

1 **Article (Discoveries section)**

2 **Title: Genome-wide SNP analysis reveals population structure and demographic**
3 **history of the Ryukyu Islanders in the southern part of the Japanese Archipelago**

4
5 **Authors:**

6 Takehiro Sato¹, Shigeki Nakagome², Chiaki Watanabe¹, Kyoko Yamaguchi¹, Akira
7 Kawaguchi¹, Kae Koganebuchi^{3,4}, Kuniaki Haneji⁵, Tetsutaro Yamaguchi⁶, Tsunehiko
8 Hanihara⁷, Ken Yamamoto⁸, Hajime Ishida¹, Shuhei Mano², Ryosuke Kimura^{1*}, Hiroki
9 Oota^{7*}

10
11 **Affiliations:**

12 ¹Department of Human Biology and Anatomy, Graduate School of Medicine, University
13 of the Ryukyus, Okinawa, 903-0215, Japan

14 ²Department of Mathematical Analysis and Statistical Inference, The Institute of
15 Statistical Mathematics, Tokyo, 190-8562, Japan

16 ³Department of Integrated Biosciences, Graduate School of Frontier Sciences, The
17 University of Tokyo, Chiba, 277-8562, Japan

18 ⁴Department of Human Anatomy, Kitasato University Graduate School of Medical
19 Sciences, Kanagawa, 252-0373, Japan

20 ⁵Medical Corporation Hayama Group, Haneji's Clinic Dentistry and Surgery, Okinawa,
21 906-0012, Japan

22 ⁶Department of Orthodontics, School of Dentistry, Showa University, Tokyo, 145-8515,
23 Japan

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

1 252-0374, Japan

2 ⁸Division of Genome Analysis, Research Center for Genetic Information, Medical
3 Institute of Bioregulation, Kyushu University, Fukuoka, 812-8582, Japan

4

5 **Abbreviated title:**

6 Population structure of the Ryukyu Islanders

7

8 *Corresponding authors:

9 Ryosuke Kimura

10 Department of Human Biology and Anatomy, Graduate School of Medicine, University
11 of the Ryukyus, Okinawa, 903-0215, Japan

12 E-mail: rkimura@med.u-ryukyu.ac.jp

13

14 Hiroki Oota

15 Department of Anatomy, Kitasato University School of Medicine, Kanagawa, 252-0374,
16 Japan

17 E-mail: hiroki_oota@med.kitasato-u.ac.jp

18

1 **Abstract**

2

3 The Ryukyu Islands are located to the southwest of the Japanese Archipelago.
4 Archaeological evidence has revealed the existence of prehistoric cultural
5 differentiation between the northern Ryukyu islands of Amami and Okinawa, and the
6 southern Ryukyu islands of Miyako and Yaeyama. To examine a genetic subdivision in
7 the Ryukyu Islands, we conducted genome-wide SNP typing of inhabitants from the
8 Okinawa Islands, the Miyako Islands, and the Yaeyama Islands. Principal component
9 and cluster analyses revealed genetic differentiation between the inhabitants of the
10 Okinawa and Miyako, as well as more recent migration from the Okinawa Islands to the
11 Yaeyama Islands. No genetic affinity was observed between aboriginal Taiwanese and
12 any of the Ryukyu populations. The genetic differentiation observed between the
13 inhabitants of the Okinawa Islands and the Miyako Islands is likely to have arisen due
14 to genetic drift rather than admixture with people from neighboring regions. Based on
15 the observed genetic differences, the inhabitants of Okinawa and Miyako islands are
16 considered to have separated during the Holocene. These findings suggest that the
17 Pleistocene inhabitants, whose bones have been found on the southern Ryukyu Islands,
18 did not make a major genetic contribution, if any, to the present-day inhabitants of the
19 southern Ryukyu Islands.

20

21 **Key words:** Japanese Archipelago, Ryukyu Islanders, population structure,
22 demographic history, genome-wide SNPs

23

24

1 **Introduction**

2

3 The Ryukyu Islands located to the southwest of the Japanese archipelago, are distributed
4 over an area of 840 km from the north to the south, and 610 km from the east to the
5 west. The archipelago consists primarily of the Amami Islands, Okinawa Islands,
6 Miyako Islands, and the Yaeyama Islands (Fig. 1). Of the numerous straits in the
7 Ryukyu Archipelago, the Kerama Gap between Okinawa-jima (the largest of the
8 Okinawa Islands) and Miyako-jima (the largest of the Miyako Islands) is the largest at
9 260 km wide (here, “-jima” is used for “Island” to distinguish it from “Islands”)
10 (Kishimoto 2003). This strait divides the Ryukyu Islands into the northern Ryukyu and
11 southern Ryukyu islands.

12 Archaeological findings show that two different cultures developed in the
13 northern and southern Ryukyus during prehistoric times (Asato 2003). The distributions
14 of these cultures reflected the geography of the region, with the Amami Islands and the
15 Okinawa Islands belonging to the northern Ryukyu cultural group, and the Miyako
16 Islands and the Yaeyama Islands belonging to the southern Ryukyu cultural group
17 (Fig.1). In northern Ryukyu, Shellmound Period have extended from 6,700 to 1,000
18 before present (BP) (Asato 2003). Based on pottery characteristics, the culture of this
19 period is considered to have been similar to that of the Jomon period on the mainland of
20 Japan (i.e. the islands of Hokkaido, Honshu, Shikoku, Kyushu: Fig. 1). These apparent
21 similarities prompted some archaeologists to propose that northern Ryukyu was settled
22 by immigrants from mainland Japan (Asato 2003). Two prehistoric periods have been
23 characterized in southern Ryukyu: the Shimotabaru Period, which spanned from 4,500
24 to 3,000 BP and was characterized by a specific style of pottery, and the Aceramic

1 Period from 2,500 to 800 BP, during which no pottery was produced in this region.
2 Since no affinity has been observed between the northern and the southern Ryukyu
3 cultures in these prehistoric periods, some archaeologists have hypothesized that there
4 was no human migration between the northern and southern Ryukyu during these
5 periods (Asato 2003). However, in the subsequent Gusuku Period (800-600 BP), the
6 cultures of both the northern and southern Ryukyus were unified after rice cultivation
7 was introduced to the Ryukyu Islands.

8 The peopling of the Ryukyu Islands is considered central to understanding the
9 origin of the Japanese people. The dual structure model of Japanese origin, which is
10 based on cranial and dental morphology, hypothesizes that (1) the first inhabitants of the
11 Japanese Archipelago originated from somewhere in Southeast Asia, and that initial
12 colonization occurred during the Upper Paleolithic age (probably through the Ryukyu
13 Islands). These colonialists are considered to have given rise to the native
14 hunter-gatherers in the Japanese Archipelago, or Jomon people. (2) This initial wave of
15 migration was then followed by a second wave of migration from northeast Asia, which
16 occurred during in and after the Aeneolithic Yayoi age. Finally, (3) the populations of
17 both lineages (Jomon and Yayoi people) gradually mixed with each other on mainland
18 Japan, except for in Hokkaido. According to this hypothesis, the marked similarities
19 between the Ainu and the Ryukyu Islanders are due to their common origin from the
20 Jomon people (Hanihara 1991). Of the numerous anthropological studies that have been
21 conducted to date to verify the dual structure model, most concur with the idea that the
22 Japanese have a dual structure and that the Ryukyu Islanders share a common origin
23 with the Ainu (Hanihara 1993; Matsumura 1994; Hammer and Horai 1995; Horai et al.
24 1996; Omoto and Saitou 1997; Tajima et al. 2002; Tajima et al. 2004; Matsumura and

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

7 Indeed, relatively few studies have examined the population structure of the
8 inhabitants of the Ryukyu Islands to date. However, of the studies that have been
9 undertaken, some have reported the existence of local variations in morphological and
10 genetic characteristics (Naito 1976; Omoto et al. 1976; Haneji et al. 2007; Toma et al.
11 2007; Matsukusa et al. 2010). However, the population structure within the Ryukyu
12 Islands remains unclear due to a lack of comprehensive data and due to
13 incompatibilities between the findings of previous studies. In the present study, we
14 performed a comprehensive genetic analysis based on genome-wide SNP data of the
15 Ryukyu Islanders. Specifically, we included individuals from the Okinawa Islands, the
16 Miyako Islands, and the Yaeyama Islands. We also discuss the demographic history of
17 the Ryukyu Islanders, and provide evidence for genetic differentiation between the
18 inhabitants of the Okinawa and Miyako island groups.

19

20

21 **Results and Discussion**

22

23 **Population structure in the Ryukyu Islanders**

24

1 Figure 2 shows the results of the principal component analysis (PCA) using 540,451
2 SNPs shared among datasets including individuals from the Okinawa Islands (OK),
3 Miyako Islands (MY), Yaeyama Islands (YE), mainland Japan (MJ), Korea (KR) and
4 two HapMap populations (Chinese Han in Beijing, CHB; Japanese in Tokyo, JPT). Four
5 main clusters (Ryukyu Islanders, mainland Japanese, Koreans, and Han Chinese) were
6 observed in PC1. The observation that almost all Japanese individuals were grouped
7 into two main clusters (Ryukyu and mainland Japan clusters) was consistent with a
8 previous study (Yamaguchi-Kabata et al. 2008). Under the dual structure model of
9 Japanese origin (Hanihara 1991), PC1 can be interpreted as being associated with the
10 Jomon and Yayoi components of individual ancestry. Meanwhile, a biplot of PC1 versus
11 PC2 produced a “U” shape (Fig. 2a), implying that PC2 is unlikely to directly reflect
12 any demographic events; this pattern is remarkably similar to a pattern of mathematical
13 artifacts that has previously been observed in a simulation of one-dimensional stepping
14 stone-like habitats (Novembre and Stephens 2008). The OK and MY individuals were
15 separated into two clusters (Okinawa and Miyako clusters) in PC 3, and YE individuals
16 were plotted between the Okinawa and Miyako clusters (Fig. 2b).

17 We then performed ADMIXTURE analysis by adding four HapMap
18 populations (CHB; JPT; Yorba in Ibadan, YRI; Utah residents with ancestry from
19 northern and western Europe, CEU). When the assumed number of ancestral
20 populations (K) was 4, the resulting bar plot revealed that the East Asian populations
21 were primarily formed by admixture of two ancestral populations at different
22 proportions (shown as green and purple in the bar plot in Fig. 3a). In addition, at $K = 4$,
23 the minimum value of the cross-validation error was observed (Fig. 3b). The relative
24 proportions of the two ancestral populations (green vs. purple) at $K = 4$ were

1 approximately 1:9 in Koreans and 5:5 in mainland Japanese, which also supports the
2 dual structure model of Japanese origins (Hanihara 1991); the two components for East
3 Asians (green and purple) may not correspond exactly to the genetic components of the
4 Jomon and Yayoi people, respectively, but they are at least likely to be associated with
5 the population structure formed by the admixture of two ancestral populations. A
6 component specific to the MY samples (shown in orange) appeared at $K = 6$, indicating
7 that the inhabitants of the Okinawa Islands and of the Miyako Islands are subdivided.
8 For the level of individual ancestry at $K = 6$, YE individuals were more similar to the
9 OK individuals compared to the MY individuals.

10 Our PCA and ADMIXTURE analysis showed that MY individuals can be
11 genetically differentiated from OK individuals and that YE individuals are genetically
12 closer to OK individuals than they are to MY individuals (Fig. 2b and Fig. 3), even
13 though the Miyako Islands are geographically located between the others (Fig.1). It has
14 also previously been suggested that people from the Miyako Islands were genetically
15 isolated from the other Ryukyu populations, based on somatometry (Naito 1976), as
16 well as 13 genetic markers associated with blood types, red cell enzymes, and serum
17 proteins (Omoto et al. 1976). Results of the PCA and ADMIXTURE analysis may
18 suggest that people of the Yaeyama Islands were affected by recent large gene flows
19 from the Okinawa Islands; according to historical records, people migrated from the
20 Okinawa Islands to the Yaeyama Islands to develop farm land, fisheries, and the
21 infrastructure during the 19th and 20th centuries (Miki 2010). These recent migrations
22 probably resulted in the genetic relationships between people of the Okinawa, Miyako,
23 and Yaeyama Islands being inconsistent with their geographic locations.

24

1 **The genetic relationship between the Ryukyu Islanders and aboriginal Taiwanese**

2
3 What caused the genetic differentiation of people of the Miyako Islands from people of
4 the Okinawa Islands? Possible explanations are isolation (genetic drift) and/or gene
5 flow from a neighboring population, possibly from Taiwan. To examine the possibility
6 of gene flow from the aboriginal Taiwanese to people of the Miyako Islands, we
7 performed PCA using 12,488 SNPs overlapping between data for aboriginal Taiwanese,
8 Ami (AM) and Atyal (AT), from the HUGO Pan-Asian SNP Consortium database and
9 our dataset. The PCA result demonstrated no genetic affinities between the aboriginal
10 Taiwanese and any of the Ryukyu individuals (Fig. 2c). This suggests that genetic
11 differentiation between the people of the Miyako Islands and of the Okinawa Islands is
12 unlikely to have resulted from gene flow from Taiwan, but rather due to genetic drift
13 after the Miyako ancestors migrated from the northern Ryukyu to the southern Ryukyu.
14 A previous study based on mtDNA, Y chromosomal STRs, and autosomal STRs also
15 showed that there was no evidence of any contribution from the aboriginal Taiwanese
16 populations to the gene pool of the Ryukyu Islanders (Matsukusa et al. 2010). Similarly,
17 using a large number of SNP markers, this study verified that there is a clear genetic gap
18 between Taiwan and the Ryukyu Islands.

20 **Demographic reconstruction among the Ryukyu Islanders**

21
22 To visualize the phylogenetic relationships between the Ryukyu Islanders and the
23 neighboring populations, we constructed a neighbor-joining tree and a neighbor-net
24 based on Nei's minimum distances between the populations (Table 1, Fig. 4, and

1 Supplementary Fig. S3). The NJ tree clearly showed the extremely low inter-population
2 genetic variances compared to intra-population genetic variances, as described in the
3 previous study (Rosenberg et al. 2002). Reticulations observed in the neighbor-net
4 indicate the occurrence of previous admixture events in these populations, suggesting
5 that mainland Japanese and Koreans are composed of two admixed ancestral
6 populations. In the neighbor-net, OK, MY, and YE formed a monophyletic Ryukyu
7 cluster. The relatively long branch to the Ryukyu cluster in the network explains the low
8 genetic diversity among the individuals composing the Ryukyu cluster. Especially, the
9 external branch to MK was the longest in the Ryukyu cluster, reflecting the low genetic
10 diversity within the Miyako population (Fig. 4).

11 We employed several assumptions to infer demographic parameters from
12 genetic diversities. To exclude SNPs that had been arisen in Eurasian populations after
13 the out of Africa migrations, only polymorphic SNP markers in YRI were used. In this
14 study, CEU, CHB, OK, and MY were included in the analysis; however, potentially
15 admixed populations (YE, MJ, and KR) were excluded from this analysis. Further, for
16 simplicity, we assumed that there was no gene flow among these populations after
17 divergence. Based on the newly calculated genetic differences within and between
18 populations (D_X and D_{XY} values), an NJ tree was constructed (Fig. 5a) and average gene
19 diversities at different times were estimated (H_{AS-EU} , H_{RY-CH} , H_{OK-MY} , and H_{OK}).

20 We considered two simple demographic models: (I) a constant population size
21 model (Fig. 5b) and (II) an exponential population growth model (Fig. 5c). The
22 exponential growth model assumes that the Ryukyu populations and Han Chinese
23 expanded at the same growth rate (α). To estimate focal parameters, we fixed the
24 divergence time between Asian and European ancestries ($T_{CH-EU} = 36,000$ BP) in both

1 models, and the initial population size of the Asian ancestry ($N_{CH0} = 1,200$) in model II,
2 all of which were estimated in a previous study (Gronau et al. 2011). Parameter
3 estimates for each model are shown in table 2. In the constant population size model, we
4 estimated the approximate divergence time between the inhabitants of the Okinawa
5 Islands and the Han Chinese (T_{OK-CH}) to be 3,000 BP (assuming 25 years/generation).
6 The divergence time between the inhabitants of the Okinawa Islands and the Miyako
7 Islands (T_{OK-MY}) was dated to approximately 130 BP. In this model, the effective
8 population size of Han Chinese from T_{CH-EU} to the present (N_{CH}) was estimated to be
9 approximately 5,000, while that in the Okinawa Islands from T_{OK-CH} to the present (N_{OK})
10 was estimated to be approximately 2,600. In the exponential growth model, T_{OK-CH} and
11 T_{OK-MY} were dated to approximately 15,000 BP and 1,600 BP, respectively. The initial
12 population size in the Okinawa Islands (N_{OK0}) was considered to be 6,426. These results
13 showed that the estimated divergence times vary markedly depending on the model.
14 When we performed the analysis using $T_{CH-EU} = 26,400$ BP and $N_{CH0} = 590$ estimated
15 by Gutenkunst et al. (2009), instead of the estimates by Gronau et al. (2011), more
16 recent (in model I) and older (in model II) divergence times were obtained, respectively
17 (Table 2).

18 In the estimation method above, parameter estimates were drastically affected
19 by the prefixed parameters and, in addition, isolation-with-migration models were not
20 applicable. To compensate for these drawbacks, we performed a kernel approximate
21 Bayesian computation (kernel-ABC) (Fukumizu et al. 2013; Nakagome et al. 2013),
22 considering possible migrations between populations and without requiring
23 pre-estimated demographic parameters. As the summary statistics in this analysis, we
24 used the frequency spectrum of haplotypes that are defined by SNPs shared between

1 populations, which represent patterns of linkage disequilibrium. Under the model
2 assuming no migration between people of the Okinawa Islands and the Miyako Islands
3 (Fig. 6a), the estimated divergence time varied from $1,800 \pm 150$ to $2,825 \pm 650$ BP
4 depending on the prior distribution of the divergence time (Table 3). When we
5 employed a model with migration (migration rate $m = 0.001$) (Fig. 6b), the divergence
6 time between people of the Okinawa Islands and the Miyako Islands was estimated to
7 be $2,175 \pm 225 - 6,925 \pm 1,925$ BP (Table 3). These results suggested that the
8 divergence event between the inhabitants of the Okinawa Islands and the Miyako
9 Islands dates from the Holocene, even when the effect of migration at $m = 0.001$ is
10 considered. Estimated effective population sizes in the kernel-ABC are shown in
11 Supplementary Table S3.

12 Human remains dating from the Pleistocene have been discovered on the
13 Ryukyu Islands. One of the most famous Pleistocene sites is the Minatogawa Fissure on
14 Okinawa-jima (northern Ryukyu) that was discovered by an amateur archaeologist,
15 Seiho Oyama. A total of four skeletons have been excavated at the site, and two
16 charcoal fragments found in the vicinity of human remains were dated to $16,600 \pm 300$
17 and 18250 ± 650 BP by ^{14}C dating (Matsu'ura 1999; Hashimoto et al. 2002). On
18 Miyako-jima (southern Ryukyu), 26,000-year-old human remains were excavated from
19 the Pinza-Abu Cave site (Sakura 1985). More recently, a 20,000-year-old human bone
20 fragment was excavated from the Shiraho-Saonetabaru Cave site on Ishigaki-jima, the
21 main island of the Yaeyama Islands (Nakagawa et al. 2010). Our estimates for the
22 parameters T_{RK-CH} and T_{OK-MY} were more recent than the ages of these Pleistocene
23 humans. If these ancient people had made major contributions to the gene pools of the
24 modern inhabitants of the islands where they were excavated, then the estimates of

1 T_{RK-CH} and T_{OK-MY} should have been dated within the Pleistocene. Therefore, our
2 findings suggest that the first inhabitants of the Ryukyu Islands during the Pleistocene
3 were unlikely to have contributed substantially, or at all, to the modern inhabitants of
4 the island. Indeed, our analysis suggested that more recent divergence would
5 sufficiently explain the population structure of the modern inhabitants of the Ryukyu
6 Islands. A previous study proposed that the first sustained colonization of the Ryukyu
7 Islands occurred during the Holocene (Takamiya 2006), and our results strongly support
8 this hypothesis.

9 The present study proposed that divergence of the modern Ryukyu Islanders
10 occurred during the Holocene, even though the T_{OK-MY} estimated in the present study
11 varied depending on which model was employed, and that this divergence could not be
12 assigned to a particular period, such as Simotabaru (4,500-3,000 BP), Aceramic
13 (2500-800 BP) or Gusuku (800-600 BP) Periods. This uncertainty in the estimation is
14 due to limitations in demographic reconstruction using SNP genotyping data.
15 Whole-genome sequencing data will enable us to more accurately estimate demographic
16 parameters. Therefore, accumulation of whole-genome sequencing data is considered
17 necessary to derive a more detailed demographic history of the Ryukyu Islanders.

18

19

20 **Materials and Methods**

21

22 **Samples**

23

24 A total of 449 Japanese individuals (OK, n = 281; MY, n=44; YE, n= 24; MJ, n=100)

1 participated in the present study; some of the MY and YE individuals are the same as
2 those examined in Matsukusa et al. (2010). Written informed consent was obtained from
3 all of the participants, and only individuals with four grandparents of the same origin
4 were used in this study. Blood or saliva samples were collected from the participants.
5 This study was approved by the ethical committees at University of the Ryukyus,
6 Showa University, and Kitasato University. Geographic locations of each population
7 sampled in this study are shown in Figure 1.

8 For comparison with our data, we used genotype data for KR individuals (n =
9 224) (Haga et al. 2013). We also downloaded genotype data for unrelated individuals of
10 JPT (n = 113), CHB (n = 137), CEU (n = 112), and YRI (n = 147) from release 23 of the
11 HapMap project phase 3 database (The International HapMap 3 Consortium 2010), and
12 for aboriginal Taiwanese (Ami, n = 10; Atyal, n = 10) from the HUGO Pan-Asian SNP
13 database (PASNP) (The HUGO Pan-Asian SNP Consortium 2009; Ngamphiw et al.
14 2011).

15

16 **DNA extraction and genotyping of genome-wide SNPs**

17

18 Genomic DNA was extracted from blood and saliva samples using Genra Puregen
19 Blood kit (Qiagen, Germantown, MD, USA) or DNA Extractor WB kit (Wako, Osaka,
20 Japan), and genotyped for 729,172 SNPs using HumanOmniExpress BeadChip system
21 (Illumina, San Diego, CA, USA). Quality control of the genotype data was carried out
22 using the following criteria. After samples with call rates lower than 0.98 were excluded,
23 those SNPs having call rates lower than 0.98 in the remaining samples were excluded.
24 To ensure that only unrelated individuals were used for subsequent analyses, one

1 individual in a pair who had a proportion of shared identical-by-descent ($\hat{\pi}$) higher
2 than 0.125 (corresponding to a third-degree relative) were excluded from the dataset.
3 We also excluded individuals with an inbreeding coefficient (F) higher than 0.0625
4 (corresponding to mating between first cousins). In addition, we performed
5 Hardy-Weinberg equilibrium test for each sample set (OK, MY, YE, MJ, KR, JPT, CHB,
6 YRI, and CEU) using the exact test which adequately controls for type I errors in both
7 large and small samples and is computationally efficient (Wigginton et al. 2005). Those
8 SNPs with P values lower than 1.0×10^{-6} were excluded from the dataset. All quality
9 control procedures were performed using PLINK 1.07 (Purcell et al. 2007). As a result,
10 694,887 SNPs for OK, 694,952 SNPs for MY and YE, 694,944 SNPs for MJ, and
11 694,902 SNPs for KR passed the quality control criteria described above. Of the
12 694,852 overlapping SNPs in the five populations, 602,317 were polymorphic
13 (Supplementary Table S1).

14 To unify strands between genotype data generated by the Illumina
15 HumanOmniExpress BeadChip and obtained from HapMap database, strand checks of
16 our genotype data were performed using “check_strands.py,” a utility program for
17 BEAGLE (Browning and Browning 2009). HapMap3 JPT and CHB data were used as a
18 reference panel for the strand checks. As part of this process, a total of 154,401 SNPs
19 were excluded due to missing HapMap3 data or strand ambiguities. Finally, 540,451
20 autosomal SNPs from 281 OK, 43 MY, 23 YE, 100 MJ, 221 KR, 113 JPT, 137 CHB,
21 111 CEU, and 139 YRI individuals passed the quality control and strand check
22 procedures.

23

24 **Detection of population outliers**

1

2 Detection of population outliers is performed generally according to results of the
3 principal component analysis (PCA) or multidimensional scaling (MDS) in recent
4 genome-wide association studies (GWASs) (Diskin et al. 2012; Nakamura et al. 2012;
5 Shi et al. 2012; Steffens et al. 2012). However, it is difficult to judge objectively how
6 many dimensions should be used to detect population outliers. In the present study, we
7 used a dimension-free procedure based on genotype frequencies to exclude population
8 outliers from a population without using data of any other population.

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

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

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

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

16 The expectation of Q can be written as

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

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

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

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

1 When the SNP density in the real data is high, then the assumption that all the
2 SNPs are independent does not hold. However, the real distribution of L , at least
3 approximately, follows a normal distribution (Supplementary Fig. S1). Therefore, we
4 can convert L to Z scores according to the observed variance in the population to obtain
5 P values. Here, we set the significance level at 0.05 and used Bonferroni correction for
6 multiple comparison of individuals.

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

17

18 **Principal component analysis**

19

20 To investigate the relationships among the East Asian populations, PCA was performed
21 using the EIGENSOFT 3.0 package (Patterson et al. 2006) with 540,451 SNP genotype
22 data for the OK, MY, YE, MJ, KR, JPT, and CHB individuals. To investigate the genetic
23 relationships between the Ryukyu Islanders and the aboriginal Taiwanese, we also
24 included Ami (AX-AM) and Atyal (AX-AT) genotype data from the HUGO Pan-Asian

1 SNP (PASNP) database (The HUGO Pan-Asian SNP Consortium 2009; Ngamphiw et al.
 2 2011). By combining all these datasets, a total of 12,484 overlapping SNPs were
 3 obtained and subject to PCA.

4

5 **Clustering analysis for estimating individual ancestry**

6

7 To estimate individual ancestry, we subjected nine sample sets (OK, MY, YE, MJ, KR,
 8 JPT, CHB, CEU, and YRI) to maximum likelihood clustering analysis using the
 9 ADMIXTURE 1.21 (Alexander et al. 2009). We ran ADMIXTURE from $K = 3$ to $K = 7$
 10 and calculated the cross validation error for each number of K . Five-fold cross
 11 validations were performed for each number of K .

12

13 **Phylogenetic analysis based on the genetic distance**

14

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

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

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

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

18 where L is the number of loci, n_X and n_Y are the numbers of individuals sampled in
 19 population X and Y , respectively, and a_X and a_Y are the allele counts observed in

1 population X and Y , respectively. A neighbor-joining (NJ) tree (Saitou and Nei 1987)
2 and neighbor-net (Bryant and Moulton 2004) were constructed based on D_X , D_{XY} , and
3 D_m using SPLITS TREE 4.12 (Huson and Bryant 2006) and MEGA 5.2 (Tamura et al.
4 2011).

5

6 **Estimation of divergence time between populations based on genetic diversity**

7

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

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

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

17 Using this approach, we tested two simple demographic models. The first
18 model assumed a constant population size for the Asian populations (model I; Fig 5b),
19 and the second assumed an exponential growth of the Asian populations (model II; Fig
20 5c). The divergence time between the Asian and European populations (T_{CH-EU}) was
21 assumed to be 24,600 or 36,000 BP in both models, and the initial population size of
22 Asians (N_{CH0}) was taken as 590 or 1,200 in model II, as these values were estimated

1 previously (Gutenkunst et al. 2009; Gronau et al. 2011). In model II, we further
2 assumed that every Asaian population has the same population growth rate α , and that
3 N_t can be defined by $N_t = N_0 \times \alpha^t$ (where N_0 is the initial population size). The
4 harmonic mean of N_t corresponds to the effective population size over a period of time.
5 By conducting coalescent simulations, we confirmed that this estimation method was
6 robust to ascertainment bias of SNP data (Supplementary Text S1, Supplementary Table
7 S2, and Supplementary Fig. S4).

9 **Bayesian inference of population demographic history**

10
11 We also inferred the demographic history of people from the Okinawa Islands and the
12 Miyako Islands by kernel approximate Bayesian computation (kernel-ABC) (Fukumizu
13 et al. 2013; Nakagome et al. 2013). ABC is a statistical framework that can be used to
14 approximate a posterior estimate without a likelihood function, and data are summarized
15 as a set of summary statistics (\mathbf{S}). The advantage of kernel-ABC is that computing costs
16 are independent of the dimension of data, which means that high-dimensional summary
17 statistics can be incorporated into the ABC analyses. We used the haplotype frequency
18 spectrum (HFS) as summary statistics, as these can account for recombination patterns
19 at a genomic segment. In this study, the genome-wide SNP data were summarized into
20 two-dimensional HFS (\mathbf{S}_{2D-HFS}). A detailed explanation of the kernel-ABC
21 methodology is given in Supplementary Materials (Supplementary Text S2,
22 Supplementary Fig. S5, S6, and S7).

1 **Acknowledgments**

2

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

14

1 **References**

2

3 Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in
4 unrelated individuals. *Genome Res* 19:1655-1664.

5 Asato S. 2003. Prehistory and proto-history in the Ryukyu island chain. History of
6 Okinawa Prefecture 2: Archaeology. Naha: Okinawa Prefecture.

7 Browning BL, Browning SR. 2009. A unified approach to genotype imputation and
8 haplotype-phase inference for large data sets of trios and unrelated individuals.
9 *Am J Hum Genet* 84:210-223.

10 Bryant D, Moulton V. 2004. Neighbor-net: an agglomerative method for the
11 construction of phylogenetic networks. *Mol Biol Evol* 21:255-265.

12 Diskin SJ, Capasso M, Schnepf RW, Cole KA, Attiyeh EF, Hou C, Diamond M,
13 Carpenter EL, Winter C, Lee H, et al. 2012. Common variation at 6q16 within
14 HACE1 and LIN28B influences susceptibility to neuroblastoma. *Nat Genet*
15 44:1126-1130.

16 Fukumine T, Hanihara T, Nishime A, Ishida H. 2006. Nonmetric cranial variation of
17 early modern human skeletal remains from Kumejima, Okinawa and the
18 peopling of the Ryukyu Islands. *Anthropol Sci* 114:141-151.

19 Fukumizu K, Song L, Gretton A. 2013. Kernel Bayes' rule: Bayesian inference with
20 positive definite kernels. *J Mach Learn Res* 14:3753-3783.

21 Gronau I, Hubisz MJ, Gulko B, Danko CG, Siepel A. 2011. Bayesian inference of
22 ancient human demography from individual genome sequences. *Nat Genet*
23 43:1031-1034.

24 Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. 2009. Inferring the

- 1 joint demographic history of multiple populations from multidimensional SNP
2 frequency data. *PLoS Genet* 5:e1000695.
- 3 Haga S, Nakaoka H, Yamaguchi T, Yamamoto K, Kim YI, Samoto H, Ohno T,
4 Katayama K, Ishida H, Park SB, et al. 2013. A genome-wide association study
5 of third molar agenesis in Japanese and Korean populations. *J Hum Genet*
6 58:799-803.
- 7 Hammer M, Horai S. 1995. Y chromosomal DNA variation and the peopling of Japan.
8 *Am J Hum Genet* 56:951-962.
- 9 Haneji K, Hanihara T, Sunakawa H, Toma T, Ishida H. 2007. Non-metric dental
10 variation of Sakishima Islanders, Okinawa, Japan: a comparative study among
11 Sakishima and neighboring populations. *Anthropol Sci* 115:35-45.
- 12 Hanihara K. 1991. Dual structure model for the population history of the Japanese. *Jpn*
13 *Rev* 2:1-33.
- 14 Hanihara T. 1993. Craniofacial features of southeast Asians and Jomonese: a
15 reconsideration of their microevolution since the Late Pleistocene. *Anthropol Sci*
16 101:25-46.
- 17 Hashimoto M, Yahagi K, Tsujimoto H, Saito T, Yoshida K. 2002. Chapter 8. In:
18 Gushichan Village Board of Education, editor. Minatogawa Fissure Site. p.
19 89-106.
- 20 Horai S, Murayama K, Hayasaka K, Matsubayashi S, Hattori Y, Fucharoen G, Harihara
21 S, Park KS, Omoto K, Pan IH. 1996. mtDNA polymorphism in East Asian
22 Populations, with special reference to the peopling of Japan. *Am J Hum Genet*
23 59:579-590.
- 24 Huson DH, Bryant D. 2006. Application of phylogenetic networks in evolutionary

- 1 studies. *Mol Biol Evol* 23:254-267.
- 2 Ishida H, Hanihara T, Kondo O, Fukumine T. 2009. Craniometric divergence history of
3 the Japanese populations. *Anthropol Sci* 117:147-156.
- 4 Jinam T, Nishida N, Hirai M, Kawamura S, Oota H, Umetsu K, Kimura R, Ohashi J,
5 Tajima A, Yamamoto T, et al. 2012. The history of human populations in the
6 Japanese Archipelago inferred from genome-wide SNP data with a special
7 reference to the Ainu and the Ryukyuan populations. *J Hum Genet* 57:787-795.
- 8 Kishimoto Y. 2003. Nature and antiquity of Rukyu island chain. *History of Okinawa*
9 *Prefecture 2: Archaeology*. Naha: Okinawa Prefecture.
- 10 Koganebuchi K, Katsumura T, Nakagome S, Ishida H, Kawamura S, Oota H,
11 Consortium TAADR. 2012. Autosomal and Y-chromosomal STR markers reveal
12 a close relationship between Hokkaido Ainu and Ryukyu islanders. *Anthropol*
13 *Sci* 120:199-208.
- 14 Matsu'ura S. 1999. A chronological review of Pleistocene human remains from the
15 Japanese Archipelago. In: K Omoto, editor. *Interdisciplinary Perspectives on the*
16 *Origins of the Japanese*. Kyoto: International Research Center for Japanese
17 *Studies*. p. 181-197.
- 18 Matsukusa H, Oota H, Haneji K, Toma T, Kawamura S, Ishida H. 2010. A genetic
19 analysis of the Sakishima islanders reveals no relationship with Taiwan
20 aborigines but shared ancestry with Ainu and main-island Japanese. *Am J Phys*
21 *Anthropol* 142:211-223.
- 22 Matsumura H. 1994. A microevolutional history of the Japanese people from a dental
23 characteristics perspective. *Anthropol Sci* 102:93-118.
- 24 Matsumura H, Hudson MJ. 2005. Dental perspectives on the population history of

- 1 Southeast Asia. *Am J Phys Anthropol* 127:182-209.
- 2 Miki K. 2010. The lineage of United States of Yaeyama. Ishigaki: Nanzansha.
- 3 Naito Y. 1976. People in Okinawa prefecture from the point of view of somatometry.
4 OKINAWA-Nature-Culture-Society-. Tokyo: Kobundo. p. 77-86.
- 5 Nakagawa R, Doi N, Nishioka Y, Nunami S, Yamauchi H, Fujita M, Yamazaki S,
6 Yamamoto M, Katagiri C, Mukai H, et al. 2010. Pleistocene human remains
7 from Shiraho-Saonetabaru Cave on Ishigaki Island, Okinawa, Japan, and their
8 radiocarbon dating. *Anthropol Sci* 118:173-183.
- 9 Nakagome S, Fukumizu K, Mano S. 2013. Kernel approximate Bayesian computation
10 in population genetic inferences. *Stat Appl Genet Mol Biol* 12:667-678.
- 11 Nakamura M, Nishida N, Kawashima M, Aiba Y, Tanaka A, Yasunami M, Nakamura H,
12 Komori A, Nakamuta M, Zeniya M, et al. 2012. Genome-wide association study
13 identifies TNFSF15 and POU2AF1 as susceptibility loci for primary biliary
14 cirrhosis in the Japanese population. *Am J Hum Genet* 91:721-728.
- 15 Nei M. 1972. Genetic distance between populations. *Am Nat* 106.
- 16 Ngamphiw C, Assawamakin A, Xu S, Shaw PJ, Yang JO, Ghang H, Bhak J, Liu E,
17 Tongshima S. 2011. PanSNPdb: the Pan-Asian SNP genotyping database. *PLoS*
18 *One* 6:e21451.
- 19 Novembre J, Stephens M. 2008. Interpreting principal component analyses of spatial
20 population genetic variation. *Nat Genet* 40:646-649.
- 21 Omoto K, Misawa S, G. I. 1976. People in Okinawa prefecture from the point of view
22 of genetic markers which is associated with blood. In: Association of Nine
23 Learned Societies, editor. OKINAWA-Nature-Culture-Society-. Tokyo:
24 Kobundo. p. 93-111.

- 1 Omoto K, Saitou N. 1997. Genetic origins of the Japanese: a partial support for the dual
2 structure hypothesis. *Am J Phys Anthropol* 102:437-446.
- 3 Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. *PLoS*
4 *Genet* 2:e190.
- 5 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar
6 P, de Bakker PI, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome
7 association and population-based linkage analyses. *Am J Hum Genet*
8 81:559-575.
- 9 Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA,
10 Feldman MW. 2002. Genetic structure of human populations. *Science*
11 298:2381-2385.
- 12 Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing
13 phylogenetic trees. *Mol Biol Evol* 4:406-425.
- 14 Sakura H. 1985. Pleistocene human fossil remains from Pinza-Abu (Goat Cave),
15 Miyako Island, Okinawa, Japan. *Reports on Excavation of the Pinza-Abu*
16 *Cave*:161-176.
- 17 Shi Y, Zhao H, Cao Y, Yang D, Li Z, Zhang B, Liang X, Li T, Chen J, Shen J, et al. 2012.
18 Genome-wide association study identifies eight new risk loci for polycystic
19 ovary syndrome. *Nat Genet* 44:1020-1025.
- 20 Steffens M, Leu C, Ruppert AK, Zara F, Striano P, Robbiano A, Capovilla G, Tinuper P,
21 Gambardella A, Bianchi A, et al. 2012. Genome-wide association analysis of
22 genetic generalized epilepsies implicates susceptibility loci at 1q43, 2p16.1,
23 2q22.3 and 17q21.32. *Hum Mol Genet* 21:5359-5372.
- 24 Tajima A, Hayami M, Tokunaga K, Juji T, Matsuo M, Marzuki S, Omoto K, Horai S.

- 1 2004. Genetic origins of the Ainu inferred from combined DNA analyses of
2 maternal and paternal lineages. *J Hum Genet* 49:187-193.
- 3 Tajima A, Pan IH, Fucharoen G, Fucharoen S, Matsuo M, Tokunaga K, Juji T, Hayami
4 M, Omoto K, Horai S. 2002. Three major lineages of Asian Y chromosomes:
5 implications for the peopling of east and southeast Asia. *Hum Genet* 110:80-88.
- 6 Takamiya H. 2006. An unusual case? Hunter-gatherer adptations to an island
7 environment: A case study from Okinawa, Japan. *JICA* 1:49-66.
- 8 Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5:
9 molecular evolutionary genetics analysis using maximum likelihood,
10 evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*
11 28:2731-2739.
- 12 The HUGO Pan-Asian SNP Consortium 2009. Mapping human genetic diversity in Asia.
13 *Science* 326:1541-1545.
- 14 The International HapMap 3 Consortium 2010. Integrating common and rare genetic
15 variation in diverse human populations. *Nature* 467:52-58.
- 16 Toma T, Hanihara T, Sunakawa H, Haneji K, Ishida H. 2007. Metric dental diversity of
17 Ryukyu Islanders: a comparative study among Ryukyu and other Asian
18 populations. *Anthropol Sci* 115:119-131.
- 19 Wigginton JE, Cutler DJ, Abecasis GR. 2005. A note on exact tests of Hardy-Weinberg
20 equilibrium. *Am J Hum Genet* 76:887-893.
- 21 Yamaguchi-Kabata Y, Nakazono K, Takahashi A, Saito S, Hosono N, Kubo M,
22 Nakamura Y, Kamatani N. 2008. Japanese population structure, based on SNP
23 genotypes from 7003 individuals compared to other ethnic groups: effects on
24 population-based association studies. *Am J Hum Genet* 83:445-456.

1 **Figure Legends**

2

3 **Fig. 1** Map of the Japanese Archipelago and neighboring regions. The Japanese
4 Archipelago consists of four main islands (Hokkaido, Honshu, Shikoku, and Kyushu),
5 the Ryukyu Islands, and numerous other islands. The Ryukyu Islands consist of the
6 Amami Islands, the Okinawa Islands, the Miyako Islands, and the Yaeyama Islands.

7

8 **Fig. 2** The results of principal component analysis using genotype data for East Asian
9 populations. **(a)** PC1 versus PC2, and **(b)** PC1 versus PC3 for 540,451 SNPs. **(c)** PC1
10 versus PC2 for 12,484 SNPs, which include aboriginal Taiwanese samples (Ami and
11 Atyal). Population codes are as follows: *OK*, Okinawa; *MY*, Miyako; *YE*, Yaeyama; *MJ*,
12 mainland Japanese; *KR*, Koreans; *JPT*, Japanese in Tokyo; *CHB*, Chinese Han in
13 Beijing; *AM*, Ami; *AT*, Atyal.

14

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

18

19 **Fig. 4** A neighbor-net based on Nei's minimum distance (D_m) among eight populations.
20 The data of MJ individuals in the present study were merged with JPT.

21

22 **Fig. 5** Models for estimating demographic parameters based on genetic diversity. T_{X-Y} is
23 the divergence time between populations X and Y . N_X and N_{X0} indicate the effective and
24 initial population sizes of population X , respectively. **(a)** Neighbor-joining tree based on

1 511,913 SNPs that are polymorphic in YRI. Branch lengths from the split between
2 populations X and Y to each external node denotes a half of the gene diversity in the
3 ancestral population at the divergence time (H_{X-Y}). The height of the triangle on the
4 external node of population X denotes a half of the gene diversity in the present
5 population X (H_X). (b) Model I assumes a constant population size. (c) Model II assumes
6 exponential growth of Asian populations. In this model, population growth rate of the
7 Ryukyu Islanders and CHB is α . The value of T_{AS-EU} that used in this study is an
8 estimate by Gutenkunst et al. (2009) or Gronau et al. (2011).

9

10 **Fig. 6** Schematic models employed for the kernel-ABC. T_{OK-MY} is the divergence time
11 between the people from the Okinawa Islands and the Miyako Islands. N_{OK} and N_{MY} are
12 the effective population sizes of the Okinawa Islands and the Miyako Islands,
13 respectively. (a) No migration model. (b) Isolation-with-migration model. In the
14 migration model, we assumed that migration rate (m) was 0.001.

15

1 **Table 1.** Genetic distances within each population and between each pair of populations.

	Genetic distances ^a							
	OK	MY	YE	ML+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

2 ^a 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.

3 Unbiased estimates were calculated for D_X due to the small sample size of the Miyako and Yaeyama populations.

4

1 **Table 2.** Parameter estimates based on genetic diversity

	Model I		Model II	
Fixed parameters				
T_{AS-EU}	1056 gens (26400 BP)	1440 gens (36000 BP)	1056 gens (26400 BP)	1440 gens (36000 BP)
N_{CHO}	-	-	590	1200
Reference	Gutenkunst et al. (2009)	Gronau et al. (2011)	Gutenkunst et al. (2009)	Gronau et al. (2011)
Estimated parameters				
N_{CH}	3534	4818	3534*	4818*
α	(1)	(1)	1.0057	1.0027
T_{RK-CH}	89 gens (2225 BP)	131 gens (3275 BP)	623 gens (15575 BP)	605 gens (15125 BP)
N_{OKO}	1935	2639	3728	6426
N_{OK}	1935	2639	13541*	15649*
T_{OK-MY}	4 gens (100 BP)	5 gens (135 BP)	159 gens (3975 BP)	63 gens (1575 BP)
N_{MYO}	1569	2140	8190	4522
N_{MY}	1569	2140	12439*	4928*

2 *Harmonic mean of population sizes is shown.

3

1 **Table 3.** Posterior estimates of the divergence time between people of the Okinawa Islands and the Miyako Islands
 2 by kernel-ABC.

Models		Prior distributions ^a		
		(μ : mean, μ^2 : variance)		
		$\mu = 50$	$\mu = 100$	$\mu = 400$
No migration	Mean generations (years) ^b	72 (1,800)	81 (2,025)	113 (2,825)
	S.D. of generations (years) ^b	6 (150)	7 (175)	26 (650)
Migration with m = 0.001	Mean generations (years) ^b	87 (2,175)	104 (2,600)	277 (6,925)
	S.D. of generations (years) ^b	9 (225)	10 (250)	77 (1,925)

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

4 ^bThe mean and S.D. are calculated from 265 posterior means of the divergence time (assuming 25 years/generation).

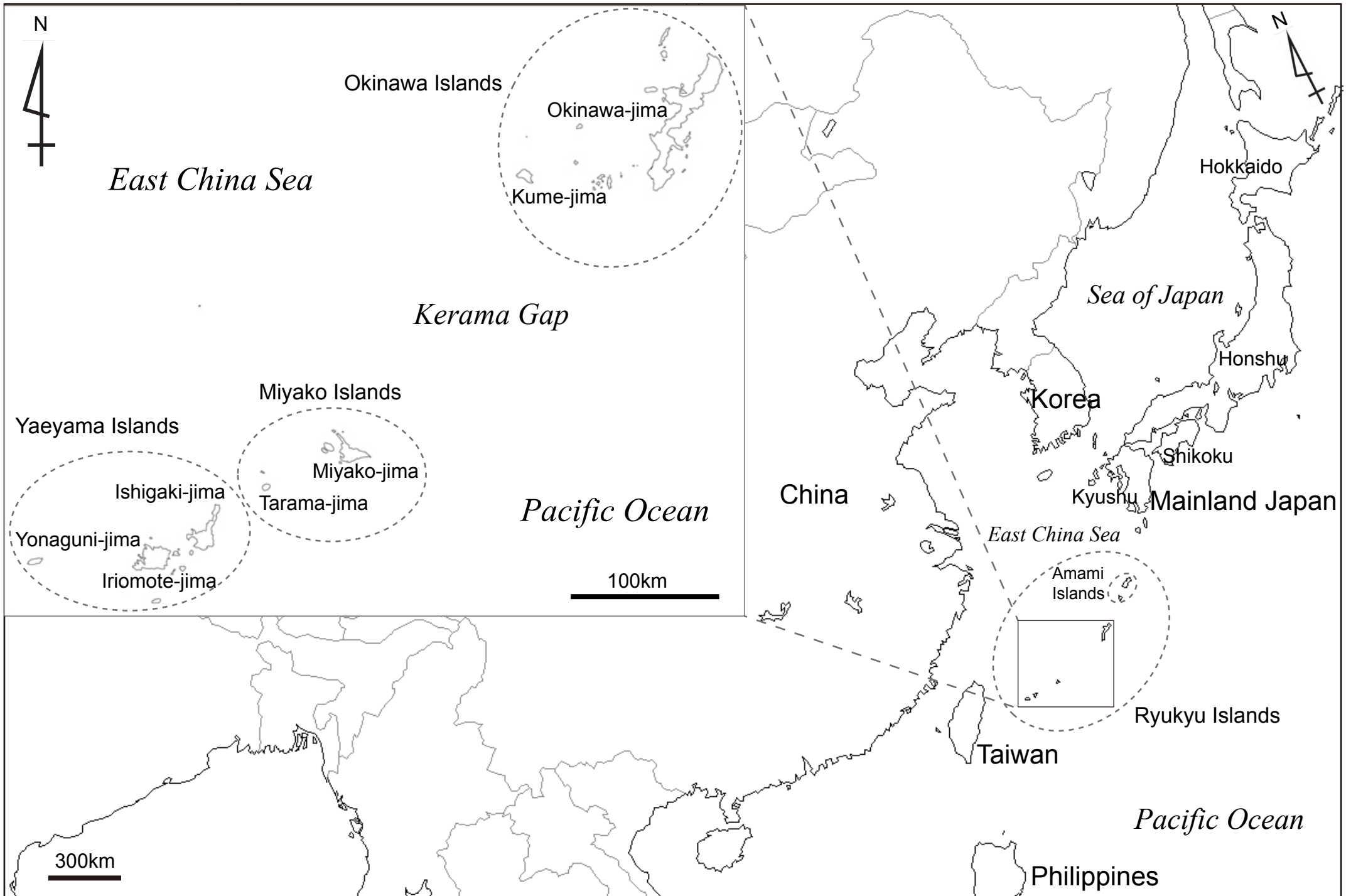


Figure 1

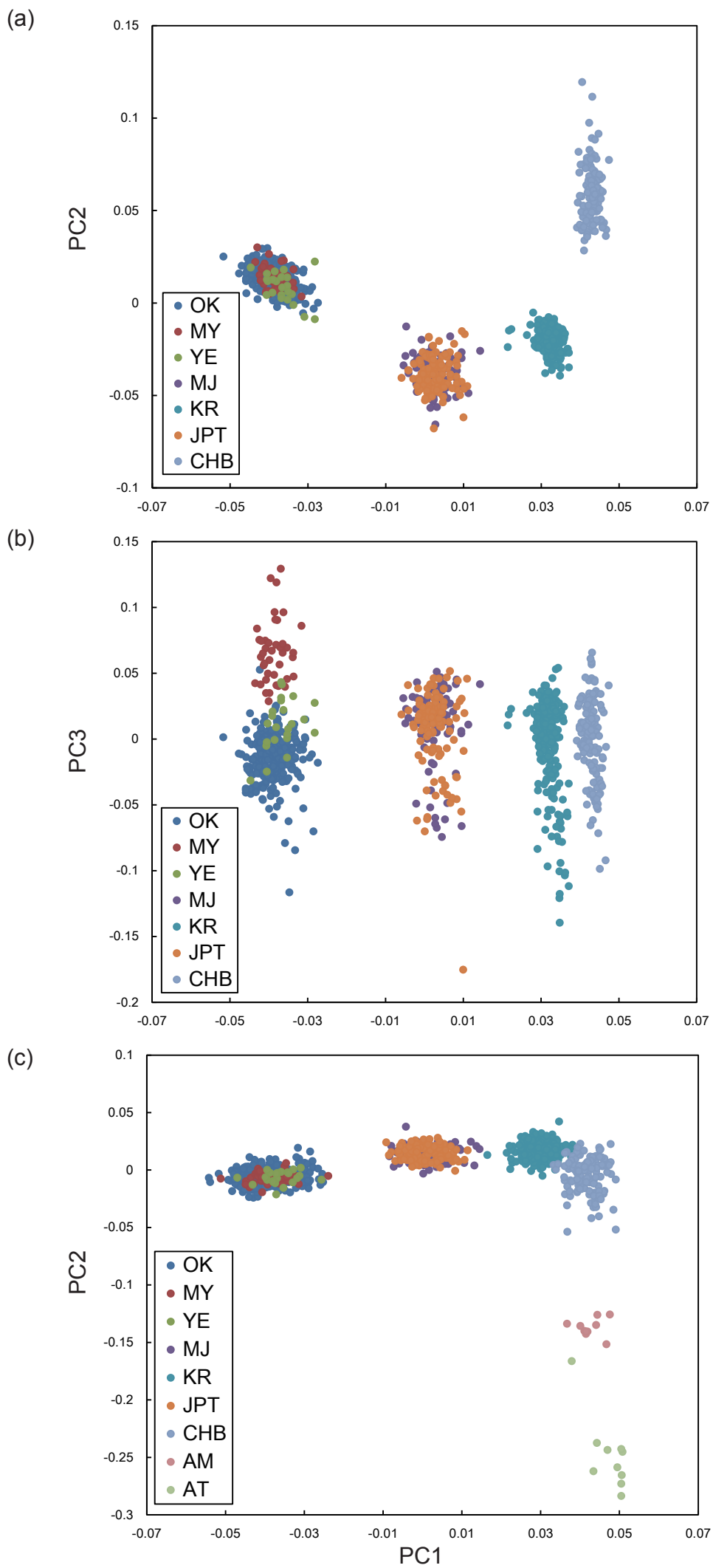


Figure 2

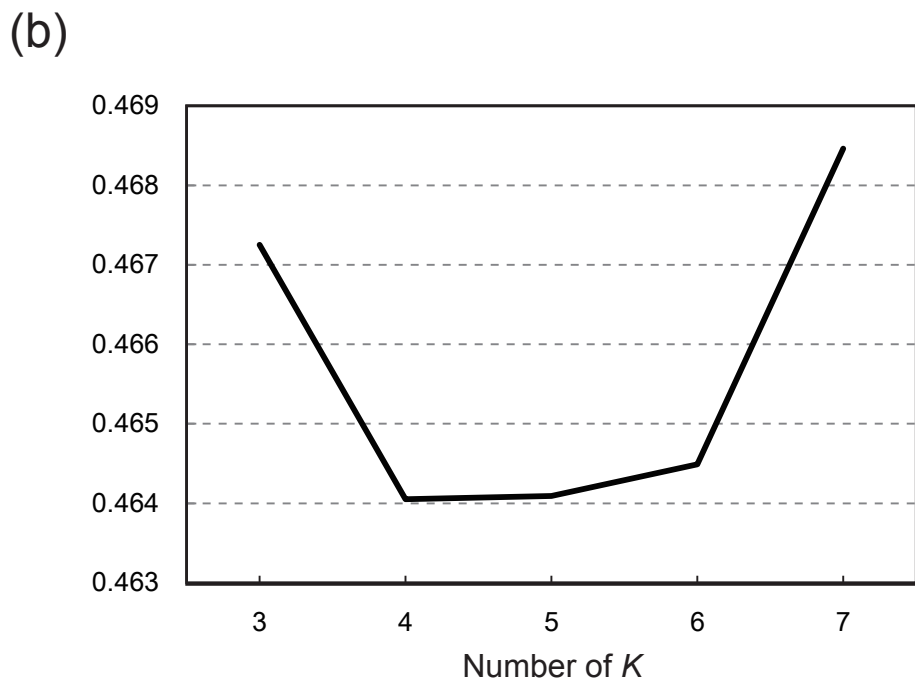
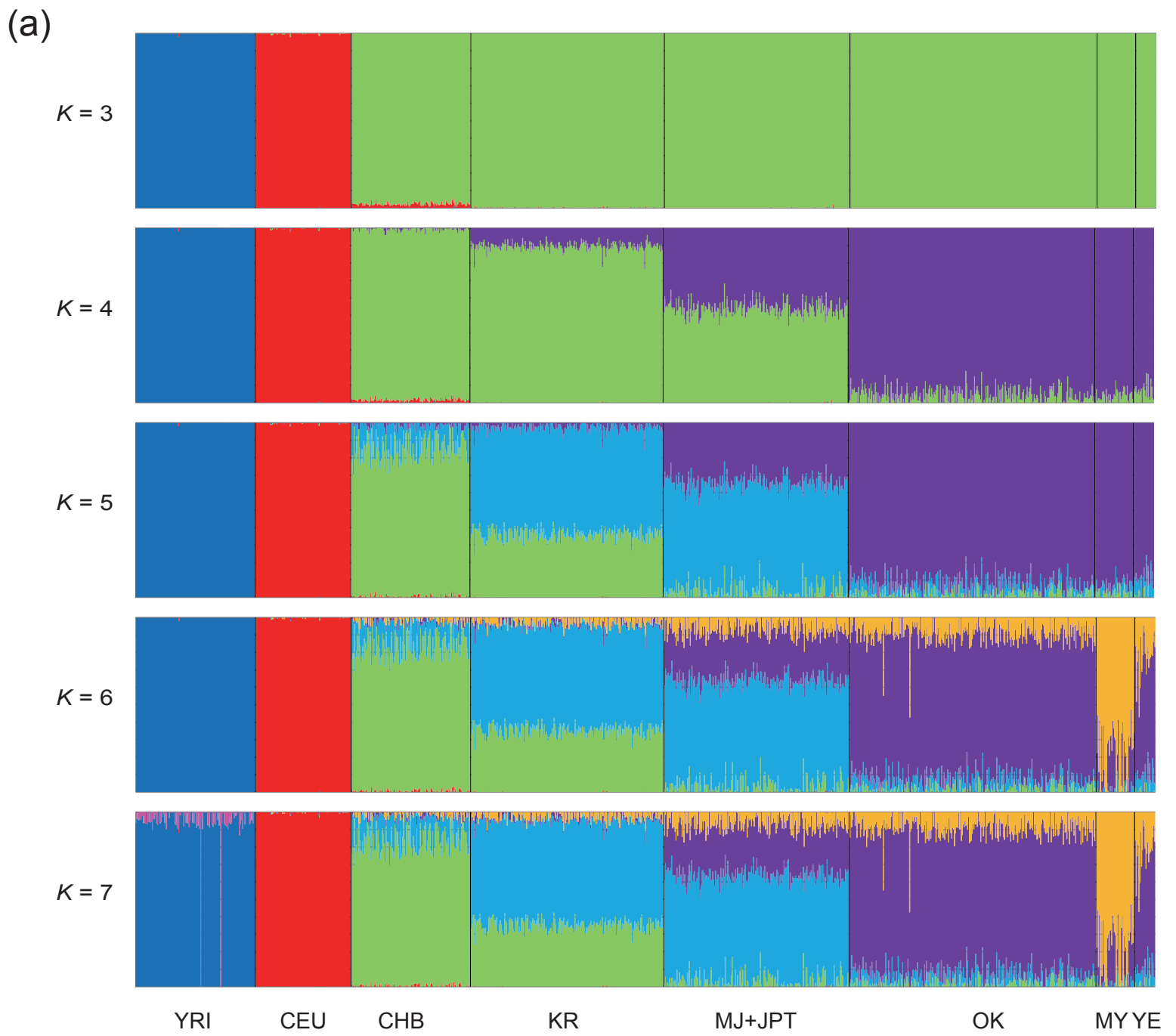
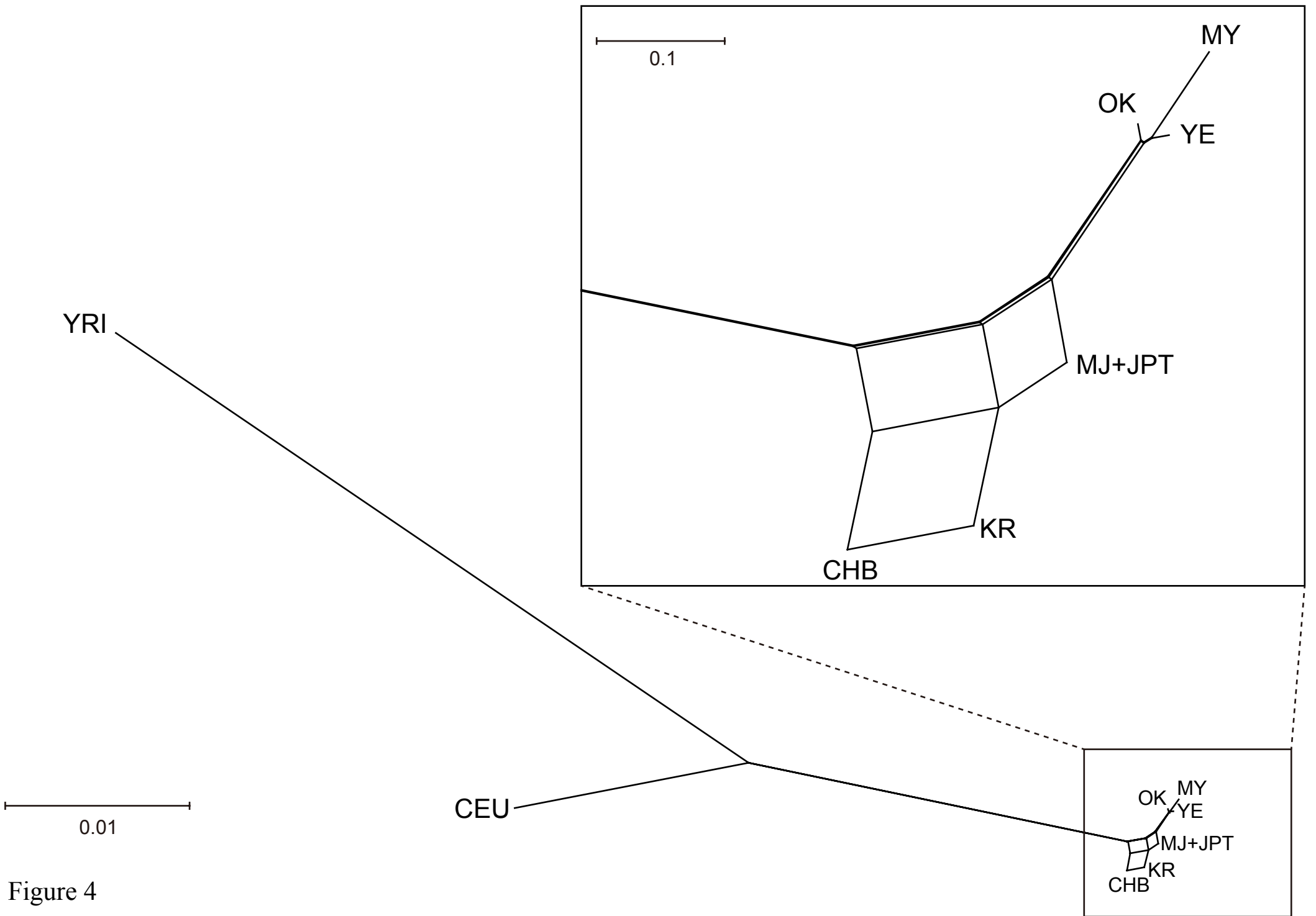


Figure 3



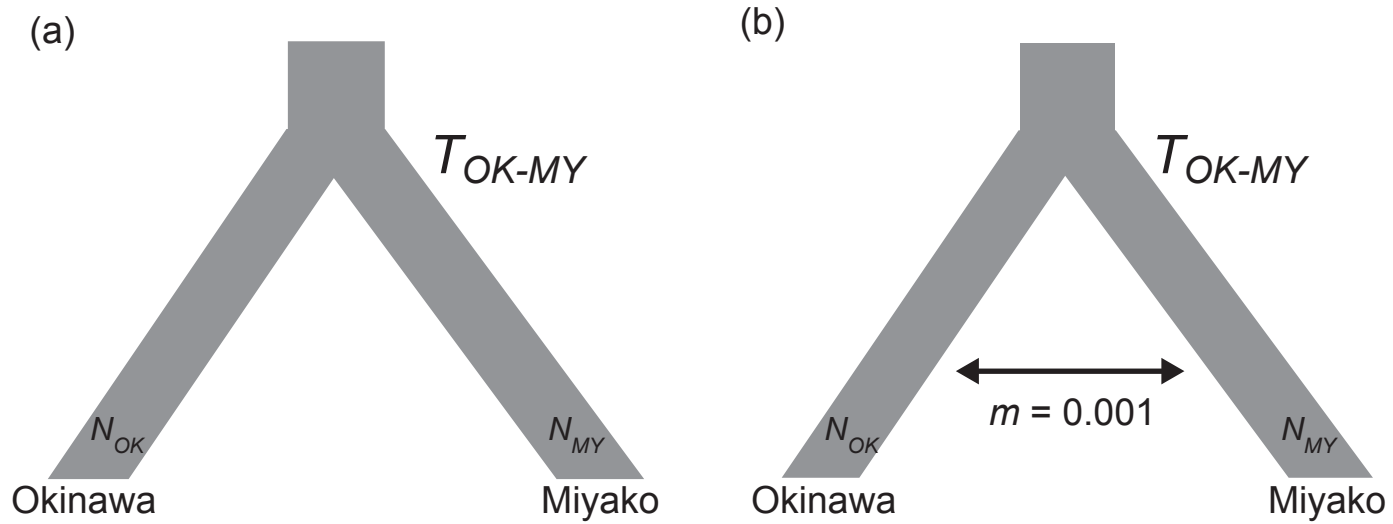


Figure 6