

ORIGINAL ARTICLE

Unique characteristics of the Ainu population in Northern Japan

Timothy A Jinam^{1,2,3}, Hideaki Kanzawa-Kiriyama^{2,4}, Ituro Inoue^{2,3}, Katsushi Tokunaga⁵, Keiichi Omoto⁶ and Naruya Saitou^{1,2,7}

Various genetic data (classic markers, mitochondrial DNAs, Y chromosomes and genome-wide single-nucleotide polymorphisms (SNPs)) have confirmed the coexistence of three major human populations on the Japanese Archipelago: Ainu in Hokkaido, Ryukyuan in the Southern Islands and Mainland Japanese. We compared genome-wide SNP data of the Ainu, Ryukyuan and Mainland Japanese, and found the following results: (1) the Ainu are genetically different from Mainland Japanese living in Tohoku, the northern part of Honshu Island; (2) using Ainu as descendants of the Jomon people and continental Asians (Han Chinese, Koreans) as descendants of Yayoi people, the proportion of Jomon genetic component in Mainland Japanese was ~18% and ~28% in Ryukyuan; (3) the time since admixture for Mainland Japanese ranged from 55 to 58 generations ago, and 43 to 44 generations ago for the Ryukyuan, depending on the number of Ainu individuals with varying rates of recent admixture with Mainland Japanese; (4) estimated haplotypes of some Ainu individuals suggested relatively long-term admixture with Mainland Japanese; and (5) highly differentiated genomic regions between Ainu and Mainland Japanese included *EDAR* and *COL7A1* gene regions, which were shown to influence macroscopic phenotypes. These results clearly demonstrate the unique status of the Ainu and Ryukyuan people within East Asia.

Journal of Human Genetics advance online publication, 16 July 2015; doi:10.1038/jhg.2015.79

INTRODUCTION

The Japanese Archipelago consists of four major islands (Hokkaido, Honshu, Shikoku and Kyushu) and many other small islands that can be grouped into nine regions (Supplementary Figure 1). Frequent waves of human migrations from the Eurasian continent to the archipelago took place from at least 30 000 years ago (YBP).¹ There were various migration routes to the archipelago.² These migrations have shaped the human population structure in the Japanese Archipelago, where there are currently three main populations: the Ainu who mainly live in Hokkaido at the northernmost island of the Archipelago; the Ryukyuan who mainly live in the Ryukyu Islands at the southern part; and the Mainland Japanese whose population size is the largest and who live in all major four islands and small islands.

From an archeological perspective, the prehistory of the Japanese Archipelago can be divided into the Paleolithic period (older than 16 000 YBP), the Jomon period (16 000–3000 YBP) and the Yayoi period (3000–1700 YBP).¹ The currently accepted model regarding the origin of Japanese populations is the dual-structure model,³ whereby the current Japanese population is the result of admixture between the early migrants (Jomon people) and later migrants (Yayoi people) and that the Ainu and the Ryukyuan are thought to retain more Jomon components than the Mainland Japanese. Subsequent studies using

mitochondrial DNA and several autosomal markers have been in general agreement with the dual-structure model, showing the admixed nature of Mainland Japanese^{4,5} and demonstrating close affinities between the Ainu and Ryukyuan populations.^{6,7}

The Japanese Archipelago Human Population Genetics Consortium⁸ produced ~900 000 genome-wide single-nucleotide polymorphism (SNP) data in the Ainu and the Ryukyuan, and through principal component analysis (PCA), and phylogenetic tree construction, demonstrated a clear genetic similarity between these two groups despite their current geographical locations at the opposite poles of the Japanese archipelago. Analysis of individual ancestry proportions and phylogenetic analysis of the Mainland Japanese also show that they carry both Ainu-Ryukyuan and continental Asian genetic components. A recent study that used a model-based approach was also in favor of the dual-structure model.⁹

Although previous studies generally support the dual-structure model, some details regarding the amount of genetic contributions from the ancestral populations were not really well defined. We also wanted to identify what kind of factors that contributed to the genetic uniqueness that was previously observed in the Ainu.⁸ Therefore, the aims of this study are to perform a test for admixture and to clarify the timing and admixture proportions in the Japanese populations.

¹Division of Population Genetics, National Institute of Genetics, Mishima, Japan; ²Department of Genetics, School of Life Science, Graduate University for Advanced Studies (SOKENDAI), Mishima, Japan; ³Division of Human Genetics, National Institute of Genetics, Mishima, Japan; ⁴Department of Anthropology, National Museum of Nature and Science, Tsukuba, Japan; ⁵Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; ⁶Department of Anthropology, Faculty of Science, The University of Tokyo, Tokyo, Japan and ⁷Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Tokyo, Japan
Correspondence: Professor N Saitou, Division of Population Genetics, National Institute of Genetics, 1111 Yata, Mishima 411-0831, Japan.
E-mail: saitounr@nig.ac.jp

Table 1 Populations used for ancestry estimation tests

Population	Geographical location	<i>n</i>	Reference	No. of SNP loci
Ainu	Hokkaido, Japan	36	Japanese Archipelago Population Genetics Consortium ⁸	641 314
Ryukyuan	Okinawa, Japan	30	Japanese Archipelago Population Genetics Consortium ⁸	
Mainland Japanese	Kanto, Japan	50	Nishida <i>et al.</i> (2008) ¹⁰	
Korean	Korea	50	Bae <i>et al.</i> ¹³	317 054
Han Chinese (CHB)	Beijing, China	42	The International HapMap Consortium ¹¹	906 600
European (CEU)	USA	50	The International HapMap Consortium ¹¹	
Yoruban (YRI)	Nigeria, Africa	50	The International HapMap Consortium ¹¹	
Malay	Singapore	50	Teo <i>et al.</i> ¹²	570 408
Indian	Singapore	50	Teo <i>et al.</i> ¹²	
Han Chinese (Chinese-Sg)	Singapore	50	Teo <i>et al.</i> ¹²	

Abbreviation: SNP, single-nucleotide polymorphism.

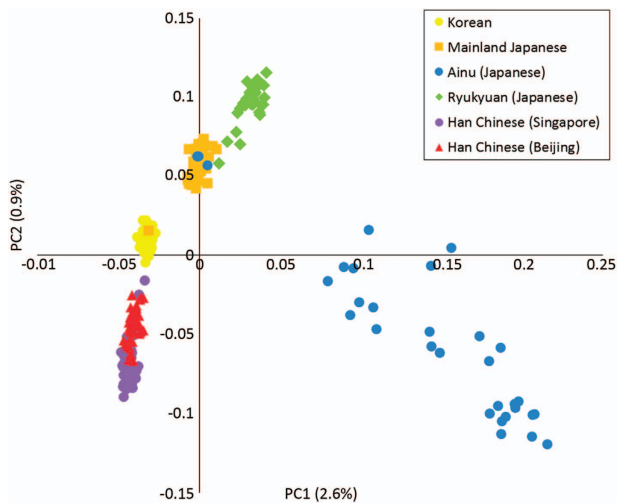


Figure 1 Principal component analysis (PCA) plot after omitting closely related Ainu individuals.

In addition, we wish to identify highly diverged genetic loci between the Ainu and Mainland Japanese.

MATERIALS AND METHODS

Sample data

We used 641 314 genome-wide SNP data from the Ainu and Ryukyans originally published by the Japanese Archipelago Human Population Genetics Consortium⁸ as well as Mainland Japanese from Nishida *et al.*¹⁰ We merged the data with those from three HapMap populations¹¹ and three Singaporean populations,¹² resulting in 431 486 overlapping SNPs. We further merged these data with genome-wide SNP data for 50 randomly sampled Korean individuals from Seoul who were originally used as a control group in a genome-wide association study.¹³ With the inclusion of the Korean data, the number of overlapping SNPs was reduced to 65 256. The list of populations used in this study is shown in Table 1. All of the SNP data used were already available in the published literature.

Data analysis

Given the rather close-knit nature of the Ainu community, we investigated the possibility that closely related individuals may be included in the data set. We estimated measures of kinship coefficients and identity-by-descent between all pairs of Ainu individuals using REAP software,¹⁴ which can be used in populations with admixed ancestry. It was shown previously that the Ainu have experienced admixture with Mainland Japanese,⁸ and we used individual ancestry information at $k=2$ from ADMIXTURE analysis¹⁵ (Supplementary Figure 2) as part of input for the software.

The relationship between Ainu and other populations from various geographical locations of the Japanese archipelago was estimated using Population Structure Prediction System for Japanese (PCAj).¹⁶ Based on the probabilistic PCA, our genotype data were used to project individuals onto a scatterplot similar to that shown by Yamaguchi-Kabata *et al.*¹⁷ The Japanese samples included in this software are part of the RIKEN Biobank collection.

To formally test whether Japanese populations are the result of admixture between ancestral Jomon and Yayoi populations, we performed the 3-population test (f_3) using the Ainu as surrogates of Jomon ancestors and continental Asians (Han Chinese, Koreans) as surrogates of Yayoi ancestors. We also performed the f_4 -ratio test to estimate the genetic contributions of source populations in the admixed populations and estimated the time since admixture occurred using rolloff. These three tests are included in the ADMIXTOOLS software package.¹⁸

To test whether the Ainu individuals that lie in intermediate positions between Ainu and Mainland Japanese clusters in the PCA plot⁸ are recently admixed individuals, we first phased haplotypes in 20 Ainu individuals, 20 Mainland Japanese individuals and 8 potentially recently admixed Ainu individuals using FastPhase¹⁹ and BEAGLE²⁰ programs. The eight possibly admixed Ainu individuals were phased together with the 20 Ainu and 20 Mainland Japanese separately (Supplementary Figure 3). Pairwise distances between all phased haplotypes were calculated to generate a distance matrix that was used to construct a neighbor-joining tree²¹ to assess the affinity of haplotypes from potentially admixed Ainu individuals.

To identify genomic regions that are highly differentiated in the Ainu, we calculated pairwise F_{st} ²² between Ainu and Mainland Japanese after omitting potentially recently admixed individuals from each population. We focused on the top 1% of highly differentiated SNPs and performed gene annotation search using GOrilla gene ontology tool²³ to find out whether these SNPs have any significant biological functions.

RESULTS

We identified five parent-offspring and two sibling pairs in the Ainu based on the values of kinship coefficient and probability of identity-by-descent=0 (Supplementary Figure 4). We therefore omitted one individual from each of the parent-offspring pairs, and used the remaining 31 individuals for PCA. The result of the new PCA using approximately 65 k SNPs is shown in Figure 1. The first principal component (PC1) separates the Ainu from the rest of East Asian populations, and the population closest to the Ainu are Ryukyans, consistent with the previous observations. The Ainu individuals are spread out in the same 'comet-like' pattern as before,⁸ but no outlier Ainu individuals were distinguished from PC2. When the Korean data set was omitted, the resulting PCA using about 430k SNPs also showed a similar pattern (Supplementary Figure 5). We therefore surmise that the five outlier Ainu individuals seen in the previous PCA plot⁸ represent an artifact due to the inclusion of very closely related individuals within the Ainu population.

The evolutionary factors that created variations explained by PC2 are difficult to conjecture. If we disregard Ryukyuan individuals, then the PC2 axis from top to bottom seems to reflect a south to north geographical cline in populations: Singapore Han Chinese, Beijing Han Chinese (CHB), Koreans and Mainland Japanese of the Japanese Archipelago. However, Ryukyuan individuals are located on the top

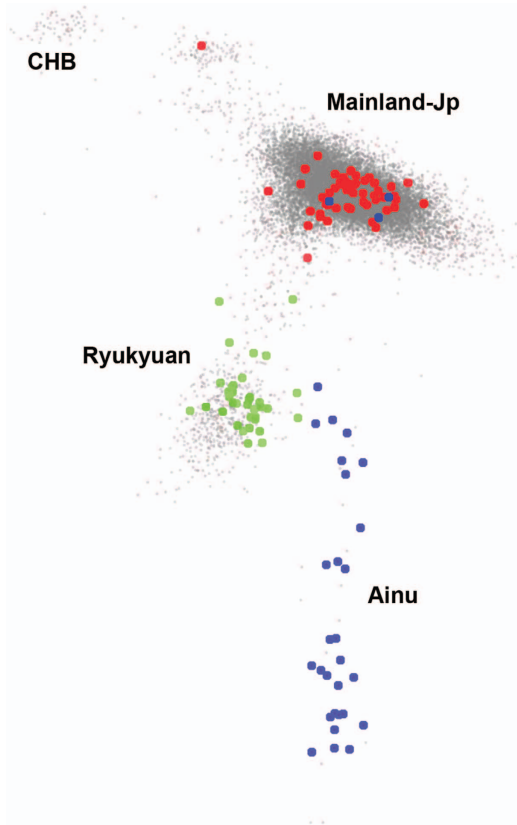


Figure 2 Scatterplot output from PCAj software. The three populations from Japanese Archipelago Human Population Genetics Consortium⁸ are shown in colored dots. Other Japanese and Chinese individuals are shown as gray dots.

part on the PC2 axis, above the Mainland Japanese of the Japanese Archipelago. A possible explanation would be some unknown populations were involved in the formation of Ryukyuan, as previously suggested.²⁴

We identified eight Ainu individuals who might be recently admixed with Mainland Japanese based on their intermediate positions between Ainu and Mainland Japanese clusters in the PCA plot (Supplementary Figure 3). If these individuals were the result of very recent admixture events, then one of the pair of chromosomes should be from an Ainu parent and the other from a Mainland Japanese parent. The neighbor joining tree of chromosome 22 haplotypes that were phased using Fastphase and BEAGLE for each of the eight admixed Ainu individuals is shown in Supplementary Figure 6. The haplotype affinities for these potentially recently admixed Ainu did not show a consistent pattern, with three individuals having haplotype affinities with the Ainu and another four individuals with affinities to Mainland Japanese haplotypes. However, in one individual (labeled 2120001B03), one of the haplotypes clustered with other Ainu haplotypes while the other clustered with Mainland Japanese haplotypes, indicating that this person might be a result of recent admixture.

The 31 Ainu individuals together with 35 Ryukyuan and 50 randomly chosen Mainland Japanese from the Kanto region were compared with other individuals from various geographical locations in the Japanese archipelago using probabilistic PCA.¹⁶ Our samples were overlaid on a scatterplot that showed major clustering between Mainland Japanese, CHB and Ryukyuan. Our Ryukyuan (green dots) and Mainland Japanese samples from the Kanto region (red dots) fall within well-defined clusters reported by Yamaguchi-Kabata *et al.*¹⁷ as seen in Figure 2. Interestingly, the Ainu individuals (blue dots) form a gradient alongside the cluster of Ryukyuan individuals on the vertical axis. There are several individuals from the RIKEN data set (identified as gray dots) that cluster with our Ainu samples. Although their specific ethnicity was not directly mentioned, they are most likely Ainu individuals based on their geographical origin, which is Hokkaido where most Ainu people currently reside. This identification of Ainu people in the RIKEN data was not reported by Yamaguchi-Kabata *et al.*¹⁷ nor Kumasaka *et al.*¹⁶

We tested the dual-structure model for the origin of modern Japanese using the 3-population test (f_3 test). We used two data sets

Table 2 Results of 3-population admixture test

Source population 1	Target population	Source population 2	f_3	Standard error	Z-score
<i>Using ~400 k SNP data set</i>					
Ainu	Mainland Japanese	CHB	-0.0070	0.000242	-28.8
Ainu	Mainland Japanese	Chinese-Sg	-0.0068	0.000258	-26.5
Ainu	Ryukyuan	CHB	-0.0089	0.000359	-24.9
Ainu	Ryukyuan	Chinese-Sg	-0.0091	0.000376	-24.1
Ainu	Ryukyuan	Mainland Japanese	-0.0046	0.000278	-16.5
<i>Using ~60 k SNP with the inclusion of Korean data</i>					
Ainu	Mainland Japanese	Korean	-0.0071	0.000248	-28.6
Ainu	Mainland Japanese	CHB	-0.0066	0.000271	-24.2
Ainu	Mainland Japanese	Chinese-Sg	-0.0064	0.000283	-22.8
Ainu	Ryukyuan	Korean	-0.0089	0.000363	-24.7
Ainu	Ryukyuan	CHB	-0.0084	0.000367	-22.9
Ainu	Ryukyuan	Chinese-Sg	-0.0086	0.000377	-22.7
Ainu	Ryukyuan	Mainland Japanese	-0.0043	0.000296	-14.7

Abbreviations: CHB, Beijing Han Chinese; SNP, single-nucleotide polymorphism.

for this test: a high density SNP data set without Korean data (~430 k SNPs) and a low density data set with Korean data (~65 k SNPs). Negative values for the f_3 test imply that the target population is the result of admixture between two source populations. The Z -score for the test can be taken as a measure of statistical significance. The combination using continental Asians (Han Chinese, Koreans) and Ainu as source populations and the Mainland Japanese and Ryukyuan as target populations showed the most significant results (Table 2). Using the high density SNP data set, the (Mainland Japanese; CHB, Ainu) combination gave the most significant result ($f_3 = -7.0 \times 10^{-3}$; Z -score = -28.8), whereas in the low density SNP data set, the combination (Mainland Japanese; Korean, Ainu) had the most significant result ($f_3 = -7.1 \times 10^{-3}$; Z -score = -28.6). Tests using Ryukyuan as target populations also showed similar patterns, but with lower Z -scores compared with Mainland Japanese. These results showed that the Mainland Japanese were the result of admixture between the ancestors of Han Chinese/Koreans and Ainu, who represent the Yayoi and Jomon peoples, respectively. This adds further support to the dual-structure model³ for the origin of Mainland Japanese.

To estimate the proportion of genetic contributions from the ancestral populations in the Japanese, we conducted the f_4 -ratio estimation test. Populations used in this test are assumed to be related to each other according to the tree shown in Figure 3. This test differs from the f_3 test in the inclusion of an outgroup, and another population that is close to one of the source populations (continental

Asians). Using various combinations of populations, the results with the most significant Z -scores are shown in Table 3. The proportion of the Jomon ancestry (ancestors of Ainu) in Mainland Japanese was estimated to be 17.8% (Z -score 72.3) when using the CHB as the other source population. When using the smaller data set with Koreans as the source population, the proportion of the Jomon ancestry in Mainland Japanese was higher at 17.9% (Z -score 64.9). The proportion of Jomon components in the Ryukyuan was from 28.4% (Z -score 43.8) when using high density SNP data and 27.8% (Z -score 40.0) when using the lower density SNP data. The higher proportion of Jomon component in the Ryukyuan compared with the Mainland Japanese was consistent with the individual ancestry estimates using ADMIXTURE (Supplementary Figure 2).

We further estimated the time since the admixture event between the Yayoi and Jomon ancestors that resulted in the Japanese populations using the rolloff program.¹⁸ Due to the presence of admixed individuals within Ainu, we decided to create three subsets of Ainu individuals with different levels of admixture (Supplementary Figure 7) to gauge the performance of the rolloff program. Because this test is based on the decay of linkage disequilibrium over time, we used the higher density SNP data set (~430 k SNP). Thus, we took CHB and the three Ainu subsets as ancestral populations, and the Mainland Japanese and Ryukyuan as admixed populations. The results in Table 4 show that the Ainu data set with the least number of admixed individuals (Ainu-15) yielded a much older time since

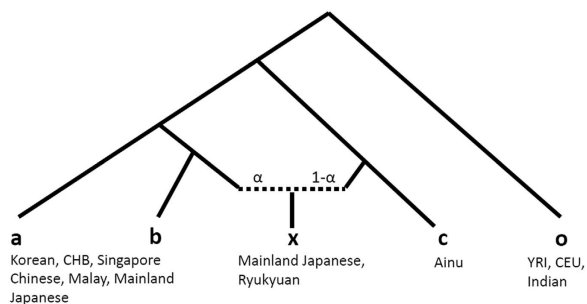


Figure 3 Schematic diagram showing the relationship between populations used for the estimation of Jomon ($1 - \alpha$) and Yayoi (α) proportions in the Mainland Japanese.

Table 4 Estimation of time since admixture for Mainland Japanese and Ryukyuan, using Han Chinese (CHB) and Ainu as source populations

Ainu dataset	Ainu ancestry (admixture results at $k=2$) (%)	Time since admixture in generations (years) ^a	
		Mainland Japanese	Ryukyuan
Ainu-15	98.2	58 (1450)	44 (1100)
Ainu-20	92.9	56 (1400)	43 (1075)
Ainu-28	82.5	55 (1375)	43 (1075)

Three Ainu data sets with different proportions of admixture were used.
^aA generation time of 25 years was used.

Table 3 Proportion of Jomon ancestry ($1 - \alpha$) estimated from the f_4 -ratio test

a	b	x	c	o	($1 - \alpha$)	Standard error	Z-score
<i>Using ~400 k SNP data set</i>							
Chinese-Sg	CHB	Mainland Japanese	Ainu	Indian	0.1780	0.0114	72.3
CHB	Chinese-Sg	Mainland Japanese	Ainu	Indian	0.1363	0.0128	67.3
CHB	Mainland Japanese	Ryukyuan	Ainu	Indian	0.2845	0.0164	43.8
Chinese-Sg	Mainland Japanese	Ryukyuan	Ainu	Indian	0.2717	0.0167	43.7
<i>Using ~60 k SNP with the inclusion of Korean data</i>							
CHB	Korean	Mainland Japanese	Ainu	Indian	0.1794	0.0126	64.9
Chinese-Sg	Korean	Mainland Japanese	Ainu	Indian	0.1744	0.0129	63.8
CHB	Chinese-Sg	Mainland Japanese	Ainu	Indian	0.1392	0.0147	58.7
Chinese-Sg	CHB	Mainland Japanese	Ainu	Indian	0.1960	0.0138	58.2
Korean	Mainland Japanese	Ryukyuan	Ainu	Indian	0.2788	0.0180	40.0
Chinese-Sg	Mainland Japanese	Ryukyuan	Ainu	Indian	0.2744	0.0190	38.2
CHB	Mainland Japanese	Ryukyuan	Ainu	Indian	0.2945	0.0190	37.2

Abbreviations: CHB, Beijing Han Chinese; SNP, single-nucleotide polymorphism. Populations labels (a, b, x, c, o) are ordered as in Figure 3.

admixture (58 generations ago or 1450 years ago, assuming a generation time of 25 years) when using Mainland Japanese as the admixed population. With the inclusion of more admixed individuals in the Ainu data set, the time since admixture became more recent (55 generations ago or 1375 years ago). The estimates using Ryukyuan as the admixed population ranged from 44 to 43 generations ago (1100–1075 years ago), which are relatively more recent compared with those for the Mainland Japanese.

We also identified SNP loci that are differentiated between the Ainu and Mainland Japanese by using pairwise F_{st} values. The pairwise F_{st} values ranged from 0 to 0.8903, with a mean of 0.0407. The majority of the SNPs (approximately 400 000) have F_{st} values of less than 0.02. We picked 6413 SNPs that were within the top 1% and had F_{st} values higher than 0.36. The F_{st} values and annotations for these top 1% SNP are listed in Supplementary Table 1. Within those top 1% of SNP, some of them were found in genes reported to be associated with facial structure in Europeans²⁵ and hair and tooth morphology in East Asians.^{26,27} The distribution of F_{st} values for SNPs in those genes are shown in Figure 4. Two out of five genes for facial morphology (*PAX3* and *COL7A1*) contain highly differentiated SNPs, as with the hair/tooth morphology gene (*EDAR*). The results of gene annotation analysis on those top 1% SNPs showed enrichment for biological processes and cellular components involving collagen (Supplementary Figure 8).

DISCUSSION

We identified several very closely related pairs of individuals within the Ainu, which was not reported previously.⁸ The inclusion of closely related individuals may produce artifactual results for individual-based clustering tests such as PCA or STRUCTURE. Although the PCA plot after omission of closely related Ainu individuals is slightly different from the previous result,⁸ the overall pattern remains unchanged. The gradient of Ainu individuals along PC1 can be explained partially by recent admixture between Ainu and Mainland Japanese parents and also intermarriage between individuals with different proportions of Ainu and Mainland Japanese ancestry.

We also had the opportunity to compare our Ainu data with those from other Japanese living in different locations in the Japanese Archipelago that was represented in the RIKEN Biobank data.¹⁶ The Ryukyuan and Mainland Japanese from Kanto region (Supplementary Figure 1) in our data overlap with the RIKEN data. Ainu individuals still form the gradient, but are positioned adjacent to the cluster of Ryukyuan populations (Figure 2). The three Ainu individuals appearing in the Mainland Japanese cluster were also observed with the RIKEN data. Although there was no major discrepancy between our data and the RIKEN data, the position of the Ainu in Figure 2 is slightly different from Figure 1. This may reflect biases in sample sizes between the two data sets. In the RIKEN data, the majority of the individuals (approximately 1000) are from

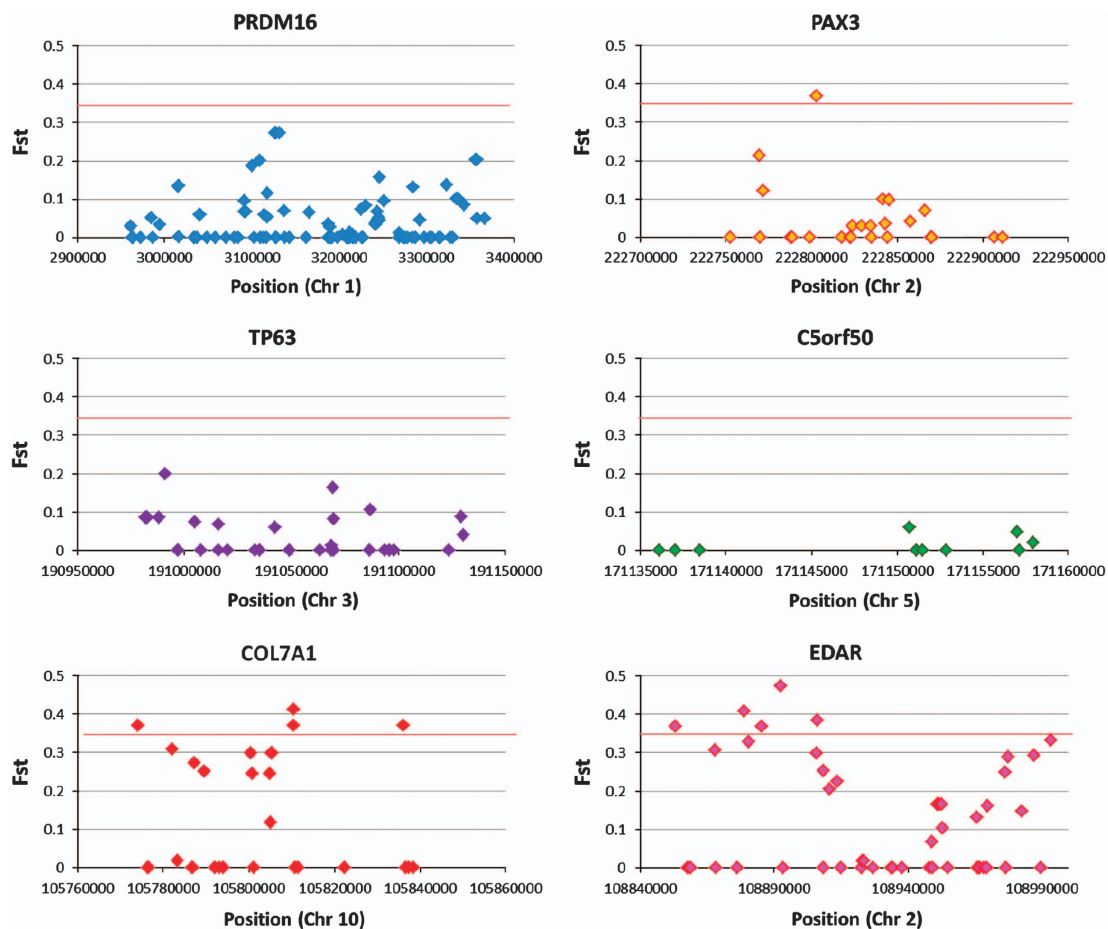


Figure 4 Distribution of differentiated SNP in the Ainu for the five genes associated with facial morphology in Europeans (Liu *et al.*²⁵) and for hair and tooth morphology (*EDAR*). The top 1% highly differentiated SNPs have F_{st} values above 0.36, represented by the red horizontal lines.

the Mainland Japanese cluster, with only a few individuals placed within our Ainu cluster.

A clear difference between the northern Tohoku region Mainland Japanese and Ainu is another important finding. Yamaguchi-Kabata *et al.*¹⁷ showed that people living in the Tohoku region are distributed slightly downward in the Mainland Japanese cluster, which overlaps with Kanto region Mainland Japanese (red color) shown in Figure 2. Historically, people in northern Tohoku area were called 'Emishi', while the Ainu people were called 'Ezo', using the same Chinese characters. It is still not clear whether the Emishi and Ezo people were genetically similar. Some cultural components such as locale names and potteries are shared between the Emishi in Tohoku region and the Ezo in Hokkaido who eventually evolved into the Ainu people.²⁸ However, the PCA shown in Figure 2 suggests that some unknown factor(s) may exist in the formation of Emishi people in the Tohoku region.

Japanese Archipelago Human Population Genetics Consortium⁸ presented a figure that represents the possible evolutionary history of populations in the Japanese Archipelago. Here we did a formal test using the 3-population, f_4 ratio estimation and rolloff tests that has been previously used in other populations.²⁹ In particular, we wanted to test the plausibility of the dual-structure model for the origin of current Japanese using the Ainu and continental Asians as ancestral populations. The results of the 3-population test show that the Mainland Japanese are indeed a result of admixture between the ancestors of the Ainu (Jomon people) and continental Asians (Yayoi People), with an estimated 18% contribution from the Jomon ancestry. Results using the Ryukyuan also indicate that they are admixed, although the Z -scores were less significant when using Mainland Japanese (Tables 2 and 3). The estimated Jomon component was higher in the Ryukyuan, at approximately 28%. This was expected, given the low but significant affinity between the Ainu and the Ryukyuan. Our Jomon ancestry estimates were lower than what was reported previously, which ranged from 30 to 40% in the Mainland Japanese and 50–60% in the Ryukyuan.^{30,31} In this study, we used high density autosomal SNP data in the Ainu, whereas previous studies used only Y-chromosome polymorphisms,³⁰ or using non-Ainu as the ancestral population.³¹ These differences in addition to other methodological parameters may lead to the differences in the estimates.

The time since admixture estimated using rolloff varied from 55 to 58 generations ago, depending on the number of admixed Ainu included in the data set. If we had more Ainu individuals who are less admixed with Mainland Japanese, then the estimate could be even older. Our estimated time since admixture of 1450 years ago may be treated as a lower bound, since rolloff assumes only one admixture event and estimates the most recent event. This time estimate coincides with the period, during 5th–8th century, in which the Japanese government expanded their territory to the northern part (Tohoku) of the Honshu Island.²⁸ Interestingly, the time since the admixture event that led to the formation of Ryukyuan was more recent compared with the Mainland Japanese. The estimated admixture time of approximately 1100 years ago corresponds to the Gusuku period of Okinawan archeological history.³² A plausible explanation may involve the following scenario: the Jomon people who already settled in Ryukyu Islands experienced admixture with migrants from the continent who themselves may have already admixed with the Jomon people in the Japanese Mainland.

We also report several SNP loci that are highly differentiated between the Ainu and the Mainland Japanese. These include two genes associated with facial structure in Europeans.²⁵ SNPs within the *EDAR* gene have been reported to be highly differentiated between Mainland

Japanese and Ryukyuan.¹⁷ Similar patterns were observed when comparing between Ainu and Mainland Japanese. However, the non-synonymous SNP rs3827760 reported by Yamaguchi-Kabata *et al.*¹⁷ was not covered on the Affymetrix 6.0 genechip. They also reported a SNP within the *ABCC11* gene that was responsible for earwax type was highly differentiated between the Ryukyuan and the Mainland Japanese. However, possibly due to poor coverage of the *ABCC11* gene in the Affymetrix genechip, all SNPs in that gene were found to be monomorphic between the Ainu and the Mainland Japanese. It is known that the Ainu have very different physical characteristics compared with Mainland Japanese, and the finding of differentiated loci associated with observable phenotypes agrees with the works done by physical anthropologists over the years.³³

To conclude, we supplemented our initial analysis on the Ainu population, and demonstrated that their ancestors contributed significantly to the genetic makeup of the current populations on the Japanese Archipelago. If we consider a very unique genetic status of Jomon people, then it is understandable that Ainu people, who inherited the highest proportion of Jomon DNA, are quite unique among all extant East Eurasians.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank Japanese Archipelago Human Population Genetics Consortium and Asian DNA Repository Consortium for permission of access to SNP data. This study was supported by a SOKENDAI Grant on modern human dispersal. Finally, we appreciate people on the Japanese Archipelago (Ainu, Ryukyuan and Mainland Japanese) who kindly donated their DNA to this and past anthropological studies.

- 1 Imamura, K. *Prehistoric Japan: New Perspectives on Insular East Asia* (University of Hawaii Press, Honolulu, HI, USA, 1996).
- 2 Saitou, N. in *Simulations, Genetics and Human Prehistory* (eds Matsumura S., Forster P. & Renfrew C.) 89–92 (McDonald Institute Monographs, Cambridge, UK, 2008).
- 3 Hanihara, K. Dual structure model for the population history of the Japanese. *Japan Rev.* **2**, 1–33 (1991).
- 4 Tanaka, M., Cabrera, V. M., González, A. M., Larruga, J. M., Takeyasu, T., Fukui, N. *et al.* Mitochondrial genome variation in eastern Asia and the peopling of Japan. *Genome Res.* **14**, 1832–1850 (2004).
- 5 Nakaoka, H., Mitsunaga, S., Hosomichi, K., Shyh-yuh, L., Sawamoto, T., Fujiwara, T. *et al.* Detection of ancestry informative HLA alleles confirms the admixed origins of Japanese population. *PLoS ONE* **8**, e60793 (2013).
- 6 Omoto, K. & Saitou, N. Genetic origins of the Japanese: a partial support for the dual structure hypothesis. *Am J Phys Anthropol.* **102**, 437–446 (1997).
- 7 Bannai, M., Ohashi, J., Harihara, S., Takahashi, Y., Juji, T., Omoto, K. *et al.* Analysis of HLA genes and haplotypes in Ainu (from Hokkaido, northern Japan) supports the premise that they descent from Upper Paleolithic populations of East Asia. *Tissue Antigens* **55**, 128–139 (2000).
- 8 Japanese Archipelago Human Population Genetics Consortium. [Jinam, T., Nishida, N., Hirai, M., Kawamura, S., 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.* **57**, 787–795 (2012).
- 9 Nakagome, S., Sato, T., Ishida, H., Hanihara, T., Yamaguchi, T., Kimura, R. *et al.* Model-based verification of hypotheses on the origin of modern Japanese revisited by Bayesian inference based on genome-wide SNP data. *Mol. Biol. Evol.* **32**, 1533–1543 (2015).
- 10 Nishida, N., Koike, A., Tajima, A., Ogasawara, Y., Ishibashi, Y., Uehara, Y., *et al.* Evaluating the performance of Affymetrix SNP Array 6.0 platform with 400 Japanese individuals. *BMC Genomics* **9**, 431–431 (2008).
- 11 The International HapMap Consortium. A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).
- 12 Teo, Y. Y., Sim, X., Ong, R. T., Tan, A. K., Chen, J., Tantoso, E. *et al.* Singapore Genome Variation Project: a haplotype map of three Southeast Asian populations. *Genome Res.* **19**, 2154–2162 (2009).
- 13 Bae, J. S., Cheong, H. S., Kim, J. H., Park, B. L., Kim, J. H., Park, T. J. *et al.* The genetic effect of copy number variations on the risk of type 2 diabetes in a Korean population. *PLoS ONE* **6**, e19091 (2011).

- 14 Thornton, T., Tang, H., Hoffman, T. J., Ochs-Balcom, H. M., Baan, B. J. & Risch, N. Estimating kinship in admixed populations. *Am. J. Hum. Genet.* **91**, 122–138 (2012).
- 15 Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
- 16 Kumasaka, N., Yamaguchi-Kabata, Y., Takahashi, A., Kubo, M., Nakamura, Y. & Kamatani, N. Establishment of a standardized system to perform population structure analyses with limited sample size or with different sets of SNP genotypes. *J. Hum. Genet.* **55**, 525–533 (2010).
- 17 Yamaguchi-Kabata, Y., Nakazono, K., Takahashi, A., Saito, S., Hosono, N., Kubo, M. *et al*. Population structure of Japanese based on SNP genotypes from 7,001 individuals in comparison to other ethnic groups: Effects on population-based association studies. *Am. J. Hum. Genet.* **83**, 445–456 (2008).
- 18 Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y. *et al*. Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
- 19 Scheet, P. & Stephens, M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am. J. Hum. Genet.* **78**, 629–644 (2006).
- 20 Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing data inference for whole genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81**, 1084–1097 (2007).
- 21 Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425 (1987).
- 22 Weir, B. S. & Cockerham, C. C. Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370 (1984).
- 23 Eden, E., Navon, R., Steinfeld, I., Lipson, D. & Yakhini, Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**, 48 (2009).
- 24 Sato, T., Nakagome, S., Watanabe, C., Yamaguchi, K., Kawaguchi, A., Koganebuchi, K. *et al*. Genome-Wide SNP analysis reveals population structure and demographic history of the Ryukyu Islanders in the Southern part of the Japanese Archipelago. *Mol. Biol. Evol.* **31**, 2929–2940 (2014).
- 25 Liu, F., van der Lijn, F., Schurmann, C., Zhu, G., Chakravarty, M. M., Hysi, P. G. *et al*. A genome-wide association study identifies five loci influencing facial morphology in Europeans. *PLoS Genet.* **8**, e1002932 (2012).
- 26 Fujimoto, A., Kimura, R., Ohashi, J., Omi, K., Yuliwulandari, R., Batubara, L. *et al*. A scan for genetic determinants of human hair morphology: EDAR is associated with Asian hair thickness. *Hum. Mol. Genet.* **17**, 835–843 (2008).
- 27 Kimura, R., Yamaguchi, T., Takeda, M., Kondo, O., Toma, T., Haneji, K. *et al*. A common variation in EDAR is a genetic determinant of shovel-shaped incisors. *Am. J. Hum. Genet.* **85**, 528–535 (2009).
- 28 Kudo, M. *Ancient History of Emishi*, (Heibon-sha, Tokyo, Japan, 2001) [in Japanese].
- 29 Moorjani, P., Thangaraj, K., Patterson, N., Lipson, M., Loh, P. R., Govindaraj, P. *et al*. Genetic evidence for recent population mixture in India. *Am. J. Hum. Genet.* **93**, 422–438 (2013).
- 30 Rasteiro, R. & Chikhi, L. Revisiting the peopling of Japan: an admixture perspective. *J. Hum. Genet.* **54**, 349–354 (2009).
- 31 He, Y., Wang, W. R. & Xu, S. The HUGO Pan-Asian SNP Consortium. Paleolithic contingent in modern Japanese: estimation and inference using genome-wide data. *Sci. Rep.* **2**, 1–7 (2012).
- 32 Tomiyama, K. *World of Ryukyu and Okinawa History*, Yoshikawa Kobunkan, Tokyo (2003) [in Japanese].
- 33 Ossenberg, N. S., Dodo, Y., Maeda, T. & Kawakubo, Y. Ethnogenesis and craniofacial change in Japan from the perspective of nonmetric traits. *Anthropol. Sci.* **114**, 99–115 (2006).

Supplementary Information accompanies the paper on Journal of Human Genetics website (<http://www.nature.com/jhg>)