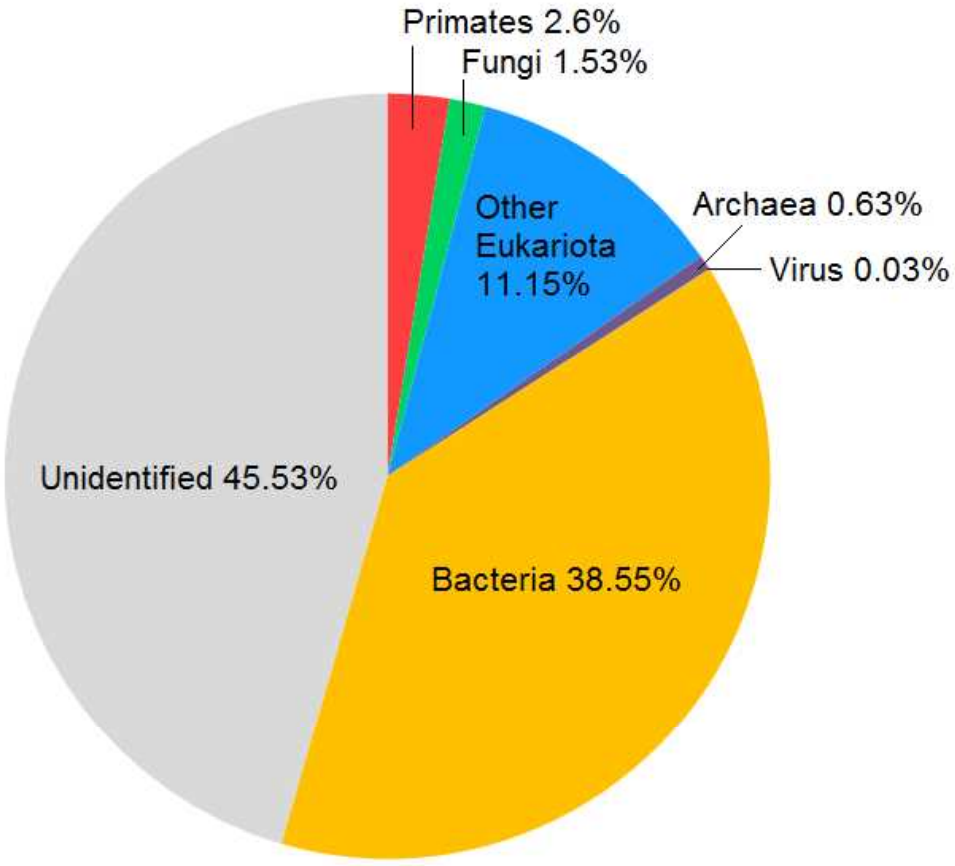


Figure 3.3: Content and distribution of Sanganji Jomon's meta-genome
About 2.6% of the sequence reads were uniquely assigned to primates with MEGABLAST.

About 0.03% to 1.77% of total reads were mapped to the human reference genome (hg19) (Table 3.3). I expected that all the five libraries have human DNA at high frequency because I successfully typed mitochondrial DNA haplotypes from those three individuals (Chapter 2). However, the ratios of human derived DNA were very small in Sanganji 131421-2(1) and 131464(2) after the filtering with mapping quality (>=30). Those libraries include singletons at high frequency. The high frequency of singletons may indicate that non-human DNA accidently mapped to human reference genome, and that the DNA libraries didn't include human DNA. In other words, the low frequency of human derived DNA implies that the amounts of modern human DNA contamination are very low in our experiments. In the following analysis, I used the data of Sanganji 131421-3(1), 131421-3(2), and 131464(1). I successfully identified sexes with the ratio of chromosome X and Y in the two samples. These results were consistent with morphological observation of the individual (see Chapter 2). The consistency among morphological and genetic identification also support the authenticity of mapped DNAs.

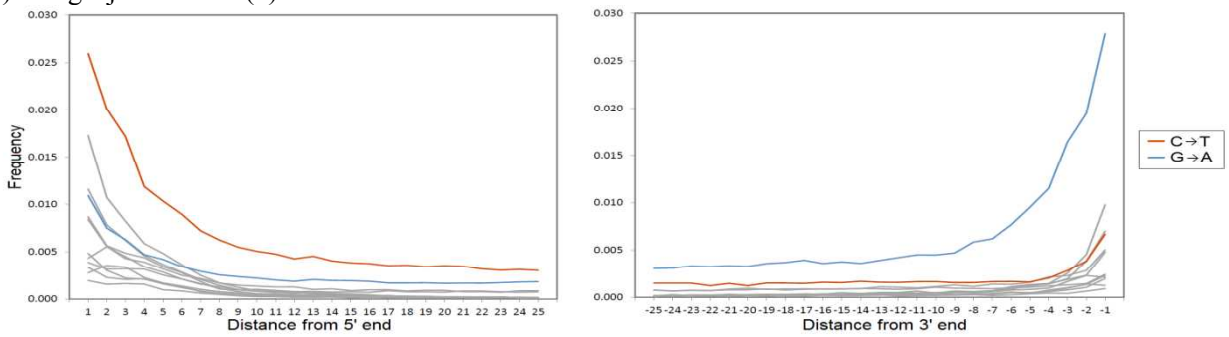
3.3.2 Confirming the authenticity with Mitochondrial DNA and sex chromosomes

To clarify whether mapped sequence reads derives from ancient DNA, I checked the effect of post-mortem misincorporation and depurination, which are characteristic of ancient DNA. C to T and G to A misincorporations were observed at high frequency in 5' and 3' ends, respectively (Figures 3.4a-3.4c). I also observed high frequencies of A and G in the next to the end of mapped reads termini, which derives from post-mortem depurination (Figures 3.4d-3.4f). These results support that the sequence reads mapped to hg19 includes ancient human DNA (it doesn't assure that all the mapped sequences are from ancient human DNA).

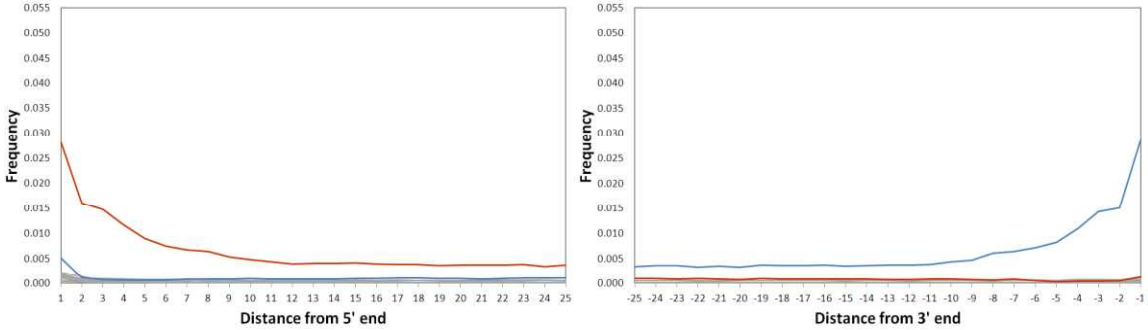
Although I used clean room to minimize modern human DNA contamination, it is difficult to remove all the contaminations because some DNA probably comes from the air or/and from reagents (e.g., I previously obtained high coverage of mitochondrial DNA sequences from 300 years old Edo era Japanese, but the haplotype was specific to Africans, not observed in

modern East Asians. The DNA sequences are probably modern human DNA contamination from reagents). To estimate the frequency of modern human DNA contaminations, I used the sequence reads mapped to mitochondrial DNA and sex chromosomes. In mitochondrial DNA, I calculated the frequencies of inconsistency with mitochondrial DNA haplotypes. In sex chromosomes, I recognized sequence reads mapped to Y chromosome in female samples as modern human DNA contaminations, and the numbers are used to estimate the frequency of the contaminations. The frequencies, estimated with mitochondrial DNA, were in between 0.00% and 7.50% (Table 3.4). The haplotypes of Sanganji 131421-3(1) and 131421-3(2) were consistent each other. However, the haplotypes of Sanganji 131464(1), haplotype N9b*, were partially different from the result of previous Sanger based haplotyping, haplotype N9b2 (Table 2.6). It is still unclear that which result is correct. The frequency, estimated with the ratio of chromosome X and Y, was 2.76% (2.41-3.11%, 95% C.I.) in Sanganji 131464(1). The frequency is very close to the frequency estimated from mtDNA haplotype (2.70%). In any cases, these results support the authenticity of sequence data.

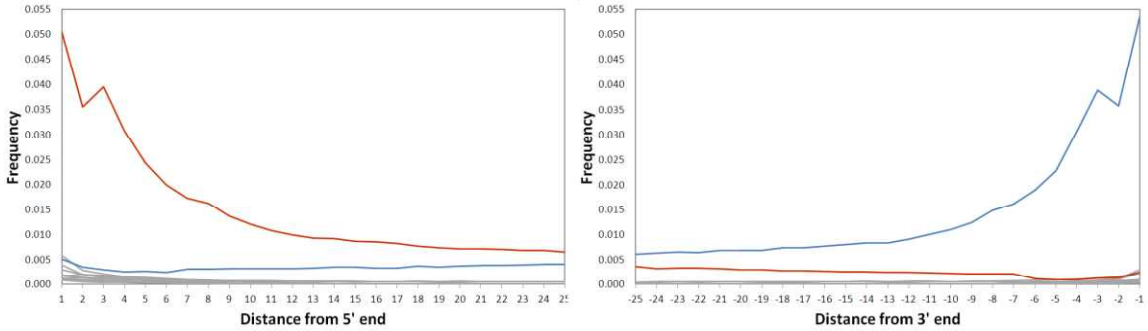
a) Sanganji 131421-3 (1)



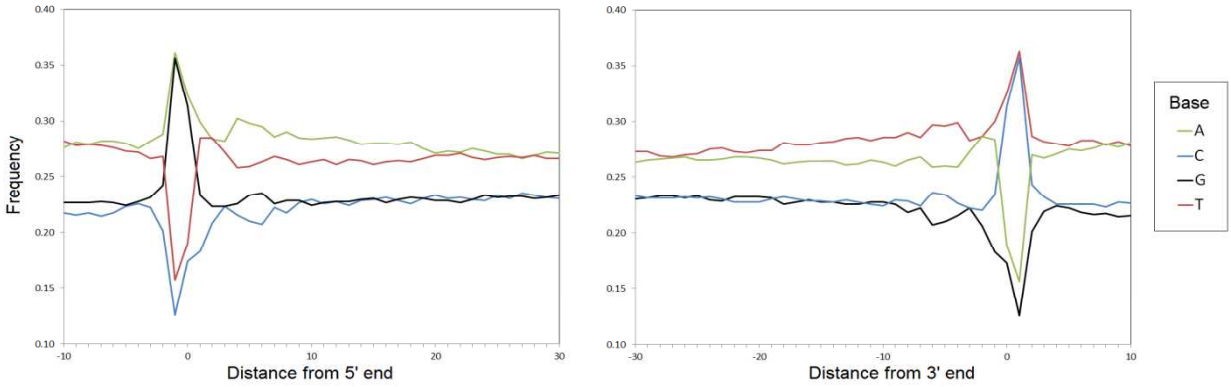
b) Sanganji 131421-3 (2)



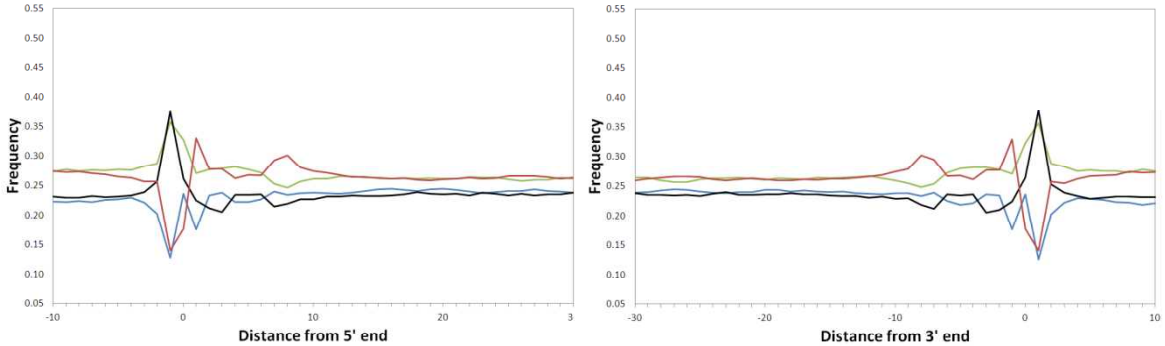
c) Sanganji 131464



d) Sanganji 131421-3 (1)



e) Sanganji 131421-3 (2)



f) Sanganji 131464

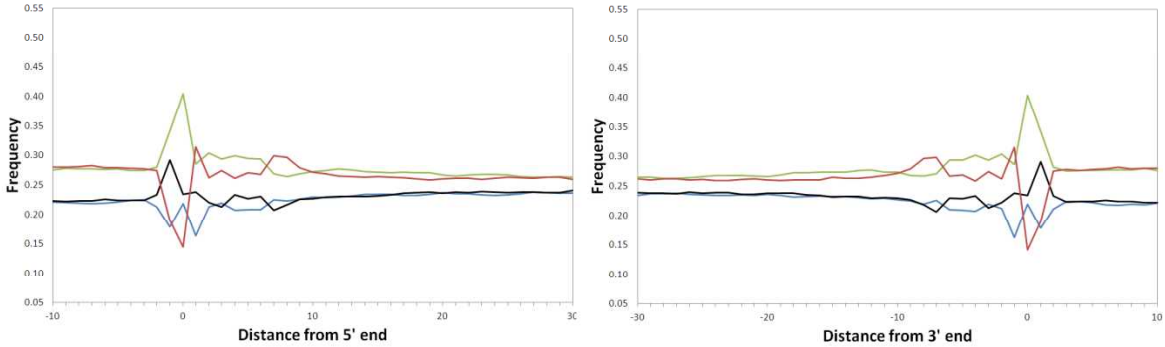


Figure 3.4: Pattern of postmortem misincorporation and depurination in three Sanganji Jomon individuals

(a-c) Red line indicates C to T misincorporation, and blue line indicates G to A misincorporation. (d-f) Base composition of the human reference genome around the 5'- and 3'-ends of the sequence reads.

Table 3.4: Contamination estimates with mitochondrial DNA and sex chromosomes

Remains	ID	Sex	*Nconsistent* *1	*Ninconsistent* *2	Total	*Fmt* *3	Number of sequence reads mapped to chromosomeX	Number of sequence reads mapped to chromosomeY	X : Y	*Fsex* *4
Sanganji shell mound	131421-2(1)	?	-	-	-	-	-	-	- - -	-
	131421-3(1)	Male	29	0	29	0.00%	3,098	303	10.2 : 1	-
	131421-3(2)	Male	37	3	40	7.50%	6,348	413	15.4 : 1	-
	131464(1)	Female	36	1	37	2.70%	16,665	23	724.6 : 1	2.76%
	131464(2)	Female	-	-	-	-	-	-	- - -	-
Aboriginal		Male	-	-	-	-	71,371,413	7,130,997	10.0 : 1	-

*1 Number of reads consistent with haplotype

*2 Number of reads inconsistent with haplotype

*3 Frequency of contamination infered from mitochondrial DNA haplotype

*4 Frequency of contamination infered from the ratio of sex chromosomes

3.3.3 Damage estimates in Sanganji Jomon

I inferred the degree of post-mortem change in Sanganji Jomon. Totally 52 million bases of the genome were available (Table 3.5). First, I aligned the genome of Sanganji Jomon and 14 modern humans (7 Africans and 7 non-Africans), and I chose variant sites common in 7 Africans but in others. Since Africans are the outgroup of non-Africans, pairwise distance among Africans and non-Africans should be same in the variant sites. In other words, excess distances are the result of post-mortem changes. The pairwise distance between Sanganji Jomon and African was much larger than the distance between modern humans and Africans. I found that Sanganji Jomon include many post-mortem changes (1.6650%) (Table 3.6). Especially, the frequency of transition type substitutions (1.4308%) is six times higher than the frequency of transversion type substitutions (0.2342%). Within transition substitution, the frequency of C to T and G to A is extremely higher than other patterns of transition type substitutions. This pattern is typical in ancient DNA (Green et al., 2010; Reich et al., 2011).

I inferred the frequency of post-mortem transversion type substitutions in ancient samples from the number of differences among 17 individuals. I considered that pairwise distances between Chimpanzee and anatomically modern humans or archaic humans are the same, and excess distances are from post-mortem changes. While the frequency was 0.1904% in Sanganji Jomons, the frequency was 0.0456% in Denisovan (Table 3.7). When I included Vindija Neanderthal, the frequencies were 0.1705%, 0.0437%, 0.0247%, in Sanganji Jomon, Denisovan, and Vindija Neanderthal, respectively. The frequencies were quite different among the ancient samples. There are two possibilities to explain the difference. The first possibility is that the frequencies of post-mortem change were truly different between the two individuals. Environmental conditions (e.g. temperature, humidity, pH value, geochemical properties of the soil) make differences the condition of ancient DNA. Another possibility is the difference of the depth in each nucleotide positions. To reduce the effect of sequence error, I chose bases having high base quality in each nucleotide positions. If the depth is low, the possibility to choose relatively low base quality will increase, which can be recognized as post-mortem changes, not as sequencing error. In other words, the effect of sequencing error to the frequencies of post-mortem

changes may be reduced if the depth is high enough.

Table 3.5: Summary of genomic sequence data overlapping among Sanganji Jomon (131421-3 and 131464), 14 modern humans, Denisovan (high coverage), and Chimpanzee (PanTro2)

	Number of bases
All sequence	52,462,354
All sequence (only one type or two types)	52,454,859
Fixed in all 17 individuals	
All	50,621,450
A	14,277,968
T	14,286,664
G	11,029,641
C	11,027,177
All variant site (only two types)	
All	1,833,409
Transversion	481,128

Table 3.6: Summary of difference between seven African and non-African individuals to infer the degree of post-mortem change in each types of bases in Sanganji Jomon

Seven African (San, Yoruba, Mbuti, ESN, GWD, MSL, and DNK)									
Common in seven African but not in others			Sanganji Jomons *1	French	Sardinian	Papuan	Han	JPT	CDX	Karitiana
A	244497	A->T	19427	861	890	910	935	1307	716	1045
		A->G	33349	4116	3376	3352	3611	2612	1837	3627
		A->C	11542	950	967	1003	1023	733	482	1048
T	246412	T->A	17587	847	879	902	862	1295	753	1026
		T->G	15984	1022	982	951	965	722	458	1066
		T->C	33467	4261	3533	3292	3426	2505	1845	3625
G	500446	G->A	278037	5103	4869	4891	4774	3050	2497	5002
		G->T	13676	1147	1154	1195	1249	2168	927	1153
		G->C	9257	821	830	855	861	795	667	850
C	499623	C->A	15933	1091	1222	1238	1184	2179	912	1221
		C->T	280590	4854	4785	4887	4695	3043	2443	5113
		C->G	4375	809	821	884	848	808	627	869
Total	1,490,978	Total	733224	25882	24308	24360	24433	21217	14164	25645
		Transition	625443	18334	16563	16422	16506	11210	8622	17367
		Transversion	107781	7548	7745	7938	7927	10007	5542	8278
Maximum differences (total)			1.7186%	0.0607%	0.0570%	0.0571%	0.0573%	0.0497%	0.0332%	0.0601%
Maximum differences (transition)			1.4659%	0.0430%	0.0388%	0.0385%	0.0387%	0.0263%	0.0202%	0.0407%
Maximum differences (transversion)			**0.2526%**	**0.0177%**	**0.0182%**	**0.0186%**	**0.0186%**	**0.0235%**	**0.0130%**	**0.0194%**
Frequency of post-mortem change (total)			1.6650%							
Frequency of post-mortem change (transition)			1.4308%							
Frequency of post-mortem change (transversion)			**0.2342%**							

*1 Merged dataset (Sanganji Jomon 131421-3 and 131464)

Table 3.7: Pair-wise distance in transversion substitution among 17 individuals at 52,454,859 nucleotide sites uncorrected for post-mortem change and the estimation of the frequency of post-mortem change in Sanganji Jomon

	Pa	San	Yor	Mbu	ESN	GWD	MSL	DNK	Den	S.J.	Fre	Sar	Pap	Han	JPT	CDX
PanTro2 (Pa)	0															
San	223,951	0														
Yoruba (Yor)	222,687	30,544	0													
Mbuti (Mbu)	222,529	30,454	28,576	0												
ESN	223,488	31,581	28,225	29,561	0											
GWD	223,908	32,019	28,785	30,005	29,616	0										
MSL	224,496	32,647	29,401	30,617	30,378	30,748	0									
DNK	222,841	30,794	27,908	28,766	28,613	29,251	29,933	0								
H-Denisovan *1 (Den)	246,044	59,817	58,593	58,353	59,434	59,958	60,432	58,773	0							
Sanganji Jomon *2 (S.J.)	321,562	129,915	127,199	127,923	127,904	128,378	129,082	127,049	157,492	0						
French (Fre)	222,045	29,876	26,798	27,868	27,779	28,353	28,941	26,776	57,689	123,245	0					
Sardinian (Sar)	222,103	29,904	26,912	28,044	28,021	28,487	29,127	26,858	57,847	123,523	22,392	0				
Papuan (Pap)	222,389	30,268	27,434	28,270	28,475	28,755	29,755	27,424	57,659	123,673	23,862	24,176	0			
Han	222,267	30,060	27,336	28,194	28,181	28,695	29,449	27,226	57,975	122,615	23,532	23,726	23,702	0		
JPT	224,270	32,203	29,331	30,207	30,006	30,438	31,368	29,061	59,944	124,136	25,377	25,641	25,619	24,267	0	
CDX	219,899	27,700	24,770	25,704	25,537	26,039	26,739	24,738	55,461	120,063	20,928	21,260	21,126	19,812	21,725	0
Karitiana	222,634	30,531	27,627	28,479	28,524	29,082	29,828	27,617	58,388	123,256	23,595	23,785	24,213	23,205	25,152	20,735
Average Pairwise Distance between modern non-Africans (APD)	222,230	30,077	27,173	28,109	28,075	28,550	29,315	27,100	Average							
Number of difference between APD and Sanganji Jomon	99,332	99,838	100,026	99,814	99,829	99,828	99,767	99,949	**99,798**							
Post-mortem change in Sanganji Jomon (%) *3	0.1902	0.1904	0.1908	0.1904	0.1904	0.1904	0.1903	0.1906	**0.1904**	Number of post-mortem change in Sanganji Jomon				**99,898** (+1,000)		
Number of difference between APD and Denisovan (n)	23,814															
Post-mortem change in Denisovan (%) *3	**0.0456**									Number of post-mortem change in Denisovan				**23,916** (+101)		

*1 High coverage of Denisovan genome

*2 Merged genome dataset with Sanganji 131421-3 and Sanganji 131464

*3 The frequency was calculated with [(number of post-mortem change)/(Total nucleotide sites-APD)]*100

3.3.4 Principal component analysis

I carried out principal component analysis (PCA) to explore the genetic relationship between the Sanganji Jomon individual and modern worldwide populations. I listed the number of SNPs used for PCA in Table 3.8. When I used worldwide populations, Sanganji Jomons grouped with other East Eurasians (Figures 3.5-3.7). Within East Eurasia, however, Sanganji Jomon was isolated from other modern East Eurasians (Figures 3.8, 3.9). The isolation indicates that the Sanganji Jomon was genetically apart from modern East Eurasians. However, modern Japanese were located in between Sanganji Jomon and other East Eurasians in PC3 of Figure 3.8b. Within East Asians, modern Japanese (JPT) was located between Sanganji Jomon and Han Chinese from Beijing (Figure 3.9). This indicates that the genetic source of immigrants from Eurasian continents in and after Yayoi era was from Northern East Asians.

Three populations, Ainu, mainland Japanese, and Ryukyuan, currently inhabit the Japanese Archipelago (Figure 2.1). I compared Sanganji Jomon with the three Japanese populations and Han Chinese from Beijing. PC1 divides Ainu and Sanganji Jomons from mainland Japanese, Ryukyuan, and Han Chinese, and PC2 divides Ainu and Sanganji Jomons (Figure 3.10). This result suggests the genetic similarity between Sanganji Jomon and modern Ainu people, followed by the Ryukyuan, but basically their genetic background was quite different from that of modern three populations inhabiting in the Japanese Archipelago.

To investigate the effect of modern human DNA contamination in statistical analysis, I carried out PCA with Sanganji 131421-3(2), which had 7.50% (1.2–13.8%, 95% C.I.) of modern human DNA contaminations. Surprisingly, I didn't observe the effect of the contamination (Appendix Figure 1).

I also investigated the genetic relationship between archaic humans and HGDP humans. PC1 divides non-Africans and archaic humans from African and Chimpanzee (Figure 3.11). This indicates that non-Africans were genetically closer to archaic humans than Africans. Vindija Neanderthal hauls Europeans and East Eurasians, while Denisovan pull not only Melanesians but also East Asians. I suppose that East Asians are genetically closer to Denisovan than Europeans.

Table 3.8: Summary of the number of overlapping SNP sites and transversion substitution sites between Sanganji Jomons and five datasets

	Sanganji 131421-3(1)	Sanganji 131421-3(2)	Sanganji 131464(1)	Sanganji 131421-3(1) and Sanganji 131464(1)
JAHPGC and HapMap CHB				
Five Ainu individuals	4,600	7,097	9,424	13,341
32 Ainu individuals	3,526	-	-	-
HGDP-CEPH	4,664	7,111	9,601	11,084
1000 genome project				
Worldwide	74,491	-	-	-
East Eurasians	61,339	-	-	-
High coverage of Denisovan and 14 present-day humans	-	-	-	481,128
Vindija Neanderthal, high coverage of Denisovan and 14 present-day humans	-	-	-	221,750

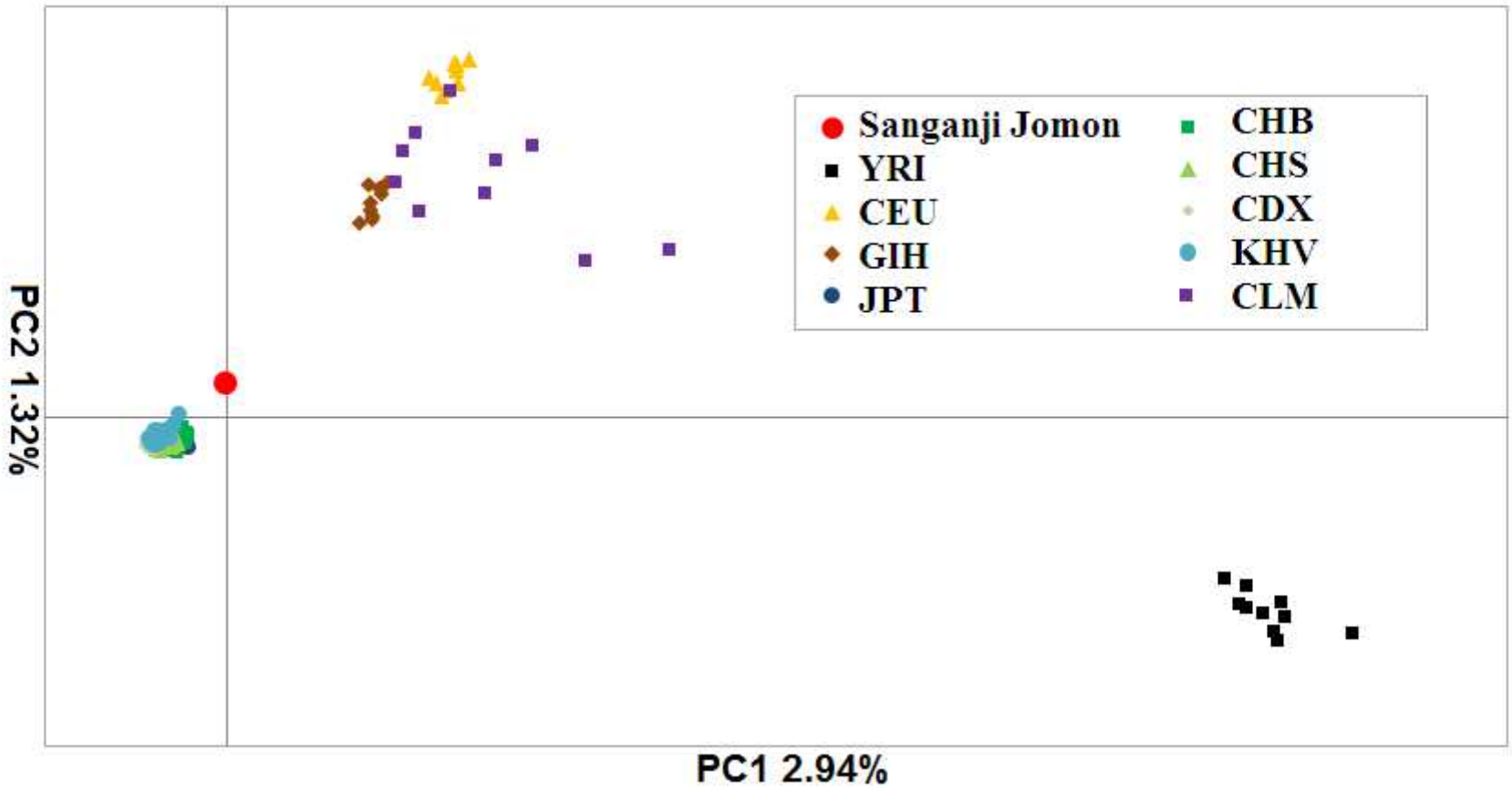


Figure 3.5: Genetic relationship among Sanganji Jomon individual and worldwide modern human individuals based on 74,491 SNPs
The abbreviation was consistent with that of 1000 genome project (YRI: Yoruba in Ibadan, Nigeria, CEU: Utah residents with Northern and Western European ancestry, GIH: Gujarati Indian in Houston,TX, JPT: Japanese Tokyo, CHB: Han Chinese, Beijing, CHS: southern Han Chinese, CDX: Chinese Dai in Xishuangbanna, China, KHV: Kinh in Ho Chi Minh City, Vietnam, CLM: Colombian in Medellin, Colombia).

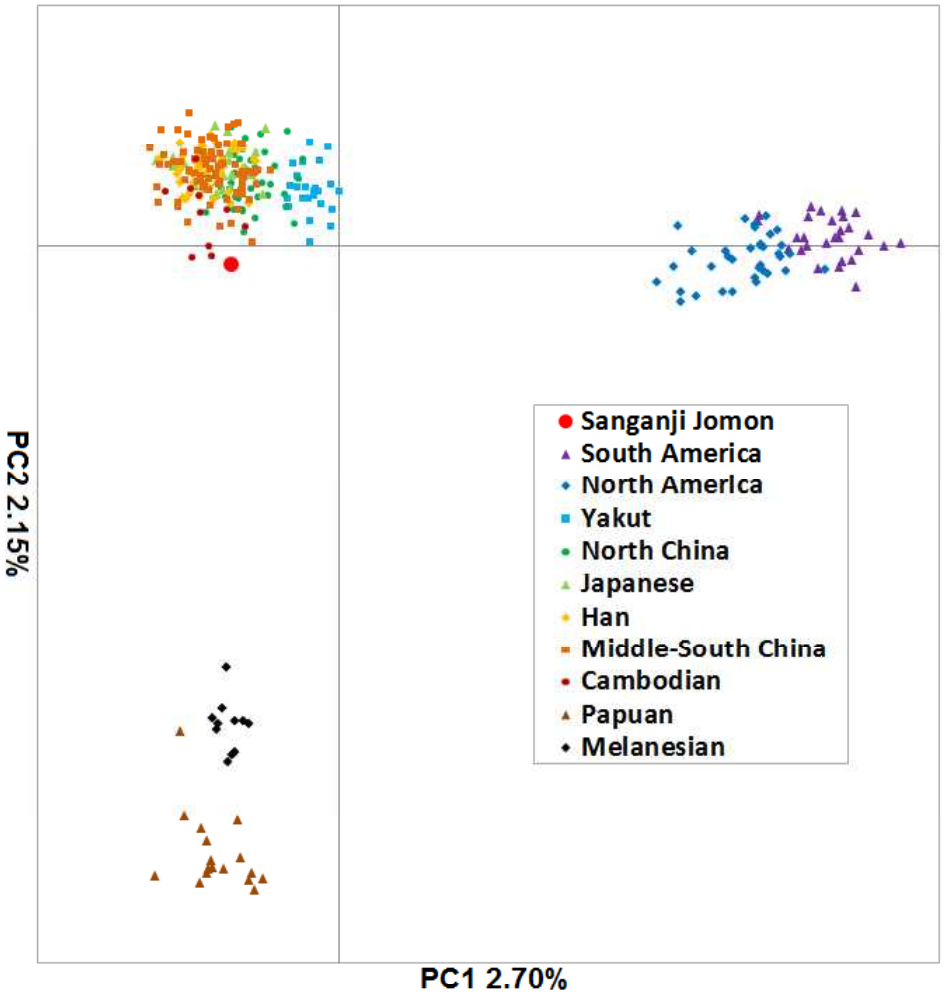


Figure 3.6: Genetic relationship among Sanganji Jomon individual 1 and the individuals of East Eurasia, Oceania, and America based on 4,319 SNPs

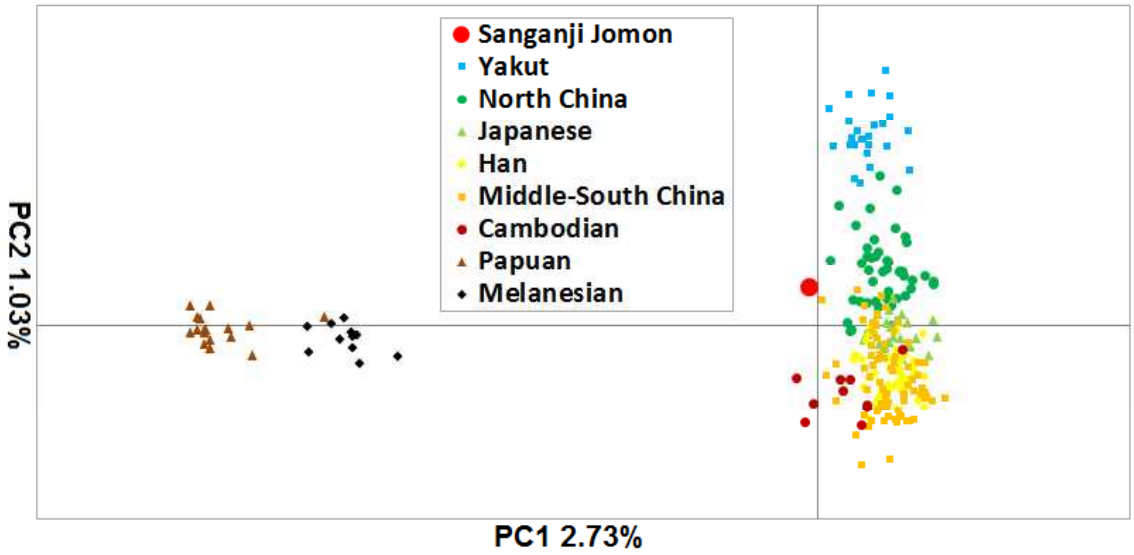
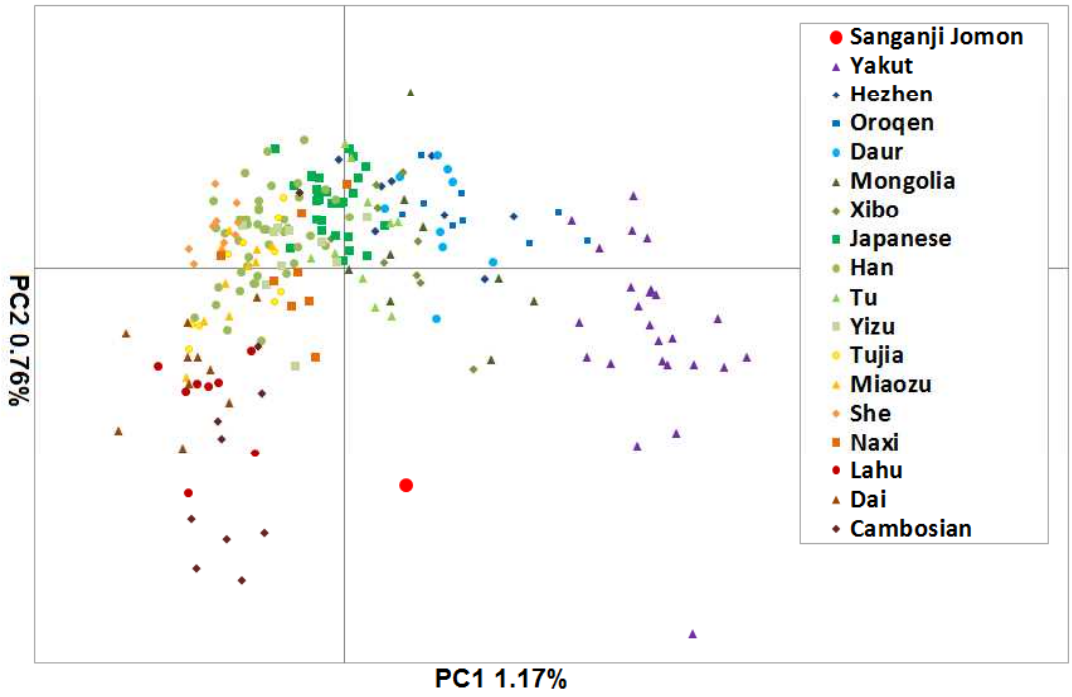


Figure 3.7: Genetic relationship among Sanganji Jomon individual 1 and the individuals of East Eurasia and Oceania based on 4,319 SNPs

a) PC1 and PC2



b) PC1 and PC3

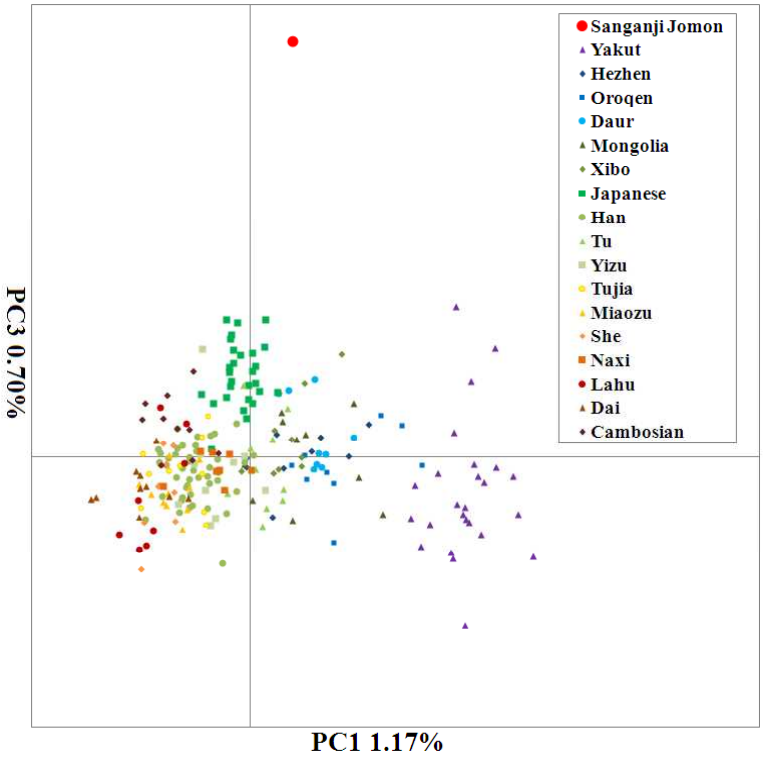


Figure 3.8: Genetic relationship among Sanganji Jomon individual and East Eurasians based on 3,921 SNPs
a) PC1 and PC2, b) PC1 and PC3

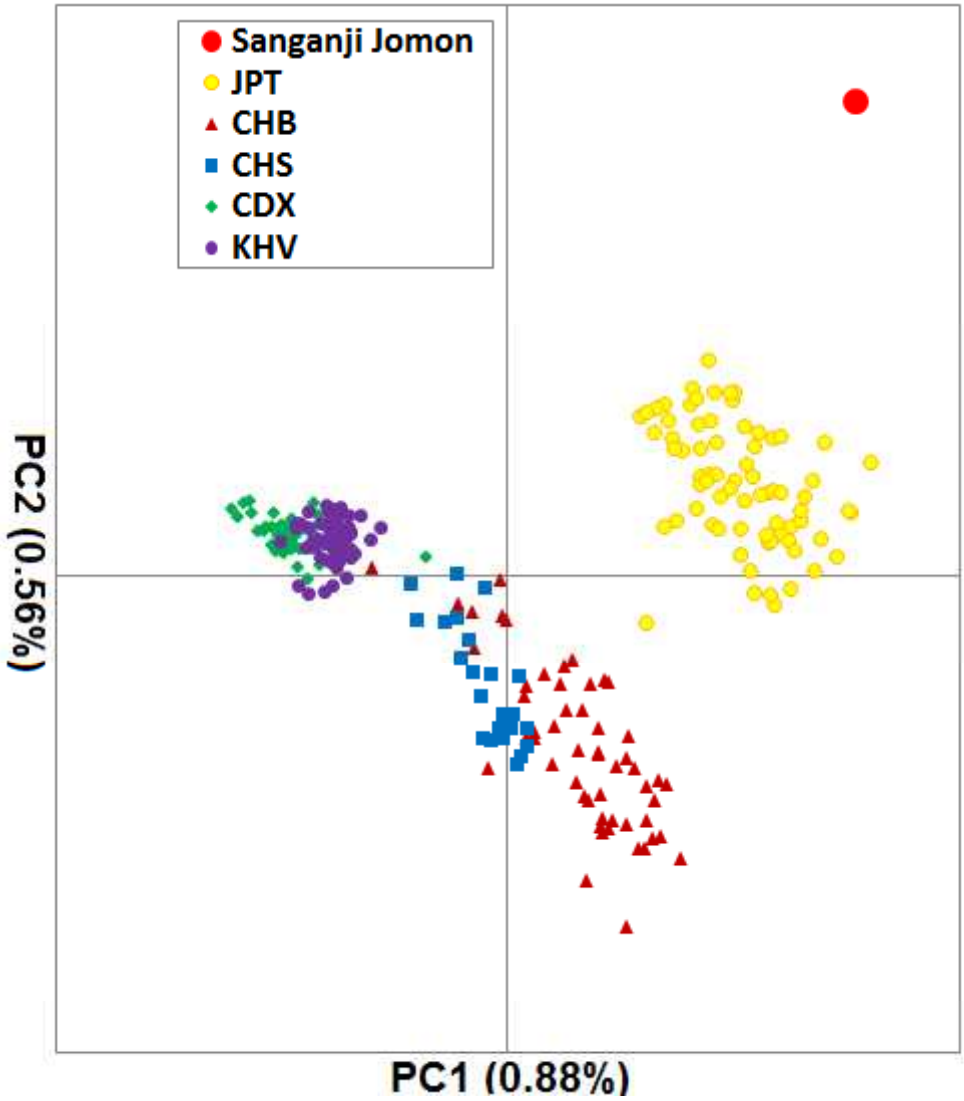
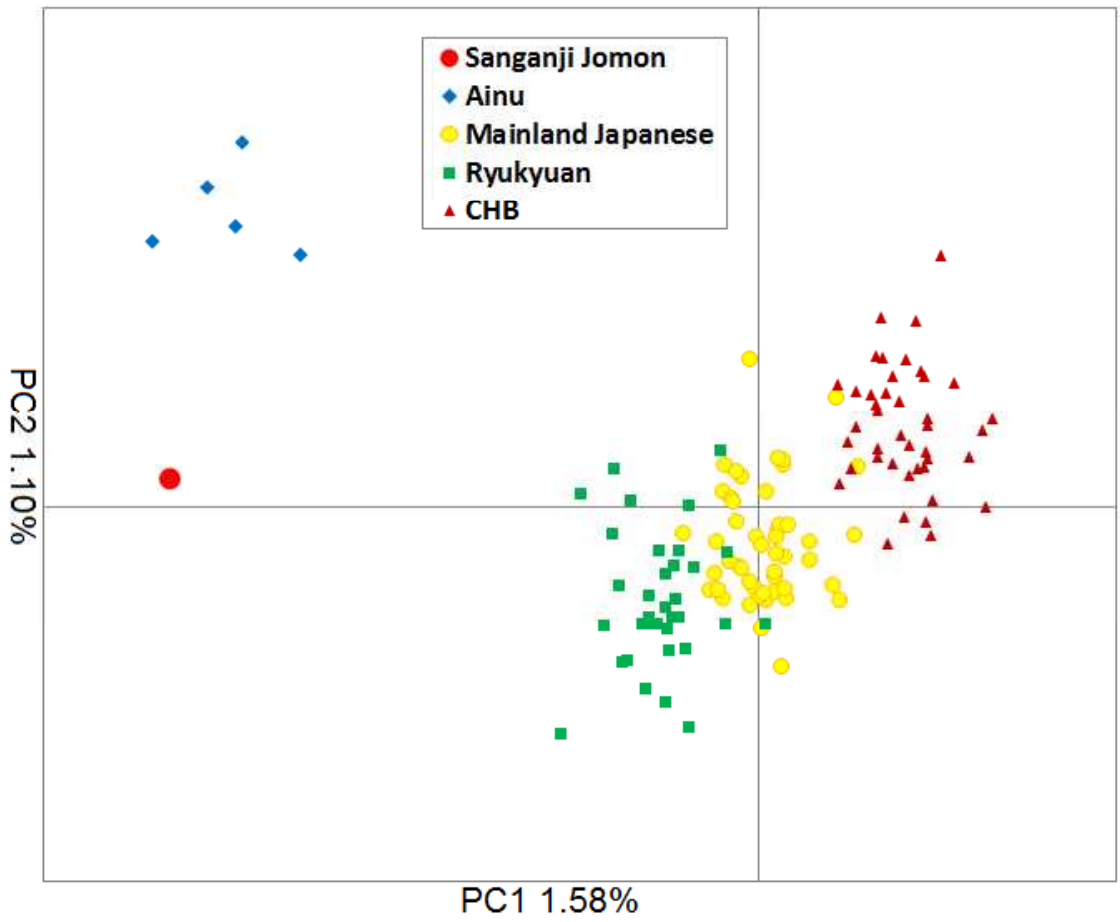


Figure 3.9: Genetic relationship among Sanganji Jomon individual and East Eurasians based on 61,339 SNPs

a)



b)

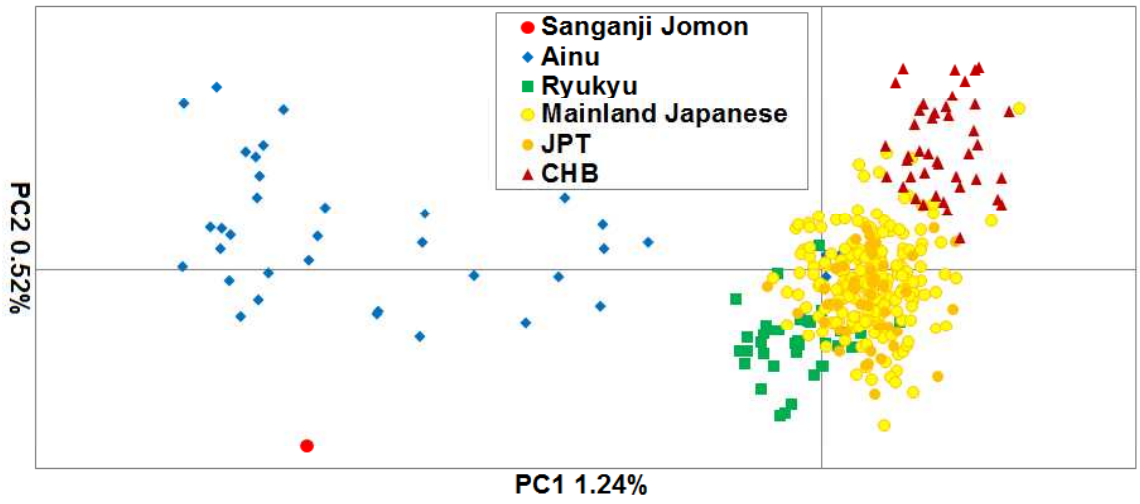


Figure 3.10: Genetic relationship among Sanganji Jomon individual and the individuals of three populations inhabits in Japanese archipelago (Ainu, mainland Japanese, and Ryukyuan) and of Chinese Beijing (CHB)

a) 4,600 SNP sites with five Ainu individuals who have non-recent admixture, b) 3,526 SNP sites with 32 Ainu individuals

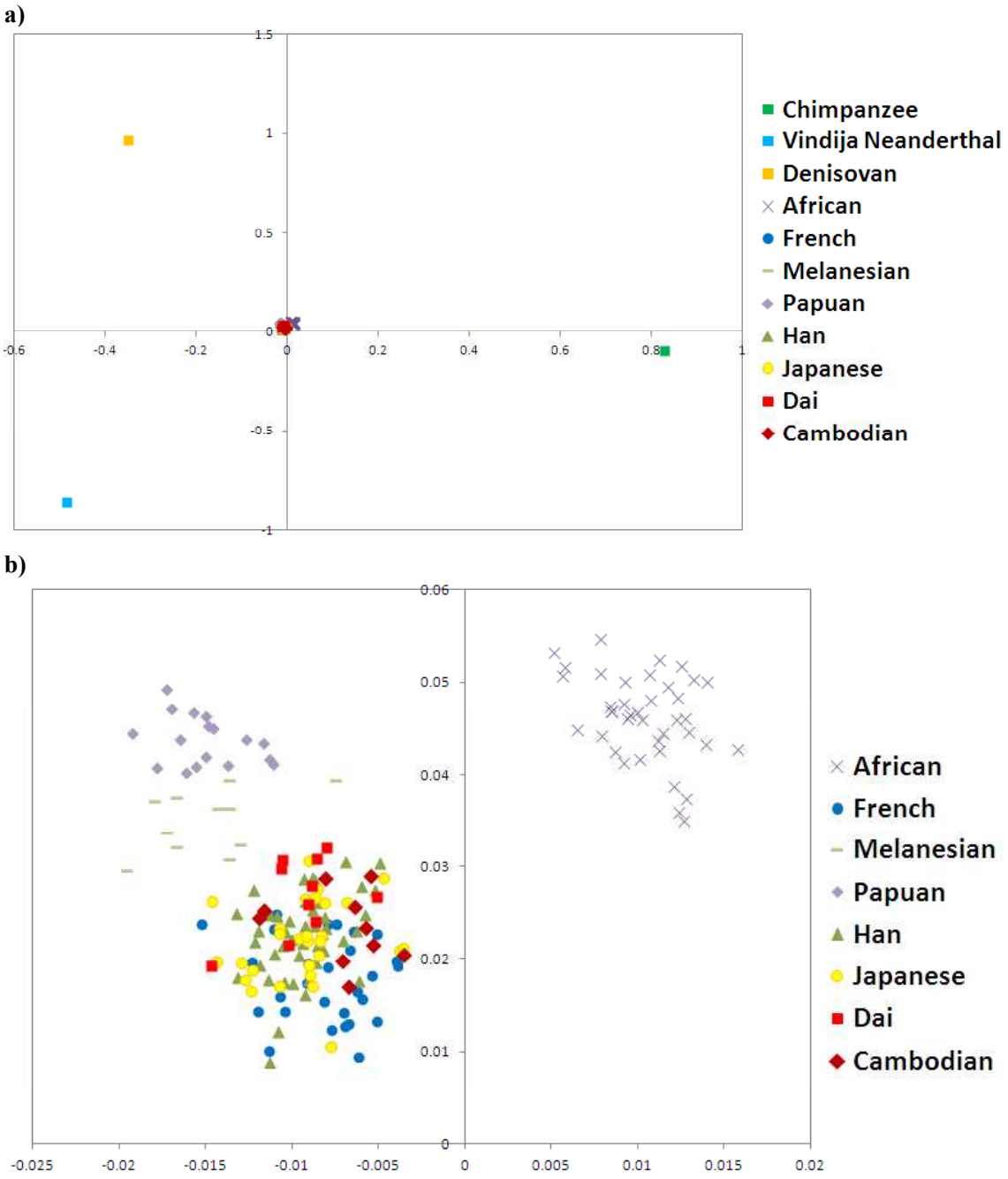
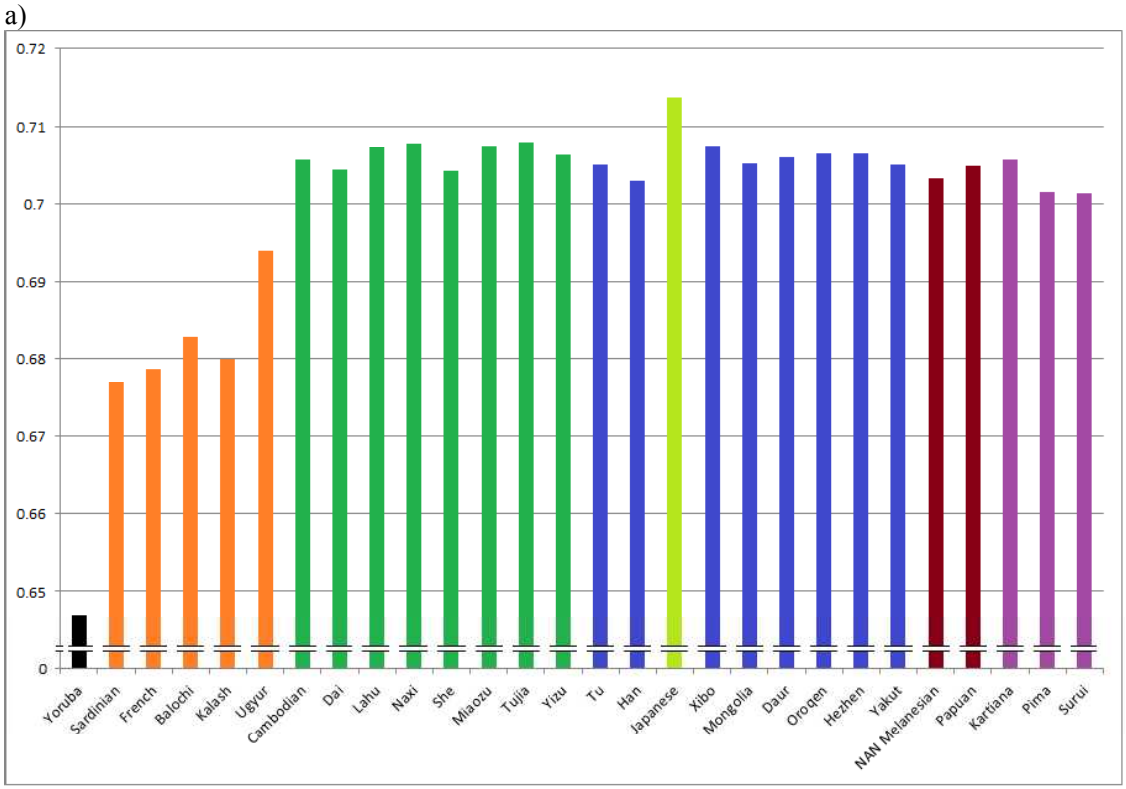


Figure 3.11: Principal Component Analysis of 70,010 SNPs known to be polymorphic among three individuals, Chimpanzee, Vindija Neanderthal, and Denisovan
38,477 SNPs were unique to Chimpanzee, 22,480 SNPs were unique to Vindija Neanderthal, and 18,019 SNPs were unique to Denisovan. a) PCA of Chimpanzee, Vindija Neanderthal, Denisovan, and modern humans. b) Magnification of the central portion of the plot.

3.3.5 Allele sharing analysis

Because the number of individuals and the coverage of genome is limited in Sanganji Jomon people, it is not suitable to interpret population differentiation between Jomon and modern populations based on *F*st value. I instead investigated the frequency of allele sharing between Sanganji Jomon individuals and modern populations to infer degree of genetic similarity between them. I observed that modern East Eurasians share more alleles with Sanganji Jomons than other worldwide populations (Figure 3.12a). Within East Eurasians, modern Japanese shares more alleles with Sanganji Jomons than other populations (Figures 3.12b, 3.13). The finding is consistent with the result of PCA. The frequency of allele sharing with Sanganji Jomon is relatively high in Northern and Southern East Eurasian ethnic minorities (such as Xibo, Tujia, and Naxi), and is low in Han Chinese, She, and Dai (although the degree of differences are not significant). Within Japanese populations, Ainu people retain Jomon component the most, and Ryukyuan was second most.

a)



b)

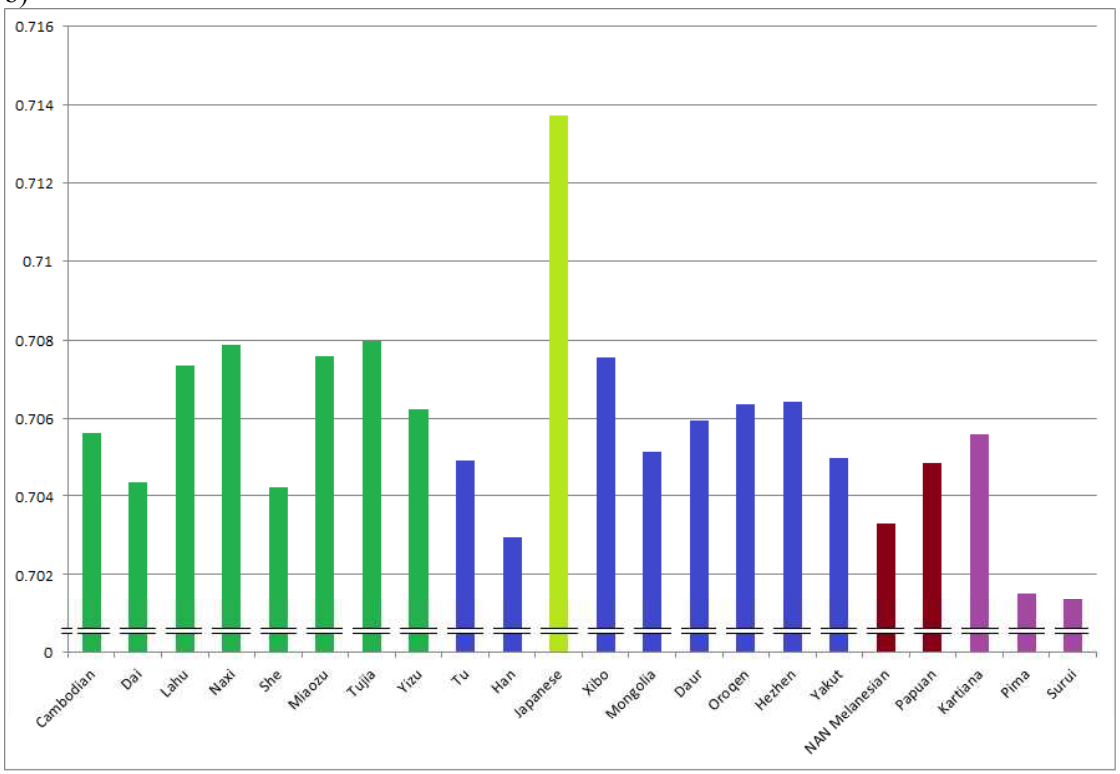


Figure 3.12: Allele sharing between Sanganji Jomon individual and worldwide populations
Color of orange, green, blue, brawn, and violet indicate Europeans, southern East Eurasians, northern East Eurasians, Oceanians, and Native Americans, respectively. a) comparison with worldwide populations, b) comparison with East Eurasians, Melanesians, and Native Americans.

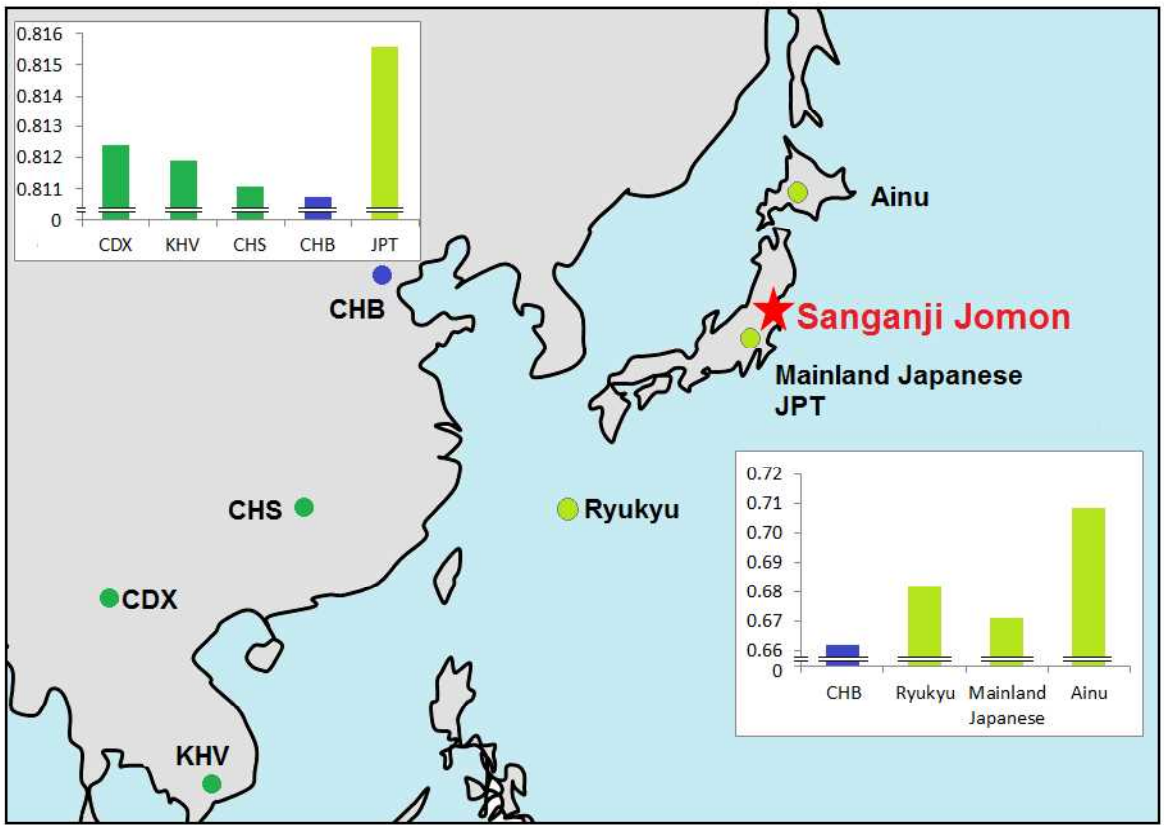


Figure 3.13: Allele sharing between Sanganji Jomon individual and continental East Eurasians and three populations inhabits in the Japanese Archipelago
Color of green and right green indicates southern East Eurasians and Japanese populations, respectively.

3.3.6 Phylogenetic tree analysis

To investigate the phylogenetic relationship between Sanganji Jomon and other humans, I constructed neighbor-joining tree. Sanganji Jomon clustered neither with Northeast nor with Southeast Asians, but was located as the outgroup of modern East Eurasians and Native American (Figures 3.14, 3.15). Branch lengths of Sanganji Jomon were long. I assume that the cause of long branch is post-mortem change, PCR error, and sequencing error. In HGDP populations, the bootstrap value between Sanganji Jomon and other East Eurasian populations were not significant (69.6%), but the value increased when I removed Japanese from the analysis (85.7%) because Japanese inherit Jomon genome and the inheritance makes the bootstrap value low. In the populations from 1000 genome project, I observed same tendency, although Colombia, Native American, clustered with European (Figure 3.15). The cluster probably indicates recent gene introgression from Europeans into Colombian. However, when I constructed the tree with genome sequences of 17 individuals, Sanganji Jomon was outgroup of modern East Asians but postdated the divergence of East Asians and Karitiana with 100% bootstrap value (Figure 3.16). The bootstrap value between Sanganji Jomons and Karitiana was instead non-significant when I removed Japanese from the analysis (79%) (Figure 3.17). Within the Japanese Archipelago, Ainu and Sanganji Jomon clustered with the 100% bootstrap probability (Figure 3.18). This result is consistent with the result of PCA (Figure 3.10).

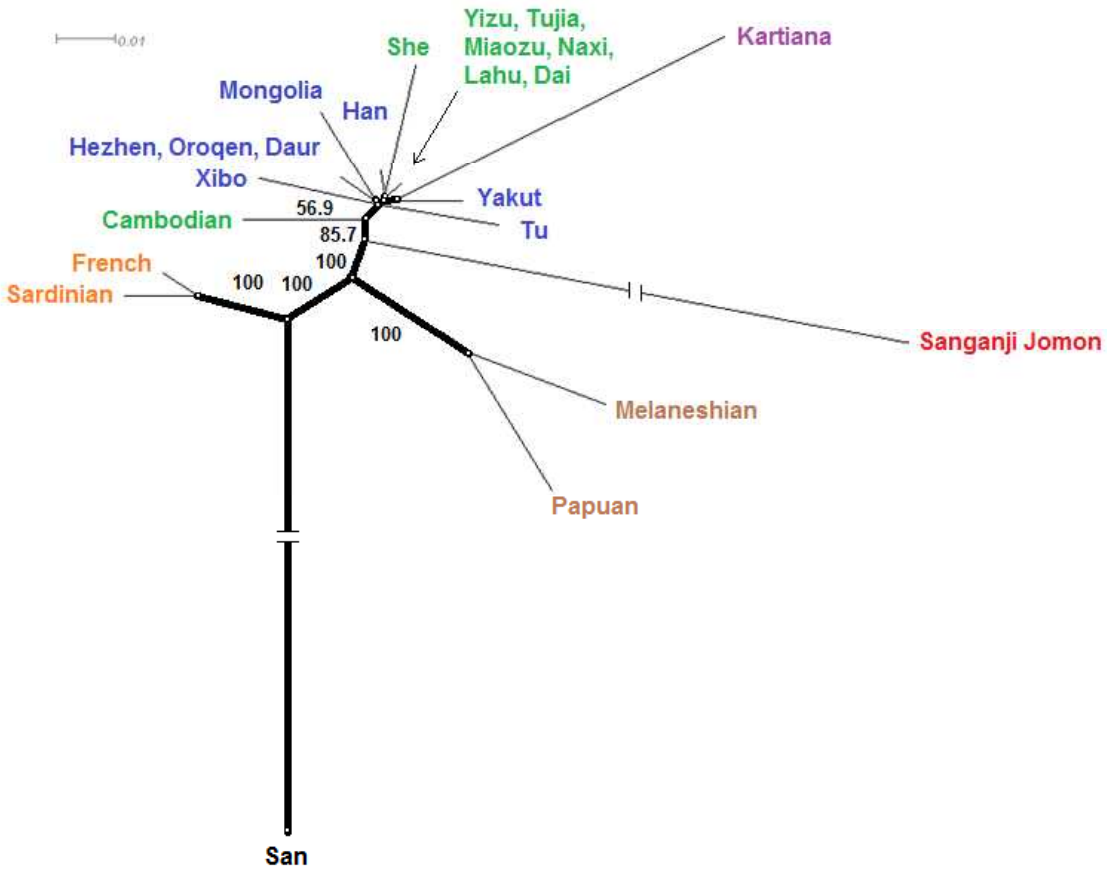


Figure 3.14: A Neighbor-joining tree of Sanganji Jomon individual and HGDP modern human populations based on allele frequencies of 4,319 SNPs
Color of blue, green, orange, and brawn indicate northern East Eurasians, southern East Eurasians, Europeans, and Oceanians, respectively.

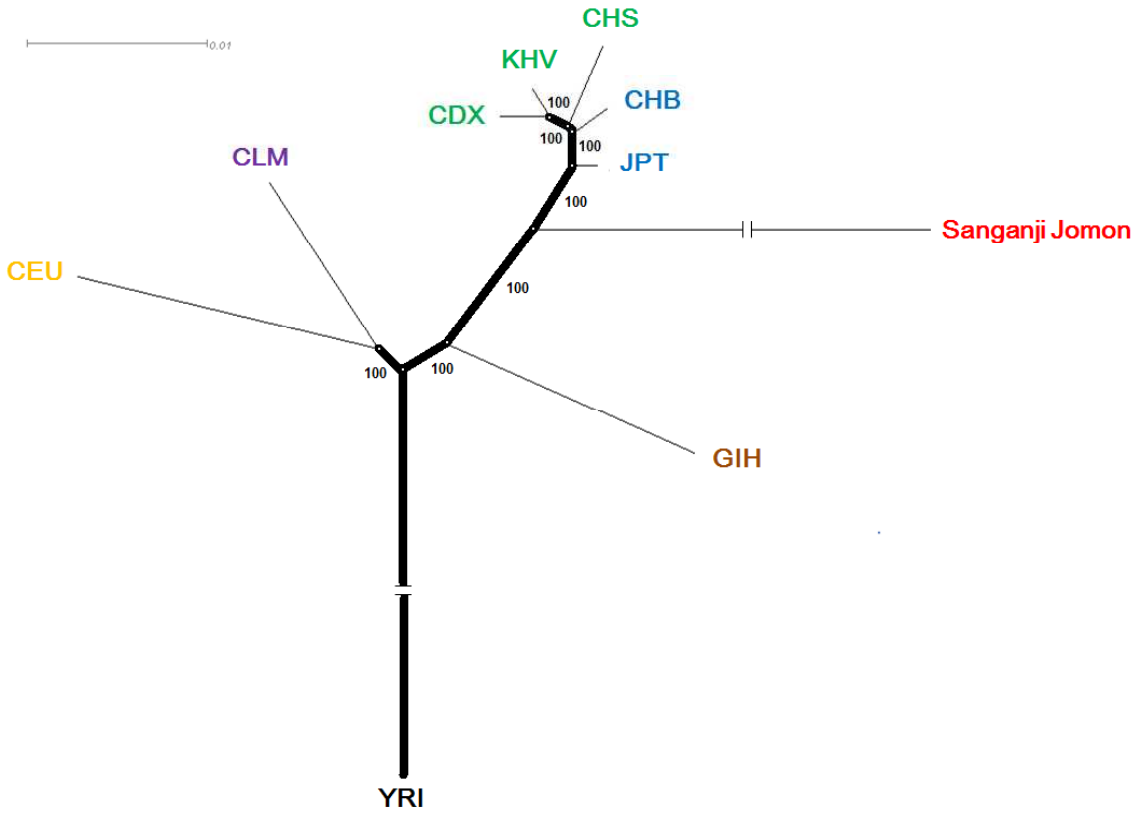


Figure 3.15: A neighbor-joining tree of Sanganji Jomon individual and worldwide modern human populations based on 74,491 SNPs
Color of blue and green indicate northern East Eurasians and southern East Eurasians, respectively.

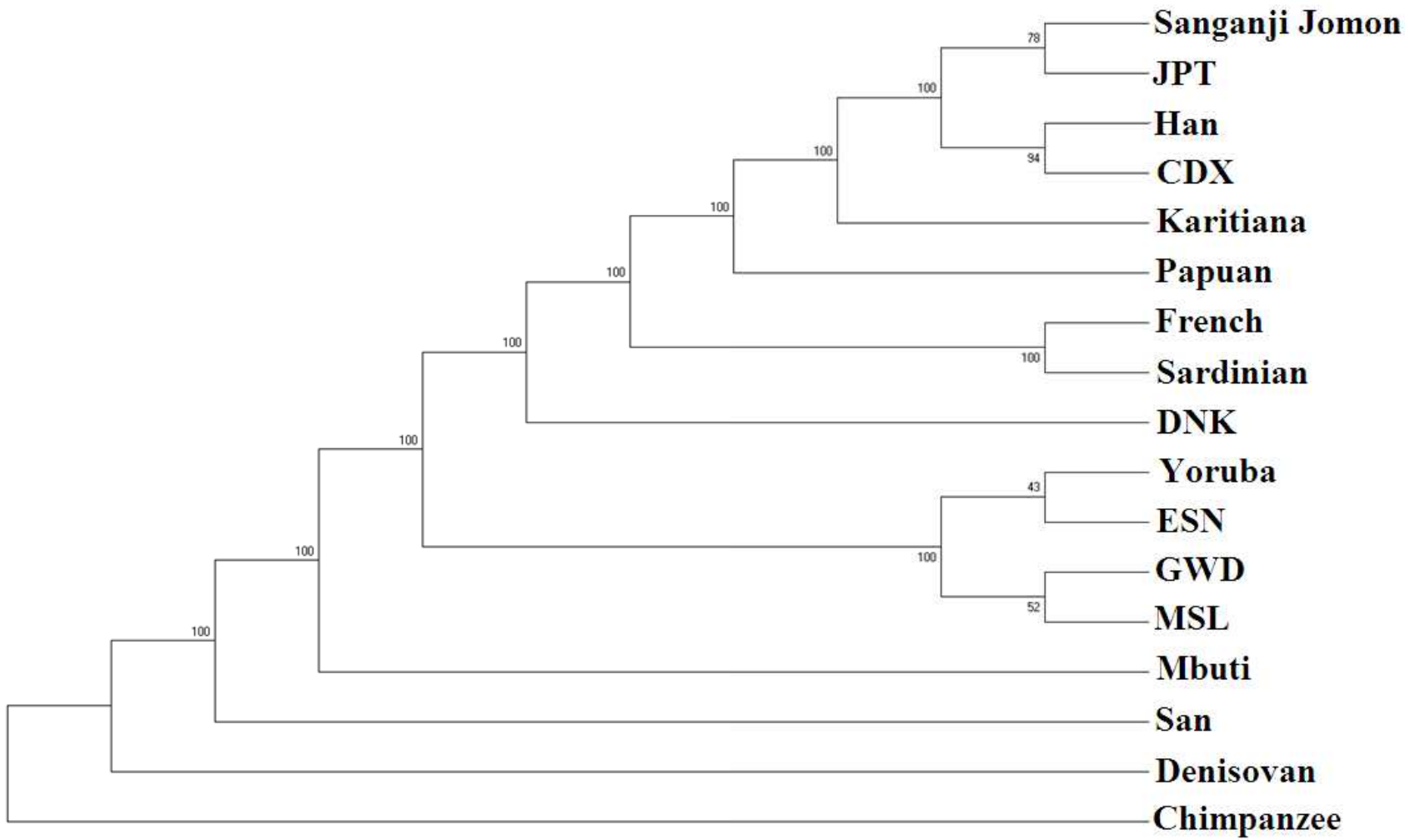


Figure 3.16: A neighbor-joining tree of Sanganji Jomon individual, 14 present-day humans, Denisovan, and Chimpanzee genome based on 481,128 transversion substitution sites

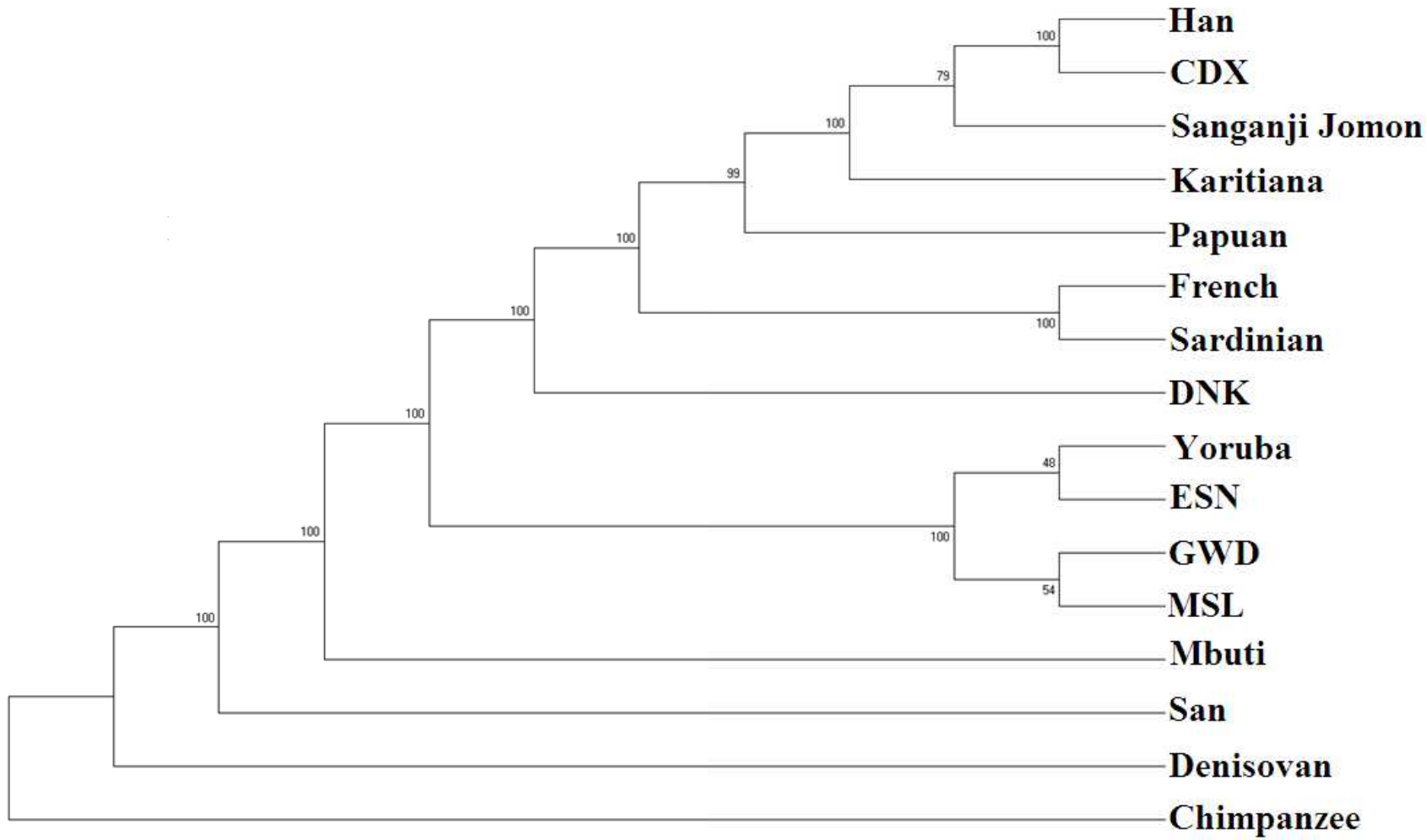


Figure 3.17: A neighbor-joining tree of Sanganji Jomon individual, 13 present-day humans, Denisovan, and Chimpanzee genome based on 481,128 transversion substitution sites

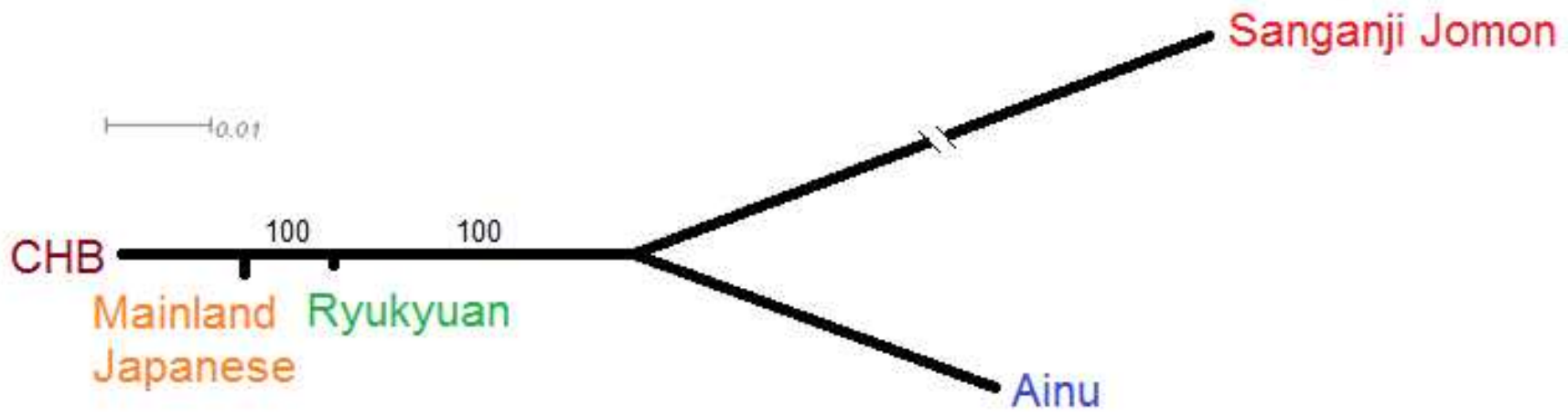


Figure 3.18: A neighbor-joining tree of Sanganji Jomon individual, three populations inhabits in Japanese archipelago, and Han Chinese Beijing based on allele frequencies of 4,600 SNPs

3.3.7 Network analysis

To take multiple human migrations into account for the establishment of East Eurasians and Sanganji Jomon people, I constructed phylogenetic network. I observed the split of (Sanganji Jomon, Japanese; other East Asians) (Figures 3.19, 3.20, 3.21). The split shared among Sanganji Jomon and Japanese suggests that modern Japanese are the admixture of Jomon people and continental populations, which is inferred from the result of PCA. The structures of network based on genomic sequences and genome wide SNPs were basically the same (Figure 3.22). Within European, French were relatively closer to East Eurasians than Sardinian. Within Melanesians, Melanesian was relatively closer to East Eurasians than Papuan.

Interestingly, Karitiana, Native American inhabiting in Brazil, has genetic similarity with Europeans. Since Karitiana is genetically intact from recent migrants from Europe, the genetic similarity among Karitiana and Europeans may suggest ancient gene flow among them. Neither the genetic similarity between European and Sanganji Jomons nor between European and East Eurasians were observed. However, the split separating European and Karitiana from other humans include Tianyuan, ancient East Asian inhabited in Beijing China about 40,000 YBP (Figure 3.23).

I investigated genetic relationship between ancient or modern humans and archaic humans, Neanderthal and Denisovan. Recent studies suggest that ancestors of non-African populations experienced admixture with Neanderthals, while Denisovan contributed 4–6% of its genetic material to the genomes of present-day Melanesians (Green et al., 2010; Rasmussen et al., 2011; Reich et al., 2011; Meyer et al., 2012). I suppose that post-mortem changes in ancient samples skew the topology of phylogenetic network, and it sometimes give false positive or negative results. I adjusted pair-wise distances to weed out the effect of post-mortem changes (Table 3.9), and constructed network tree based on both uncorrected and corrected pair-wise distances to compare the difference of the results. I observed that most of the splits were consistent with each other, although the widths of splits were little bit different. I observed splits not only separating Vindija Neanderthal and Non-African including Sanganji Jomon from Africans but also separating Denisovan and Papuan from other humans (Figures 3.19, 3.20). In

addition, ambiguous but clear splits separating Denisovan, Sanganji Jomon, Japanese, and Dai from other modern humans were observed (Figure 3.20). The splits of blue arrow probably corresponds to the genetic tie between Vindija Neanderthal and non-African, and the splits of red arrow perhaps corresponds to the genetic tie between Denisovan and Papuan, Sanganji Jomon, Japanese, and Dai, respectively. These results imply that Vindija Neanderthal contributed their genetic material to Sanganji Jomon, and Denisovan contributed their genetic material to not only Papuan but also Sanganji Jomon, Japanese, and Dai.

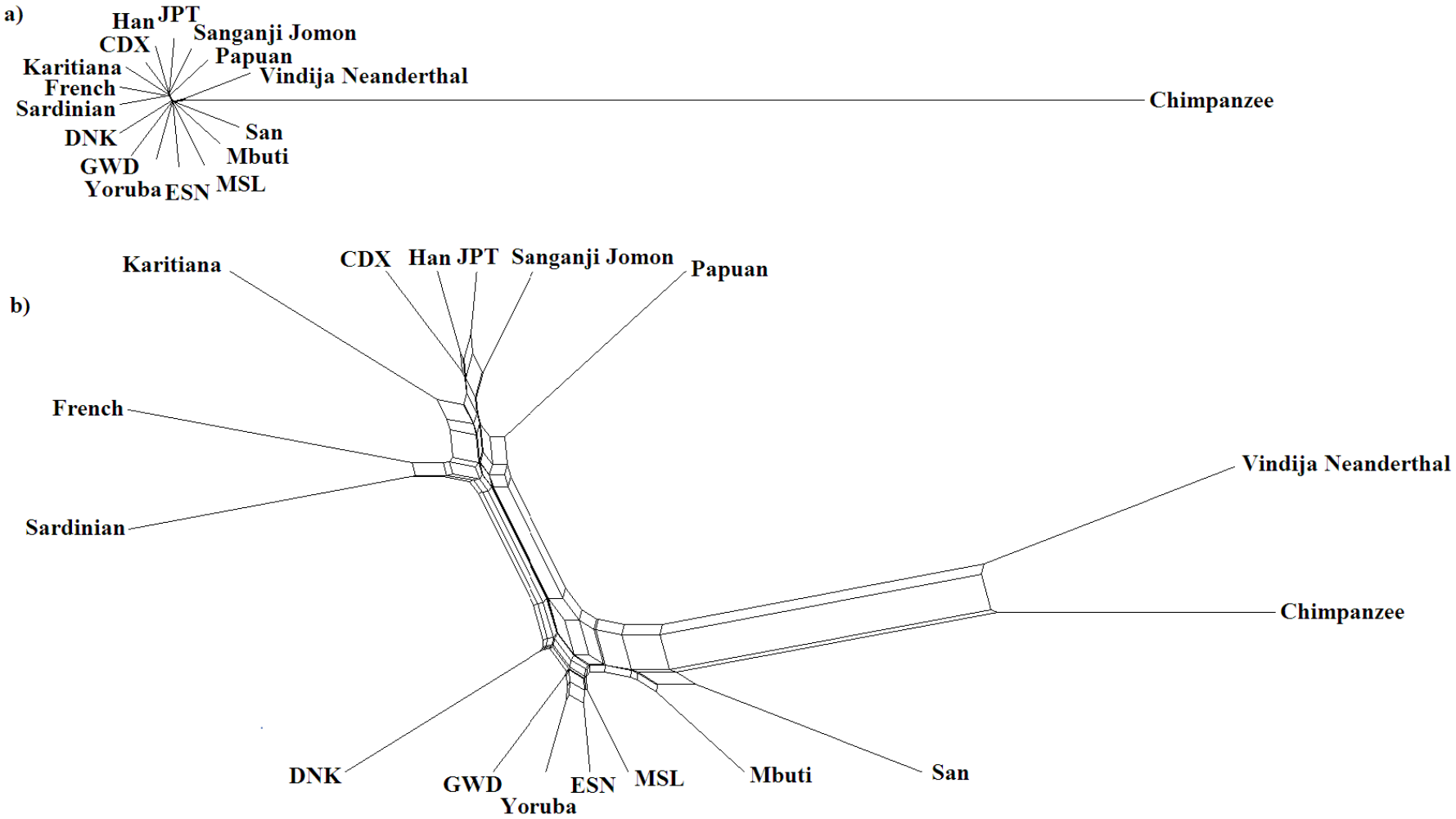


Figure 3.19: A corrected phylogenetic network of Sanganji Jomon individual, 14 present-day humans, Vindija Neanderthal, and Chimpanzee genome based on 481,128 transversion substitution sites

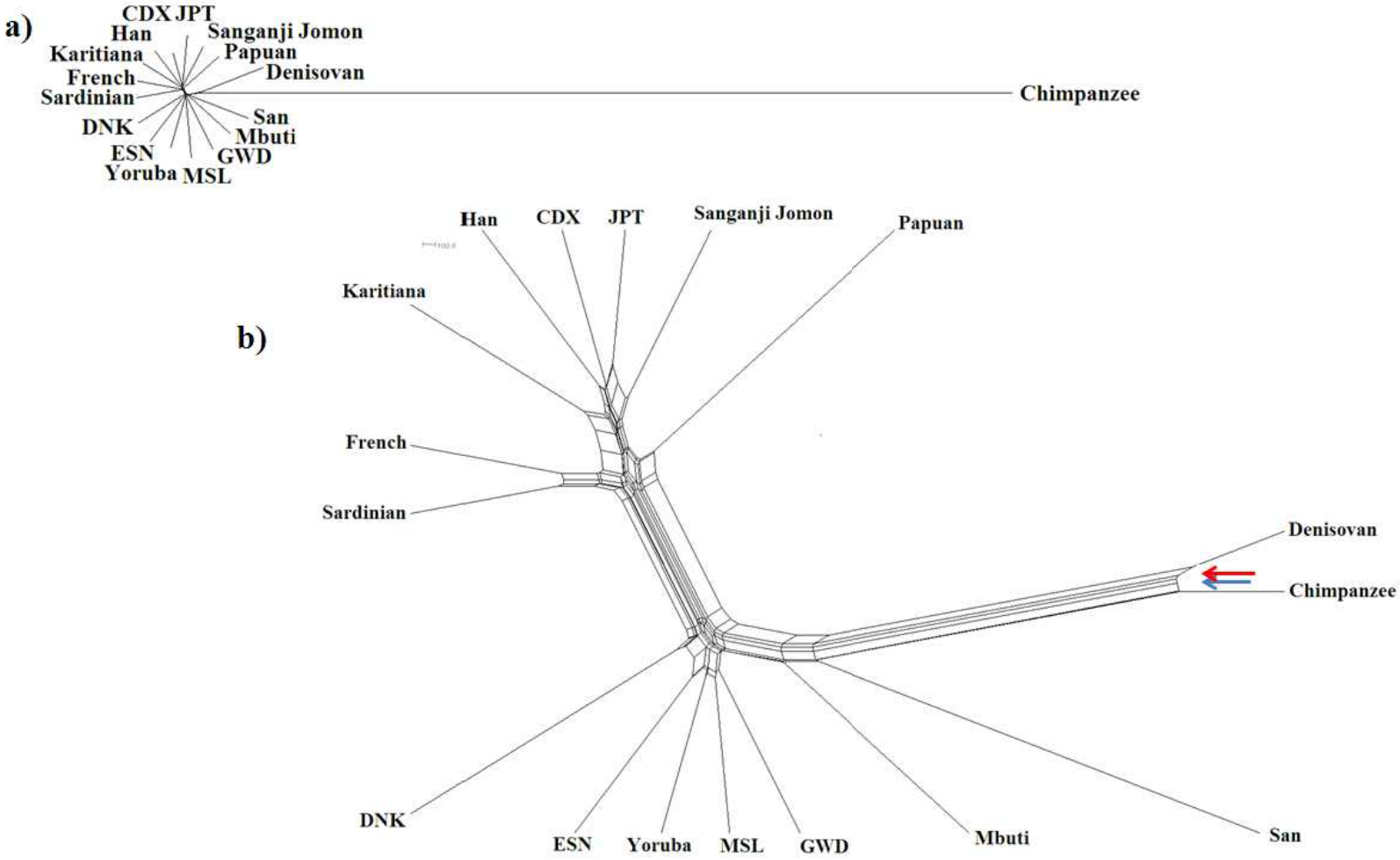


Figure 3.20: A corrected phylogenetic network of Sanganji Jomon individual, 14 present-day humans, Denisovan, and Chimpanzee genome based on 481,128 transversion substitution sites

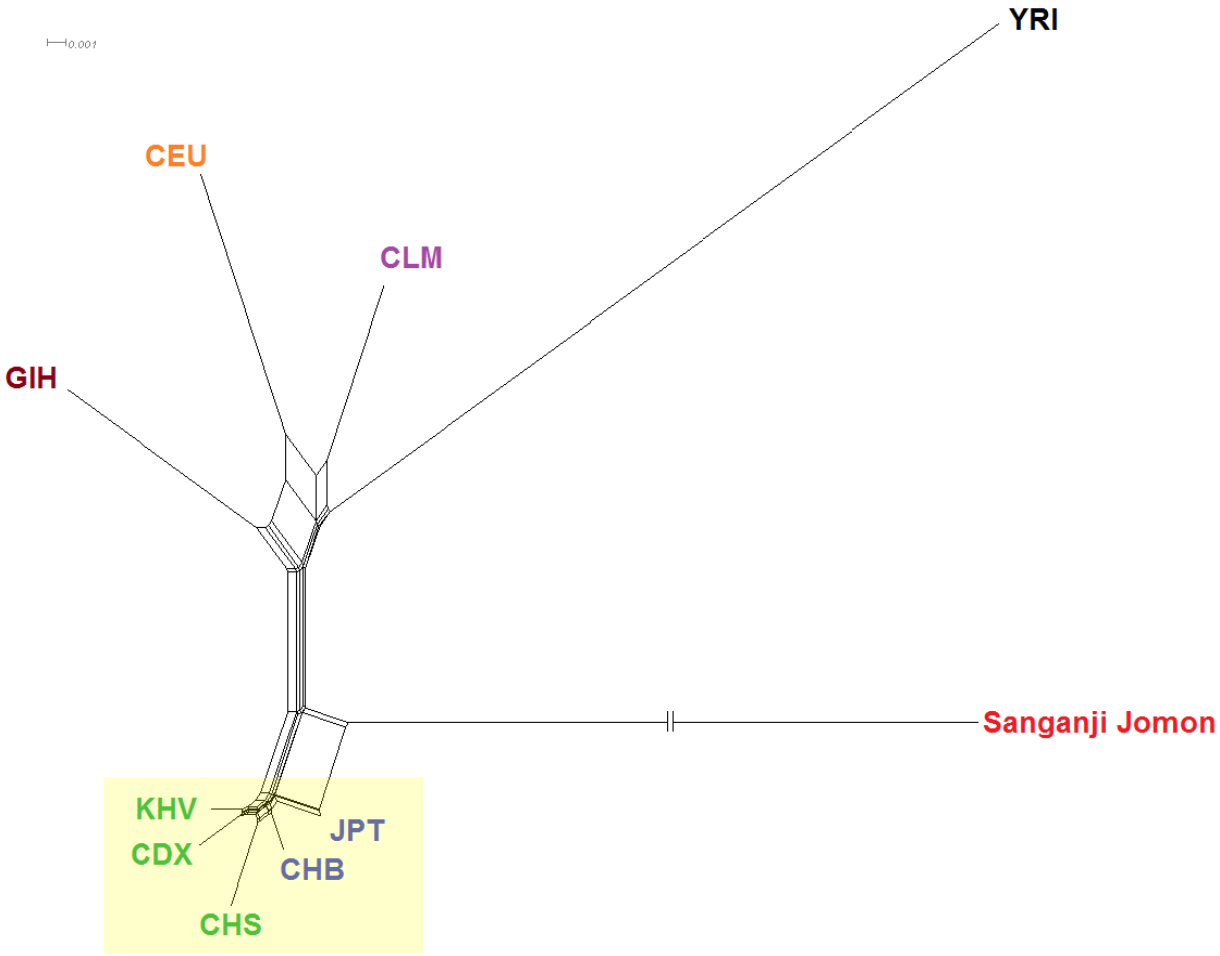


Figure 3.21: A phylogenetic network of Sanganji Jomon individual and worldwide modern human populations based on 74,491 SNPs
Color of blue and green indicate northern East Eurasians and southern East Eurasians, respectively.

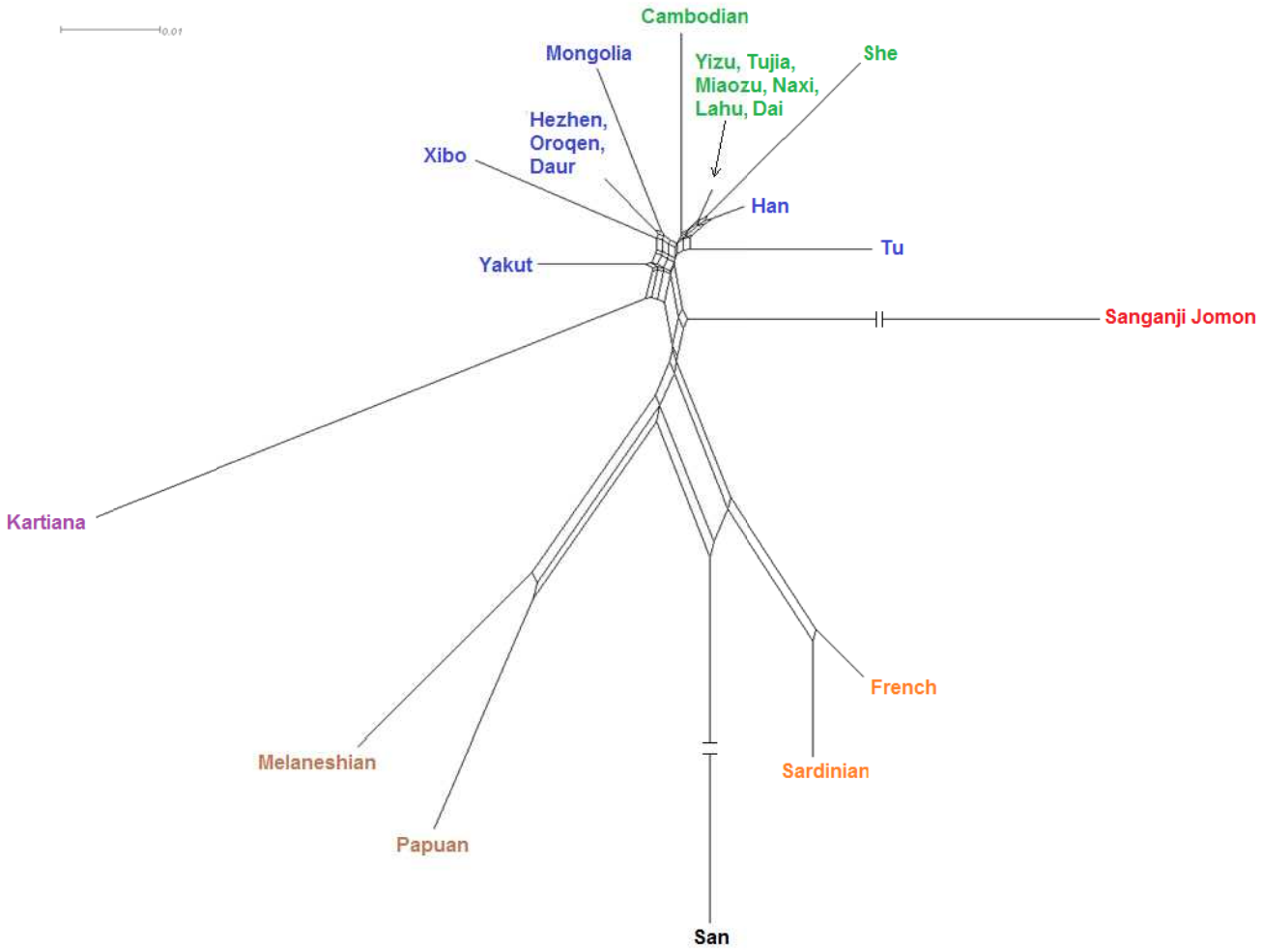


Figure 3.22: A phylogenetic network of Sanganji Jomon individual and modern human populations based on allele frequencies of 4,319 SNPs
Color of blue, green, orange, and brawn indicate northern East Eurasians, southern East Eurasians, Europeans, and Oceanians, respectively.

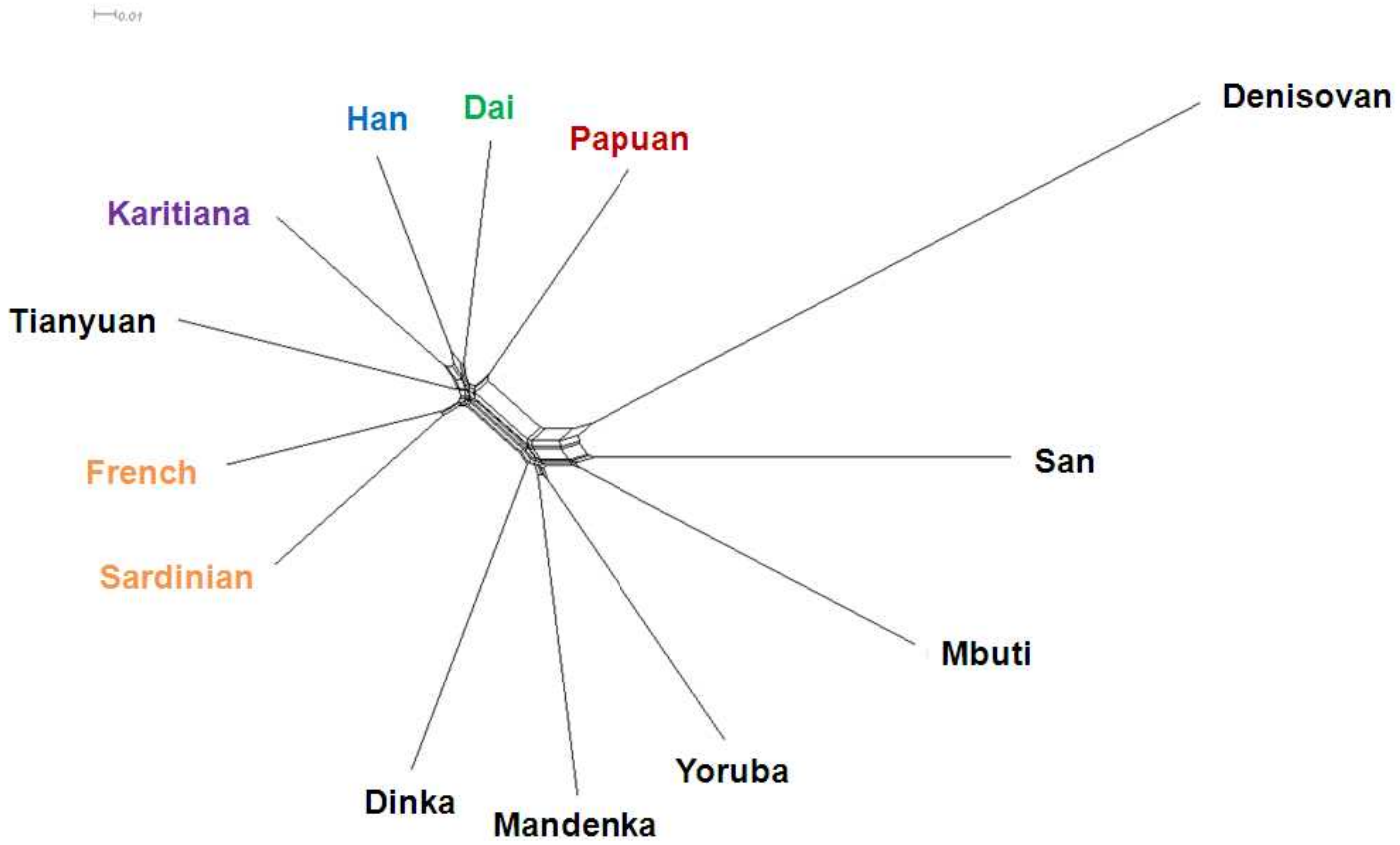


Figure 3.23: A phylogenetic network of Tianyuan individual, archaic Denisovan individual, and 11 modern human individuals based on pairwise distances

Pairwise distances reported in Fu et al. (2013) were used to construct phylogenetic network. Northern East Eurasians were colored with blue, and Southern East Eurasians were colored with green.

Table 3.9: Pair-wise distance in transversion substitution among 17 individuals at 52,454,859 nucleotide sites corrected for post-mortem change

	Pa	San	Yor	Mbu	ESN	GWD	MSL	DNK	Den	S.J.	Fre	Sar	Pap	Han	JPT	CDX
PanTro2 (Pa)	0															
San	223,951	0														
Yoruba (Yor)	222,687	30,544	0													
Mbuti (Mbu)	222,529	30,454	28,576	0												
ESN	223,488	31,581	28,225	29,561	0											
GWD	223,908	32,019	28,785	30,005	29,616	0										
MSL	224,496	32,647	29,401	30,617	30,378	30,748	0									
DNK	222,841	30,794	27,908	28,766	28,613	29,251	29,933	0								
H-Denisovan *1 (Den)	**222,226**	**35,914**	**34,690**	**34,450**	**35,531**	**36,055**	**36,529**	**34,870**	0							
Sanganji Jomon *2 (S.J.)	**222,111**	**30,098**	**27,377**	**28,102**	**28,083**	**28,558**	**29,264**	**27,227**	**33,783**	0						
French (Fre)	222,045	29,876	26,798	27,868	27,779	28,353	28,941	26,776	**33,786**	**23,415**	0					
Sardinian (Sar)	222,103	29,904	26,912	28,044	28,021	28,487	29,127	26,858	**33,944**	**23,693**	22,392	0				
Papuan (Pap)	222,389	30,268	27,434	28,270	28,475	28,755	29,755	27,424	**33,756**	**23,843**	23,862	24,176	0			
Han	222,267	30,060	27,336	28,194	28,181	28,695	29,449	27,226	**34,072**	**22,785**	23,532	23,726	23,702	0		
JPT	224,270	32,203	29,331	30,207	30,006	30,438	31,368	29,061	**36,041**	**24,306**	25,377	25,641	25,619	24,267	0	
CDX	219,899	27,700	24,770	25,704	25,537	26,039	26,739	24,738	**31,558**	**20,233**	20,928	21,260	21,126	19,812	21,725	0
Karitiana	222,634	30,531	27,627	28,479	28,524	29,082	29,828	27,617	**34,485**	**23,426**	23,595	23,785	24,213	23,205	25,152	20,735

*1 High coverage of Denisovan genome

*2 Merged genome dataset with Sanganji 131421-3 and Sanganji 131464

3.3.8 *D*-statistics

In addition to network analysis, I carried out *D*-statistic analysis to detect the gene flow among ancient and modern humans. When I compute $D_{standard}$(Chimpanzee, Africans; Non-Africans, Sanganji Jomons) and $D_{standard}$(Chimpanzee, Europeans; East Asians and Native Americans, Sanganji Jomons), I obtained negative $D_{standard}$-value and a significant $Z_{standard}$-score (Table 3.10). However, when I take the effect post-mortem change into account in the calculation, I obtained positive $D_{revision}$-value and non-significant $Z_{revision}$–score. These results mean that the effect of post-mortem change is so large, and the revision is important.

I tried to investigate the genetic relationship between Sanganji Jomon and archaic humans. When I compute $D_{revision}$(Chimpanzee, Vindija Neanderthal; Africans, Sanganji Jomons and non-Africans), I obtain positive $D_{revision}$-value and a significant $Z_{revision}$-score (Table 3.11). This indicates that Sanganji Jomons as well as present-day non-Africans were genetically closer to Vindija Neanderthal than Africans. Significant $Z_{revision}$(Chimpanzee, Vindija Neanderthal; French or Han, Sanganji Jomons) indicates that Sanganji Jomon were more genetically close to Vindija Neanderthal. This pattern was observed in Papuan and Dai as well as Sanganji Jomon. The phylogenetic network in Figure 3.19 also showed that Papuan and East Eurasians including Sanganji Jomons were genetically closer to Vindija Neanderthal than European and Africans. When I compute $D_{revision}$(Chimpanzee, Denisovan; H2, H3), I observed that Sanganji Jomon and Papuan were significantly and genetically closer to Denisovan than other present-day humans (Table 3.10). The genetic relationship suggests that there were gene flow between Denisovan and the ancestors of not only Papuan but also Sanganji Jomons. Japanese and Dai also had relatively high $Z_{revision}$-score than French, which may correspond to Denisovan gene flow into Japanese and Dai. But the $Z_{revision}$-scores were not significant. For the revision of *Z*-score, I assumed that the frequency of post-mortem change was 0.0456% and 0.1904% in Denisovan and Sanganji Jomon, respectively, and the frequencies strongly correlate to the $Z_{revision}$-scores. I investigated how much do the change of the frequencies change the $Z_{revision}$-scores. I observed that if the P_j was larger than 0.11%, $Z_{revision}$-score of Denisovan-Sanganji Jomons became significant. Therefore, I suppose that Sanganji Jomon retain Denisovan DNA than other East Eurasians.

Another interesting viewpoint is the genetic relationship within Non-African people. I observed that Japanese include many unique derived-alleles (O_{AAAB} or O_{AABA}) than other non-Africans, and share many derived-alleles with Sanganji Jomons (O_{AABB}) (Table 3.10). This indicates that Japanese have high genetic diversity, and the diversity may be caused by the genetic introduction from Jomon people. The feature observed in Japanese is consistent with previous study (The 1000 Genomes Project Consortium, 2012). Network analyses showed the gene flow among French and Karitiana (Figures 3.19, 3.20, 3.22). My *D*-statistic analysis also showed the relationship. This finding is consistent with Patterson et al., (2011).

Table 3.10: Summary of *D*-statistic analysis of Sanganji Jomons, 14 present-day humans, Denisovan, and Chimpanzee

Outgroup	H1	H2	H3	O_{AAAA}	O_{AAAB}	O_{AABA}	O_{ABAA}	O_{ABBB}	O_{AABB}	O_{ABAB}	O_{ABBA}	$D_{standard}$	$Z_{standard}$	SE $(D_{standard}/Z_{standard})$	$D_{revision}$	$Z_{revision}$
(Chimpanzee, Denisovan, African, Non-African)																
Chimpanzee	Denisovan	San	Sanganji_Jomon	50,725,773	111,899	14,256	39,091	203,193	4,606	1,864	1,896	-0.0085	-0.480	0.0177	0.1270	7.1695
Chimpanzee	Denisovan	San	French	50,825,445	12,227	14,244	39,197	203,442	4,618	1,758	1,647	0.0326	1.870	0.0174	0.0595	3.4127
Chimpanzee	Denisovan	San	Sardinian	50,825,381	12,291	14,200	39,218	203,413	4,662	1,737	1,676	0.0179	1.013	0.0177	0.0443	2.5050
Chimpanzee	Denisovan	San	Papuan	50,825,270	12,402	14,262	39,004	203,436	4,600	1,951	1,653	0.0827	4.856	0.0170	0.1094	6.4247
Chimpanzee	Denisovan	San	Han	50,825,238	12,434	14,197	39,201	203,414	4,665	1,754	1,675	0.0231	1.308	0.0177	0.0494	2.7992
Chimpanzee	Denisovan	San	CDX	50,827,603	10,069	14,273	39,200	203,486	4,589	1,755	1,603	0.0453	2.376	0.0191	0.0732	3.8391
Chimpanzee	Denisovan	San	JPT	50,823,155	14,517	14,294	39,211	203,441	4,568	1,744	1,648	0.0283	1.625	0.0174	0.0549	3.1514
Chimpanzee	Denisovan	San	Karitiana	50,824,824	12,848	14,221	39,196	203,386	4,641	1,759	1,703	0.0162	0.897	0.0181	0.0421	2.3291
Chimpanzee	Denisovan	ESN	Sanganji_Jomon	50,726,787	111,308	13,242	39,314	203,376	5,197	1,681	1,673	0.0024	0.129	0.0186	0.1592	8.5556
Chimpanzee	Denisovan	ESN	French	50,826,722	11,573	13,167	39,400	203,605	5,272	1,595	1,444	0.0497	2.701	0.0184	0.0806	4.3788
Chimpanzee	Denisovan	ESN	Sardinian	50,826,350	11,745	13,231	39,422	203,577	5,208	1,573	1,472	0.0332	1.763	0.0188	0.0634	3.3681
Chimpanzee	Denisovan	ESN	Papuan	50,826,211	11,884	13,321	39,191	203,583	5,118	1,804	1,466	0.1034	5.671	0.0182	0.1336	7.3270
Chimpanzee	Denisovan	ESN	Han	50,826,218	11,877	13,217	39,392	203,565	5,222	1,603	1,484	0.0386	2.083	0.0185	0.0685	3.6985
Chimpanzee	Denisovan	ESN	CDX	50,828,701	9,394	13,175	39,415	203,661	5,264	1,580	1,388	0.0647	3.349	0.0193	0.0971	5.0286
Chimpanzee	Denisovan	ESN	JPT	50,824,266	13,829	13,183	39,430	203,620	5,256	1,565	1,429	0.0454	2.474	0.0184	0.0763	4.1593
Chimpanzee	Denisovan	ESN	Karitiana	50,825,852	12,243	13,193	39,403	203,553	5,246	1,592	1,496	0.0311	1.685	0.0185	0.0608	3.2928
(Chimpanzee, Denisovan, Non-African, Non-African)																
Chimpanzee	Denisovan	French	Sanganji_Jomon	50,729,672	110,017	10,357	39,480	203,693	6,488	1,364	1,507	-0.0497	-2.527	0.0197	0.1279	6.5040
Chimpanzee	Denisovan	French	Papuan	50,829,038	10,651	10,494	39,392	203,935	6,351	1,452	1,265	0.0688	3.355	0.0205	0.1044	5.0900
Chimpanzee	Denisovan	French	Han	50,829,058	10,631	10,377	39,598	203,922	6,468	1,246	1,278	-0.0127	-0.604	0.0210	0.0225	1.0691
Chimpanzee	Denisovan	French	JPT	50,827,108	12,581	10,341	39,624	203,965	6,504	1,220	1,235	-0.0061	-0.299	0.0204	0.0300	1.4700
Chimpanzee	Denisovan	French	CDX	50,831,546	8,143	10,330	39,596	203,993	6,515	1,248	1,207	0.0167	0.776	0.0215	0.0545	2.5337
Chimpanzee	Denisovan	French	Karitiana	50,828,802	10,887	10,243	39,639	203,940	6,602	1,205	1,260	-0.0223	-1.055	0.0211	0.0133	0.6284

Table 3.10: Continued

Outgroup	H1	H2	H3	O_{AAAA}	O_{AAAB}	O_{AABA}	O_{ABAA}	O_{ABBB}	O_{AABB}	O_{ABAB}	O_{ABBA}	$D_{standard}$	$Z_{standard}$	$SE_{(D_{standard}/Z_{standard})}$	$D_{revision}$	$Z_{revision}$
(Chimpanzee, African, Non-African, Non-African)																
Chimpanzee	San	French	Sanganji_Jomon	50,754,977	109,665	9,887	14,175	206,083	4,098	1,716	1,977	-0.0707	-4.116	0.0172	0.0330	1.9212
Chimpanzee	San	French	Papuan	50,854,364	10,278	9,910	14,066	206,211	4,075	1,825	1,849	-0.0065	-0.391	0.0166	-	-
Chimpanzee	San	French	Han	50,854,546	10,096	9,893	14,110	206,298	4,092	1,781	1,762	0.0054	0.322	0.0168	-	-
Chimpanzee	San	French	JPT	50,852,519	12,123	9,847	14,213	206,331	4,138	1,678	1,729	-0.0149	-0.860	0.0173	-	-
Chimpanzee	San	French	CDX	50,856,979	7,663	9,824	14,163	206,347	4,161	1,728	1,713	0.0044	0.243	0.0181	-	-
Chimpanzee	San	French	Karitiana	50,854,271	10,371	9,749	14,170	206,306	4,236	1,721	1,754	-0.0095	-0.545	0.0174	-	-
Chimpanzee	San	Han	Sanganji_Jomon	50,755,021	109,418	9,843	14,335	206,262	4,345	1,537	1,817	-0.0835	-4.610	0.0181	0.0302	1.6699
Chimpanzee	San	Han	Papuan	50,854,281	10,158	9,993	14,118	206,282	4,195	1,754	1,797	-0.0121	-0.715	0.0169	-	-
Chimpanzee	San	Han	JPT	50,852,809	11,630	9,557	14,367	206,504	4,631	1,505	1,575	-0.0227	-1.244	0.0182	-	-
Chimpanzee	San	Han	CDX	50,857,248	7,191	9,555	14,341	206,544	4,633	1,531	1,535	-0.0013	-0.070	0.0186	-	-
Chimpanzee	San	Han	Karitiana	50,854,252	10,187	9,768	14,273	206,428	4,420	1,599	1,651	-0.0160	-0.877	0.0182	-	-
(African, Non-African, Non-African, Non-African)																
San	French	Papuan	Sanganji_Jomon	50,951,221	109,354	9,839	9,429	15,967	2,288	2,306	2,174	0.0295	1.853	0.0159	0.0336	2.1074
San	French	Han	Sanganji_Jomon	50,951,683	109,161	9,377	9,600	16,199	2,481	2,074	2,003	0.0174	1.021	0.0170	0.0217	1.2735
San	French	JPT	Sanganji_Jomon	50,949,823	109,027	11,237	9,628	16,376	2,615	1,897	1,975	-0.0201	-1.208	0.0166	-0.0161	-0.9669
San	French	CDX	Sanganji_Jomon	50,954,174	109,152	6,886	9,565	16,286	2,490	1,987	2,038	-0.0127	-0.745	0.0170	-0.0087	-0.5076
San	French	Karitiana	Sanganji_Jomon	50,951,360	109,217	9,700	9,367	16,170	2,425	2,103	2,236	-0.0307	-1.903	0.0161	-0.0270	-1.6752
San	French	Papuan	Han	51,051,085	9,490	9,759	9,478	15,945	2,368	2,257	2,196	0.0137	0.827	0.0166	-	-
San	French	CDX	Han	51,054,138	9,188	6,706	9,654	16,304	2,670	1,898	2,020	-0.0311	-1.803	0.0172	-	-
San	French	Karitiana	Han	51,051,311	9,266	9,533	9,369	16,101	2,592	2,101	2,305	-0.0463	-2.73	0.0170	-	-
San	French	Han	JPT	51,049,728	11,116	9,122	9,585	16,262	2,736	2,089	1,940	0.037	2.259	0.0164	-	-
San	French	CDX	JPT	51,052,158	11,108	6,632	9,546	16,345	2,744	2,006	1,979	0.0068	0.409	0.0166	-	-
San	French	Karitiana	JPT	51,049,270	11,307	9,580	9,365	16,246	2,545	2,105	2,160	-0.0129	-0.754	0.0171	-	-
San	French	Karitiana	CDX	51,053,749	6,828	9,577	9,346	16,200	2,548	2,124	2,206	-0.0189	-1.164	0.0162	-	-

Table 3.10: Continued

Outgroup	H1	H2	H3	O_{AAAA}	O_{AAAB}	O_{AABA}	O_{ABAA}	O_{ABBB}	O_{AABB}	O_{ABAB}	O_{ABBA}	$D_{standard}$	$Z_{standard}$	$SE_{(D_{standard}/Z_{standard})}$	$D_{revision}$	$Z_{revision}$
San	Papuan	Han	Sanganji_Jomon	50,951,360	109,203	9,290	9,923	16,223	2,457	2,032	2,090	-0.0141	-0.837	0.0168	-0.0103	-0.6139
San	Papuan	JPT	Sanganji_Jomon	50,949,494	109,039	11,156	9,957	16,370	2,621	1,885	2,056	-0.0434	-2.586	0.0168	-0.0399	-2.3748
San	Papuan	CDX	Sanganji_Jomon	50,953,841	109,190	6,809	9,898	16,306	2,470	1,949	2,115	-0.0408	-2.34	0.0174	-0.0374	-2.1427
San	Papuan	Karitiana	Sanganji_Jomon	50,950,854	109,218	9,796	9,873	16,153	2,442	2,102	2,140	-0.0089	-0.527	0.0169	-0.0053	-0.3112
San	Papuan	CDX	Han	51,053,923	9,108	6,640	9,869	16,335	2,639	1,978	2,086	-0.0266	-1.581	0.0168	-	-
San	Papuan	Karitiana	Han	51,050,870	9,202	9,693	9,810	16,148	2,545	2,165	2,145	0.0046	0.281	0.0164	-	-
San	Papuan	Han	JPT	51,049,470	11,093	9,063	9,843	16,314	2,684	2,112	1,999	0.0275	1.682	0.0163	-	-
San	Papuan	CDX	JPT	51,051,921	11,110	6,612	9,843	16,422	2,667	2,004	1,999	0.0013	0.076	0.0171	-	-
San	Papuan	Karitiana	JPT	51,048,867	11,205	9,666	9,768	16,219	2,572	2,207	2,074	0.0311	1.897	0.0164	-	-
San	Papuan	Karitiana	CDX	51,053,353	6,719	9,678	9,742	16,188	2,560	2,233	2,105	0.0295	1.816	0.0162	-	-
ESN	French	Papuan	Sanganji_Jomon	50,953,139	109,381	9,975	9,413	14,049	2,304	2,170	2,147	0.0053	0.327	0.0162	0.0084	0.5186
ESN	French	Han	Sanganji_Jomon	50,953,660	109,172	9,454	9,568	14,222	2,513	1,997	1,992	0.0012	0.073	0.0164	0.0045	0.2719
ESN	French	JPT	Sanganji_Jomon	50,951,929	109,068	11,185	9,626	14,270	2,617	1,949	1,934	0.0039	0.228	0.0171	0.0072	0.4191
ESN	French	CDX	Sanganji_Jomon	50,956,268	109,188	6,846	9,558	14,192	2,497	2,027	2,002	0.0062	0.372	0.0167	0.0094	0.5667
ESN	French	Karitiana	Sanganji_Jomon	50,953,372	109,257	9,742	9,364	14,158	2,428	2,061	2,196	-0.0317	-1.935	0.0164	-0.0289	-1.7658
ESN	French	Papuan	Han	51,053,032	9,488	9,800	9,367	13,998	2,479	2,216	2,198	0.0041	0.256	0.0160	-	-
ESN	French	CDX	Han	51,056,184	9,272	6,648	9,629	14,258	2,695	1,956	1,936	0.0052	0.286	0.0182	-	-
ESN	French	Han	JPT	51,051,755	11,077	9,242	9,596	14,235	2,725	1,969	1,979	-0.0026	-0.146	0.0178	-	-
ESN	French	CDX	JPT	51,054,335	11,121	6,662	9,609	14,228	2,681	1,976	1,966	0.0025	0.149	0.0168	-	-
ESN	French	Karitiana	JPT	51,051,348	11,281	9,649	9,389	14,168	2,521	2,036	2,186	-0.0356	-2.015	0.0177	-	-
ESN	French	Karitiana	CDX	51,055,795	6,834	9,661	9,385	14,154	2,509	2,040	2,200	-0.0377	-2.208	0.0171	-	-
ESN	French	Karitiana	Han	51,053,267	9,362	9,565	9,356	14,145	2,605	2,069	2,209	-0.0327	-1.964	0.0166	-	-
ESN	Papuan	Han	Sanganji_Jomon	50,953,248	109,151	9,304	9,980	14,335	2,400	2,018	2,142	-0.0298	-1.801	0.0165	-0.0272	-1.6423
ESN	Papuan	JPT	Sanganji_Jomon	50,951,481	109,047	11,071	10,074	14,383	2,504	1,970	2,048	-0.0194	-1.178	0.0165	-0.0166	-1.0092
ESN	Papuan	CDX	Sanganji_Jomon	50,955,831	109,178	6,721	9,995	14,316	2,373	2,037	2,127	-0.0216	-1.292	0.0167	-0.0189	-1.1308
ESN	Papuan	Karitiana	Sanganji_Jomon	50,952,764	109,208	9,788	9,972	14,243	2,343	2,110	2,150	-0.0094	-0.569	0.0165	-0.0066	-0.4000
ESN	Papuan	CDX	Han	51,055,860	9,149	6,539	9,953	14,398	2,555	2,079	2,045	0.0082	0.471	0.0174	-	-
ESN	Papuan	Karitiana	Han	51,052,741	9,231	9,658	9,882	14,277	2,473	2,200	2,116	0.0195	1.191	0.0164	-	-
ESN	Papuan	Han	JPT	51,051,389	11,010	9,139	9,962	14,395	2,565	2,036	2,082	-0.0112	-0.685	0.0164	-	-
ESN	Papuan	CDX	JPT	51,053,959	11,050	6,569	9,985	14,384	2,525	2,047	2,059	-0.0029	-0.177	0.0164	-	-
ESN	Papuan	Karitiana	JPT	51,050,824	11,148	9,704	9,913	14,262	2,427	2,169	2,131	0.0088	0.556	0.0158	-	-
ESN	Papuan	Karitiana	CDX	51,055,311	6,661	9,698	9,869	14,230	2,433	2,213	2,163	0.0114	0.663	0.0172	-	-

Table 3.11: Summary of *D*-statistic analysis of Sanganji Jomons, 14 present-day humans, Vindija Neanderthal, Denisovan, and Chimpanzee

Outgroup	H1	H2	H3	O_{AAAA}	O_{AAAB}	O_{AABA}	O_{ABAA}	O_{ABBB}	O_{AABB}	O_{ABAB}	O_{ABBA}	$D_{standard}$	$Z_{standard}$	$SE_{(D_{standard}/Z_{standard})}$	$D_{revision}$	$Z_{revision}$
(Chimpanzee, Vindija Neanderthal, African, Non-African)																
Chimpanzee	Vindija Neanderthal	San	Sanganji Jomon	23842134	45777	5757	11808	89800	1587	804	712	0.0611	1.915	0.0319	0.1925	6.0322
Chimpanzee	Vindija Neanderthal	San	French	23882970	4941	5793	11874	89882	1551	738	630	0.079	2.805	0.0282	0.0952	3.3805
Chimpanzee	Vindija Neanderthal	San	Sardinian	23883037	4874	5759	11865	89858	1585	747	654	0.0665	2.272	0.0293	0.0821	2.8036
Chimpanzee	Vindija Neanderthal	San	Papuan	23882881	5030	5768	11823	89880	1576	789	632	0.1105	4.076	0.0271	0.1266	4.6707
Chimpanzee	Vindija Neanderthal	San	Han	23882876	5035	5734	11880	89855	1610	732	657	0.054	1.893	0.0285	0.0696	2.4392
Chimpanzee	Vindija Neanderthal	San	JPT	23881825	6086	5771	11866	89873	1573	746	639	0.0774	2.784	0.0278	0.0931	3.3490
Chimpanzee	Vindija Neanderthal	San	CDX	23883892	4019	5771	11861	89934	1573	751	578	0.1302	4.821	0.0270	0.1479	5.4779
Chimpanzee	Vindija Neanderthal	San	Karitiana	23882742	5169	5755	11904	89871	1589	708	641	0.0497	1.744	0.0285	0.0656	2.3030
Chimpanzee	Vindija Neanderthal	ESN	Sanganji Jomon	23842293	45553	5598	11863	89882	1811	722	657	0.0477	1.422	0.0335	0.1920	5.7225
Chimpanzee	Vindija Neanderthal	ESN	French	23883166	4680	5597	11918	89953	1812	667	586	0.0647	2.196	0.0295	0.0822	2.7905
Chimpanzee	Vindija Neanderthal	ESN	Sardinian	23883150	4696	5646	11928	89948	1763	657	591	0.0529	1.738	0.0304	0.0703	2.3100
Chimpanzee	Vindija Neanderthal	ESN	Papuan	23882990	4856	5659	11847	89931	1750	738	608	0.0966	3.597	0.0269	0.1134	4.2237
Chimpanzee	Vindija Neanderthal	ESN	Han	23882995	4851	5615	11937	89939	1794	648	600	0.0385	1.374	0.0280	0.0556	1.9841
Chimpanzee	Vindija Neanderthal	ESN	JPT	23881992	5854	5604	11927	89961	1805	658	578	0.0648	2.26	0.0287	0.0823	2.8716
Chimpanzee	Vindija Neanderthal	ESN	CDX	23884021	3825	5642	11893	89993	1767	692	546	0.118	4.259	0.0277	0.1368	4.9391
Chimpanzee	Vindija Neanderthal	ESN	Karitiana	23882912	4934	5585	11948	89942	1824	637	597	0.0325	1.085	0.0300	0.0496	1.6565
(Chimpanzee, Vindija Neanderthal, Non-African, Non-African)																
Chimpanzee	Vindija Neanderthal	French	Sanganji Jomon	23843687	45076	4204	11921	90021	2288	583	599	-0.0129	-0.341	0.0378	0.1487	3.9317
Chimpanzee	Vindija Neanderthal	French	Papuan	23884391	4372	4258	11945	90110	2234	559	510	0.0458	1.435	0.0319	0.0662	2.0735
Chimpanzee	Vindija Neanderthal	French	Han	23884431	4332	4179	11997	90080	2313	507	540	-0.0315	-1.015	0.0310	-0.0124	-0.4000
Chimpanzee	Vindija Neanderthal	French	JPT	23883403	5360	4193	11983	90098	2299	521	522	-0.0009	-0.028	0.0321	0.0187	0.5804
Chimpanzee	Vindija Neanderthal	French	CDX	23885463	3300	4200	11960	90141	2292	544	479	0.0635	2.079	0.0305	0.0854	2.7973
Chimpanzee	Vindija Neanderthal	French	Karitiana	23884352	4411	4145	12036	90111	2347	468	509	-0.042	-1.211	0.0347	-0.0217	-0.6264
Chimpanzee	Vindija Neanderthal	Han	Sanganji Jomon	23843719	44891	4172	11960	90027	2473	577	560	0.0157	0.392	0.0401	0.1904	4.7549
Chimpanzee	Vindija Neanderthal	Han	Papuan	23884373	4237	4276	11958	90090	2369	579	497	0.0762	2.332	0.0327	0.0971	2.9709
Chimpanzee	Vindija Neanderthal	Han	JPT	23883486	5124	4110	12067	90149	2535	470	438	0.0353	0.994	0.0355	0.0588	1.6562
Chimpanzee	Vindija Neanderthal	Han	CDX	23885563	3047	4100	12025	90173	2545	512	414	0.1058	3.289	0.0322	0.1312	4.0775
Chimpanzee	Vindija Neanderthal	Han	Karitiana	23884343	4267	4154	12058	90100	2491	479	487	-0.0083	-0.243	0.0342	0.0130	0.3811

Table 3.11: Continued

Outgroup	H1	H2	H3	O_{AAAA}	O_{AAAB}	O_{AABA}	O_{ABAA}	O_{ABBB}	O_{AABB}	O_{ABAB}	O_{ABBA}	$D_{standard}$	$Z_{standard}$	$SE_{(D_{standard}/Z_{standard})}$	$D_{revision}$	$Z_{revision}$
(Chimpanzee, Denisovan, African, Non-African)																
Chimpanzee	Denisovan	San	Sanganji Jomon	23837262	45873	5746	16680	89561	1826	708	723	-0.0104	-0.378	0.0275	0.1338	4.8642
Chimpanzee	Denisovan	San	French	23878133	5002	5760	16711	89621	1812	677	663	0.0105	0.375	0.0280	0.0387	1.3833
Chimpanzee	Denisovan	San	Sardinian	23878202	4933	5736	16700	89607	1836	688	677	0.0081	0.297	0.0273	0.0358	1.3111
Chimpanzee	Denisovan	San	Papuan	23878082	5053	5742	16622	89626	1830	766	658	0.0759	2.868	0.0265	0.1043	3.9417
Chimpanzee	Denisovan	San	Han	23878046	5089	5721	16710	89614	1851	678	670	0.006	0.211	0.0284	0.0338	1.1914
Chimpanzee	Denisovan	San	CDX	23879053	4082	5733	16700	89668	1839	688	616	0.0553	1.926	0.0287	0.086	2.9962
Chimpanzee	Denisovan	San	JPT	23876982	6153	5751	16709	89625	1821	679	659	0.015	0.545	0.0275	0.0431	1.5643
Chimpanzee	Denisovan	San	Karitiana	23877942	5193	5734	16704	89622	1838	684	662	0.0164	0.583	0.0281	0.0446	1.5863
(Chimpanzee, Denisovan, Non-African, Non-African)																
Chimpanzee	Denisovan	French	Sanganji Jomon	23838798	45095	4210	16810	89705	2604	564	593	-0.0249	-0.823	0.0303	0.1578	5.2154
Chimpanzee	Denisovan	French	Papuan	23879556	4337	4268	16780	89798	2546	594	500	0.0859	2.722	0.0316	0.124	3.9282
Chimpanzee	Denisovan	French	Han	23879561	4332	4206	16867	89785	2608	507	513	-0.0058	-0.175	0.0331	0.0313	0.9431
Chimpanzee	Denisovan	French	JPT	23878520	5373	4213	16866	89796	2601	508	502	0.006	0.2	0.0300	0.0435	1.4505
Chimpanzee	Denisovan	French	CDX	23880574	3319	4212	16849	89831	2602	525	467	0.0585	1.841	0.0318	0.0998	3.1417
Chimpanzee	Denisovan	French	Karitiana	23879502	4391	4174	16886	89818	2640	488	480	0.0084	0.243	0.0346	0.0481	1.3907
Chimpanzee	Denisovan	Han	Sanganji Jomon	23838813	44954	4195	16866	89755	2745	514	537	-0.0218	-0.654	0.0333	0.1846	5.5386
Chimpanzee	Denisovan	Han	Papuan	23879525	4242	4299	16806	89818	2641	574	474	0.0954	2.905	0.0328	0.1356	4.1294
Chimpanzee	Denisovan	Han	JPT	23878617	5150	4116	16936	89860	2824	444	432	0.0137	0.395	0.0347	0.0578	1.6660
Chimpanzee	Denisovan	Han	CDX	23880678	3089	4108	16910	89886	2832	470	406	0.073	2.09	0.0349	0.1210	3.4631
Chimpanzee	Denisovan	Han	Karitiana	23879495	4272	4181	16906	89832	2759	474	460	0.0149	0.428	0.0348	0.0567	1.6277

3.4 Discussion

3.4.1 History of Sanganji Jomon people

I observed that Sanganji Jomons were the member of modern East Eurasians (Figures 3.5, 3.6, 3.7). Morphological evidence suggests that Jomon people originated from Southeast Asia (Turner, 1987, 1990; Hanihara, 1991; Matsumura, 2007; Matsumura et al., 2009, Yamaguchi, 1999), while archaeology, genetics, and recent studies of cranial morphology suggest that the Jomon people were of northern origin (Nei, 1995; Imamura, 1996; Omoto and Saitou, 1997; Hanihara and Ishida, 2009; Nakashima et al., 2010). To elucidate their origin, I carried out PCA and phylogenetic analysis, and investigated whether Sanganji Jomon phylogenetically cluster with modern Northeast Asians or Southeast Asians. Interestingly, within East Eurasians, Sanganji Jomons were genetically apart from all other East Eurasians, and they didn't cluster with both modern Northeast Asians and Southeast Asians in PCA. These isolations from modern East Eurasians may indicates that Sanganji Jomons had been genetically isolated from continental East Eurasians for long time, or the genetic background of modern East Eurasians dramatically changed after that the Neolithic period started. I think that the former idea is correct, but it is necessary to sequence the ancient genome of Paleolithic and Neolithic continental people to answer which idea is correct.

The phylogenetic tree shows that Sanganji Jomons diverged from the ancestors of modern East Eurasians (Figures 3.13, 3.14, 3.16). PCA shows that Sanganji Jomon was located in between Cambodian and Yakut (Figure 3.8), and it looks like that Sanganji Jomons were the admixed populations between Cambodian and Yakut. I think that it is artifact of PCA, because Sanganji Jomons was outgroup of all modern East Eurasians in the phylogenetic tree. Importantly, the phylogenetic relationship implies that modern genetic population structure in East Eurasians was constructed after the divergence between the ancestors of Sanganji Jomon and modern East Eurasians. If not, Sanganji Jomon should be classified into the group of either Northern or Southern Asians in the phylogenetic tree. The results of PCA show that modern Japanese are between Sanganji Jomon and other continental populations, especially modern Han Chinese through different datasets (Figure 3.9). In addition, the result of network analysis shows

the genetic relationship between not only Japanese and Chinese but also Japanese and Sanganji Jomons. This suggests that modern Japanese are the result of admixture between Jomon people and Northeast Asians. The candidate of later genetic source is migrants from the continental East Asia to the Japanese Archipelago during and after the Yayoi period who introduced rice agriculture.

The divergence between Sanganji Jomons and Native Americans were unclear in my phylogenetic analysis, because in genome wide SNP data, Sanganji Jomons were outgroup of both East Eurasians and Native Americans (85.7% bootstrap value, not significant), while in genome sequence data, the divergence between Sanganji Jomons and modern East Eurasians postdated the divergence of Native American from East Eurasians (79% bootstrap value, not significant). It indicates that the events of human migration into or within East Eurasia and into American continents are not simple, and it is necessary to consider more complex history to clear the genetic relationship among them. Network tree, which describe the complexity, shows that Sanganji Jomon was located in the ancestors of both East Asians and Native Americans (Figures 3.19, 3.20, 3.22). The phylogenetic relationship may suggest that Sanganji Jomons diverged from the ancestors of modern East Asians and Native Americans. If this idea is correct, the divergence time between Sanganji Jomon and the ancestors of East Eurasian and Native American would predate 15,000 YBP, because the ancestors of modern Native Americans migrated into the North American continent about that period. Assuming that modern continental Asian populations are the descendants of people from the Early Upper Paleolithic period, the ancestors of 3,000 year old Sanganji Jomon people had migrated into Japanese Archipelago during the Upper Paleolithic period. Interestingly, I observed splits of ((European, Native American) (Sanganji Jomon, East Eurasian)) (Figures 3.19, 3.20, 3.22, 3.23). *D*-statistic analysis also shows not significant but negative *D*-value in *D*(Africans, French; Native American, East Asians) (Table 3.10). This suggests two possibilities: one is the gene flow between Sanganji Jomon and East Eurasians after the first divergence from the ancestry of East Eurasian and Native Americans; another is the gene flow between Native American and European after the split between East Eurasian and Native Americans. The phylogenetic complexity might make ambiguous result of phylogenetic tree.

Gene flow between ancient Europeans and Native American is plausible to explain the result of Figure 3.19 and 3.20. Because Karitiana is genetically intact from the recent immigrants from Europe, the splits were not the result genetic introgression from recent European migrants. Patterson et al. (2012) mentioned the possibility that ancient northern Eurasian population contributed genetic material both to the ancestral population of the Americas, and also to the ancestral population of northern Europe. A draft genome of Siberian boy from 24,000 YBP suggests that western Eurasian genetic signatures in modern-day Native Americans derived not only from post-Columbian admixture, but also from a mixed ancestry of the First Americans (Raghavan et al., 2013). The position of Sanganji Jomon in Figure 3.19 and 3.20 seems to indicate no gene flow with Europeans. So, Sanganji Jomons were mostly genetically derived from Southeast Asia.

I also would like to discuss the genetic relationship between Sanganji Jomons and 40,000 YBP East Asian Tianyuan individual from Beijing, China. Since Fu et al. (2013) reported the pairwise nucleotide differences among Tianyuan, 10 present-day humans, and Denisovan, I constructed a phylogenetic network (Figure 3.23). Tianyuan was located in the root of modern East Asians and Native American. However, surprisingly, I observed a split suggesting the genetic similarity between Tianyuan and Europeans. Raghavan et al. (2013) investigated the genetic relationship between Tianyuan and Europeans based on D-statistical analysis, but no genetic relationships were observed between them. We should be noted that these results are based on only chromosome 21. As Tianyuan has mtDNA haplotype B, which is the typical haplotype in East Asians, Tianyuan is basically a member of East Asians, but the genetic relationship between them is still unclear. This result is based on chromosome 21. So, it is essential to accumulate additional genomic information from Tianyuan.

3.4.2 History of Ainu, modern Japanese, and Ryukyuan

Hanihara's dual structure model is a standard theory to explain the history of modern three populations inhabit in Japanese Archipelago. This theory predicts that mainland Japanese more admixed with agricultural people than Ainu and Ryukyuan in and after the Yayoi period.

Recent genetic studies 'indirectly' support this model (e.g. Yamaguchi-Kabata et al., 2008; Jinam et al., 2012). Our PCA and phylogenetic tree show that the ancestor of Ainu and Ryukyuan were less admixed with migrants from Eurasian continent, and retains more genetic components of the Jomon people (Figures 3.10, 3.17). I also found that the Ainu and Ryukyuan share many alleles with Sanganji Jomon (Supplementary Figure 11). These results are consistent with the dual structure model (Hanihara, 1991; Yamaguchi, 1999) and other recent morphologic and genetic studies (Fukase et al., 2012, Jinam et al., 2012). However, the dual structure model doesn't fully explain the results of PCA, because PC2 of Figure 1c, which is independent to the signal of admixture with Yayoi and later continental migrants, divides Ainu people and Sanganji Jomon. There are three possibilities to explain this pattern. The first possibility is that ancestors of the Ainu people experienced admixture with other populations after the Jomon period, and they are genetically differentiated from Jomon people. The candidate population is Okhotsk people, who inhabited the Northeastern costal area of the Hokkaido region from the 5th to 13th century (Figure 1.3). They were morphologically close to modern Southern Siberians, Nivkhi and Ulchi people (Ishida, 1988, 1996; Kozintsev, 1990, 1992; Komesu et al., 2008). Recent studies suggest that the Okhotsk people merged with descendants of Hokkaido Jomon people and became ancestors of the modern Ainu (Sato et al., 2009; Kaburagi et al., 2010; Kazuta et al., 2011; Dodo et al., 2012). The second possibility is population structure within the Jomon people, whereby the Sanganji Jomon (Tohoku region) and Hokkaido Jomon (ancestors of modern Ainu) may exhibit some degree of genetic heterogeneity (Figure 2.1). Since I observed the heterogeneity of mtDNA between Tohoku Jomon and Hokkaido Jomons (Table 2.9), it is one of plausible reasons. The third possibility is genetic drift in the ancestors of modern Ainu. This intriguing question will hopefully be clarified by genome sequencing of the Hokkaido Jomon and Okhotsk people.

I also tried to discuss more about the genetic relationship between Sanganji Jomons and modern Japanese with mitochondrial DNA. I currently classified the haplotype of Sanganji 131421-3 and 131464 into haplogroup N9b*. The sequences of the two Jomon individuals were rarely observed in the haplotype N9b of modern Japanese (Table 2.7). It may imply that the haplotype N9b in modern Japanese are not from Northern Jomon people, but from other Jomon

people inhabited in different regions (e.g., Southern part of Japanese Archipelago). The sequencing of southern Jomons is essential for more understanding the history of modern Japanese.

3.4.3 Genetic relationship between Sanganji Jomon and archaic humans

Another point of interest is whether the ancestors of Jomon people experienced gene flow with archaic humans. Ancestors of non-African populations experienced admixture with Neanderthals, whereas Denisovan DNA sequences are observed in modern Papuan, Melanesian, Australian Aboriginal, and Southeast Asian islanders but not in modern East Asians and Native Americans (Green et al., 2010; Rasmussen et al., 2011; Reich et al., 2011; Meyer et al., 2012). I tried to investigate the existence of gene flow in Sanganji Jomons with phylogenetic network analysis and *D*-statistic analysis. I found that Non-African individuals including Sanganji Jomons have Vindija Neanderthal DNA, and all the Africans don't have the signal of Neanderthal gene flow. Network analysis demonstrates that there was weak gene flow between Denisovan and Sanganji Jomon, and *D*-statistic analysis show the genetic contribution of Denisovan to Sanganji Jomon at significant level. I observed same pattern in Japanese and Dai, but the $Z_{revision}$-score was not significant in *D*-statistic analysis. However, Jinam et al. (unpublished) shows Ainu and Japanese were significantly and genetically closer to Denisovan than Europeans. So, these are plausible results. Meyer et al. (2012) suggests that excess archaic DNA in East Asians does not appear to be due to Denisovan gene flow into the ancestors of East Asians since the excess archaic material is more closely related to Vindija Neanderthal than to Denisovan. However, our $Z_{revision}$(Chimpanzee, Denisovan; H2, Sanganji Jomon) were higher than $Z_{revision}$(Chimpanzee, Vindija Neanderthal; H2, Sanganji Jomon) (Table 3.11). Furthermore, in PCA analysis, East Eurasians get closer to Denisovan than Europeans (Figure 3.18). I consider that excess archaic DNA in East Asians was from not Vindija Neanderthal but Denisovan. This result is from only one Jomon population, and it is difficult to take the effect of post-mortem change into account to detect gene-flow between archaic humans and Jomon people. To reinforce the existence of gene flow and to overcome the two problems, I need to use other Jomons, who have low post-mortem

changes or /and more genomic information.

My study clarified the genetic relationship among Sanganji Jomons, present-day humans, and archaic humans. The genomic information of Sanganji Jomons was only 2% of total genome, but I cultivated new field in the study of Jomon origin and the history of modern Japanese and modern East Eurasian. Furthermore, the genetic relationship between Sanganji Jomons and Denisovan can also be a hint to understand the history of East Eurasian.

Chapter 4

Nuclear DNA analysis of other Jomon individuals

4.1 Introduction

I discussed the uniqueness of Sanganji Jomon people in previous chapter based on their nuclear genome. But we don't know whether the Sanganji Jomons were genetically a standard of all the Jomon people. The results of Sanganji based on mtDNA haplotyping suggests the heterogeneity of Jomon people among Hokkaido, Tohoku, and Kanto Regions, and it implies that the history of the Jomon people is more complex than previously considered (Chapter 2; Adachi et al., 2011). Morphological evidence suggests that there were phenotypic differences among regions in the early stages of Jomon era, but the differences were canceled in the later stages of Jomon era. Our knowledge of other Jomon peoples is limited in mitochondrial DNA, and we don't know the genetic background of the Jomon people from northern and southern part of Japanese Archipelago. Therefore, further data accumulation and ancient DNA analysis based on larger samples with both adequate temporal control and more extensive geographical regions are necessary to clarify an apparently complex Jomon population history. Especially, it is possible that Jomon people originated from multiple regions. My previous analysis based on nuclear DNA, discussed in Chapter 3, suggests that the origin of Jomon people were basically from Southeast Asia, although it is unclear whether the ancestor of Jomon people passed northeast Asia. The result of Jomon mitochondrial DNA haplotyping implies their Northern origin. However, Japanese Archipelago connected to Eurasia in last glacial maximum at Hokkaido region and Kyusyu region, northern and southern part of Japanese Archipelago, respectively, and it seems that Araya type burin and non-Araya type burin entered into Japanese archipelago from north and south, respectively (Figure 1.2). So, it is thus possible that the two or more genetically different migrants, who contributed to the formation of Jomon people, entered into Japanese Archipelago via both Korean Peninsula and Sakhalin before Jomon era. In addition, the genomic DNA analysis of late Paleolithic humans as well as Neolithic Jomon people also extremely

important to investigate the early stages of human migration into Japanese Archipelago and the genetic relationship between Paleolithic and Neolithic Japanese. It is difficult to obtain the genomic DNA data from Late Paleolithic humans, but it is valuable to investigate the genome sequences of Upper Paleolithic Japanese.

I analyzed four additional Jomon individuals from Shikkariabe, Yukura, Odake, and Daizenno-minami archeological sites, to investigate the genetic relationship among Jomon people inhabited different regions and era. Shikkariabe and Daizenno-Minami Jomon were from late Jomon period, while Yukura and Odake Jomon belongs to early Jomon era, which is important to investigate the genetic difference between early stage and late stage of Jomon people. Furthermore, I analyzed an Upper Paleolithic Ryukyuan from Shiraho-Saonetabaru, Ishigaki Island, Ryukyu Archipelago. Recent studies suggest that Upper Paleolithic Ryukyuan were morphologically distinct from Neolithic Mainland Jomon people (Saso et al., 2011; Suwa et al., 2011). So, analysis of the Upper Paleolithic Ryukyuan may clarify the relationship between them in genetic level.

4.2 Material and Methods

4.2.1 Sampling, DNA extraction, library preparations, sequencing, mapping and contamination estimates

I listed sample informations in Table 4.1. Figure 4.1 show the geographic locations of each sample. Dr. Ken-ichi Shinoda kindly provided these precious Jomon and Paleolithic human DNA samples for my Ph.D. study. It should be noted that Professor Noboru Adachi at Yamanashi University Medical School determined mtDNA haplogroups of Shikkariabe, Yugura, and Shiraho-Saonetabaru DNA samples. For the genome analyses, I mostly used the samples that mitochondrial DNA haplotypes were successfully decided. For Shikkariabe, Yugura, and Shiraho-Saonetabaru, the workflow of DNA extractions were carried out by Professor Noboru Adachi at University of Yamanashi. For Odake and Daizenno-Minami Jomons, I extracted DNA at National Science Museum, Tsukuba. Library preparation, sequencing, mapping to human reference genome, and contamination estimates were the same with Chapter 3.2.

Table 4.1: Sample information of four Jomon individuals and an Upper Paleolithic Ryukyuan

Sites	ID	Dating
Shikkariabe	F12	Middle Jomon (4,000 calBP)
Yugura	LLM3	Early Jomon (about 8,000 calBP)
Daizenno-Minami	5_3	Late Jomon
Odake	32	Early Jomon
Shiraho-Saonetabaru	-	Upper Paleolithic (about 20,000 YBP)

Figure 4.1: Map of the Japanese Archipelago and geographic locations of Shikkariabe, Yugura, Daizenno-Minami, and Odake Jomon, and Shiraho-Saonetabaru investigated in the present study

4.2.2 Phylogenetic analysis on mitochondrial DNA and nuclear DNA

I aligned the mtDNA sequences with MEGA 5.0 (Tamura et al., 2011), and constructed neighbor-Joining tree. I used rCRS (Revised Cambridge Reference Sequence) as outgroup (Andrews et al., 1999). I ignored gap sites for analysis. In nuclear DNA, I used the workflow described in Chapter 3.2.8.

4.2.3 Principal Component Analysis and investigating heterogeneity of populations

I did PCA by using the dataset of HGDP-CEPH, Pan Asian SNP Consortium, and JAHPGC to investigate the genetic relationships among the three Jomons and modern humans (The HUGO Pan-Asian SNP Consortium, et al., 2009). Figure 4.2 shows geographical locations of PASNP populations from Southeast Asia. The data preparation steps were the same with that described in chapter 3.2.6.1. I listed the number of SNPs overlapping between Jomon people and modern humans in Table 4.2. To investigate the relationship among three Jomon individuals, Shikkariabe, Yukura, and Sanganji Jomons, I included two Jomons into one PCA. I also carried out PCA by using Shikkariabe Jomon and Shiraho-Saonetabaru to investigate the genetic relationship between Jomon and Upper Paleolithic Ryukyuan.

To investigate whether Jomon people were heterogeneous among regions, we calculated pair-wise distances of in-group pairs of individuals in Jomon and modern populations. I used genome wide SNP data for the calculation of pair-wise distances. I assumed that each population was homogeneous. I compared the distances with that of other populations. I suppose that the distance among two Jomon individuals was small as well as other populations if Jomon people were heterogeneous.

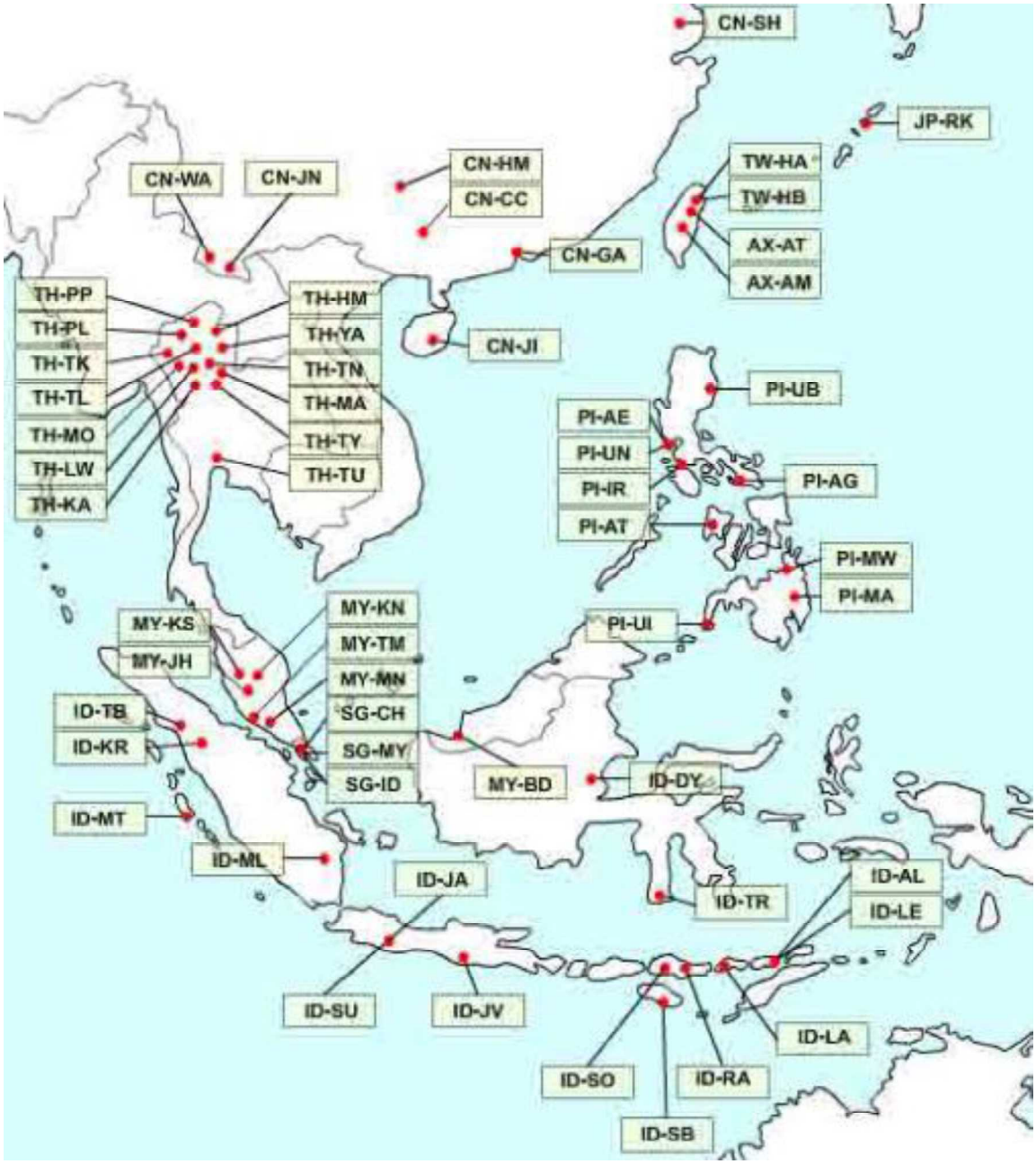


Figure 4.2: Geographical locations of PASNP populations from Southeast Asia
(from Adbulla et al., 2009)

Table 4.2: Summary of the number of overlapping SNPs between ancient samples and three datasets

	Shikkariabe	Yugura	Daizenno-Minami	Odake	Shiraho-Saonetabaru	Shikkariabe & Sanganji Jomon (131421-3(1), 131464(1))	Shikkariabe & Sanganji Jomon (131421-3(2)	Shikkariabe & Yugura	Shikkariabe & Shiraho-Saonetabaru
JAHPGC and HapMap CHB Five Ainu individuals	574,168	62,027	430	-	227	12,022	6,486	56,228	218
HGDP-CEPH	563,824	62,355	607	-	209	10,033	-	56,481	-
Pan-Asian SNP Consortium	46,035	-	-	-	-	-	-	-	-
Vindija Neanderthal, high coverage of Denisovan and 14 present-day humans	-	-	-	-	-	221,750	-	-	-

I examined the genetic relationship between Shikkariabe Jomon and archaic humans with PCA. Protocol of the PCA is different from a standard PCA. I used HGDP-CEPH datasets for the analysis. Since the depth of genome is low, I randomly chose one allele in each SNP sites and duplicated it to prepare homozygous diploid datasets in each present-day humans and ancient humans. I first carried out a PCA on chimpanzee, Neandertal and Denisova, without using data from Jomon and present-day humans at all. Using the SNP weights from the PCA, I can then project the Jomon and present-day humans onto the plane of PC1 and PC2. This allows us to explore the relationship of diverse present-day humans relative to archaic humans and chimpanzee, and to test if the genetic differences among present-day humans are correlated to the differences among these non-modern humans. I also considered the effect of post-mortem changes in the analysis, but the effect is dramatically small and mostly negligible. I also investigated the genetic relationship between Ainu people and archaic humans with same analysis without using Shikkariabe Jomon individual.

4.2.4 Network analysis of nuclear genome

I constructed phylogenetic network based on the genomic sequence of Shikkariabe Jomon, Sanganji Jomon, 14 present-day humans, two archaic humans, and Chimpanzee. I used the workflow described in chapter 3.3.7.

4.2.5 *D*-statistic analysis

I calculated $D_{standard}$, $Z_{standard}$, $D_{revision}$ and $Z_{revision}$ with (3.2.8.9) and (3.2.8.10). I used genome sequence of Shikkariabe Jomon, Sanganji Jomons, 14 present-day humans, Denisovan, Vindija-Neanderthal, and Chimpanzee. I used the nucleotide positions which all the 19 individuals have, and only transversion substitution is used for the analysis. Since I sequenced 80% of Shikkariabe Jomon genome, I also tried the *D*-statistic analysis based on genome wide SNP data, HGDP-CEPH. I also considered the effect of post-mortem changes in the dataset, but the effect is dramatically small and mostly negligible.

4.3 Results

4.3.1 Sequencing of four Jomon individual genomes

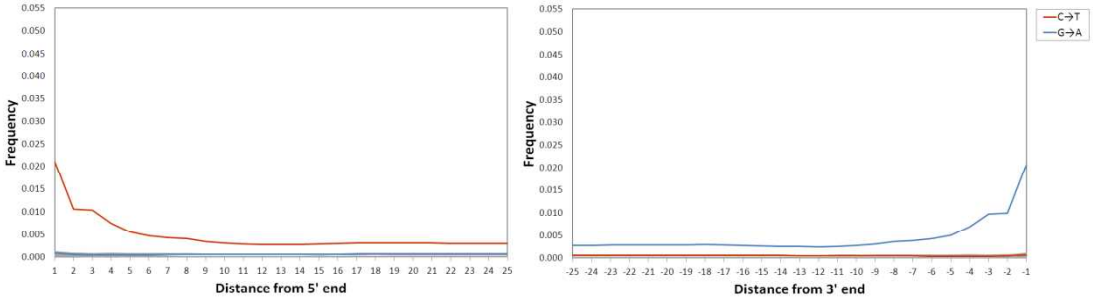
From five ancient samples, I obtained 455,403,880 sequence reads (Table 4.3). About 13% of the total reads were from an individual in Yugura Jomon, while 71.71% of the sequence reads were from an individual in Shikkariabe Jomon. Coverage of the human genome was 9.55% and 79.86%, respectively. On the other hand, the frequencies of human DNA were quite small in Odake Jomon, Daizenno-minami Jomon, and Shiraho-Saonetabaru (0.09-0.15%). It seems that DNAs from the samples inhabited northern regions were well preserved. It indicates that the temperature is one of an important factor for DNA preservation. Although the frequencies are small in later three samples, many of sequence reads remained after the filtering with Mapq30, and the frequency of singleton is also small unlike Sanganji 131421-2(1) and 131464(2). This may indicate that the mapped reads are from humans. With the help of Dr. Kirill Kryukov, I observed C to T misincorporation and the specific cleavage patterns characteristic to ancient DNA (Figure 4.3), and it indicates that all the four samples include Jomon DNA. For the following statistical analysis, I basically used genome sequences determined for Shikkariabe Jomon and Yugura Jomon.

Table 4.3: Summary of genomic sequence data from four Jomon individuals and one Upper Paleolithic Ryukyuan

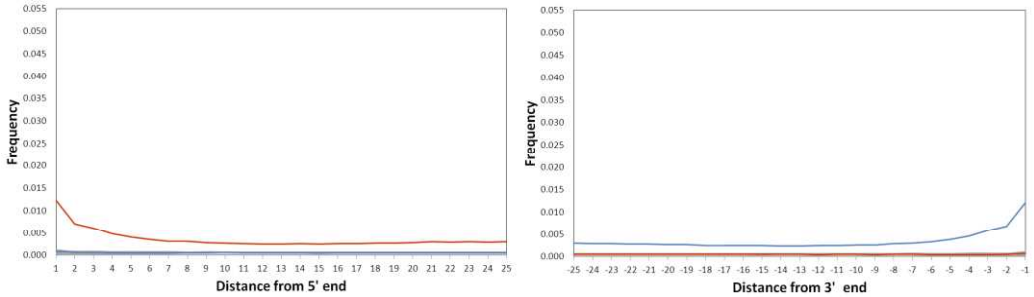
Sample	Shikkariabe	Yugura	Daizenno-Minami	Odake	Shiraho-Saonetabaru
Illumina index	2	19	-	7	4
Total reads	230,983,518	102,461,218	80,304,280	19,695,396	21,959,468
Hits to hg19 (%)	165,645,814 (71.71%)	13,335,221 (13.01%)	117,869 (0.15%)	16,937 (0.09%)	23,498 (0.11%)
Non-restriction enzyme treatment	-	-	60,530 (0.11%)	-	-
restriction enzyme treatment	-	-	57,339 (0.24%)	-	-
Total reads (>= Mapq30) (%)	135,001,878 (58.45%)	11,243,578 (10.97%)	93,001 (0.12%)	13,332 (0.07%)	18,528 (0.08%)
Properly paired (%)	88,904,476 (65.85%)	7,600,237 (67.70%)	17,422 (18.73%)	6,160 (46.20%)	12,366 (66.74%)
With intself and mate mapped	132,416,738	11,149,482	91,216	11,433	16,938
Singletons (%)	2,585,027 (1.91%)	94,091 (0.84%)	1,785 (1.92%)	1,899 (14.24%)	1,590 (8.58%)
With mapped to a different chr	730,902	39,963	53	42	29
Hits after rmdup (%)	107,774,742 (46.66%)	5,563,726 (5.43%)	64,583 (0.08%)	10,580 (0.05%)	12,903 (0.06%)
Bases covering hg19 (only autosome) (%)	2,300,762,201 (79.86%)	275,115,295 (9.55%)	2,413,028 (0.08%)	613,338 (0.02%)	820519 (0.03%)
Coverage of mtDNA (%)	16568 (99.99%)	16275 (98.23%)	1,629 (9.83%)	930 (5.6%)	3969 (23.95%)
Sequence reads mapped to X-chromosome *1	2,013,244	90,264	1,101	133	211
Sequence reads mapped to Y-chromosome *1	1,631	67	4	2	2
Relative frequency of base pair mapped to X and Y (X : Y)	1234.4 : 1	1347.2 : 1	275.3 : 1	66.5 : 1	105.5 : 1
Sex identification based on X : Y	Female	Female	Female	Female?	Female
Sex identification based on morphological study	-	-	-	-	(Female?)

*1 For sex identification and contamination estimates, we conuted only the numbers of pairwise-forward reads.

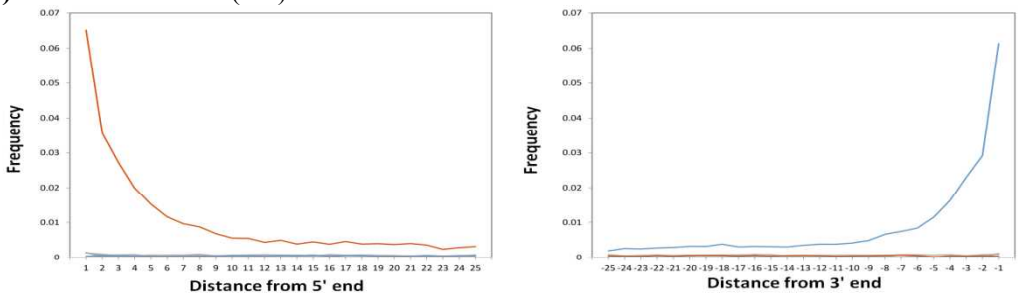
a) Shikkariabe (F12)



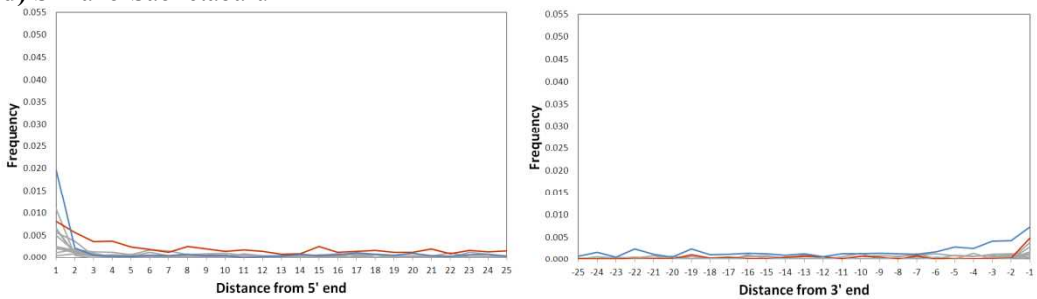
b) Yugura (LLM3)



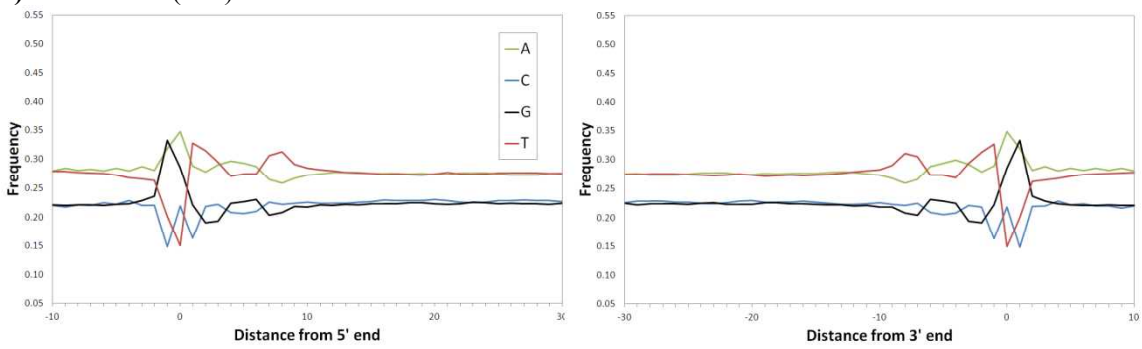
c) Daizenno-Minami (5-3)



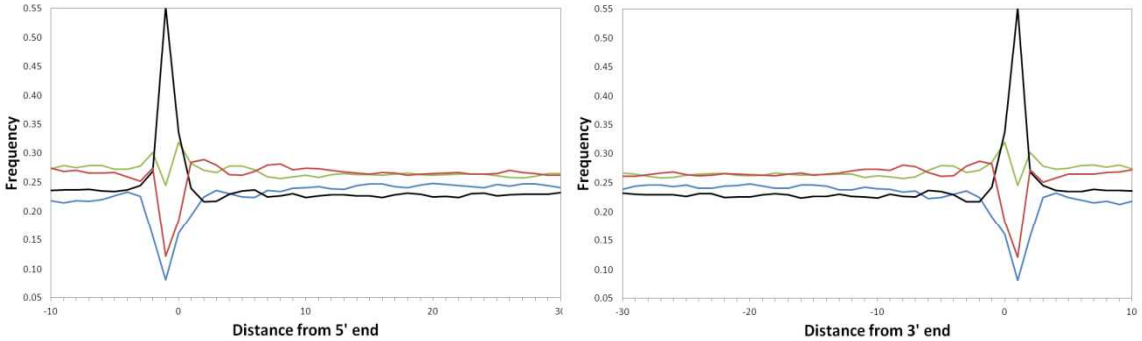
d) Shiraho-Saonetabaru



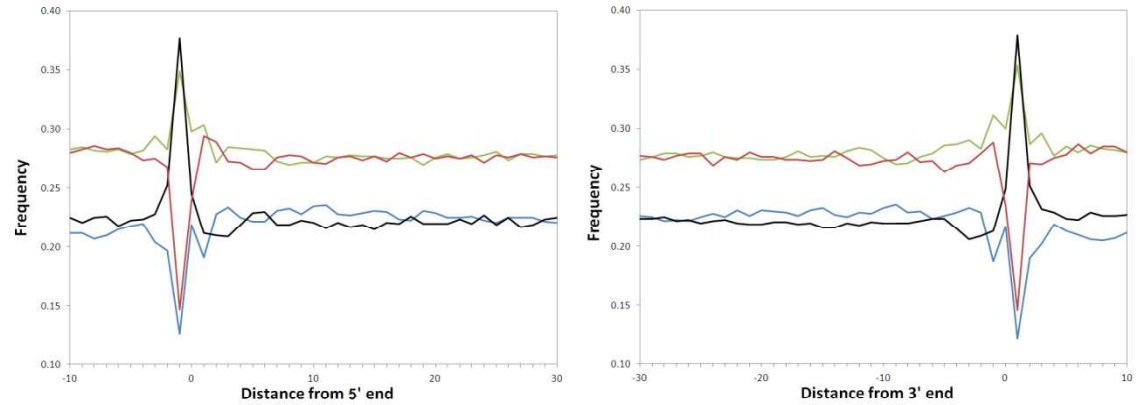
e) Shikkariabe (F12)



f) Yugura (LLM3)



g) Daizenno-Minami (5-3)



h) Shiraho-Saonetabaru

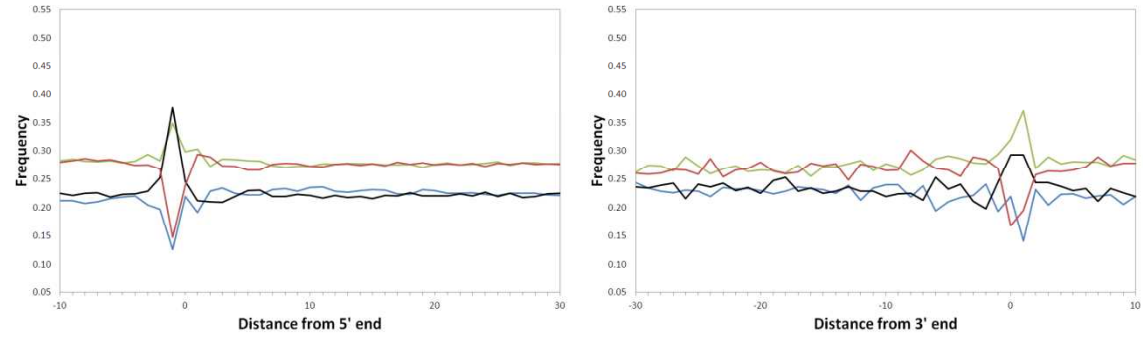


Figure 4.3: Pattern of postmortem misincorporation and depurination in Shikkariabe, Yugura, Daizenno-Minami Jomon and Shiraho-Saonetabaru individuals
a-d) Red line indicates C to T misincorporation, and blue line indicates G to A misincorporation. e-h) Base composition of the human reference genome around the 5'- and 3'-ends of the sequence reads.

4.3.2 MtDNA haplotyping and tree analysis

I determined the mtDNA haplotypes of two individuals. Shikkariabe Jomon individual and Yugura Jomon individual were classified into haplotype D4h2 and D4b2, respectively. Some modern Japanese have the two haplotypes. Since I obtained the complete mtDNA sequences of Shikkariabe Jomon at high sequence depth, I constructed a phylogenetic tree of haplotype D4h2 sequences. The phylogenetic tree suggests that Shikkariabe Jomon individual was basal of haplotype D4h2, or she diverged from other D4h2 sub-groups linking to the lineage of modern Japanese (Figure 4.4). Combination of the mutation in *T961C*, *G8269A*, *T13879C*, *A15236G*, *C16278T!*, and *T16325C* specify haplotype D4h2. However, Shikkariabe Jomon didn't have mutation in *T961C* and *G8269A*. Haplotype D4b2 have sub-haplotype D4b2a and D4b2b, and modern Japanese belongs to either D4b2a or D4b2b (Table 4.4). However, Yugura Jomon was classified into ancestral haplotype, D4b2*.

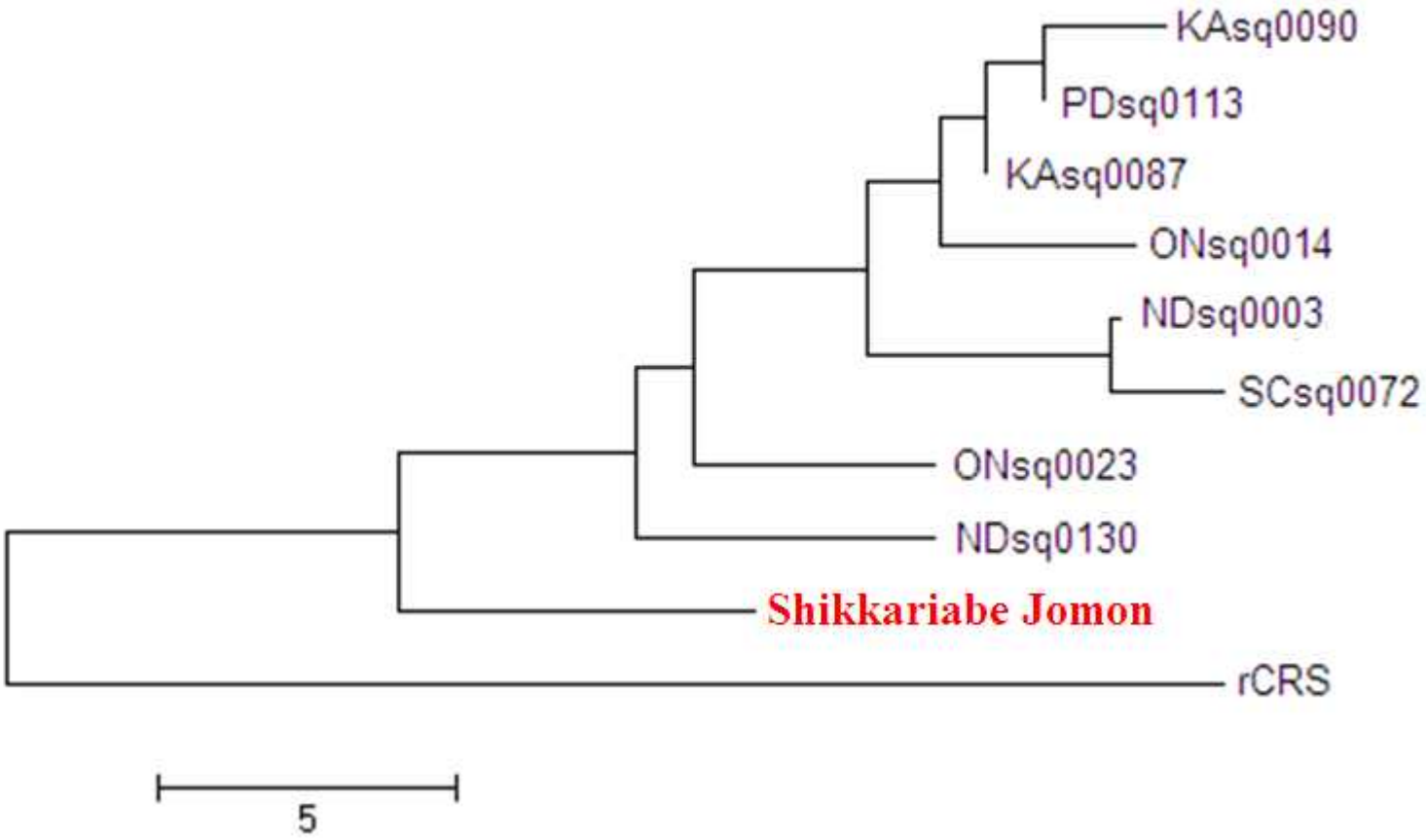


Figure 4.4: P**hylogenetic tree haplotype D4h2 on Shikkariabe Jomon and modern Japanese**
rCRS is revised Cambridge Reference Sequence.

Table 4.4: Summary of mitochondrial DNA haplogroup D4h2 on Yukura Jomon and 68 modern Japanese individuals.

	SNP		146	150	194	199	709	1382	3010	4824	4883	5178	8020	8251	8414	8701	8964	9296	9540	9624	10104	10398	10400	10873	12358	12705	13708	14470	14905	14568	14783	15343	15301	15324	16183	16187	16189	16223	16227	16278	16290	16291	16311	16344	16355	16362	16519
	rCRS	Haplotype	T	C	C	T	C	A	C	A	C	C	G	G	C	A	C	C	T	T	C	A	C	T	A	C	C	T	A	C	T	C	C	A	A	C	T	C	A	C	C	C	T	C	C	T	T
Ancient human	**LLM3 (Yukura Jomon)**	MD4b2*	T	C	C	T	G	C	^	^	T	^	^	G	T	G	T		C	^	C	G	T	C	^	T	G	T	^	T	C	^	^	^	^	C	T	T	^	C	C	C	T	C	C	C	T
Modern human	AT008253	MD4b2	T	C	C	C	G	C	A	G	T	A	A	A	T	G	T	C	C	A	T	G	T	T	A	T	G	T	A	T	C	A	A	A	A	C	T	T	G	T	C	C	T	C	C	C	T
	AT008858	MD4b2a	T	C	C	T	G	C	A	A	T	A	A	G	T	G	T	C	C	A	C	G	T	C	A	T	A	T	A	T	C	A	A	G	A	C	T	T	A	C	C	C	T	C	T	C	T
	AT008838	MD4b2a	T	C	C	C	G	C	A	A	T	A	A	A	T	G	T	C	C	A	T	G	T	C	A	T	C	T	A	T	C	A	A	G	A	C	C	T	A	C	C	C	T	C	C	C	T
	AT008804	MD4b2a	T	C	C	T	G	C	A	A	T	A	A	A	T	G	T	C	C	A	T	G	T	C	A	T	G	T	A	T	C	A	A	G	A	C	T	T	A	C	C	C	T	C	C	C	T
	AT008709	MD4b2a	T	C	C	C	G	C	A	A	T	A	A	A	T	G	T	C	C	A	T	G	T	C	A	T	G	T	A	T	C	A	A	G	C	C	C	T	A	C	C	C	C	C	C	C	T
	AT008703	MD4b2a	T	C	C	T	G	C	A	A	T	A	A	A	T	G	T	C	C	A	T	G	T	C	A	T	G	T	A	T	C	A	A	G	A	C	C	T	A	C	C	C	T	C	C	C	T
	AT008685	MD4b2a	T	C	C	T	G	C	A	A	T	A	A	G	T	G	T	C	C	A	C	G	T	C	A	T	A	T	A	T	C	A	A	G	A	C	T	T	A	C	C	C	T	C	T	C	T
	AT008676	MD4b2a	T	C	C	T	G	C	^	^	T	^	^	G	T	G	T	C	C	^	C	G	T	C	^	T	A	T	^	T	C	^	^	G	^	C	T	T	^	C	C	C	T	C	C	C	T
	AT008656	MD4b2a	T	C	C	C	G	C	A	A	T	A	A	A	T	G	T	C	C	A	T	G	T	C	A	T	G	T	A	T	C	A	A	G	A	C	C	T	A	C	C	C	T	C	C	C	T
	AT008634	MD4b2a	T	C	C	T	G	C	A	A	T	A	A	G	T	G	T	C	C	A	C	G	T	C	A	T	A	T	A	T	C	A	A	G	A	C	T	T	A	C	C	C	T	C	T	C	T
	AT008583	MD4b2a	T	C	C	T	G	C	A	A	T	A	A	A	T	G	T	C	C	A	T	G	T	C	A	T	G	T	A	T	C	A	A	G	A	C	C	T	A	C	C	C	T	C	C	C	T
	AT008502	MD4b2a	T	C	C	T	G	C	^	^	T	^	^	G	T	G	T	C	C	^	C	G	T	C	^	T	A	T	^	T	C	^	^	G	^	C	T	T	^	C	T	C	T	C	C	C	T
	AT008463	MD4b2a	T	C	C	T	G	C	A	A	T	A	A	A	T	G	T	C	C	A	T	G	T	C	A	T	G	T	A	T	C	A	A	G	A	C	C	T	A	C	C	C	T	C	C	C	T
	AT008448	MD4b2a	T	C	C	C	G	C	A	A	T	A	A	A	T	G	T	C	C	A	T	G	T	C	A	T	G	T	A	T	C	A	A	G	A	C	C	T	A	C	C	C	T	C	C	C	T
	AT008434	MD4b2a	T	C	C	T	G	C	A	A	T	A	A	G	T	G	T	C	C	A	C	G	T	C	A	T	G	T	A	T	C	A	A	G	A	C	T	T	A	C	C	C	T	C	C	C	T
	AT008303	MD4b2a	T	C	C	T	G	C	A	A	T	A	A	A	T	G	T	C	C	A	T	G	T	C	A	T	C	C	A	T	C	A	A	G	A	C	C	T	A	C	C	C	T	C	C	C	T
	AT008287	MD4b2a	T	C	C	C	G	C	A	A	T	A	A	A	T	G	T	C	C	A	T	G	T	C	A	T	G	T	A	T	C	A	A	G	C	C	C	T	A	C	C	C	T	C	C	C	T
	AT008286	MD4b2a	T	C	C	T	G	C	A	A	T	A	A	G	T	G	T	C	C	A	C	G	T	C	A	T	A	T	A	T	C	A	A	G	A	C	T	T	A	C	T	C	T	C	C	C	T
	AT008277	MD4b2a	T	C	C	T	G	C	A	A	T	A	A	A	T	G	T	C	C	A	T	G	T	C	A	T	G	T	A	T	C	A	A	G	A	C	C	T	A	T	C	C	T	C	C	C	T
	AT008256	MD4b2a	T	C	C	C	G	C	A	G	T	A	A	A	T	G	T	C	C	A	T	G	T	C	A	T	G	T	A	T	C	A	A	G	A	T	T	T	A	C	T	C	C	C	C	C	T
	AT010721	MD4b2a	T	C	C	C	G	C	^	^	T	^	^	^	T	G	T	C	C	^	T	G	T	C	^	T	G	T	^	T	C	^	^	G	^	C	C	T	^	C	C	C	T	C	C	C	T
	AT010741	MD4b2a	T	C	C	C	G	C	A	A	T	A	A	A	T	G	T	C	C	A	T	G	T	C	A	T	G	T	A	T	C	A	A	G	A	C	C	T	A	C	C	C	T	C	C	C	T
	AT010697	MD4b2a	T	C	C	T	G	C	A	A	T	A	A	A	T	G	T	C	C	A	T	G	T	C	A	T	G	T	A	T	C	A	A	G	A	C	T	T	A	C	C	C	T	C	C	C	T
	AT008361	MD4b2b	T	C	C	T	G	C	A	A	T	A	A	G	T	G	T	T	C	A	C	G	T	C	A	T	G	T	A	T	C	A	A	A	A	C	T	T	A	C	C	C	T	C	C	C	T
	AT008480	MD4b2b	T	C	C	T	G	C	^	^	T	^	^	G	T	G	T	T	C	^	C	G	T	C	^	T	G	T	^	T	C	^	^	^	^	C	T	T	^	C	C	C	T	C	C	C	T
	AT008300	MD4b2b	T	C	C	T	G	C	A	A	T	A	A	G	T	G	T	T	C	A	C	G	T	C	A	T	G	T	A	T	C	A	A	A	A	C	T	T	A	C	C	C	T	C	C	C	T
	AT010673	MD4b2b	T	C	C	T	G	C	A	A	T	A	A	G	T	G	T	T	C	A	C	G	T	C	A	T	G	T	A	T	C	A	A	A	A	C	T	T	A	C	C	C	T	C	C	C	T
	AT008908	MD4b2b	T	C	T	T	G	C	A	A	T	A	A	G	T	G	T	T	C	A	C	G	T	C	A	T	G	T	A	T	C	A	A	A	A	C	T	T	A	C	C	C	T	C	C	C	T
	AT008905	MD4b2b	T	C	T	T	G	C	A	A	T	A	A	G	T	G	T	T	C	A	C	G	T	C	A	T	G	T	A	T	C	A	A	A	A	C	T	T	A	C	C	C	T	C	C	C	T
	AT008851	MD4b2b	T	C	T	T	G	C	A	A	T	A	A	G	T	G	T	T	C	A	C	G	T	C	A	T	G	T	A	T	C	A	A	A	A	C	T	T	A	C	C	C	T	C	C	C	T
	AT008795	MD4b2b	T	C	T	T	G	C	A	A	T	A	A	G	T	G	T	T	C	A	C	G	T	C	A	T	G	T	A	T	C	A	A	A	A	C	T	T	A	C	C	C	T	C	C	C	T
	AT008773	MD4b2b	T	C	T	T	G	C	A	A	C	A	A	G	T	G	T	T	C	A	C	G	T	C	A	T	G	T	A	T	C	A	A	A	A	C	T	T	A	C	C	C	T	C	C	C	T
	AT008768	MD4b2b	C	C	T	T	G	C	A	A	T	A	A	G	T	G	T	T	C	A	C	G	T	C	A	T	G	T	A	T	C	A	A	A	A	C	T	T	A	C	C	C	T	C	C	C	T
	AT008757	MD4b2b	T	C	T	T	G	C	^	^	T	^	^	G	T	G	T	T	C	^	C	G	T	C	^	T	G	T	^	T	C	^	^	^	^	C	T	T	^	C	C	C	T	C	C	C	T
	AT008749	MD4b2b	T	C	T	T	G	C	A	A	T	A	A	G	T	G	T	T	C	A	C	G	T	C	A	T	G	T	A	T	C	A	A	A	A	C	T	T	A	C	C	C	T	C	C	C	T
	AT008746	MD4b2b	T	C	T	T	G	C	A	A	T	A	A	G	T	G	T	T	C	A	C	G	T	C	A	T	G	T	A	T	C	A	A	A	A	C	T	T	A	C	C	C	T	C	C	C	T
	AT008732	MD4b2b	T	C	T	T	A	C	A	A	T	A	A	G	T	G	T	T	C	A	C	G	T	C	A	T	G	T	A	T	C	A	A	A	A	C	T	T	A	C	C	C	T	C	C	C	T
	AT008697	MD4b2b	T	C	T	T	G	C	^	^	T	^	^	G	T	G	T	T	C	^	C	G	T	C	^	T	G	T	^	T	C	^	^	^	^	C	T	T	^	C	C	C	T	C	C	C	T
	AT008684	MD4b2b	T	T	T	T	G	C	A	A	T	A	A	G	T	G	T	T	C	A	C	G	T	C	G	T	G	T	A	T	C	A	A	A	A	C	T	T	A	C	C	C	T	C	C	C	T
	AT008579	MD4b2b	T	C	T	T	G	C	A	A	T	A	A	G	T	G	T	T	C	A	C	G	T	C	A	T	G	T	A	T	C	A	A	A	A	C	T	T	A	C	C	C	T	C	C	C	T
	AT008574	MD4b2b	T	C	T	T	G	C	A	A	T	A	A	G	T	G	T	T	C	A	C	G	T	C	A	T	G	T	A	T	C	A	A	A	A	C	T	T	A	C	C	C	T	C	C	C	T
	AT008549	MD4b2b	T	C	T	T	C	C	A	A	T	A	A	C	T	G	T	T	C	A	C	G	T	C	A	T	C	T	C	T	C	A	A	A	A	C	T	T	A	C	C	C	T	C	C	C	T
	AT008526	MD4b2b	T	C	T	T	G	C	A	A	T	A	A	G	T	G	T	T	C	A	C	G	T	C	A	T	G	T	A	T	C	A	A	A	A	C	T	T	A	C	C	C	T	C	C	C	T
	AT008523	MD4b2b	T	C	T	T	G	C	A	A	T	A	A	G	T	G	T	T	C	A	C	G	T	C	A	T	G	T	A	T	C	A	A	A	A	C	T	T	A	C	C	C	T	C	C	C	T
	AT008520	MD4b2b	T	C	T	T	G	C	A	A	T	A	A	G	T	G	T	T	C	A	C	G	T	C	A	T	G	T	A	T	C	A	A	A	A	C	T	T	A	C	C	C	T	C	C	C	T
	AT008306	MD4b2b	T	C	T	T	G	C	A	A	T	A	A	G	T	G	T	T	C	A	C	G	T	C	A	T	G	T	A	T	C	A	A	A	A	C	T	T	A	C	C	C	T	C	C	C	T
	AT008499	MD4b2b	T	T	T	T	G	C	^	^	T	^	^	G	T	G	T	T	C	^	C	G	T	C	^	T	G	T	^	T	C	^	^	^	^	C	T	T	^	C	T	C	T	C	C	C	T
	AT008477	MD4b2b	T	C	T	C	G	C	A	A	T	A	A	G	T	G	T	T	C	A	C	G	T	C	A	T	A	T	A	T	C	A	A	A	A	C	T	T	A	C	C	C	T	C	C	C	T
	AT008467	MD4b2b	T	C	T	T	G	C	A	A	T	A	A	G	T	G	T	T	C	A	C	G	T	C	A	T	G	T	A	T	C	A	A	A	A	C	T	T	A	C	T	C	T	C	C	C	T
	AT008437	MD4b2b	T	T	T	T	G	C	A	A	T	A	A	G	T	G	T	T	C	A	C	G	T	C	G	T	G	T	A	T	C	A	A	A	A	C	T	T	A	C	C	C	T	C	C	C	T
	AT008411	MD4b2b	C	C	T	T	G	C	^	^	T	^	^	G	T	G	T	T	C	^	C	G	T	C	^	T	G	T	^	T	C	^	^	^	^	C	T	T	^	C	C	C	T	C	C	C	T
	AT008392	MD4b2b	T	C	T	T	G	C	A	A	T	A	A	G	T	G	T	T	C	A	C	G	T	C	A	T	G	T	A	T	C	A	A	A	A	C	T	T	A	C	C	C	T	C	C	C	T
	AT008388	MD4b2b	T	C	T	T	G	C	A	A	T	A	A	G	T	G	T	T	C	A	C	G	T	C	A	T	A	T	A	T	C	A	A	A	T	C	T	T	A	C	C	C	T	C	C	C	T
	AT008344	MD4b2b	T	C	T	T	G	C	A	A	T	A	A	G	T	G	T	T	C	A	C	G	T	C	A	T	G	T	A	T	C	A	A	A	A	C	T	T	A	C	C	C	T	C	C	C	T
	AT008339	MD4b2b	T	C	T	T	C	C	A	A	T	A	A	C	T	G	T	T	C	A	C	G	T	C	A	T	C	T	C	T	C	A	A	A	A	C	T	T	A	C	C	C	T	C	C	C	T
	AT008333	MD4b2b	T	C	T	T	G	C	A	A	T	A	A	G	T	G	T	T	C	A	C	G	T	C	A	T	G	T	A	T	C	A	A	A	A	C	T	T	A	C	C	C	T	C	C	C	T
	AT008331	MD4b2b	T	C	T	T	G	C	A	A	T	A	A	G	T	G	T	T	C	A	C	G	T	C	A	T	G	T	A	T	C	A	A	A	A	C	T	T	A	C	C	C	T	C	C	C	T
	AT008326	MD4b2b	T	C	T	T	G	C	A	A	T	A	A	G	T	G	T	T	C	A	C	G	T	C	A	T	G	T	A	T	C	A	A	A	A	C	T	T	A	C	C	C	T	C	C	C	T
	AT008308	MD4b2b	T	C	T	C	G	C	A	A	T	A	A	G	T	G	T	T	C	A	C	G	T	C	G	T	G	T	A	T	C	A	A	A	A	C	T	T	A	C	C	C	T	C	C	C	T
	AT008302	MD4b2b	T	C	T	T	G	C	^	^	T	^	^	G	T	G	T	T	C	^	C	G	T	C	^	T	G	T	^	T	C	^	^	^	^	C	T	T	^	C	C	C	T	C	C	C	T
	AT008293	MD4b2b	T	C	T	T	G	C	A	A	T	A	A	G	T	G	T	T	C	A	C	G	T	C	A	T	G	T	A	T	C	A	A	A	A	C	T	T	A	C	C	C	T	C	C	C	T
	AP008792	MD4b2b	C	C	T	T	G	C	A	A	T	A	A	G	T	G	T	T	C	A	C	G	T	C	A	T	G	T	A	T	C	A	A	A	A	C	T	T	A	C	C	C	T	C	C	C	T
	AP008264	MD4b2b	T	C	T	T	G	C	A	A	T	A	A	G	T	G	T	T	C	A	C	G	T	C	A	T	G	T	A	T	C	A	A	A	A	C	T	T	A	C	C	C	T	C	C	C	T
	AP010666	MD4b2b	C	C	T	T	G	C	^	^	T	^	^	G	T	G	T	T	C	^	C	G	T	C	^	T	G	T	^	T	C	^	^	^	^	C	T	T	^	C	C	C	T	C	C	C	T
	AP010678	MD4b2b	T	C	T	T	G	C	A	A	T	A	A	G	T	G	T	T	C	A	C	G	T	C	A	T	G	T	A	T	C	A	A	A	A	C	T	T	A	C	C	C	T	C	C	C	T
	AP010694	MD4b2b	T	C	T	T	G	C	A	A	T	A	A	G	T	G	T	T	C	A	C	G	T	C	A	T	G	T	A	T	C	A	A	A	A	C	T	T	A	C	C	C	T	C	C	C	T
	AP010739	MD4b2b	T	C	T	T	G	C	A	A	T	A	A	G	T	G	T	T	C	A	C	G	T	C	A	T	G	T	A	T	C	A	A	A	A	C	T	T	A	C	C	C	T	C	C	C	T

4.3.3 Contamination estimates

The frequency of modern human DNA contamination was 1.48-30.08% (Table 4.5). The frequencies of contamination were relatively high in Odake Jomon, Daizenno-minami Jomon, and Shiraho-Saonetabaru sample, because the DNAs were highly degraded and sensitive to low amount of the contaminations. The contamination frequencies, estimated from mitochondrial DNA and nuclear DNA, were very close each other.

Table 4.5: Contamination estimates with mitochondrial DNA and sex chromosomes

Remains	ID	Sex *1	*Nconsistent* *2	*Ninconsistent* *3	Total	*Fmt* *4	Number of sequence reads mapped to chromosomeX	Number of sequence reads mapped to chromosomeY	X:Y	*Fsex* *5
Shikkaribe	F12	Female	9,060	142	9,202	1.54%	2,013,244	1,631	1234.4 : 1	1.62%
Yugura	LLM3	Female	99	2	101	1.98%	90,264	67	1347.2 : 1	1.48%
Daizenno-Minami	5_3	Female	-	-	-	-	1,101	4	275.3 : 1	7.27%
Odake	32	Female?	-	-	-	-	133	2	66.5 : 1	30.08%
Shiraho-Saonetabaru	-	Female	-	-	-	-	211	2	105.5 : 1	18.96%
Aboriginal		Male	-	-	-	-	71,371,413	7,130,997	10.0 : 1	-

*1 Sexes were clasified by the ratio of chromosome X and Y.
*2 Number of reads consistent with haplotype
*3 Number of reads inconsistent with haplotype
*4 Frequency of contamination infered from mitochondrial DNA haplotype
*5 Frequency of contamination infered from the ratio of sex chromosomes

4.3.4 Estimation the frequency of post-mortem change

I inferred the frequency of post-mortem changes on Shikkariabe Jomon, Sanganji Jomon, Denisovan, and Vindija Neanderthal. About 24 million nucleotide positions were available for the estimation (Table 4.6). The frequency of post-mortem transversion-type substitution is only 0.0196% in the Shikkariabe Jomon genome (Table 4.7). The frequency is one tenth of the frequency in Sanganji Jomon genomes. The frequencies may be proportional to the residual rate of human DNA and to the coverage of the genome, which reduce the effect of sequencing error. The frequencies of post-mortem changes were 0.0247% and 0.0437% in Vindija Neanderthal and Denisovan, respectively, and the frequency on Vindija Neanderthal is consistent with the estimation by Reich et al. (2011).

Table 4.6: Summary of genomic sequence data overlapping among Shikkariabe Jomon, Sanganji Jomon, 14 modern humans, Vindija Neanderthal, Denisovan, and Chimpanzee

	Number of bases
All sequence	24,006,855
All sequence (only one type or two types)	23,998,379
All variant site (only two types)	
All	1,033,301
Transversion	221,750

Table 4.7: Pair-wise distance in transversion substitution among 19 individuals at 23,998,379 nucleotide sites uncorrected for post-mortem change and the estimation of the frequency of post-mortem change in Shikkariabe Jomon and Sanganji Jomon

	Pa	San	Yor	Mbu	ESN	GWD	MSL	DNK	Den	VNean	S.A.J	S.J.	Fre	Sar	Pap	Han	JPT	CDX
PanTro2 (Pa)																		
San	97,856																	
Yoruba (Yor)	97,396	12,366																
Mbuti (Mbu)	97,282	12,176	11,540															
ESN	97,948	12,946	11,582	12,136														
GWD	98,058	13,132	11,908	12,320	12,514													
MSL	98,352	13,410	12,170	12,630	12,744	13,020												
DNK	97,549	12,529	11,409	11,739	11,901	12,181	12,483											
H-Denisovan *1 (Den)	107,672	24,960	24,582	24,422	25,080	25,266	25,518	24,697										
Vindija Neanderthal (VNean)	103,124	19,956	19,460	19,454	19,994	20,238	20,440	19,725	28,174									
Shikkariabe Jomon (S.A.J)	101,905	16,735	15,685	16,093	16,267	16,471	16,761	15,770	28,855	23,723								
Sanganji Jomon *2 (S.J.)	137,968	53,050	52,046	52,368	52,530	52,724	53,032	52,113	65,102	59,884	53,489							
French (Fre)	97,112	12,102	10,936	11,330	11,530	11,712	11,982	10,941	24,188	18,996	14,201	50,462						
Sardinian (Sar)	97,064	12,034	10,882	11,324	11,590	11,780	12,050	10,953	24,146	18,978	14,243	50,532	9,038					
Papuan (Pap)	97,275	12,219	11,079	11,479	11,861	11,889	12,351	11,176	24,163	19,061	14,296	50,615	9,699	9,745				
Han	97,232	12,158	11,078	11,420	11,714	11,920	12,174	11,111	24,320	19,182	13,825	50,200	9,558	9,586	9,589			
JPT	98,278	13,242	12,198	12,584	12,694	12,866	13,290	12,163	25,342	20,164	14,719	51,072	10,596	10,674	10,657	10,142		
CDX	96,277	11,119	10,065	10,477	10,705	10,869	11,115	10,156	23,237	18,031	12,838	49,141	8,523	8,637	8,578	8,073	9,187	
Karitiana	97,337	12,273	11,077	11,463	11,753	11,965	12,349	11,258	24,397	19,303	14,034	50,467	9,533	9,525	9,794	9,387	10,463	8,424
Average pairwise distance between modern individuals (APD)	97,225	12,164	11,045	11,440	11,692	11,857	12,187	11,108	Average									
Number of difference between APD and Shikkariabe Jomon	4,680	4,571	4,640	4,653	4,575	4,614	4,574	4,662	4,621									
Number of difference between APD and Sanganji Jomon	40,743	40,886	41,001	40,928	40,838	40,867	40,845	41,005	40,889									
Post-mortem change in Shikkariabe Jomon (%)	0.0196	0.0191	0.0193	0.0194	0.0191	0.0192	0.0191	0.0194	**0.0193**	Number of post-mortem change in Shikkariabe Jomon				**4,625** (+4)				
Post-mortem change in Sanganji Jomon (%) *3	0.1705	0.1705	0.1709	0.1706	0.1703	0.1704	0.1703	0.1709	**0.1705**	Number of post-mortem change in Sanganji Jomon				**40,927** (+38)				
Number of difference between APD and Denisovan (n)	10,447																	
Number of difference between APD and Vindija Neanderthal (n)	5,899																	
Post-mortem change in Denisovan (%) *3	**0.0437**									Number of post-mortem change in Denisovan				**10,489** (+42)				
Post-mortem change in Vindija Neanderthal (%) *3	**0.0247**									Number of post-mortem change in Vindija Neanderthal				**5,923** (+24)				

*1 High coverage of Denisovan genome

*2 Merged genome dataset with Sanganji 131421-3 and Sanganji 131464

*3 The frequency was calculated with [(number of post-mortem change)/(Total nucleotide sites)]*100

4.3.5 Principal component analysis and investigating heterogeneity of populations

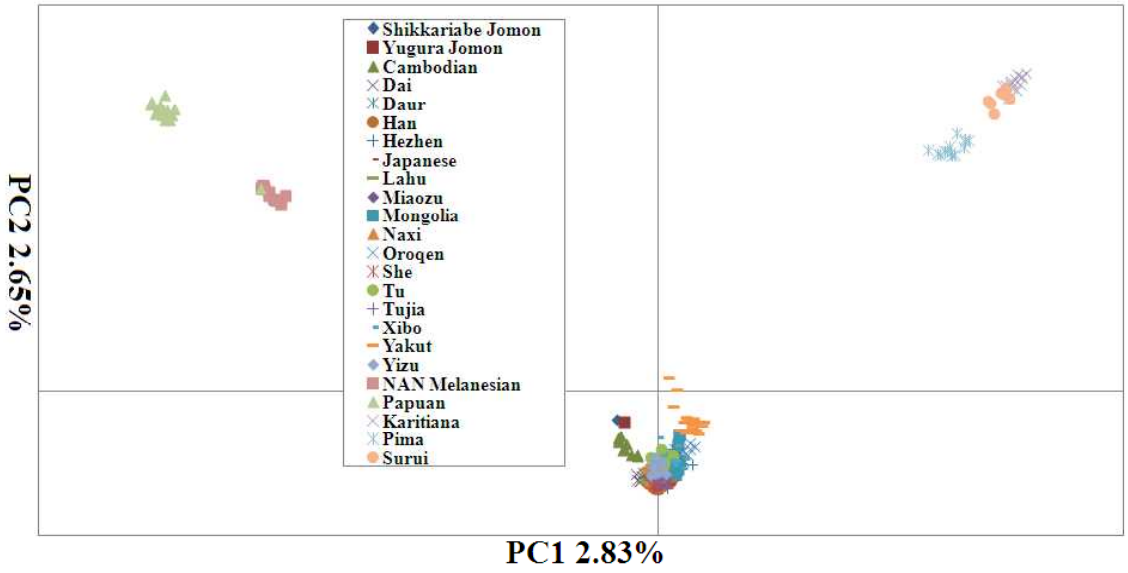
To investigate the genetic relationship among two Jomons and HGDP-CEPH East Eurasians, Melanesians, and Native Americans I carried out PCA. I observed that two Jomon individuals clustered with present-day East Eurasians, but they were relatively closer to Melanesians than other East Eurasians (Figure 4.5a). When I carried out PCA on two Jomons and modern East Eurasians, both Shikkariabe and Yugura Jomon individuals were isolated from modern continental East Eurasians (Figure 4.5b). These results are consistent with the result using Sanganji Jomons (see Chapter 3). Since the number of SNPs was larger than that of Sanganji Jomon, we can clearly see the differences between the two Jomons and East Eurasians. Two Jomons were from different regions and periods, but they were plotted to similar positions. Even in the Japanese Archipelago, all three Jomons neighbored each other, and they apart from all the three present-day populations who inhabit in the Japanese Archipelago (Figure 4.6). The genetic similarity among Jomon individuals was observed not only in PC1 and PC2 but also PC3 and PC4, and I didn't find the evidence that Jomon people were heterogeneous in terms of regions and era. I also investigated pair-wise distances of Shikkariabe Jomon and Yugura Jomon with JAHPGC dataset, and I found that their distances (mean = 0.257) were smaller than those for other populations (Table 4.8). This result also supports that the two Jomon individuals were homogeneous each other.

I also investigated the genetic relationship between Shikkariabe Jomon and PASNP human minorities from southern part of East Eurasians and Melanesians. I observed that Shikkariabe Jomon was isolated from other East Eurasians, but Shikkariabe Jomon was relatively closer to Austronesian Philippino and Taiwanese than Sino-Tibetan Chinese, Taiwanese, and Korean (Figure 4.7).

I carried out PCA with Shikkariabe Jomons, Upper Paleolithic Ryukyuan, and the individuals from three Japanese populations and Chinese Beijing. Against my expectations, the number of SNPs was limited, but I tried to investigate the genetic relationship between Jomon people and Paleolithic Ryukyuan. PC1 separated Shikkariabe Jomon and Paleolithic Ryukyuan from other modern humans (Figure 4.8). Since informative SNPs were limited, it is premature to

extract meaningful result from Figure 4.8, but it is possible that both Jomon people and Paleolithic Ryukyuan shared the same ancestors migrated from Southeast Asia into East Asia.

a)



b)

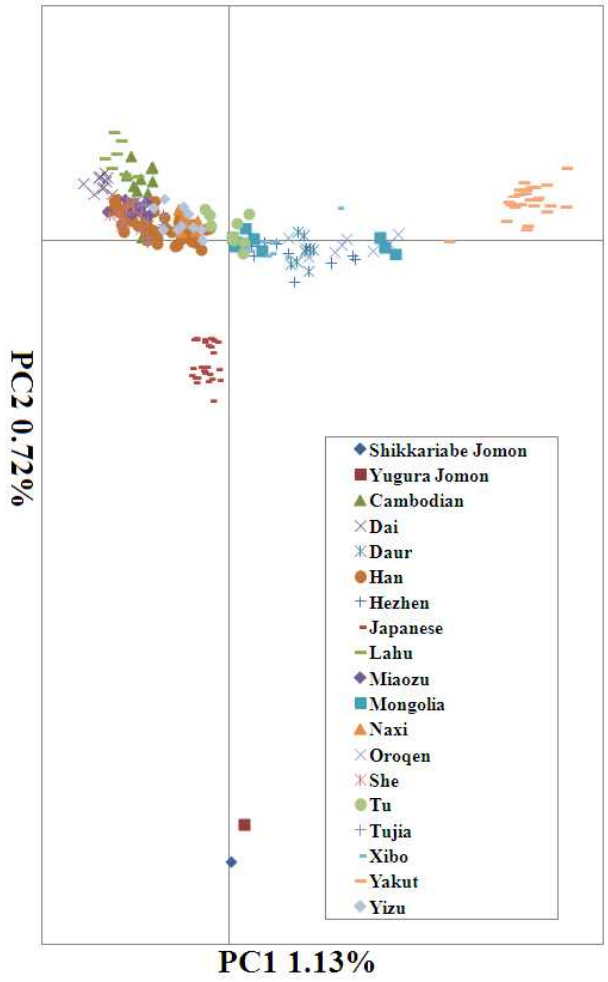
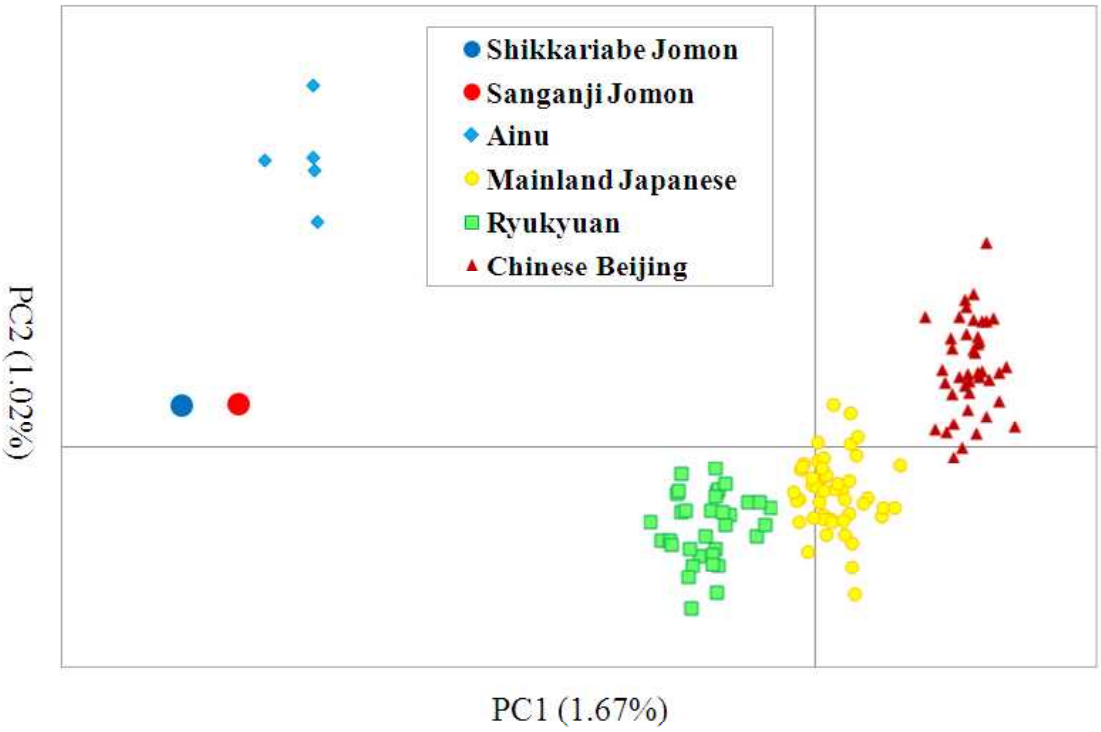


Figure 4.5: Genetic relationship among Three Jomons and HGDP-CEPH populations
a) Shikkariabe Jomon, Yugura Jomon, and HGDP-CEPH East Eurasians, Melanesians, and Native Americans, b) Shikkariabe Jomon, Yugura Jomon, and HGDP-CEPH East Eurasians.

a)



b)

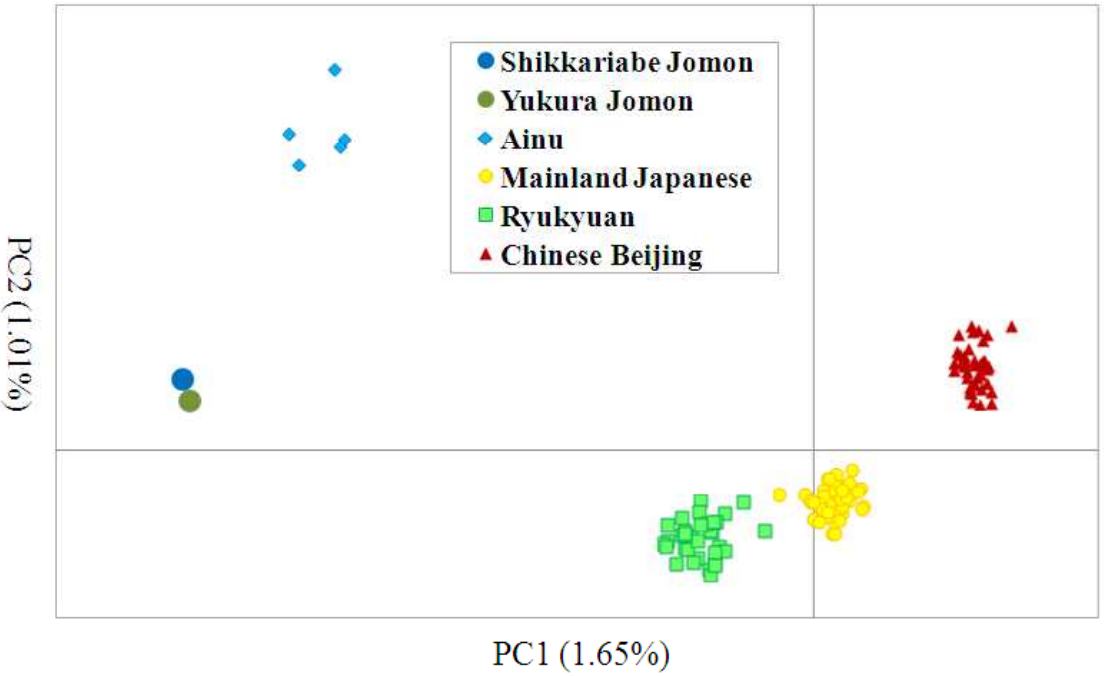
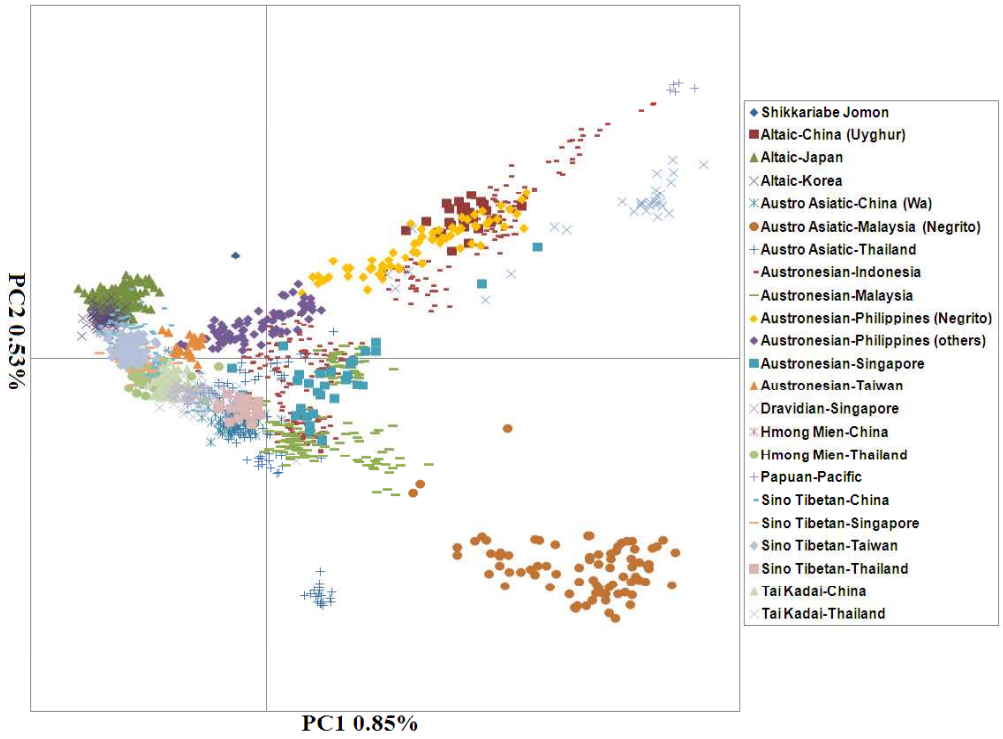


Figure 4.6: Genetic relationship among three Jomon individuals and the individuals of three populations inhabits in Japanese archipelago (Ainu, mainland Japanese, and Ryukyuan) and of Chinese Beijing (CHB)
a) Shikkariabe Jomon and Sanganji Jomons, b) Shikkariabe Jomon and Yugura Jomon.

a)



b)

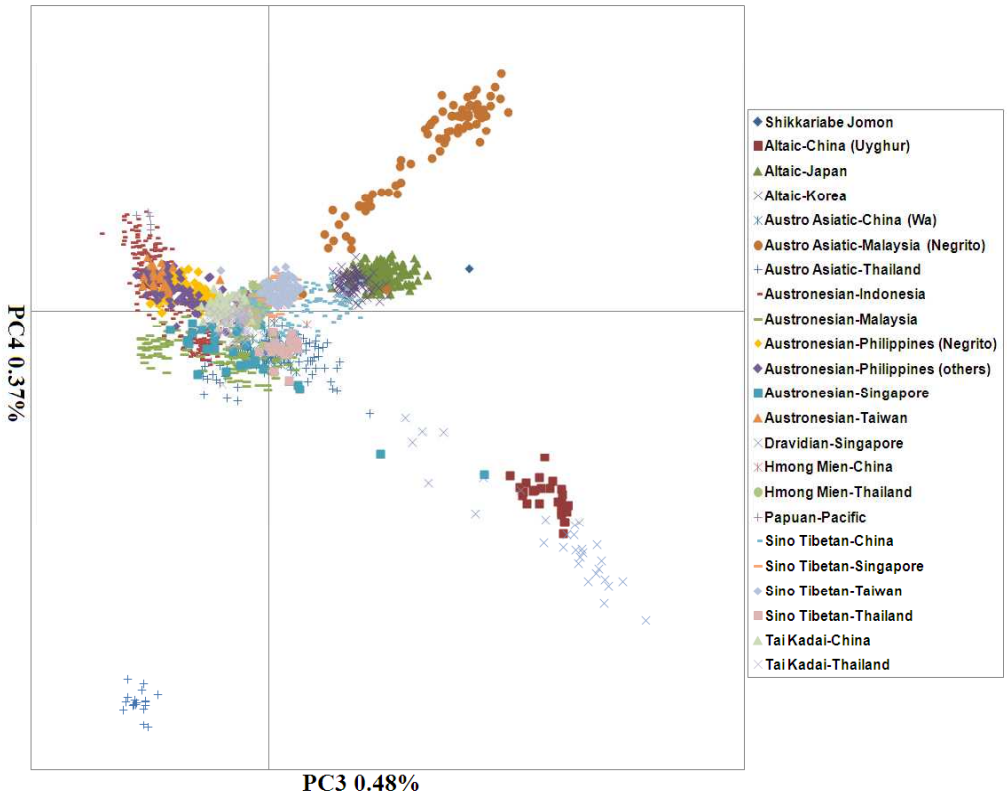


Figure 4.7: Genetic relationship among Shikkariabe Jomon and PASNP East Eurasians and Melanesian
a) result of PC1 and PC2, b) result of PC3 and PC4.

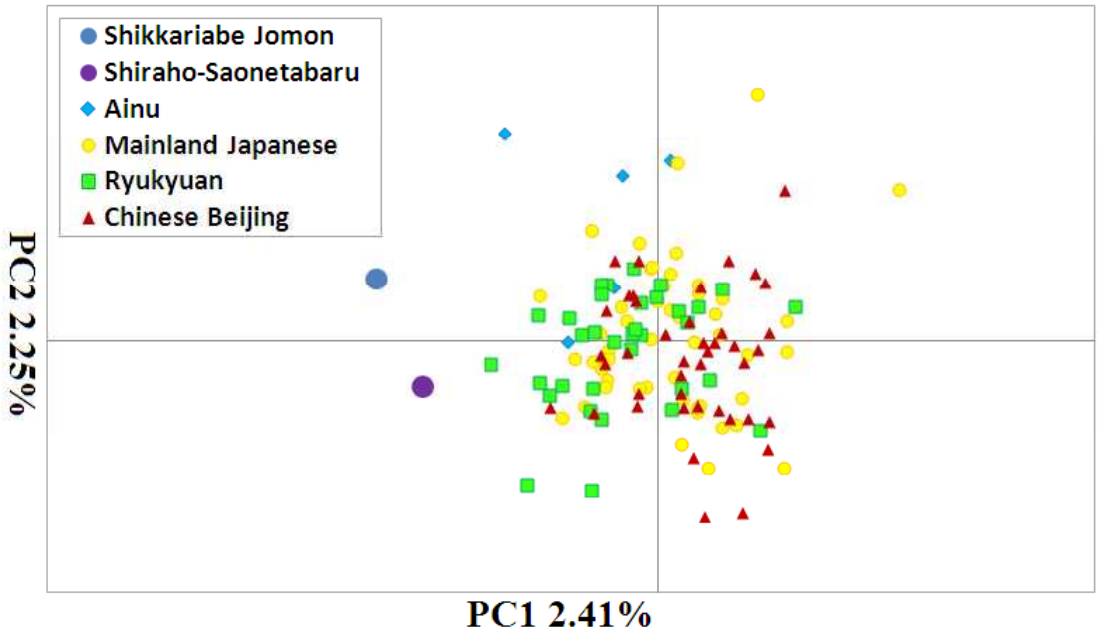


Figure 4.8: Genetic relationship among Shikkariabe Jomon, Shiraho-Saonetabaru individual, and the individuals of three populations inhabits in Japanese archipelago (Ainu, mainland Japanese, and Ryukyuan) and of Chinese Beijing (CHB)

Table 4.8: Pair-wise distance among two Jomon, Ainu, Ryukyuan, mainland Japanese, and Chinese Beijing

Region	Individual ID		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Shikkariabe Jomon	#Shikkariabe F12	1																					
Yugura Jomon	#Yugawa LLM3	2	0.257																				
Ainu	#2120001A02	3	0.286	0.289																			
	#2120001A05	4	0.281	0.284	0.272																		
	#2120001B11	5	0.286	0.286	0.276	0.277																	
	#2120002A01	6	0.291	0.296	0.284	0.281	0.289																
	#2120002A02	7	0.288	0.289	0.275	0.269	0.282	0.281															
Ryukyuan	#3040001A01	8	0.327	0.327	0.325	0.323	0.324	0.326	0.321														
	#3040001A04	9	0.328	0.327	0.325	0.323	0.323	0.327	0.319	0.317													
	#3040001A10	10	0.325	0.329	0.322	0.324	0.321	0.328	0.322	0.322	0.317												
	#3040001B01	11	0.326	0.328	0.320	0.323	0.322	0.328	0.324	0.318	0.321	0.321											
	#3040001B07	12	0.328	0.329	0.325	0.327	0.328	0.328	0.323	0.322	0.320	0.319	0.319										
Mainland Japanese	#2200001A01	13	0.316	0.313	0.316	0.317	0.319	0.327	0.317	0.321	0.321	0.321	0.316	0.322									
	#2200001A02	14	0.316	0.316	0.321	0.317	0.320	0.325	0.317	0.321	0.325	0.321	0.322	0.320	0.316								
	#2200001A03	15	0.316	0.314	0.317	0.314	0.321	0.323	0.317	0.323	0.315	0.317	0.320	0.319	0.319	0.318							
	#2200001A05	16	0.314	0.313	0.318	0.317	0.319	0.322	0.315	0.321	0.316	0.317	0.319	0.319	0.315	0.318	0.317						
	#2200001A06	17	0.313	0.314	0.320	0.318	0.322	0.321	0.319	0.322	0.322	0.322	0.321	0.324	0.315	0.318	0.316	0.318					
Chinese Beijing	#NA18524_GW6_A	18	0.344	0.342	0.335	0.334	0.337	0.338	0.332	0.328	0.327	0.326	0.326	0.327	0.329	0.331	0.328	0.327	0.328				
	#NA18529_GW6_A	19	0.340	0.337	0.332	0.334	0.334	0.334	0.326	0.321	0.323	0.321	0.323	0.322	0.324	0.325	0.319	0.324	0.324	0.321			
	#NA18537_GW6_A	20	0.339	0.340	0.332	0.333	0.336	0.336	0.332	0.323	0.324	0.323	0.322	0.326	0.324	0.326	0.325	0.323	0.327	0.321	0.320		
	#NA18540_GW6_A	21	0.339	0.339	0.331	0.330	0.331	0.333	0.328	0.322	0.320	0.322	0.322	0.326	0.322	0.328	0.323	0.326	0.322	0.325	0.320	0.319	
	#NA18542_GW6_A	22	0.336	0.340	0.331	0.331	0.332	0.335	0.331	0.321	0.323	0.324	0.323	0.326	0.323	0.325	0.325	0.327	0.325	0.326	0.320	0.321	0.325

I also carried out PCA with Chimpanzee, Vindija Neanderthal, Denisovan, Shikkariabe Jomon, and modern humans to investigate the genetic relationship between the Shikkariabe Jomon and archaic humans. A total of 53,783 SNP sites were available. Present-day non-Africans were genetically closer to archaic humans than African people (Figure 4.9), and Vindija Neanderthal hauls Shikkariabe Jomon as well as other Eurasians and Native Americans. Denisovan pulls all East Asians and Melanesians, especially the contribution of Denisovan gene flow seems to be strong in Melanesians. This result is consistent with the result of Reich et al. (2011). Shikkariabe Jomon was within the cluster of East Eurasians, and no strong Denisovan gene flow was observed. If Shikkariabe Jomon and Sanganji Jomon were homogeneous, the result is inconsistent with the result of *D*-statistic analysis in Chapter 3. I conducted the same analysis by substituting Ainu people for Shikkariabe Jomon. A total of 16,496 SNP sites were available, and all the five individuals were within the cluster of modern European, East Eurasians and Native Americans (Figure 4.10). Because of the limitation of available SNPs, the separation between Europeans and East Eurasians was not observed. Since Ainu people seem to inherit many of their genetic components from Jomon people, the result is consistent with the result of Shikkariabe Jomon.

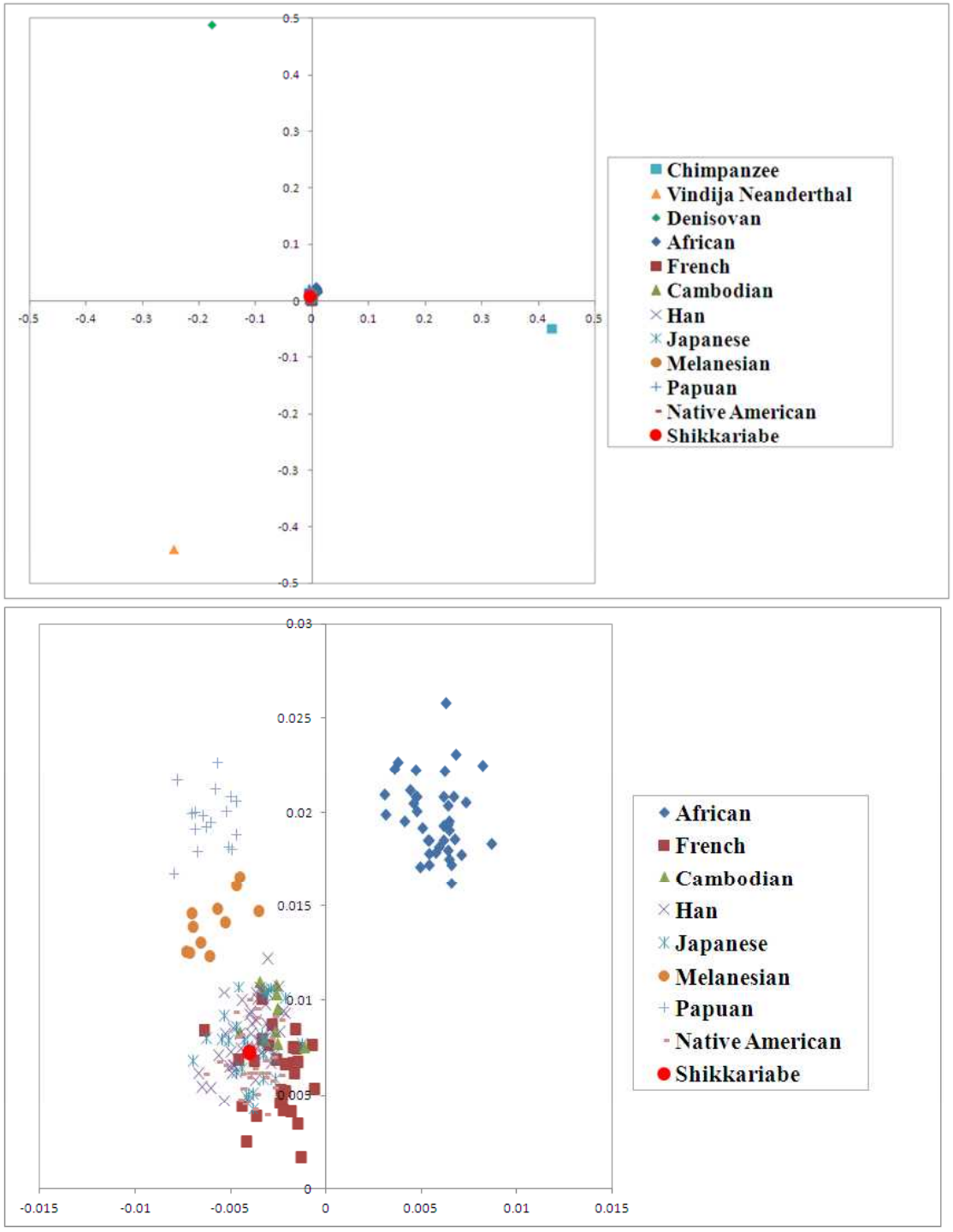


Figure 4.9: Principal Component Analysis of 53,783 SNPs known to be polymorphic among three individuals, Chimpanzee, Vindija Neanderthal, and Denisovan
26,051 SNPs were unique to Chimpanzee, 15,181 SNPs were unique to Vindija Neanderthal, and 12,551 SNPs were unique to Denisovan. a) PCA of Chimpanzee, Vindija Neanderthal, Denisovan, and modern humans. b) Magnification of the central portion of the plot.

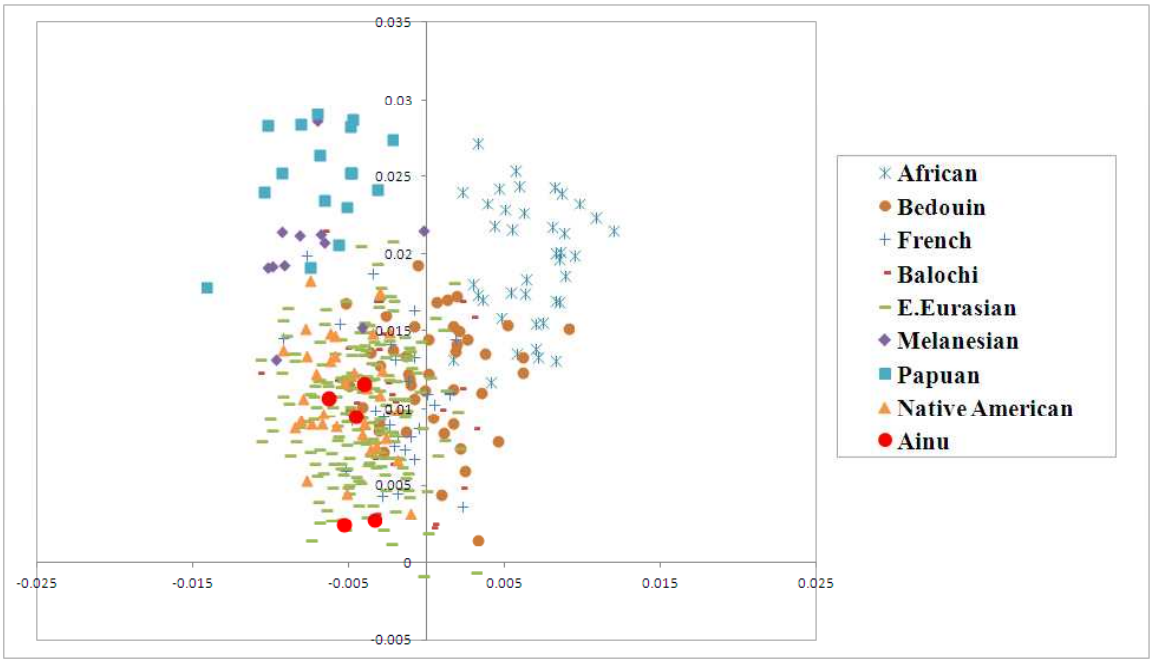


Figure 4.10: Principal Component Analysis of 20,843 SNPs known to be polymorphic among three individuals, Chimpanzee, Vindija Neanderthal, and Denisovan

10,053 SNPs were unique to Chimpanzee, 5,941 SNPs were unique to Vindija Neanderthal, and 4,849 SNPs were unique to Denisovan. I magnified the central portion of the plot.

4.3.6 Phylogenetic tree and Network analysis

It has been shown that Sanganji Jomon and Shikkariabe Jomon are genetically close in chapter 4.3.5. To investigate phylogenetic relationship between the two Jomons and among the two Jomons and modern or archaic humans, I constructed phylogenetic tree and network. Sanganji Jomon and Shikkariabe Jomon clustered at the 100% bootstrap value (Figure 4.11). It indicates that the two Jomons belong to same lineage in East Asia, which is consistent with the result of PCA (Figure 4.11). In network analysis, Shikkariabe Jomon and Sanganji Jomon shared a long split (network not shown). The long split may indicate that Shikkariabe Jomon and Sanganji Jomon were genetically isolated from other populations for a long time.

No European gene flow was observed in Shikkariabe Jomons as well as Sanganji Jomons. Within the population of Japanese Archipelago, Shikkariabe Jomon was a little closer to modern Japanese than Sanganji Jomons. It is possible that this is caused by the effect of post-mortem changes. I thus corrected the pair-wise distances with equation (3.6) to minimize the effect of post-mortem changes in network analysis (Table 4.9). The corrected phylogenetic network shown in Figure 4.12 showed the same topology with the network using uncorrected distances. The splits separating Vindija Neanderthal and Shikkariabe Jomon from Africans suggest the gene flow of Vindija Neanderthal into the ancestors of Shikkariabe Jomon. The result is consistent with the result of PCA (Figure 4.9). However, the Denisovan gene flow into Shikkariabe Jomon was not observed (Figure 4.13), which is incongruent result with phylogenetic network described in Chapter3. It is possible that the phylogenetic network doesn't show all the genetic relationship. Actually, when I used only four individuals for network analysis, I observed splits separating Shikkariabe Jomon and Denisovan from Chimpanzee and another present-day human (e.g. Figure 4.14a). The amount of Denisovan DNA seems to be high in Papuan than Shikkariabe Jomon (Figure 4.14b).

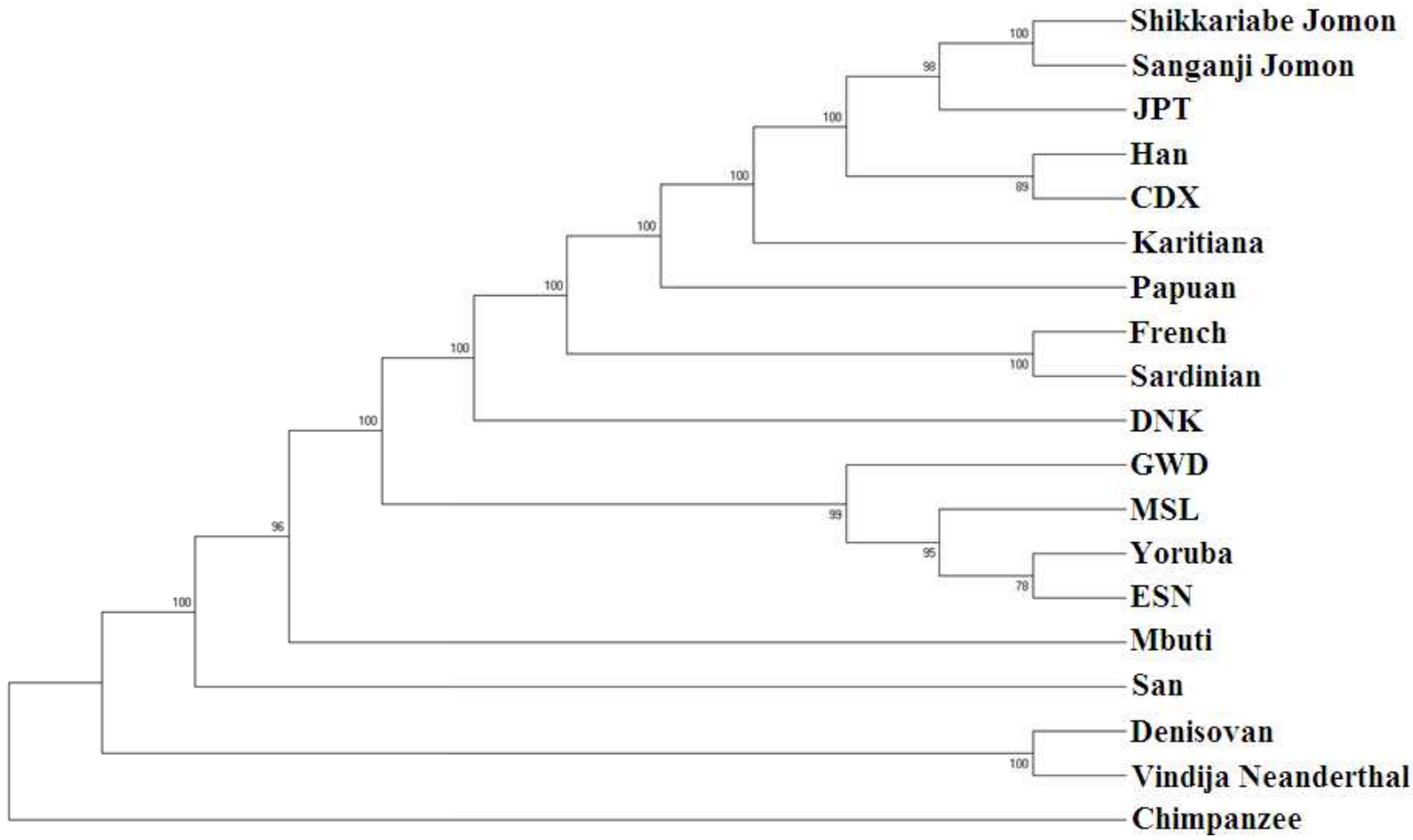


Figure 4.11: A neighbor-joining tree of Shikkariabe Jomon, Sanganji Jomon, 14 present-day humans, Vindija Neanderthal, Denisovan, and Chimpanzee genome based on 23,998,379 transversion substitution sites

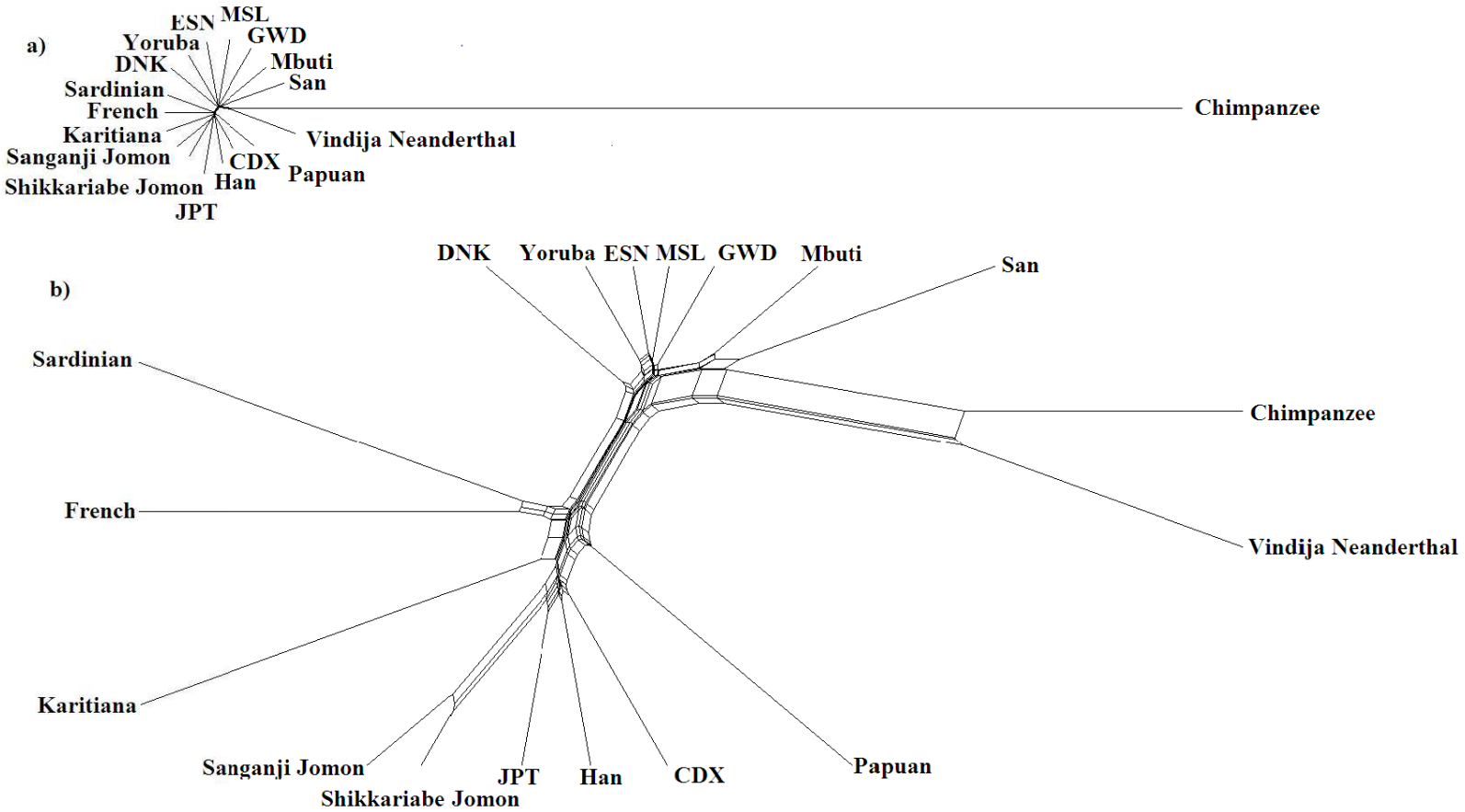


Figure 4.12: A corrected phylogenetic network of Shikkariabe Jomon, Sanganji Jomon, 14 present-day humans, Vindija Neanderthal, and Chimpanzee genome based on 23,998,379 transversion substitution sites

Table 4.9: Pair-wise distance in transversion substitution among 19 individuals at 23,998,379 nucleotide sites corrected for post-mortem change

	Pa	San	Yor	Mbu	ESN	GWD	MSL	DNK	Den	VNean	S.A.J	S.J.	Fre	Sar	Pap	Han	JPT	CDX
PanTro2 (Pa)																		
San	97,856																	
Yoruba (Yor)	97,396	12,366																
Mbuti (Mbu)	97,282	12,176	11,540															
ESN	97,948	12,946	11,582	12,136														
GWD	98,058	13,132	11,908	12,320	12,514													
MSL	98,352	13,410	12,170	12,630	12,744	13,020												
DNK	97,549	12,529	11,409	11,739	11,901	12,181	12,483											
H-Denisovan *1 (Den)	**97,227**	**14,479**	**14,101**	**13,941**	**14,599**	**14,785**	**15,037**	**14,216**										
Vindija Neanderthal (VNean)	**97,220**	**14,032**	**13,536**	**13,530**	**14,070**	**14,314**	**14,516**	**13,801**	**11,767**									
Shikkariabe Jomon (S.A.J)	**97,292**	**12,106**	**11,055**	**11,464**	**11,638**	**11,842**	**12,132**	**11,140**	**13,745**	**13,170**								
Sanganji Jomon *2 (S.J.)	**97,217**	**12,153**	**11,148**	**11,470**	**11,633**	**11,827**	**12,135**	**11,215**	**13,727**	**13,065**	**7,955**							
French (Fre)	97,112	12,102	10,936	11,330	11,530	11,712	11,982	10,941	**13,707**	**13,072**	**9,571**	**9,561**						
Sardinian (Sar)	97,064	12,034	10,882	11,324	11,590	11,780	12,050	10,953	**13,665**	**13,054**	**9,613**	**9,631**	9,038					
Papuan (Pap)	97,275	12,219	11,079	11,479	11,861	11,889	12,351	11,176	**13,682**	**13,137**	**9,666**	**9,714**	9,699	9,745				
Han	97,232	12,158	11,078	11,420	11,714	11,920	12,174	11,111	**13,839**	**13,258**	**9,195**	**9,299**	9,558	9,586	9,589			
JPT	98,278	13,242	12,198	12,584	12,694	12,866	13,290	12,163	**14,861**	**14,240**	**10,089**	**10,172**	10,596	10,674	10,657	10,142		
CDX	96,277	11,119	10,065	10,477	10,705	10,869	11,115	10,156	**12,755**	**12,106**	**8,208**	**8,238**	8,523	8,637	8,578	8,073	9,187	
Karitiana	97,337	12,273	11,077	11,463	11,753	11,965	12,349	11,258	**13,916**	**13,379**	**9,404**	**9,566**	9,533	9,525	9,794	9,387	10,463	8,424

*1 High coverage of Denisovan genome

*2 Merged genome dataset with Sanganji 131421-3 and Sanganji 131464

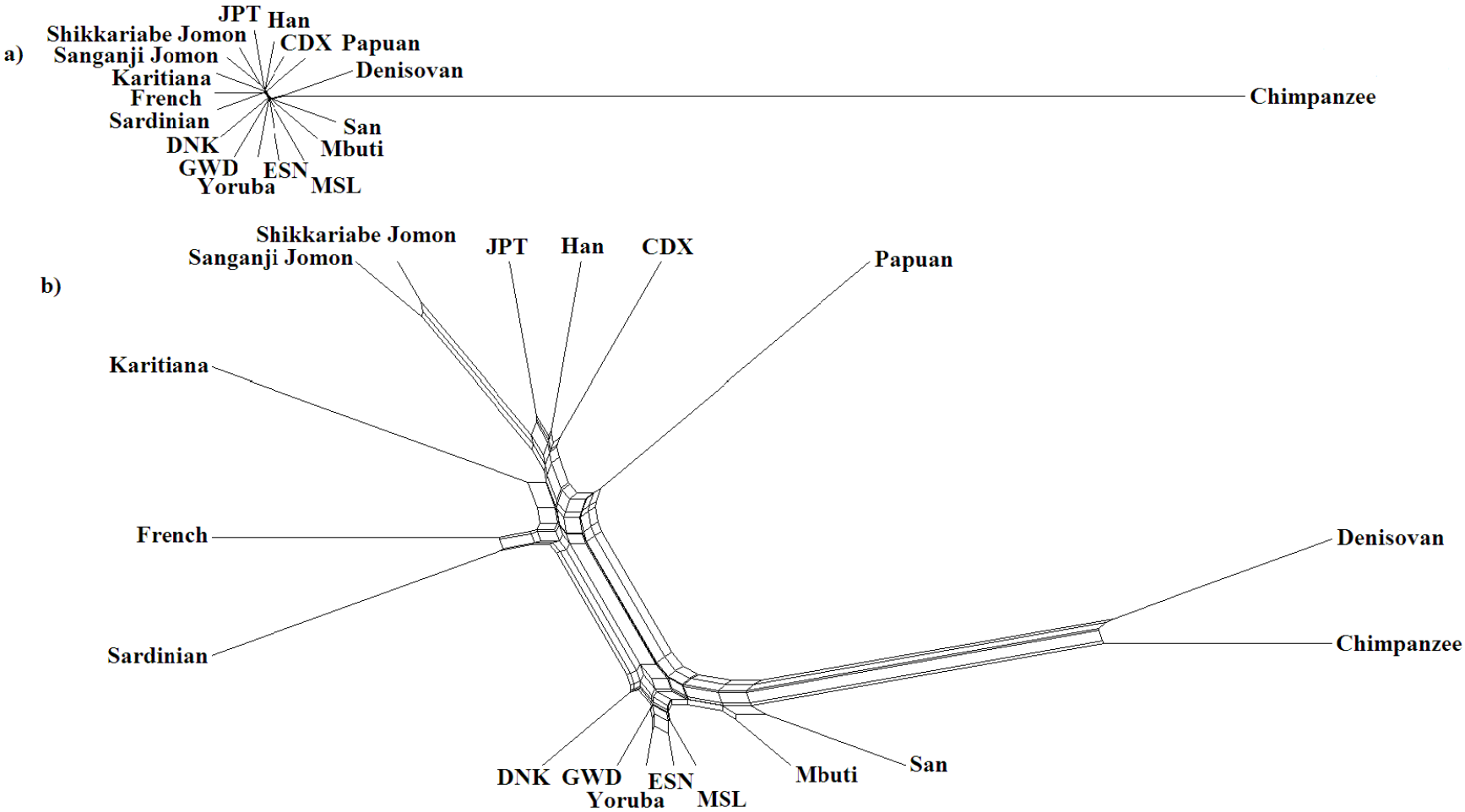


Figure 4.13: A corrected phylogenetic network of Shikkariabe Jomon, Sanganji Jomon, 14 present-day humans, Denisovan, and Chimpanzee genome based on 23,998,379 transversion substitution sites

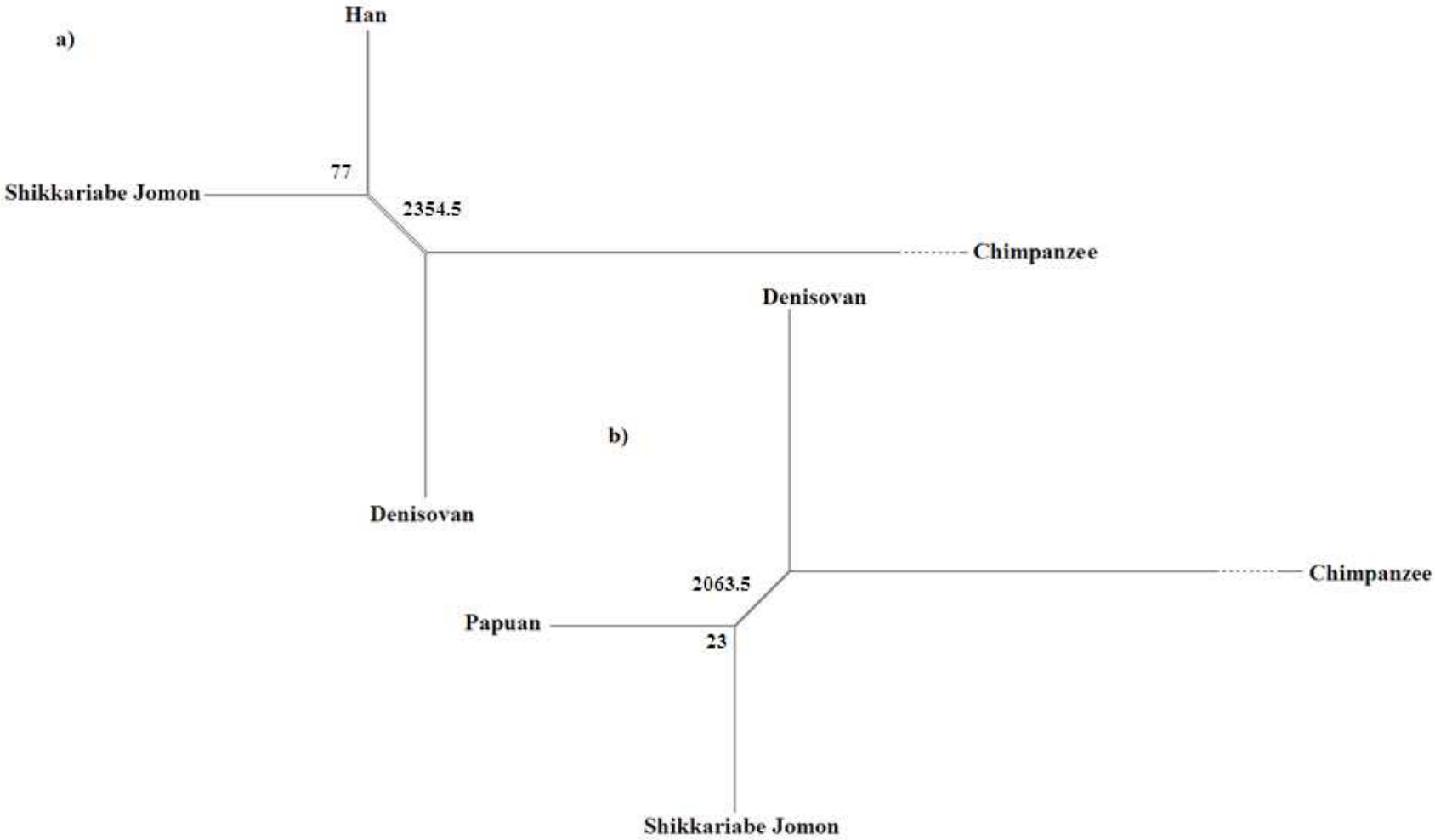


Figure 4.14: A corrected phylogenetic network of four individuals
a) Shikkariabe Jomon was genetically closer to Denisovan than Han, b) Papuan was genetically closer to Denisovan than Shikkariabe Jomon.

4.3.7 *D*-statistic analysis

To detect gene flow among anatomically modern humans and archaic humans, I did *D*-statistic analysis. Based on genome sequences, significant positive $D_{revision}$–score and $Z_{revision}$–values were observed for the quartet of [Chimpanzee, Vindija Neanderthal; San/French/Han, Shikkariabe Jomon] (Table 4.10). The positive values indicate that Shikkariabe Jomon retain more archaic DNA than San, French, and Han. When I computed $D_{standard}$(Chimpanzee, Denisovan; H2, H3), I observed that Shikkariabe Jomon and Papuan had positive $D_{revision}$–score and significant $Z_{standard}$–values. Taking post-mortem change into account, not only Shikkariabe Jomon and Papuan but also Sanganji Jomon and Dai showed significant $Z_{revision}$–values. The $Z_{revision}$–value of Sanganji Jomon was relatively higher than Shikkariabe Jomon. It is possible that I overestimated the frequency of post-mortem changes in Sanganji Jomons. The $Z_{revision}$(Chimpanzee, Denisovan; French, Shikkariabe) was larger than $Z_{revision}$(Chimpanzee, Vindija Neanderthal; French, Shikkariabe). This suggests that excess archaic DNA components in the Shikkariabe Jomon genome than European was from Denisovan, not Vindija Neanderthal. This result is consistent with the result of network analysis (Figure 4.14), but is inconsistent with principal component analysis (Figure 4.9).

I also used genome wide SNP data instead of genomic sequencing data to detect gene flow with *D*-statistic analysis. A total of 285,716 SNP sites were available. I detected the gene flow between Vindija Neanderthal and Shikkariabe Jomon as well as other Non-Africans, while no Denisovan gene flow was detected (Table 4.10, Appendix Table 1). Even if I take the effect of post-mortem changes into account, the $Z_{revision}$(Chimpanzee, Denisovan; French, Shikkariabe Jomon) was not significant. This result is inconsistent with the result based on genomic sequences. There are two possibilities to explain this inconsistency: one is that the SNP sites are for investigating the differentiation among modern humans, not for that between modern human and archaic humans, and the dataset doesn't include informative SNP sites to detect Denisovan gene flow into East Eurasians. The second possibility is that post-mortem changes, low coverage of the genome inducing sequencing error, and short sequence reads mismapping to the reference genome caused false-positive signals in Shikkariabe Jomon. However, the inconsistency was

observed not only for Shikkariabe Jomon but also for Dai. The similar pattern was observed in Cambodian (Reich et al., 2011) (Appendix Table 2). In addition, low frequency of post-mortem changes and small numbers of singleton in mapped reads support weak or negligible effect in *D*-statistic analysis. So, I suppose that first assumption is correct.

Table 4.10: Summary of *D*-statistic analysis of Shikkariabe Jomon, HGDP French, Denisovan, and Chimpanzee based on genome wide SNP data

Outgroup	H1	H2	H3	O_{AAAA}	O_{AAAB}	O_{AABA}	O_{ABAA}	O_{ABBB}	O_{AABB}	O_{ABAB}	O_{ABBA}	$D_{standard}$	$Z_{standard}$	SE $(D_{standard}/Z_{standard})$	$D_{revision}$	$Z_{revision}$
Chimpanzee	Denisovan	French_HGDP00511	Shikkariabe Jomon	120063	35875	37335	14676	17678	37156	8342	8307	0.0027	0.321	0.00841	0.0022	0.2655
Chimpanzee	Denisovan	French_HGDP00511	Cambodian_HGDP00711	120233	35705	37485	14565	17548	37006	8453	8437	0.0002	0.028	0.00714	-	-
Chimpanzee	Denisovan	French_HGDP00511	Dai_HGDP01307	119932	36006	37549	14476	17604	36942	8542	8381	0.0085	1.066	0.00797	-	-
Chimpanzee	Denisovan	French_HGDP00511	Han_HGDP00785	120134	35804	37251	14560	17643	37240	8458	8342	0.0067	0.859	0.00780	-	-
Chimpanzee	Denisovan	French_HGDP00511	Japan_HGDP00757	119981	35957	37306	14585	17529	37185	8433	8456	-0.0008	-0.095	0.00842	-	-
Chimpanzee	Denisovan	French_HGDP00511	Melanesian_HGDP00661	121429	34509	38706	14395	17819	35785	8623	8166	0.0282	3.471	0.00812	-	-
Chimpanzee	Denisovan	French_HGDP00511	Papuan_HGDP00544	121319	34619	38408	14389	17793	36083	8629	8192	0.0259	3.129	0.00828	-	-
Chimpanzee	Denisovan	French_HGDP00511	Karitiana_HGDP01006	120194	35744	36585	14695	17802	37906	8323	8183	0.0071	0.894	0.00794	-	-

4.4 Discussion

Both mitochondrial DNA and nuclear genome data showed the uniqueness of Shikkariabe Jomon and Yugura Jomon. In mitochondrial DNA, Shikkariabe Jomon was classified into haplotype D4h2, but she phylogenetically diverged from other D4h2-type individuals in ancient times (Figure 4.4). Haplotype D4h2 was observed in Ulchi from southern Siberian and Hokkaido Jomon at high frequencies, and haplotype D4b is observed in Tibetan, Chukchi, Altai, and Koryak at high frequencies (18.5%, 6.67%, 3.57%, and 2.58%, respectively). This connection supports that northern to central Jomon people originated from northeast Asia. Previous studies showed that many northern Jomon individuals were classified into ancestral haplotypes or rare haplotypes, N9b*, M7a*, M7a2, while modern Japanese mainly have haplotypes N9b1, N9b2, N9b3, and M7a1. One possibility of this result is that not northern but central or southern part of Jomon people mainly contribute to the genetic formation of modern Japanese. This is mere guesswork, and it is necessary to clarify the haplotype of central and southern Jomon people.

I obtained the 80% of Jomon genome from Shikkariabe Jomon individual. I could use the Shikkariabe Jomon genome as reference Jomon genome. The high coverage of Jomon genome made it possible to investigate the genetic relationship among Jomon people from different regions and era, based on genome sequences. Introduction of two different microblade stone tools into the Japanese Archipelago via Sakhalin and Korean Peninsula suggests that multiple migrants genetically contributed to form Jomon people (Figure 1.2). Based on the limb bone and cranial metric analyses, Takigawa (2006) demonstrated that there are considerable interregional variations among Jomon people, implying their heterogeneity (i.e., multiple origins). I also showed the genetic heterogeneity between Hokkaido&Tohoku Jomon and Kanto Jomon based on the frequency of mitochondrial DNA haplotypes (see Chapter 2). However, I observed no genetic heterogeneity between northern and central Jomon in PCA (Figure 4.6). The pair-wise distances based on genome wide SNPs also showed that Shikkariabe Jomon individual and Yugura Jomon individual were genetically close each other than other in-pair-distances of Ainu, Japanese, Ryukyuan, and Chinese Beijing. Since I failed to obtain high coverage of

Daizenno-Minami Jomon and Odake Jomon genomes, I couldn't compare the genetic relationship between Tohoku Jomon and coastal to Kanto Jomon people. The genetic relationship with modern East Eurasians was basically the same with the result of Sanganji Jomons. This indicates that the Shikkariabe Jomon and Yugura Jomon as well as Sanganji Jomons were genetically unique in East Eurasians, but the Jomon people were genetically close each other even if the region and period are different.

Using principal component analysis (PCA), I tried to investigate the genetic relationship among present-day Japanese, Jomons, and Upper Paleolithic Ryukyuan. The numbers of SNPs were limited, and it is difficult to say much about the relationships. But the separation of Jomon and Upper Paleolithic Ryukyuan from other Japanese in PC1 may indicates the genetic difference between present-day humans and Upper Paleolithic Ryukyuan as well as Jomons, and Jomon people and Upper Paleolithic Ryukyuan were genetically close each other. To investigate more detail of the genetic relationship, we need to accumulate more genome sequences from the Paleolithic humans.

My *D*-statistic analysis implies that Shikkariabe Jomon showed gene flow not only from Vindija Neanderthal but also from Denisovan, and the result is consistent with the result of Sanganji Jomons. The evidence of the Denisovan gene flow wasn't detected from genome wide SNPs, which is consistent with modern Cambodian and Dai. The inconsistency between the result based on genome-wide SNP data and the result based on genomic sequencing data may indicate that genome-wide SNP data is not always suitable to detect the introduction of archaic DNA, and genomic sequences are better for that. Presence or absence of the gene flow is a matter of controversy, but the observation of Denisovan gene flow into Jomon people is one of the keystones to explain the history of modern East Eurasians. Taking into account the result of a recent morphological study (Matsumura, 2006) and the distribution of Denisovan genetic material in East Eurasians and Oceanians, parsimonious explanations for my data will be as follows: first, Denisovan genetic material was introduced into early Southeast Asians, and was widespread in Northeast Asia and Oceania (ancestors of Jomon and Papuan); second, a later wave of an ancestral population(s) with no Denisovan admixture spread from Northeast Asia into

Southeast Asia and contributed significantly to the ancestry of present-day East Eurasians. However, this can't explain the origin of the non-admixed populations. Furthermore, the 40,000 year old Tianyuan individual who inhabited near Beijing, China, experienced no gene flow with Denisovan. This implies a complex history in early human migrations into East Asia. But the analysis to detect Denisovan gene flow into Tianyuan was based on genome-wide SNPs specific to archaic humans. I suppose that Tianyuan SNPs didn't show the evidence of Denisovan gene flow, but instead, if Tianyuan genomic sequences become available, it is possible that we can detect Denisovan DNA material in Tianyuan. Denisovan DNA of modern Japanese seemed to be diluted with the admixture between Jomon people and agricultural people ($Z_{revision}$ = 3.9190 in Shikkariabe Jomon, $Z_{revision}$ = 5.2154 in Sanganji Jomon, but $Z_{revision}$ = 1.4505 in JPT). It is possible that multiple human migrations contributed to the formation of the early stage of East Eurasians. This relates to the question of the genetic relationship between Upper Paleolithic Ryukyuan (e.g. Minatogawa people) and Neolithic Jomon people. Although I pointed out the possibility that Jomon people and Upper Paleolithic Ryukyuan have some genetic relationship, recent studies based on morphology suggest that Upper Paleolithic Ryukyuan were morphologically distinct from Neolithic Mainland Jomon people (Saso et al., 2011; Suwa et al., 2011). More ancient DNA analysis in early East Eurasians will show more detail history of human migration in East Eurasia.

I observed that Jomon people were genetically quite unique in East Eurasia, and it suggests that Jomon people were the descendants of early stage of humans migrated into East Asia. Therefore, Jomon people are very important populations to clarify how early humans migrated into East Asia. I obtained draft sequence of Shikkariabe Jomon genome, and this made it possible to compare the Jomon genome with other ancient and modern human genomes. This result is expected to accelerate the understanding the history of not only modern Japanese and Jomon people but also East Eurasians and Native Americans. I sequenced part of the genome from several Jomon individuals to investigate whether Jomon people were single origin or multiple origins, and whether they were homogeneous or heterogeneous populations. In addition, it is important to discuss the genetic relationship between Jomon people and Upper Paleolithic

Japanese. In this study, I successfully obtained 820Kb of genome sequences from 20,000 year-old Upper Paleolithic Ryukyuan, and I could investigate the genetic relationship based on those genome sequences. Finally, one of the most important discoveries is the existence of genetic relationship between Jomon people and archaic humans. I observed the genetic relationship between not only Jomon and Vindija Neanderthal but also Jomon and Denisovan. Since the current conclusion is only based on partial sequences of the Jomon genomes, further analyses with more genomic information and other statistical methods is essential in future studies. The existence of Denisovan DNA material in Jomon people will be one of the center of debate connected to not only the history of Japanese but also the origin and population structure of East Eurasian.

Reference

Adachi N., Shinoda K., and Umetsu K. (2008) Mitochondrial DNA analysis of the Jomon skeletons excavated from the Tohoku region of Japan. **DNA Polymorphism**, vol. 16, pp. 287-290 (in Japanese).

Adachi N., Shinoda K., and Umetsu K. (2009a) Mitochondrial DNA analysis of the Jomon skeletons excavated from the northern part of Japan. **DNA Polymorphism**, vol. 17, pp. 265-269 (in Japanese).

Adachi N., Shinoda K., Umetsu K., and Matsumura H. (2009b) Mitochondrial DNA Analysis of Jomon Skeletons From the Funadomari Site, Hokkaido, and Its Implication for the Origins of Native American. **American Journal of Physical Anthropology**, vol. 138, pp. 255-265.

Adachi N., Shinoda K., Umetsu K., Kitano T., Matsumura H., Fujiyama R., Sawada J., and Tanaka M. (2011) Mitochondrial DNA Analysis of Hokkaido Jomon Skeletons: Remnants of Archaic Maternal Lineages at the Southwester n Edge of Former Beringia. **American Journal of Physical Anthropology**, vol. 146, pp. 346-360.

Adachi N., Sawada J., Yoneda M., Kobayashi K., and Itoh S. (2013) Mitochondrial DNA analysis of the human skeleton of the initial Jomon phase excavated at the Yugura cave site, Nagano, Japan. **Anthropological Science**, vol. 121, no. 2, pp. 137-143.

Andrews R.M., Kubacka I., Chinnery P.F., Lightowlers R.N., Turnbull D.M., and Howell N. (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. **Nature Genetics**, vol. 23, no. 2, pp. 147.

Bryant, D. and Moulton, V. (2004) Neighbor-net: an agglomerative method for the construction of phylogenetic networks. **Molecular Biology and Evolution**, vol. 21, pp. 255–265.

Clark P.U., Dyke A.S., Shakun J.D., Carlson A.E., Clark J., Wohlfarth B., Mitrovica J.X., Hostetler S.W., and McCabe A.M. (2009) The Last Glacial Maximum. **Science**, vol. 325, pp. 710-714.

Database of Single Nucleotide Polymorphisms (dbSNP). Bethesda (MD): National Center for Biotechnology Information, National Library of Medicine. (dbSNP Build 135: {build 135}). Available from: http://www.ncbi.nlm.nih.gov/SNP/

Dodo Y. (1982) A metric analysis of Jomon crania from the Tohoku district. **The Journal of Anthropological Society of Nippon**, vol. 90 (Suppl.), pp. 119–128 (in Japanese with English summary).

Dodo Y., Kawakubo Y., Sawada J. and Ishida H. (2012) The Ainu and their neighbors as seen from the perspective of nonmetric cranial trait variation:I. The Hokkaido Ainu as a population among East and Northeast Asian peoples. **Anthropological Science (Japanese series),** vol. 120, pp. 1-13.

Endo Y. and Endo B. (1979) Catalogue of skeletal remains from Neolithic Jomon period in Japan preserved in The University Museum, The University of Tokyo. The University Museum, The University of Tokyo, **Material Reports**, vol. 3 (in Japanese).

Excoffier L., Laval G., and Schneider S. (2005) Arlequin ver. 3.0: An intergrated software pakage for population genetics data analysis. **Evolutionary Bioinfomatics Online**, vol. 1, pp. 47–50.

Felsenstein, J. (1989) PHYLIP - Phylogeny Inference Package (Version 3.2). **Cladistics**, vol. 5**,** pp. 164-166.

Fu Q., Meyer M., Gao X., Stenzel U., Burbano H.A., Kelso J., and Pääbo S. (2013) DNA analysis of an early modern human from Tianyuan Cave, China. **Proceedings of the National Academy of Sciences of the United States of America**, vol. 110, pp. 2223-2227.

Fukase H., Wakebe T., Tsumoto T., Saiki K., Fujita M., and Ishida H. (2012) Facial characteristics of the prehistoric and early-modern inhabitants of the Okinawa islands in comparison to the contemporary people of Honshu. **Anthropological Science**, vol. 120, pp. 23-32.

Fukushima Museum (1988) Sanganji shell mound, investigative report from Fukushima Museum, No. 17, (in Japanese).

Fuller D. Q., Qin L., Zheng Y., Zhao Z., Chen X., Hosaya L.A., and Sun G. (2009) The Domestication Process and Domestication Rate in Rice: Spikelet Bases from the Lower Yangtze. **Science**, vol. 323, pp. 1607-1610.

Funahashi K. (2010) Ritual Tooth Extraction and Social Groupings. Suirensha, Tokyo (in Japanese).

Green R.E., Krause J., Briggs A.W., Maricic T., Stenzel U., Kircher M., Patterson N., Li H., Zhai W., Fritz M.H., Hansen N.F., Durand E.Y., Malaspinas A., Jensen J.D., Marquws-Bonet T., Alkan C., Prüfer K., Meyer M., Hernán A. Burbano, Good J.M, Schultz R., Aximu-Petri A., Butthof A., Höber B., Höffner B., Siegemund M., Weihmann A., Nusbaum C., Lander E.S., Russ C., Novod N., Affourtit J., Egholm M., Verna C., Rudan P., Brajkovic D., Kucan Ž., Gušic I., Doronichev V.B., Golovanova L.V., Lalueza-Fox C., Rasilla M., Fortea J., Rosas A., Schmitz R.W., Johnson P.L.F., Eichler E.E., Falush D., Birney E., Mullikin J.C., Slatkin M., Nielsen R., Kelso J., Lachmann M., Reich D., and Pääbo S. (2010) A Draft Sequence of the Neandertal Genome. **Science**, vol. 328, pp. 710-722.

Haak W., Balanovsky O., Sanchez J.J., Koshel S., Zaporozhchenko V., Adler C.J., Sarkissian C.S.I., Brandt G., Schwarz C., Nicklisch N., Dresely V., Fritsch B., Balanovska E., Villems R., Meller H., Alt K.W., Cooper A., and the Genographic Consortium (2010) Ancient DNA from European Early Neolithic Farmers Reveals Their Near Eastern Affinities. **PLoS Biology**, vol. 8, no. 11, e1000536.

Hanihara K. (1991) Dual structure model for the population history of the Japanese. **Japan Review**, vol. 2, pp. 1–33.

Hanihara T. and Ishida H. (2009) Regional difference in craniofacial diversity and the population history of Jomon Japan. **American Journal of Physical Anthropology**, vol. 139, pp. 311-322.

Harunari H. (2002) A Study of Jomon Society. Hanawa Syobo Company Limited, Tokyo (in Japanese).

Harunari H. and Imamura M. (2004) Real Time of Yayoi Period. Gakusei-sha (in Japanese).

Huson D. H. and Bryant D. (2006) Application of Phylogenetic Networks in Evolutionary Studies. **Molecular Biology and Evolution**, vol. 23, no. 2, pp. 254-267.

Horai S., Hayasaka K., Murayama K., Wate N., Koike H., and Nakai N. (1989) DNA amplification from ancient human skeletal specimens and their sequence analysis. **Proceedings of Japan Academy**, vol. 65, pp. 229–233.

Horai S., Kondo R., Murayama K., Hayashi S., Koike H., and Nakai N. (1991) Phylogenetic affiliation of ancient and contemporary humans inferred from mitochondrial DNA.

Philosophical Transactions of the Royal Society of London B, vol. 333, pp. 409–417.

Horai S., Murayama K., Hayasaka K., Matsubayashi S., Hattori Y., Fucharoen G., Harihara S., Park K.S., Omoto K., and Pan I.H. (1996) mtDNA polymorphism in East Asian Populations, with special reference to the peopling of Japan. **The American Journal of Human Genetics**, vol. 59, pp. 579–590.

Huang X., Kurata N., Wei X., Wang Z.X., Wang A., Zhao Q., Zhao Y., Liu K., Lu H., Li W., Guo Y., Lu Y., Zhou C., Fan D., Weng Q., Zhu C., Huang T., Zhang L.,Wang Y., Feng L., Furuumi H., Kubo T., Miyabayashi T., Yuan X., Xu Q., Dong G., Zhan Q., Li C., Fujiyama A., Toyoda A., Lu T., Feng Q., Qian Q., Li J., and Han B. (2012) A map of rice genome variation reveals the origin of cultivated rice. **Nature**, vol. 490, pp. 497-501.

Huson, D. H. and Bryant, D. (2006) Application of phylogenetic networks in evolutionary studies. **Molecular Biology and Evolution**. vol. 23, pp. 254–267.

Imamura K. (1996) Prehistoric Japan: new perspectives on insular East Asia. Honolulu: University of Hawaii press.

Ishida H. (1988) Morphological studies of Okhotsk crania from Omisaki. Hokkaido. **Journal of Anthropological Society of Nippon**, vol. 96, pp. 17–45.

Ishida H. (1996) Metric and nonmetric cranial variation of the pre-historic Okhotsk people. **Anthropological Science**, vol. 104, pp. 233–258.

Jacob T. (1967) Some Problems Pertaining to the Racial History of the Indonesian Region: A Study of Human Skeletal and Dental Remains from Several Prehistoric Sites in Indonesia and Malaysia. PhD dissertation, University of Utrecht.

Japanese Archipelago Human Population Genetics Consortium (Jinam T. A., Nishida N., Hirai M., Kawamura S., Oota H., Umetsu K., Kimura R., Ohashi J., Tajima A., Yamamoto T., Tanabe H., Mano S., Suto Y., Kaname T., Naritomi K., Yanagi K., Niikawa N., Omoto K., Tokunaga K., and Saitou N.) (2012) The history of human populations in the Japanese Archipelago inferred from genome-wide SNP data with a special reference to the Ainu and the Ryukyuan populations. **Journal of Human Genetics**, vol. 57, pp. 787-795.

Kaburagi M. Ushida H., Goto M., and Hanihara T. (2010) Comparative studies of the Ainu, their

ancestors, and neighbors: assessment based on metric and nonmetric dental data. **Anthropological Science**, vol. 118, pp. 95-106.

Kanzawa-Kiriyama H., Saso A., Suwa G., and Saitou N. (2013) Ancient mitochondrial DNA sequences of Jomon teeth samples from Sanganji, Tohoku district, Japan. **Anthropological Science**, vol. 121, no 2, pp. 89-103.

Kazuta H., Sato T., Dodo Y., Matsumura H., Amano T., Ishisa H., and Masuda R (2011) Genotype frequencies of the *ABCC11* gene in 2000–3000-year-old human bones from the Epi-Jomon and Jomon sites in Hokkaido, Japan. **Anthropological Science**, vol. 119, pp. 81-86.

Kobayashi K. (2008) The period of Jomon pottery in eastern Japan. Complete guide to Jomon pottery, **UMpromotion**, pp. 896-903 (in Japanese).

Komesu A., Hanihara T., Amano T., Ono H., Yoneda M., Dodo Y., Fukumine T., and Ishida H. (2008) Nonmetric cranial variation in human skeleton remains associated with Okhotsk culture. **Anthropological Science**, vol. 116, pp. 33–47.

Kondo O. (1994) The Skulls of Ubayama Shell-mounds. II. An analysis of intra- and inter-regional variation of the Jomon population. **Anthropological Science**, vol. 102, pp. 59–74.

Kozintsev A.G. (1990) Ainu, Japanese, their ancestors and neigh-bors: cranioscopic data. Journal of Anthropological Society of Nippon, vol. 98, pp. 247–267.

Kozintsev A.G. (1992) Prehistoric and recent populations of Japan: multivariate analysis of cranioscopic data. Arctic Anthropology, vol. 29, pp. 104–111.

Lambeck K., Esat T.M., and Potter E.K. (2002) Links between climate and sea levels for the past three million years. **Nature**, vol 419, pp. 199-206.

Lee H.Y., Yoo J.E., Park M.J., Chung U., and Shin K.J. (2006) Mitochondrial DNA control region sequences in Korean: identification of useful variable sites and phylogenetic analysis for mtDNA data quality control. **International Journal of Legal Medicine**, vol. 120, pp. 5–14.

Li H. and Durbin R. (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform. **Bioinformatics**, vol. 25, pp. 1754-60.

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin

R. and 1000 Genome Project Data Processing Subgroup (2009) The Sequence alignment/map (SAM) format and SAMtools. **Bioinformatics**, vol. 25, pp. 2078-9.

Li, J.Z., Absher D.M., Tang H., Southwick A.M., Casto A.M., Ramachandran S., Cann H.M. Barsh G.S., Feldman M., Cavalli-Sforza L.L., and Myers R.M. (2008) Worldwide human relationships inferred from genome-wide patterns of variation. **Science**, vol. 319, pp. 1100–1104.

Maruyama S., Minaguchi K., and Saitou N. (2003) Sequence polymorphisms of the mitochondrial DNA control region and phylogenetic analysis of mtDNA lineages in the Japanese population. **International Journal of Legal Medicine**, vol. 117, pp. 218–225.

Matsukusa H., Oota H., Haneji K., Toma T., Kawamura S., and Ishida H. (2010) A Genetic Analysis of the Sakishima Islanders Reveals No Relationship With Taiwan Aborigines but Shared Ancestry With Ainu and Main-Island Japanese. **American Journal of Physical Anthropology**, vol. 142, pp. 211-223.

Matsumura H. (2007) Non-metric dental trait variation among local sites and regional groups of the Neolithic Jomon period, Japan. **Anthropological Science**, vol. 115, pp. 25-33.

Matsumura H., Ishida H., Amano T., Ono H., and Yoneda M. (2009) Biological affinities of Okhotsk-culture people with East Siberians and Arctic people based on dental characteristics. **Anthropological Science**, vol. 117, no. 2, pp. 121-132.

Melchior L., Lynnerup N., Siegismund H. R., Kivisild T. and Dissing J (2010) Genetic Diversity among Ancient Nordic Populations. **PLoS ONE**, vol. 5, no. 7, e11898.

Meyer M., Kircher M., Gansauge M., Li H., Racimo F., Mallick S., Schraiber J.G., Jay F., Prüfer K., Filippo C., Sudmant P.H., Alkan C., Fu Q., Do R., Rohland N., Tandon A., Siebauer M., Green R.E., Bryc K., Briggs A.W., Stenzel U., Dabney J., Schendure J., Kitzman J., Hammer M.F., Shunkov M.V., Derevianko A.P., Patterson N., Andres A.M., Eichler E.E., Slatkin M., Reich D., Kelso J., and Pääbo S. (2012) A High-Coverage Genome Sequence from an Archaic Denisovan Individual. **Science**, vol. 338, pp. 222-226.

Michalakis, Y. and Excoffier, L. (1996) A generic estimation of population subdivision using distances between alleles with special reference to microsatellite loci. **Genetics**, vol. 142, pp.

1061-1064.

Miller W., Drautz D.I., Ratan A., Pusey B., Qi J., Lesk A.M., Tomsho L.P., Packard M.D., Zhao F., Sher A., Tikhonov A., Raney B., Patterson N., Lindblad-Toh K., Lander E.S., Knight J.R., Irzyk G.P., Fredrikson K.M., Harkins T.T., Sheridan S., Pringle T., and Schuster C. (2008) Sequencing the nuclear genome of the extinct woolly mammoth. **Nature**, vol. 456, pp. 387-390.

Nakahashi T. (2005) Origin of Japanese (in Japanese). Kodansya.

Nakashima A., Ishida H., Shigematsu M., Goto M., and Hanihara T. (2010) Nonmetric Cranial Variation of Jomon Japan: Implications for the Evolution of Eastern Asian Diversity of Eastern Asian Diversity. **American Journal of Human Biology**, vol. 22, pp. 782-790.

Nei M. (1995) The origins of human populations: genetic, linguistic, and archaeological data. In: Brenner S., Hanihara K., editors. The origin and past of modern human as viewed from DNA. World Scientific, Singapore, pp 71–91.

Omono K. and Saitou N. (1997) Genetic Origins of the Japanese: A Partial Support for the Dual Structure Hypothesis. **American Journal of Physical Anthropology**, vol. 102, pp. 437-446.

Ossenberg N.S., Dodo Y., Maeda T., and Kawakubo Y. (2006) Ethnogenesis and craniofacial change in Japan from the perspective of nonmetric traits. **Anthropological Science**, vol. 114, pp. 99-115.

Pääbo S., Poinar H., Serre D., Jaenicke-Després V., Hebler J., Rohland N., Kuch M., Krause J., Linda V., and Hofreiter M. (2004) Genetic Analysis from Ancient DNA. **Annual Review of Genetics**, vol. 38, pp. 645-679.

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

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**, vol. 112, 145037.

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R., Bender D., Maller J., Sklar P., de Bakker P.I.W., Daly M.J., and Sham P.C. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. **American Journal of Human Genetics**,

vol. 81, pp. 559–575.

Raghavan M., Skoglund P., Graf K.E., Metspalu M., Albrechtsen A., Moltke I., Rasmussen S., Stafford Jr T.W., Orlando L., Metspalu E., Karmin M., Tambets K., Rootsi S., Mägi R., Campos P.F., Balanovska E., Balanovsky O., Khusnutdinova E., Litvinov S., Osipova L.P., Fedorova S.A., Voevoda M.I., DeGiorgio M., Sicheritz-Ponten T., Brunak S., Demeshchenko S., Kivisild T., Villems R., Nielsen R., Jakobsson M., and Willerslev E. (2013) Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. **Nature**, Published online 20 November.

Rasmussen M., Li Y., Lindgreen S., Pedersen J.S., Albrechtsen A., Moltke L., Metspalu M., Metspalu E., Kivisild T., Gupta R., Bertalan M., Nielsen K., Gilbert M.T.P., Wang Y., Raghavan M., Campos P.F., Kamp H.M., Wilson A.S., Gledhill A., Tridico S., Bunce M., Lorenzen E.D., Binladen J., Guo X., Zhao J., Zhang X., Zhang H., Li Z., Chen M., Orlando L., Kristiansen K., Bak M., Tommerup N., Bendixen C., Pierre T.L., Grønnow B., Meldgaard M., Andreasen C., Fedorova S.A., Osipova L.P., Higham T.F.G., Ramsey C.B., Hansen T.v.O., Nielsen F.C., Crawford M.H., Brunak S., Sicheritz-Ponte´n T., Villems R., Nielsen R., Krogh A., Wang J., Willerslev E. (2010) Ancient human genome sequence of an extinct Palaeo-Eskimo. **Nature**, vol. 463, pp. 757-762.

Rasmussen M., Guo X., Wang Y., Lohmueller K.E., Rasmussen S., Albrechtsen A., Skotte L., Lindgreen S., Metspalu M., Jombart T., Kivisild T., Zhai W., Eriksson A., Manica A., Orlando L., de la Vega F., Tridico S., Metspalu E.,Nielsen K., ávila-Arcos M.C., Víctor Moreno-Mayar J., Muller C., Dortch J., Gilbert M.T.P., Lund O., Wesolowska A.,Karmin M., Weinert L.A., Wang B., Li J., Tai S., Xiao F., Hanihara T., van Driem G., Jha AR., Ricaut F., De knijff P., B migliano A., Gallego-Romero I., Kristiansen K., Lambert D.M., Brunak S., Forster P., Brinkmann B., Nehlich O.,Bunce M., Richards M., Gupta R., Bustamante C.D., Krogh A., Foley R.A., Lahr M.M., Balloux F., Sicheritz-Pontén T.,Villems R., Nielsen R., Jun W., Willerslev E. (2011) An Aboriginal Australian Genome Reveals Separate Human Dispersals into Asia. **Science**, vol. 334, pp. 94-98.

R Development Core Team (2010) R: A language and environment forstatistical computing. R

Foundation for Statistical Computing,Vienna, Austria.

Reich D., Green R.E., Kircher M., Krause J., Patterson N., Durand E.Y., Viola B., Briggs A.W., Stenzel U., Johnson P.L., Maricic T., Good J.M., Marques-Bonet T., Alkan C., Fu Q., Mallick S., Li H., Meyer M., Eichler E.E., Stoneking M., Richards M., Talamo S., Shunkov M.V., Derevianko A.P., Hublin J.J., Kelso J., Slatkin M., Paabo S. (2010) Genetic history of an archaic hominin group from Denisova Cave in Siberia. **Nature**, vol. 468, pp. 1053-1060.

Reich D., Patterson N., Kircher M., Delfin F., Nandineni M.R., Pugach I., Ko A.M.S., Ko Y.C., Jinam T.A., Phipps M.E., Saitou N., Wollstein A., Kayser M., Paabo S., and Stoneking (2011) Denisova Admixture and the First Modern Human Dispersals into Southeast Asia and Oceania. **The American Journal of Human Genetics**, vol. 89, pp. 516-528.

Rohland N. and Hofreiter M. (2007) Comparison and optimization of ancient DNA extraction. **Bio Techniques**, vol. 42, pp. 343-352.

Saitou N. and Nei M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. **Molecular Biology and Evolution**, vol. 4, pp. 406-425.

Saso A. Matsukawa S. and Gen S. (2011) Comparative analysis of the glabellar region morphology of the late Pleistocene Minatogawa crania: a three-dimensional approach. **Anthropological Science** vol. 119, pp. 113-121.F

Sato T., Amano T., Ono H., Ishida H., Kodera H., Matsumura H., Yoneda M., and Masuda R. (2009) Mitochondrial DNA haplogrouping of the Okhotsk people based on analysis of ancient DNA: an intermediate of gene flow from the continental Sakhalin people to the Ainu. **Anthropological Science**, vol. 117, no. 3, pp. 171-180.

Scally A. and Durbin R. (2012) Revising the human mutation rate: implications for understanding human evolution. **Nature Reviews**, vol. 13, pp. 745-753.

Schurr T.G., Sukernik R.I., Starikovskaya E.B., and Wallace D.C. (1999) Mitochondrial DNA variation in Koryaks and Itel'men: population replacement in the Okhotsk sea-Bering sea region during Neolithic. **American Journal of Physical Anthropology**, vol. 108, pp. 1–39.

Shinoda K. and Kanai S. (1999) Intracemetry genetic analysis at the Nakazuma Jomon site in Japan by mitochondrial DNA sequencing. **Anthropological Science**, vol. 107, pp. 129–140.

Shinoda K. (2003) DNA analysis of the Jomon skeletal remains excavated from Shimo-Ohta shell midden, Chiba prefecture. **Report for Sohnan Research Institute for Cultural Properties**, vol. 50, pp. 201–205 (in Japanese).

Skoglund P. and Jakobsson M. (2011) Archaic human ancestry in East Asia. **Proceedings of the National Academy of Sciences of the United States of America**, vol. 108, no. 45, pp. 18301-18306.

Starikovskaya E.B., Sukernik R.I., Schurr T.G., Kogelnik A., and Wallace D.C. (1998) Mitochondrial DNA diversity in Chukchi and Siberian Eskimos: implications for the genetic history of ancient Beringia and the peopling of the New World. **The American Journal of Human Genetics**, vol. 63, pp. 1473–1491.

Starikovskaya E.B., Sukernik R.I., Derbeneva O.A., Volodko N.V., Ruiz-Pesini E., Torroni A., Brown M.D., Lott M.T., Hosseini S.H., Huoponen K., and Wallace D.C. (2005) Mitochondrial DNA diversity in indigenous populations of the southern extent of Siberia, and the origins of native American haplogroups. **Annals of Human Genetics**, vol. 69, pp. 67–89.

Suwa G., Fukase H., Kono T. R., Kubo D., and Fujita M. (2011) Mandibular tooth root size in modern Japanese, prehistoric Jomon, and Late Pleistocene Minatogawa human fossils. **Anthropological Science**, vol. 119, pp. 159-171.

Tajima A., Hayami M., Tokunaga K., Juji T., Matsuo M., Marzuki S., Omoto K., and Horai S. (2004) Genetic origins of the Ainu inferred from combined DNA analyses of maternal and paternal lineages. **Journal of Human Genetics**, vol. 49, pp. 187-193.

Takigawa W. (2006) Inter-regional Variation of Metric Traits of Limb Bones in Jomon and Modern Japanese. Anthropological Science, vol. 114, no. 2, pp. 101-129 (in Japanese).

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. **Molecular Biology and Evolution**, **vol.** 28, pp. 2731-2739.

Tanaka M., Cabrera V.M., González A.M., Larruga J.M., Takeyasu T., Fuku N., Guo L.J., Hirose R., Fujita Y., Kurata M., Shinoda K., Umetsu K., Yamada Y., Oshida Y., Sato Y., Hattori N., Mizuno Y., Arai Y., Hirose N., Ohta S., Ogawa O., Tanaka Y., Kawamori R., Shamoto-Nagai

M., Maruyama W., Shimokawa H., Suzuki R., and Shimodaira H. (2004) Mitochondrial Genome Variation in Eastern Asia and the Peopling of Japan. **Genome Research**, vol. 14, pp. 1832–1850.

The 1000 Genomes Project Consortium (2012) An integrated map of genetic variation from 1,092 human genomes. **Nature**, vol. 491, pp. 56-65.

The Chimpanzee Sequencing and Analysis Consortium. (2005) Initial sequence of the chimpanzee genome and comparison with the human genome. **Nature**, vol. 437, pp. 69-87.

The International HapMap Consortium (2005) A haplotype map of the human genome. **Nature**, vol. 437, pp. 1299–1320.

Trejaut J.A., Kivisild T., Loo J.H., Lee C.L., He C.L., Hsu C.J., Li Z.Y., and Lin M. (2005) Traces of Archaic Mitochondrial Lineages Persist in Austronesian-Speaking Formosan populations. **PLoS Biology**, vol. 3, no. 8, e247.

Turner C.G. II (1987) Late Pleistocene and Holocene population history of East Asia based on dental variation. **American Journal of Physical Anthropology**, vol. 73, pp. 305–321.

Turner C.G. II (1990) Major features of sundadonty and sinodonty, including suggestions about East Asian microevolution, population history, and late Pleistocene relationships with Australian Aboriginals. **American Journal of Physical Anthropology**, vol. 82, pp. 295-317.

Umetsu K., Tanaka M., Yuasa I., Adachi N., Miyoshi A., Kashimura S., Park K.S., Wei Y.H., Watanabe G., and Osawa M. (2005) Multiplex amplified product-length polymorphism analysis of 36 mitochondrial single-nucleotide polymorphisms for haplogrouping of East Asian populations. **Electrophoresis**, vol., 26, pp. 91-98.

Van Oven M. and Kayser M. (2008) Updated Comprehensive Phylogenetic Tree of Global Human Mitochondrial DNA Variation. **Human Mutation**, vol., 30, E386-E394.

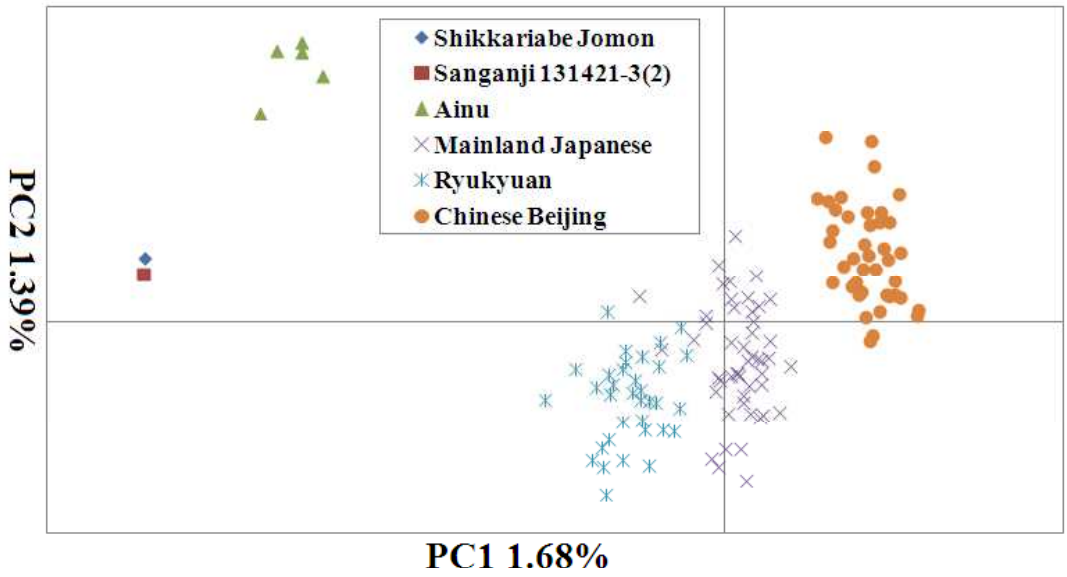
Weir B.S. and Cockerham C.C. (1984) Estimating F-statistics for the analysis of population structure. **Evolution**, vol., 38, pp. 1358-1370.

Yao Y.G., Kong Q.P., Bandelt H.J., Kivisild T., and Zhang Y.P. (2002) Phylogeographic Differentiation of Mitochondrial DNA in Han Chinese. **The American Journal of Human Genetics**, vol. 70, pp. 635–651.

Yamaguchi B. (1982) A review of the osteological characteristics of the Jomon population in prehistoric Japan. **The Journal of Anthropological Society of Nippon**, vol. 90 (Suppl.), pp. 77–90.

Yamaguchi B. (1999) Establishment of Japanese. Misuzu Shobo (in Japanese).

Yamaguchi-Kabata Y., Nakazono K., Takahashi A., Saito S., Hosono N., Kubo M., Nakamura Y., and Kamatani N. (2008) Japanese Population Structure, Based on SNP Genotypes from 7003 Individuals Compared to Other Ethnic Groups: Effects on Population-Based Association Studies. **The American Journal of Human Genetics**, vol. 83, pp. 445–456.



Appendix Figure 1: Genetic relationship among Sanganji 131421-3(2) and the individuals of three populations inhabits in Japanese archipelago (Ainu, mainland Japanese, and Ryukyuan) and of Chinese Beijing (CHB)

Appendix Table 1: Summary of *D*-statistic analysis of Shikkariabe Jomon, HGDP populations, Vindija Neanderthal, Denisovan, and Chimpanzee based on genome wide SNP data

Outgroup	H1	H2	H3	D-statistic	Z-score	SE
Chimpanzee	Denisovan	San_HGDP00991	Shikkariabe Jomon	0.0001	0.007	0.01429
Chimpanzee	Denisovan	San_HGDP00991	Bedouin_HGDP00607	-0.0013	-0.174	0.00747
Chimpanzee	Denisovan	San_HGDP00991	French_HGDP00511	-0.0021	-0.268	0.00784
Chimpanzee	Denisovan	San_HGDP00991	Balochi_HGDP00052	-0.0024	-0.319	0.00752
Chimpanzee	Denisovan	San_HGDP00991	Cambodian_HGDP00711	-0.0019	-0.234	0.00812
Chimpanzee	Denisovan	San_HGDP00991	Dai_HGDP01307	0.0054	0.712	0.00758
Chimpanzee	Denisovan	San_HGDP00991	Han_HGDP00785	0.0041	0.567	0.00723
Chimpanzee	Denisovan	San_HGDP00991	Japan_HGDP00757	-0.0027	-0.355	0.00761
Chimpanzee	Denisovan	San_HGDP00991	Melanesian_HGDP00661	0.022	2.791	0.00788
Chimpanzee	Denisovan	San_HGDP00991	Papuan_HGDP00544	0.0199	2.617	0.00760
Chimpanzee	Denisovan	San_HGDP00991	Karitiana_HGDP01006	0.0041	0.545	0.00752
Chimpanzee	Vindija Neanderthal	San_HGDP00991	Shikkariabe Jomon	0.0358	4.489	0.00798
Chimpanzee	Vindija Neanderthal	San_HGDP00991	Bedouin_HGDP00607	0.03	3.839	0.00781
Chimpanzee	Vindija Neanderthal	San_HGDP00991	French_HGDP00511	0.0339	4.613	0.00735
Chimpanzee	Vindija Neanderthal	San_HGDP00991	Balochi_HGDP00052	0.0277	3.832	0.00723
Chimpanzee	Vindija Neanderthal	San_HGDP00991	Cambodian_HGDP00711	0.0289	3.728	0.00775
Chimpanzee	Vindija Neanderthal	San_HGDP00991	Dai_HGDP01307	0.043	5.637	0.00763
Chimpanzee	Vindija Neanderthal	San_HGDP00991	Han_HGDP00785	0.0327	4.219	0.00775
Chimpanzee	Vindija Neanderthal	San_HGDP00991	Japan_HGDP00757	0.0346	4.743	0.00729
Chimpanzee	Vindija Neanderthal	San_HGDP00991	Melanesian_HGDP00661	0.0506	6.446	0.00785
Chimpanzee	Vindija Neanderthal	San_HGDP00991	Papuan_HGDP00544	0.0392	4.941	0.00793
Chimpanzee	Vindija Neanderthal	San_HGDP00991	Karitiana_HGDP01006	0.0358	5.029	0.00712
Chimpanzee	Denisovan	French_HGDP00511	Shikkariabe Jomon	0.0027	0.321	0.00841
Chimpanzee	Denisovan	French_HGDP00511	Cambodian_HGDP00711	0.0002	0.028	0.00714
Chimpanzee	Denisovan	French_HGDP00511	Dai_HGDP01307	0.0085	1.066	0.00797
Chimpanzee	Denisovan	French_HGDP00511	Han_HGDP00785	0.0067	0.859	0.00780
Chimpanzee	Denisovan	French_HGDP00511	Japan_HGDP00757	-0.0008	-0.095	0.00842
Chimpanzee	Denisovan	French_HGDP00511	Melanesian_HGDP00661	0.0282	3.471	0.00812
Chimpanzee	Denisovan	French_HGDP00511	Papuan_HGDP00544	0.0259	3.129	0.00828
Chimpanzee	Denisovan	French_HGDP00511	Karitiana_HGDP01006	0.0071	0.894	0.00794
Chimpanzee	Vindija Neanderthal	French_HGDP00511	Shikkariabe Jomon	0.0026	0.314	0.00828
Chimpanzee	Vindija Neanderthal	French_HGDP00511	Bedouin_HGDP00607	-0.0049	-0.592	0.00828
Chimpanzee	Vindija Neanderthal	French_HGDP00511	Balochi_HGDP00052	-0.0073	-0.919	0.00794
Chimpanzee	Vindija Neanderthal	French_HGDP00511	Cambodian_HGDP00711	-0.0062	-0.77	0.00805
Chimpanzee	Vindija Neanderthal	French_HGDP00511	Dai_HGDP01307	0.0101	1.262	0.00800
Chimpanzee	Vindija Neanderthal	French_HGDP00511	Han_HGDP00785	-0.0017	-0.214	0.00794
Chimpanzee	Vindija Neanderthal	French_HGDP00511	Japan_HGDP00757	0.0009	0.107	0.00841
Chimpanzee	Vindija Neanderthal	French_HGDP00511	Melanesian_HGDP00661	0.0193	2.417	0.00799
Chimpanzee	Vindija Neanderthal	French_HGDP00511	Papuan_HGDP00544	0.0061	0.744	0.00820
Chimpanzee	Vindija Neanderthal	French_HGDP00511	Karitiana_HGDP01006	0.0029	0.375	0.00773
Chimpanzee	San_HGDP00991	Shikkariabe Jomon	French_HGDP00511	0.0071	0.993	0.00715
Chimpanzee	San_HGDP00991	Shikkariabe Jomon	Bedouin_HGDP00607	0.0016	0.236	0.00678
Chimpanzee	San_HGDP00991	Shikkariabe Jomon	Balochi_HGDP00052	0.0054	0.812	0.00665
Chimpanzee	San_HGDP00991	Shikkariabe Jomon	Cambodian_HGDP00711	-0.0002	-0.029	0.00690
Chimpanzee	San_HGDP00991	Shikkariabe Jomon	Cambodian_HGDP00715	-0.0053	-0.748	0.00709
Chimpanzee	San_HGDP00991	Shikkariabe Jomon	Cambodian_HGDP00720	-0.0017	-0.249	0.00683
Chimpanzee	San_HGDP00991	Shikkariabe Jomon	Dai_HGDP01307	-0.0045	-0.628	0.00717
Chimpanzee	San_HGDP00991	Shikkariabe Jomon	Dai_HGDP01311	-0.0023	-0.343	0.00671
Chimpanzee	San_HGDP00991	Shikkariabe Jomon	Dai_HGDP01315	-0.0041	-0.592	0.00693
Chimpanzee	San_HGDP00991	Shikkariabe Jomon	Han_HGDP00785	-0.0064	-0.924	0.00693
Chimpanzee	San_HGDP00991	Shikkariabe Jomon	Han_HGDP00822	0.0003	0.043	0.00698
Chimpanzee	San_HGDP00991	Shikkariabe Jomon	Han_HGDP01290	0	-0.006	0.00000
Chimpanzee	San_HGDP00991	Shikkariabe Jomon	Japan_HGDP00757	-0.0025	-0.357	0.00700
Chimpanzee	San_HGDP00991	Shikkariabe Jomon	Japan_HGDP00769	-0.0026	-0.364	0.00714
Chimpanzee	San_HGDP00991	Shikkariabe Jomon	Japan_HGDP00828	0.0001	0.012	0.00833

Appendix Table 1: Continued

Outgroup	H1	H2	H3	D-statistic	Z-score	SE
Chimpanzee	San_HGDP00991	Shikkariabe Jomon	Melanesian_HGDP00661	-0.0226	-3.334	0.00678
Chimpanzee	San_HGDP00991	Shikkariabe Jomon	Melanesian_HGDP00787	-0.0145	-2.014	0.00720
Chimpanzee	San_HGDP00991	Shikkariabe Jomon	Melanesian_HGDP01027	-0.0125	-1.788	0.00699
Chimpanzee	San_HGDP00991	Shikkariabe Jomon	Papuan_HGDP00544	-0.009	-1.304	0.00690
Chimpanzee	San_HGDP00991	Shikkariabe Jomon	Papuan_HGDP00549	-0.0127	-1.726	0.00736
Chimpanzee	San_HGDP00991	Shikkariabe Jomon	Papuan_HGDP00554	-0.0112	-1.576	0.00711
Chimpanzee	San_HGDP00991	Shikkariabe Jomon	Karitiana_HGDP01006	-0.006	-0.854	0.00703
Chimpanzee	San_HGDP00991	Shikkariabe Jomon	Karitiana_HGDP01014	-0.0113	-1.668	0.00677
Chimpanzee	San_HGDP00991	Shikkariabe Jomon	Karitiana_HGDP01019	-0.01	-1.49	0.00671
Chimpanzee	French_HGDP00511	Shikkariabe Jomon	Bedouin_HGDP00607	0.0367	5.352	0.00686
Chimpanzee	French_HGDP00511	Shikkariabe Jomon	Balochi_HGDP00052	0.0471	6.825	0.00690
Chimpanzee	French_HGDP00511	Shikkariabe Jomon	Cambodian_HGDP00711	-0.0087	-1.295	0.00672
Chimpanzee	French_HGDP00511	Shikkariabe Jomon	Cambodian_HGDP00715	-0.0021	-0.301	0.00698
Chimpanzee	French_HGDP00511	Shikkariabe Jomon	Cambodian_HGDP00720	-0.0023	-0.311	0.00740
Chimpanzee	French_HGDP00511	Shikkariabe Jomon	Dai_HGDP01307	-0.0086	-1.237	0.00695
Chimpanzee	French_HGDP00511	Shikkariabe Jomon	Dai_HGDP01311	0.0001	0.017	0.00588
Chimpanzee	French_HGDP00511	Shikkariabe Jomon	Dai_HGDP01315	-0.0042	-0.591	0.00711
Chimpanzee	French_HGDP00511	Shikkariabe Jomon	Han_HGDP00785	0.0013	0.188	0.00691
Chimpanzee	French_HGDP00511	Shikkariabe Jomon	Han_HGDP00822	-0.0038	-0.547	0.00695
Chimpanzee	French_HGDP00511	Shikkariabe Jomon	Han_HGDP01290	-0.0072	-0.961	0.00749
Chimpanzee	French_HGDP00511	Shikkariabe Jomon	Japan_HGDP00757	-0.0034	-0.486	0.00700
Chimpanzee	French_HGDP00511	Shikkariabe Jomon	Japan_HGDP00769	-0.006	-0.832	0.00721
Chimpanzee	French_HGDP00511	Shikkariabe Jomon	Japan_HGDP00828	-0.0073	-0.996	0.00733
Chimpanzee	French_HGDP00511	Shikkariabe Jomon	Melanesian_HGDP00661	-0.0338	-4.679	0.00722
Chimpanzee	French_HGDP00511	Shikkariabe Jomon	Melanesian_HGDP00787	-0.0267	-3.709	0.00720
Chimpanzee	French_HGDP00511	Shikkariabe Jomon	Melanesian_HGDP01027	-0.0308	-4.106	0.00750
Chimpanzee	French_HGDP00511	Shikkariabe Jomon	Papuan_HGDP00544	-0.0273	-3.924	0.00696
Chimpanzee	French_HGDP00511	Shikkariabe Jomon	Papuan_HGDP00549	-0.0389	-5.483	0.00709
Chimpanzee	French_HGDP00511	Shikkariabe Jomon	Papuan_HGDP00554	-0.0358	-5.05	0.00709
Chimpanzee	French_HGDP00511	Shikkariabe Jomon	Karitiana_HGDP01006	0.026	3.565	0.00729
Chimpanzee	French_HGDP00511	Shikkariabe Jomon	Karitiana_HGDP01014	0.0198	2.72	0.00728
Chimpanzee	French_HGDP00511	Shikkariabe Jomon	Karitiana_HGDP01019	0.0189	2.623	0.00721
Chimpanzee	Bedouin_HGDP0060	Shikkariabe Jomon	Balochi_HGDP00052	0.0568	7.802	0.00728
Chimpanzee	Bedouin_HGDP0060	Shikkariabe Jomon	Cambodian_HGDP00711	0.0013	0.189	0.00688
Chimpanzee	Bedouin_HGDP0060	Shikkariabe Jomon	Cambodian_HGDP00715	0.0002	0.033	0.00606
Chimpanzee	Bedouin_HGDP0060	Shikkariabe Jomon	Cambodian_HGDP00720	0.0019	0.25	0.00760
Chimpanzee	Bedouin_HGDP0060	Shikkariabe Jomon	Dai_HGDP01307	-0.0007	-0.094	0.00745
Chimpanzee	Bedouin_HGDP0060	Shikkariabe Jomon	Dai_HGDP01311	0.0024	0.316	0.00759
Chimpanzee	Bedouin_HGDP0060	Shikkariabe Jomon	Dai_HGDP01315	-0.0028	-0.388	0.00722
Chimpanzee	Bedouin_HGDP0060	Shikkariabe Jomon	Han_HGDP00785	0.0074	1.015	0.00729
Chimpanzee	Bedouin_HGDP0060	Shikkariabe Jomon	Han_HGDP00822	0	-0.001	0.00000
Chimpanzee	Bedouin_HGDP0060	Shikkariabe Jomon	Han_HGDP01290	0.0072	0.956	0.00753
Chimpanzee	Bedouin_HGDP0060	Shikkariabe Jomon	Japan_HGDP00757	0.0006	0.082	0.00732
Chimpanzee	Bedouin_HGDP0060	Shikkariabe Jomon	Japan_HGDP00769	0.0065	0.85	0.00765
Chimpanzee	Bedouin_HGDP0060	Shikkariabe Jomon	Japan_HGDP00828	-0.0036	-0.502	0.00717
Chimpanzee	Bedouin_HGDP0060	Shikkariabe Jomon	Melanesian_HGDP00661	-0.0257	-3.436	0.00748
Chimpanzee	Bedouin_HGDP0060	Shikkariabe Jomon	Melanesian_HGDP00787	-0.0161	-2.14	0.00752
Chimpanzee	Bedouin_HGDP0060	Shikkariabe Jomon	Melanesian_HGDP01027	-0.027	-3.435	0.00786
Chimpanzee	Bedouin_HGDP0060	Shikkariabe Jomon	Papuan_HGDP00544	-0.0229	-3.004	0.00762
Chimpanzee	Bedouin_HGDP0060	Shikkariabe Jomon	Papuan_HGDP00549	-0.0333	-4.283	0.00777
Chimpanzee	Bedouin_HGDP0060	Shikkariabe Jomon	Papuan_HGDP00554	-0.0271	-3.442	0.00787
Chimpanzee	Bedouin_HGDP0060	Shikkariabe Jomon	Karitiana_HGDP01006	0.0098	1.317	0.00744
Chimpanzee	Bedouin_HGDP0060	Shikkariabe Jomon	Karitiana_HGDP01014	0.0051	0.695	0.00734
Chimpanzee	Bedouin_HGDP0060	Shikkariabe Jomon	Karitiana_HGDP01019	0.0017	0.239	0.00711

Appendix Table 1: Continued

Outgroup	H1	H2	H3	D-statistic	Z-score	SE
Chimpanzee	Balochi_HGDP00052	Shikkariabe Jomon	Cambodian_HGDP00711	-0.0004	-0.06	0.00667
Chimpanzee	Balochi_HGDP00052	Shikkariabe Jomon	Cambodian_HGDP00715	-0.0016	-0.239	0.00669
Chimpanzee	Balochi_HGDP00052	Shikkariabe Jomon	Cambodian_HGDP00720	0.0099	1.451	0.00682
Chimpanzee	Balochi_HGDP00052	Shikkariabe Jomon	Dai_HGDP01307	0.0015	0.209	0.00718
Chimpanzee	Balochi_HGDP00052	Shikkariabe Jomon	Dai_HGDP01311	-0.0014	-0.188	0.00745
Chimpanzee	Balochi_HGDP00052	Shikkariabe Jomon	Dai_HGDP01315	-0.0006	-0.081	0.00741
Chimpanzee	Balochi_HGDP00052	Shikkariabe Jomon	Han_HGDP00785	0.0053	0.744	0.00712
Chimpanzee	Balochi_HGDP00052	Shikkariabe Jomon	Han_HGDP00822	0.0033	0.48	0.00688
Chimpanzee	Balochi_HGDP00052	Shikkariabe Jomon	Han_HGDP01290	0.0117	1.605	0.00729
Chimpanzee	Balochi_HGDP00052	Shikkariabe Jomon	Japan_HGDP00757	0.0028	0.386	0.00725
Chimpanzee	Balochi_HGDP00052	Shikkariabe Jomon	Japan_HGDP00769	0.0009	0.114	0.00789
Chimpanzee	Balochi_HGDP00052	Shikkariabe Jomon	Japan_HGDP00828	-0.0002	-0.029	0.00690
Chimpanzee	Balochi_HGDP00052	Shikkariabe Jomon	Melanesian_HGDP00661	-0.0318	-4.292	0.00741
Chimpanzee	Balochi_HGDP00052	Shikkariabe Jomon	Melanesian_HGDP00787	-0.0285	-3.693	0.00772
Chimpanzee	Balochi_HGDP00052	Shikkariabe Jomon	Melanesian_HGDP01027	-0.0296	-3.891	0.00761
Chimpanzee	Balochi_HGDP00052	Shikkariabe Jomon	Papuan_HGDP00544	-0.0263	-3.606	0.00729
Chimpanzee	Balochi_HGDP00052	Shikkariabe Jomon	Papuan_HGDP00549	-0.0434	-5.546	0.00783
Chimpanzee	Balochi_HGDP00052	Shikkariabe Jomon	Papuan_HGDP00554	-0.0358	-4.538	0.00789
Chimpanzee	Balochi_HGDP00052	Shikkariabe Jomon	Karitiana_HGDP01006	0.0236	3.315	0.00712
Chimpanzee	Balochi_HGDP00052	Shikkariabe Jomon	Karitiana_HGDP01014	0.0092	1.211	0.00760
Chimpanzee	Balochi_HGDP00052	Shikkariabe Jomon	Karitiana_HGDP01019	0.0139	1.982	0.00701
Chimpanzee	Papuan_HGDP00544	Shikkariabe Jomon	Cambodian_HGDP00711	0.0063	0.83	0.00759
Chimpanzee	Papuan_HGDP00544	Shikkariabe Jomon	Cambodian_HGDP00715	-0.0007	-0.103	0.00680
Chimpanzee	Papuan_HGDP00544	Shikkariabe Jomon	Cambodian_HGDP00720	0.0055	0.734	0.00749
Chimpanzee	Papuan_HGDP00544	Shikkariabe Jomon	Dai_HGDP01307	0.0143	1.972	0.00725
Chimpanzee	Papuan_HGDP00544	Shikkariabe Jomon	Dai_HGDP01311	0.0226	2.951	0.00766
Chimpanzee	Papuan_HGDP00544	Shikkariabe Jomon	Dai_HGDP01315	0.0124	1.68	0.00738
Chimpanzee	Papuan_HGDP00544	Shikkariabe Jomon	Han_HGDP00785	0.025	3.474	0.00720
Chimpanzee	Papuan_HGDP00544	Shikkariabe Jomon	Han_HGDP00822	0.02	2.64	0.00758
Chimpanzee	Papuan_HGDP00544	Shikkariabe Jomon	Han_HGDP01290	0.0047	0.62	0.00758
Chimpanzee	Papuan_HGDP00544	Shikkariabe Jomon	Japan_HGDP00757	0.0059	0.771	0.00765
Chimpanzee	Papuan_HGDP00544	Shikkariabe Jomon	Japan_HGDP00769	0.0114	1.562	0.00730
Chimpanzee	Papuan_HGDP00544	Shikkariabe Jomon	Japan_HGDP00828	0.0098	1.298	0.00755
Chimpanzee	Papuan_HGDP00544	Shikkariabe Jomon	Melanesian_HGDP00661	0.1058	12.592	0.00840
Chimpanzee	Papuan_HGDP00544	Shikkariabe Jomon	Melanesian_HGDP00787	0.1156	13.828	0.00836
Chimpanzee	Papuan_HGDP00544	Shikkariabe Jomon	Melanesian_HGDP01027	0.1117	13.795	0.00810
Chimpanzee	Papuan_HGDP00544	Shikkariabe Jomon	Karitiana_HGDP01006	-0.0171	-2.241	0.00763
Chimpanzee	Papuan_HGDP00544	Shikkariabe Jomon	Karitiana_HGDP01014	-0.0254	-3.303	0.00769
Chimpanzee	Papuan_HGDP00544	Shikkariabe Jomon	Karitiana_HGDP01019	-0.0156	-2.068	0.00754
Chimpanzee	San_HGDP00991	Han_HGDP00785	French_HGDP00511	0.0129	1.919	0.00672
Chimpanzee	San_HGDP00991	Han_HGDP00785	Bedouin_HGDP00607	0.007	1.062	0.00659
Chimpanzee	San_HGDP00991	Han_HGDP00785	Balochi_HGDP00052	0.0112	1.625	0.00689
Chimpanzee	San_HGDP00991	Han_HGDP00785	Cambodian_HGDP00711	0.0064	0.879	0.00728
Chimpanzee	San_HGDP00991	Han_HGDP00785	Cambodian_HGDP00715	0.0009	0.14	0.00643
Chimpanzee	San_HGDP00991	Han_HGDP00785	Cambodian_HGDP00720	0.0048	0.682	0.00704
Chimpanzee	San_HGDP00991	Han_HGDP00785	Dai_HGDP01307	0.0021	0.29	0.00724
Chimpanzee	San_HGDP00991	Han_HGDP00785	Dai_HGDP01311	0.0041	0.588	0.00697
Chimpanzee	San_HGDP00991	Han_HGDP00785	Dai_HGDP01315	0.0023	0.327	0.00703
Chimpanzee	San_HGDP00991	Han_HGDP00785	Japan_HGDP00757	0.0042	0.575	0.00730
Chimpanzee	San_HGDP00991	Han_HGDP00785	Japan_HGDP00769	0.0039	0.517	0.00754
Chimpanzee	San_HGDP00991	Han_HGDP00785	Japan_HGDP00828	0.0067	0.919	0.00729
Chimpanzee	San_HGDP00991	Han_HGDP00785	Melanesian_HGDP00661	-0.0167	-2.437	0.00685
Chimpanzee	San_HGDP00991	Han_HGDP00785	Melanesian_HGDP00787	-0.0085	-1.207	0.00704
Chimpanzee	San_HGDP00991	Han_HGDP00785	Melanesian_HGDP01027	-0.0065	-0.925	0.00703
Chimpanzee	San_HGDP00991	Han_HGDP00785	Papuan_HGDP00544	-0.0029	-0.405	0.00716
Chimpanzee	San_HGDP00991	Han_HGDP00785	Papuan_HGDP00549	-0.0073	-1.022	0.00714
Chimpanzee	San_HGDP00991	Han_HGDP00785	Papuan_HGDP00554	-0.0053	-0.748	0.00709

Appendix Table 1: Continued

Outgroup	H1	H2	H3	D-statistic	Z-score	SE
Chimpanzee	San_HGDP00991	Han_HGDP00785	Karitiana_HGDP01006	-0.0004	-0.058	0.00690
Chimpanzee	San_HGDP00991	Han_HGDP00785	Karitiana_HGDP01014	-0.0052	-0.741	0.00702
Chimpanzee	San_HGDP00991	Han_HGDP00785	Karitiana_HGDP01019	-0.0042	-0.594	0.00707
Chimpanzee	Pygmy_HGDP00449	Han_HGDP00785	French_HGDP00511	0.0015	0.223	0.00673
Chimpanzee	Pygmy_HGDP00449	Han_HGDP00785	Bedouin_HGDP00607	0.0019	0.287	0.00662
Chimpanzee	Pygmy_HGDP00449	Han_HGDP00785	Balochi_HGDP00052	0.0008	0.124	0.00645
Chimpanzee	Pygmy_HGDP00449	Han_HGDP00785	Cambodian_HGDP00711	-0.0023	-0.315	0.00730
Chimpanzee	Pygmy_HGDP00449	Han_HGDP00785	Cambodian_HGDP00715	-0.0001	-0.018	0.00556
Chimpanzee	Pygmy_HGDP00449	Han_HGDP00785	Cambodian_HGDP00720	0.0082	1.146	0.00716
Chimpanzee	Pygmy_HGDP00449	Han_HGDP00785	Dai_HGDP01307	0.008	1.102	0.00726
Chimpanzee	Pygmy_HGDP00449	Han_HGDP00785	Dai_HGDP01311	0.0069	0.932	0.00740
Chimpanzee	Pygmy_HGDP00449	Han_HGDP00785	Dai_HGDP01315	0.0073	1.052	0.00694
Chimpanzee	Pygmy_HGDP00449	Han_HGDP00785	Japan_HGDP00757	0.0006	0.081	0.00741
Chimpanzee	Pygmy_HGDP00449	Han_HGDP00785	Japan_HGDP00769	-0.0039	-0.549	0.00710
Chimpanzee	Pygmy_HGDP00449	Han_HGDP00785	Japan_HGDP00828	0.0003	0.042	0.00714
Chimpanzee	Pygmy_HGDP00449	Han_HGDP00785	Melanesian_HGDP00661	-0.0161	-2.265	0.00711
Chimpanzee	Pygmy_HGDP00449	Han_HGDP00785	Melanesian_HGDP00787	-0.0075	-1.012	0.00741
Chimpanzee	Pygmy_HGDP00449	Han_HGDP00785	Melanesian_HGDP01027	-0.0048	-0.66	0.00727
Chimpanzee	Pygmy_HGDP00449	Han_HGDP00785	Papuan_HGDP00544	-0.0157	-2.225	0.00706
Chimpanzee	Pygmy_HGDP00449	Han_HGDP00785	Papuan_HGDP00549	-0.0107	-1.535	0.00697
Chimpanzee	Pygmy_HGDP00449	Han_HGDP00785	Papuan_HGDP00554	-0.0071	-1	0.00710
Chimpanzee	Pygmy_HGDP00449	Han_HGDP00785	Karitiana_HGDP01006	-0.0038	-0.536	0.00709
Chimpanzee	Pygmy_HGDP00449	Han_HGDP00785	Karitiana_HGDP01014	-0.0075	-1.034	0.00725
Chimpanzee	Pygmy_HGDP00449	Han_HGDP00785	Karitiana_HGDP01019	-0.0029	-0.399	0.00727
Chimpanzee	Yoruba_HGDP00920	Han_HGDP00785	French_HGDP00511	0.0029	0.45	0.00644
Chimpanzee	Yoruba_HGDP00920	Han_HGDP00785	Bedouin_HGDP00607	0.0115	1.811	0.00635
Chimpanzee	Yoruba_HGDP00920	Han_HGDP00785	Balochi_HGDP00052	0.0031	0.492	0.00630
Chimpanzee	Yoruba_HGDP00920	Han_HGDP00785	Cambodian_HGDP00711	-0.0017	-0.252	0.00675
Chimpanzee	Yoruba_HGDP00920	Han_HGDP00785	Cambodian_HGDP00715	-0.0021	-0.309	0.00680
Chimpanzee	Yoruba_HGDP00920	Han_HGDP00785	Cambodian_HGDP00720	0.0069	0.995	0.00693
Chimpanzee	Yoruba_HGDP00920	Han_HGDP00785	Dai_HGDP01307	0.0055	0.766	0.00718
Chimpanzee	Yoruba_HGDP00920	Han_HGDP00785	Dai_HGDP01311	-0.0002	-0.023	0.00870
Chimpanzee	Yoruba_HGDP00920	Han_HGDP00785	Dai_HGDP01315	0.0029	0.438	0.00662
Chimpanzee	Yoruba_HGDP00920	Han_HGDP00785	Japan_HGDP00757	0.0004	0.055	0.00727
Chimpanzee	Yoruba_HGDP00920	Han_HGDP00785	Japan_HGDP00769	0.0009	0.127	0.00709
Chimpanzee	Yoruba_HGDP00920	Han_HGDP00785	Japan_HGDP00828	0.0046	0.669	0.00688
Chimpanzee	Yoruba_HGDP00920	Han_HGDP00785	Melanesian_HGDP00661	-0.0112	-1.583	0.00708
Chimpanzee	Yoruba_HGDP00920	Han_HGDP00785	Melanesian_HGDP00787	-0.0132	-1.987	0.00664
Chimpanzee	Yoruba_HGDP00920	Han_HGDP00785	Melanesian_HGDP01027	-0.0051	-0.7	0.00729
Chimpanzee	Yoruba_HGDP00920	Han_HGDP00785	Papuan_HGDP00544	-0.01	-1.488	0.00672
Chimpanzee	Yoruba_HGDP00920	Han_HGDP00785	Papuan_HGDP00549	-0.0147	-2.173	0.00676
Chimpanzee	Yoruba_HGDP00920	Han_HGDP00785	Papuan_HGDP00554	-0.0069	-0.997	0.00692
Chimpanzee	Yoruba_HGDP00920	Han_HGDP00785	Karitiana_HGDP01006	-0.0006	-0.08	0.00750
Chimpanzee	Yoruba_HGDP00920	Han_HGDP00785	Karitiana_HGDP01014	-0.0105	-1.531	0.00686
Chimpanzee	Yoruba_HGDP00920	Han_HGDP00785	Karitiana_HGDP01019	-0.0027	-0.386	0.00699
Chimpanzee	French_HGDP00511	Han_HGDP00785	Bedouin_HGDP00607	0.0358	5.405	0.00662
Chimpanzee	French_HGDP00511	Han_HGDP00785	Balochi_HGDP00052	0.0463	6.816	0.00679
Chimpanzee	French_HGDP00511	Han_HGDP00785	Cambodian_HGDP00711	-0.0107	-1.594	0.00671
Chimpanzee	French_HGDP00511	Han_HGDP00785	Cambodian_HGDP00715	-0.0038	-0.613	0.00620
Chimpanzee	French_HGDP00511	Han_HGDP00785	Cambodian_HGDP00720	-0.004	-0.599	0.00668
Chimpanzee	French_HGDP00511	Han_HGDP00785	Dai_HGDP01307	-0.0106	-1.616	0.00656
Chimpanzee	French_HGDP00511	Han_HGDP00785	Dai_HGDP01311	-0.0014	-0.211	0.00664
Chimpanzee	French_HGDP00511	Han_HGDP00785	Dai_HGDP01315	-0.0058	-0.873	0.00664
Chimpanzee	French_HGDP00511	Han_HGDP00785	Japan_HGDP00757	-0.0049	-0.737	0.00665
Chimpanzee	French_HGDP00511	Han_HGDP00785	Japan_HGDP00769	-0.0077	-1.177	0.00654
Chimpanzee	French_HGDP00511	Han_HGDP00785	Japan_HGDP00828	-0.0087	-1.296	0.00671

Appendix Table 1: Continued

Outgroup	H1	H2	H3	D-statistic	Z-score	SE
Chimpanzee	French_HGDP00511	Han_HGDP00785	Melanesian_HGDP00661	-0.0357	-5.094	0.00701
Chimpanzee	French_HGDP00511	Han_HGDP00785	Melanesian_HGDP00787	-0.0283	-4.095	0.00691
Chimpanzee	French_HGDP00511	Han_HGDP00785	Melanesian_HGDP01027	-0.0325	-4.303	0.00755
Chimpanzee	French_HGDP00511	Han_HGDP00785	Papuan_HGDP00544	-0.0287	-4.313	0.00665
Chimpanzee	French_HGDP00511	Han_HGDP00785	Papuan_HGDP00549	-0.0407	-5.839	0.00697
Chimpanzee	French_HGDP00511	Han_HGDP00785	Papuan_HGDP00554	-0.0369	-5.456	0.00676
Chimpanzee	French_HGDP00511	Han_HGDP00785	Karitiana_HGDP01006	0.025	3.697	0.00676
Chimpanzee	French_HGDP00511	Han_HGDP00785	Karitiana_HGDP01014	0.0191	2.948	0.00648
Chimpanzee	French_HGDP00511	Han_HGDP00785	Karitiana_HGDP01019	0.018	2.698	0.00667
Chimpanzee	Bedouin_HGDP0060	Han_HGDP00785	Balochi_HGDP00052	0.0502	7.451	0.00674
Chimpanzee	Bedouin_HGDP0060	Han_HGDP00785	Cambodian_HGDP00711	-0.0063	-0.908	0.00694
Chimpanzee	Bedouin_HGDP0060	Han_HGDP00785	Cambodian_HGDP00715	-0.0073	-1.077	0.00678
Chimpanzee	Bedouin_HGDP0060	Han_HGDP00785	Cambodian_HGDP00720	-0.0057	-0.802	0.00711
Chimpanzee	Bedouin_HGDP0060	Han_HGDP00785	Dai_HGDP01307	-0.0086	-1.23	0.00699
Chimpanzee	Bedouin_HGDP0060	Han_HGDP00785	Dai_HGDP01311	-0.0053	-0.74	0.00716
Chimpanzee	Bedouin_HGDP0060	Han_HGDP00785	Dai_HGDP01315	-0.011	-1.541	0.00714
Chimpanzee	Bedouin_HGDP0060	Han_HGDP00785	Japan_HGDP00757	-0.0072	-1.006	0.00716
Chimpanzee	Bedouin_HGDP0060	Han_HGDP00785	Japan_HGDP00769	-0.0012	-0.171	0.00702
Chimpanzee	Bedouin_HGDP0060	Han_HGDP00785	Japan_HGDP00828	-0.0114	-1.642	0.00694
Chimpanzee	Bedouin_HGDP0060	Han_HGDP00785	Melanesian_HGDP00661	-0.0331	-4.656	0.00711
Chimpanzee	Bedouin_HGDP0060	Han_HGDP00785	Melanesian_HGDP00787	-0.0237	-3.333	0.00711
Chimpanzee	Bedouin_HGDP0060	Han_HGDP00785	Melanesian_HGDP01027	-0.0347	-4.822	0.00720
Chimpanzee	Bedouin_HGDP0060	Han_HGDP00785	Papuan_HGDP00544	-0.0303	-4.461	0.00679
Chimpanzee	Bedouin_HGDP0060	Han_HGDP00785	Papuan_HGDP00549	-0.0413	-5.712	0.00723
Chimpanzee	Bedouin_HGDP0060	Han_HGDP00785	Papuan_HGDP00554	-0.0345	-4.789	0.00720
Chimpanzee	Bedouin_HGDP0060	Han_HGDP00785	Karitiana_HGDP01006	0.0023	0.333	0.00691
Chimpanzee	Bedouin_HGDP0060	Han_HGDP00785	Karitiana_HGDP01014	-0.0021	-0.31	0.00677
Chimpanzee	Bedouin_HGDP0060	Han_HGDP00785	Karitiana_HGDP01019	-0.0058	-0.832	0.00697
Chimpanzee	Balochi_HGDP00052	Han_HGDP00785	Cambodian_HGDP00711	-0.006	-0.816	0.00735
Chimpanzee	Balochi_HGDP00052	Han_HGDP00785	Cambodian_HGDP00715	-0.0074	-1.066	0.00694
Chimpanzee	Balochi_HGDP00052	Han_HGDP00785	Cambodian_HGDP00720	0.0051	0.739	0.00690
Chimpanzee	Balochi_HGDP00052	Han_HGDP00785	Dai_HGDP01307	-0.004	-0.573	0.00698
Chimpanzee	Balochi_HGDP00052	Han_HGDP00785	Dai_HGDP01311	-0.0072	-0.988	0.00729
Chimpanzee	Balochi_HGDP00052	Han_HGDP00785	Dai_HGDP01315	-0.006	-0.865	0.00694
Chimpanzee	Balochi_HGDP00052	Han_HGDP00785	Japan_HGDP00757	-0.0029	-0.407	0.00713
Chimpanzee	Balochi_HGDP00052	Han_HGDP00785	Japan_HGDP00769	-0.0049	-0.689	0.00711
Chimpanzee	Balochi_HGDP00052	Han_HGDP00785	Japan_HGDP00828	-0.0056	-0.783	0.00715
Chimpanzee	Balochi_HGDP00052	Han_HGDP00785	Melanesian_HGDP00661	-0.0374	-5.256	0.00712
Chimpanzee	Balochi_HGDP00052	Han_HGDP00785	Melanesian_HGDP00787	-0.0341	-4.714	0.00723
Chimpanzee	Balochi_HGDP00052	Han_HGDP00785	Melanesian_HGDP01027	-0.0356	-4.807	0.00741
Chimpanzee	Balochi_HGDP00052	Han_HGDP00785	Papuan_HGDP00544	-0.0318	-4.509	0.00705
Chimpanzee	Balochi_HGDP00052	Han_HGDP00785	Papuan_HGDP00549	-0.0494	-6.628	0.00745
Chimpanzee	Balochi_HGDP00052	Han_HGDP00785	Papuan_HGDP00554	-0.0411	-5.39	0.00763
Chimpanzee	Balochi_HGDP00052	Han_HGDP00785	Karitiana_HGDP01006	0.0189	2.719	0.00695
Chimpanzee	Balochi_HGDP00052	Han_HGDP00785	Karitiana_HGDP01014	0.0042	0.604	0.00695
Chimpanzee	Balochi_HGDP00052	Han_HGDP00785	Karitiana_HGDP01019	0.009	1.284	0.00701
Chimpanzee	Papuan_HGDP00544	Han_HGDP00785	Cambodian_HGDP00711	-0.0191	-2.637	0.00724
Chimpanzee	Papuan_HGDP00544	Han_HGDP00785	Cambodian_HGDP00715	-0.0263	-3.939	0.00668
Chimpanzee	Papuan_HGDP00544	Han_HGDP00785	Cambodian_HGDP00720	-0.02	-2.831	0.00706

Appendix Table 1: Continued

Outgroup	H1	H2	H3	D-statistic	Z-score	SE
Chimpanzee	Papuan_HGDP00544	Han_HGDP00785	Dai_HGDP01307	-0.0114	-1.715	0.00665
Chimpanzee	Papuan_HGDP00544	Han_HGDP00785	Dai_HGDP01311	-0.0025	-0.349	0.00716
Chimpanzee	Papuan_HGDP00544	Han_HGDP00785	Dai_HGDP01315	-0.0134	-1.942	0.00690
Chimpanzee	Papuan_HGDP00544	Han_HGDP00785	Japan_HGDP00757	-0.0199	-2.682	0.00742
Chimpanzee	Papuan_HGDP00544	Han_HGDP00785	Japan_HGDP00769	-0.0146	-2.148	0.00680
Chimpanzee	Papuan_HGDP00544	Han_HGDP00785	Japan_HGDP00828	-0.0165	-2.383	0.00692
Chimpanzee	Papuan_HGDP00544	Han_HGDP00785	Melanesian_HGDP00661	0.0838	10.373	0.00808
Chimpanzee	Papuan_HGDP00544	Han_HGDP00785	Melanesian_HGDP00787	0.0926	11.362	0.00815
Chimpanzee	Papuan_HGDP00544	Han_HGDP00785	Melanesian_HGDP01027	0.0891	11.256	0.00792
Chimpanzee	Papuan_HGDP00544	Han_HGDP00785	Karitiana_HGDP01006	-0.0427	-5.802	0.00736
Chimpanzee	Papuan_HGDP00544	Han_HGDP00785	Karitiana_HGDP01014	-0.0509	-6.996	0.00728
Chimpanzee	Papuan_HGDP00544	Han_HGDP00785	Karitiana_HGDP01019	-0.0413	-5.456	0.00757

Appendix Table 2: Comparison of *Z*-score estimated from Sequencing data and SNP data (from Reich et al., 2011)

Table 2. Denisovan Admixture $p_D(X)$ Estimated from Sequencing versus Genotyping Data

		Sequencing Data			Genotyping Data		
Sample	**HGDP ID for Sequence Data**	**Estimated Ancestry**	**Standard Error in the Estimate**	**Z Score**	**Estimated Ancestry**	**Standard Error in the Estimate**	**Z Score**
Papuan	HGDP00542	105%	9%	11.8	100%	n/a	n/a
New Guinea Highlander		104%	9%	11.7	100%	n/a	n/a
Bougainville	HGDP00491	83%	10%	8.3	82%	5%	15.9
Mamanwa		28%	10%	2.9	49%	5%	9.2
Cambodian	HGDP00711	19%	9%	2.0	−3%	3%	−0.8
Karitiana	HGDP00998	9%	12%	0.7	4%	6%	0.7
Mongolian	HGDP01224	−6%	12%	−0.5	3%	3%	1.1

For the sequencing data, we present the ratio f_4(Yoruba, Denisova; Han, X)/f_4(Yoruba, Denisova; Han, Papuan2), estimating the proportion of Denisova ancestry in a population X as a fraction of that in the Papuan2 sample (for the first line, the Papuan sample in the numerator is Papuan1 HGDP000551). For the genotyping data, we present the ratio f_4(YRI, Denisova; CHB, X)/f_4(YRI, Denisova; CHB, Papuan). No standard errors are given for the genotyping-based estimates in the first two rows because the Papuans and New Guineans are the reference populations, and so by definition those fractions are 100%.