ARTICLE



Approximate Bayesian computation analysis of EST-associated microsatellites indicates that the broadleaved evergreen tree *Castanopsis sieboldii* survived the Last Glacial Maximum in multiple refugia in Japan

K. Aoki¹ · I. Tamaki² · K. Nakao³ · S. Ueno⁴ · T. Kamijo⁵ · H. Setoguchi⁶ · N. Murakami⁷ · M. Kato⁶ · Y. Tsumura^{4,5}

Received: 12 October 2017 / Revised: 10 June 2018 / Accepted: 19 June 2018 / Published online: 2 August 2018 © The Genetics Society 2018

Abstract

Climatic changes have played major roles in plants' evolutionary history. Glacial oscillations have been particularly important, but some of their effects on plants' populations are poorly understood, including the numbers and locations of refugia in Asian warm temperate zones. In the present study, we investigated the demographic history of the broadleaved evergreen tree species *Castanopsis sieboldii* (Fagaceae) during the last glacial period in Japan. We used approximate Bayesian computation (ABC) for model comparison and parameter estimation for the demographic modeling using 27 EST-associated microsatellites. We also performed the species distribution modeling (SDM). The results strongly support a demographic scenario that the Ryukyu Islands and the western parts in the main islands (Kyushu and western Shikoku) were derived from separate refugia and the eastern parts in the main islands and the Japan Sea groups were diverged from the western parts prior to the coldest stage of the Last Glacial Maximum (LGM). Our data indicate that multiple refugia survived at least one in the Ryukyu Islands, and the other three regions of the western and eastern parts and around the Japan Sea of the main islands of Japan during the LGM. The SDM analysis also suggests the potential habitats under LGM climate conditions were mainly located along the Pacific Ocean side of the coastal region. Our ABC-based study helps efforts resolve the demographic history of a dominant species in warm temperate broadleaved forests during and after the last glacial period, which provides a basic model for future phylogeographical studies using this approach.

Introduction

Glacial cycles during the Quaternary period have played important roles in shaping the diversity patterns of

Electronic supplementary material The online version of this article (https://doi.org/10.1038/s41437-018-0123-9) contains supplementary material, which is available to authorized users.

- ¹ United Graduate School of Child Development, Osaka University, Suita, Osaka 565-0871, Japan
- ² Gifu Academy of Forest Science and Culture, Mino, Gifu 501-3714, Japan
- ³ Kansai Research Center, Forestry and Forest Products Research Institute, Forest Research and Management Organization, Kyoto 612-0855, Japan

SPRINGER NATURE

contemporary populations (Hewitt 2000, 2004), even in the warm temperate zone where the land was not covered with ice sheets (Tsukada 1974; Minato and Ijiri 1976). Severe climatic oscillations drove repeated north–south and highland–lowland migrations of communities and ecosystems, thereby strongly affecting the genetic structure and the distribution of organisms they hosted (reviewed in Taberlet et al. 1998; Hewitt 1999, 2004). Thus, current geographic patterns of genetic diversity within species reflect their

- ⁴ Department of Forest Molecular Genetics and Biotechnology, Forestry and Forest Products Research Institute, Forest Research and Management Organization, Tsukuba, Ibaraki 305-8687, Japan
- ⁵ Faculty of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki 305-8572, Japan
- ⁶ Graduate School of Human and Environmental Studies, Kyoto University, Kyoto 606-8501, Japan
- ⁷ Makino Herbarium, Tokyo Metropolitan University, Hachioji, Tokyo 192-0397, Japan

S. Ueno saueno@ffpri.affrc.go.jp

responses to expansions and contractions of their habitats associated with the glacial cycles, particularly the numbers, sizes, and locations of refugia; the timing and severity of bottlenecks; and the rates of recolonization.

Molecular phylogeographic studies of the genetic structure of extant populations have provided detailed insights into apparent glacial and postglacial distributions of numerous species (Avise 2000). Key issues for phylogeographic analyses include the numbers and locations of refugia, which are still uncertain, despite intensive recent investigations, especially in Europe (Provan and Bennett 2008; Svenning et al. 2008; Stewart et al. 2010) and North America (Soltis et al. 2006; Shafer et al. 2010). In Europe, Svenning et al. (2008) found indications suggesting that temperate trees were probably largely confined to the traditionally recognized southern refugia in and around Mediterranean lowlands, but several boreal tree species may have been widespread not only in these southern areas but also in several areas of northern Europe during the Last Glacial Maximum (LGM). Existence of previously unknown, or cryptic, northern refugia have also been proposed (Birks and Willis 2008; Provan and Bennett 2008; Rull 2009; Stewart et al. 2010; Parducci et al. 2012; Tzedakis et al. 2013; de Lafontaine et al. 2014), based on the phylogeographic patterns of several species together with evidence from pollen and plant macrofossils. The eastern edge of East Asia, including East China, Japan, and the Korean Peninsula, harbors diverse of the temperate flora, and the effects of the locations of refugia throughout Quaternary glacial cycles on the current genetic structure are important (Qiu et al. 2011). For East Asian temperate plants, the population fragmentation occurred caused by fluctuations in sea level throughout the Quaternary, and these taxa restricted to distinct refugia (Qiu et al. 2011). Recent phylogeographic studies of the warm temperate trees in this region also detected the multiple localized refugia (Huang et al. 2002; Cheng et al. 2005, 2006; Wu et al. 2006; Lee et al. 2013, 2014; Xu et al. 2015), though most of them were based on a single locus (chloroplast DNA). Tests of refugia models using multilocus genetic data based on estimated impacts of demographic events and the associated genetic processes to patterns of species' genetic variation can help to understand more detailed history during and after the glacial periods.

Advances in population genetics have led to the development of statistical approaches based on coalescent models for estimating population genetic parameters, reconstructing demographic histories, and testing related hypotheses using Bayesian or likelihood-based frameworks (Knowles and Maddison 2002). In these approaches, a hypothesis is considered more probable than postulated alternatives if all the available data are more likely under the applied model and underlying assumptions (Hickerson et al. 2010). In addition, a more versatile tool for phylogeographic analysis, termed the approximate Bayesian computation (ABC), has been recently introduced. This "likelihood-free" method relatively easily enables the formulation of model-based inferences in a Bayesian setting for evaluating the demographic or evolutionary scenarios (Pritchard et al. 1999). Because it relies on only summary statistics (Beaumont et al. 2002) rather than evaluations of the overall probability of the data, we can make models flexibly to conform to the situations and compare them without much computational effort. ABC has been successfully used very recently to estimate the demographic parameters relating to various complex demographic scenarios and to assess the effects of the past climatic oscillations (Gao et al. 2012; Budde et al. 2013; Li et al. 2013; Liu et al. 2014a, b; Louis et al. 2014; Wang et al. 2016; Yoichi et al. 2016; Chen et al. 2017; Ren et al. 2017).

Following biogeographic scenarios have been proposed to describe the glacial and postglacial history of Japan's warm temperate ecosystems based on fossilized pollen data and reconstructions of past climate. The pollen record indicates that refugial populations of broadleaved evergreen forests existed in the warmer southern end of Kyushu in the main islands as well as in the Ryukyu Islands and migrated northward from these southern refugia after the LGM (Tsukada 1974; Matsuoka and Miyoshi 1998). However, based on historical climate data and tolerance of cool temperature of the component species of these forests, some ecologists have proposed that these forests might also have survived in multiple refugia along the Pacific coasts of the main islands as well as on the southern end of Kyushu during the glacial periods (Maeda 1980; Kamei and Research Group for the Biogeography from Würm Glacial 1981; Nakanishi 1996; Hattori 2002). However, it has been difficult to resolve the detailed demographic history of the plant species currently inhabiting the broadleaved evergreen forests, especially Castanopsis-type forest, because of the limited fossilized pollen records and of the relatively slow molecular evolution of chloroplast DNA in plants (Wolfe et al. 1987; Chase et al. 1993; Aoki et al. 2005) and the extremely low levels of intraspecific variation in the chloroplast DNA of Japanese broadleaved evergreen species (Aoki et al. 2003, 2004a, 2016). Application of the ABC method to reconstruct the demography of the dominant species in the type of the forest can help to understand detailed history from the information of the locations of refugia, the diverged time and the population size changes.

In the present study, we examined the demographic history of *Castanopsis sieboldii* (Fagaceae), the dominant



Fig. 1 Locations of the 33 sampled *Castanopsis sieboldii* populations. Numbers correspond to the population numbers in Table S1. The dotted line indicates the coastline during the LGM, about 18,000–24,000 years ago

tree species in Japan's broadleaved evergreen forests, by conducting the ABC method partially using the expressed sequence tag (EST)-associated microsatellite genetic variation analyzed in Aoki et al. (2014). For this, we first used an ABC framework to identify the most likely model of demographic history among the several alternative scenarios based on Bayesian clustering of current genetic variation. We then applied the model to estimate the demographic parameters describing effective population sizes of the species at several points in time and the recent migrations in their demographic histories. Finally, we combined the results of the species distribution modeling (SDM) with the ABC analyses and those of previous phylogeographic studies of other organisms inhabiting the same climatic zone to reconstruct the glacial and postglacial history of Japan's broadleaved evergreen forests.

Materials and methods

Plant materials

We used fresh or silica-gel-dried leaves of 792 *C. sieboldii* trees collected from 33 populations (Table S1, Fig. 1) covering most of the species' natural distribution in Japan. *C. sieboldii* is represented by two varieties, var. *sieboldii* and var. *lutchuensis* and they are not distributed sympatrically; *C. sieboldii* var. *sieboldii* is found in the main islands whereas *C. sieboldii* var. *lutchuensis* in the Ryukyu Islands and Kyushu (southern regions from Amamioshima) (Yamazaki and Mashiba 1987b). These varieties have morphologically different; seed size and shape of *C. sieboldii* var. *sieboldii* var. *sieboldii* var. *sieboldii* var. *sieboldii* var. *sieboldii* var. *sieboldii* of *C. sieboldii* var. *sieboldii* however, intermediate morphologies are often observed (Yamazaki and Mashiba 1987a).

C. sieboldii var. sieboldii is sometimes codistributed with its congeneric species Castanopsis cuspidata var. cuspidata in the main islands of Japan. Hybrids with intermediate morphology sometimes occur, especially at sites where the two species coexist (Kobayashi and Hiroki 2003). Our previous study by Aoki et al. (2014) examined the epidermis of the leaves from each sampled individual, since this is the most effective way of discriminating the two species and hybrids (Yamazaki and Mashiba 1987b). In this study, we identified individuals with double epidermal cell layers as C. sieboldii var. sieboldii, with no ancestral contributions from C. cuspidata var. cuspidata. In subsequent analyses, we used the six C. sieboldii var. lutchuensis populations in the Ryukyu Islands and the 27 C. sieboldii var. sieboldii populations in the main islands, which do not contain any individuals of C. cuspidata var. cuspidata.

The value of minimum temperature of the coldest month of the altitudinal distribution of Castanopsis was estimated to be 0-1 °C (Hattori 1985). Castanopsis is insect pollinated plant species, which is difficult to detect the fossilized pollen in the past. Fossilized pollen of Castanopsis trees at Ryukyu Islands existed at the LGM (Kuroda 1998). The fossilized pollen records of *Castanopsis*-type broadleaved evergreen tree genus (i.e., Myrica and, or Podocarpus) at the LGM existed in southwestern Kyushu (at the mean frequency of several to 10%), northern Kyushu and Japan Sea (several %) as well as at Ryukyu Islands (10%) (Kuroda 1998; Takahara and Takeoka 1992). There were several fossilized pollen records of Castanea-Castanopsis (less than 1% to several %) at the LGM in Japan Sea (Takahara and Takeoka 1992; Miyoshi et al. 1999) and along the Pacific coasts (Kanauchi 2005), however, we cannot determine these pollens as Castanopsis.

Fagaceae acorns including *Castanopsis* are dispersed by transporting and caching by animals, especially jay and rodents, and jay participates in longer distance dispersals, for example, 100–1900 m (Darleyhill and Johnson 1981), 250–1000 m (Gómez 2003) and 3–500 m (Pons and Pausas 2007) than rodents, 4–37 m (Jensen and Nielsen 1986) and 30–40 m (Miyaki and Kikuzawa 1988).

EST-SSR analysis

Total DNA was extracted from each leaf sample using either the CTAB (hexadecyltrimethylammonium bromide) method (Murray and Thompson 1980) or the method of Doyle and Doyle (1987) after removing the polysaccharides using HEPES buffer (pH 8.0) (Setoguchi and Ohba 1995). We determined the genotypes of each sample based on 32 pairs of nuclear microsatellite markers (expressed sequence tags-simple sequence repeats, EST-SSRs) used in the study of Aoki et al. (2014). The DNA at each EST-SSR locus was amplified using a QIAGEN Multiplex PCR Kit following the manufacturer's recommended protocol. PCR products were detected using a PRISM 3100 sequencer in conjunction with GENESCAN software, and genotypes were scored using GENOTYPER software (both supplied by Applied Biosystems).

Choosing suitable loci for ABC analyses

Because we need to simulate neutral loci without null alleles in the ABC analysis in order to estimate the past population demography, we checked the neutrality and the frequencies of null alleles of each locus. First, we converted the genotype data from size numbers to repeat numbers; in the course of this procedure, indels were not taken into consideration. For these alleles with indels other than SSRs, we modified the allelic calls to the nearest alleles in size. Second, we estimated the frequencies of null alleles at each SSR locus within a population using INEst 2.0 (Chybicki 2014). The unique feature of this program is taking into account a possibility of inbreeding within a population during the estimation of null allele frequencies. We removed the two loci with the null allele frequencies of values >0.05. Third, we compared the distribution of the F_{ST} values over all loci to their expected distributions under an island model with the assumption of neutrality using the LOSITAN program (Antao et al. 2008), based on fdist as described by Beaumont and Nichols (1996). To calculate the approximate P values for each locus, 10,000 independent loci were generated and the simulated F_{ST} distribution was compared with the observed F_{ST} values. This made it possible to identify outliers in a one-step process by defining them as observed F_{ST} values falling outside the 99% confidence interval for the simulated group. We removed three outlier loci in the process. In total, we removed the five EST-SSR loci (two loci with high null allele frequencies and three outlier loci) in these steps from the 32 loci of which C. sieboldii genotypes were scored and used the remaining 27 loci in the subsequent analyses.

Genetic diversity

We calculated the total number of alleles (N_A), the average gene diversity within populations (H_S) (Nei 1978), the observed heterozygosity (H_O), and the F_{IS} fixation indices (Weir and Cockerham 1984) across all populations at each locus to estimate departures from the Hardy–Weinberg equilibrium (HWE). The significance of deviations of F_{IS} from zero was evaluated by permutation tests with sequential Bonferroni correction.

We calculated the average N_A , the unbiased heterozygosity (Nei 1978), and the allelic richness (El Mousadik



Fig. 2 The compared four candidate models for the population history of *Castanopsis sieboldii*. R, W, J, and E correspond to the Ryukyu Islands, the West, the Japan Sea, and the East groups, respectively, in Figs. 3 and 4

and Petit 1996), based on data from a minimum sample size of 17 per population, using MSA (Dieringer and Schlotterer 2003) and FSTAT version 2.9.3.2 (Goudet 2002) software.

Genetic structure

We used the genotype data, based on the 27 EST-SSR loci, of 792 *C. sieboldii* trees collected from 33 sites. We estimated the genetic structure among populations by constructing a neighbor-joining (NJ) tree (Saitou and Nei 1987) based on D_A distances (Nei et al. 1983) between all pairs of populations using MSA and PHYLIP (Felsenstein 1989). We used 10 *C. cuspidata* populations as outgroups that do not contain any individual of *C. sieboldii* var. *sieboldii* populations (the populations consisting of only *C. cuspidata* type of leaf epidermal cells in the study of Aoki et al. (2014)).

We used the Bayesian clustering method to elucidate the genetic structure among the C. sieboldii populations using TESS version 2.3 (François et al. 2006; Chen et al. 2007; Durand et al. 2009), because TESS may detect less clusters than STRUCTURE and is suitable for the data in the presence of clines and the sampling covered the species' geographic distribution (François et al. 2006; Chen et al. 2007; Durand et al. 2009). A total of 100 simulations were run for each value of K (2–10) with 50,000 Markov chain Monte Carlo samplings after 10,000 burn-in iterations using the admixture model. Following the manuals (Francois et al. 2006; Chen et al. 2007; Durand et al. 2009), we filtered the candidate runs by choosing to keep 20% of them with the lowest deviance information criterion values and estimated Q-matrices of each K value from a mean of the permuted matrices across replicates using CLUMPP (Jakobsson and Rosenberg 2007). The most appropriate cluster number (K)was selected using the ΔK criterion of Evanno et al. (2005).

Estimation of population demographic history

We used the four distinct genetic groups, the Ryukyu Islands, the West (western parts of the main islands along the Pacific Coast), the East (eastern parts of the main islands along the Pacific Coast), and the Japan Sea (Japan Sea side of the main islands) shown by the NJ and Bayesian clustering methods of C. sieboldii for the subsequent ABC analyses. Four probable alternative scenarios for population diversifications in four genetic groups of C. sieboldii were then constructed (Fig. 2). Model 1 assumed one refugia in the southern Kyushu as well as another in the Ryukyu Islands. This model corresponds to the most classic history supported by the fossilized pollen record of the broadleaved evergreen trees at LGM (Tsukada 1974). Therefore, we assumed that the Ryukyu Islands and the West groups were derived from the refugia and the other two groups were recently diverged from the West. Model 2 assumed that as large refugia existed in the Pacific Ocean side of the coastal region in the East as in the West and the Ryukyu Islands. This Model corresponds to the alternative history supported by the past climatic conditions and tolerance of cool temperatures of the broadleaved evergreen trees (Hattori 2002, Kamei and Research Group for the Biogeography from Würm Glacial 1981). So we modeled that the Ryukyu Islands, the West and the East groups were derived from each refugia, and the Japan Sea group was recently diverged from the West group. Model 3 was based on the information of the phylogeographic tree that was constructed by mitochondrial DNA sequence of Curculio hilgendorfi, which is the host-specific predator of C. sieboldii acorns (Aoki et al. 2008). The phylogeographic tree was chosen because the association between the host plant C. sieboldii and its hostspecific insect is known to be very tight (Aoki et al. 2005). Moreover, the geographic patterns of the genetic differentiation of these species seemed to be similar (Aoki et al. 2008, 2014). This model assumes that first, the populations of the Ryukyu Islands diversified from the other three groups, then the East diversified from the other two groups, and finally, the West and the Japan Sea diversified. Model 4 was constructed based on the result of the TESS analysis at K=2, where the West and the Japan Sea showed genetic admixture between ancestries dominated in the Ryukyu Islands and the East. We assumed that all the four groups exponentially grew with gene flow (except for the model 4) from the time after the four groups were established (from

the present to T_1). We assumed no migration in the model 4, because in this model, observed multiple ancestries in the West and the Japan Sea groups were solely explained by genetic admixture at T_1 . Symmetric migrations were considered only between the Ryukyu Island and the West, between the West and the Japan Sea, and between the West and the East groups, because these three group pairs showed relatively low F_{ST} values ($F_{ST} < 0.02$; Figure S1) and adjacently located. Moreover, when the source-sink relationships existed, asymmetric migration from sink to source in the direction of coalescence was assumed (e.g., from the Japan Sea to the West, and from the East to the West in model 1, and from the Japan Sea to the West in model 2). In the preliminary analyses, we tested without exponential growth or without migration (other than model 4) models, but supports of such models were much lower (data were not shown). We conducted coalescent simulations for each four models, compared the four models, and estimated the parameters using the ABC technique (Beaumont et al. 2002).

We used randomly selected 100 diploid individuals per group to reduce the computational costs. To check the sampling biases, we drew randomly selected 100 individuals in each group and calculated the summary statistics 100 times. The average and standard deviation among the 27 loci for the N_A , the heterozygosity, and the allele size range in each of the four groups were calculated. Pair-wise F_{ST} and F_{ST} among the total four groups for overall loci were also calculated. A total of 31 summary statistics for the observed data were calculated using arlsumstat version 3.5.2 (Excoffier and Lischer 2010). We compared them and confirmed that sampling biases were very small.

We assumed that current effective population size (N_{CUR}) was common among the four groups because the average and standard deviation values of the N_A , the heterozygosity, and the allele size range were not so different among the four groups (Figure S1), and this indicated that these four groups had experienced a similar recent population history. Moreover, each group was assumed to have shrunk at the rate G from the current to T_1 , i.e., to have grown from T_1 to the current $[N_T = N_{\text{CUR}} \exp (G \times T); N_T$ is an effective population size at time T]. Prior for G was drawn from uniform distribution from -0.01 to 0. The population size for the ancestral population among all the four groups (N_{ANC}) was also set. Priors for the N_{CUR} , and N_{ANC} were drawn from uniform distribution from 1 to 10⁴ independently. The unit of all the population sizes was scaled in the unit of the number of diploid individuals and converted into the number of haploid individuals by multiplying with two when applying a coalescent simulator. The time when the demographic events occurred (T) was scaled in the unit for generations. Priors for Ts were drawn from uniform distribution from 1 to 10^4 simultaneously, and $T_1 < T_2 < T_3$ was assumed. The number of migrants per generation (*Nm*) was used as migration parameter and the value divided by N_{CUR} (i.e., migration rate) was passed to the coalescent simulator. A single common value was used for all migration events. Prior for *Nm* was drawn from uniform distribution from 0.1 to 20. In model 4, P_{ij} indicates the proportion of genes that coalesce from the group *i* to *j* within the group *i* at T_1 . Priors for P_{WR} and P_{JR} were drawn from uniform distribution from 0 to 1 and then P_{WE} and P_{JE} were 1— P_{WR} and 1— P_{JR} , respectively. W, R, J, and E denote the West, the Ryukyu Islands, the Japan Sea, and the East groups, respectively.

Generalized stepwise mutation (GSM) model was used for the microsatellite mutation model (Estoup et al. 2002). The GSM model has two parameters, mutation rate (μ) and geometric parameter (P_{GSM}). The value of P_{GSM} can range from 0 to 1, and $P_{\text{GSM}} = 0$ indicates the strict stepwise mutation model. To accommodate the variation among loci, locus-specific μ and P_{GSM} were generated as a random effect variable. We applied the slightly modified method used by Excoffier et al. (2005). The value of μ for microsatellites is generally in the range from 5×10^{-5} to 5×10^{-3} , and several studies have used 5×10^{-4} (Estoup and Angers 1998; Estoup et al. 2002). In this study, we set the mean value of μ to 5×10^{-4} and each locus value was assumed to follow a Gamma distribution with shape and rate parameters. Prior distribution of the shape parameter was drawn from uniform distribution from 0.5 to 2.0 and that of the *rate* parameter was calculated by *shapel*(mean value of μ). Prior of the mean value of P_{GSM} among the loci was drawn from uniform distribution from 0 to 0.8, and each locus value was assumed to follow beta distribution with α and β parameters. According to Excoffier et al. (2005), $\alpha = 0.5 +$ 199 (mean P_{GSM}) and $\beta = \alpha (1 - \text{mean } P_{\text{GSM}})/(\text{mean } P_{\text{GSM}})$ were used.

Coalescent simulations were replicated 1×10^4 times in each model using program fastsimcoal2 version 2.5.2.21 (Excoffier and Foll 2011). A total of 31 summary statistics that were the same to the ones calculated by the observed dataset were calculated using arlsumstat. Since there were too many numbers of summary statistics, we consistently used machine learning approach to reduce the effects of curse of dimensionality in model choice and parameter estimation. ABC model choice via random forest (ABC-RF) implemented in the abcrf package in R was conducted (Pudlo et al. 2016). ABC-RF allows us to compare models with relatively small number of simulations and thus greatly contributes to save time. One thousand trees were constructed and the best model was selected by classification votes of random forest. Posterior probability of the best model and confusion matrix among compared models were also calculated. We replicated 2×10^6 simulations for the best model and posterior distributions of parameters were estimated by a neural network regression method



Fig. 3 Neighbor-joining tree based on Nei's genetic distances (D_A) among the 33 *Castanopsis sieboldii* populations over 27 EST-SSR loci. Numbers after the outgroup species, *C. cuspidata*, correspond to

the population numbers in the study of Aoki et al. (2014). Values in italics are percentages of 1000 bootstrap replicates supporting the respective nodes

implemented in the abc package in R (Blum and Francois 2010; Csillery et al. 2012). Tolerance value was set to $5 \times$ 10^{-4} to obtain the closest 1000 simulated datasets to the observed dataset. The number of neural networks was set to 20. When estimating the posterior distributions by the regression method, to prevent the estimated posterior falling outside the lower or upper bounds of the priors, the logit transformation option of the abc function was used. Posterior mode and the 95% highest posterior density (HPD) were calculated using density and HPDinterval functions in the base and coda packages, respectively, in R (Plummer et al. 2006). Time was converted to years, assuming that the generation time of C. sieboldii is 25 or 100 years per generation for short and long expected values. Short and long generation time values assume the average age to start reproduction (Avise 2000) and the intermediate age between the age at maturity and the maximum lifespan (Petit and Hampe 2006).

Finally, a posterior model check was performed to assess the goodness-of-fit to the data estimated from the posterior distributions (Csillery et al. 2012). A total of 1000 new genotype datasets were generated using parameters randomly sampled from the posterior distributions. Summary statistics were calculated and compared with the observed data. To validate the parameter estimation, posterior quantiles for each parameter were calculated with 1000 pseudo observed data set (Cook et al. 2006; Wegmann et al. 2010). When the parameter estimation is unbiased, distribution of posterior quantiles can be approximated by a uniform distribution. To test the deviation from a uniform distribution, a Kolmogorov–Smirnov test was conducted using R.

Species distribution modeling

To assess the refugia of *C. seiboldii* under LGM (21,000 years BP) climate conditions, the geographical distribution of potential habitats were estimated using the existing species distribution model of the target species (Nakao et al. 2014). The model was constructed with the presence/ absence records of the species as a response variables, and four climatic factors (Bio6, mean of daily minimum temperature of winter; Bio10, mean temperature of warmest quarter; and Bio19,



Fig. 4 Genetic relationships among the 33 *Castanopsis sieboldii* populations estimated using the Bayesian clustering method, TESS (Durand et al. 2009)

precipitation of coldest quarter) as explanatory variables (Nakao et al. 2014). Climatic variables at a spatial resolution of 2.5 min were used as explanatory variables, which were extracted from the WorldClim database (http://www.worldclim.org). We also predicted potential habitats under LGM climate conditions using three LGM climate data (i.e., MIROC, CCSM, MPI-ESM) from WorldClim.

Results

Genetic diversity

The EST-SSR loci were highly polymorphic (Table S2). Across all populations, the $F_{\rm IS}$ values deviated significantly and positively from zero at two loci. High levels of genetic diversity within populations were also observed in each population (mean numbers of alleles, expected heterozygosity, and allelic richness were 5.226, 0.574, and 4.888, respectively; Table S1). Of the four genetic groups of the ABC analyses, the average allelic richness among each populations was highest in the group of the Ryukyu Islands (5.337), followed by those in the West (5.191), the East (4.680), and the Japan Sea (4.659). The amount of unique alleles to each group per population was also highest in the group of the Ryukyu Islands (2.667), followed by those in the West (1.000), the East (0.846), and the Japan Sea (0.625).

Genetic structure

The NJ tree of *C. sieboldii* with *C. cuspidata* populations as outgroups contained four clusters corresponding to populations from the Ryukyu Islands and the West and the East and the Japan Sea in the main islands (Fig. 3).

The Bayesian clustering of the *C. sieboldii* populations indicated that ΔK was highest when K = 2; however, fractions of ancestry from three or four clusters (K = 3 or 4) were also found (Fig. 4). Black cluster dominated in the populations of the Ryukyu Islands (97%), while white cluster dominated in the populations of the East (98%) but was also present in the populations around the Japan Sea (88%) at K = 2. Light gray clusters at K = 3 and 4 dominated in the populations around the Japan Sea. Populations of the West and around the Japan Sea showed genetic admixture between three clusters (the West, black (47%), white (15%), and light gray (36%) and the Japan Sea, black (2%), white (6%), and light gray (75%) at K = 4).

The NJ tree and the admixture analyses using TESS of *C. sieboldii* showed that there were almost shared four distinct genetic groups (the Ryukyu Islands, the West, the East, and the Japan Sea). Only one population of No. 20 was assigned to the Japan Sea in the NJ tree, while to the West in the TESS analysis. We assigned the No. 20 population to the West genetic groups in the subsequent ABC analyses because No. 20 geographically located around the Pacific Ocean and was not suitable for including in the Japan Sea

genetic group. The populations of No. 13, 14, and 19, which located geographically intermediated between the eastern and western parts of the main islands and showed intermediate cluster between the eastern and western clusters by the NJ tree, were assigned to the East genetic groups in the subsequent ABC analyses because the black cluster (8%, 9%, and 11%, respectively at No. 13, 14, and 19) that dominated in the populations of the Ryukyu Islands was less frequent and the white cluster (64%, 65%, and 42%, respectively) that dominated in the populations of the East was more frequent in the TESS analysis.

Model comparison and parameter estimation for the demographic modeling

Prior to the explanation for the results of model comparison and parameter estimation, in this paragraph, we summarized the validations of model comparison and parameter estimation. Confusion matrix of model comparison among the four demographic models is shown in Table S3. Classification error rates of each model ranged 0.010-0.238 and overall classification error rate was 0.116 (Table 1 and S3). These low classification errors indicate that the wrong model is rarely selected with high confidence. Distributions of predicted summary statistics using posterior distribution of the best model were well congruent with the observed ones (Figure S2). Therefore, we concluded that the goodness-of-fit of the best model to the observed data was well. Posterior quantiles of each parameter were not significantly deviated from a uniform distribution except for P_{GSM} (Figure S3). Distribution of posterior quantile of P_{GSM} was biased toward 1.0. These results suggest that the parameters except for P_{GSM} can be estimated without bias, but P_{GSM} may be underestimated. However, as the P_{GSM} is a nuisance parameter, it will not become a big problem when we consider the following results.

Model 1 was selected as the best model among the compared four candidate models with a posterior probability of 0.906 (Table 1). Posterior distributions of all parameters for the best model were clearly different from prior ones (Table 2; Figure S4). All posterior distributions of parameters showed a clear single peak. Posterior modes (95% HPD) of N_{CUR} and N_{ANC} were 5463 (2349–9239) and

 Table 1 Fraction of votes, classification error rate, and posterior

 probability (PP) of the best model, which were estimated by random

 forest composed of 1000 trees based on a trained set of

 10,000 simulations

Fraction	of votes		Classification error rate	PP of the best model	
Model 1	Model 2	Model 3	Model 4		
0.634	0.212	0.152	0.002	0.116	0.906

The best model is shown in bold

 Table 2 Prior and posterior distributions of parameters in the best model (model 1)

Parameter	Prior	Posterior		
		Mode	95% HPD ^a	
			Lower	Upper
N _{CUR}	Unif (1, 10 ⁴)	5463	2349	9239
N _{ANC}	Unif (1, 10 ⁴)	1517	13	8809
$G(\times 10^{-3})$	Unif (-0.01, 0)	-1.89	-4.65	-0.39
T_1	Unif (1, 10 ⁴) ^b	431	95	2453
T_2	Unif (1, 10 ⁴) ^b	1184	110	8311
Nm	Unif (0.1, 20)	9.70	6.08	17.82
shape	Unif (0.5, 2)	0.88	0.50	1.67
mean P _{GSM}	Unif (0, 0.8)	0.217	0.077	0.352

Units of the effective population size and time parameters are the numbers of diploid individuals and generations, respectively

 N_{CUR} and N_{ANC} current and ancestral effective population sizes, respectively, *G* growth rate, *T* the time when the demographic events occurred, *Nm* the number of migrants per generation; shape and mean, P_{GSM} parameters for the mutation model of microsatellite

^a95% highest posterior density

 ${}^{\rm b}T_1 < T_2$ was assumed

1517 (13–8809), respectively. Posterior mode (95% HPD) of the effective population size at T_1 estimated using *G* was 1468 (989–3228; Figure S5).

Posterior modes (95% HPD) of T_1 and T_2 were 431 (95–2453) and 1184 (110–8311) generations ago, respectively. When the short generation time (25 years per generation) was applied, posterior modes of T_1 and T_2 were 10,775 and 29,600 years ago, respectively. When the long generation time (100 years per generation) was applied, posterior modes of T_1 and T_2 were 43,100 and 118,400 years ago, respectively.

Posterior mode (95% HPD) of *Nm* was 9.70 (6.08–17.82). According to Sewall Wright's one migrant per generation theory, one migrant is sufficient to prevent complete differentiation of the populations (Wright 1931). As *Nm* was significantly larger than 1.0, we considered that migrations were effective.

Species distribution modeling

Potential habitats under the present climate conditions were mainly located in the lowland around the coastal regions. Potential habitats under LGM climate conditions were mainly located in the Pacific Ocean side (southern to northwestern Kyushu, southern Shikoku to western Kii Pen. and around Izu Pen.), including the area of below sea level under current climate conditions (Fig. 5). Fig. 5 Potential habitats of *Castanopsis sieboldii* estimated by species distribution models at the present and the Last Glacial Maximum (LGM). The potential habitats under LGM were shown by the average of the three Global Climate Models (i.e., MIROC, CCSM, MPI-ESM) (Figure S6)



Discussion

The numbers and the locations of the glacial refugia in several areas of the world, including Japan, are key phylogeographic concerns and have been intensely debated. The ABC analysis of the *C. sieboldii* genetic dataset provided much stronger support for the scenario that the

Ryukyu Islands and the West groups were derived from separate refugia and the East and the Japan Sea groups were diverged from the West. Available fossil pollen data for broadleaved evergreen trees during the LGM (Matsuoka and Miyoshi 1998) suggest that refugia for *Castanopsis*type broadleaved evergreen forests were present in the Ryukyu Islands and the southwestern Kyushu in the main islands. The higher levels of genetic diversity of *C. sieboldii*, the ancient diversification from ABC analysis and higher probability of potential habitats under LGM climate conditions from SDM analysis are observed in the populations of the Ryukyu Islands and the West group containing Kyushu, suggesting that these populations have remained sufficiently large for ancestral polymorphism to be retained from the glacial periods up to the present day.

The time when the East and the Japan Sea groups diverged from an ancestral population of the West were estimated to be 431 generations ago. Under the coalescent theory, basic demographic parameters related to population size, timescale, and migration must be scaled by the rate of genetic drift or by the mutation rate (Hey and Nielsen 2004; Wakeley 2008). In the case of timescale, the parameter is expressed in terms of mutations as the number of generations × mutation rate (Hey and Nielsen 2004). In this study, we thus fixed the value of mutation rate at the moderate level (5×10^{-4}) and estimated the number of generations. As we used EST-SSR and in general mutation rate for EST-SSR is considered to be lower than that for genomic-SSR (Cubry et al. 2014), the estimated values of T_1 and T_2 might be underestimated. Moreover, lacking accurate information on the generation time for Castanopsis, we assumed generation times of 25 and 100 years in ABC analyses. When we consider the generation time, it is important to distinguish among age at maturity, generation time and lifespan (Petit and Hampe 2006). For instance, according to the first complete lifetable for a tree, the estimated generation time of the palm tree Enterpe globosa is 101 years, intermediate between its age at maturity (50 years) and maximum observed lifespan (156 years) (van Valen 1975). Thus, we choose more probable 100-year generation time, which provides that the East and the Japan Sea groups were diverged from the West at 43,100 years ago. The time predates the coldest stage around 18,000 years ago (Maeda 1980; Tsukada 1983; Martinson et al. 1987) of the LGM. These results suggest that the current populations of C. sieboldii descended from at least four isolated populations that were established prior to the LGM, however, the groups of the eastern parts and around the Japan Sea were more recently established from ancient southern refugia. Thus, these results provide evidence for the presence of LGM refugia in the regions of the eastern parts and around the Japan Sea, as well as the Ryukyu Islands and the western parts in Japan.

There was clear evidence of about four times expansion in the historical effective sizes from the glacial periods to the present (Figure S5). Chronological data of the Earth's climate (Martinson et al. 1987) and pollen records from the Japanese forests (Maeda 1980; Tsukada 1983) suggest that the Japanese climate was coldest around 18,000 years ago, started to warm about 12,000 years ago, and warmest about 6000–8000 years ago. The result suggests that the smaller populations of the four genetic groups may have survived in each glacial refugium for several glacial and interglacial periods in the Quaternary period, and when the climate began to get warm after the LGM, these populations regained and expanded their population sizes from the coastal regions to the inland areas. Migrations also occurred between adjacent populations from the southwestern to the eastern regions in the main islands of Japan from the LGM to the present.

In the main islands of Japan, a similar genetic differentiation between the western and eastern populations that was observed in C. sieboldii has also been detected in several other species (both plants and animals mentioned later) of the broadleaved evergreen forests. Geographical boundaries between the western and eastern populations are often reportedly located in a region spanning the main islands of Shikoku and Chugoku, the western tip of Honshu to Kii Peninsula, for example, Elaeocarpus sylvestris var. ellipticus (Aoki et al. 2004b), Photinia glabra (Aoki et al. 2006), Myrsine seguinii (Aoki et al. 2011), and Camellia japonica (Ueno 2015) for plant species, C. hilgendorfi; a specific predator of C. sieboldii seeds (Aoki et al. 2008) and Rhynchaenus dorsoplanatus; a specific leaf miner of C. sieboldii leaves (Aoki et al. 2010) for insect species. Genetic boundaries between several species of warm to cool temperate zones in Japan also appear to be located in this region, for example, Pinus (Miyata and Ubukata 1994; Iwaizumi et al. 2013), Cerasus (Tsuda et al. 2009), Zanthoxylum (Yoshida et al. 2010), Padus, Carpinus, and Magnolia (Iwasaki et al. 2010, 2012), and Acer (Yoshimaru and Matsumoto 2015) for plant species; Curculio (Aoki et al. 2009) for insect species; and Cervus (Nagata et al. 1999), Macaca (Kawamoto et al. 2007), Petaurista (Oshida et al. 2009), Ursus (Yasukochi et al. 2009), and Lepus (Nunome et al. 2010) for mammal species. The common geographic differentiation among multiple codistributed taxa may help efforts to elucidate the relative influence of major historical events such as climate changes during the glacial and interglacial periods (Avise 2000). The common genetic uniqueness of the eastern populations observed in various component species of broadleaved evergreen forests also suggests the eastern refugia in the main islands.

There is no clear palynological evidence at the LGM that considerable number of *Castanopsis*-type broadleaved evergreen trees survived the LGM in the eastern regions of the main islands (Figure S7). However, frequencies of fossilized pollen of broadleaved evergreen trees in the eastern parts of Japan began to markedly increase at the southern ends of the Kii, Izu, and Boso Peninsulas (located sequentially eastward along the Pacific Coast, close to the route of the warm Kuroshio Current) from 7500, 8500, and 7500 years ago, respectively, approximately 500-1500 years after similar increases at the southern ends of Kyushu (Matsushita 1992). Paleontological evidence also indicates that several other broadleaved evergreen trees were present on the Boso Peninsula about 9000 years ago (Okazaki et al. 2011). Moreover, the genetic diversity and the distinct chloroplast DNA haplotypes of these types of trees suggest the existence of refugia around the Kii Peninsula (Aoki et al. 2004b; Liu et al. 2013). Potential habitats of C. sieboldii under LGM climate conditions also located in the Pacific Ocean side including the area of below sea level under current climate conditions. Thus, some places in the southern tips of these peninsulas along the Pacific Coast are possible candidate locations of the cryptic eastern refugia in Japan.

Genetic boundaries between the Japan Sea and the Pacific Ocean populations of various plant species of temperate zones in Japan have been observed, for example, Cryptomeria (Tsumura et al. 2007), Fagus (Fujii et al. 2002; Hiraoka and Tomaru 2009), and Euonymus (Iwasaki et al. 2012). Fossil pollen evidence of Cryptomeria at the LGM also existed in the Japan Sea region (Kawamura 1977). There were fossilized pollen records of Castanea-Castanopsis (less than 1% to several %) at the LGM in the Japan Sea (Takahara and Takeoka 1992; Miyoshi et al. 1999). These information and the ABC-based demographic analyses of C. sieboldii suggest that Castanopsis forests could have survived in the small refugia along the Japan Sea coasts, though lack of SDM support for the refugia of this region because of the SDM model is too coarse to take into account topographic effects on microclimate.

Conclusion

The results of this study may help efforts to resolve the demographic dynamics for species of warm temperate ecosystems during the Quaternary period in Japan. Our ABC-based simulations of demographics and SDM of the dominant tree Castanopsis (Fagaceae), serving here as an indicator for the behavior of species in broadleaved evergreen forests, indicate that refugia were located at least one in the Ryukyu Islands and the other three regions of the western and eastern parts and around the Japan Sea of the main islands in Japan at the LGM. The results also provide ABC-based indications of demographic expansions driven by glacial oscillations. Our analyses provide foundations for further ABC-based explorations of the demographic history of species in warm temperate broadleaved forests during and after the last glacial period and a basic model for future phylogeographical studies using this approach.

Data archiving

Data available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.5sb1219.

Acknowledgements The authors are grateful to Hiroshi Yoshimaru, Yoshiaki Tsuda, Akihiro Seo, Kentaro Kimura, Hiroshi Takada, Sota Matsunaga, Ryouichi Kusano, Tsai-Wen Hsu, and Ho-Ming Chang for collecting the samples. We also thank Yuriko Taguchi and Yasuyuki Komatsu for the laboratory work and Lerma San Jose Maldia for the valuable comments on ABC analysis. This research was partly supported by a grant for research on Genetic Guidelines for Restoration Programs using Genetic Diversity Information from the Ministry of Environment, Japan, Grants-in-Aid from the Japan Society for the Promotion of Science (Nos. 1701416, 21770087, 2240030, 15J40011, and 18K06381 to KA), and the Research Project "A new cultural and historical exploration into human-nature relationships in the Japanese archipelago" of the Research Institute for Humanity and Nature, Kyoto, Japan.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: a workbench to detect molecular adaptation based on a F_{st}-outlier method. BMC Bioinformatics 9:323
- Aoki K, Hattori T, Murakami N (2004a) Intraspecific sequence variation of chloroplast DNA among the component species of evergreen broad-leaved forests in Japan II. Acta Phytotaxon Geobot 55:125–128
- Aoki K, Kato M, Murakami N (2005) Mitochondrial DNA of phytophagous insects as a molecular tool for phylogeographic study of host plants. Acta Phytotaxon Geobot 56:55–69
- Aoki K, Kato M, Murakami N (2008) Glacial bottleneck and postglacial recolonization of a seed parasitic weevil, *Curculio hilgendorfi*, inferred from mitochondrial DNA variation. Mol Ecol 17:3276–3289
- Aoki K, Kato M, Murakami N (2009) Phylogeographical patterns of a generalist acorn weevil: insight into the biogeographical history of broadleaved deciduous and evergreen forests. BMC Evol Biol 9:103
- Aoki K, Kato M, Murakami N (2011) Phylogeography of phytophagous weevils and plant species in broadleaved evergreen forests: a congruent genetic gap between western and eastern parts of Japan. Insects 2:128–150
- Aoki K, Matsumura T, Hattori T, Murakami N (2006) Chloroplast DNA phylogeography of *Photinia glabra* (Rosaceae) in Japan. Am J Bot 93:1852–1858
- Aoki K, Murakami N, Kato M (2010) Phylogeography of a specialist leaf-mining weevil, *Rhynchaenus dorsoplanatus* (Coleoptera: Curculionidae), associated with *Castanopsis* species. Ann Entomol Soc Am 103:379–388
- Aoki K, Suzuki T, Hsu T-W, Murakami N (2004b) Phylogeography of the component species of broad-leaved evergreen forests in Japan, based on chloroplast DNA. J Plant Res 117:77–94
- Aoki K, Suzuki T, Murakami N (2003) Intraspecific sequence variation of chloroplast DNA among the component species of evergreen broad-leaved forests in Japan. J Plant Res 116:337–344

- Aoki K, Ueno S, Kamijo T et al. (2014) Genetic differentiation and genetic diversity of *Castanopsis* (Fagaceae), the dominant tree species in Japanese broadleaved evergreen forests, revealed by analysis of EST associated microsatellites. PLoS One 9: e87429
- Aoki K, Ueno S, Kamijo T et al. (2016) Detecting east-west genetic differentiation in *Castanopsis* (Fagaceae) on the main islands of Japan and north-south on the Ryukyu Islands, based on chloroplast haplotypes. Plant Syst Evol 302:1093–1107
- Avise JC (2000) Phylogeography: the history and formation of species.. Harvard University Press, Cambridge, Massachusetts
- Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. Proc R Soc Lond B Biol Sci 263:1619–1626
- Beaumont MA, Zhang WY, Balding DJ (2002) Approximate Bayesian computation in population genetics. Genetics 162:2025–2035
- Birks HJB, Willis KJ (2008) Alpines, trees, and refugia in Europe. Plant Ecol Divers 1:147–160
- Blum M, Francois O (2010) Non-linear regression models for approximate Bayesian computation. Stat Comput 20:63–73.
- Budde KB, Gonzalez-Martinez SC, Hardy OJ, Heuertz M (2013) The ancient tropical rainforest tree *Symphonia globulifera* L. f. (Clusiaceae) was not restricted to postulated Pleistocene refugia in Atlantic Equatorial Africa. Heredity 111:66–76
- Chase MW, Soltis DE, Olmstead RG et al. (1993) Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. Ann Missouri Bot Gard 80:528–580
- Chen C, Durand E, Forbes F, François O (2007) Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. Mol Ecol Notes 7:747–756
- Chen C, Lu RS, Zhu SS, Tamaki I, Qiu YX (2017) Population structure and historical demography of *Dipteronia dyeriana* (Sapindaceae), an extremely narrow palaeoendemic plant from China: implications for conservation in a biodiversity hot spot. Heredity 119:95–106
- Cheng YP, Hwang SY, Chiou WL, Lin TP (2006) Allozyme variation of populations of *Castanopsis carlesii* (Fagaceae) revealing the diversity centres and areas of the greatest divergence in Taiwan. Ann Bot 98:601–608
- Cheng YP, Hwang SY, Lin TP (2005) Potential refugia in Taiwan revealed by the phylogeographical study of *Castanopsis carlesii* Hayata (Fagaceae). Mol Ecol 14:2075–2085
- Chybicki IJ (2014) INEST 2.0. wwwukwedupl/pracownicy/strona/ igor_chybicki/software_ukw/.
- Cook SR, Gelman A, Rubin DB (2006) Validation of software for Bayesian models using posterior quantiles. J Comput Graph Stat 15:675–692
- Csillery K, Francois O, Blum MGB (2012) abc: an R package for approximate Bayesian computation (ABC). Methods Ecol Evol 3:475–479
- Cubry P, Pujade-Renaud V, Garcia D et al. (2014) Development and characterization of a new set of 164 polymorphic EST-SSR markers for diversity and breeding studies in rubber tree (*Hevea brasiliensis* Müll. Arg.). Plant Breed 133:419–426
- Darleyhill S, Johnson WC (1981) Acorn dispersal by the blue jay (*Cyanocitta cristata*). Oecologia 50:231–232
- de Lafontaine G, Amasifuen Guerra CA, Ducousso A, Petit RJ (2014) Cryptic no more: soil macrofossils uncover Pleistocene forest microrefugia within a periglacial desert. New Phytol 204:715–729
- Dieringer D, Schlotterer C (2003) Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. Mol Ecol Notes 3:167–169
- Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf material. Phytochemistry 19:11–15
- SPRINGER NATURE

- Durand E, Chen C, François O (2009) TESS version 2.3: free computer program that implements a Bayesian clustering algorithm for spatial population genetic studies.
- El Mousadik A, Petit RJ (1996) High level of genetic differentiation for allelic richness among populations of the argan tree [Argania spinosa (L.) Skeels] endemic to Morocco. Theor Appl Genet 92:832–839.
- Estoup A, Angers B (1998) Microsatellites and minisatellites for molecular ecology: theoretical and empirical considerations. In: Carvalho G (ed) Advances in molecular ecology. NATO Press, Amsterdam, Holland, p 55–86
- Estoup A, Jarne P, Cornuet J-M (2002) Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. Mol Ecol 11:1591–1604
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620
- Excoffier L, Estoup A, Cornuet J-M (2005) Bayesian analysis of an admixture model with mutations and arbitrarily linked markers. Genetics 169:1727–1738
- Excoffier L, Foll M (2011) fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex scenarios. Bioinfomatics 27:1332–1334
- Excoffier L, Lischer H (2010) Arlquin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10:564–567
- Felsenstein J (1989) PHYLIP—Phylogeny Inference Package (version 3.2). Cladistics 5:164–166
- François O, Ancelet S, Guillot G (2006) Bayesian clustering using hidden Markov random fields in spatial population genetics. Genetics 174:805–816
- Fujii N, Tomaru N, Okuyama K, Koike T, Mikami T, Ueda K (2002) Chloroplast DNA phylogeography of *Fagus crenata* (Fagaceae) in Japan. Plant Syst Evol 232:21–33
- Gao J, Wang B, Mao I-F, Ingvarsson P, Zeng Q-Y, Wang X-R (2012) Demography and speciation history of the homoploid hybrid pine *Pinus densata* on the Tibetan Plateau. Mol Ecol 21:4811–4827
- Gómez JM (2003) Spatial patterns in long-distance dispersal of *Quercus ilex* acorns by jays in a heterogeneous landscape. Ecography 26:573–584
- Goudet J (2002) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3.2) http://www2.unil.ch/ popgen/softwares/fstat.htm. University of Lausanne, Lausanne, Switzerland.
- Hattori T (1985) Synecological study on the lucidophyllous forest of *Castanopsis-Persea* type in Japan proper. Bull Kobe Geobot Soc 1:1–98
- Hattori T (2002) Chapter 7. In: Yahara T, Kawakubo N (eds) Biology of conservation and restoration. Bun-ichi, Tokyo, p 203–222
- Hewitt G (2000) The genetic legacy of the Quaternary ice ages. Nature 405:907–913
- Hewitt GM (1999) Post-glacial re-colonization of European biota. Biol J Linn Soc 68:87–112
- Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. Philos Trans R Soc Lond Ser B Biol Sci 359:183–195
- Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. Genetics 167:747–760
- Hickerson MJ, Carstens BC, Cavender-Bares J et al. (2010) Phylogeography's past, present, and future: 10 years after Avise, 2000. Mol Phylogenet Evol 54:291–301
- Hiraoka K, Tomaru N (2009) Population genetic structure of *Fagus japonica* revealed by nuclear microsatellite markers. Int J Plant Sci 170:748–758

- Huang SSF, Hwang SY, Lin TP (2002) Spatial pattern of chloroplast DNA variation of *Cyclobalanopsis glauca* in Taiwan and east Asia. Mol Ecol 11:2349–2358
- Iwaizumi MG, Tsuda Y, Ohtani M, Tsumura Y, Takahashi M (2013) Recent distribution changes affect geographic clines in genetic diversity and structure of *Pinus densiflora* natural populations in Japan. Ecol Manag 304:407–416
- Iwasaki T, Aoki K, Seo A, Murakami N (2012) Comparative phylogeography of four component species of deciduous broad-leaved forests in Japan based on chloroplast DNA variation. J Plant Res 125:207–221
- Iwasaki T, Tono A, Aoki K, Seo A, Murakami N (2010) Phylogeography of *Carpinus japonica* and *Carpinus tschonoskii* (Betulaceae) growing in Japanese deciduous broad-leaved forests, based on chloroplast DNA variation. Acta Phytotaxon Geobot 61:1–20
- Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801–1806
- Jensen TS, Nielsen OF (1986) Rodents as seed dispersers in a heath oak wood succession. Oecologia 70:214–221
- Kamei T, Research Group for the Biogeography from Würm Glacial (1981) Fauna and flora of the Japanese Islands in the last glacial time. Quat Res 20:191–205
- Kanauchi A (2005) Pollen analysis of the Jaishi oike Moordeposit, in the southern part of the Izu Peninsula. Sundai Shigaku (Sundai Hist Rev) 125:119–130
- Kawamoto Y, Shotake T, Nozawa K et al. (2007) Postglacial population expansion of Japanese macaques (*Macaca fuscata*) inferred from mitochondrial DNA phylogeography. Primates 48:27–40
- Kawamura T (1977) Pollen analytical studies on the distribution of *Cryptomeria japonica* D. Don I. Akita Prefecture. Pollen Sci 11:8–20
- Knowles LL, Maddison WP (2002) Statistical phylogeography. Mol Ecol 11:2623–2635
- Kobayashi S, Hiroki S (2003) Patterns of occurrence of hybrids of *Castanopsis cuspidata* and *C. sieboldii* in the IBP Minamata Special Research Area, Kumamoto Prefecture, Japan. J Phytogeogr Taxon 51:63–67
- Kuroda T (1998) Chapter II-9. In: Yasuda Y, Miyoshi N (eds) The illustrated vegetation history of the Japanese Archipelago. Asakura-shoten, Tokyo, p 162–175
- Lee JH, Lee DH, Choi BH (2013) Phylogeography and genetic diversity of East Asian *Neolitsea sericea* (Lauraceae) based on variations in chloroplast DNA sequences. J Plant Res 126:193–202
- Lee JH, Lee DH, Choi IS, Choi BH (2014) Genetic diversity and historical migration patterns of an endemic evergreen oak, *Quercus acuta*, across Korea and Japan, inferred from nuclear microsatellites. Plant Syst Evol 300:1913–1923
- Li L, Abbott RJ, Liu BB et al. (2013) Pliocene intraspecific divergence and Plio-Pleistocene range expansions within *Picea likiangensis* (Lijiang spruce), a dominant forest tree of the Qinghai-Tibet Plateau. Mol Ecol 22:5237–5255
- Liu B, Abbott RJ, Lu Z, Tian B, Liu J (2014a) Diploid hybrid origin of Ostryopsis intermedia (Betulaceae) in the Qinghai-Tibet Plateau triggered by Quaternary climate change. Mol Ecol 23:3013–3027
- Liu C, Tsuda Y, Shen H, Hu L, Saito Y, Ide Y (2014b) Genetic structure and hierarchical population divergence history of *Acer mono* var. *mono* in south and northeast China. PLoS One 9: e87187
- Liu H-Z, Takeichi Y, Kamiya K, Harada K (2013) Phylogeography of *Quercus phillyraeoides* (Fagaceae) in Japan as revealed by chloroplast DNA variation. J For Res 18:361–370

- Louis M, Fontaine MC, Spitz J et al. (2014) Ecological opportunities and specializations shaped genetic divergence in a highly mobile marine top predator. Proc R Soc Lond B Biol Sci 281:20141558
- Maeda Y (1980) Jomon no umi to mori. Soju-shobo, Tokyo
- Martinson DG, Pisias NG, Hays JD, Imbrie J, Moore TC, Shackleton Jr NJ (1987) Age dating and the orbital theory of the Ice Ages: development of a high-resolution 0 to 300,000-year chronos-tratigraphy. Quat Res 27:1–29
- Matsuoka K, Miyoshi N (1998) Chapter III-4. In: Yasuda Y, Miyoshi N (eds) The illustrated vegetation history of the Japanese Archipelago. Asakura-shoten, Tokyo, p 224–236
- Matsushita M (1992) Lucidophyllous forest development along the Pacific Coast of the Japanese Archipelago during the Holocene. Quat Res 31:375–387
- Minato M, Ijiri S (1976) The Japanese archipelago, 3rd edn. Iwanamishoten, Tokyo
- Miyaki M, Kikuzawa K (1988) Dispersal of *Quercus mongolica* acorns in a broadleaved deciduous forest 2. Scatterhoarding by mice. Ecol Manag 25:9–16
- Miyata M, Ubukata M (1994) Genetic variation of allozymes in natural stands of Japanese black pine. J Jpn Soc 76:445–455
- Miyoshi N, Fujiki T, Morita Y (1999) Palynology of a 250-m core from Lake Biwa: a 430,000-year record of glacial-interglacial vegetation change in Japan. Rev Palaeobot Palynol 104:267–283
- Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res 8:4321–4432
- Nagata J, Masuda R, Tamate HB et al. (1999) Two genetically distinct lineages of the sika deer, *Cervus nippon*, in Japanese islands: comparison of mitochondrial D-loop region sequences. Mol Phylogenet Evol 13:511–519
- Nakanishi H (1996) Plant species with northbound distribution in western-Kyushu, Japan: definition, composition and origin. Acta Phytotaxon Geobot 47:113–124
- Nakao K, Higa M, Tsuyama I et al. (2014) Changes in the potential habitats of 10 dominant evergreen broad-leaved tree species in the Taiwan-Japan archipelago. Plant Ecol 215:639–650
- Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583–590
- Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. J Mol Evol 19:153–170
- Nunome M, Torii H, Matsuki R, Kinoshita G, Suzuki H (2010) The influence of Pleistocene refugia on the evolutionary history of the Japanese hare, *Lepus brachyurus*. Zool Sci 27:746–754
- Okazaki H, Kobayashi M, Momohara A et al. (2011) Early Holocene coastal environment change inferred from deposits at Okinoshima archeological site, Boso Peninsula, central Japan. Quat Int 230:87–94
- Oshida T, Masuda R, Ikeda K (2009) Phylogeography of the Japanese giant flying squirrel, *Petaurista leucogenys* (Rodentia: Sciuridae): implication of glacial refugia in an arboreal small mammal in the Japanese Islands. Biol J Linn Soc 98:47–60
- Parducci L, Jorgensen T, Tollefsrud MM et al. (2012) Glacial survival of boreal trees in Northern Scandinavia. Science 335:1083–1086
- Petit RJ, Hampe A (2006) Some evolutionary consequences of being a tree. Annu Rev Ecol Evol Syst 37:187–214
- Plummer M, Best N, Cowles K, Vines K (2006) CODA: Convergence Diagnosis and Output Analysis for MCMC. R News 6:7–11
- Pons J, Pausas JG (2007) Acorn dispersal estimated by radio-tracking. Oecologia 153:903–911
- Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW (1999) Population growth of human Y chromosomes: a study of Y chromosome microsatellites. Mol Biol Evol 16:1791–1798
- Provan J, Bennett KD (2008) Phylogeographic insights into cryptic glacial refugia. Trends Ecol Evol 23:564–571

- Pudlo P, Marin J-M, Estoup A, Cornuet J-M, Gautier M, Robert CP (2016) Reliable ABC model choice via random forests. Bioinfomatics 32:859–866
- Qiu YX, Fu CX, Comes HP (2011) Plant molecular phylogeography in China and adjacent regions: tracing the genetic imprints of Quaternary climate and environmental change in the world's most diverse temperate flora. Mol Phylogenet Evol 59:225–244
- Ren GP, Mateo RG, Liu JQ et al. (2017) Genetic consequences of Quaternary climatic oscillations in the Himalayas: *Primula tibetica* as a case study based on restriction site-associated DNA sequencing. New Phytol 213:1500–1512
- Rull V (2009) Microrefugia. J Biogeogr 36:481-484
- Saitou N, Nei M (1987) The neighbor-joining method; a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425
- Setoguchi H, Ohba H (1995) Phylogenetic relationships in Crossostylis (Rhizophoraceae) inferred from restriction site variation of chloroplast DNA. J Plant Res 108:87–92
- Shafer ABA, Cullingham CI, Cote SD, Coltman DW (2010) Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. Mol Ecol 19:4589–4621
- Soltis DE, Morris AB, McLachlan JS, Manos PS, Soltis PS (2006) Comparative phylogeography of unglaciated eastern North America. Mol Ecol 15:4261–4293
- Stewart JR, Lister AM, Barnes I, Dalen L (2010) Refugia revisited: individualistic responses of species in space and time. Proc R Soc Lond B Biol Sci 277:661–671
- Svenning JC, Normand S, Kageyama M (2008) Glacial refugia of temperate trees in Europe: insights from species distribution modelling. J Ecol 96:1117–1127
- Taberlet P, Fumagalli L, Wust-Saucy A-G, Cosson J-F (1998) Comparative phylogeography and postglacial colonization routes in Europe. Mol Ecol 7:453–464
- Takahara H, Takeoka M (1992) Vegetation history since the last glacial period in the Mikata lowland, the Sea of Japan area, western Japan. Ecol Res 7:371–386
- Tsuda Y, Kimura M, Kato S, Katsuki T, Mukai Y, Tsumura Y (2009) Genetic structure of *Cerasus jamasakura*, a Japanese flowering cherry, revealed by nuclear SSRs: implications for conservation. J Plant Res 122:367–375
- Tsukada M (1974) Paleoecology. II. Synthesis. Kyoritsu, Tokyo
- Tsukada M (1983) Vegetation and climate during the last glacial maximum in Japan. Quat Res 19:212–235
- Tsumura Y, Kado T, Takahashi T, Tani N, Ujino-Ihara T, Iwata H (2007) Genome scan to detect genetic structure and adaptive genes of natural populations of *Cryptomeria japonica*. Genetics 176:2393–2403
- Tzedakis PC, Emerson BC, Hewitt GM (2013) Cryptic or mystic? Glacial tree refugia in northern Europe. Trends Ecol Evol 28:696–704

- Ueno S (2015) Camellia japonica. In: Tsumura Y, Suyama Y (eds) Chizu de wakaru jumoku no shubyo ido gaidorain. Bunichi, Tokyo
- van Valen L (1975) Life, death, and energy of a tree. Biotropica 7:259–269
- Wakeley J (2008) Coalescent theory. Roberts and Company Publishers, Colorado, U.S.
- Wang Q, Liu JQ, Allen GA et al. (2016) Arctic plant origins and early formation of circumarctic distributions: a case study of the mountain sorrel, *Oxyria digyna*. New Phytol 209:343–353
- Wegmann D, Leuenberger C, Neuenschwander S, Excoffier L (2010) ABCtoolbox: a versatile toolkit for approximate Bayesian computations. BMC Bioinformatics 11:116
- Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370
- Wolfe KH, Li W-H, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Biol J Linn Soc 84:9054–9058
- Wright S (1931) Evolution in Mendelian populations. Genetics 16:97–159
- Wu SH, Hwang CY, Lin TP, Chung JD, Cheng YP, Hwang SY (2006) Contrasting phylogeographical patterns of two closely related species, *Machilus thunbergii* and *Machilus kusanoi* (Lauraceae), in Taiwan. J Biogeogr 33:936–947
- Xu J, Deng M, Jiang X-L, Westwood M, Song Y-G, Turkington R (2015) Phylogeography of *Quercus glauca* (Fagaceae), a dominant tree of East Asian subtropical evergreen forests, based on three chloroplast DNA interspace sequences. Tree Genet Genomes 11:805.
- Yamazaki T, Mashiba S (1987a) A taxonomical revision of *Castanopsis cupidata* (Thunb.) Schottky and the allies in Japan, Korea and Taiwan. 1. J Jpn Bot 62:289–298
- Yamazaki T, Mashiba S (1987b) A taxonomical revision of *Castanopsis cupidata* (Thunb.) Schottky and the allies in Japan, Korea and Taiwan. 2. J Jpn Bot 62:332–339
- Yasukochi Y, Nishida S, Han SH, Kurosaki T, Yoneda M, Koike H (2009) Genetic structure of the asiatic black bear in Japan using mitochondrial DNA analysis. J Hered 100:297–308
- Yoichi W, Tamaki I, Sakaguchi S, Song JS, Yamamoto S, Tomaru N (2016) Population demographic history of a temperate shrub, *Rhododendron weyrichii* (Ericaceae), on continental islands of Japan and South Korea. Ecol Evol 6:8800–8810
- Yoshida T, Nagai H, Yahara T, Tachida H (2010) Genetic structure and putative selective sweep in the pioneer tree, *Zanthoxylum* ailanthoides. J Plant Res 123:607–616
- Yoshimaru H, Matsumoto A (2015) Acer palmatum. In: Tsumura Y, Suyama Y (eds) Chizu de wakaru jumoku no shubyo ido gaidorain. Bunichi, Tokyo