Comprehensive analysis of Japanese population history by detecting ancestry-marker polymorphisms without using ancestral genomic information

Yusuke Watanabe[1,2], Jun Ohashi[1]

[1]Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Tokyo 113-0033, Japan

[2]Genome Medical Science Project Toyama Project, National Center for Global Health and Medicine, Tokyo 162-8655, Japan.

Correspondence:

Jun Ohashi, Ph.D.

Phone: +81-3-5841-8395 (ext. 28395)

Abstract

Modern Japanese are considered to derive from a mixture of two major ancestral populations: the indigenous Jomon people and immigrants from continental East Asia. Since most of the existing methods for detecting genetic components from ancestral populations require their genomes, ancestral genomic components in Japanese could not detected so far due to the lack of precisely sequenced ancient Jomon genomes. To overcome the difficulty, we developed a reference-free detection method using a novel summary statistic, the ancestry-marker index (*AMI*). We applied the *AMI* to modern Japanese samples from the 1000 Genomes Project and identified 208,648 ancestry-marker SNPs that were likely derived from the Jomon people (Jomon-derived SNPs). Comparing the Jomon allele score detected in this study with modern Japanese and two ancient Jomon individuals showed that the Jomon derived SNPs were detected with high accuracy by the *AMI* in real data, and that the Jomon derived SNPs were detected by several tens of times from a single Jomon individual by the *AMI*. The analysis of Jomon-derived SNPs in 10,842 modern Japanese individuals recruited from all the 47 prefectures of Japan showed that the genetic differences among the prefectures were mainly caused by differences in the admixture proportion of the Jomon people, due to the difference of population size of immigrants in the final Jomon to the Yayoi period. We also confirmed the presence of the Jomon alleles around phenotype associated SNPs characteristic of East Asians to clarify whether these phenotypes of modern Japanese were derived from the Jomon people.

Introduction

Modern Japanese populations are divided into three main populations: the Ainu, who live mainly in Hokkaido; the Ryukyuan, who live mainly in Okinawa; and mainland Japanese, who live in Honshu, Shikoku, and Kyushu (Fig. 1). As a powerful hypothesis of the formation processes of Japanese populations, a dual structure model(Hanihara 1991) was proposed based on morphology. This model assumes that Japanese originated through a mixture of the Jomon people, who settled in the Japanese archipelago during the Jomon period (from 16,500 YBP to 2,800 YBP)(Habu 2004; Mizoguchi 2013; Fujio 2015), and the immigrants came to the Japanese archipelago from continental East Asia around the beginning of the Yayoi period (around 2,800 YBP)(Fujio 2015). According to this model, compared to mainland Japanese, the Ainu and the Ryukyuan were genetically less influenced by immigrants. Findings from genetical studies support the dual structure model(Jinam et al. 2012; Jinam et al. 2015; Nakagome et al. 2015; Kanzawa-Kiriyama et al. 2017; Kanzawa-Kiriyama et al. 2019). Whole-genome analyses extracted from the remains of the Jomon people have revealed that the Jomon were highly differentiated from other East Asians, being genetically closely related to the Ainu/Ryukyuan, and that 10–20% of genomic components found in mainland Japanese are derived from the Jomon people(Kanzawa-Kiriyama et al. 2017; Kanzawa-Kiriyama et al. 2019). However, it is still unknown whether there are regional differences in the admixture degree between the Jomon people and immigrants in mainland Japanese and how they admixed in the Japanese Archipelago. To solve these problems, it is necessary

to comprehensively detect genomic components or single nucleotide polymorphisms (SNPs) in mainland Japanese that were derived from Jomon people and analyze them in detail at the prefectural level.

In populations derived from a mixture of two source populations, recombination between haplotypes from different source populations inevitably occurs after the admixture event. As a result, haplotypes from two ancestral populations are patchily present in the chromosomes of admixed populations, and the alleles in the haplotypes from each ancestral population are in linkage disequilibrium (LD) with each other (Supplementary Fig. 1). Most existing methods(Price et al. 2009; Baran et al. 2012; Churchhouse and Marchini 2013; Guan 2014; Hui et al. 2017) for estimating the local ancestry of genomes using the LD state require genome information of ancestral populations or of modern populations as proxies of ancestral populations. In the Japanese population, only a few Jomon individuals(Kanzawa-Kiriyama et al. 2017; Mccoll et al. 2018; Kanzawa-Kiriyama et al. 2019; Gakuhari et al. 2020) have been sequenced so far. Furthermore, the sequence depths of these samples are low, except for the Funadomari Jomon(Kanzawa-Kiriyama et al. 2019) people excavated from Rebun Island in Hokkaido, making it difficult to obtain sufficient genome information to accurately estimate the local ancestry of modern Japanese (i.e., the genomes of many Jomon individuals are required to perform highly accurate studies). Therefore, at present, it is not possible to estimate the local ancestry of modern mainland Japanese using previous methods. In this study, we developed a method using a new summary statistic, the Ancestry-Marker Index (*AMI*), to

88 detect ancestry-marker SNPs derived from the Jomon people in modern
89 mainland Japanese that does not require genomes obtained from Jomon skeletal
90 specimens. Since the Jomon people have been found to be highly differentiated
91 from other East Asians(Kanzawa-Kiriyama et al. 2017; Kanzawa-Kiriyama et al.
92 2019), they are expected to have had specific variants that were not found in
93 other East Asians. Thus, the modern mainland Japanese also likely have specific
94 variants derived from the Jomon people. The *AMI* detects the Jomon-derived
95 variants based on the LD between Japanese specific variants. Based on the *AMI*,
96 we could successfully extract the Jomon-derived variants from real genomic data
97 of the Japanese. Using Jomon-derived variants, we clarified regional differences
98 among modern Japanese and estimated the population frequencies of alleles
99 associated with common phenotypes, which are characteristic of East Asians, in
100 the Jomon people. Based on these results, we propose a model of admixture
101 between the Jomon people and immigrants from continental East Asia in the
102 Japanese Archipelago.
103
104 **Results**
105
106 **Performance of *AMI***
107
108 To confirm the usefulness of the *AMI*, we performed a coalescent
109 simulation assuming a mixture of the Jomon people and continental East Asians
110 (Supplementary Fig. 2). The preliminary 10 Mb simulation suggested that modern

111 Japanese still have haplotypes of several megabases that are derived from the
112 Jomon people (see Supplementary Fig. 3 and Supplementary Texts).
113
114 There are three types of Japanese specific variants: (type 1) Jomon-
115 derived variants; (type 2) variants derived from continental East Asians; and (type
116 3) novel variants (see Supplementary Fig. 4 and Methods). Our aim is to
117 distinguish Jomon-derived variants (type 1) from (type 2) and (type 3). In the
118 subsequent 1 Mb simulation, Japanese specific variants (types 1, 2, and 3) were
119 extracted from each genealogy. The distributions of *AMI* for Jomon-derived
120 variants (type 1) and other variant types (types 2 and 3) are shown in Fig. 2 (a).
121 It was found that Jomon-derived variants (type 1) had larger *AMI* values than the
122 other Japanese specific variants (types 2 and 3). The receiver operating
123 characteristic (ROC) analysis showed that Jomon-derived variants (type 1) could
124 be distinguished from the other Japanese specific variants (types 2 and 3) by the
125 *AMI* (area under the curve [AUC] = 0.91; Fig. 2 (b)). The Youden index, a
126 measure of the cutoff value, was 28.0374. We performed further simulations,
127 varying the split time between the Jomon people and continental East Asians or
128 the effective population size in simulation, to confirm robustness of *AMI* to
129 different population history. Though the value of the Youden index varied
130 depending on the population history assumed, the Jomon-derived variants could
131 be accurately detected (Supplementary Fig. 5). According to Fig. 2, we set the
132 threshold for detecting Jomon-derived variants at $AMI > 28.0374$.
133

We also attempted to detect Jomon SNPs by S*(Plagnol and Wall 2006; Vernot and Akey 2014), a reference-free method previously proposed, and found that S* was unable to detect Jomon SNPs, perhaps due to an insufficient number of Jomon-derived specific variants of mainland Japanese (Supplementary Fig. 6 and Supplementary Texts).

Detection of Jomon variants in real data

Using the data set of 87 KPGP Koreans(Zhang et al. 2014) and 26 global populations of 1KG(Auton et al. 2015), approximately 1.7 million SNPs were found to be specific to mainland Japanese (1KG JPT). Of these 1.7 million SNPs, 208,648 SNPs exceeding the threshold of *AMI* were regarded as Jomon-derived SNPs. Jomon-derived SNPs were distributed throughout the autosomal genome (Supplementary Fig. 7). Hereafter, at the Jomon-derived SNPs, an allele only found in the mainland Japanese population is called a “Jomon allele.”

To examine the detection accuracy of Jomon-derived SNPs, we calculated the Jomon allele score (*JAS*, see Methods section for details) for the Ikawazu(Mccoll et al. 2018; Gakuhari et al. 2020)/Funadomari(Kanzawa-Kiriyama et al. 2019) Jomon individuals and mainland Japanese. If Jomon-derived SNPs were properly detected by the *AMI*, the *JAS* of the Ikawazu or Funadomari Jomon were expected to be higher than those of mainland Japanese. Of the JPT mainland Japanese, NA18976 was genetically close to continental East Asians in PCA (Supplementary Fig. 8) and was expected to have a lower

158 *JAS*. The distribution of the *JAS* is shown in Supplementary Fig. 9. The mean
159 *JAS* of 103 mainland Japanese individuals, excluding NA18976, was 0.0164. As
160 expected, NA18976 had the lowest *JAS*, 0.00269, which was much lower than
161 that of the other mainland Japanese. The *JAS* in the Ikawazu/Funadomari Jomon
162 were 0.0523 and 0.0555, respectively, indicating that the Jomon alleles were
163 found more frequently in Jomon people than in the modern mainland Japanese.
164 These results suggest that the *AMI* could detect SNPs derived from the Jomon
165 people. In addition, the *JAS* values were only a few % for both Jomon individuals,
166 which suggests that the number of Jomon-specific variants obtained from *AMI*
167 analyses of modern Japanese were several tens of times greater than that
168 obtained from the whole genome sequence of a single Jomon individual.
169
170 **Detection of regional genetic differences in mainland Japanese by Jomon-**
171 **derived SNPs**
172
173 ***JAS* by region and prefecture**
174 Previous prefecture-scale population studies showed that the Tohoku, Kanto,
175 and Kyushu populations (Fig. 1) are genetically more closely related to the
176 Ryukyuans, while the Kinki and Shikoku populations are more closely related to
177 continental East Asians(Yamaguchi-Kabata et al. 2008; Watanabe et al. 2020).
178 Based on these facts, we hypothesized that the genetic regional differences
179 among the modern mainland Japanese are caused by regional geographical
180 differences in the admixture proportion of the Jomon and immigrants from
181 continental East Asia. To verify this, we calculated the average *JAS* for each

geographic region and prefecture from imputed genotypes of 3,917 Jomon-derived SNPs of 10,842 Japanese individuals previously used for regional population genetic analysis(Watanabe et al. 2020). We removed the Hokkaido samples, which were largely affected by the immigration of Japanese after the Meiji period, for subsequent analysis and a total of 10,412 samples were used. The samples of each prefecture except for Hokkaido were divided into ten regions: Tohoku, Kanto, Hokuriku, Chubu, Tokai, Kinki, Chugoku, Shikoku, Kyushu, and Okinawa in accordance with a previous study(Koyama 1979) (Fig. 1 and Supplementary Table 1). The *JAS*s in these ten geographical regions are presented in Fig. 3 (a) and Supplementary Table 2. We found that the *JAS* was the highest in Okinawa (0.0255), followed by Tohoku (0.0189) and Kanto (0.018), and the lowest in Kinki (0.0163), followed by Shikoku (0.016). In prefecture scale, the average *JAS* in mainland Japan tended to be higher in prefectures located in the northernmost and southernmost parts of mainland Japan (Fig. 3 (b) and Supplementary Table 3). The *JAS* was especially high in Aomori (0.0192), Iwate (0.0195), Fukushima (0.0187), and Akita (0.0186) prefectures of the Tohoku region, as well as Kagoshima Prefecture (0.0186) in Kyushu. Japanese individuals in these prefectures are considered to possess more Jomon-derived genomic components than those in other prefectures. Prefectures with lower *JAS*s were in the Kinki and Shikoku regions, including Wakayama (0.0157), Nara (0.0189), Kochi (0.016), Tokushima (0.0161), and Mie (0.0161). These populations are considered to have more genomic components derived from continental East Asians. The *JAS* of each prefecture and the principal component 1 (PC1) value, which was obtained from the principal component analysis (PCA)

206 of a previous study by the allele frequency of autosomal 183,708 SNPs in each
207 prefecture(Watanabe et al. 2020) are plotted in Fig. 3 (c). The *JAS* was strongly
208 correlated with PC1 (R = 0.91, two-sided t-test $P = 2.2 \times 10^{-16}$). The geographic
209 distribution was not changed by tighter cutoff values (AMI > 100) for the detection
210 of Jomon-derived SNPs by *AMI* (Supplementary Figure 10 and Supplementary
211 Texts).
212
213 To confirm the results obtained from *JAS*, *f3*(each prefecture; Jomon,
214 CHB)(Patterson et al. 2012) were also calculated (Supplementary Fig. 11). The
215 distribution of *f3*, shown in Supplementary Fig. 11 (a) and (b) is similar to the
216 distribution of the *JAS*s (Fig. 3 (b)), with *f3* being generally small in prefectures
217 with high *JAS*s, such as Kagoshima Prefecture in Kyushu and Aomori and Iwate
218 Prefectures in Tohoku, and was large in the prefectures of regions with low *JAS*s,
219 such as Kinki and Shikoku. Based on these two findings: a correlation between
220 the *JAS* and the PC1 of Fig. 3 (c) and the concordance between the geographical
221 distribution of *f3* and *JAS*, it is strongly suggested that the genetic regional
222 differences of modern Japanese can be explained mainly by regional
223 geographical differences in the admixture proportions of the Jomon people. It
224 should be emphasized that the admixture, although to varying degrees, widely
225 occurred throughout the Japanese archipelago. Notably, prefectures in the
226 Tohoku region showed higher *JAS*s than those in the Kyushu region. However,
227 *f3* values were lower in Kyushu than in Tohoku. Since the Jomon-derived SNPs
228 were detected in 1KG JPT (Japanese living in Tokyo), the specific variants
229 possessed by the Jomon people of the Kyushu region may not have been

230 detected, and thus the *JAS* in Kyushu may have been underestimated. In other
231 words, these results could reflect differences in the genetic background of Jomon
232 people in Tohoku and Kyushu. Overall, the geographical gradient of *f3* in
233 mainland Japan was more consistent with the *JAS* than the distances from the
234 locations (Funadomari and Ikawazu) where Jomon samples were taken.
235
236 We assumed that the regional differences in the *JAS* were related to regional
237 differences in population size during the Jomon period. Therefore, we examined
238 the correlation between *JAS* and three indexes related to the Jomon population
239 size. The *JAS* of each prefecture was significantly correlated with the number of
240 archeological sites from the Jomon period ($R = 0.69$, two-sided *t*-test $P = 1.27\times10^{-7}$; Fig. 4 (a)). The *JAS* of each region also correlated with the population size
242 estimated from the number of archeological sites in the Late Jomon period ($R = 0.7$, two-sided *t*-test $P = 3.6\times10^{-2}$; Fig. 4 (b)). Moreover, the *JAS* of each
244 prefecture was strongly correlated with $\log_{10}$(number of archeological sites in the
245 Yayoi period/number of archeological sites in the Late Jomon period) ($R = -0.64$,
246 two-sided *t*-test $P = 2.08\times10^{-6}$; Fig. 4 (c)). It is considered that Figs. 4 (a) and (b)
247 correspond to Jomon population size in the Jomon period and Fig. 4 (c)
248 corresponds to the population growth rate occurring from the Late Jomon period
249 to the Yayoi period. The correlation between *JAS* and population size in each
250 region suggests that the smaller the population size in the Jomon period, the
251 lower *JAS* in modern mainland Japan (i.e., the higher contribution of genomic
252 components of immigrants from continental East Asia).
253

To summarize the above results, we can conclude that genetic differences among the regions of the modern Japanese population were mainly caused by differences in the admixture proportion of the Jomon people and that differences in the admixture proportion were caused by differences in the population sizes in each region during the Final Jomon period. Regarding these, previous morphological analyses showed that, of several populations in Japan, the Hokkaido Ainu and contemporary Kinki populations had contrasting cranial morphologies, while other modern regional populations were intermediate, with the Tohoku population being relatively similar in morphology to the Hokkaido Ainu(Hanihara 1985; Hanihara 1991). Archeological evidence suggests that immigrants from continental East Asia first reached northern Kyushu(Habu 2004), which seems contradictory considering that the *JAS* was lower in the Kinki and Shikoku regions than in northern Kyushu. The reason for this could be that the Kinki and Shikoku regions had a smaller population size during the Final Jomon period (Fig. 4), and thus, the proportion of genomic components derived from the immigrants became larger than in the other regions. In this study, we could clearly evaluate the similarity of local populations to the Jomon people using the Jomon-derived SNPs, and could clarify the main cause of genetic differences among the regional populations in mainland Japan.

PCA of prefectures according to Jomon allele frequency

A PCA was conducted for 46 Japanese prefectures using Jomon allele frequencies (Supplementary Fig. 12 (a)). This PCA demonstrated that Okinawa

Prefecture was separated from the other prefectures by PC1. Next, 45 prefectures in mainland Japan (Okinawa was excluded from further analyses based on Supplementary Fig. 12 (a)) were analyzed in PCA (Supplementary Fig. 12 (b)). The PC1 showed that prefectures in the Tohoku and Kanto regions, where higher *JAS*s were observed among Japanese prefectures, were greatly differentiated from prefectures in Kinki and Shikoku, where lower *JAS*s were observed (Fig. 3). The *JAS* was strongly correlated with the PC1 of Supplementary Fig. 12 (b) (R = -0.94, $P < 2.2\times10^{-16}$; Supplementary Fig. 13), which shows that PC1 reflects the ancestry proportion of Jomon people in each prefecture. In contrast, the PC2 of Supplementary Fig. 12 (b) was strongly correlated with both the latitude and longitude of each prefecture (latitude: R = 0.78, two-sided t-test $P = 2.30\times10^{-10}$, longitude: R = 0.66, two-sided t-test $P = 9.31\times10^{-7}$; Supplementary Fig. 14 (a) and (b)). These results indicate that the PC2 is determined by the geographical location of each prefecture, which might reflect that the genetic background of the Jomon people may differ according to the geographical locations in the Japanese archipelago. Previous studies have shown regional differences in the skeletal morphology of the Jomon people in the Japanese archipelago(Kaifu 1995; Maeda 2002; Takigawa 2006; Fukase et al. 2012; Kondo et al. 2017). For example, Kondo et al. suggested that Jomon craniofacial morphology, especially in the neurocranium, exhibit a northeast-to-southwest geographical cline across the Japanese archipelago(Kondo et al. 2017). To the best of our knowledge, including studies with ancient Jomon genomes, this is the first genome-wide study to refer to the genetic regional differences among the Jomon people in the Japanese archipelago.

Haplotype structures composed of Jomon-derived SNPs in genes associated with characteristic phenotypes of East Asians

To estimate the phenotype frequencies in the Jomon people, we investigated the haplotype structures of four genes (*ABCC11*, *EDAR*, *ALDH2*, and *ADH1B*), each having a nonsynonymous SNP associated with characteristic phenotypes of East Asians(Harada et al. 1981; Edenburg H.J. and Bosron W.F. 1997; Yoshiura et al. 2006; Fujimoto et al. 2008; Kimura et al. 2009). The derived alleles of these four nonsynonymous SNPs are associated with the following phenotypes: *ABCC11* rs17822931: dry ear wax(Yoshiura et al. 2006), *EDAR* rs3827760: thicker hair(Fujimoto et al. 2008) and shovel-shaped incisors(Kimura et al. 2009), and *ALDH2* rs671 and *ADH1B* rs1229984: lower alcohol tolerance(Harada et al. 1981; Edenburg H.J. and Bosron W.F. 1997). Haplotype structures composed of Jomon-derived SNPs are shown in Fig. 5. The haplotypes in each region could be classified into four types according to the presence or absence of the derived allele associated with the phenotype, and the composition of Jomon-derived alleles. The frequency of each haplotype in the Japanese population is presented in Table 1. Here, the haplotypes containing the Jomon-derived SNPs are called “Jomon-derived haplotypes.” For *ABCC11* and *EDAR*, Jomon-derived haplotypes were observed for both ancestral and derived alleles of the phenotype-associated SNPs. In *ABCC11* (Fig. 5 (a)), the frequencies of the Jomon-derived haplotypes in mainland Japanese were 10.6% for the ancestral allele (wet ear wax) and 24% for the derived allele (dry ear wax).

In *EDAR* (Fig. 5 (b)), the Jomon-derived haplotype frequencies were 14.9% for the ancestral allele (thinner hair and non-shovel-shaped incisors) and 17.3% for the derived allele (thicker hair and shovel-shaped incisors) in mainland Japanese. The haplotypes containing ancestral and derived alleles of the phenotype-associated SNPs had different Jomon alleles. Thus, it is unlikely that these haplotypes were generated by recombination in the Japanese population after the admixture between the Jomon and immigrants from continental East Asia. The present results suggest that modern Japanese have derived alleles from both the Jomon people and immigrants from continental East Asia in *EDAR* and *ABCC11*. Previous studies examining ancient DNA of the Hokkaido Jomon population obtained from archeological sites showed that the derived allele (dry ear wax) of *ABCC11* was present in the Hokkaido Jomon population at a frequency of 47.6%(Sato et al. 2009; Kazuta et al. 2011). As for *EDAR*, although the frequency of the derived allele (shovel-shaped incisors and thicker hair) in the Jomon people has not been estimated, it has been shown that shovel-shaped incisors were found at a frequency of 68.9% in the Jomon people(Matsumura 1994). The results of these previous studies are consistent with our results. In *ALDH2* rs671 (Fig. 5 (c)), the Jomon-derived haplotypes containing the derived allele of phenotype-associated SNPs were found to be rare (2.4%) in modern Japanese, and the number of Jomon alleles per Jomon-derived haplotype containing the derived allele was very small. This suggests that the Jomon people had few derived alleles (lower alcohol tolerance) of rs671, and most of the derived alleles found in modern Japanese originated from continental East Asians. In *ADH1B* rs1229984 (Fig. 5 (d)), the total frequency of the Jomon-derived

haplotypes was relatively lower than that of the other three genes. The frequencies of the Jomon-derived haplotype were 3.8% for the ancestral allele (higher alcohol tolerance) and 4.8% for the derived allele (lower alcohol tolerance). In addition, when we calculated the number of Jomon-derived SNPs per 1 Mb at the genome-wide scale (Supplementary Fig. 15), we found that the number of Jomon-derived SNPs was especially small in the region around *ADH1B* (red dashed line). Therefore, regarding *ADH1B*, it is possible that the Jomon people possessed the derived allele of rs1229984 at a higher frequency compared to that of *ALDH2,* but both the Jomon-derived haplotypes with ancestral and derived alleles may have been lost after the admixture in Japanese. Koganebuchi et al., (Koganebuchi et al. 2017) previously estimated that most of the derived alleles in *ALDH2* originated from immigrants from continental East Asia, which agrees with our results, while they concluded that the genetic contribution of immigrants was small for *ADH1B*, which contradicts the results of this study. Their study(Koganebuchi et al. 2017) assumed that, among mainland Japanese, the population in northern Kyushu had a relatively large genetic contribution from immigrants, but this assumption is inconsistent with the *JAS* estimated in the present study (Fig. 3 (b)). Thus, it is more plausible that *ADH1B* haplotypes of mainland Japanese were introduced mainly by immigrants from continental East Asia, regardless of the allelic status (ancestral or derived) of rs1229984.

Discussion

In this study, we developed the *AMI* as a summary statistic to detect the Jomon-derived variants in modern Japanese without requiring any genomic sequences from the former. The computer simulation showed that *AMI* can detect ancestral variants with high accuracy, even in an admixed population whose source populations diverged tens of thousands of years ago. Since we were able to detect Jomon-derived SNPs by the *AMI* even changing the population history in the simulations, the present approach using the *AMI* is likely to be applied to other admixed populations which source population whose source populations diverged relatively recently. Potential applications include the population of Madagascar(Pierron et al. 2017; Pierron et al. 2018) and the current Polynesian population(Skoglund et al. 2016; Pugach et al. 2018), which were formed around hundreds to thousands years ago by population admixture. As exemplified by these cases, the genetic diversity of modern humans has been greatly influenced by population admixture events(Lazaridis et al. 2014; Nielsen et al. 2017). The *AMI* will be a powerful tool for clarifying the population history of not only the Japanese but also other admixed populations.

As for the process of population formation in the Japanese archipelago from the Late Jomon period to the present, we propose a model, which is shown in Fig. 6. From the Late to Final Jomon period, the Jomon people settled down in mainland Japan, and the population size or the population density of the Jomon people varied among regions. According to Koyama 1979(Koyama 1979), based

395 on the number of archeological sites, it was estimated that the population sizes
396 in the Tohoku and Kanto regions were relatively large at 43,800 and 52,100,
397 respectively, while those in the Kinki and Shikoku regions were relatively small at
398 4,400 and 2,700, respectively. Thus, in the Kinki and Shikoku regions, modern
399 Japanese have lower degrees of genomic components derived from the Jomon
400 people. In the Final Jomon period, continental East Asians arrived in northern
401 Kyushu and started to admix with the Jomon people in all regions of mainland
402 Japan. During the Yayoi period, the population size of immigrants was relatively
403 increased in the Kinki and Shikoku regions, where the populations were small at
404 the end of the Jomon period. Further analyses of ancient human DNA from the
405 Final Jomon period to the Yayoi period will allow the verification of the Japanese
406 population history model proposed in this study.

Methods

Coalescent simulation by Msprime

To investigate the characteristics of the Jomon-derived autosomal genomic components of mainland Japanese, we conducted a coalescent simulation assuming the admixture of the Jomon and continental East Asians using msprime(Kelleher et al. 2016) (Supplementary Fig. 2). A feature of the msprime program is that it specifies the time and population where the mutation and coalescence occurred. Our simulation code was made with reference to a previous study(Browning et al. 2018). The split between the Jomon ancestors and continental East Asians was set to 1,200 generations ago (30,000 YBP), according to the divergence time estimated in Kanzawa-Kiriyama et al., (Kanzawa-Kiriyama et al. 2019) (between 18,000 YBP and 38,000 YBP) and the beginning of the Jomon period (around 16,000 YBP)(Habu 2004). Migration from continental East Asia to mainland Japan was set between 120 and 80 generations ago, with reference to the beginning of the Yayoi period, around 2,800 years ago(Fujio 2015). The total admixture proportion of the Jomon people in the modern mainland Japanese was set to 12%(Kanzawa-Kiriyama et al. 2017). The effective population size was set to 5,000 for both populations. The mutation rate and recombination rate were set to 1.2×10^{-8} per bp per generation and 1.3×10^{-8} per bp per generation, respectively(International Human Genome Sequencing Consortium 2001; Altshuler et al. 2010; Kong et al. 2012; Scally and Durbin 2012).

This study aimed to detect Jomon-derived variants based on LD among Japanese specific variants. There are three types of Japanese specific variants: (type 1) Jomon-derived variants, which appeared in the Jomon lineage before the admixture event; (type 2) variants derived from continental East Asians, which appeared in the continental East Asian populations and were moved into the Japanese lineages through the admixture, but were eventually lost in the East Asian population; and (type 3) novel variants that appeared only in Japanese lineages after the admixture. It is worth noting that Japanese specific variants generated earlier than the split time of the Jomon people and the continental East Asians were classified as Jomon-derived variants (type 1). Of these Japanese specific variants, the Jomon-derived variants (type 1) are considered to be accumulated on the same haplotype or to be in strong LD with each other (Supplementary Fig. 4 (b)). We compared the LD status of three types of Japanese specific variants by coalescent simulations. The origin of each haplotype of mainland Japanese can be estimated by coalescent time to the haplotypes of the Jomon or continental East Asians by msprime simulations. That is, if a haplotype of a mainland Japanese sample coalesced with haplotypes of Jomon samples earlier than the admixture of the Jomon people and continental East Asians, the haplotype is inferred to be derived from Jomon. To extract the three types of Japanese specific variants (i.e., variants not found in samples from continental East Asians), 3,000 replicates of 1 Mb simulations were performed. We sampled 200 haplotypes from each of the four populations (modern mainland Japanese, modern continental East Asians, Jomon people 120 generations ago,

and continental East Asians 120 generations ago) to detect variants observed in modern mainland Japanese but not seen in continental East Asians. Each Japanese specific variant was classified into (type 1) the Jomon-derived variant, (type 2) the continental East Asian-derived variant, and (type 3) the novel variant based on when and in which lineage the mutation occurred (Supplementary Fig. 4 (a)). For each type of the Japanese specific variant, the ancestry marker index (*AMI*) was calculated as:

$$AMI = \frac{\{Number\ of\ variants\ with\ linkage\ disequilibrium\ coefficients\ (r^2)\ > 0.01\}}{(Number\ of\ Japanese\ specific\ variants\ per\ KB)}$$

Jomon-derived variants are expected to have higher *AMI* values. The performance of the *AMI* was verified by receiver operating characteristic (ROC) analysis using the ROCR package in R. The threshold to detect Jomon-derived variants was determined based on the Youden Index.

Detection and verification of Jomon-derived SNPs on autosomes using real data

Detection of Jomon-derived variants in real data

Jomon-derived SNPs were inferred from the whole genome sequence data from 26 populations from different parts of the world, including mainland Japanese (JPT) and four continental East Asian populations (CHB, CHS, CDX, and KHV), obtained from the 1000 Genomes Project Phase III (1KG)(Auton et al. 2015), and 87 individuals from the Korean Personal Genome Project(Zhang et al. 2014). In this study, only biallelic SNPs were used. Prior to extracting the Jomon-

479 derived SNPs, we performed a principal component analysis (PCA) in PLINK
480 (version 1.9)(Purcell et al. 2007) using 1KG mainland Japanese (JPT) and Han
481 Chinese (CHB) data. During this analysis, we found that one JPT individual
482 (NA18976) was close to the continental East Asians (Supplementary Fig. 8), so
483 NA18976 was excluded from subsequent analyses. First, 1,784,634 SNPs
484 specific to 1KG JPT were detected using VCFtools v0.1.13(Danecek et al. 2011).
485 Next, LD coefficients (r^2) were calculated between the Japanese specific SNPs
486 located within 1 Mb from each other with the --hap-r2 option of VCFtools in
487 combination with the --ld-window-bp option. The number of SNPs with $r^2 > 0.01$
488 was counted for each Japanese specific SNP. The density of Japanese specific
489 variants per 1 kb of each chromosome was calculated using the --SNPdensity
490 option of VCFtools, and the *AMI* was calculated for each Japanese specific SNP.
491 To eliminate the possibility of sequence errors, regions with a density of Japanese
492 specific variants per kb below a mean of - 1sd of each chromosome were
493 excluded from the analysis. In this analysis, we assumed that the number of
494 Japanese specific variants per kb, which is the denominator of the *AMI*, is
495 constant for each chromosome (i.e., the numerator of the *AMI* was normalized
496 for each chromosome). Based on the threshold set by the ROC analysis of
497 simulated Japanese specific variants, we inferred variants originating from the
498 Jomon people.
499
500 **Verification of Jomon-derived SNPs based on whole-genome sequence**
501 **data from Jomon remains**

For the verification of Jomon-derived SNPs based on the whole genome sequence data, the "Jomon allele score" (*JAS*) was calculated for the Ikawazu(Mccoll et al. 2018; Gakuhari et al. 2020) and Funadomari(Kanzawa-Kiriyama et al. 2019) Jomon, as well as for 104 individuals from the 1KG JPT. The *JAS* was calculated using the following formula:

$$JAS = \frac{(Jomon\ derived\ allele\ count)}{2*(\text{Number of total Jomon}-\text{derived SNPs})}.$$

The BAM file of the Ikawazu Jomon was provided by Hiroki Ota of Tokyo University, Tokyo, Japan, and Takashi Gakuhari of Kanazawa University, Ishikawa, Japan. The BAM file of the Funadomari Jomon was provided by Naruya Saito from the National Institute of Genetics, Shizuoka, Japan, and Hideaki Kanzawa-Kiriyama from the National Museum of Nature and Science, Tokyo, Japan. The genotypes of Ikawazu Jomon and Funadomari Jomon samples were called by the UnifiedGenotyper tool in the GenomeAnalysisToolkit version 3.6(McKenna et al. 2010). For the Ikawazu Jomon, the --mbq 30 --ploidy 2 --output_mode EMIT_ALL_CONFIDENT_SITES options were specified. For the Funadomari Jomon, the options described in the original paper were specified. Jomon SNPs were subjected to LD pruning by the --indep-pairwise command of PLINK (--indep-pairwise 1000 200 0.8). In addition, only the Jomon-derived SNPs of depth ≥ 6 in the Ikawazu and Funadomari Jomon were used for the calculation of the *JAS*. As a result, 4,458 SNPs were used to calculate *JAS*.

Detection of regional genetic differences in mainland Japanese by Jomon-derived SNPs

526 **Sample data**

527 We used 183,708 SNPs from 10,842 individuals from the Japanese
528 archipelago published by Watanabe et al.(Watanabe et al. 2020). All the
529 individuals investigated in this study were customers of the Japanese Direct to
530 Consumer (DTC) genetic-testing service, HealthData Lab (Yahoo! Japan
531 Corporation, Tokyo, Japan). They were provided an agreement, and informed
532 consent was obtained for their data to be used for research. In this study, the
533 Japanese archipelago was divided roughly into eleven regions (Fig. 1 and
534 Supplementary Table 1): Hokkaido (430 individuals), Tohoku (746 individuals),
535 Kanto (3,990 individuals), Hokuriku (431 individuals), Chubu (410 individuals),
536 Tokai (933 individuals), Kinki (1,861 individuals), Chugoku (600 individuals),
537 Shikoku (314 individuals), Kyushu (1,016 individuals), and Okinawa (111
538 individuals). All statistical analyses were conducted at the Yahoo! Japan
539 Corporation, with personal information of the customers completely hidden. We
540 obtained approval from the Ethics Committee of the Yahoo! Japan Corporation.
541

542 **Imputation of genotypes of Jomon-derived SNPs**

543 Haplotype phasing and genotype imputation were performed using
544 EAGLE2(Loh et al. 2016) and Minimac3(Das et al. 2016), respectively, with whole
545 genome sequence data of 413 mainland Japanese(Watanabe et al. 2019)
546 phased by SHAPEIT2(Delaneau et al. 2013). After the imputation, Jomon-derived
547 SNPs with high imputation quality ($R^2 > 0.8$) were extracted. Also, LD pruning
548 was performed with PLINK (--indep-pairwise 1000 200 0.1), and a total of 3,917
549 Jomon-derived SNPs were used for the analysis.

550

551 **Geographical distribution of the Jomon allele score**

552 In subsequent analyses, individuals from Hokkaido that were largely affected
553 by immigration after the Meiji period were excluded. Using 3,917 Jomon-derived
554 SNPs, we calculated the *JAS* for individuals of each prefecture and compared
555 them between regions and prefectures.

556

557 ***f3*-testing of prefectural populations in Japan**

558 The *f3*-test(Patterson et al. 2012) was carried out in order to examine the
559 relatedness between contemporary populations of each prefecture in Japan and
560 the Funadomari Jomon or Ikawazu Jomon. Each Jomon sample and the 1KG
561 CHB were set as the source populations of admixture, and each prefecture of
562 Japan was set as the target population, (described as *f3*(each prefecture; Jomon,
563 CHB)) . LD pruning was carried out on whole genome SNPs common in the
564 Japanese, 1KG CHB, Funadomari Jomon, and Ikawazu Jomon by PLINK (--
565 indep-pairwise 1000 200 0.1), with 17,492 SNPs being used for subsequent
566 analyses. For the Funadomari and Ikawazu Jomon people, VCF files were
567 converted to PED files using VCFtools and combined with the PED file of the
568 Japanese. The PED files were converted to eigenstrat format with the
569 Admixtools(Patterson et al. 2012) convertf command, and then the *f3*-test was
570 conducted with the Admixtools qp3Pop command.

571

572 **Examination of correlations between population size during the Jomon**
573 **period and *JAS* in each prefecture**

We compared the population size estimated from the number of archeological sites in each prefecture, assuming that the population size per archeological site was constant in each prefecture during the same period. We examined the correlations between (a) the average *JAS* in each prefecture and the number of archeological sites from the Jomon period (obtained from the Statistical report of buried cultural properties, Agency of Cultural Affairs, Japan; https://www.bunka.go.jp/seisaku/bunkazai/shokai/pdf/h29_03_maizotokei.pdf), (b) the average *JAS* in each region and the population size estimated from the number of archeological sites in the Late Jomon period(Koyama 1979), and (c) the average *JAS* in each prefecture and the $\log_{10}$(number of archeological sites in the Yayoi period/number of archeological sites in the Late Jomon period)(Koyama 1979). Finally, (a) and (c) were plotted for each prefecture, while (b) was plotted for each region because data for each prefecture were not available. Correlation test was conducted by R cor.test function (df = 43).

PCA based on the Jomon allele frequencies of each prefecture

The Jomon allele frequency was calculated for 50 randomly sampled individuals from each prefecture using VCFtools version 0.1.13. A PCA was performed based on the Jomon allele frequency using R version 3.6.0. Correlation test between PC1 and JAS, and PC2 and longitude/latitude were conducted by R cor.test function (df = 43).

Analysis of haplotype structures composed of Jomon-derived SNPs in four genes associated with characteristic phenotypes of East Asians

598 We investigated the haplotype structures composed of Jomon-derived SNPs
599 in the regions surrounding SNPs associated with characteristic phenotypes of
600 East Asians (*ABCC11*: rs17822931(Yoshiura et al. 2006), *EDAR*:
601 rs3827760(Fujimoto et al. 2008; Kimura et al. 2009), *ALDH2*: rs671(Harada et al.
602 1981; Oota et al. 2004; Li et al. 2009; Luo et al. 2009; Koganebuchi et al. 2017),
603 and *ADH1B*: rs1229984(Edenburg H.J. and Bosron W.F. 1997; Osier et al. 1999;
604 Osier et al. 2002; Han et al. 2007; Li et al. 2007)). The --IMPUTE option of
605 VCFtools was used to extract haplotypes composed of Jomon-derived SNPs in
606 regions 500 kb upstream and downstream (1 Mb in total) of the SNPs associated
607 with phenotypes from the phased 1KG JPT whole genome dataset.
608
609 **Data availability**
610 The individual genotypes of 10,842 Japanese analyzed in this study are
611 not available to avoid personal identification. The list of Jomon-derived SNPs
612 detected in this study, and the allele frequencies of Jomon-derived SNPs in each
613 Japanese prefecture are available from the corresponding author upon request.
614 Our custom code for msprime simulation are available from the corresponding
615 author upon request.
616

617 Altshuler DL, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, Clark AG,
618 Collins FS, De La Vega FM, Donnelly P, Egholm M, et al. 2010. A map of
619 human genome variation from population-scale sequencing. Nature
620 467:1061–1073.
621 Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR,
622 Chakravarti A, Clark AG, Donnelly P, Eichler EE, et al. 2015. A global
623 reference for human genetic variation. Nature 526:68–74.
624 Baran Y, Pasaniuc B, Sankararaman S, Torgerson DG, Gignoux C, Eng C,
625 Rodriguez-Cintron W, Chapela R, Ford JG, Avila PC, et al. 2012. Fast and
626 accurate inference of local ancestry in Latino populations. Bioinformatics
627 28:1359–1367.
628 Browning SR, Browning BL, Zhou Y, Tucci S, Akey JM. 2018. Analysis of
629 Human Sequence Data Reveals Two Pulses of Archaic Denisovan
630 Admixture. Cell 173:1–9.
631 Churchhouse C, Marchini J. 2013. Multiway Admixture Deconvolution Using
632 Phased or Unphased Ancestral Panels. Genet. Epidemiol. 37:1–12.
633 Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA,
634 Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call
635 format and VCFtools. Bioinformatics 27:2156–2158.
636 Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, Vrieze SI, Chew
637 EY, Levy S, McGue M, et al. 2016. Next-generation genotype imputation
638 service and methods. Nat. Genet. 48:1284–1287.
639 Delaneau O, Zagury JF, Marchini J. 2013. Improved whole-chromosome
640 phasing for disease and population genetic studies. Nat. Methods 10:5–6.

641 Edenburg H.J., Bosron W.F. 1997. Alcohol dehydrogenases. In:
642 Comprehensive Toxicology, Vol. 3: Biotransformation. (Guengerich F.P,
643 editor.). New York: Pergamon Elsevier Science
644 Fujimoto A, Kimura R, Ohashi J, Omi K, Yuliwulandari R, Batubara L, Mustofa
645 MS, Samakkarn U, Settheetham-ishida W, Ishida T, et al. 2008. A scan for
646 genetic determinants of human hair morphology: EDAR is associated with
647 Asian hair thickness. Hum. Mol. Genet. 17:835–843.
648 Fujio S. 2015. History of Yayoi Period. Tokyo: Kodansha
649 Fukase H, Wakebe T, Tsurumoto T, Saiki K, Fujita M, Ishida H. 2012.
650 Geographic variation in body form of prehistoric Jomon males in the
651 Japanese archipelago: Its ecogeographic implication. Am. J. Phys.
652 Anthropol. 149:125–135.
653 Gakuhari T, Nakagome S, Rasmussen S, Allentoft ME, Sato T, Chuinneagáin
654 BN, Matsumae H, Koganebuchi K, Schmidt R, Mizushima S, et al. 2020.
655 Ancient Jomon genome sequence analysis sheds light on migration
656 patterns of early East Asian. Commun. Biol. 3:1–10.
657 Guan Y. 2014. Detecting structure of haplotypes and local ancestry. Genetics
658 196:625–642.
659 Habu J. 2004. Ancient Jomon of Japan. Cambridge University Press Available
660 Han Y, Gu S, Oota H, Osier M V., Pakstis AJ, Speed WC, Kidd JR, Kidd KK.
661 2007. Evidence of positive selection on a class I ADH locus. Am. J. Hum.
662 Genet. 80:441–456.
663 Hanihara K. 1985. Geographic variation of modern Japanese crania and its
664 relationship to the origin of Japanese. Homo 36:1–10.

665 Hanihara K. 1991. Dual structure model for the population history of the
666 Japanese. Japan Rev. 2:1–33.
667 Harada S, Agarwal DP, Goedde HW. 1981. Aldehyde Dehydrogenase
668 Deficiency As Cause of Facial Flushing Reaction To Alcohol in Japanese.
669 Lancet 318:982.
670 Hui D, Fang Z, Lin J, Duan Q, Li Y, Hu M, Chen W. 2017. LAIT: A local ancestry
671 inference toolkit. BMC Genet. 18:1–5.
672 International Human Genome Sequencing Consortium. 2001. Initial sequencing
673 and analysis of the human genome International Human Genome. Nature
674 409:860–921.
675 Jinam T, Nishida N, Hirai M, Kawamura S, Oota H, Umetsu K, Kimura R,
676 Ohashi J, Tajima A, Yamamoto T, et al. 2012. The history of human
677 populations in the Japanese Archipelago inferred from genome-wide SNP
678 data with a special reference to the Ainu and the Ryukyuan populations. J.
679 Hum. Genet. 57:787–795.
680 Jinam TA, Kanzawa-Kiriyama H, Inoue I, Tokunaga K, Omoto K, Saitou N.
681 2015. Unique characteristics of the Ainu population in Northern Japan. J.
682 Hum. Genet. 60:565–571.
683 Kaifu Y. 1995. Regional variation in mandibular morphology of the Jomon
684 people. Bull Natl Sci Mus 21:37–50.
685 Kanzawa-Kiriyama H, Jinam TA, Kawai Y, Sato T, Hosomichi K, Tajima A,
686 Adachi N, Matsumura H, Kryukov K, Saitou N, et al. 2019. Late Jomon
687 male and female genome sequences from the Funadomari site in
688 Hokkaido , Japan. Anthropol. Sci. 127:83–108.

689 Kanzawa-Kiriyama H, Kryukov K, Jinam TA, Hosomichi K, Saso A, Suwa G,
690 Ueda S, Yoneda M, Tajima A, Shinoda K, et al. 2017. A partial nuclear
691 genome of the Jomons who lived 3000 years ago in Fukushima, Japan. J.
692 Hum. Genet. 62:213–221.
693 Kazuta H, Sato T, Dodo Y, Matsumura H, Amano T, Ishida H, Masuda R. 2011.
694 Genotype frequencies of the ABCC11 gene in 2000-3000-year-old human
695 bones from the Epi-Jomon and Jomon sites in Hokkaido, Japan. Anthropol.
696 Sci. 119:81–86.
697 Kelleher J, Etheridge AM, McVean G. 2016. Efficient Coalescent Simulation and
698 Genealogical Analysis for Large Sample Sizes. PLoS Comput. Biol. 12:1–
699 22.
700 Kimura R, Yamaguchi T, Takeda M, Kondo O, Toma T, Haneji K, Hanihara T,
701 Matsukusa H, Kawamura S, Maki K, et al. 2009. A Common Variation in
702 EDAR Is a Genetic Determinant of Shovel-Shaped Incisors. Am. J. Hum.
703 Genet. 85:528–535.
704 Koganebuchi K, Haneji K, Toma T, Joh K, Soejima H, Fujimoto K, Ishida H,
705 Ogawa M, Hanihara T, Harada S, et al. 2017. The allele frequency of
706 ALDH2*Glu504Lys and ADH1B*Arg47His for the Ryukyu islanders and
707 their history of expansion among East Asians. Am. J. Hum. Biol. 29:1–14.
708 Kondo O, Fukase H, Fukumoto T. 2017. Regional variations in the Jomon
709 population revisited on craniofacial morphology. Anthropol. Sci. 125:85–
710 100.
711 Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G,
712 Gudjonsson SA, Sigurdsson A, Jonasdottir Aslaug, Jonasdottir Adalbjorg,

et al. 2012. Rate of de novo mutations and the importance of father-s age to disease risk. Nature 488:471–475.

Koyama S. 1979. Jomon Subsistence and Population. Senri Ethnol. Stud. 2:1–65.

Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, Kirsanow K, Sudmant PH, Schraiber JG, Castellano S, Lipson M, et al. 2014. Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature 513:409–413.

Li H, Borinskaya S, Yoshimura K, Kal'ina N, Marusin A, Stepanov VA, Qin Z, Khaliq S, Lee MY, Yang Y, et al. 2009. Refined geographic distribution of the oriental ALDH2* 504Lys (nee 487Lys) variant. Ann. Hum. Genet. 73:335–345.

Li H, Mukherjee N, Soundararajan U, Tárnok Z, Barta C, Khaliq S, Mohyuddin A, Kajuna SLB, Mehdi SQ, Kidd JR, et al. 2007. Geographically separate increases in the frequency of the derived ADH1B*47His allele in eastern and western Asia. Am. J. Hum. Genet. 81:842–846.

Loh PR, Danecek P, Palamara PF, Fuchsberger C, Reshef YA, Finucane HK, Schoenherr S, Forer L, McCarthy S, Abecasis GR, et al. 2016. Reference-based phasing using the Haplotype Reference Consortium panel. Nat. Genet. 48:1443–1448.

Luo HR, Wu GS, Pakstis AJ, Tong L, Oota H, Kidd KK, Zhang YP. 2009. Origin and dispersal of atypical aldehyde dehydrogenase ALDH2*487Lys. Gene 435:96–103.

736 Maeda T. 2002. Mandibular ramus morphology of the Jomon people from
737 Hokkaido. Anthropological. Anthropol. Sci. (Japanese Ser. 110:27–40.
738 Matsumura H. 1994. A Microevolutional History of the Japanese People from a
739 Dental Characteristics Perspective. Anthropol. Sci. 102:93–118.
740 Mccoll H, Racimo F, Vinner L, Demeter F, Gakuhari T, Moreno-mayar JV,
741 Driem G Van, Wilken UG, Seguin-orlando A, De C, et al. 2018. The
742 prehistoric peopling of Southeast Asia. Science (80-.). 361:88–92.
743 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A,
744 Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The genome
745 analysis toolkit: A MapReduce framework for analyzing next-generation
746 DNA sequencing data. Genome Res. 20:1297–1303.
747 Mizoguchi K. 2013. The archaeology of Japan : from the earliest rice farming
748 villages to the rise of the state. Cambrige: Cambrige University Press
749 Nakagome S, Sato T, Ishida H, Hanihara T, Yamaguchi T, Kimura R, Mano S,
750 Oota H, Omoto K, Tokunaga K, et al. 2015. Model-based verification of
751 hypotheses on the origin of modern Japanese revisited by Bayesian
752 inference based on genome-wide SNP data. Mol. Biol. Evol. 32:1533–
753 1543.
754 Nielsen R, Akey JM, Jakobsson M, Pritchard JK, Tishkoff S, Willerslev E. 2017.
755 Tracing the peopling of the world through genomics. Nature 541:302–310.
756 Oota H, Pakstis AJ, Bonne-Tamir B, Goldman D, Grigorenko E, Kajuna SLB,
757 Karoma NJ, Kungulilo S, Lu RB, Odunsi K, et al. 2004. The evolution and
758 population genetics of the ALDH2 locus: Random genetic drift, selection,
759 and low levels of recombination. Ann. Hum. Genet. 68:93–109.

760 Osier M, Pakstis AJ, Kidd JR, Lee JF, Yin SJ, Ko HC, Edenberg HJ, Lu RB,
761 Kidd KK. 1999. Linkage disequilibrium at the ADH2 and ADH3 loci and risk
762 of alcoholism. Am. J. Hum. Genet. 64:1147–1157.
763 Osier M V., Pakstis AJ, Soodyall H, Comas D, Goldman D, Odunsi A, Okonofua
764 F, Parnas J, Schulz LO, Bertranpetit J, et al. 2002. A global perspective on
765 genetic variation at the ADH genes reveals unusual patterns of linkage
766 disequilibrium and diversity. Am. J. Hum. Genet. 71:84–99.
767 Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T,
768 Webster T, Reich D. 2012. Ancient admixture in human history. Genetics
769 192:1065–1093.
770 Pierron D, Heiske M, Razafindrazaka H, Pereda-Loth V, Sanchez J, Alva O,
771 Arachiche A, Boland A, Olaso R, Deleuze JF, et al. 2018. Strong selection
772 during the last millennium for African ancestry in the admixed population of
773 Madagascar. Nat. Commun. 9:1–9.
774 Pierron D, Heiske M, Razafindrazaka H, Rakoto I, Rabetokotany N,
775 Ravololomanga B, Rakotozafy LMA, Rakotomalala MM, Razafiarivony M,
776 Rasoarifetra B, et al. 2017. Genomic landscape of human diversity across
777 Madagascar. Proc. Natl. Acad. Sci. U. S. A. 114:E6498–E6506.
778 Plagnol V, Wall JD. 2006. Possible ancestral structure in human populations.
779 PLoS Genet. 2:e105.
780 Price AL, Tandon A, Patterson N, Barnes KC, Rafaels N, Ruczinski I, Beaty TH,
781 Mathias R, Reich D, Myers S. 2009. Sensitive detection of chromosomal
782 segments of distinct ancestry in admixed populations. PLoS Genet.
783 5:e1000519.

784 Pugach I, Duggan AT, Merriwether DA, Friedlaender FR, Friedlaender JS,
785 Stoneking M. 2018. The gateway from near into remote oceania: New
786 insights from genome-wide data. Mol. Biol. Evol. 35:871–886.
787 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller
788 J, Sklar P, De Bakker PIW, Daly MJ, et al. 2007. PLINK: A tool set for
789 whole-genome association and population-based linkage analyses. Am. J.
790 Hum. Genet. 81:559–575.
791 Sato T, Amano T, Ono H, Ishida H, Kodera H, Matsumura H, Yoneda M,
792 Masuda R. 2009. Allele frequencies of the ABCC11 gene for earwax
793 phenotypes among ancient populations of Hokkaido, Japan. J. Hum.
794 Genet. 54:409–413.
795 Scally A, Durbin R. 2012. Revising the human mutation rate: Implications for
796 understand ing human evolution. Nat. Rev. Genet. 13:745–753.
797 Skoglund P, Posth C, Sirak K, Spriggs M, Valentin F, Bedford S, Clark GR,
798 Reepmeyer C, Petchey F, Fernandes D, et al. 2016. Genomic insights into
799 the peopling of the Southwest Pacific. Nature 538:510–513.
800 Takigawa W. 2006. Inter-regional variation of metric traits of limb bones in
801 Jomon and modern Japanese. Anthropological. Anthropol. Sci. (Japanese
802 Ser. 114:101–129.
803 Vernot B, Akey JM. 2014. Resurrecting Surviving Neandeltal Linages from
804 Modern Human Genomes. Science (80-.). 343:1017–1021.
805 Watanabe Y, Isshiki M, Ohashi J. 2020. Prefecture-level population structure of
806 the Japanese based on SNP genotypes of 11,069 individuals. J. Hum.
807 Genet. In press.

808 Watanabe Y, Naka I, Khor S-S, Sawai H, Hitomi Y, Tokunaga K, Ohashi J.
809 2019. Analysis of whole Y-chromosome sequences reveals the Japanese
810 population history in the Jomon period. Sci. Rep. 9:8556.
811 Yamaguchi-Kabata Y, Nakazono K, Takahashi A, Saito S, Hosono N, Kubo M,
812 Nakamura Y, Kamatani N. 2008. Japanese Population Structure, Based on
813 SNP Genotypes from 7003 Individuals Compared to Other Ethnic Groups:
814 Effects on Population-Based Association Studies. Am. J. Hum. Genet.
815 83:445–456.
816 Yoshiura KI, Kinoshita A, Ishida T, Ninokata A, Ishikawa T, Kaname T, Bannai
817 M, Tokunaga K, Sonoda S, Komaki R, et al. 2006. A SNP in the ABCC11
818 gene is the determinant of human earwax type. Nat. Genet. 38:324–330.
819 Zhang W, Meehan J, Su Z, Ng HW, Shu M, Luo H, Ge W, Perkins R, Tong W,
820 Hong H. 2014. Whole genome sequencing of 35 individuals provides
821 insights into the genetic architecture of Korean population. BMC
822 Bioinformatics 15:1–13.
823

Acknowledgments

We are grateful to the individuals who participated in the study. We would like to express our deepest gratitude to Mr. Masahiro Inoue, Shota Arichi, and Akito Tabira who obtained the genotype data and provided the technical environment for analyzing them. We would like to thank Dr. Hiroki Ota of Tokyo University, Tokyo, Japan, and Dr. Takashi Gakuhari of Kanazawa University, Ishikawa, Japan, for providing us the BAM file of the IKawazu Jomon. We also thank Dr. Naruya Saito from the National Institute of Genetics, Shizuoka, Japan, and Dr. Hideaki Kanzawa-Kiriyama from the National Museum of Nature and Science, Tokyo, Japan for providing us the BAM file of the Funadomari Jomon. This study was partly supported by Grant-in-Aid for Scientific Research (B) (18H02514) and Grant-in-Aid for Scientific Research on Innovative Areas (19H05341) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan. Computations were partially performed on the NIG supercomputer at ROIS National Institute of Genetics.

Author contributions

Y.W. and J.O. conceived the study. Y.W. designed and conducted the data analyses. Y.W. performed the computer simulations. Y.W. wrote the manuscript with support from J.O. J.O. supervised the project. All authors read and approved the final manuscript.

Competing interests

847 The authors declare no competing interests.

Table 1 Frequencies of the Jomon-derived haplotypes specified by phenotypes associated SNPs in *ABCC11*, *EDAR*, *ALDH2*, and *ADH1B*.

(a) *ABCC11*	rs17822931		
	Ancestral	Derived	Total
Jomon-derived haplotype	0.106	0.240	0.346
Non-Jomon–derived haplotype	0.014	0.639	0.653
Total	0.120	0.879	

(b) *EDAR*	rs3827760		
	Ancestral	Derived	Total
Jomon-derived haplotype	0.149	0.173	0.322
Non-Jomon–derived haplotype	0.048	0.630	0.678
Total	0.197	0.803	

(c) *ALDH2*	rs671		
	Ancestral	Derived	Total
Jomon-derived haplotype	0.125	0.024	0.149
Non-Jomon–derived haplotype	0.635	0.216	0.851
Total	0.760	0.240	

(d) *ADH1B*	rs1229984		
	Ancestral	Derived	Total
Jomon-derived haplotype	0.038	0.048	0.086
Non-Jomon–derived haplotype	0.231	0.683	0.914
Total	0.269	0.731	

Figures

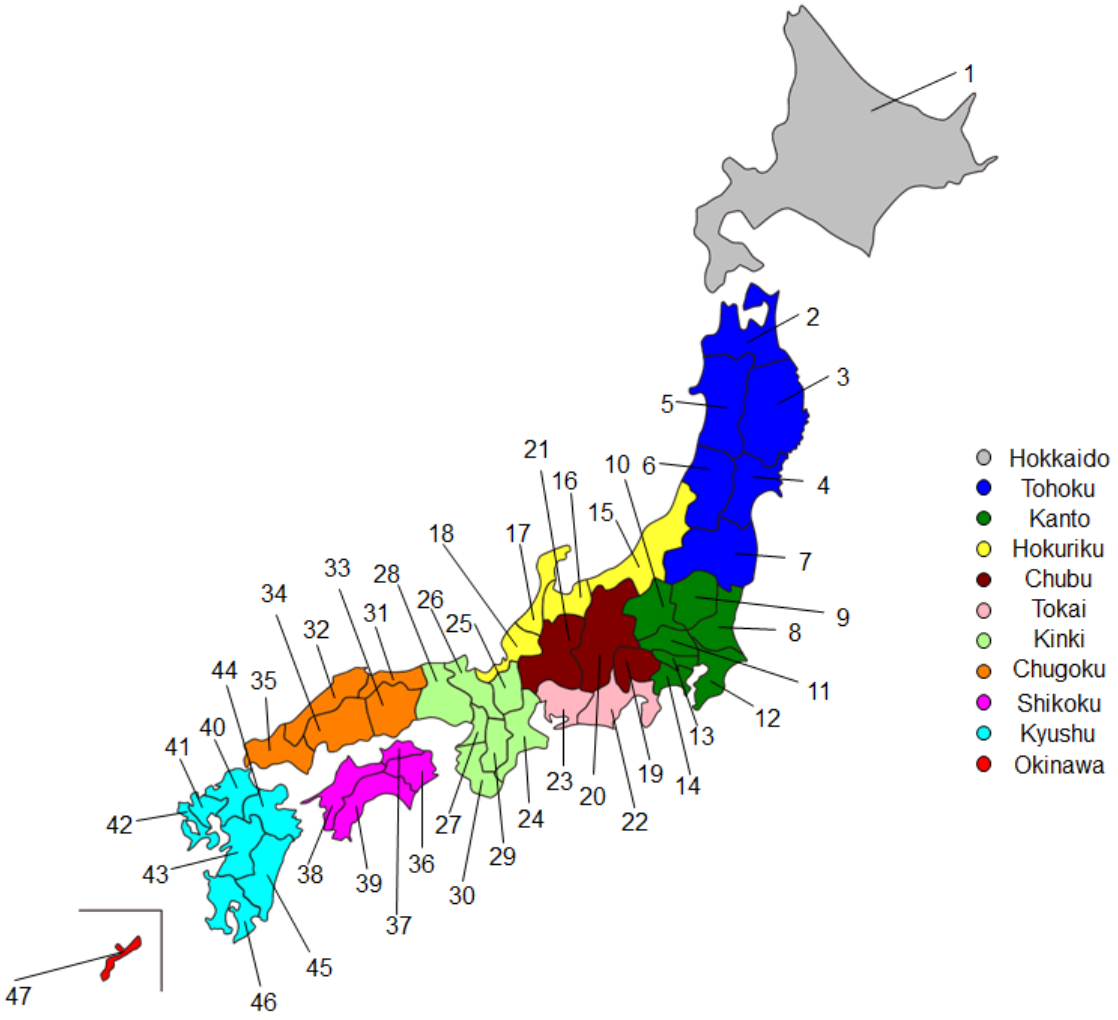
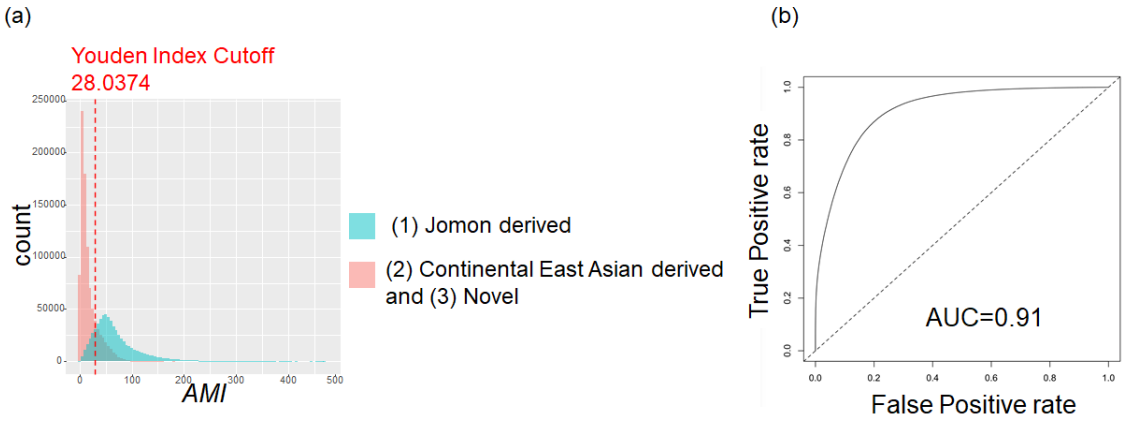


Fig. 1 Map of the Japanese prefectures. The prefectures of Japan are divided into eleven regions. The prefecture numbers in Supplementary Table 1 are indicated (the corresponding prefecture names are given in Supplementary Table 1). In this study, “mainland Japanese” means the Japanese people except for individuals from Hokkaido and Okinawa.



878

879

880 **Fig. 2 (a) Distribution of *AMI* simulated by msprime.** The histogram of *AMI* for
881 the Jomon-derived variants (type 1) and the other variants (types 2 and 3) are
882 shown. The red dashed line indicates the threshold of *AMI* (28.0374) obtained
883 from ROC analysis for the detection of the Jomon-derived variants (type 1). **(b)**
884 **ROC curve illustrating the performance of the *AMI* for the detection of the**
885 **Jomon-derived SNPs.** The ROC curve was drawn based on the simulated data
886 shown in Fig. 2 (a). The *AMI* showed high accuracy (AUC = 0.91) for
887 discriminating the Jomon-derived variants (type 1) from the other variants (types
888 2 and 3).

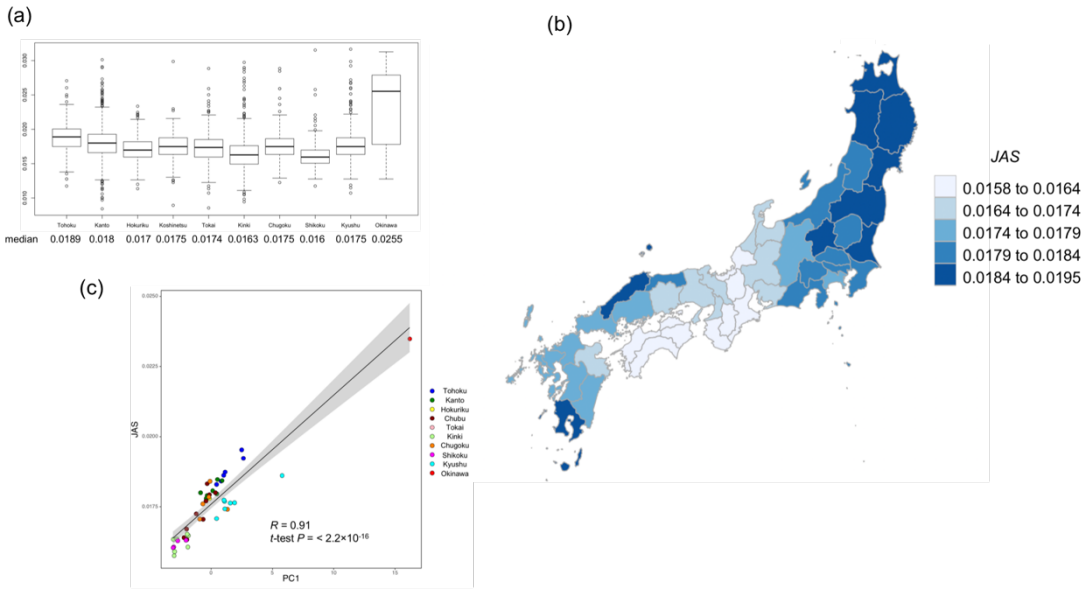


Fig. 3 (a) Distribution of *JAS* in ten regions. The boxplot of the *JAS* is presented for each of the ten regions, excluding Hokkaido. **(b) *JAS* of each prefecture in Mainland Japan.** The average *JAS* by prefecture was calculated. Hokkaido and Okinawa prefectures are not illustrated. The prefecture with the higher average *JAS* is illustrated by the darker color. **(c) Relationship between the PC1 of the PCA performed in a previous study by the allele frequency of autosomal 183,708 SNPs in each prefecture and average *JAS*.** Each prefecture was colored according to the region of Japan in Fig. 1. Horizontal axe: PC1, vertical axe: average *JAS*. Pearson's correlation coefficients (*R*), *P* values, regression lines and 95% CI are shown.

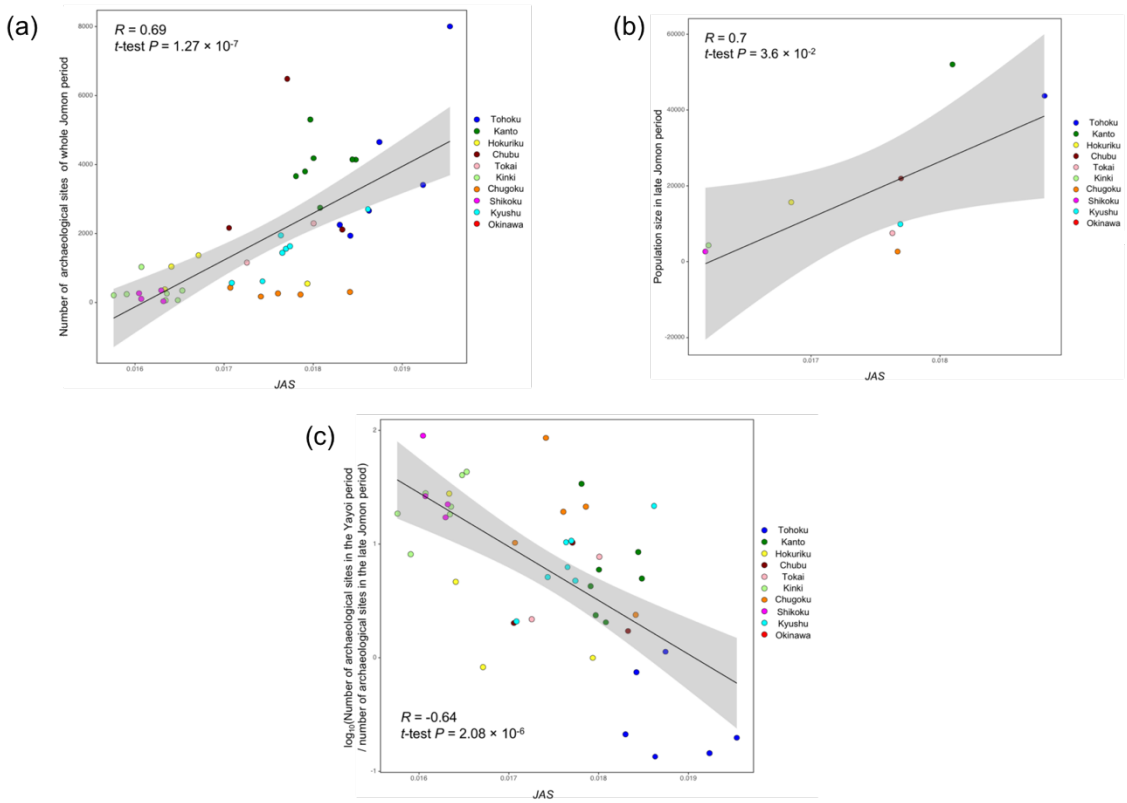


Fig. 4 Relationship between the *JAS* and values associated with the population size of prefectures in the Jomon to Yayoi periods. The horizontal axis shows the average *JAS* and the vertical axis shows the (a) number of archaeological sites of the whole Jomon period, (b) population size in the Late Jomon period, and (c) $\log_{10}$ (number of archaeological sites in the Yayoi period/number of archaeological sites in the Late Jomon period). Pearson's correlation coefficients (*R*), *P* values, regression lines and 95% CI are shown in each figure. Each prefecture is colored according to the region in Fig. 1.

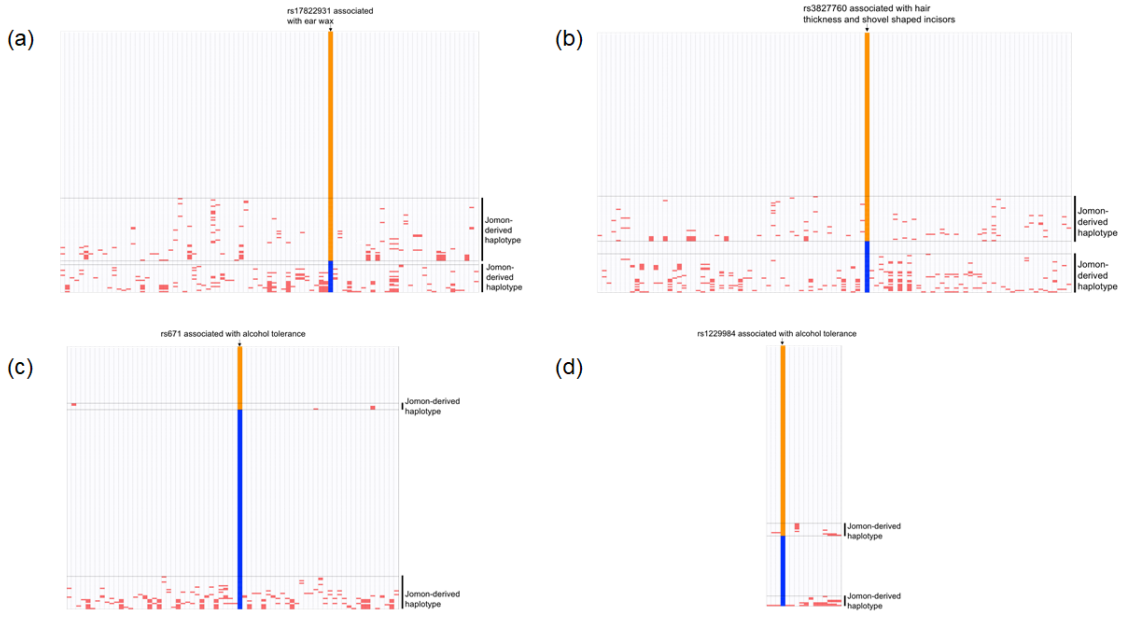
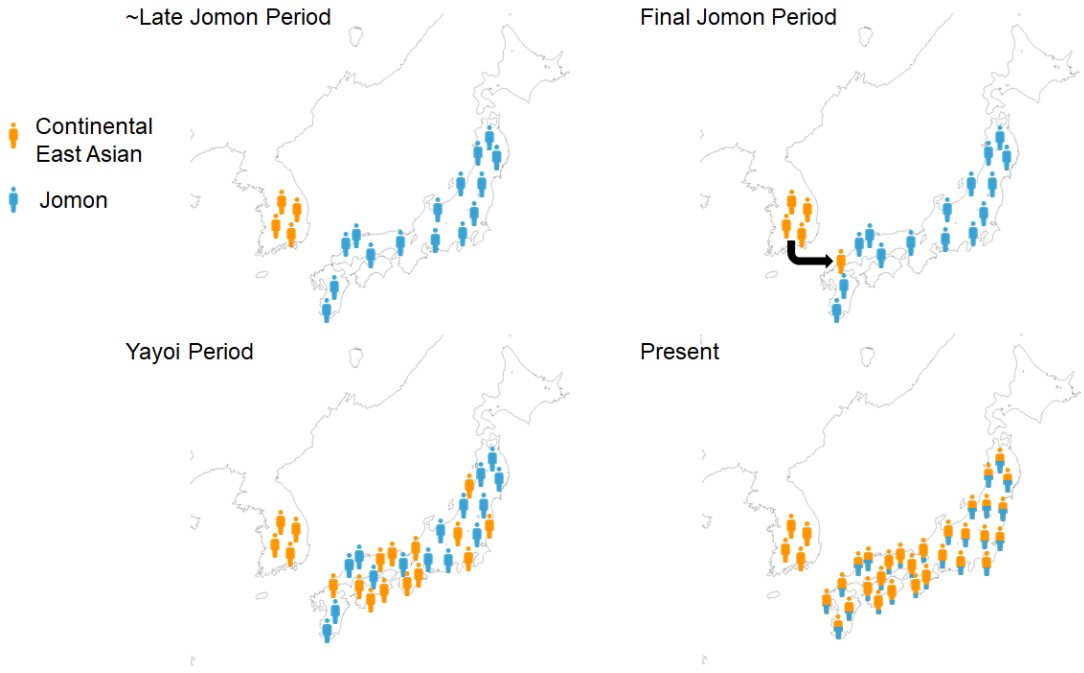


Fig. 5 Haplotype structures composed of Jomon-derived SNPs in four genes associated with the characteristic phenotypes of East Asians. The haplotype structures surrounding four nonsynonymous SNPs, (a) rs17822931 in *ABCC11* (associated with ear wax type), (b) rs3827760 in *EDAR* (associated with hair thickness and shovel-shaped incisors), (c) rs671 in *ALDH2* (associated with alcohol tolerance), and (d) rs1229984 in *ADH1B* (associated with alcohol tolerance) are illustrated. Each horizontal line represents each haplotype, and each vertical line represents each of the Jomon-derived SNP or phenotype-associated SNP. The derived alleles of the SNPs associated with phenotypes are shown in orange, and the ancestral alleles are shown in blue. The red color represents the Jomon allele, i.e., horizontal lines containing the red colored grid indicate the Jomon-derived haplotypes.



925

926

927 **Fig. 6 The formation process of the Japanese population from the Late**
928 **Jomon period to the present.** From the Late to the Final Jomon period, the
929 Jomon people settled down in Mainland Japan, and the population size varied
930 between regions. In the Final Jomon period, the continental East Asians arrived
931 in northern Kyushu and then admixed with Jomon people in all the regions of
932 Mainland Japan. In regions such as Kinki and Shikoku, where the population size
933 was smaller at the end of the Jomon period, modern Japanese have lower
934 degrees of genome components derived from the Jomon people.