

Genome-Wide SNP Analysis Reveals Population Structure and Demographic History of the Ryukyu Islanders in the Southern Part of the Japanese Archipelago

Takehiro Sato,¹ Shigeki Nakagome,² Chiaki Watanabe,¹ Kyoko Yamaguchi,¹ Akira Kawaguchi,¹ Kae Koganebuchi,^{3,4} Kuniaki Haneji,⁵ Tetsutaro Yamaguchi,⁶ Tsunehiko Hanihara,⁷ Ken Yamamoto,^{8,9} Hajime Ishida,¹ Shuhei Mano,² Ryosuke Kimura,^{*1} and Hiroki Oota^{*7}

¹Department of Human Biology and Anatomy, Graduate School of Medicine, University of the Ryukyus, Okinawa, Japan

²Department of Mathematical Analysis and Statistical Inference, The Institute of Statistical Mathematics, Tokyo, Japan

³Department of Integrated Biosciences, Graduate School of Frontier Sciences, The University of Tokyo, Chiba, Japan

⁴Department of Human Anatomy, Kitasato University Graduate School of Medical Sciences, Kanagawa, Japan

⁵Haneji Oral Surgery Clinic, Medical Corporation Hayamakai, Okinawa, Japan

⁶Department of Orthodontics, School of Dentistry, Showa University, Tokyo, Japan

⁷Department of Anatomy, Kitasato University School of Medicine, Kanagawa, Japan

⁸Division of Genome Analysis, Research Center for Genetic Information, Medical Institute of Bioregulation, Kyushu University, Fukuoka, Japan

⁹Department of Medical Chemistry, Kurume University School of Medicine, Fukuoka, Japan

***Corresponding author:** E-mail: rkimura@med.u-ryukyu.ac.jp; hiroki_oota@med.kitasato-u.ac.jp.

Associate editor: Naruya Saitou

Abstract

The Ryukyu Islands are located to the southwest of the Japanese archipelago. Archaeological evidence has revealed the existence of prehistoric cultural differentiation between the northern Ryukyu islands of Amami and Okinawa, and the southern Ryukyu islands of Miyako and Yaeyama. To examine a genetic subdivision in the Ryukyu Islands, we conducted genome-wide single nucleotide polymorphism typing of inhabitants from the Okinawa Islands, the Miyako Islands, and the Yaeyama Islands. Principal component and cluster analyses revealed genetic differentiation among the island groups, especially between Okinawa and Miyako. No genetic affinity was observed between aboriginal Taiwanese and any of the Ryukyu populations. The genetic differentiation observed between the inhabitants of the Okinawa Islands and the Miyako Islands is likely to have arisen due to genetic drift rather than admixture with people from neighboring regions. Based on the observed genetic differences, the divergence time between the inhabitants of Okinawa and Miyako islands was dated to the Holocene. These findings suggest that the Pleistocene inhabitants, whose bones have been found on the southern Ryukyu Islands, did not make a major genetic contribution, if any, to the present-day inhabitants of the southern Ryukyu Islands.

Key words: Japanese archipelago, Ryukyu Islanders, population structure, demographic history, genome-wide SNPs.

Introduction

The Ryukyu Islands located to the southwest of the Japanese archipelago are distributed over an area of 610 km from the north to the south, and 840 km from the east to the west. The archipelago consists primarily of the Amami Islands, Okinawa Islands (OK), Miyako Islands (MY), and the Yaeyama Islands (YE) (fig. 1). Of the numerous straits in the Ryukyu Archipelago, the Kerama Gap between Okinawa-jima (the largest of the OK) and Miyako-jima (the largest of the MY) is the largest at 260 km wide (here, “-jima” is used for “Island” to distinguish it from “Islands”) (Kishimoto 2003). This strait divides the Ryukyu Islands into the northern Ryukyu and southern Ryukyu islands.

Archaeological findings show that two different cultures developed in the northern and southern Ryukyus during

prehistoric times (Asato 2003). The distributions of these cultures reflected the geography of the region, with the Amami Islands and the OK belonging to the northern Ryukyu cultural group, and the MY and the YE belonging to the southern Ryukyu cultural group (fig. 1). In northern Ryukyu, Shellmound Period has extended from 6,700 to 1,000 before present (BP) (Asato 2003). Based on pottery characteristics, the culture of this period is considered to have been similar to that of the Jomon period on the mainland of Japan (i.e., the islands of Hokkaido, Honshu, Shikoku, Kyushu: fig. 1). These apparent similarities prompted some archaeologists to propose that northern Ryukyu was settled by immigrants from mainland Japan (MJ) (Asato 2003). Two prehistoric periods have been characterized in southern Ryukyu: The Shimotabaru Period, which spanned from 4,500 to 3,000 BP

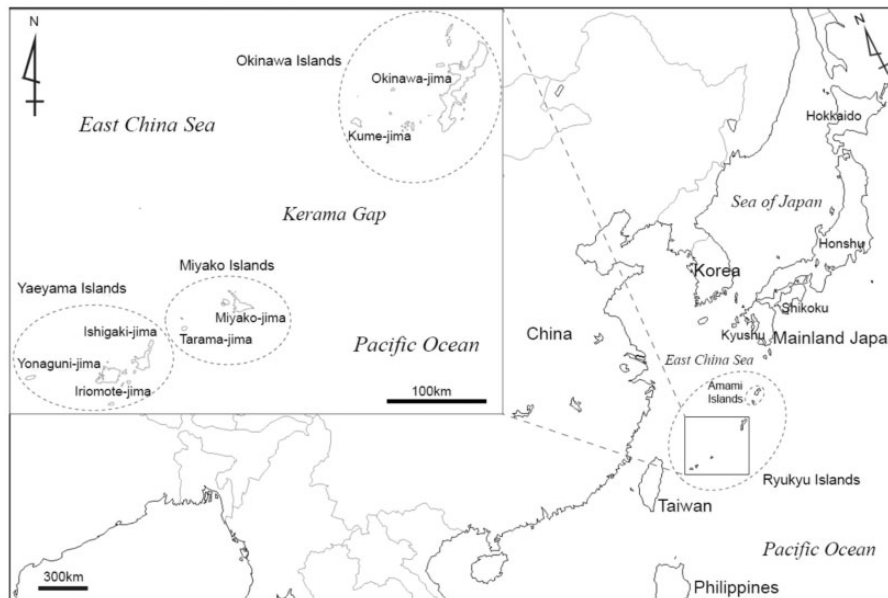


Fig. 1. Map of the Japanese archipelago and neighboring regions. The Japanese archipelago consists of four main islands (Hokkaido, Honshu, Shikoku, and Kyushu), the Ryukyu Islands, and numerous other islands. The Ryukyu Islands consist of the Amami Islands, the OK, the MY, and the YE.

and was characterized by a specific style of pottery, and the Aceramic Period from 2,500 to 800 BP, during which no pottery was produced in this region. As no affinity has been observed between the northern and the southern Ryukyu cultures in these prehistoric periods, some archaeologists have hypothesized that there was no human migration between the northern and southern Ryukyu during these periods (Asato 2003). However, in the subsequent Gusuku Period (800–600 BP), the cultures of both the northern and southern Ryukyus were unified after rice cultivation was introduced to the Ryukyu Islands.

The peopling of the Ryukyu Islands is considered central to understanding the origin of the Japanese people. The dual structure model of Japanese origin, which is based on cranial and dental morphology, hypothesizes that 1) the first inhabitants of the Japanese archipelago originated from somewhere in Southeast Asia, and that initial colonization occurred during the Upper Paleolithic age (probably through the Ryukyu Islands). These colonialists are considered to have given rise to the native hunter-gatherers in the Japanese Archipelago, or Jomon people. 2) This initial wave of migration was then followed by a second wave of migration from northeast Asia, which occurred during in and after the Aeneolithic Yayoi age. Finally, 3) the populations of both lineages (Jomon and Yayoi people) gradually mixed with each other on MJ, except for in Hokkaido. According to this hypothesis, the marked similarities between the Ainu and the Ryukyu Islanders are due to their common origin from the Jomon people (Hanihara 1991). Of the numerous anthropological studies that have been conducted to date to verify the dual structure model, most concur with the idea that the Japanese have a dual structure and that the Ryukyu Islanders share a common origin with the Ainu (Hanihara 1993; Matsumura 1994; Hammer and Horai 1995; Horai

et al. 1996; Omoto and Saitou 1997; Tajima et al. 2002; Tajima et al. 2004; Matsumura and Hudson 2005; Fukumine et al. 2006; Ishida et al. 2009; Matsukusa et al. 2010; Jinam et al. 2012; Koganebuchi et al. 2012). Nonetheless, the origin and migration routes employed by the Jomon people are still open to debate. In these previous studies, individuals from Okinawa-jima have been considered to be representative of the Ryukyu Islanders, but no consideration has been given to the geographical and cultural boundary that exists between the northern and southern Ryukyus.

Indeed, relatively few studies have examined the population structure of the inhabitants of the Ryukyu Islands to date. Of the studies that have been undertaken, some have reported the existence of local variations in morphological and genetic characteristics (Naito 1976; Omoto et al. 1976; Haneji et al. 2007; Toma et al. 2007; Matsukusa et al. 2010). However, the population structure within the Ryukyu Islands remains unclear due to a lack of comprehensive data and due to incompatibilities between the findings of previous studies. In this study, we performed a comprehensive genetic analysis based on genome-wide single nucleotide polymorphism (SNP) data of the Ryukyu Islanders. Specifically, we included individuals from the OK, the MY, and the YE. We also discuss the demographic history of the Ryukyu Islanders and provide evidence for genetic differentiation between the inhabitants of the Okinawa and Miyako island groups.

Results and Discussion

Population Structure in the Ryukyu Islanders

After the quality control and outlier filtering of the genome-wide SNP data (see Materials and Methods, [supplementary figs. S1 and S2](#), and [table S1, Supplementary Material](#) online), population genetics-based analyses were performed. [Figure 2](#) shows the results of principal component analysis (PCA)

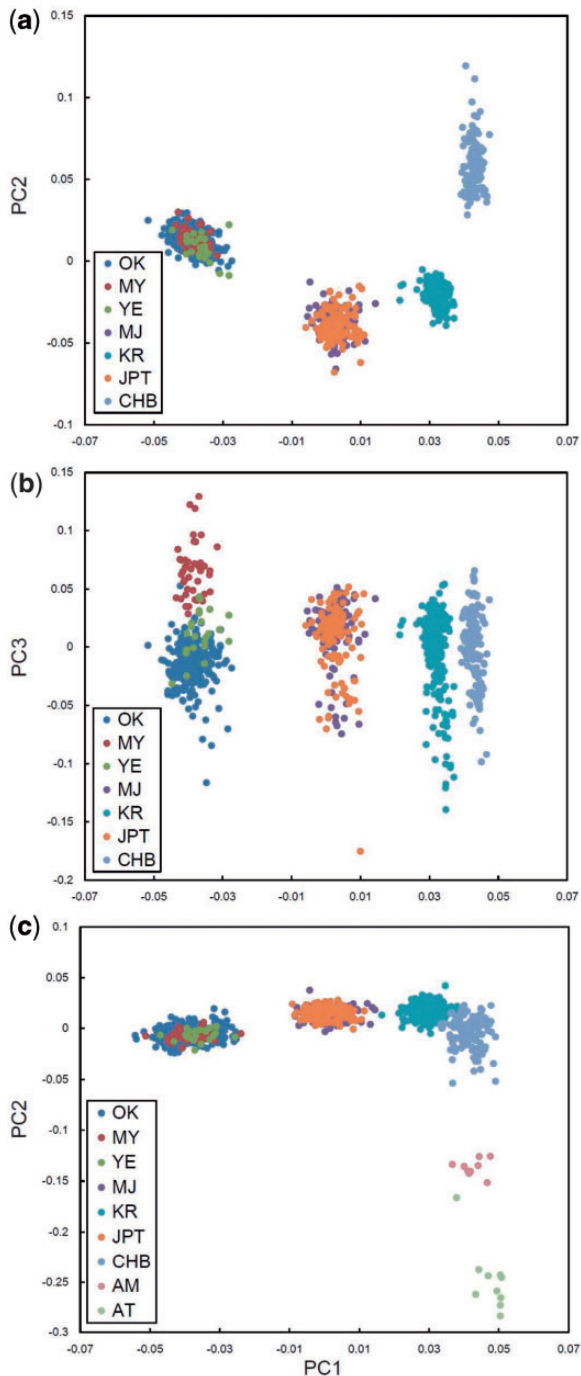


Fig. 2. The results of PCA using genotype data for East Asian populations. (a) PC1 versus PC2, and (b) PC1 versus PC3 for 540,451 SNPs. (c) PC1 versus PC2 for 12,484 SNPs, which include aboriginal Taiwanese samples (Ami and Atyal). Population codes are as follows: OK, Okinawa; MY, Miyako; YE, Yaeyama; MJ, mainland Japanese; KR, Koreans; AM, Ami; AT, Atyal.

using 540,451 SNPs shared among data sets including individuals from the OK, MY, YE, MJ, Korea (KR), and two HapMap populations (Chinese Han in Beijing, CHB; Japanese in Tokyo, JPT). Four main clusters (Ryukyu Islanders, mainland Japanese, Koreans, and Han Chinese) were observed in PC1. The observation that almost all Japanese individuals were grouped into two main clusters (Ryukyu and MJ clusters) was consistent

with a previous study (Yamaguchi-Kabata et al. 2008). Under the dual structure model of Japanese origin (Hanihara 1991), PC1 can be interpreted as being associated with the Jomon and Yayoi components of individual ancestry. Meanwhile, a biplot of PC1 versus PC2 produced a “U” shape (fig. 2a), implying that PC2 is unlikely to directly reflect any demographic events; this pattern is remarkably similar to a pattern of mathematical artifacts that has previously been observed in a simulation of one-dimensional stepping stone-like habitats (Novembre and Stephens 2008). The OK and MY individuals were separated into two clusters (Okinawa and Miyako clusters) in PC3, and YE individuals were plotted between the Okinawa and Miyako clusters (fig. 2b).

We then performed ADMIXTURE analysis by adding four HapMap populations (CHB; JPT; Yorba in Ibadan, YRI; Utah residents with ancestry from northern and western Europe, CEU). The resulting bar plot revealed that the East Asian populations were primarily formed by admixture of two ancestral populations at different proportions (shown as green and purple in the bar plot in fig. 3a) when the assumed number of ancestral populations (K) was 4, at which the minimum value of the cross-validation error was observed (fig. 3b). The relative proportions of the two ancestral populations (green vs. purple) at $K = 4$ were approximately 1:9 in Koreans and 5:5 in mainland Japanese, which also supports the dual structure model of Japanese origins (Hanihara 1991); the two components for East Asians (green and purple) may not correspond exactly to the genetic components of the Jomon and Yayoi people, respectively, but they are at least likely to be associated with the population structure formed by the admixture of two ancestral populations. A component specific to the MY samples (shown in orange) appeared at $K = 6$, indicating that the inhabitants of the OK and of the MY are subdivided. For the level of individual ancestry at $K = 6$, YE individuals were more similar to the OK individuals compared with the MY individuals.

Our PCA and ADMIXTURE analysis showed that MY individuals can be genetically differentiated from OK individuals and that YE individuals are genetically closer to OK individuals than they are to MY individuals (figs. 2b and 3), even though the MY are geographically located between the others (fig. 1). This is also confirmed by the genetic distance between populations (table 1). It has also previously been suggested that people from the MY were genetically isolated from the other Ryukyu populations, based on somatometry (Naito 1976), as well as 13 genetic markers associated with blood types, red cell enzymes, and serum proteins (Omoto et al. 1976). Results of the PCA and ADMIXTURE analysis possibly indicate that people of the YE were affected by gene flows from the OK. However, D statistic did not support the admixture in YE: D (CHB, OK; MY, YE) did not significantly depart from zero (table 2). Instead, we found that genetic diversity (D_X) within MY was smaller than those within OK and within YE (table 1). This suggests that extreme genetic drift in people of the MY due to small effective population size promoted a large genetic differentiation from the other populations and caused the observed population structure, which

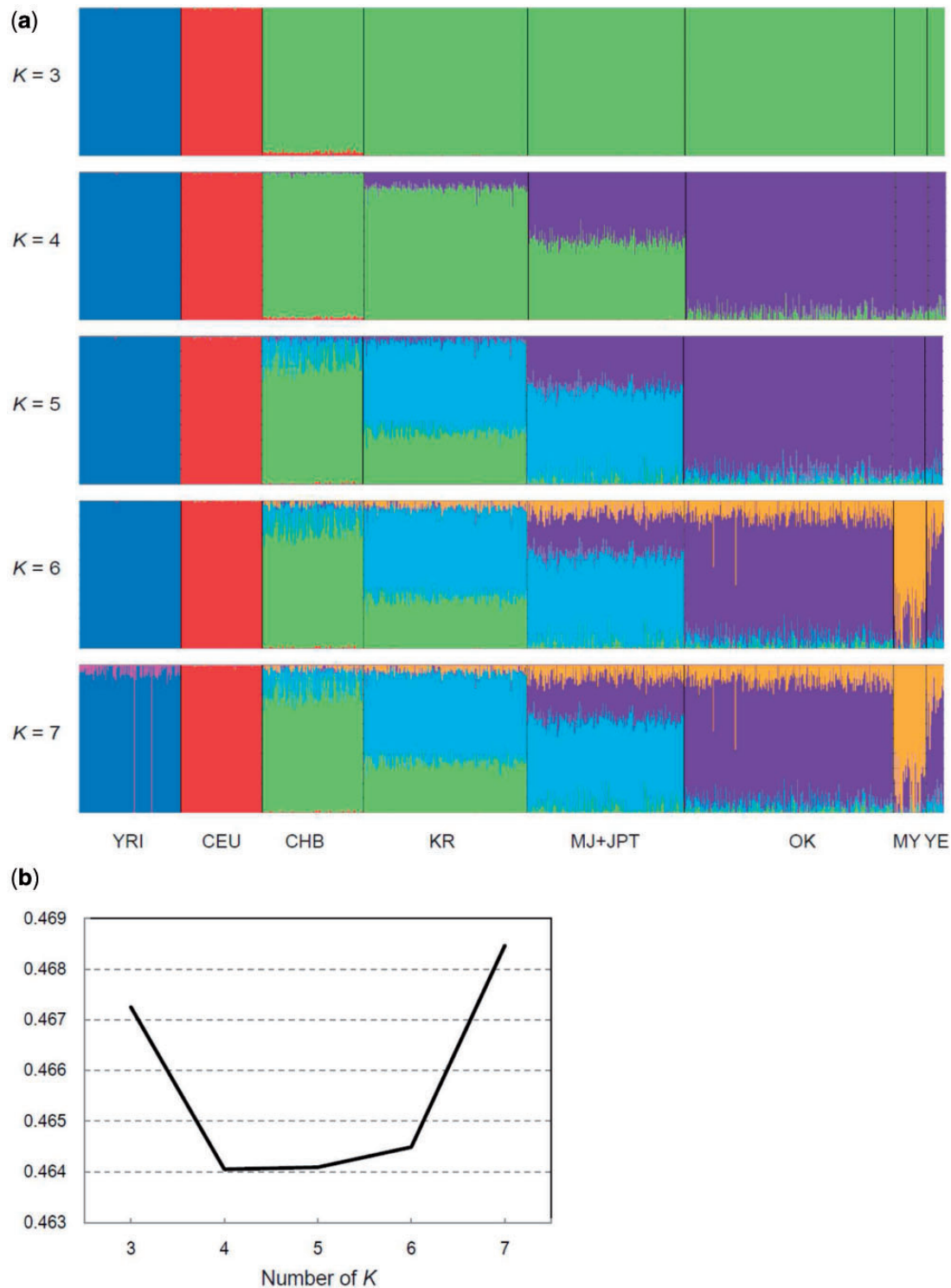


Fig. 3. The results of ADMIXTURE analysis. (a) Stacked barplots showing individual ancestry. (b) Cross-validation errors. Five-fold cross validations were performed for each number of K .

seems to be incompatible with geographic distributions of the Ryukyu populations.

The Genetic Relationship between the Ryukyu Islanders and Aboriginal Taiwanese

As described above, our data demonstrated the genetic differentiation between the island groups in the Ryukyu.

Alternative possible explanation instead of genetic drift is gene flow from a neighboring population, such as the aboriginal Taiwanese. To examine this possibility, we added data for aboriginal Taiwanese, Ami (AM) and Atyal (AT), from the HUGO Pan-Asian SNP (PASNP) Consortium (2009) database, and performed PCA and ADMIXTURE analysis using 12,488 SNPs overlapping between the data sets. The results of PCA (fig. 2c) and ADMIXTURE analysis (supplementary fig. S3,

Table 1. Genetic Distances within Each Population and between Each Pair of Populations.

	Genetic Distances							
	OK	MY	YE	MJ + JPT	KR	CHB	CEU	YRI
Okinawa	0.2755	0.0010	0.0004	0.0016	0.0038	0.0050	0.0368	0.0652
Miyako	0.2758	0.2741	0.0010	0.0024	0.0045	0.0057	0.0375	0.0660
Yaeyama	0.2757	0.2756	0.2751	0.0017	0.0039	0.0050	0.0368	0.0654
Mainland Japan + JPT	0.2780	0.2781	0.2779	0.2773	0.0008	0.0020	0.0361	0.0645
Korea	0.2800	0.2800	0.2798	0.2779	0.2769	0.0007	0.0364	0.0648
HapMap3 CHB	0.2820	0.2821	0.2819	0.2800	0.2785	0.2787	0.0355	0.0638
HapMap3 CEU	0.3251	0.3251	0.3249	0.3253	0.3254	0.3254	0.3011	0.0541
HapMap3 YRI	0.3518	0.3519	0.3518	0.3520	0.3521	0.3520	0.3535	0.2977

NOTE.— D_x values are shown on the diagonal, D_{xy} values are shown below the diagonal, and D_m values are shown above the diagonal. Unbiased estimates were calculated for D_x due to the small sample size of the Miyako and Yaeyama populations.

Supplementary Material online) demonstrated no genetic affinities between the aboriginal Taiwanese and any of the Ryukyu individuals. In addition, we calculated D statistics to assess the admixture. D (CEU, AM; MY, OK) and D (CEU, AT; MY, OK) did not significantly depart from zero (table 2). These results suggest that genetic differentiation between the people of the MY and of the OK are unlikely to have resulted from gene flow from Taiwan, but rather due to genetic drift after the Miyako ancestors migrated from the northern Ryukyu to the southern Ryukyu. A previous study based on mtDNA, Y chromosomal STRs, and autosomal STRs also showed that there was no evidence of any contribution from the aboriginal Taiwanese populations to the gene pool of the Ryukyu Islanders (Matsukusa et al. 2010). Similarly, using a large number of SNP markers, this study verified that there is a clear genetic gap between Taiwan and the Ryukyu Islands.

Demographic Reconstruction among the Ryukyu Islanders

To visualize the phylogenetic relationships between the Ryukyu Islanders and the neighboring populations, we constructed a neighbor-joining (NJ) tree and a neighbor-net based on D_x , D_{xy} , and Nei's minimum distances between the populations (table 1, fig. 4a, and supplementary fig. S4, Supplementary Material online). The NJ tree clearly showed the extremely low interpopulation genetic variances compared with intrapopulation genetic variances, as described in the previous study (Rosenberg et al. 2002). Reticulations observed in the neighbor-net (fig. 4a) indicate the occurrence of past admixture events in these populations, suggesting that mainland Japanese and Koreans are composed of two admixed ancestral populations. D (CEU, OK; CHB, MJ) and D (CEU, OK; CHB, KR) supported the admixture events in mainland Japanese and Koreans (table 1), and the admixture graph also indicated multiple origins of mainland Japanese (fig. 4b and supplementary fig. S5, Supplementary Material online). In the neighbor-net and admixture graph (fig. 4), OK, MY, and YE formed a monophyletic Ryukyu cluster. The relatively long branch to the Ryukyu cluster in the network explains the low genetic diversity among the individuals composing the Ryukyu cluster. Especially, the external branch to MK was

Table 2. D Statistics.

	D Statistic	Z Score
D (CHB, OK; MY, YE)	−0.0002	−0.622
D (CEU, AM; MY, OK)	0.0021	1.178
D (CEU, AT; MY, OK)	0.0016	0.899
D (CEU, OK; CHB, MJ)	0.0171	41.697
D (CEU, OK; CHB, KR)	0.0092	31.658

the longest in the Ryukyu cluster, reflecting the low genetic diversity within the Miyako population. Assuming simple demographic models, we estimated demographic parameters by two different approaches, moment-based and Bayesian methods, especially focusing on the divergence between the Okinawa and Miyako Islanders.

In the moment-based method, we employed several assumptions to infer demographic parameters from genetic diversities. To exclude SNPs that had been arisen in Eurasian populations after the out of Africa migrations, only polymorphic SNP markers in YRI were used. CEU and CHB were considered as outgroups of the Ryukyu populations. Further, for simplicity, we assumed that there was no gene flow among these populations after divergence. Based on the newly calculated genetic differences within and between populations (D_x and D_{xy} values), an NJ tree was constructed (fig. 5a) and average gene diversities at different times were estimated (H_{CH-EU} , H_{RY-CH} , H_{OK-MY} , and H_{OK}).

We considered two simple demographic models: I) a constant population size model (fig. 5b) and II) an exponential population growth model (fig. 5c). The exponential growth model assumes that the Ryukyu populations and Han Chinese expanded at the same growth rate (α). To estimate focal parameters, we fixed the divergence time between Asian and European ancestries ($T_{CH-EU} = 36,000$ BP) in both models, and the initial population size of the Asian ancestry ($N_{CH0} = 1,200$) in model II, all of which were estimated in a previous study (Gronau et al. 2011). Parameter estimates for each model are shown in table 3. In the constant population size model, we estimated the approximate divergence time between the inhabitants of the OK and the Han Chinese (T_{OK-CH}) to be 3,000 BP (assuming 25 years/generation). The divergence time between the inhabitants of the OK

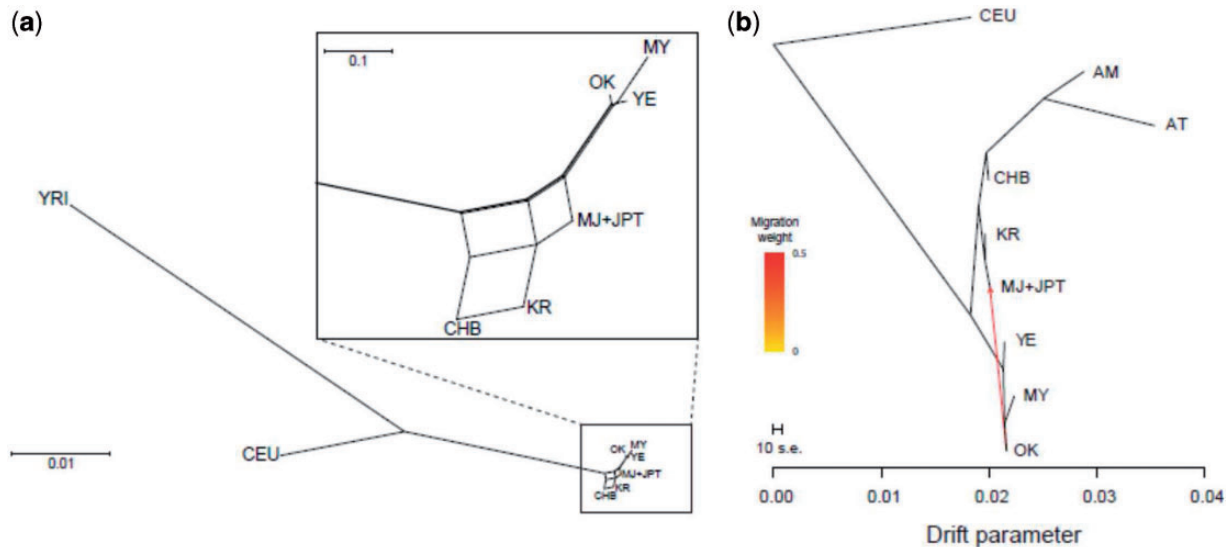


Fig. 4. Phylogenetic relationships among populations. (a) A neighbor-net based on Nei's minimum distance (D_m) among eight populations. (b) An admixture graph allowing one migration event (shown as a red arrow). The data of MJ individuals in this study were merged with JPT.

and the MY (T_{OK-MY}) was dated to approximately 130 BP. In this model, the effective population size of Han Chinese from T_{CH-EU} to the present (N_{CH}) was estimated to be approximately 5,000, whereas that in the OK from T_{OK-CH} to the present (N_{OK}) was estimated to be approximately 2,600. In the exponential growth model, T_{OK-CH} and T_{OK-MY} were dated to approximately 15,000 and 1,600 BP, respectively. The initial population size in the OK (N_{OK0}) was considered to be 6,426. These results showed that the estimated divergence times vary markedly depending on the model. When we performed the analysis using $T_{CH-EU} = 26,400$ BP and $N_{CH0} = 590$ estimated by Gutenkunst et al. (2009), instead of the estimates by Gronau et al. (2011), more recent (in model I) and older (in model II) divergence times were obtained, respectively (table 3). Using simulation data, we also confirmed that the moment-based method is robust even under relatively strong ascertainment biases (see supplementary text S1, table S2, and fig. S6, Supplementary Material online).

In the moment-based method above, parameter estimates were drastically affected by the prefixed parameters and, in addition, isolation-with-migration models were not applicable. To compensate for these drawbacks, we also apply kernel approximate Bayesian computation (kernel-ABC) (Fukumizu et al. 2013; Nakagome et al. 2013), taking account of possible migrations between the Okinawa and Miyako populations as shown in figure 6. As the summary statistics in this analysis, we used the frequency spectrum of haplotypes that are defined by SNPs shared between populations, which represent patterns of linkage disequilibrium.

We tested 12 alternative models that vary in migration rate (m) and prior distribution of divergence time (T) (table 4, fig. 6, and supplementary table S3, Supplementary Material online). Then, the best fitting model was chosen from these models using approximated Bayes factor (aBF). The model that has $m = 0$ and the prior of log-normal distribution (LN)

with the mean ($\mu = 50$) and the variance (μ^2), LN(50, 50²), shows the highest approximated marginal likelihood (aML) among the models. Based on the criteria of BF (Kass and Raftery 1995), this model is strongly or decisively more likely than the other models, except for the model with $m = 0.001$ and LN(100, 100²). This means that we can reject models with higher migration rates ($m = 0.01$ and 0.1). The divergence time between the Okinawa and Miyako Islanders under the likely models with $m = 0$ and LN(50, 50²) and with $m = 0.001$ and LN(100, 100²) was estimated to be $1,800 \pm 150$ and $2,600 \pm 250$ BP, respectively (table 4 and supplementary table S3, Supplementary Material online). These results suggested that the divergence event between the Okinawa and the Miyako Islanders dates from the Holocene, even when the effect of migration is considered.

Remarks on the Peopling of the Ryukyu Islands

Human remains dating from the Pleistocene have been discovered on the Ryukyu Islands. One of the most famous Pleistocene sites is the Minatogawa Fissure on Okinawa-jima (northern Ryukyu) that was discovered by an amateur archaeologist, Seiho Oyama (Suzuki 1975). A total of four skeletons have been excavated at this site, and two charcoal fragments found in the vicinity of human remains were dated to $16,600 \pm 300$ and $18,250 \pm 650$ BP by ¹⁴C dating (Matsu'ura 1999; Hashimoto et al. 2002). On Miyako-jima (southern Ryukyu), 26,000-year-old human remains were excavated from the Pinza-Abu Cave site (Sakura 1985). More recently, a 20,000-year-old human bone fragment was excavated from the Shiraho-Saonetabaru Cave site on Ishigaki-jima, the main island of the YE (Nakagawa et al. 2010). Our estimates for the parameters T_{RK-CH} and T_{OK-MY} were more recent than the ages of these Pleistocene humans. If these ancient people had made marked contributions to the gene pools of the modern inhabitants of the islands where they were excavated, then the estimates of T_{RK-CH} and T_{OK-MY}

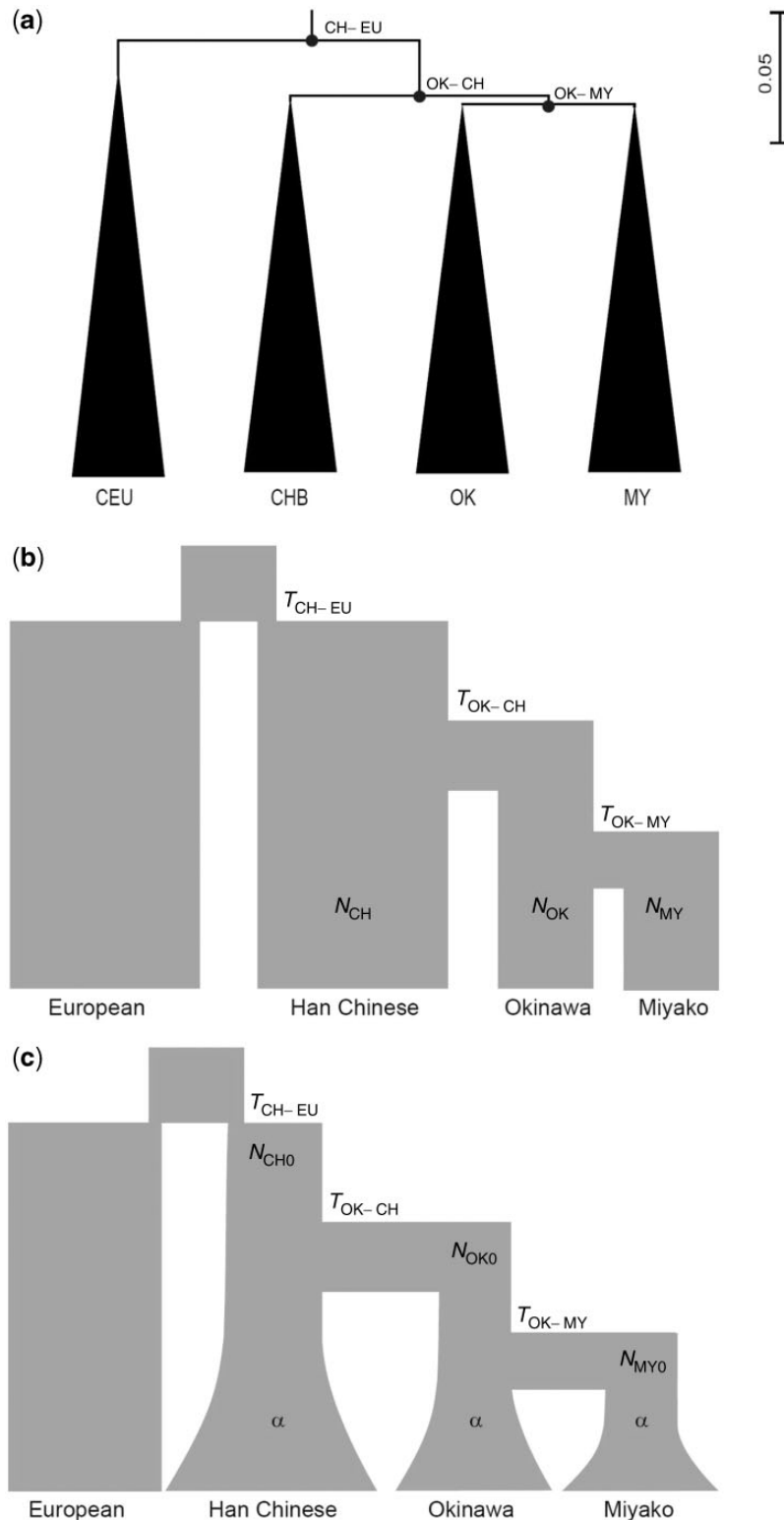


Fig. 5. Models for estimating demographic parameters based on genetic diversity. T_{X-Y} is the divergence time between populations X and Y . N_X and N_{X0} indicate the effective and initial population sizes of population X , respectively. (a) NJ tree based on 511,913 SNPs that are polymorphic in YRI. Branch lengths from the split between populations X and Y to each external node denote a half of the gene diversity in the ancestral population at the divergence time (H_{X-Y}). The height of the triangle on the external node of population X denotes a half of the gene diversity in the present population X (H_X). (b) Model I assumes a constant population size. (c) Model II assumes exponential growth of Asian populations. In this model, population growth rate of the Ryukyu Islanders and CHB is α . The value of T_{CH-EU} that used in this study is an estimate by Gutenkunst et al. (2009) or Gronau et al. (2011).

Table 3. Parameter Estimates Based on Genetic Diversity.

	Model I		Model II	
Fixed parameters				
T_{CH-EU}	1,056 gens (26,400 BP)	1,440 gens (36,000 BP)	1,056 gens (26,400 BP)	1,440 gens (36,000 BP)
N_{CHO}	—	—	590	1,200
Reference	Gutenkunst et al. (2009)	Gronau et al. (2011)	Gutenkunst et al. (2009)	Gronau et al. (2011)
Estimated parameters				
N_{CH}	3,534	4,818	3,534 ^a	4,818 ^a
α	(1)	(1)	1.0057	1.0027
T_{RK-CH}	89 gens (2,225 BP)	131 gens (3,275 BP)	623 gens (15,575 BP)	605 gens (15,125 BP)
N_{OK0}	1,935	2,639	3,728	6,426
N_{OK}	1,935	2,639	13,541 ^a	15,649 ^a
T_{OK-MY}	4 gens (100 BP)	5 gens (135 BP)	159 gens (3,975 BP)	63 gens (1,575 BP)
N_{MY0}	1,569	2,140	8,190	4,522
N_{MY}	1,569	2,140	12,439 ^a	4,928 ^a

^aHarmonic mean of population sizes is shown.

should have been dated within the Pleistocene. Therefore, our findings suggest that the first inhabitants of the Ryukyu Islands during the Pleistocene were unlikely to have contributed substantially, or at all, to the modern inhabitants of the island. Indeed, our analysis suggested that more recent divergence would sufficiently explain the population structure of the modern inhabitants of the Ryukyu Islands. A previous study proposed that the first sustained colonization of the Ryukyu Islands occurred during the Holocene (Takamiya 2006), and our results strongly support this hypothesis.

This study proposed that large migrations of the modern Ryukyu Islanders occurred during the Holocene, even though the T_{OK-MY} estimated in our analyses varied depending on which model was employed, and that these migrations could not be assigned to a particular period, such as Simotabaru (4,500–3,000 BP), Aceramic (2,500–800 BP), or Gusuku (800–600 BP) periods. This uncertainty in the estimation is due to limitations in demographic reconstruction using SNP genotyping data. Whole-genome sequencing data will enable us to more accurately estimate demographic parameters. Therefore, accumulation of whole-genome sequencing data is considered necessary to derive a more detailed demographic history of the Ryukyu Islanders.

Materials and Methods

Samples

A total of 449 Japanese individuals (OK, $n = 281$; MY, $n = 44$; YE, $n = 24$; MJ, $n = 100$) participated in this study; some of the MY and YE individuals are the same as those examined in Matsukusa et al. (2010). Written informed consent was obtained from all of the participants, and only individuals with four grandparents of the same origin were used in this study. Blood or saliva samples were collected from the participants. This study was approved by the ethical committees at University of the Ryukyus, Showa University, and Kitasato University. Geographic locations of each population sampled in this study are shown in figure 1.

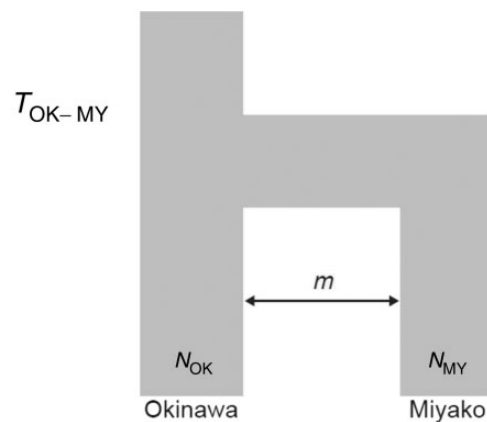


FIG. 6. Schematic models employed for the kernel-ABC. T_{OK-MY} is the divergence time between the people from the OK and the MY. N_{OK} and N_{MY} are the effective population sizes of the OK and the MY, respectively. We tested 12 demographic models that assumed the migration rate $m = 0, 0.001, 0.01, \text{ or } 0.1$ in combination with the prior distribution of divergence time $T \sim \text{LN}(\mu = 50, \mu^2 = 50^2)$, $\text{LN}(\mu = 100, \mu^2 = 100^2)$, or $\text{LN}(\mu = 400, \mu^2 = 400^2)$.

For comparison with our data, we used genotype data for KR individuals ($n = 224$) (Haga et al. 2013). We also downloaded genotype data for unrelated individuals of JPT ($n = 113$), CHB ($n = 137$), CEU ($n = 112$), and YRI ($n = 147$) from release 23 of the HapMap project phase 3 database (International HapMap 3 Consortium 2010), and for aboriginal Taiwanese (Ami, $n = 10$; Atyal, $n = 10$) from the HUGO PASNP database (HUGO Pan-Asian SNP Consortium 2009; Ngamphiw et al. 2011).

DNA Extraction and Genotyping of Genome-Wide SNPs

Genomic DNA was extracted from blood and saliva samples using Genra Puregen Blood kit (Qiagen, Germantown, MD) or DNA Extractor WB kit (Wako, Osaka, Japan), and genotyped for 729,172 SNPs using HumanOmniExpress BeadChip system (Illumina, San Diego, CA). Quality control of the genotype data was carried out using the following

Table 4. Approximated Bayes Factors and Estimated Divergence Times under the 12 Models Using the Kernel-ABC Method.

Models		Prior Distributions ^a (μ : mean; μ^2 : variance)			
		$\mu = 50$	$\mu = 100$	$\mu = 400$	
Migration Rates (m)					
	0	aBF (log ₁₀ -scaled) ^b	0	2.014	14.076
		Posterior estimates: Generations ^c	72 (6)	81 (7)	113 (26)
10 ⁻³		Posterior estimates: Years ^c	1,800 (150)	2,025 (175)	2,825 (650)
		aBF (log ₁₀ -scaled) ^b	1.756	0.322	10.117
		Posterior estimates: Generations ^c	87 (9)	104 (10)	277 (77)
10 ⁻²		Posterior estimates: Years ^c	2,175 (225)	2,600 (250)	6,925 (1,925)
		aBF (log ₁₀ -scaled) ^b	17.84	17.322	15.844
		Posterior estimates: Generations ^c	280 (25)	482 (42)	931 (120)
10 ⁻¹		Posterior estimates: Years ^c	7,070 (620)	12,058 (1,060)	23,275 (2,991)
		aBF (log ₁₀ -scaled) ^b	177.701	163.353	178.458
		Posterior estimates: Generations ^c	11 (14)	19 (23)	119 (108)
		Posterior estimates: Years ^c	275 (350)	475 (575)	2,975 (2,700)

^aPrior distributions for divergence time are given by a log-normal distribution with mean (μ) and variance (μ^2).

^bThe aBF values (log₁₀-scaled) are calculated for each model against the highest aML model ($\mu = 50$ and $m = 0$).

^cThe mean and SD are calculated from 265 posterior means of the divergence time (assuming 25 years/generation).

criteria. After samples with call rates lower than 0.98 were excluded, those SNPs having call rates lower than 0.98 in the remaining samples were excluded. To ensure that only unrelated individuals were used for subsequent analyses, one individual in a pair who had a proportion of shared identical-by-descent ($\hat{\pi}$) higher than 0.125 (corresponding to a third-degree relative) was excluded from the data set. We also excluded individuals with an inbreeding coefficient (F) higher than 0.0625 (corresponding to mating between first cousins). In addition, we performed Hardy–Weinberg equilibrium test for each sample set (OK, MY, YE, MJ, KR, JPT, CHB, YRI, and CEU) using the exact test which adequately controls for type I errors in both large and small samples and is computationally efficient (Wigginton et al. 2005). Those SNPs with P values lower than 1.0×10^{-6} were excluded from the data set. All quality control procedures were performed using PLINK 1.07 (Purcell et al. 2007). As a result, 694,887 SNPs for OK, 694,952 SNPs for MY and YE, 694,944 SNPs for MJ, and 694,902 SNPs for KR passed the quality control criteria described above. Of the 694,852 overlapping SNPs in the five populations, 602,317 were polymorphic (supplementary table S1, Supplementary Material online).

To unify strands between genotype data generated by the Illumina HumanOmniExpress BeadChip and obtained from HapMap database, strand checks of our genotype data were performed using “check_strands.py,” a utility program for BEAGLE (Browning and Browning 2009). HapMap3 JPT and CHB data were used as a reference panel for the strand checks. As part of this process, a total of 154,401 SNPs were excluded due to missing HapMap3 data or strand ambiguities. Finally, 540,451 autosomal SNPs from 281 OK, 43 MY, 23 YE, 100 MJ, 221 KR, 113 JPT, 137 CHB, 111 CEU, and 139 YRI individuals passed the quality control and strand check procedures.

Detection of Population Outliers

Detection of population outliers is performed generally according to results of the PCA or multidimensional scaling in

recent genome-wide association studies (Diskin et al. 2012; Nakamura et al. 2012; Shi et al. 2012; Steffens et al. 2012). However, it is difficult to judge objectively how many dimensions should be used to detect population outliers. In this study, we used a dimension-free procedure based on genotype frequencies to exclude population outliers from a population without using data of any other population.

Let p_{0i} , p_{1i} , and p_{2i} be the frequencies of genotypes AA, AB, and BB for the i th SNP, where A and B are major and minor alleles, respectively. The probability that an individual has his/her genotype G_i at the i th SNP, x_i , can be estimated using the genotype frequency in the population,

$$\hat{x}_i = p_{0i}I(G_i = AA) + p_{1i}I(G_i = AB) + p_{2i}I(G_i = BB) \quad (1)$$

where I is 0 or 1. When all of the SNPs are independent, that is, the SNPs are not affected by linkage disequilibrium, the whole profile probability, Q , can be expressed by the equation:

$$Q = \prod_i \hat{x}_i. \quad (2)$$

The expectation of Q can be written as

$$\hat{E}(Q) = \prod_i M_i, \quad (3)$$

Where $M_i = p_{0i}^{p_{0i}} \times p_{1i}^{p_{1i}} \times p_{2i}^{p_{2i}}$, which is the geometric mean of genotype frequencies at the i th SNP.

Now we propose a statistic to detect outliers of a population as the equation:

$$L := \ln Q - \ln \hat{E}(Q). \quad (4)$$

When the SNP density in the real data is high, then the assumption that all the SNPs are independent does not hold. However, the real distribution of L , at least approximately, follows a normal distribution (supplementary fig S1, Supplementary Material online). Therefore, we can convert L to Z scores according to the observed variance in the

population to obtain P values. Here, we set the significance level at 0.05 and used Bonferroni correction for multiple comparison of individuals.

The results identified one OK individual, two KR individuals, three JPT individuals, two CHB individuals, three CEU individuals, and three YRI individuals were detected as population outliers, and these individuals were excluded from subsequent analyses. One of the three JPT outliers was considered to be cryptic Korean-Japanese as they were grouped in the Korean cluster in PCA using genotype data before outliers were excluded (supplementary fig. S2, Supplementary Material online). Our method successfully detected population outliers in a single population. As a result, 280 OK individuals, 43 MY individuals, 23 YE individuals, 100 MJ individuals, 219 KR individuals, 110 JPT individuals, 135 CHB individuals, 108 CEU individuals, and 136 YRI individuals were used for subsequent analyses.

Analyses for Inferring Population Structure

To investigate the relationships among the East Asian populations, PCA was performed using the EIGENSOFT 3.0 package (Patterson et al. 2006) with 540,451 SNP genotype data for the OK, MY, YE, MJ, KR, JPT, and CHB individuals. To investigate the genetic relationships between the Ryukyu Islanders and the aboriginal Taiwanese, we also included Ami (AX-AM) and Atyal (AX-AT) genotype data from the HUGO PASNP database (HUGO Pan-Asian SNP Consortium 2009; Ngamphiw et al. 2011). By combining all these data sets, a total of 12,484 overlapping SNPs were obtained and subject to PCA.

To estimate individual ancestry, we subjected nine sample sets (OK, MY, YE, MJ, KR, JPT, CHB, CEU, and YRI) to maximum-likelihood clustering analysis using the ADMIXTURE 1.21 (Alexander et al. 2009). We ran ADMIXTURE from $K = 3$ to 7 and calculated the cross validation error for each number of K . Five-fold cross validations were performed for each number of K . To assess possible admixtures between populations, we also calculated D statistics using ADMIXTOOLS 1.1 (Patterson et al. 2012).

Phylogenetic Analysis Based on the Genetic Distance

Unbiased estimates of the averaged genetic distances within populations (D_X), between populations (D_{XY}), and the Nei's minimum genetic distance (D_m) (Nei 1972) were calculated according to the following equations:

$$D_X = \frac{1}{L} \sum_i^L \frac{2a_{Xi}(2n_{Xi} - a_{Xi})}{2n_{Xi}(2n_{Xi} - 1)} \quad (5)$$

$$D_{XY} = \frac{1}{L} \sum_i^L \left(\frac{a_{Xi}}{2n_{Xi}} + \frac{a_{Yi}}{2n_{Yi}} - 2 \frac{a_{Xi}a_{Yi}}{2n_{Xi}2n_{Yi}} \right) \quad (6)$$

$$D_m = D_{XY} - \frac{D_X + D_Y}{2} \quad (7)$$

where L is the number of loci; n_X and n_Y are the numbers of individuals sampled in population X and Y , respectively; and a_X and a_Y are the allele counts observed in population X and Y , respectively. An NJ tree (Saitou and Nei 1987) and neighbor-net (Bryant and Moulton 2004) were constructed based on D_X , D_{XY} , and D_m using SPLITS TREE 4.12 (Huson and Bryant 2006) and MEGA 5.2 (Tamura et al. 2011). In addition, admixture graph was also constructed using TreeMix 1.12 (Pickrell and Pritchard 2012).

Estimation of Divergence Time between Populations Based on Genetic Diversity

In this analysis, we used SNPs that are polymorphic in YRI and then constructed an NJ tree based on D_{XY} (fig. 5a). Using the branch lengths of the NJ tree, demographic parameters, such as divergence times between populations and population sizes were calculated according to the following equation:

$$\frac{H_{T_2}}{H_{T_1}} = \prod_{t=T_1}^{T_2} \left(1 - \frac{1}{2N_t} \right), \quad (8)$$

where N_t is the population size at the t th generation, and H_{T_1} and H_{T_2} are estimates of the gene diversity at the T_1 th and T_2 th generations, respectively. The gene diversity at the divergence time between populations X and Y (H_{X-Y}) corresponds to twice the branch length from the split to each external node (fig. 5a). The gene diversity in the present population is D_X .

Using this approach, we tested two simple demographic models. The first model assumed a constant population size for the Asian populations (model I; fig. 5b), and the second assumed an exponential growth of the Asian populations (model II; fig. 5c). The divergence time between the Asian and European populations (T_{CH-EU}) was assumed to be 24,600 or 36,000 BP in both models, and the initial population size of Asians (N_{CH0}) was taken as 590 or 1,200 in model II, as these values were estimated previously (Gutenkunst et al. 2009; Gronau et al. 2011). In model II, we further assumed that every Asian population has the same population growth rate α , and that N_t can be defined by $N_t = N_0 \times \alpha^t$ (where N_0 is the initial population size). The harmonic mean of N_t corresponds to the effective population size over a period of time. By conducting coalescent simulations, we confirmed that this estimation method was robust to ascertainment bias of SNP data (supplementary text S1, table S2, and fig. S6, Supplementary Material online).

Bayesian Inference of Population Demographic History

We also inferred the demographic history of people from the OK and the MY by assuming an isolation-with-migration model that implements kernel-ABC (Fukumizu et al. 2013; Nakagome et al. 2013). ABC is a statistical framework that can be used to approximate a posterior estimate without a likelihood function, and data are summarized as a set of summary statistics (\mathcal{S}). The advantage of kernel-ABC is that computing costs are independent of the dimension of data, which means

that high-dimensional summary statistics can be incorporated into the ABC analyses. We used the haplotype frequency spectrum (HFS) as summary statistics, as these can account for recombination patterns at a genomic segment. In this study, the genome-wide SNP data were summarized into two-dimensional HFS (S_{2D-HFS}). To simplify the demographic model, we assumed the constant population size of OK from an infinite past and fixed migration rates between OK and MY ($m = 0, 0.001, 0.01, \text{ and } 0.1$) (fig. 6). A detailed explanation of the kernel-ABC methodology is given in the [supplementary text S2](#) and [figs. S7–S10, Supplementary Material](#) online.

Supplementary Material

[Supplementary texts S1 and S2](#), [tables S1–S3](#), and [figures S1–S10](#) are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

The authors thank all of the anonymous volunteers who donated DNA to this study. They are also grateful to Ms Michiko Komesu, Dr Takashi Toma, Mr Kei-Ichi Moromizato, and Ms Misaki Shindo for their assistance with sample collection. This work was supported by Grant-in-Aid for the Fellows of the Japan Society for the Promotion of Science (25-2544 to T.S. and 24-3234 to S.N.). This study was also supported in part by the Rising Star Program at University of the Ryukyus (to R.K.), Grant-in-Aid for Young Scientists (A) (22687023 to R.K.), Grant-in-Aid for Scientific Research (B) (22370087 to H.I.; 21370108 and 24370099 to H.O.) from the Japan Society for the Promotion of Science, and Grant-in-Aid for Scientific Research on Innovative Areas “Genome Science” (221S0002) from Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

- Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19:1655–1664.
- Asato S. 2003. Prehistory and proto-history in the Ryukyu island chain. History of Okinawa Prefecture 2: archaeology. Naha: Okinawa Prefecture. (in Japanese).
- Browning BL, Browning SR. 2009. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet.* 84:210–223.
- Bryant D, Moulton V. 2004. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Mol Biol Evol.* 21: 255–265.
- Diskin SJ, Capasso M, Schnepf RW, Cole KA, Attiyeh EF, Hou C, Diamond M, Carpenter EL, Winter C, Lee H, et al. 2012. Common variation at 6q16 within HACE1 and LIN28B influences susceptibility to neuroblastoma. *Nat Genet.* 44:1126–1130.
- Fukumine T, Hanihara T, Nishime A, Ishida H. 2006. Nonmetric cranial variation of early modern human skeletal remains from Kumejima, Okinawa and the peopling of the Ryukyu Islands. *Anthropol Sci.* 114: 141–151.
- Fukumizu K, Song L, Gretton A. 2013. Kernel Bayes' rule: Bayesian inference with positive definite kernels. *J Mach Learn Res.* 14: 3753–3783.
- Gronau I, Hubisz MJ, Gulko B, Danko CG, Siepel A. 2011. Bayesian inference of ancient human demography from individual genome sequences. *Nat Genet.* 43:1031–1034.
- Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* 5:e1000695.
- Haga S, Nakaoka H, Yamaguchi T, Yamamoto K, Kim YI, Samoto H, Ohno T, Katayama K, Ishida H, Park SB, et al. 2013. A genome-wide association study of third molar agenesis in Japanese and Korean populations. *J Hum Genet.* 58:799–803.
- Hammer M, Horai S. 1995. Y chromosomal DNA variation and the peopling of Japan. *Am J Hum Genet.* 56:951–962.
- Haneji K, Hanihara T, Sunakawa H, Toma T, Ishida H. 2007. Non-metric dental variation of Sakishima Islanders, Okinawa, Japan: a comparative study among Sakishima and neighboring populations. *Anthropol Sci.* 115:35–45.
- Hanihara K. 1991. Dual structure model for the population history of the Japanese. *Jpn Rev.* 2:1–33.
- Hanihara T. 1993. Craniofacial features of southeast Asians and Jomonese: a reconsideration of their microevolution since the Late Pleistocene. *Anthropol Sci.* 101:25–46.
- Hashimoto M, Yahagi K, Tsujimoto H, Saito T, and Yoshida K. 2002. Geochemical analysis of Minatogawa Fissure site. In: Gushikami Village Board of Education, editor. Minatogawa Fissure Site. Chapter 8. Gushikami: Gushikami Village. pp. 89–106. (in Japanese).
- Horai S, Murayama K, Hayasaka K, Matsubayashi S, Hattori Y, Fuchareon G, Harihara S, Park KS, Omoto K, Pan IH. 1996. mtDNA polymorphism in East Asian Populations, with special reference to the peopling of Japan. *Am J Hum Genet.* 59:579–590.
- HUGO Pan-Asian SNP Consortium. 2009. Mapping human genetic diversity in Asia. *Science* 326:1541–1545.
- Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol.* 23:254–267.
- International HapMap 3 Consortium. 2010. Integrating common and rare genetic variation in diverse human populations. *Nature* 467: 52–58.
- Ishida H, Hanihara T, Kondo O, Fukumine T. 2009. Craniometric divergence history of the Japanese populations. *Anthropol Sci.* 117: 147–156.
- Jinam T, Nishida N, Hirai M, Kawamura S, Oota H, Umetsu K, Kimura R, Ohashi J, Tajima A, Yamamoto T, et al. 2012. The history of human populations in the Japanese Archipelago inferred from genome-wide SNP data with a special reference to the Ainu and the Ryukyuan populations. *J Hum Genet.* 57:787–795.
- Kass RE, Raftery AE. 1995. Bayes factors. *J Am Stat Assoc.* 90:773–795.
- Kishimoto Y. 2003. Nature and antiquity of Ryukyu island chain. History of Okinawa Prefecture 2: archaeology. Naha: Okinawa Prefecture. (in Japanese).
- Koganebuchi K, Katsumura T, Nakagome S, Ishida H, Kawamura S, Oota H, Consortium TAADR. 2012. Autosomal and Y-chromosomal STR markers reveal a close relationship between Hokkaido Ainu and Ryukyu islanders. *Anthropol Sci.* 120:199–208.
- Matsukusa H, Oota H, Haneji K, Toma T, Kawamura S, Ishida H. 2010. A genetic analysis of the Sakishima islanders reveals no relationship with Taiwan aborigines but shared ancestry with Ainu and mainland Japanese. *Am J Phys Anthropol.* 142:211–223.
- Matsumura H. 1994. A microevolutional history of the Japanese people from a dental characteristics perspective. *Anthropol Sci.* 102:93–118.
- Matsumura H, Hudson MJ. 2005. Dental perspectives on the population history of Southeast Asia. *Am J Phys Anthropol.* 127:182–209.
- Matsu'ura S. 1999. A chronological review of Pleistocene human remains from the Japanese Archipelago. In: Omoto K, editor. Interdisciplinary perspectives on the origins of the Japanese. Kyoto: International Research Center for Japanese Studies. p. 181–197.
- Naito Y. 1976. People in Okinawa prefecture from the point of view of somatometry. In: Association of Nine Learned Societies, editor. OKINAWA-Nature-Culture-Society. Tokyo: Kobundo. p. 77–86. (in Japanese).
- Nakagawa R, Doi N, Nishioka Y, Nunami S, Yamauchi H, Fujita M, Yamazaki S, Yamamoto M, Katagiri C, Mukai H, et al. 2010. Pleistocene human remains from Shiraho-Saonetabaru Cave on Ishigaki Island, Okinawa, Japan, and their radiocarbon dating. *Anthropol Sci.* 118:173–183.

- Nakagome S, Fukumizu K, Mano S. 2013. Kernel approximate Bayesian computation in population genetic inferences. *Stat Appl Genet Mol Biol.* 12:667–678.
- Nakamura M, Nishida N, Kawashima M, Aiba Y, Tanaka A, Yasunami M, Nakamura H, Komori A, Nakamura M, Zeniya M, et al. 2012. Genome-wide association study identifies TNFSF15 and POU2AF1 as susceptibility loci for primary biliary cirrhosis in the Japanese population. *Am J Hum Genet.* 91:721–728.
- Nei M. 1972. Genetic distance between populations. *Am Nat.* 106: 283–292.
- Ngamphiw C, Assawamakin A, Xu S, Shaw PJ, Yang JO, Ghang H, Bhak J, Liu E, Tongsimma S. 2011. PanSNPdb: the Pan-Asian SNP genotyping database. *PLoS One* 6:e21451.
- Novembre J, Stephens M. 2008. Interpreting principal component analyses of spatial population genetic variation. *Nat Genet.* 40: 646–649.
- Omoto K, Misawa S, Ishimoto G. 1976. People in Okinawa prefecture from the point of view of genetic markers which is associated with blood. In: Association of Nine Learned Societies, editor. OKINAWA-Nature-Culture-Society. Tokyo: Kobundo. p. 93–111. (in Japanese).
- Omoto K, Saitou N. 1997. Genetic origins of the Japanese: a partial support for the dual structure hypothesis. *Am J Phys Anthropol.* 102:437–446.
- Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D. 2012. Ancient admixture in human history. *Genetics* 192:1065–1093.
- Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. *PLoS Genet.* 2:e190.
- Pickrell JK, Pritchard JK. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* 8: e1002967.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 81:559–575.
- Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovskiy LA, Feldman MW. 2002. Genetic structure of human populations. *Science* 298:2381–2385.
- Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol.* 4:406–425.
- Sakura H. 1985. Pleistocene human fossil remains from Pinza-Abu (Goat Cave), Miyako Island, Okinawa, Japan. In: Okinawa Prefecture editor. Reports on Excavation of the Pinza-Abu Cave. Naha: Okinawa Prefecture. p. 161–176. (in Japanese).
- Shi Y, Zhao H, Cao Y, Yang D, Li Z, Zhang B, Liang X, Li T, Chen J, Shen J, et al. 2012. Genome-wide association study identifies eight new risk loci for polycystic ovary syndrome. *Nat Genet.* 44:1020–1025.
- Steffens M, Leu C, Ruppert AK, Zara F, Striano P, Robbiano A, Capovilla G, Tinuper P, Gambardella A, Bianchi A, et al. 2012. Genome-wide association analysis of genetic generalized epilepsies implicates susceptibility loci at 1q43, 2p16.1, 2q22.3 and 17q21.32. *Hum Mol Genet.* 21:5359–5372.
- Suzuki H. 1975. Discoveries of the fossil man from Okinawa Island. *J Anthropol Soc Nippon.* 83:113–124. (in Japanese with English abstract)
- Tajima A, Hayami M, Tokunaga K, Juji T, Matsuo M, Marzuki S, Omoto K, Horai S. 2004. Genetic origins of the Ainu inferred from combined DNA analyses of maternal and paternal lineages. *J Hum Genet.* 49: 187–193.
- Tajima A, Pan IH, Fucharoen G, Fucharoen S, Matsuo M, Tokunaga K, Juji T, Hayami M, Omoto K, Horai S. 2002. Three major lineages of Asian Y chromosomes: implications for the peopling of east and southeast Asia. *Hum Genet.* 110:80–88.
- Takamiya H. 2006. An unusual case? Hunter-gatherer adaptations to an island environment: a case study from Okinawa, Japan. *J Island Coast Archaeol.* 1:49–66.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* 28:2731–2739.
- Toma T, Hanihara T, Sunakawa H, Haneji K, Ishida H. 2007. Metric dental diversity of Ryukyu Islanders: a comparative study among Ryukyu and other Asian populations. *Anthropol Sci.* 115: 119–131.
- Wigginton JE, Cutler DJ, Abecasis GR. 2005. A note on exact tests of Hardy-Weinberg equilibrium. *Am J Hum Genet.* 76:887–893.
- Yamaguchi-Kabata Y, Nakazono K, Takahashi A, Saito S, Hosono N, Kubo M, Nakamura Y, Kamatani N. 2008. Japanese population structure, based on SNP genotypes from 7003 individuals compared to other ethnic groups: effects on population-based association studies. *Am J Hum Genet.* 83:445–456.