

**An update and frequency distribution of Y chromosome haplogroups in modern
Japanese males**

Running title: Y chromosome haplogroups

Makoto Inoue¹, Youichi Sato^{1*}

¹ Department of Pharmaceutical Information Science, Tokushima University Graduate
School of Biomedical Sciences, Tokushima, 770-8505, Japan

*Address correspondence to: Youichi Sato, Ph.D.

Department of Pharmaceutical Information Science, Tokushima University Graduate
School of Biomedical Sciences

1-78-1 Sho-machi, Tokushima, Tokushima 770-8505, Japan

Phone: +81-88-633-7253; FAX: +81-88-633-7253

E-mail: youichi.sato@tokushima-u.ac.jp

Abstract

Japanese males belong to the Y chromosome C1a1, C2, D1a2a, D1a2a-12f2b, O1b2, O1b2a1a1, O2a2b1, and O2a1b haplogroups. Notably, the regional frequency of each haplogroup is homogeneous. Owing to recent developments in genome sequencing technology, the phylogenetic tree of Y chromosome haplogroups is updated annually. Therefore, in this study, we aimed to provide an update on the Y chromosome haplogroups of modern Japanese males and examine their regional distributions. Using 1,640 samples of Japanese males from seven Japanese cities (Nagasaki, Fukuoka, Tokushima, Osaka, Kanazawa, Kawasaki, and Sapporo), haplogroups C1a1, C2, D1a2a, D1a2a-12f2b, O1b2, and O1b2a1a1 were updated based on the latest phylogenetic tree. Haplogroup C1a1 was mainly classified into C1a1a1a and C1a1a1b subgroups; C1a1a1b was more common in Tokushima and Osaka than in the other regions. Haplogroup C2 was mainly classified into C2a, C2b1a1a, C2b1a1b, C2b1a2, and C2b1b subgroups and exhibited frequency differences in Osaka. Haplogroup D1a2a was classified into D1a2a1c1 and D1a2a2 subgroups, and its frequency varied between Tokushima and Osaka. Haplogroup D1a2a-12f2b was classified into D1a2a1a2b1a1a and D1a2a1a3 subgroups; however, no significant frequency differences were observed. Haplogroup O1b2 was classified into O1b2a1a2a1a, O1b2a1a2a1b, and O1b2a1a3 subgroups, with frequency differences

between Nagasaki and Kanazawa. Haplogroup O1b2a1a1 was mainly classified into O1b2a1a1a, O1b2a1a1b, and O1b2a1a1c subgroups; however, no significant frequency differences were observed. Our findings suggest that gene flow in the Kinki region is caused by human migration.

Keywords: Y chromosome haplogroup; modern Japanese male; phylogenetic tree; regional distribution

Introduction

The human Y chromosome comprises approximately 50 million base pairs and 106 protein-coding genes [1]. It has a pseudo-autosomal region (PAR) homologous to the X chromosome and male-specific region on the human Y chromosome (MSY), consisting of hetero- and euchromatin [2]. The Y chromosome is passed on from the father to the son in the same manner because only the PAR region of the Y chromosome recombines with that of the X chromosome. Therefore, the Y chromosome in modern humans is a good resource for studying male population genetics, as the DNA of the ancestor is transmitted to the offspring in its original form. The Y chromosome is divided into various groups called haplogroups based on a combination of multiple mutations. Tilford et al. first reported the sequence of the human Y chromosome [3], leading to the subsequent identification of many DNA polymorphisms. Polymorphisms in 47z [4] and SRY 465 [5] have been reported in Japan. In 2002, Hammer et al. summarized the worldwide classification of the Y chromosome and constructed a global phylogenetic tree for males [6], facilitating the study of human global evolution.

Haniwara K. proposed the “dual-structure model,” which is a working hypothesis in Japanese anthropology [7]. This model, which is widely accepted, indicates that the formation of the modern Japanese population is the result of the mixing of

indigenous Jomon and migrating Yayoi lineages. Indigenous Jomon people, originating in Southeast Asia, migrated to the Japanese archipelago from Eurasia via Sakhalin, Kuril Islands, and the Korean Peninsula from the mouth of the Yangtze River or around Taiwan northward across the sea, extending their range of residence from Okinawa to Hokkaido (approximately 40,000–3,000 years ago). The Yayoi people from Northeast Asia, who adopted rice cultivation and agricultural techniques, migrated to the Japanese archipelago from northern Kyushu via the Korean Peninsula and spread to Kyushu, Shikoku, and Honshu (approximately 3,000 years ago in the 3rd century AD). These findings suggest a geographic gradient among the genetically distinct Japanese populations.

Sato et al. analyzed Y chromosome haplogroups in 2,390 samples collected from seven cities (Nagasaki, Fukuoka, Tokushima, Osaka, Kanazawa, Kawasaki, and Sapporo) to identify the possible geographic gradients in Japanese males [8]. They revealed that the Japanese males could be classified into eight haplogroups (C1, C3, D2*, D2a1, O2b*, O2b1, O3a3c, and O3a4). However, no significant regional differences in haplogroup frequencies have been reported.

Recent technological advances have facilitated the whole-genome sequencing of the Japanese population. The International Society annually updates the phylogenetic tree of the Y-chromosome haplogroup for the Study of Genetic Genealogy (ISOGG), and the

haplogroups are renamed accordingly. Haplogroup C1 has been renamed as C1a1, C3 as C2, D2* as D1a2a, O2b* as O1b2, O2b1 as O1b2a1a1, O3a3c as O2a2b1, and O3a4 as O2a1b. In this study, we sought to determine the Y chromosome haplogroups in Japanese males (previously assigned to haplogroups C, D, and O1b2) in more detail to clarify the structure of the phylogenetic tree of the modern Japanese male population.

Materials and methods

Subjects

Of the 2,390 samples collected from the Japanese males residing in Nagasaki, Fukuoka, Tokushima, Osaka, Kanazawa, Kawasaki, and Sapporo (Supplementary Figure S1) for the Y chromosome haplogroup study reported by Sato et al. in 2014 [8], residual DNA samples corresponding to 1,640 individuals were used in the present study. Table 1 summarizes the number of Y chromosome haplogroups used for the analysis in each city. This study was approved by the Ethics Committee of Tokushima University (approval number: R1-19). All participants provided informed consent prior to their inclusion in the study.

Y chromosome haplogroups

Y-chromosome haplogroups were determined based on a phylogenetic tree published by ISOGG (ISOGG Y-DNA Haplogroup Tree Ver. 15.73; July 2020; <https://isogg.org/tree/>). The haplogroup, D2a1 marker 12f2b, was previously deleted by ISSOG. Therefore, we defined D2a1 as D1a2a-12f2b in this study. Y-chromosome haplogroup branch markers were determined using polymerase chain reaction (PCR)-restriction fragment length polymorphism, TaqMan PCR, and Sanger sequencing. All primer sets, PCR annealing

temperatures, restriction enzymes, and genotyping methods used in this study are listed in Supplementary Tables S1–S3.

Data analyses

To compare the Y-chromosome haplogroup diversity among cities, Arlequin Ver. 3.5.2.2 was used to calculate the F-statistics (F_{ST}) as an indicator of genetic differentiation among populations [9]. Under the null hypothesis of no difference between groups, the number of permutations was set to 100, and the significance level was set to 0.05. Each haplogroup frequency was used as an input file for statistical calculations to derive the null distribution of paired F_{ST} values. The estimated year of divergence of the Y-chromosomal haplogroup was obtained using Y-Tree (ver. 11.04) from the yfull.com database (<https://www.yfull.com/tree/>).

Results

Haplogroup C

The haplogroups of 110 individuals belonging to the Y chromosome haplogroup, C1a1 (formerly C1), were analyzed based on the phylogenetic tree published by ISOGG. C1a1 was classified into three downstream subgroups: C1a1a (1.8%), C1a1a1a (73.6%), and C1a1a1b (24.5%; Figure 1A and Supplementary Table S4). Evaluation of the regional frequencies of these three groups revealed that the frequencies of haplogroups C1a1a1a and C1a1a1b differed between Tokushima and Osaka. Frequencies of C1a1a1a were higher in Nagasaki (90.0%), Fukuoka (100%), Kanazawa (90.0%), Kawasaki (78.9%), and Sapporo (66.7%) compared with those in Tokushima (56.5%) and Osaka (54.5%). In contrast, the frequencies of C1a1a1b were lower in Nagasaki (0%), Fukuoka (0%), Kanazawa (10.0%), Kawasaki (21.1%), and Sapporo (28.6%) compared with those in Tokushima (43.5%) and Osaka (45.5%) (Figure 1B and Supplementary Table S4). To compare the diversity among the seven populations, a pairwise comparison of F_{ST} values was performed based on the frequency of haplogroup C1a1 lineages. The results showed significant differences ($P < 0.05$) between Tokushima and Nagasaki ($F_{ST} = 0.229$), Tokushima and Kanazawa ($F_{ST} = 0.207$), Osaka and Nagasaki ($F_{ST} = 0.261$), and Osaka and Kanazawa ($F_{ST} = 0.249$) regions (Table 2).

Next, haplogroup analysis was performed on 130 individuals belonging to the Y chromosome haplogroup, C2 (formerly C3). C2 was further divided into six subgroups: C2a (14.6%), C2b1 (0.8%), C2b1a1a (36.9%), C2b1a1b (4.6%), C2b1a2 (24.6%), and C2b1b (18.5%; Figure 1A and Supplementary Table S4). Regional comparisons showed that the C2a, C2b1a1a, C2b1a2, and C2b1b subgroups exhibited frequency variations in Osaka. Frequencies of C2a were lower in Nagasaki (6.7%), Fukuoka (0%), Tokushima (11.8%), Kanazawa (15.6%), Kawasaki (9.5%), and Sapporo (20.0%) than in Osaka (42.9%). Moreover, frequencies of C2b1a2 were lower in Nagasaki (13.3%), Fukuoka (25.0%), Tokushima (11.8%), Kawasaki (19.0%), and Sapporo (23.3%) than in Osaka (42.9%) and Kanazawa (37.5%). In contrast, C2b1a1a was commonly observed in Nagasaki (53.3%), Fukuoka (50.0%), Tokushima (35.3%), Kanazawa (34.4%), Kawasaki (47.6%), and Sapporo (30.0%), but not in Osaka (0%). C2b1b was commonly observed in Nagasaki (26.7%), Fukuoka (25.0%), Tokushima (29.4%), Kanazawa (9.4%), Kawasaki (23.8%), and Sapporo (16.7%), but not in Osaka (0%) (Figure 1C and Supplementary Table S4). The F_{ST} values differed significantly ($P < 0.05$) between the Osaka and Nagasaki ($F_{ST} = 0.249$), Osaka and Fukuoka ($F_{ST} = 0.200$), Osaka and Tokushima ($F_{ST} = 0.139$), and Osaka and Kawasaki ($F_{ST} = 0.192$) regions (Table 2).

Haplogroup D

A detailed haplogroup analysis was performed on 316 individuals belonging to the Y chromosome haplogroup, D1a2a (formerly D2*). D1a2a was subclassified into 13 subgroups: D1a2a1 (0.3%), D1a2a1c (2.2%), D1a2a1c1 (4.1%), D1a2a1c1a (8.9%), D1a2a1c1a1 (4.7%), D1a2a1c1a1a (3.5%), D1a2a1c1a1b (1.9%), D1a2a1c1a1b1 (19.0%), D1a2a1c1b (1.3%), D1a2a1c1b1 (8.5%), D1a2a1c1c (6.6%), D1a2a1c2 (5.7%), and D1a2a2 (32.3%; Figure 2A and Supplementary Table S5). Three men were included in haplogroup D1a2a (0.9%). Regional comparisons showed that among these 13 groups, the frequencies of D1a2a2 were lower in Nagasaki (14.7%), Fukuoka (23.5%), Kanazawa (33.3%), Kawasaki (29.3%), and Sapporo (32.1%) compared with those in Tokushima (45.5%) and Osaka (55.6%; Figure 2B and Supplementary Table S5). The F_{ST} values differed significantly ($P < 0.05$) between the Tokushima and Nagasaki ($F_{ST} = 0.037$), Osaka and Nagasaki ($F_{ST} = 0.081$), and Osaka and Fukuoka ($F_{ST} = 0.073$) regions (Table 2).

Next, haplogroups of 380 individuals belonging to the Y chromosome haplogroup, D1a2a1-12f2b (formerly D2a1), were analyzed. D1a2a1-12f2b was classified into 11 subgroups: D1a2a1a2b (12.9%), D1a2a1a2b1 (0.5%), D1a2a1a2b1a (1.6%), D1a2a1a2b1a1 (6.8%), D1a2a1a2b1a1a (21.1%), D1a2a1a2b1a1a1 (15.8%),

D1a2a1a2b1a1a1a (13.7%), D1a2a1a2b1a1a3 (7.6%), D1a2a1a2b1a1a9 (4.2%), D1a2a1a2b1a1b (2.4%), and D1a2a1a3 (13.4%; Figure 3A and Supplementary Table S5). Comparisons of the regional frequencies of these 11 groups revealed that the frequency of D1a2a1a2b1a1a1 was high in Fukuoka, whereas that of D1a2a1a2b1a1a1a was high in Osaka (Figure 3B). However, their F_{ST} values were not significantly different (Table 2).

Haplogroup O1b2

Haplogroups of 214 individuals belonging to the Y chromosome haplogroup, O1b2 (formerly O2b*), were analyzed. O1b2 was classified into nine subgroups: O1b2a (4.7%), 1b2a1a (4.7%), O1b2a1a2a (0.5%), O1b2a1a2a1 (24.3%), O1b2a1a2a1a (28.0%), O1b2a1a2a1b (0.5%), O1b2a1a2a1b1 (14.0%), O1b2a1a3 (16.4%), and O1b2a1b (0.5%; Figure 4A and Supplementary Table S6). Fourteen men were included in haplogroup O1b2 (6.5%). Regional comparisons showed that among these 10 groups, the frequencies of O1b2a1a2a1 were higher in Nagasaki (31.3%) and Fukuoka (37.5%) than in Kanazawa (Figure 4B). Haplogroup O1b2a1a2a1a was more prevalent in Osaka (46.7%) than in the other regions. The F_{ST} values differed significantly ($P < 0.05$) between Nagasaki and Kanazawa ($F_{ST} = 0.042$).

Next, a detailed haplogroup analysis was performed on 490 individuals

belonging to the Y chromosome haplogroup, O1b2a1a1 (formerly O2b1). O1b2a1a1 was classified into three subgroups: O1b2a1a1a (34.3%), O1b2a1a1b (23.9%), and O1b2a1a1c (10.4%; Figure 4A and Supplementary Table S6). Only 154 men were included in haplogroup O1b2a1a1 (31.4%). Notably, no significant differences were observed in the regional frequencies of these four haplogroups (Figure 4C). Their F_{ST} values also showed no significant differences (Table 2).

Discussion

People belonging to haplogroups D and C are suggested to have arrived in Japan 2,000 and 12,000 years ago, respectively [10], in accordance with the dual-structure model. Lineage C has been mainly classified into C1 and C2 subgroups, but the present study showed that it branches out further.

Haplogroup C is widespread in East Asia and Siberia and has been observed at low frequencies in Oceania, Europe, and the Americas. It is believed that the ancestors of modern humans left Africa and crossed the Eurasian continent via three routes: west, north, and south. Specific subhaplogroups were identified in each region. Haplogroup C1a1 is restricted to Japan and is thought to have diverged before entering Japan. Haplogroup C2 diverged into C2a, which is common in Central Asia, Northeast Asia, and North America, and C2b, which is common in China, Mongolia, and the Korean Peninsula; part of C2b is thought to have reached Japan [11-13]. The C1a1 lineage, which is endemic to Japan, and its sister group, the C1a2 lineage, are found at low frequencies in Europe [14]. In contrast, the early branch, C1b, is found along the southern coasts of India, Australia, and Indonesia [15,16]. The major sub-haplogroups of the C1 lineage in Japan are C1a1a1a, mainly found in Tokushima and Osaka, and C1a1a1b, mainly found in Nagasaki, Fukuoka, Kanazawa, and Sapporo. The F_{ST} values showed significant

differences between Tokushima–Nagasaki and Kanazawa and between Osaka–Nagasaki and Kanazawa regions. C1a1 is unique to Japan, and C1a1a, which did not diverge into C1a1a1a and C1a1a1b, was found at low frequencies in Sapporo and Nagasaki, suggesting that C1a1, which is evenly distributed in Japan, diverged into downstream haplogroups when males migrated mainly from Tokushima and Osaka. The major C2 lineage sub-haplogroups in Japan include C2a, C2b1a1, C2b1a2, and C2b1b, all of which show variations in their haplogroup frequencies in Osaka. The F_{ST} values showed marked differences between Osaka and the surrounding areas (Nagasaki, Fukuoka, Tokushima, and Kawasaki), suggesting that population divergence and migration centered around Osaka. Haplogroup frequency variation was observed only in Osaka. As Osaka is one of the largest cities in Japan, this finding suggests that modern gene flow occurs at a high frequency in this city.

Based on Y-Tree (ver. 11.04) from YFull (<https://www.yfull.com/tree/>), a database of Y chromosome sequences from around the world, the divergence (95% CI) of the single nucleotide polymorphism was estimated to have occurred 45,300 years before present (ybp) (49,400–41,300) for C1a1 and 48,800 ybp (51,300–46,400) for C2. Other sub-haplogroup divergence years were 4,500 ybp (5,500–3,600) for C1a1a1a and C1a1b, 34,100 ybp (37,000–31,300) for C2a, 10,300 ybp (11,200–9,400) for C2b1a1 and C2b1a2,

and 11,000 ybp (12,000–10,000) for C2b1b. Jomon lineage began to spread to Japan at least 20,000 years ago [10], suggesting that the divergence of downstream haplogroups occurred within Japan.

In addition to Japan, haplogroup D has been identified in the Tibetan Plateau, Andaman Islands, and specific regions of Africa, and D1a2a is the best-characterized haplogroup in Japan. Haplogroup D2 was only observed in some parts of Africa, such as Nigeria [17], whereas D1, which expanded into Eurasia, diverged into D1a1, which is common in Central Asia and Tibet, and D1a2, which is found in Japan [10]. Haplogroup D1a2b, the sister group of D1a2a, is the most common haplogroup in the Andaman Islands [18].

The present study confirmed that haplogroup D1a2a was classified into several downstream sub-haplogroups, including D1a2a1c1, D1a2a1c2, and D1a2a2. Haplogroup D1a2a2 was more frequently observed in Tokushima and Osaka than in the other areas, whereas the other haplogroups were evenly distributed among all regions. As the D1a2a lineage is a haplogroup unique to Japan, the divergence of this sub-lineage may have occurred within Japan, and it was uniformly distributed throughout the country. The uneven distribution of D1a2a2 may be attributed to Tokushima and Osaka being the starting points of population movement and divergence and/or the expansion of the

residential area owing to the rise of the O lineage, which later entered Japan.

D1a2a-12f2b lineage formed three clusters with several downstream haplogroups: D1a2a1a2b1a1a, D1a2a1a2b1a1b, and D1a2a1a3. No significant regional differences were observed among the haplogroups. Haplogroups D1a2a and D1a2a-12f2b may have simultaneously entered Japan; however, some groups were evenly distributed in Japan, whereas others showed a frequency bias.

According to the Y-tree, the estimated divergence (95% CI) of D1a2a1 and D1a2a2 was 21,200 ybp (23,100–15,000), whereas that of D1a2a1a and D1a2a1c was 17,600 ybp (20,300–15,000). Haplogroup D1a2a diverged by 45,200 ybp (48,500–42,000) and was only observed in Japan, indicating that the D1a2a lineage diverged extensively in Japan.

Haplogroup O is the largest haplogroup in East Asia that is suggested to have arrived in Japan approximately 4,000 years ago [10]. The ancestral lineage NO expanded into Eurasia via the northern route after leaving Africa and diverged into haplogroups O and N. Haplogroup O was broadly classified as O2, which flourished in the Yellow River Basin in northern China, and O1, which flourished in the Yangtze River Basin in southern China. As haplogroup O1b2, derived from O1, is a common lineage in Japan, China, Manchuria, and Korea [19], it may have migrated north to Japan via the Korean Peninsula.

A portion of haplogroup O2, comprising most present-day Han Chinese people, may have reached Japan after branching off as O2a1b and O2a2b1.

The O1b2 lineage formed three clusters, several downstream haplogroups under O1b2a1a2a1a, O1b2a1a2a1b, and O1b2a1a3. O1b2a1a1 lineage also formed three clusters: several downstream haplogroups under O1b2a1a1a, O1b2a1a1b, and O1b2a1a1c. No marked differences in haplogroup frequencies were detected by region, suggesting that the major downstream lineages of O1b2 in Japan are homogeneous among modern Japanese males, possibly due to the divergence of the O1b2a1a1 lineage via the Korean Peninsula and surrounding areas.

Years of divergence (95% CI) estimated by the Y-tree were 28,000 ybp (30,400–26,000) for O1b2 and 5,500 ybp (6,500–4,600) for O1b2a1a1. The year of divergence of the identified O1b2a1a1a, O1b2a1a1b, O1b2a1a1c, and O1b2a1a2a1a haplogroups was 3,400 ybp (4,500–2,500), considering that the Yayoi lineage began to spread into Japan at least 4,000 years ago [10]. These findings suggest that O1b2 diverged sub-linearly in Eurasia and spread uniformly after its influx into Japan.

In conclusion, the results of the present study enabled us to update the Y chromosome haplogroups C1a1, C2, D1a2, D1a2-12f2b, O1b2, and O1b2a1a1 lineages into 3, 6, 14, 11, 10, and 4 haplogroups, respectively. In addition, we confirmed the

existence of several clusters with concentrated frequencies for each haplogroup. Haplogroups C1a1, C2, and D1a2, which are unique to Japan, showed marked differences in their frequencies, suggesting that lineage divergence and population movement within Japan occurred mainly around Tokushima and Osaka. For haplogroup C and D lineages, we detected differences in the diversity of the Japanese male population that have not been previously reported [8]. The Japanese male is believed to exist within the larger framework of a dual structure but has retained genetic structural diversity owing to population shifts, genetic flotation, and the influx of different genes. DNA analysis has recently revealed that the Japanese males migrated from the Korean Peninsula and continental East Asia following the Kofun period (3rd century AD) and that the Okhotsk cultural people also migrated to Hokkaido from the north, leading to the three-step migration model theory [20,21]. Future studies should further elucidate the diversity of Japanese males by identifying the Japanese sub-lineages of haplogroup O3 and examining their frequencies in different regions.

Acknowledgments

We are grateful to Prof. Teruaki Iwamoto helping with sample collection. This study was supported by the Grants-in-Aid for Scientific Research on Innovative Areas "Yaponesian

Genome" (19H05346) from the Japan Society for the Promotion of Science. We thank

Editage for the English language editing.

Conflicts of interest

The authors declare that there are no conflicts of interest.

References

1. Rhie A, Nurk S, Cechova M, Hoyt SJ, Taylor DJ, Altemose N, et al. The complete sequence of a human Y chromosome. *Nature*. 2023. doi:10.1038/s41586-023-06457-y.
2. Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, et al. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature*. 2003;423:825-837. doi: 10.1038/nature01722.
3. Tilford CA, Kuroda-Kawaguchi T, Skaletsky H, Rozen S, Brown LG, Rosenberg M, et al. A physical map of the human Y chromosome. *Nature*. 2001;409:943-945. doi: 10.1038/35057170.
4. Nakahori Y, Tamura T, Yamada M, Nakagome Y. Two 47z [DXYS5] RFLPs on the X and the Y chromosome. *Nucleic Acids Res*. 1989;17:2152. doi: 10.1093/nar/17.5.2152.
5. Shinka T, Tomita K, Toda T, Kotliarova SE, Lee J, Kuroki Y, et al. Genetic variations on the Y chromosome in the Japanese population and implications for modern human Y chromosome lineage. *J Hum Genet*. 1999;44:240-245. doi: 10.1007/s100380050151.

6. Y Chromosome Consortium. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res.* 2002;12:339-348. doi: 10.1101/gr.217602.

7. Hanihara K. (1991) Dual structure model for the population history of the Japanese. *Japan Review*, 2: 1–33. <https://www.jstor.org/stable/25790895>

8. Sato Y, Shinka T, Ewis AA, Yamauchi A, Iwamoto T, Nakahori Y. Overview of genetic variation in the Y chromosome of modern Japanese males. *Anthropol.Sci.* 2014;122, 131-136. doi: <https://doi.org/10.1537/ase.140709>

9. Excoffier L, Lischer HE. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour.* 2010;10:564-567. doi: 10.1111/j.1755-0998.2010.02847.x.

10. Hammer MF, Karafet TM, Park H, Omoto K, Harihara S, Stoneking M, Horai S. Dual origins of the Japanese: common ground for hunter-gatherer and farmer Y chromosomes.

J Hum Genet. 2006;51:47-58. doi: 10.1007/s10038-005-0322-0.

11. Zegura SL, Karafet TM, Zhivotovsky LA, Hammer MF. High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y chromosomes into the Americas. Mol Biol Evol. 2004;21:164-175. doi: 10.1093/molbev/msh009.

12. Pakendorf B, Novgorodov IN, Osakovskij VL, Danilova AP, Protod'jakonov AP, Stoneking M. Investigating the effects of prehistoric migrations in Siberia: genetic variation and the origins of Yakuts. Hum Genet. 2006;120:334-53. doi: 10.1007/s00439-006-0213-2.

13. Zhabagin M, Sabitov Z, Tarlykov P, Tazhigulova I, Junissova Z, Yerezhepov D, et al. The medieval Mongolian roots of Y-chromosomal lineages from South Kazakhstan. BMC Genet. 2020;21(Suppl 1):87. doi: 10.1186/s12863-020-00897-5.

14. Scozzari R, Massaia A, D'Atanasio E, Myres NM, Perego UA, Trombetta B, et al. Molecular dissection of the basal clades in the human Y chromosome phylogenetic tree. *PLoS One*. 2012;7:e49170. doi: 10.1371/journal.pone.0049170.
15. Sengupta S, Zhivotovsky LA, King R, Mehdi SQ, Edmonds CA, Chow CE, et al. Polarity and temporality of high-resolution y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists. *Am J Hum Genet*. 2006;78:202-221. doi: 10.1086/499411.
16. Gayden T, Cadenas AM, Regueiro M, Singh NB, Zhivotovsky LA, Underhill PA, et al. The Himalayas as a directional barrier to gene flow. *Am J Hum Genet*. 2007;80:884-894. doi: 10.1086/516757.
17. Haber M, Jones AL, Connell BA, Asan, Arciero E, Yang H, et al. A Rare Deep-Rooting D0 African Y-Chromosomal Haplogroup and Its Implications for the Expansion of Modern Humans Out of Africa. *Genetics*. 2019;212:1421-1428. doi: 10.1534/genetics.119.302368.

18. Thangaraj K, Singh L, Reddy AG, Rao VR, Sehgal SC, Underhill PA, et al. Genetic affinities of the Andaman Islanders, a vanishing human population. *Curr Biol.* 2003;13:86-93. doi: 10.1016/s0960-9822(02)01336-2.

19. Kim SH, Kim KC, Shin DJ, Jin HJ, Kwak KD, Han MS, et al. High frequencies of Y-chromosome haplogroup O2b-SRY465 lineages in Korea: a genetic perspective on the peopling of Korea. *Investig Genet.* 2011;2:10. doi: 10.1186/2041-2223-2-10.

20. Japanese Archipelago Human Population Genetics Consortium; Jinam T, Nishida N, Hirai M, Kawamura S, Oota H, Umetsu K, et al. 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. *J Hum Genet.* 2012;57:787-795. doi: 10.1038/jhg.2012.114.

21. Jinam TA, Kanzawa-Kiriyama H, Saitou N. Human genetic diversity in the Japanese Archipelago: dual structure and beyond. *Genes Genet Syst.* 2015;90:147-152. doi: 10.1266/ggs.90.147.

Figure legends

Figure 1. Phylogenetic tree and frequency of Y chromosome haplogroup C lineages (A).

Regional frequencies of Y chromosome haplogroup C1a1 (B) and C2 (C) lineages.

Figure 2. Phylogenetic tree and frequency of Y chromosome haplogroup D1a2a lineages

(A) and frequency by region (B).

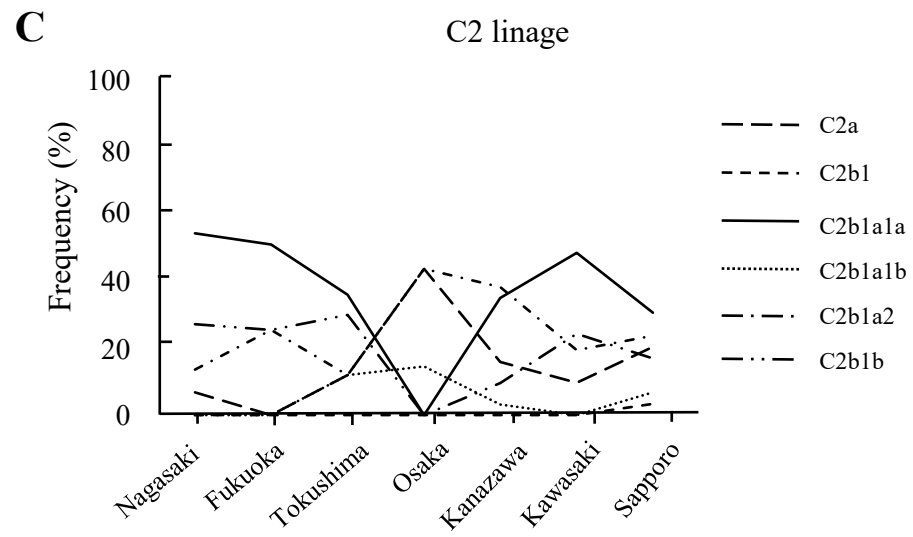
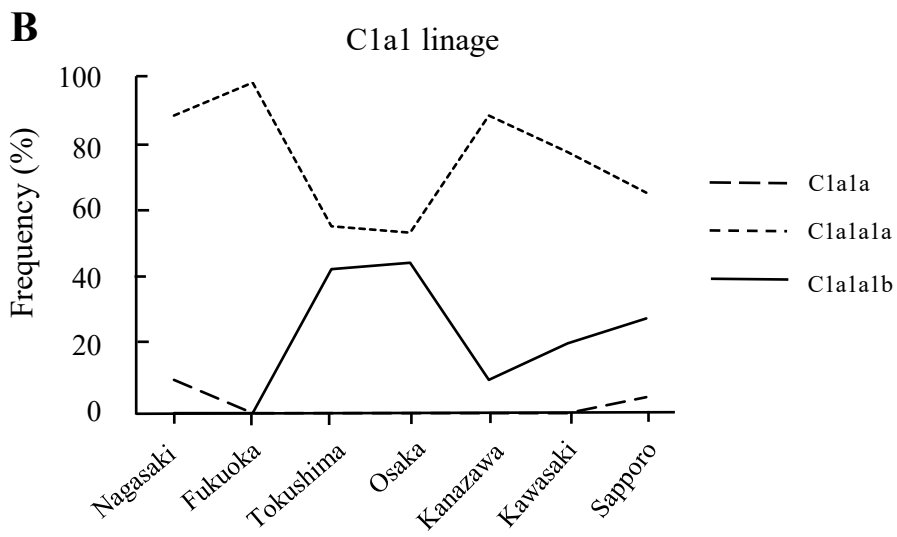
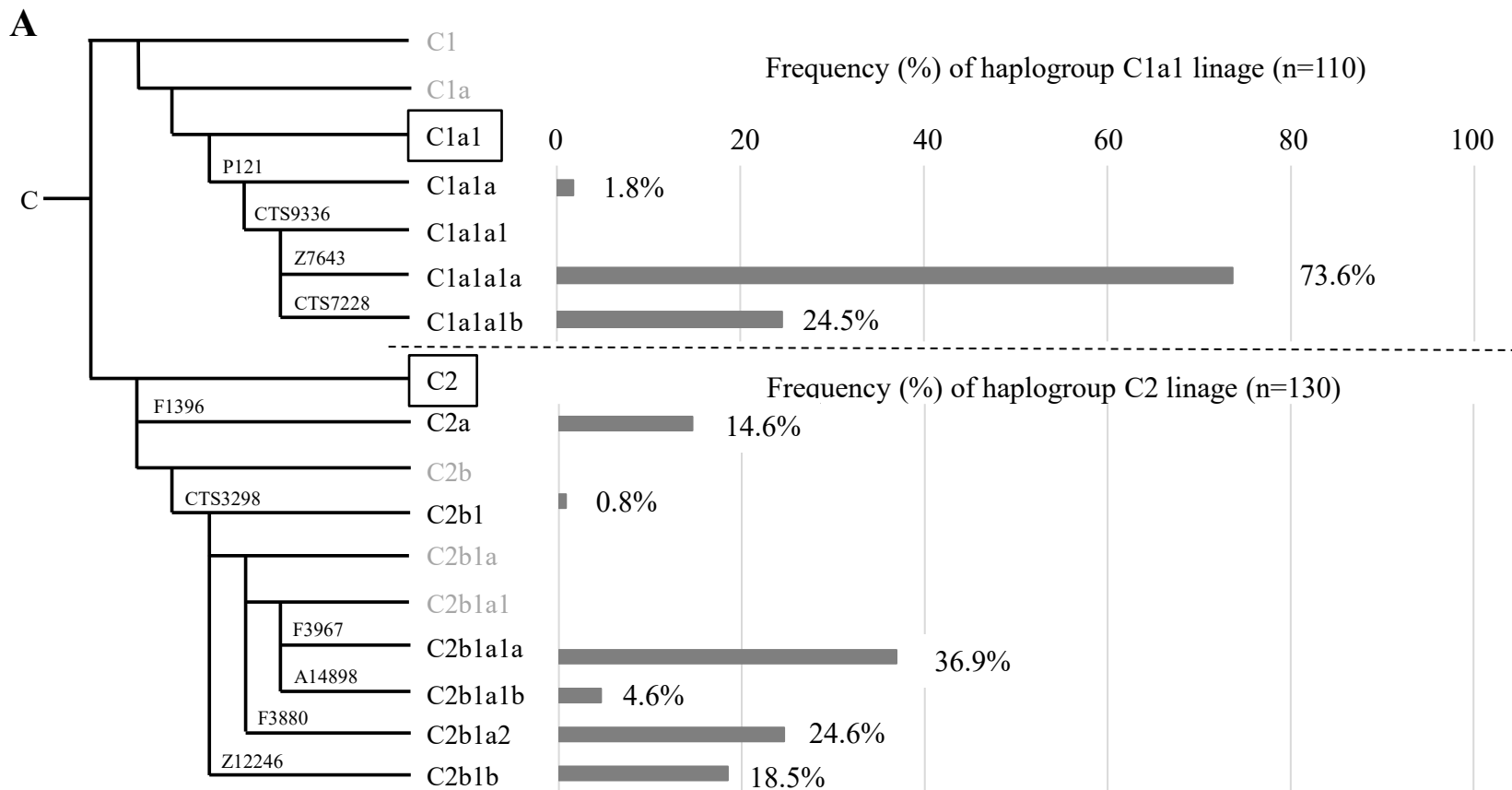
Figure 3. Phylogenetic tree and frequency of Y chromosome haplogroup D1a2a-12f2b

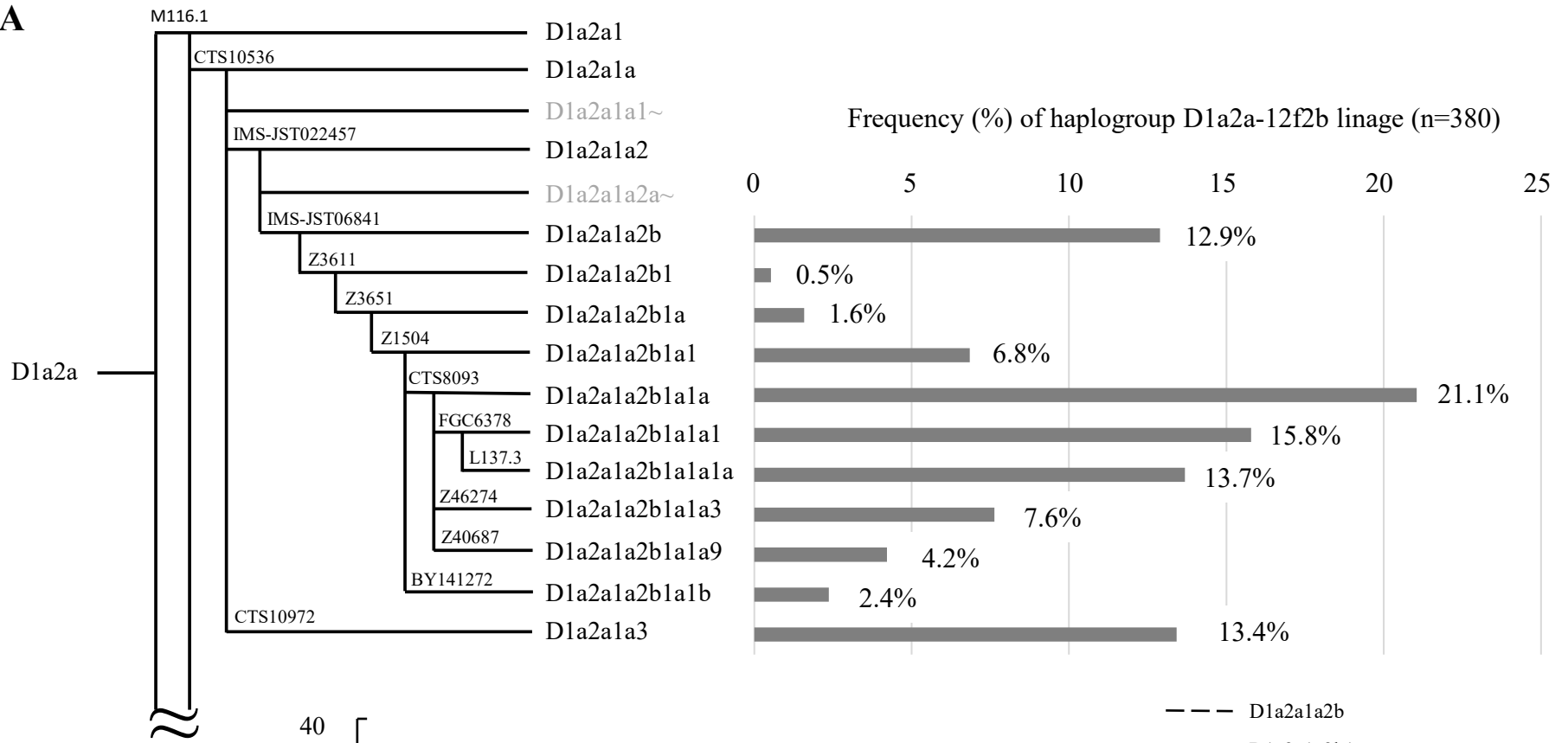
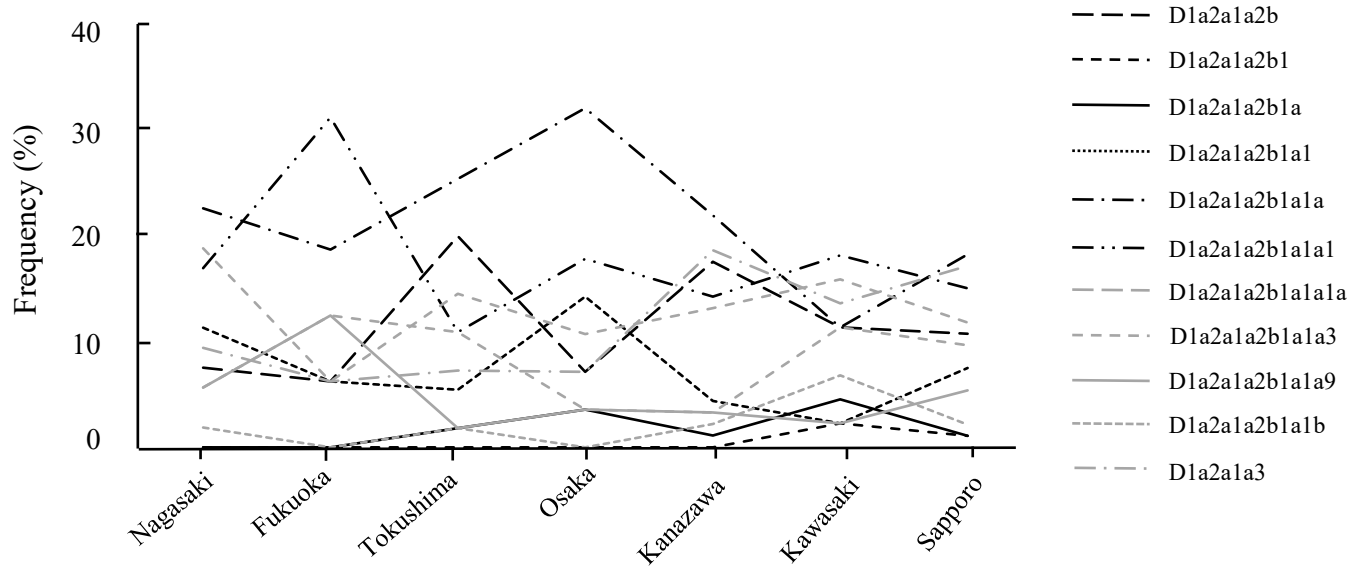
lineages (A) and frequency by region (B).

Figure 4. Phylogenetic tree and frequency of Y chromosome haplogroup O1b2 lineages

(A). Regional frequencies of Y chromosome haplogroups O1b2 (B), and O1b2a1a1 (C)

lineages.



A**B**

A

Frequency (%) of haplogroup O1b2 (n=214) and O1b2a1a1 (n=490) lineages

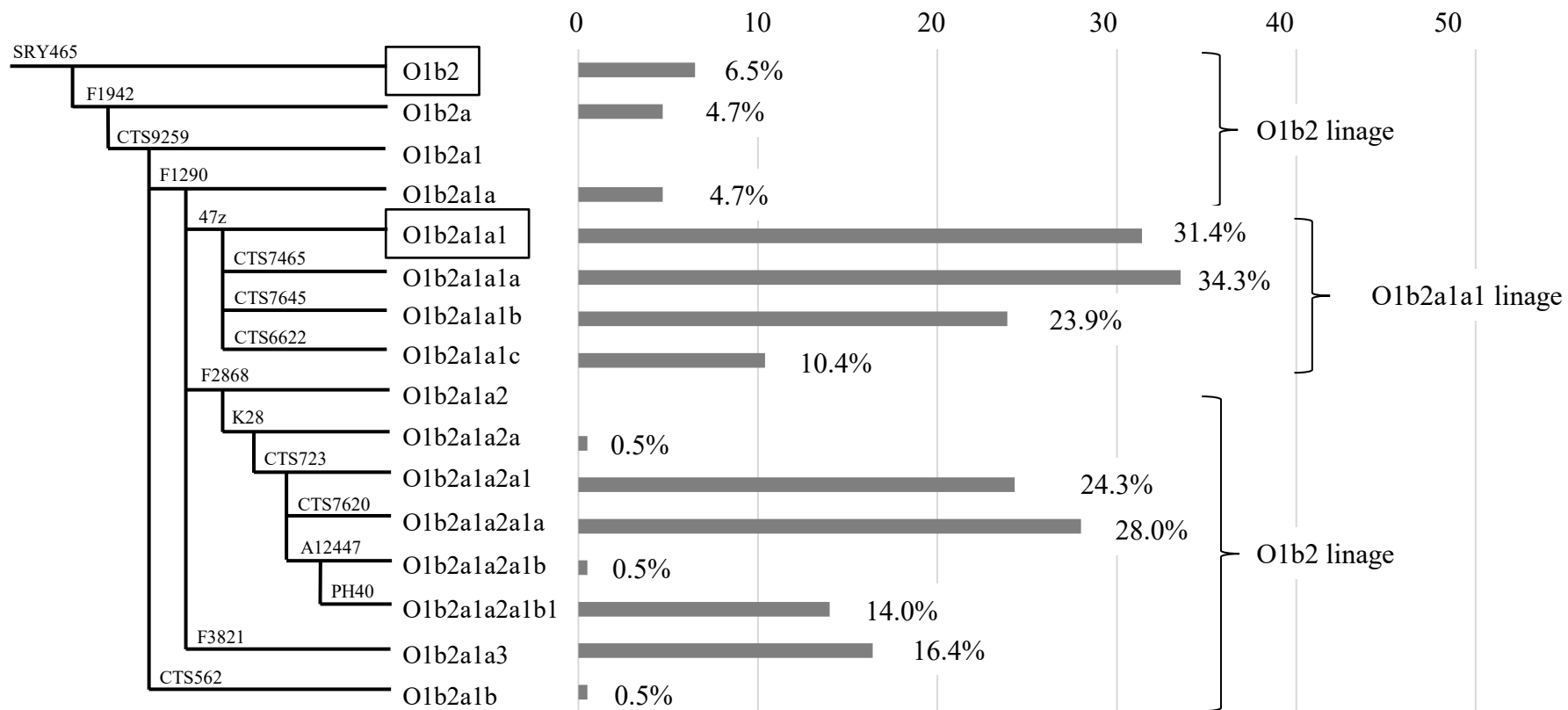
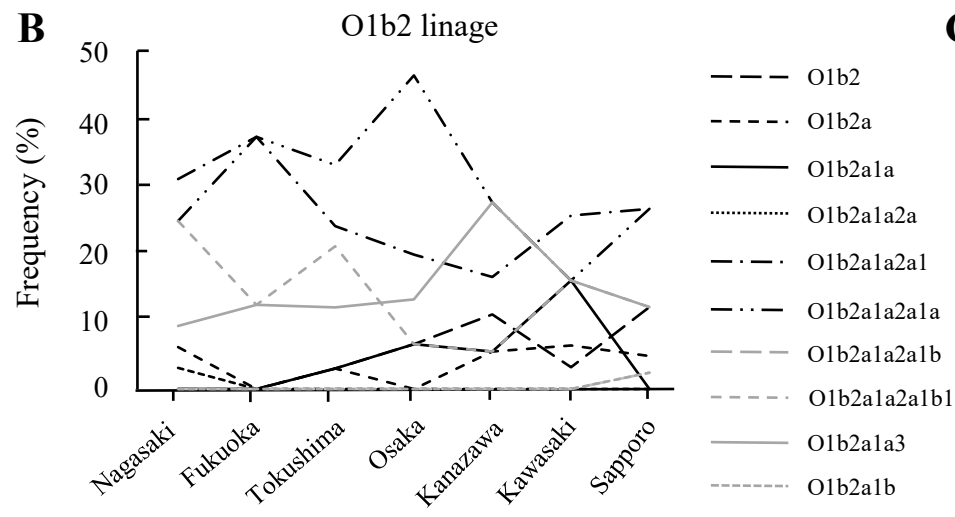
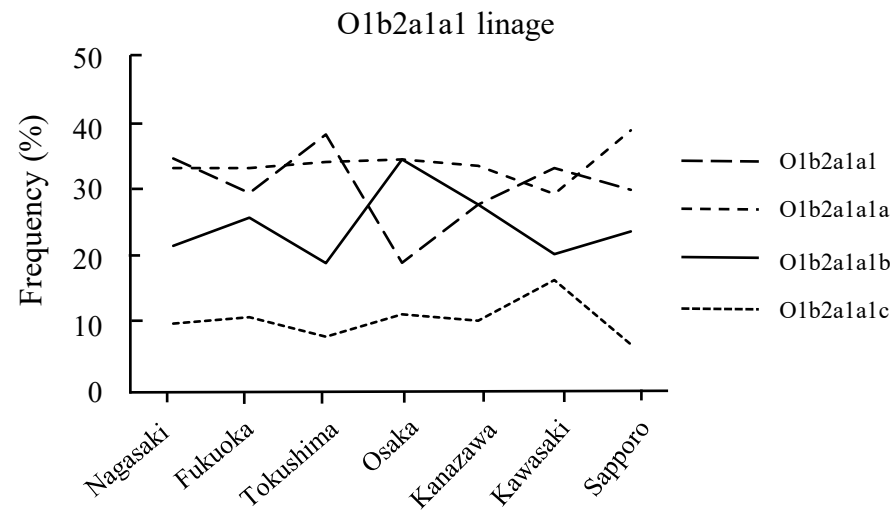
**B****C**

Table 1. Number of Y chromosome haplogroups per city used in the analysis.

	Nagasaki	Fukuoka	Tokushima	Osaka	Kanazawa	Kawasaki	Sapporo	All
C1a1 (C1)	10	6	23	11	20	19	21	110
C2 (C3)	15	8	17	7	32	21	30	130
D1a2a (D2*)	34	17	33	18	78	58	78	316
D1a2a-12f2b (D2a1)	53	16	55	28	91	44	93	380
O1b2 (O2b*)	32	8	33	15	54	31	41	214
O1b2a1a1 (O2b1)	69	27	73	26	104	78	113	490
Total	213	82	234	105	379	251	376	1640

Parentheses indicate the former name of the haplogroup according to the Y Chromosome Consortium nomenclature.

Table 2. Pairwise F_{ST} values (below diagonal line) and P values (above diagonal line) or comparison of the haplogroup C1a1, C2, D1a2a, D1a2a-12f2b, O1b2, and O1b2a1a1 lineage frequencies in Japanese male population from the seven cities.

C1a1	Nagasaki	Fukuoka	Tokushima	Osaka	Kanazawa	Kawasaki	Sapporo
Nagasaki	—	0.991	0.036*	0.036*	0.720	0.270	0.144
Fukuoka	-0.059	—	0.126	0.072	0.991	0.586	0.252
Tokushima	0.229*	0.257	—	0.991	0.009*	0.216	0.468
Osaka	0.261*	0.305	-0.071	—	0.027*	0.243	0.514
Kanazawa	-0.024	-0.043	0.207*	0.249*	—	0.342	0.180
Kawasaki	0.033	0.049	0.062	0.065	-0.006	—	0.514
Sapporo	0.089	0.122	-0.011	-0.026	0.077	-0.024	—
C2	Nagasaki	Fukuoka	Tokushima	Osaka	Kanazawa	Kawasaki	Sapporo
Nagasaki	—	0.991	0.676	0.009*	0.090	0.991	0.288
Fukuoka	-0.089	—	0.757	0.027*	0.496	0.991	0.604
Tokushima	-0.030	-0.050	—	0.036*	0.171	0.730	0.694
Osaka	0.249*	0.200*	0.139*	—	0.180	0.000*	0.207
Kanazawa	0.043	-0.017	0.031	0.058	—	0.234	0.577
Kawasaki	-0.054	-0.083	-0.029	0.192*	0.014	—	0.477
Sapporo	0.014	-0.021	-0.020	0.045	-0.012	-0.005	—
D1a2a	Nagasaki	Fukuoka	Tokushima	Osaka	Kanazawa	Kawasaki	Sapporo
Nagasaki	—	0.297	0.045*	0.009*	0.117	0.171	0.144
Fukuoka	0.006	—	0.081	0.027*	0.252	0.306	0.333
Tokushima	0.037*	0.04	—	0.865	0.559	0.198	0.468
Osaka	0.081*	0.073*	-0.024	—	0.180	0.162	0.162
Kanazawa	0.011	0.006	-0.006	0.012	—	0.703	0.838
Kawasaki	0.016	0.004	0.007	0.022	-0.006	—	0.703
Sapporo	0.015	0.005	-0.001	0.014	-0.007	-0.005	—
D1a2a-12f2b 1	Nagasaki	Fukuoka	Tokushima	Osaka	Kanazawa	Kawasaki	Sapporo
Nagasaki	—	0.595	0.496	0.892	0.351	0.468	0.613
Fukuoka	-0.011	—	0.297	0.748	0.216	0.613	0.514
Tokushima	-0.001	0.008	—	0.496	0.432	0.297	0.405
Osaka	-0.016	-0.013	-0.002	—	0.342	0.171	0.432
Kanazawa	0.002	0.010	-0.002	0.004	—	0.306	0.676
Kawasaki	-0.001	-0.01	0.003	0.015	0.001	—	0.802
Sapporo	-0.005	-0.008	0.001	0.000	-0.004	-0.008	—

Table 2. Continued.

O1b2	Nagasaki	Fukuoka	Tokushima	Osaka	Kanazawa	Kawasaki	Sapporo
Nagasaki	—	0.856	0.847	0.252	0.009*	0.432	0.550
Fukuoka	-0.055	—	0.937	0.856	0.414	0.604	0.946
Tokushima	-0.020	-0.063	—	0.730	0.126	0.324	0.712
Osaka	0.020	-0.067	-0.020	—	0.405	0.153	0.505
Kanazawa	0.042*	-0.010	0.016	-0.003	—	0.234	0.225
Kawasaki	-0.000	-0.024	0.001	0.023	0.013	—	0.396
Sapporo	-0.005	-0.052	-0.013	-0.010	0.005	0.002	—
O1b2a1a1	Nagasaki	Fukuoka	Tokushima	Osaka	Kanazawa	Kawasaki	Sapporo
Nagasaki	—	0.937	0.946	0.288	0.676	0.865	0.739
Fukuoka	-0.023	—	0.694	0.802	0.990	0.820	0.856
Tokushima	-0.013	-0.016	—	0.207	0.297	0.550	0.531
Osaka	0.002	-0.026	0.017	—	0.874	0.2973	0.486
Kanazawa	-0.006	-0.023	0.002	-0.016	—	0.414	0.757
Kawasaki	-0.010	-0.019	-0.005	0.005	-0.002	—	0.243
Sapporo	-0.007	-0.020	-0.003	-0.005	-0.005	0.003	—

* $P < 0.05$