Title page

Title: Population genetic structure and demography of *Magnolia kobus*: variety *borealis* is not supported genetically

Authors: Ichiro Tamaki [1], Naomichi Kawashima [2], Suzuki Setsuko [3], Jung-Hyun Lee [4], Akemi Itaya [5], Kyohei Yukitoshi [2], and Nobuhiro Tomaru [2]

[1] Gifu Academy of Forest Science and Culture, 88 Sodai, Mino, Gifu 501-3714, Japan

[2] Graduate School of Bioagricultural Sciences, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan

[3] Department of Forest Molecular Genetics and Biotechnology, Forestry and Forest Products Research Institute, Forest Research and Management Organization, 1 Matsunosato, Tsukuba, Ibaraki 305-8687, Japan

[4] Department of Biology Education, Chonnam National University, 77 Yongbong-ro, Buk-gu, Gwangju 500-757, Republic of Korea

[5] Graduate School of Bioresources, Mie University, 1577 Kurimamachiya, Tsu, Mie 514-8507, Japan

Abstract

Species delimitations by morphological and by genetic markers are not always congruent. *Magnolia kobus* consists of two morphologically different varieties, *kobus* and *borealis*. The latter variety is characterized by larger leaves than the former. For the conservation of *M. kobus* genetic resources in natural forests, the relationships between morphological and genetic variation should be clarified. We investigated variations in nuclear microsatellites, chloroplast DNA (cpDNA) sequences and leaf morphological traits in 23 populations of *M. kobus* over the range of species. Two genetically divergent lineages, northern and southern were detected and their geographical boundary was estimated to be at 39˚N. The northern lineage consisted of two genetic clusters and a single cpDNA haplotype, while the southern one had multiple genetic clusters and cpDNA haplotypes. The northern lineage showed significantly lower genetic diversity than the southern. Approximate Bayesian computation indicated that the northern and southern lineages had experienced, respectively, population expansion and long-term stable population size. The divergence time between the two lineages was estimated to be 565,000 years ago and no signature of migration between the two lineages after divergence was detected. Ecological niche modeling showed that the potential distribution area in northern Japan at the last glacial maximum was very small. It is thus considered that the two lineages have experienced different population histories over several glacial-inter-glacial cycles. Individuals of populations in the central to northern part of Honshu on the Sea of Japan side and in Hokkaido had large leaf width and area. These leaf characteristics corresponded with those of variety *borealis*. However, the delimitation of the northern

and southern lineages detected by genetic markers (39˚N) was not congruent with that detected by leaf morphologies (36˚N). It is therefore suggested that variety *borealis* is not supported genetically and the northern and southern lineages should be considered separately when identifying conservation units based not on morphology but on genetic markers.

Keywords Approximate Bayesian computation · Chloroplast DNA sequences · Conservation · Ecological niche modeling · Leaf morphology · Microsatellites

Introduction

Tree species that are widely distributed along the Japanese archipelago show significant genetic differentiation for neutral genetic markers between the Sea of Japan and Pacific Ocean sides (*Fagus crenata*, Hiraoka and Tomaru 2009; *Cryptomeria japonica*, Tsumura et al. 2014) and/or between the north and south (*Kalopanax septemlobus*, Sakaguchi et al. 2011; *Quercus aliena*, San Jose-Maldia et al. 2017; *Magnolia salicifolia*, Tamaki et al. 2018). However, the boundaries of genetic differentiation are not always the same among species. This may be mainly due to the differences in locations of refugia during the glacial period among species. Broad-leaved tree species growing along the Sea of Japan side of the Japanese archipelago are often characterized by large-wide-thin leaves, while related species growing in the Pacific Ocean side are characterized by small-narrow-thick leaves (Hotta 1974). The main factor generating these differences is considered to be adaptation to dryness on the Pacific Ocean side during the flushing period (Hotta 1974). Even within a species that is widely distributed along the Japanese archipelago, latitudinal clines of leaf area and leaf width can be detected (Hagiwara 1977; Koyama et al. 2002; Tamaki et al. 2018). Provenance tests on *Fagus crenata* have proved that variations in leaf morphology and physiology are based not on phenotypic plasticity but on the genetic make-up of individuals (Hashizume et al. 1997; Koike and Maruyama 1998). However, delimitation based on morphological traits and that determined from genetic structure do not always accord (Duminil and Di Michele 2009). It is therefore necessary to clarify the relationships between morphological traits reflecting physiological adaptation to

environment, and genetic structure, when considering conservation of genetic resources of forest trees that are broadly distributed along the Japanese archipelago.

Magnolia kobus DC., which belongs to the Magnoliaceae, is a major tree species in temperate forests, and it is distributed in Hokkaido, Honshu and Kyushu Islands of Japan and Jeju Island of Korea (Ueda 2006). There are two varieties, *kobus* and *borealis*. Variety *borealis*, which is characterized by larger leaves and flowers than variety *kobus*, is distributed from central to northern Honshu on the Sea of Japan side and on Hokkaido (Ohashi 2015). According to Ohashi (2015), the ranges of leaf length for varieties *kobus* and *borealis* overlap each other (6–15 cm and 10–20 cm, respectively), whereas those of leaf width do not (3–6 cm and 6–10 cm, respectively). Thus, the leaf width could become a key to distinguish the two varieties. However, Callaway (1994) points out that the morphologies of variety *borealis* are not always consistent even within an individual and thus recognizing it as a separate variety does not appear justified. Moreover, the Flora of Japan, which is one of the most authoritative catalogs of Japanese plants, does not treat the variety *borealis* as a distinct variety and treats as one of the synonyms (Ueda 2006), and this may be due to its morphological ambiguity. Recently, Tamaki et al. (2018) have reported that *M. salicifolia*, which is a species related to *M. kobus*, diverges both morphologically and genetically between northern and southern lineages. Accordingly, also in *M. kobus*, the relationships between morphological and genetic variations should be clarified.

M. kobus is popular as an ornamental tree due to its beautiful flowers and tolerance of vehicle emissions, and it is planted at roadsides all over Japan. It is also

planted when restoring natural deciduous broad-leaved forests (Takasuna and Takayama 2011). The presence within a natural forest of seedlings that have escaped from trees planted near the forest is frequently reported (Fujii 1997; Ishida et al. 2008; Tamaki et al. 2016). Planting without considering the origin of individuals may cause serious genetic disruptions of the genetic resources of *M. kobus* growing in natural forests (Lefèvre 2004; Potts et al. 2003). As the Japanese archipelago is latitudinally long, there is some degree of latitudinal climatic heterogeneity among the habitats of tree species. Moreover, there are climatic differences between the Sea of Japan side and the Pacific Ocean side. Studies of tree species broadly distributed in the Japanese archipelago have reported environmental incongruence among trees that were planted in different sites from those where they originated. Reciprocal transplanting of *Pinus densiflora* lineages between its natural northern and southern habitats revealed that the southern lineage was at a disadvantage in terms of survival and growth when it was transplanted northwards (Nagamitsu et al. 2015). Similarly, a reciprocal transplant of *F. crenata* lineages between the Sea of Japan side and the Pacific Ocean side indicated that both lineages had significant home site advantages with respect to both survival and growth (Koyama 2012). The direction and extent of environmental incongruences brought about by genetic disturbance vary among tree species. It is therefore necessary to determine conservation units carefully, taking into account genetic and ecological information about the target species.

In this study we investigated genetic variation in nuclear microsatellites and chloroplast DNA sequences, and leaf morphological traits, in *M. kobus* populations across

the distribution range. We performed approximate Bayesian computation and ecological niche modeling. The specific objectives of this study are 1) to clarify the genetic diversity and structure of *M. kobus* natural populations over the range of the species; 2) to assess the existence of variety *borealis* based on leaf morphological and genetic traits and 3) to infer how best to conserve genetic resources of *M. kobus*.

Materials and methods

Sample collection

We sampled 10 to 20 leaves per individual for DNA extraction and measurement of leaf morphology from 23 *Magnolia kobus* populations, which cover its entire distribution range (Fig. 1 and Table 1). We sampled leaves from trees more than 20 m apart, so as not to sample leaves from the same clones, because *M. kobus* can propagate clonally by sprouting and/or layering. Two to four shoots other than water sprouts and very short shoots were cut from a sun-lit surface of a tree crown and the second or subsequent leaves from the top of each shoot were collected. Leaves were transported to the laboratory under cool conditions. After scanning for leaf shape, the leaves were stored at -30˚C until required for DNA extraction. However, in population 23 (Jeju) the leaves collected were dried using silica gel and stored at room temperature prior to DNA extraction. We could not collect enough leaves for morphological measurement in this population, so the leaves were used only in genetic analysis.

DNA extraction, genotyping and sequencing

Genomic DNA was extracted using a hexadecyltrimethylammonium bromide (CTAB) method (Murray and Thompson 1980) with minor modifications. Fourteen nuclear microsatellites (nSSRs) developed for *M. obovata*, M6D8 (Isagi et al. 1999), and for *M. stellata*, stm0002, stm0114, stm0163, stm0184, stm0200, stm0214, stm0223, stm0246, stm0251, stm0353, stm0383, stm0423 and stm0448 (Setsuko et al. 2005), were amplified using a Multiplex PCR Kit (QIAGEN) with a GeneAmp PCR System 9700 (Applied Biosystems) following the manufacturer's manual. The PCR products were separated by electrophoresis on a 3100-Avant Genetic Analyzer (Applied Biosystems). Genotypes were determined using GeneMapper version 4.0 (Applied Biosystems). All genotype data were converted from fragment size to number of repeats. Before genetic analysis, in order to remove those loci with high frequencies of null alleles, we calculated the null allele frequency in each population for each locus with INEst version 1.1, which can estimate null allele frequency separately from the effect of inbreeding (Chybicki and Burczyk 2009). We calculated average null allele frequency among populations for each locus. Apart from locus stm0223, all the loci showed a null allele frequency of less than 7%, so we used the remaining 13 loci in the following analyses. Four non-coding chloroplast DNA (cpDNA) regions, *trn*S–*trn*G (Shaw et al. 2005), *trn*T–*psb*D (Shaw et al. 2007), *trn*T–*trn*L (Shaw et al. 2005; Taberlet et al. 1991) and *rpl*36–*inf*A–*rps*8–*rpl*14 (Shaw et al. 2007), were sequenced from 1 to 4 individuals in each population in the same way as described in Tamaki et al. (2018).

Analysis of genetic diversity and differentiation

For each nSSR locus across all populations, the number of alleles (A), average gene diversity within population (H_S), gene diversity in the total population (H_T) and Weir and Cockerham's F_{ST} were calculated. Hedrick's standardized G_{ST} [G'_{ST}; Hedrick (2005)] and Jost's D, which is another population differentiation measure (Jost 2008), were also manually calculated. The significance of population differentiation at each locus was evaluated by a randomization test. For each population over all nSSR loci, allelic richness (A_R) based on four diploid individuals, expected heterozygosity (H_E) and fixation index (F_{IS}) were calculated. The significance of departure from Hardy-Weinberg equilibrium in each population was evaluated by a permutation test. As STRUCTURE analysis detected two major genetic clusters, the 23 populations were divided into northern (populations 1 to 10) and the southern (11 to 23) lineages (see details in Results). Differences in A_R, H_E and F_{IS} between the two lineages were evaluated by randomization tests. The above summary statistics except for G'_{ST} and D were calculated by FSTAT version 2.9.3.2 (Goudet 1995). The presence of an isolation by distance pattern, which is a significant correlation between geographic and genetic distances, was investigated by the Mantel test with R package ade4 version 1.7.11 (Chessel et al. 2004). Kilometers on the log scale and F_{ST} / (1 - F_{ST}) between population pairs were used as geographic and genetic distances, respectively. Population-based principal component analysis using allele frequency data was conducted using R package ade4.

Genetic structure among populations was investigated with the model based

clustering method implemented in STRUCTURE version 2.3.4 (Falush et al. 2003; Pritchard et al. 2000). The admixture and correlated allele frequency models were used. As suggested by Wang (2017), different α values for each genetic cluster were estimated and a low initial value of $\alpha = 0.05$ was applied. Different numbers of genetic clusters (K) from 1 to 16 were tested. The first 40,000 iterations were discarded as a burn-in period and then 40,000 iterations were used for the estimation of membership of each genetic cluster for each individual. The estimations of parameters were repeated 10 times for each K. CLUMPAK was used to check the multimodality within the same K (Kopelman et al. 2015). LargeKGreedy option was selected and the number of repeats was set to 500. To estimate the optimal K, the log probability of data and ΔK for each K were estimated with R package corrsieve version 1.6.8 (Campana et al. 2011; Evanno et al. 2005). Analysis of molecular variance (AMOVA) was performed with Arlequin version 3.5.2 (Excoffier and Lischer 2010). Genetic variation was hierarchically divided into three layers, which were the lineages inferred by STRUCTURE analysis, populations and individuals, and variance components for each layer and related Φ-statistics were calculated. The significance of each Φ-statistic was evaluated by the permutation test implemented in Arlequin.

CpDNA sequences were edited and assembled with DNA baser version 3 (Heracle BioSoft SRL), and then aligned with the MUSCLE algorithm in MEGA version 5.1 (Edgar 2004; Tamura et al. 2011). Mono- or di-nucleotide repeats in the sequences were omitted from subsequent analysis to avoid the possibility of homoplasy. CpDNA haplotypes were determined and a network among them was constructed using TCS

version 1.21 (Clement et al. 2000). The number of polymorphic sites and the mean number of pairwise differences were calculated, and Tajima's test for selective neutrality (Tajima 1989) was performed with Arlequin.

Analysis of variation in leaf morphology

We used 9.8 leaves on average per individual tree, and a total of 4,260 leaves for analysis of leaf morphology. The length and width of each leave were measured and their average values in each individual were calculated. Numerical conversion of leaf shape into elliptic Fourier descriptors and measurement of leaf area were conducted with SHAPE version 1.3 (Iwata and Ukai 2002). Principal component analysis of leaf shape variables was conducted by SHAPE. Principal components (PCs) whose contribution to the total variance of data was more than 5% (PC1, PC2 and PC3) were used (see details in Results). Because PC3 represented the asymmetry of leaf shape and positive and negative values therefore probably had no biological meaning, and also to ensure normality, log-transformed absolute values of PC3 were used.

To evaluate the effects of environmental factors and population history on leaf morphological traits, Bayesian linear mixed model was constructed. For response variables, PC1, PC2, PC3 and leaf area were used. In this analysis, we just intended to evaluate whether environmental factors affected leaf morphological traits or not and did not intend to specify what environmental factors affected them. Thus, all 19 bioclimatic variables were downloaded from WorldClim (http://www.worldclim.com) and principal component analysis with the data for the variables was conducted using prncomp function

of R. Principal components (BioPCs) estimated by 19 bioclimatic variables were used as explanatory variables for environmental effects. Only BioPCs whose contribution to the total variance of data were more than 5% (BioPC1, BioPC2, BioPC3 and BioPC4) were used (see details in Results). Membership coefficient of the northern lineage estimated by STRUCTURE at $K = 2$ (Q) was also used as an explanatory variable for population history. In the preliminary analysis, we calculated correlation coefficients between BioPCs and Q. Because a correlation coefficient between BioPC1 and Q was -0.823, we removed BioPC1 from the following analysis. On the other hand, absolute values of the other correlation coefficients were less than 0.213. Thus, finally, we used BioPC2, BioPC3, BioPC4 and Q for explanatory variables. Replicates within an individual were treated as a random effect. Constructed Bayesian linear mixed model was as follows:

$$Y_{\mathrm{Exp}}[i] = \beta_0[j] + \beta_{\mathrm{BioPC2}} \times \mathrm{BioPC2}[i] + \beta_{\mathrm{BioPC3}} \times \mathrm{BioPC3}[i] + \beta_{\mathrm{BioPC4}} \times \mathrm{BioPC4}[i] + \beta_{\mathrm{Q}} \times \mathrm{Q}[i]$$

$$\mathrm{Y}_{\mathrm{Obs}}[i] \sim \mathrm{Normal}\ (Y_{\mathrm{Exp}}[i], \sigma_{\mathrm{All}})$$

$$\beta_0[j] \sim \mathrm{Normal}\ (\beta_{0\_\mathrm{Mean}}, \sigma_{\mathrm{Individual}})$$

Italic and roman characters indicate the parameters to be estimated and those with observed values, respectively. Indices i and j are leaf and individual IDs, respectively. Y_{Exp} and $\mathrm{Y}_{\mathrm{Obs}}$ are expected and observed values of response variables, respectively. $\beta_0[j]$ is an intercept and the other βs are regression coefficients. σs are standard deviations and assumed to take a value greater than zero. Non-informative priors were used for all parameters. Stan via rstan package version 2.18.2 of R was used to estimate posterior distributions of parameters (Stan Development Team 2018). Four independent Markov

chains were run. Each Markov chain was constructed 4,000 iterations and the first 1,000 iterations were discarded as a burn-in period. Posterior distributions were sampled by 10 steps and, finally, 300 × 4 = 1,200 samples were used to estimate the posterior mode and 95% highest posterior density (HPD). The posterior mode was estimated using the density function of R. The 95% HPD was estimated using coda package version 0.19.1 in R (Plummer et al. 2006). We considered βs significant if the 95% HPD did not overlap zero.

Inference of population demography

To infer the population demography of the two distinct lineages, the northern and southern lineages, detected by STRUCTURE analysis, an approximate Bayesian computation (ABC) approach was applied. We used only pure populations without admixture from the recessive lineage (less than 5%) and populations 10, 11 and 17 were thus removed from this analysis. To reduce the computational cost and uncertainty of parameter estimation, the parameters of the demographic model were sequentially inferred (Chen et al. 2017; Hiraoka et al. 2018; Tamaki et al. 2018). First, population size change models were applied for each lineage and then models of population divergence between the two lineages were applied using information obtained from the population size change model. We used 13 nSSR and 3,929 bp cpDNA sequences, from which insertions/deletions (indels) were removed. In population size change analysis, the average and standard deviation of the number of alleles, heterozygosity and allele size range for nSSRs, and the number of polymorphic sites and mean number of pairwise nucleotide differences for cpDNA sequences, were calculated. A total of eight summary statistics was used. In

population divergence analysis, in addition to the same summary statistics as for population size change analysis, allele size range for samples overall, F_{ST} of overall loci for nSSR and cpDNA sequences were calculated. A total of 19 summary statistics was used. The software arlsumstat version 3.5.2 was used for calculation of summary statistics (Excoffier and Lischer 2010).

Three distinct population size change models were built (Fig. 2A). 1) The standard neutral model (SNM) assumed that the effective population size was constant from the current to the past and had one structural parameter, current effective population size (N_{CUR}). The unit of N_{CUR} was the number of diploid individuals. 2) the population growth model (PGM) assumed that the current effective population size shrank exponentially towards the past with the rate G [$Nt = N_{CUR} \times \exp(G \times t)$; Nt was the effective population size at time t], i.e. there was exponential growth from past to present. PGM had two structural parameters, N_{CUR} and G. The unit of the time parameter was generations. 3) The size reduction model (SRM) assumed that the effective population size changed at time T and had three structural parameters N_{CUR}, T and relative ancestral effective population size [RN_{ANC}; $RN_{ANC} > 1$ and ancestral population size (N_{ANC}) = $N_{CUR} \times RN_{ANC}$]. Prior distributions of parameters of the three models are shown in Table S1. We simulated nSSR and cpDNA sequence data simultaneously under these three models by fastsimcoal2 version 2.5.2.21 (Excoffier and Foll 2011). As fastsimcoal2 used the number of gene copies as the unit of effective population size, twice the value of N_{CUR} was passed to it when we simulated the nSSR data. On the other hand, the raw value of N_{CUR} was passed to it when we simulated the cpDNA sequence data because *M. kobus* is

monoecious. A generalized stepwise mutation model (GSM) was used as a mutation model for nSSRs (Estoup et al. 2002). GSM has two parameters, mutation rate per generation (μ) and the geometric parameter (P_{GSM}). P_{GSM} ranges from 0 to 1 and represents the proportion of mutations that change allele sizes by more than one step; a value of zero means a strict stepwise mutation model. We simulated 13 independent loci. The prior distribution for the mean value of μ among 13 loci was drawn from a log-uniform distribution from 10^{-5} to 10^{-3} and each locus value of μ was randomly drawn from a gamma distribution with *shape* and *rate* parameters. The prior distribution of the *shape* parameter was drawn from a uniform distribution from 0.5 to 5 and the *rate* parameter was calculated by *shape* / the mean value of μ. The prior distribution of the mean value of P_{GSM} among the 13 loci was drawn from a uniform distribution from 0 to 1 and each locus value of P_{GSM} was randomly drawn from a beta distribution with *a* and *b* parameters. The values of *a* and *b* were calculated by 0.5 + 199 × the mean value of P_{GSM} and $a \times (1 - \text{the mean value of } P_{GSM}) / \text{the mean value of } P_{GSM}$, respectively, according to Excoffier et al. (2005). For cpDNA sequences, we simulated 3,929 bp sequences and the value of mutation rate for cpDNA was set to 2.0×10^{-9} substitutions per site per generation (Muse 2000; Sakaguchi et al. 2012).

All priors were generated using R version 3.5.0 (R Core Team 2018). The three size change models were simulated 10,000 times and summary statistics were calculated by arlsumstat. The three models were compared using the ABC random forest (ABC-RF) approach implemented in abcrf package version 1.7 of R (Pudlo et al. 2016). ABC-RF can yield similar results to those of the conventional ABC with much smaller numbers of

simulations and thus greatly reduce the computational time required. The number of trees in the random forest was set to 1,000. The classification error and posterior probability of the best model were calculated. For the best model, 5×10^5 simulations were conducted and posterior distributions of parameters were estimated using neural network regression implemented in abc package version 2.1 in R with the logit transformation option (Csillery et al. 2012). The tolerance value was set to 0.005 and 2,500 summary statistics nearest to the observed data were used for the parameter estimation. The posterior mode and 95% HPD were estimated in the same way as Bayesian linear mixed model analysis.

Taking into account the results of the population size change models, four different population divergence models were constructed (Fig. 2B). 1) The isolation model (ISM) assumed that the two lineages diverged at time T_{DIV} and it had four structural parameters, N_N, N_S, T_{DIV} and G. N_N and N_S were the effective population sizes of the northern and the southern lineage, respectively. The value of G was fixed at -1.56×10^{-4} using the value of the mode estimated in the population size change analysis in order to reduce the computational costs (see details in Results). 2) The isolation with migration model (IMM) assumed that there were migrations between lineages after divergence and it had six structural parameters, N_N, N_S, T_{DIV}, G and numbers of migrants per generation from the northern to the southern and from the southern to the northern lineage in the backward-in-time direction (Nm_{NS} and Nm_{SN}, respectively). When running simulations, Nm_{NS} and Nm_{SN} were divided by N_N and N_S, respectively, then the migration rates calculated were passed to fastsimcoal2. In angiosperms, the migration rate measured in the nuclear genome reflects both pollen and seed dispersals, whereas that in the

chloroplast genome reflects only seed dispersal because the chloroplast genome is generally maternally-transmitted. When simulating cpDNA sequences, we thus multiplied migration rates by a coefficient β, which ranges from 0 to 1, in order to take account of the reduction in the migration rate for the chloroplast genome. The prior distribution of β was drawn from a uniform distribution from 0 to 1. 3) and 4) IMM models with one way migration from the northern to the southern lineage (IMM_{NS}) and from the southern to the northern lineage (IMM_{SN}) were also defined. For the mutation parameters of the four population divergence models, the same settings as for population size change analysis were used. All prior distributions for parameters are listed in Table S1. Model choice and parameter estimation of the best model were conducted in the same way as population size change analysis except for the number of simulations (1.5×10^6) and tolerance value (0.002; 3,000 summary statistics nearest to the observed data) when parameter estimation was carried out.

To confirm that the model fit the observed data, posterior predictive simulations using 1,000 randomly drawn posterior samples were conducted for both population size change and population divergence analyses (Gelman et al. 2014). Summary statistics were calculated and compared to the observed data.

Ecological niche modeling

Ecological niche modeling was performed to infer the possible distribution ranges of *M. kobus* in the last glacial maximum (LGM; 21 kya) and last inter-glacial (LIG; 130 kya) with the maximum entropy method implemented in Maxent version 3.3.3k (Phillips et al.

2006). We used 101 location data for sites where the occurrence of *M. kobus* was recorded. These location data consisted of the 23 populations sampled in this study, our field observations and records from the Global Biodiversity Information Facility [GBIF.org (16 April 2017) GBIF Occurrence Download http://doi.org/10.15468/dl.ifurvo]. All records from GBIF were carefully checked against satellite images on Google Maps (http://maps.google.com) and ambiguous or erroneous location data were removed. A current distribution model was constructed with six bioclimatic variables, annual mean temperature (bio1), mean temperature of warmest quarter (bio10), mean temperature of coldest quarter (bio11), annual precipitation (bio12), precipitation in warmest quarter (bio18) and precipitation in coldest quarter (bio19), at a resolution of 2.5 arc-minutes, as used in a study of *Magnolia salicifolia* (Tamaki et al. 2018), a species growing in a similar climate zone. Validation of the model was performed using 100 replicates of cross-validation procedures, with 25% of the data for model testing, implemented in Maxent. Assuming temporal stability of ecological niche for *M. kobus*, the model constructed was applied to LGM and LIG climatic layers to predict the past distributions of the species. The model for interdisciplinary research on climate [MIROC; Hasumi and Emori (2004)] and the community climate system model [CCSM; Collins et al. (2006)] were used to predict the distributions during the LGM. All data for bioclimatic variables used in this modelling were obtained from WorldClim. To determine the coastal line at the LGM, we obtained the ETOPO1 Global Relief Model (https://doi.org/10.7289/V5C8276M) and only predicted areas higher than -130 m from the present level were clipped out.

Results

Genetic diversity and differentiation

Among the 13 nuclear microsatellite loci across 23 populations, the number of alleles (A) ranged from 12 to 37 with an average value of 24.6 and the average gene diversity within populations (H_S) ranged from 0.388 to 0.911 with an average value of 0.762 (Table 2). The values of F_{ST}, G'_{ST} and Jost's D over the 13 loci were 0.119, 0.504 and 0.439, respectively. All of the 13 loci showed significant population differentiation. Among the 23 populations over the 13 loci, allelic richness (A_R) based on four individuals ranged from 3.34 to 5.30 with an average value of 4.48 and expected heterozygosity (H_E) ranged from 0.580 to 0.859 with an average value of 0.764 (Table 1). Fixation index (F_{IS}) ranged from 0.008 to 0.340 within populations and its value over all populations was 0.097. Twelve of the 23 populations showed significant deviation from Hardy-Weinberg equilibrium.

The log probability of data in each K estimated by STRUCTURE analysis increased with increasing K until it reached a plateau at $K = 14$ (Fig. 3A and B). However, ΔK was highest at $K = 2$. We therefore considered that $K = 2$ and 14 were the optimal Ks. By using CLUMPAK, we checked the multimodality within the same K from $K = 2$ to 14 and found multiple modes except when $K = 2$ and 7. We carefully checked changes in cluster distribution along K, determined appropriate cluster distribution at each K by basically choosing major modes except when $K = 8$ and 12, and constructed a series of

barplots for membership coefficients (Fig. 3C). The distribution of genetic clusters at K = 2 showed clear separation between the northern and southern regions (Figs. 1 and 3C). Accordingly, we classified the 23 populations into northern (populations 1 to 10) and southern lineages (11 to 23). The northern and southern lineages were dominated by clusters 1 and 2, respectively. The value of F_{ST} between each cluster and the ancestral population was ca. one hundred times greater for cluster 1 (0.102) than for cluster 2 (0.001). Populations 10, 11 and 17 showed more than 5% genetic admixture of the recessive cluster. At K = 14, although populations 1 to 10 were dominated by the mixture of two clusters, most of the remaining 13 populations were dominated by one of the other 12 clusters (Fig. 3C). The genetic clusters that dominated in populations 11, 20 and 23 showed larger values of F_{ST} between each cluster and the ancestral population (> 0.33).

Population based principal component analysis also showed that populations in the northern lineage aggregated on the principal component axes (Fig. S1). Populations 20 and 23 were located apart from the other populations on the axes. A_R and H_E were higher in the southern lineage than in the northern lineage with the exceptions of populations 20 and 23 (Table 1 and Fig. S2). Average values of A_R and H_E in the southern lineage were significantly higher than those in the northern one (permutation test, P = 0.017 and 0.016 for A_R and H_E, respectively; Table 1). The difference in F_{IS} between the two lineages was not significant.

Significant isolation by distance (IBD) patterns were detected over all 23 populations (R^2 = 0.400 and P < 0.001) and in both the northern and the southern lineages (R^2 = 0.150 and 0.263, and P = 0.009 and 0.003, respectively; Fig. S3). However, the

strength of IBD in the northern lineage was less than that in the southern one.

A total of 3,932 bp-length of aligned cpDNA sequences was obtained. Four substitutions were detected within the species and three haplotypes (H, I and J) were determined (Fig. 1 and Table S2). Although only haplotype H was detected in the northern lineage, all three haplotypes were detected in the southern one. In the southern lineage, the number of polymorphic sites, mean number of pairwise differences and Tajima's *D* were 4, 1.311 and 0.695, respectively. The result of Tajima's test for selective neutrality was not significant. All of the three haplotype sequences were deposited in the DDBJ/EMBL/GenBank database (LC421491–LC421502).

AMOVA was performed with three layers, between lineages, among populations within lineages and among individuals within populations (Table 3). Both nSSR and cpDNA haplotypes showed significant divergence between lineages with Φ_{CT} values of 0.058 and 0.308, respectively.

Leaf morphological variation

Based on the definitions of varieties *kobus* and *borealis* by Ohashi (2015) and the distributions of average values of leaf length and width within trees in each population, we classified populations into two varieties (Table 1 and Fig. 4). Populations 1–10, 12–14, and 18 were classified into variety *borealis*. Populations 11, 15–17 and 20 were classified into variety *kobus*. However, because the distributions of average values of leaf width and length for populations 19, 21 and 22 fell into just the boundary between the two varieties, we could not determine varieties for these three populations. Moreover, we

examined three principal components (PCs) estimated by SHAPE which were more than 5% contributions to the overall variance (Table S3). PC1 reflected relative leaf width, for which populations in central to northern Honshu Island and Hokkaido Island had large values (Fig. 5). PC2 reflected the position of the maximum width and populations in Kyushu Island had small values. PC3 reflected leaf curvature, with positive and negative values indicating left and right, respectively. As symmetry of curvature had no ecological meanings, the values of PC3 were log-absolute transformed and they were shown on the map. However, no geographical tendencies were observed. Populations located on the Sea of Japan side from central to northern Japan and on Hokkaido Island had large leaf area values. The populations classified as variety *borealis* showed large values in both PC1 and leaf area.

Four BioPCs were more than 5% contributions to the overall variance (Fig S4). All explanatory variables (BioPC2, BioPC3, BioPC4 and Q) were significant in the Bayesian linear mixed effect models of PC1 and leaf area (Table 4). Only two explanatory variables related to environmental factors (BioPC2 and BioPC3) were significant in the model of PC2. However, no explanatory variables in the model of PC3 were significant.

Population demography

In the model choice among the three population size change models, the population growth model (PGM) and the standard neutral model (SNM) were selected as the best models for, respectively, the northern and the southern lineage, with high posterior probabilities of 0.973 and 0.837, and low classification error rates of 0.193 and 0.192

(Table 5). All posterior distributions of the parameters of the best models showed clear single peaks (Fig. S5 and Table 6).

Using the information obtained from population size change analyses, we constructed four population divergence models with different migration patterns. Model choice among the models was carried out and the isolation model (ISM) was selected as the best model with posterior probability 0.842 (Table 5). Although the classification error rate for population divergence analysis was a little high (0.323), since the probabilities that the other models were wrongly assigned to ISM were very low (from 0.005 to 0.068; Table S4), we concluded that the correct model was selected by ABC-RF. All posterior distributions of the best model showed clear single peaks (Fig. S6). Estimated posterior modes (95% HPD) of current effective population sizes for the northern and southern lineages (N_N and N_S) were 73,700 (30,100–142,000) and 120,300 (69,700–149,100), respectively (Table 6). The values were similar between the two lineages. The posterior mode (95% HPD) of divergence time (T_{DIV}) was 11,300 (4,700–32,000) generations ago.

Posterior predictive checking of the best models for both population size change and population divergence analyses showed good fits of the estimated models to the observed data (Figs S7 and S8).

Ecological niche modeling

The accuracy of ecological niche modeling was high with the average ± standard deviation of area under the curve (AUC) being 0.987 ± 0.002. The predicted distributions projected onto the current climate were a good fit to the species' range except for Shikoku

Island, where the species was predicted to be present currently as well as in LGM and LIG, but has no current populations of *M. kobus* (Figs. 1 and 6). The climate variable that made the greatest contribution to the total variance was the precipitation in the coldest quarter (bio19, 58.1%). Using this model, potential distribution maps for the LGM (MIROC and CCSM) and LIG were created (Fig. 6).

Although under both LGM climate scenarios, several potential distribution areas in coastal regions south of 36˚N for both Sea of Japan and Pacific Ocean sides were predicted with high probability ($P > 0.6$), the distribution areas with high probability north of 36˚N were much smaller in CCSM than in MIROC. In Jeju, Honshu north of 41˚N and Hokkaido, no potential distribution areas with $P > 0.3$ were detected. The predicted potential distribution areas in LIG and present were very similar.

Discussion

What factors contributed to producing the current genetic variation in *Magnolia kobus*?

Magnolia kobus showed a hierarchical genetic structure. At the highest level, two lineages, the northern and southern lineages, were identified and their geographical boundary was estimated to be at 39˚N. The northern lineage consisted of two genetic clusters, whereas the southern one contained multiple genetic clusters. The level of genetic variation in nSSRs was significantly lower in the northern lineage than in the southern one. The

cpDNA haplotype of the northern lineage was also fixed; there was only one haplotype, H. The F_{ST} value between the northern lineage and the common ancestral population at $K = 2$ was more than one hundred times larger than that between the southern one and the ancestral population. Ecological niche modeling predicted that the probabilities of occurrence for this species in the northern area of Honshu (> 39˚N) and in Hokkaido during the LGM were very low. These results suggest that genetic variation in the northern lineage has been affected by severe genetic drift during one or more past glacial periods. The significant genetic differentiation between the northern and southern lineages, and the lower level of genetic diversity in the northern lineage, that we observed in *M. kobus* are common trends among temperate tree species that are widely distributed along the Japanese archipelago (Hiraoka and Tomaru 2009; Sakaguchi et al. 2011; San Jose-Maldia et al. 2017; Tamaki et al. 2018; Tsumura et al. 2007; Tsumura et al. 2014). However, the locations of boundaries between northern and southern lineages are different among species. For example, the lineages of *Fagus crenata* and *Cryptomeria japonica* diverged between Sea of Japan and Pacific Ocean sides (Hiraoka and Tomaru 2009; Tsumura et al. 2014), those of *Kalopanax septemlobus* diverged gradually from central Honshu to Hokkaido (Sakaguchi et al. 2011), and those of *M. salicifolia* and *Quercus aliena* clearly diverged in the central Honshu (San Jose-Maldia et al. 2017; Tamaki et al. 2018). The boundary of two lineages of *M. kobus* is located in the northern part of Honshu and is rather geographically close to that of two lineages within *Pinus pumila* (Tani et al. 1996) and within *Betula maximowicziana* (Tsuda and Ide 2005, 2010), which are cold tolerant tree species and distributed further north.

The possibility of the existence of multiple past refugia in the southern part of Japan is supported by high genetic diversity in nSSRs, the existence of multiple genetic clusters in the southern lineage, endemic cpDNA haplotypes (haplotypes I and J) and several areas with high probabilities of potential distribution during the LGM ($P > 0.6$) on both Sea of Japan and Pacific Ocean sides in southern Japan. Although most of the populations of the southern lineage showed high genetic diversity and low F_{ST} values between each cluster and the ancestral population, two populations (20 and 23) did not. In the case of population 23 (Jeju), this population might have been affected by a severe historical isolation effect, as is suggested by the low probability of occurrence and the discontinuity with surrounding populations during both glacial and inter-glacial periods that were predicted by ecological niche modeling. Populations of *Rhododendron weyrichii*, which is a shrub growing in the warm temperate zone from Jeju to southern parts of Honshu Islands, on Jeju Island also showed long-term low effective population size and severe isolation as predicted by ABC and ecological niche modeling, respectively (Yoichi et al. 2016). Like *R. weyrichii*, *M. kobus* may have survived through a glacial period only in very restricted areas within Jeju Island.

From the comparison of three population size change models, histories of exponential growth from a small number of founders and stable population size were estimated for the northern and the southern lineage, respectively. These population size change histories selected by ABC are closely congruent with the hypotheses discussed in the paragraphs above based on diversity indexes and predictions from ecological niche modeling. The posterior mode (95% HPD) of divergence time estimated by the isolation

model was 11,300 (4,700–32,000) generations ago. To convert the divergence time into years ago, we must assume a generation time for *M. kobus*. It is reported that *M. kobus* starts flowering after 10–30 years from the seedling stage under garden conditions (Callaway 1994). As far as we know, there is no information about it in the wild. However, Takahashi et al. (2006) reported that average ± SD values for initial flowering and fruiting ages in the tree type of *Magnolia salicifolia*, which is a species related to *M. kobus*, were 17.6 ± 6.57 and 20.4 ± 6.70, respectively. Moreover, these authors reported that its maximum lifespan was more than 100 years and most individuals whose age was more than 50 years were well-grown trees at the canopy layer in the populations they studied. *M. kobus* can generally grow larger than *M. salicifolia*. With these factors taken into account, the generation time of *M. kobus* is considered to be 50 years or more. If we assume the generation time to be 50 years, the divergence time can be estimated at 565,000 (235,000–1,600,000) years ago. As we have made two major assumptions, of mutation rate and generation time, when estimating the divergence time, and since the estimated value has some degree of uncertainty denoted by the HPD, we should be cautions about its interpretation (Tsuda et al. 2016); however, the outcome suggests that the two lineages of *M. kobus* may have experienced several glacial-interglacial cycles after their divergence. *M. salicifolia*, which is also widely distributed in the Japanese archipelago, showed a population demographic history similar to that of the northern and the southern lineages, and a more distant divergence time than that in this study, but the extent of migrations between lineages after divergence was significant (Tamaki et al. 2018). These findings indicate that the current genetic diversities of temperate tree species

that are broadly distributed in the Japanese archipelago may have been created not only by the effect of the most recent glacial period but also by the cumulative effects of several previous glacial periods. Model choice in population divergence analysis did not show significant migration between the two lineages after divergence. It can therefore be inferred that the refugia of the two lineages were different. Ecological niche modeling predicted a potential distribution area ($0.3 < P \leq 0.6$) in LGM near the coast from 39˚–40˚N on the Sea of Japan side, where the tail of the current distribution of the northern lineage is located, and it is considered to have been the glacial refugium for the northern lineage (Fig. 6). The possibility of a refugium near the coast of northern Japan on the Sea of Japan side was also reported for *C. japonica* (Kimura et al. 2014). In our study populations near the boundary (10 and 11) showed more than 5% genetic admixture of the recessive cluster. This may be evidence of recent hybridization between the two lineages due to secondary contact. Similar admixture patterns have been also reported in *B. maximowicziana* almost at the same region (Tsuda et al. 2015). Moreover, the cited authors have proved that the observed admixture pattern in *B. maximowicziana* was created not by secondary contact but by simple population split (Tsuda et al. 2015). To clarify whether the observed admixture pattern in *M. kobus* was created by secondary contact or simple population split, more intensive population sampling near the boundary would be needed.

Relationship between geographical patterns of genetic variation and of leaf morphological traits

Populations from central Honshu on the Sea of Japan side to Hokkaido (1–10, 12–14 and 18) had wide leaves and large area. Populations with these leaf morphological traits well corresponded to *M. kobus* var. *borealis* (Ohashi 2015). Although the changes in leaf morphological traits were to some extent continuous and we could not identify varieties for some southern populations (Fig. 4), it was possible to confirm the existence of *M. kobus* var. *borealis*, which were distributed from Sea of Japan side of central to northern Honshu and Hokkaido, based on these morphological characters. However, the genetic delimitation between lineages occurred at 39˚N, further north than the morphological one (36˚N). Because variety *borealis* recognized based on morphological characters comprises not only the northern lineage but also the southern one, we can conclude that it is not supported genetically and systematically.

Variety *borealis* appeared across both the northern and southern lineages. However, ABC analysis detected no significant historical migration between lineages after divergence. Population admixture was only detected in populations 11 and 17 for the southern lineage, but these populations showed leaf morphology of variety *kobus*. Therefore, it is unlikely that introgression has contributed to shape the morphology of variety *borealis* in some southern populations. We detected significant effects of environments on relative leaf width (PC1), the position of the maximum leaf width (PC2) and leaf area. Moreover, we also detected significant effects of population history on relative leaf width and leaf area. These results suggest that variation in the position of the maximum leaf width has been shaped by only natural selection, while variations in relative leaf width and leaf area have been shaped by both natural selection and population

demographic history. However, it is possible that variations in these leaf morphological traits affected by natural selection are expressed through phenotypic plasticity and functional gene products (Ramírez-Valiente et al. 2010). In order to distinguish them a provenance test will be required.

Conservation implications

Conservation units should be determined for *M. kobus* because there is clear genetic divergence between the northern and the southern lineages at ca. 39˚N. Since the northern lineage consisted of single / two genetic clusters in nSSR (at $K \leq 6$ / $K > 6$, respectively), only one cpDNA haplotype and single variety, and the level of isolation by distance was relatively low, seed and/or seedling transfer may be permitted within the area north of 39˚N, although the information about genetic differentiation of adaptive traits would be ideally needed to determine the range of seed and/or seedling transfer. On the other hand, since the southern lineage contained multiple genetic clusters, most populations showed clear genetic structure dominated by a single distinct genetic cluster in nSSR, composition of cpDNA haplotype were different among populations and there were multiple varieties, it is obvious that seed/seedling transfer should be restricted even within the southern lineage and the optimal seed source would be the nearest natural population. In the southern lineage (< 39˚N), for example, prefectural level conservation units may be practicable (prefectural borders are shown by dotted lines in Fig. 1). Many studies have reported the existence in natural forests of *M. kobus* seedlings that have escaped from trees planted near the forest (Fujii 1997; Ishida et al. 2008; Tamaki et al. 2016). To prevent

genetic disturbance and to conserve genetic resources in natural forests, the provenance of seeds/seedlings should be considered carefully even when planting roadside trees especially in the southern distribution area.

Acknowledgements

We thank Yoichi Watanabe of Chiba University for his help in genetic data analyses, and the members of the Laboratory of Forest Ecology and Physiology of Nagoya University, and Kazunori Takahashi of Forestry and Forest Products Research Institute, for their assistance in sampling material. We also thank anonymous reviewers for their helpful comments on the previous manuscript.

Conflict of interest

The authors declare no conflicts of interest.

Funding

This research was founded by a grant for research on Development of Evaluation

Methods of Genetic Diversity in Broad-leaved Trees (2010–2012) from Forestry Agency, Ministry of Agriculture, Forestry and Fisheries, Japan.

References

Callaway DJ (1994) The world of Magnolias. Timber Press, Oregon, USA

Campana MG, Hunt HV, Jones H, White J (2011) *Corrsieve*: software for summarizing and evaluating Structure output. Molecular Ecology Resources 11:349–352

Chen C, Lu R, Zhu S, Tamaki I, Qiu Y (2017) Population structure and historical demography of *Dipteronia dyeriana* (Sapindaceae), an extremely narrow palaeoendemic plant from China: implications for conservation in a biodiversity hotspot. Heredity 119:95–106

Chessel D, Dufour AB, Thioulouse J (2004) The ade4 package-I: One-table methods. R News 4:5–10

Chybicki IJ, Burczyk J (2009) Simultaneous estimation of null alleles and inbreeding coefficients. Journal of Heredity 100:106–113

Clement M, Posada D, Crandall K (2000) TCS: a computer program to estimate gene genealogies. Molecular Ecology 9:1657–1660

Collins WD, Bitz CM, Blackmon ML, Bonan GB, Bretherton CS, Carton JA, Chang P, Doney SC, Hack JJ, Henderson TB, Kiehl JT, Large WG, McKenna DS, Santer BD, Smith RD (2006) The community climate system model version 3 (CCSM3). Journal of Climate 19:2122–2143

Csillery K, Francois O, Blum MGB (2012) abc: an R package for approximate Bayesian computation (ABC). Methods in Ecology and Evolution 3:475–479

Duminil J, Di Michele M (2009) Plant species delimitation: a comparison of morphological and molecular markers. Plant Biosystems 143:528–542

Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32:1792–1797

Estoup A, Jarne P, Cornuet J-M (2002) Homoplasy and mutation model at microsatellite

loci and their consequences for population genetics analysis. Molecular Ecology 11:1591–1604

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 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 HE (2010) Arlquin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources 10:564–567

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164:1567–1587

Fujii T (1997) Soil seed banks of fragmented forests. Human and Nature 8:113–124 (in Japanese with English abstract)

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2014) Bayesian Data Analysis, 3rd edn. CRC Press, Boca Raton, Florida, US

Goudet J (1995) FSTAT (version 1.2): a computer program to calculate *F*-statistics. Journal of Heredity 86:485–486

Hagiwara S (1977) Clines on leaf size of beech *Fagus crenata*. Species Biological Research 1:39–51 (in Japanese)

Hashizume H, Lee NH, Yamamoto F (1997) Variation in the leaf shape of planted trees of *Fagus crenata* Blume among provinances. Applied Forest Science 6:115–118 (in Japanese with English abstract)

Hasumi H, Emori S (2004) K-1 Coupled GCM (MIROC) Description. Center for Climate System Research, University Tokyo, National Institute for Environmental Studies, Frontier Research Center for Global Change, Tokyo Japan

Hedrick PW (2005) A standardized genetic differentiation measure. Evolution 59:1633–1638

Hiraoka K, Tomaru N (2009) Genetic divergence in nuclear genomes between

1 populations of *Fagus crenata* along the Japan Sea and Pacific sides of Japan.
2 Journal of Plant Research 122:269–282
3 Hiraoka Y, Tamaki I, Watanabe A (2018) The origin of populations of *Toxicodendron*
4 *succedaneum* on mainland Japan revealed by genetic variation in chloroplast and
5 nuclear DNA. Journal of Plant Research 131:225–238
6 Hotta M (1974) Evolutionary biology in plants III: history and geography of plants.
7 Sanseido, Tokyo, Japan (in Japanese)
8 Isagi Y, Kanazashi T, Suzuki W, Tanaka H, Abe T (1999) Polymorphic microsatellite
9 DNA markers for *Magnolia obovata* Thunb. and their utility in related species.
10 Molecular Ecology 8:698–700
11 Ishida H, Toi K, Takeda Y, Hattori T (2008) Invasion of fragmented, secondary deciduous
12 oak forests by greening/garden trees in urban areas of Japan. Japanese Journal of
13 Conservation Ecology 13:1–16 (in Japanese with English abstract)
14 Iwata H, Ukai Y (2002) SHAPE: A computer program package for quantitatibe evaluation
15 of biological shapes based on elliptic Fourier descriptors. Journal of Heredity
16 93:384–385
17 Jost L (2008) G_{ST} and its relatives do not measure differentiation. Molecular Ecology
18 17:4015–4026
19 Kimura MK, Uchiyama K, Nakao K, Moriguchi Y, San Jose-Maldia L, Tsumura Y (2014)
20 Evidence for cryptic nothern refugia in the last glacial period in *Cryptomeria*
21 *japonica*. Annals of Botany 114:1687-1700
22 Koike T, Maruyama Y (1998) Comparative ecophysiology of the leaf photosynthetic traits
23 in Japanese beech growth in provenances facing the Pacific Ocean and Sea of
24 Japan. Journal of Phytogeography and Taxonomy 46:23–28 (in Japanese with
25 English abstract)
26 Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) CLUMPAK:
27 a program for identifying clustering modes and packaging population structure
28 inferences across *K*. Molecular Ecology Resources 15:1179–1191
29 Koyama Y (2012) A conservation genetic study to define conservation units of *Fagus*
30 *crenata*. PhD thesis, Nagoya University, Nagoya, Japan
31 Koyama Y, Yagihashi T, Migita C, Tanaka N (2002) Geographical gradient of leaf area of
32 *Fagus crenata* in the Nagano region, Central Japan. Japanese Journal of Forest
33 Environment 44:31–33 (in Japanese with English abstract)

Lefèvre F (2004) Human inpacts on forest genetic resources in the temperate zone: an updated review. Forest Ecology and Management 197:257–271

Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acids Research 8:4321–4326

Muse SV (2000) Examining rates and patterns of nucleotide substitution in plants. Plant Molecular Biology 42:25–43

Nagamitsu T, Shimada K, Kanazashi A (2015) A reciprocal transplant trial suggests a disadvantage of northward seed transfer in survival and growth of Japanese red pine (*Pinus densiflora*) treea. Tree Genetics & Genomes 11:813

Ohashi H (2015) Magnoliaceae. In: Ohashi H, Kadota Y, Murata J, Yonekura K, Kihara H (eds) Wild Flowers of Japan, vol 1. Heibonsha, Tokyo, Japan, pp 71–74 (in Japanese)

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecological Modelling 190:231–259

Plummer M, Best N, Cowles K, Vines K (2006) CODA: Convergence Diagnosis and Output Analysis for MCMC. R News 6:7–11

Potts BM, Barbour RC, Hingston AB, Vaillancourt RE (2003) Genetic pollution of native eucalypt gene pools—identifying the risks. Australian Journal of Botany 51:1–25

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945-959

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

R Core Team (2018) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Ramírez-Valiente JA, Sánchez-Gómez D, Aranda I, Valladares F (2010) Phenotypic plasticity and local adaptation in leaf ecophysiological traits of 13 contrasting cork oak populations under different water availabilities. Tree Physiology 30:618–627

Sakaguchi S, Qiu Y-X, Liu Y-H, Qi X-S, Kim S-H, Han J, Takeuchi Y, Worth JRP, Yamasaki M, Sakurai S, Isagi Y (2012) Climate oscillation during the Quaternary associated with landscape heterogeneity promoted allopatric lineage divergence of a temperate tree *Kalopanax septemlobus* (Araliaceae) in East Asia. Molecular Ecology 21:3823–3838

Sakaguchi S, Takeuchi Y, Yamazaki M, Sakurai S, Isagi Y (2011) Lineage admixture

during postglacial range expansion is responsible for the increased gene diversity of *Kalopanax septemlobus* in a recently colonised territory. Heredity 107:338-348

San Jose-Maldia L, Matsumoto A, Ueno S, Kanazashi A, Kanno M, Namikawa K, Yoshimaru H, Tsumura Y (2017) Geographic patterns of genetic variation in nuclear and chloroplast genomes of two related oaks (*Quercus aliena* and *Q. serrata*) in Japan: implications for seed and seedling transfer. Tree Genetics & Genomes 13:121

Setsuko S, Ueno S, Tsumura Y, Tomaru N (2005) Development of microsatellite markers in *Magnolia stellata* (Magnoliaceae), a threatened Japanese tree. Conservation Genetics 6:317–320

Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, Miller J, Siripun KC, Winder CT, Schilling EE, Small RL (2005) The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. American Journal of Botany 92:142-166

Shaw J, Lickey EB, Schilling EE, Small RL (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. American Journal of Botany 94:275-288

Stan Development Team (2018) RStan: the R interface to Stan. R package version 2.18.2. http://mc-stan.org/. Accessed April, 12th 2019

Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Molecular Biology 17:1105–1109

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595

Takahashi K, Hoshizaki K, Masaki T, Osumi K (2006) Comparative analysis of size structure and life span between *Magnolia salicifolia* and the dwarf sub-species newly found. Kanto Journal of Forest Research 57:113–114 (in Japanese)

Takasuna H, Takayama H (2011) Growth of the seedlings of eight deciduous broad-leaved tree species and change of vegetation structure during fifteen years. Journal of Japanese Society of Revegetation Technology 37:108–113 (in Japanese with English abstract)

Tamaki I, Kawashima N, Setsuko S, Itaya A, Tomaru N (2018) Morphological and genetic

1 divergence between two lineages of *Magnolia salicifolia* (Magnoliaceae) in Japan.
2 Biological Journal of the Linnean Society 125:475–490
3 Tamaki I, Mizuno M, Yanagisawa N, Tsuda K, Nakagawa Y, Itaya A (2016) Tree
4 community structure and species diversity of naturally regenerated forests
5 remaining in urban green areas of northeastern Nagoya, Japan. Japanese Journal
6 of Conservation Ecology 21:93–102 (in Japanese with English abstract)
7 Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5:
8 Molecular evolutionary genetic analysis using maximum likelihood, evolutionary
9 distance, and maximum parsimony methods. Molecular Biology and Evolution
10 28:2731–2739
11 Tani N, Tomaru N, Araki M, Ohba K (1996) Genetic diversity and differentiation in
12 populations of Japanese stone pine (Pinus pumila) in Japan. Canadian Journal of
13 Forest Research 26:1454–1462
14 Tsuda Y, Chen J, Stocks M, Kallman T, Sonstebo JH, Parducci L, Semerikov V, Sperisen
15 C, Politov D, Ronkainen T, Valiranta M, Vendramin GG, Tollefsrud MM, Lascoux
16 M (2016) The extent and meaning of hybridization and introgression between
17 Siberian spruce (*Picea obovata*) and Norway spruce (*Picea abies*): cryptic refugia
18 as stepping stones to the west? Molecular Ecology 25:2773–2789
19 Tsuda Y, Ide Y (2005) Wide-range analysis of genetic structure of Betula
20 maximowicziana, a log-lived pioneer tree species and noble hardwood in the cool
21 temperate zone of Japan. Molecular Ecology 14:3929–3941
22 Tsuda Y, Ide Y (2010) Chloroplast DNA phylogeography of *Betula maximowicziana*, a
23 long-lived pioneer tree species and noble hardwood in Japan. Journal of Plant
24 Research 123:343–353
25 Tsuda Y, Nakao K, Ide Y, Tsumura Y (2015) The population demography of *Betula*
26 *maximowicziana*, a cool-temperate tree species in Japan, in relation to the last
27 glacial period: its admixture-like genetic structure is the result of simple
28 population splitting not admixing. Molecular Ecology 24:1403–1418
29 Tsumura Y, Kado T, Takahashi T, Tani N, Ujino-Ihara T, Iwata H (2007) Genome scan to
30 detect genetic structure and adaptive genes of natural populations of *Cryptomeria*
31 *japonica*. Genetics 176:2393-2403
32 Tsumura Y, Uchiyama K, Moriguchi Y, Kimura M, Ueno S, Ujino-Ihara T (2014) Genetic
33 differentiation and evolutionary adaptation in *Cryptomeria japonica*. G3 Genes

Genomics Genetics 4:2389-2402
Ueda K (2006) Magnoliaceae. In: Iwatsuki K, Boufford DE, Ohba H (eds) Flora of Japan IIa. Kodansha Scientific, Tokyo, Japan, pp 231-234
Wang J (2017) The computer program STRUCTURE for assigning individuals to populations: easy to use but easier to misuse. Molecular Ecology Resources 17:981–990
Yoichi W, Tamaki I, Sakaguchi S, Song J-S, Yamamoto S, Tomaru N (2016) Population demographic history of a temperate shrub, *Rhododendron weyrichii* (Ericaceae), on continental islands of Japan and South Korea. Ecology and Evolution 6:8800–8810

Figure legends

Fig. 1 Distribution ranges of *Magnolia kobus* (gray area), the locations of the 23 populations sampled (black dots), proportions of genetic clusters detected by STRUCTURE for nuclear microsatellites (pie chart), chloroplast DNA haplotypes detected (bold type letter) and the network they formed with outgroup data [H, I and J were found in *M. kobus*, and A–G were found in its congener, *M. salicifolia* (Tamaki et al. 2018)]. Dotted lines within the Japanese archipelago indicate prefectural borders.

Fig. 2 Comparison of three population size change (A) and four population divergence (B) models. SNM, standard neutral model; PGM, population growth model; SRM, size reduction model; ISM, isolation model; IMM, isolation with migration model; IMM_{NS}, model of isolation with one way migration from the northern to the southern lineages; IMM_{SN}, model of isolation with one way migration from the southern to the northern lineages. Direction of migration is backward-in-time.

Fig. 3 Changes in log probability of data (A) and ΔK (B) along the number of genetic clusters (K) in STRUCTURE analysis of 453 *Magnolia kobus* individuals sampled from 23 populations. Distributions of genetic clusters in each individual from $K = 2$ to 14 (C).

Fig. 4 Distributions of average values of leaf length and width within trees for 22 *Magnolia kobus* populations (the Jeju population, No. 23, was excluded from this

analysis). Arrows indicate the ranges of leaf length and width for varieties *kobus* and *borealis* according to Ohashi (2015). The population numbers (see Table 1) and variety names are shown on each panel. We classified populations into two varieties based on the distributions of average values of leaf length and width. “intermediate” indicates that we could not determine varieties because the distributions were intermediate between those for two varieties.

Fig. 5 Geographical changes in leaf shape and area for 22 *Magnolia kobus* populations (the Jeju population, No. 23, was excluded from this analysis). Leaf shape was extracted with an elliptic Fourier method and converted into principal components (PCs) by SHAPE. Only PCs whose contribution to the overall variance was more than 5% are shown.

Fig. 6 Inferred potential areas for *Magnolia kobus* at the present, the last inter-glacial (LIG, 130 kya) and the last glacial maximum (LGM, 21 kya) based on the community climate system model (CCSM) and the model for interdisciplinary research on climate (MIROC). *P* indicates probability of occurrence.

Table 1 Location, sample size, genetic variation and leaf size of 23 *Magnolia kobus* populations

Population													Leaf	
No.	Name	Latitude	Longitude	Lineage [a]	Variety [b]	N_n	N_c	N_m	A_R	H_E	F_{IS} [c]		Length (cm)	Width (cm)
1	Nakagawa	44.78	142.29	Northern	*borealis*	14	2	14	4.02	0.700	0.074	N.S.	12.9	6.7
2	Chimikepp lake	43.66	143.88	Northern	*borealis*	27	2	27	4.24	0.734	0.076	N.S.	14.1	7.3
3	Nopporo	43.05	141.51	Northern	*borealis*	24	2	24	4.33	0.735	0.079	N.S.	13.4	7.4
4	Furano	43.25	142.41	Northern	*borealis*	25	2	26	4.26	0.726	0.085	*	15.3	7.7
5	Ashoro	43.30	143.50	Northern	*borealis*	7	2	8	4.06	0.731	0.098	N.S.	14.8	7.6
6	Tomakomai	42.68	141.60	Northern	*borealis*	23	2	23	4.12	0.700	0.087	*	13.2	7.2
7	Erimo	42.05	143.29	Northern	*borealis*	18	2	18	4.06	0.703	0.106	*	12.3	6.5
8	Towada lake	40.45	140.84	Northern	*borealis*	21	2	21	4.11	0.708	0.081	N.S.	12.7	6.7
9	Akita	39.84	140.12	Northern	*borealis*	12	2	12	4.54	0.758	0.096	N.S.	13.2	7.1
10	Morioka	39.72	141.20	Northern	*borealis*	15	2	15	4.26	0.732	0.106	*	12.8	7.3
11	Arasawa	38.55	140.68	Southern	*kobus*	30	2	30	4.84	0.798	0.191	***	10.2	5.5
12	Kakudayama	37.77	138.82	Southern	*borealis*	22	2	22	5.02	0.832	0.012	N.S.	13.8	7.6
13	Kurohime	36.82	138.16	Southern	*borealis*	19	2	19	4.85	0.793	0.082	*	12.0	7.0
14	Kurikara	36.66	136.84	Southern	*borealis*	17	1	17	5.05	0.848	0.088	*	13.9	7.4
15	Hokyosan	36.17	140.13	Southern	*kobus*	28	2	25	5.30	0.845	0.041	N.S.	9.9	5.3
16	Nagara	35.45	140.21	Southern	*kobus*	7	2	7	4.41	0.789	0.026	N.S.	10.1	5.3
17	Yamanaka lake	35.41	138.87	Southern	*kobus*	31	2	30	5.06	0.825	0.011	N.S.	10.8	5.0

18	Biwa lake	35.41	136.01	Southern	*borealis*	28	2	28	4.61	0.794	0.008	N.S.	13.5	6.8
19	Toyooka	35.48	134.80	Southern	intermediate	5	2	5	4.76	0.857	0.340	***	12.2	6.5
20	Kirigaya	34.71	132.19	Southern	*kobus*	29	2	28	3.68	0.676	0.125	***	12.1	5.9
21	Kujuzan	33.12	131.24	Southern	intermediate	28	2	28	5.09	0.837	0.075	**	12.4	6.0
22	Otori-kyo	31.52	131.00	Southern	intermediate	7	2	7	4.98	0.859	0.245	***	12.8	6.3
23	Jeju	33.43	126.63	Southern	–	16	4	–	3.34	0.580	0.146	**	–	–
Average / overall														
	Northern					18.6	2.0	18.8	4.20[d]	0.723[e]	0.087[f]			
	Southern					20.5	2.1	20.5	4.69[d]	0.795[e]	0.081[f]			
	All					19.7	2.0	19.7	4.48	0.764	0.097			

N_n, number of individuals for analysis of nuclear microsatellite; N_c, number of individuals for analysis of chloroplast DNA sequences; N_m, number of individuals for analysis of leaf morphology; A_R, allelic richness based on four diploid individuals; H_E, expected heterozygosity; F_{IS}, fixation index.

[a] Lineages were determined by STRUCTURE analysis.

[b] Varieties were determined by leaf morphology.

[c] The significance of departure from Hardy-Weinberg equilibrium was tested by randomization test. *P*-values were adjusted by Bonferroni correction. N.S. not significant; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

[d] Southern > northern (permutation test, $P = 0.017$).

[e] Southern > northern (permutation test, $P = 0.016$).

[f] The difference in F_{IS} between the two lineages was not significant (permutation test).

Table 2 Genetic diversity at 13 nuclear microsatellite loci across 23 *Magnolia kobus* populations

Locus	*A*	H_S	H_T	F_{ST}	G'_{ST}	*D*
M6D8[a]	24	0.796	0.917	0.125	0.670	0.620
stm0002[b]	19	0.761	0.892	0.154	0.636	0.573
stm0114[b]	17	0.388	0.585	0.350	0.559	0.337
stm0163[b]	21	0.823	0.893	0.080	0.457	0.413
stm0184[b]	30	0.839	0.924	0.097	0.593	0.552
stm0200[b]	33	0.781	0.828	0.062	0.265	0.224
stm0214[b]	19	0.740	0.867	0.150	0.580	0.511
stm0246[b]	34	0.911	0.952	0.045	0.503	0.482
stm0251[b]	16	0.642	0.738	0.125	0.377	0.280
stm0353[b]	22	0.860	0.927	0.076	0.534	0.500
stm0383[b]	36	0.880	0.936	0.059	0.520	0.488
stm0423[b]	37	0.815	0.935	0.134	0.718	0.678
stm0448[b]	12	0.672	0.813	0.185	0.547	0.449
Average / overall	24.6	0.762	0.862	0.119	0.504	0.439

A, number of alleles; H_S, average gene diversity within populations; H_T, gene diversity in the total population; F_{ST}, Weir & Cockerham's F_{ST}; G'_{ST}, Hedrick's standardized G_{ST}; *D*, Jost's *D*.

[a] Isagi et al. (1999).

[b] Setsuko et al. (2005).

Table 3 Results from analysis of molecular variance for nuclear microsatellites and chloroplast DNA haplotypes

	Nuclear microsatellites			Chloroplast DNA haplotypes		
Layer	Variance component (%)	Φ–statistics		Variance component (%)	Φ–statistics	
Between lineages	5.8	$\Phi_{CT} = 0.058$	***	30.8	$\Phi_{CT} = 0.308$	**
Among populations within lineages	8.8	$\Phi_{SC} = 0.094$	***	43.4	$\Phi_{SC} = 0.627$	**
Among individuals within populations	85.4	$\Phi_{ST} = 0.146$	***	25.8	$\Phi_{ST} = 0.741$	***

** $P < 0.01$, *** $P < 0.001$.

Table 4 Posterior mode (95% highest posterior density) of parameters for the Bayesian linear mixed effect model explaining leaf morphological traits for *Magnolia kobus*. Leaf shape was extracted with an elliptic Fourier method and converted into principal components (PCs) by SHAPE. β_{0_Mean} is the average intercept among individuals. The other βs are regression coefficients. βs that were significantly deviated from 0 are shown in bold. σs are standard deviation parameters with a normal distribution in the model.

Parameter	PC1	PC2	log (|PC3|)	Leaf area
β_{0_Mean}	**-0.0086**	0.0013	**-4.176**	**46.40**
	(-0.0145 – -0.0024)	(-0.0019 – 0.0039)	**(-4.227 – -4.106)**	**(44.73 – 47.97)**
β_{BioPC2}	**-0.0042**	**-0.0057**	-0.0294	**-3.814**
	(-0.0067 – -0.0021)	**(-0.0068 – -0.0046)**	(-0.0494 – 0.0006)	**(-4.362 – -3.078)**
β_{BioPC3}	**0.0030**	**0.0031**	-0.0206	**-3.133**
	(0.0005 – 0.0055)	**(0.0019 – 0.0042)**	(-0.0379 – 0.0116)	**(-3.805 – -2.529)**
β_{BioPC4}	**-0.0066**	-0.0011	-0.0027	**3.283**
	(-0.0107 – -0.0007)	(-0.0030 – 0.0013)	(-0.0487 – 0.0455)	**(2.102 – 4.686)**
β_{Q}	**0.0223**	-0.0004	-0.0843	**13.06**
	(0.0131 – 0.0317)	(-0.0047 – 0.0043)	(-0.1956 – 0.0051)	**(10.80 – 16.00)**
$\sigma_{Individual}$	0.0448	0.0204	0.308	11.61
	(0.0415 – 0.0479)	(0.0188 – 0.0219)	(0.269 – 0.362)	(10.80 – 12.59)
σ_{All}	0.0406	0.0227	1.059	13.10
	(0.0398 – 0.0417)	(0.0222 – 0.0232)	(1.036 – 1.082)	(12.80 – 13.38)

BioPC, principal component estimated by 19 bioclimatic variables; Q, membership coefficient of the northern lineage estimated by STRUCTURE at $K = 2$.

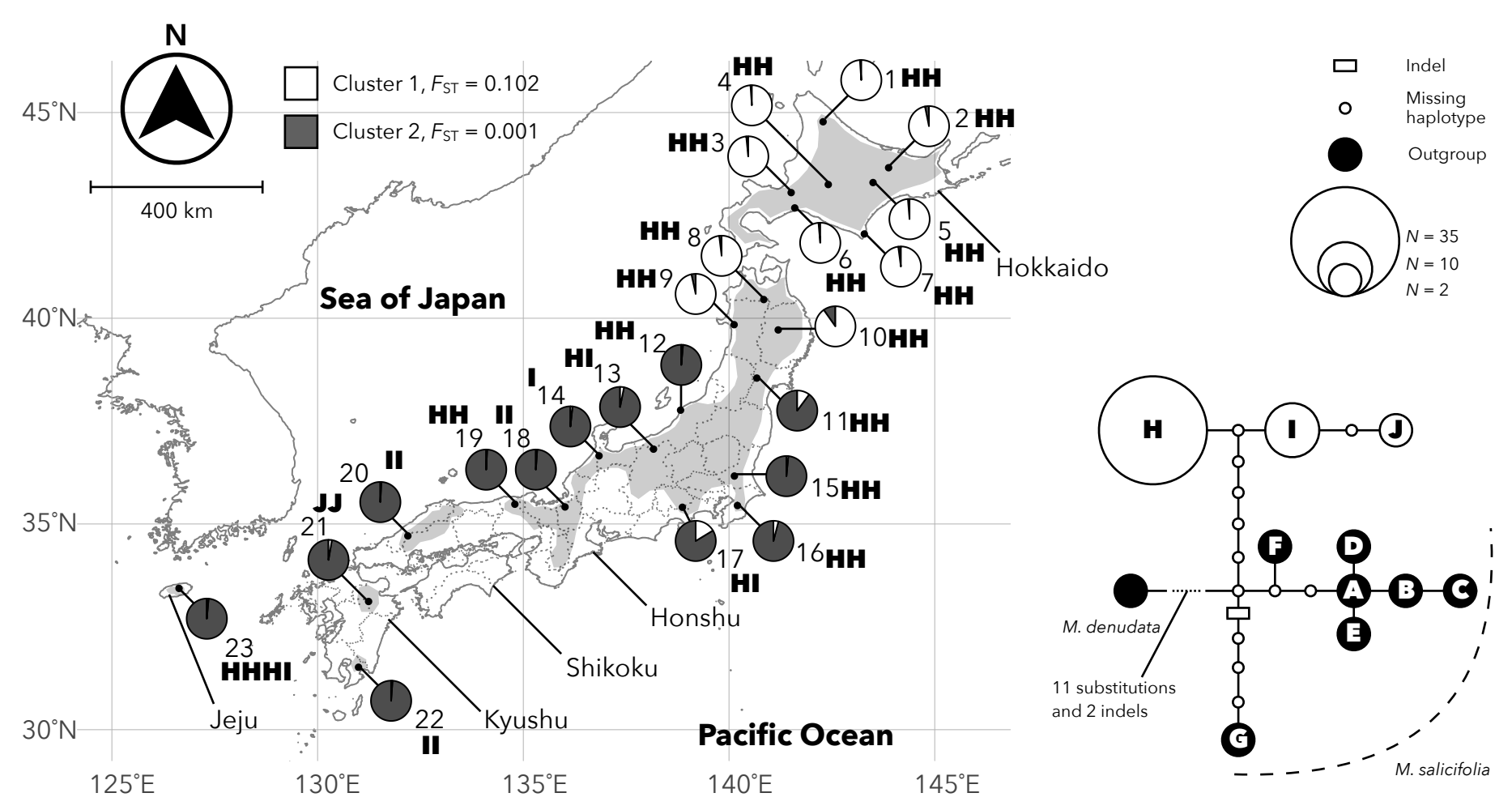
Table 5 Classification error rate, proportion of votes by random forest (RF) composed of 1,000 trees based on a trained set of 10,000 simulations, best model (shown in bold) selected by RF and its posterior probability

Analysis	Lineage	Classification error rate	Proportion of votes by RF							Posterior probability
			SNM	PGM	SRM	ISM	IMM	IMM_{NS}	IMM_{SN}	
Population size change	Northern	0.193	0.032	**0.962**	0.006	–	–	–	–	0.973
	Southern	0.192	**0.852**	0.012	0.136	–	–	–	–	0.837
Population divergence		0.323	–	–	–	**0.609**	0.019	0.043	0.329	0.842

Table 6 Posterior mode (95% highest posterior density) of parameters for population size change and population divergence models

	Population size change		Population divergence
Lineage	Northern	Southern	
Best model	PGM	SNM	ISM
N_{CUR} ($\times 10^4$)	0.95	5.63	–
	(0.18–10.17)	(2.91–14.49)	
G ($\times 10^{-4}$)	-1.56	–	Fixed to -1.56
	(-9.36–-0.24)		
N_N ($\times 10^4$)	–	–	7.37
			(3.01–14.20)
N_S ($\times 10^4$)	–	–	12.03
			(6.97–14.91)
T_{DIV} ($\times 10^4$)	–	–	1.13
			(0.47–3.20)
mean μ ($\times 10^{-4}$)	2.35	0.90	0.59
	(0.45–8.53)	(0.40–3.37)	(0.27–1.83)
shape	1.90	2.48	1.52
	(0.73–4.23)	(1.21–4.89)	(0.68–4.04)
mean P_{GSM}	0.585	0.421	0.440
	(0.488–0.664)	(0.192–0.554)	(0.237–0.571)

PGM, population growth model; SNM, standard neutral model; ISM, isolation model. N_{CUR}, current effective population size; G, population growth rate; N_N and N_S, current effective population size of the northern and southern lineages, respectively; T_{DIV}, divergence time; mean μ, *shape* and mean P_{GSM}, parameters of mutation model for nuclear microsatellites. Direction of migration is backward-in-time. The unit of effective population size is the number of diploid individuals. A negative value of G indicates exponential population growth from the past to the present. The unit of T_{DIV} is generations ago.


N
400 km
Cluster 1, F_{ST} = 0.102
Cluster 2, F_{ST} = 0.001
45°N
40°N
35°N
30°N
125°E
130°E
135°E
140°E
145°E
Sea of Japan
Pacific Ocean
Honshu
Shikoku
Kyushu
Jeju
Hokkaido
1 HH
2 HH
3 HH
4 HH
5 HH
6 HH
7 HH
8 HH
9 HH
10 HH
11 HH
12 HH
13 HI
14 I
15 HH
16 HH
17 HI
18 II
19 HH
20 II
21 JJ
22 II
23 HHHI
Indel
Missing haplotype
Outgroup
N = 35
N = 10
N = 2
H
I
J
F
D
A
B
C
E
G
M. denudata
11 substitutions and 2 indels
M. salicifolia

a

Past

Current

SNM

N_{CUR}

PGM

G

N_{CUR}

SRM

N_{ANC}

T

N_{CUR}

b

Past

Current

ISM

T_{DIV}

G

N_N

N_S

IMM

Nm_{NS}

Nm_{SN}

T_{DIV}

G

N_N

N_S

IMM_{NS}

Nm_{NS}

T_{DIV}

G

N_N

N_S

IMM_{SN}

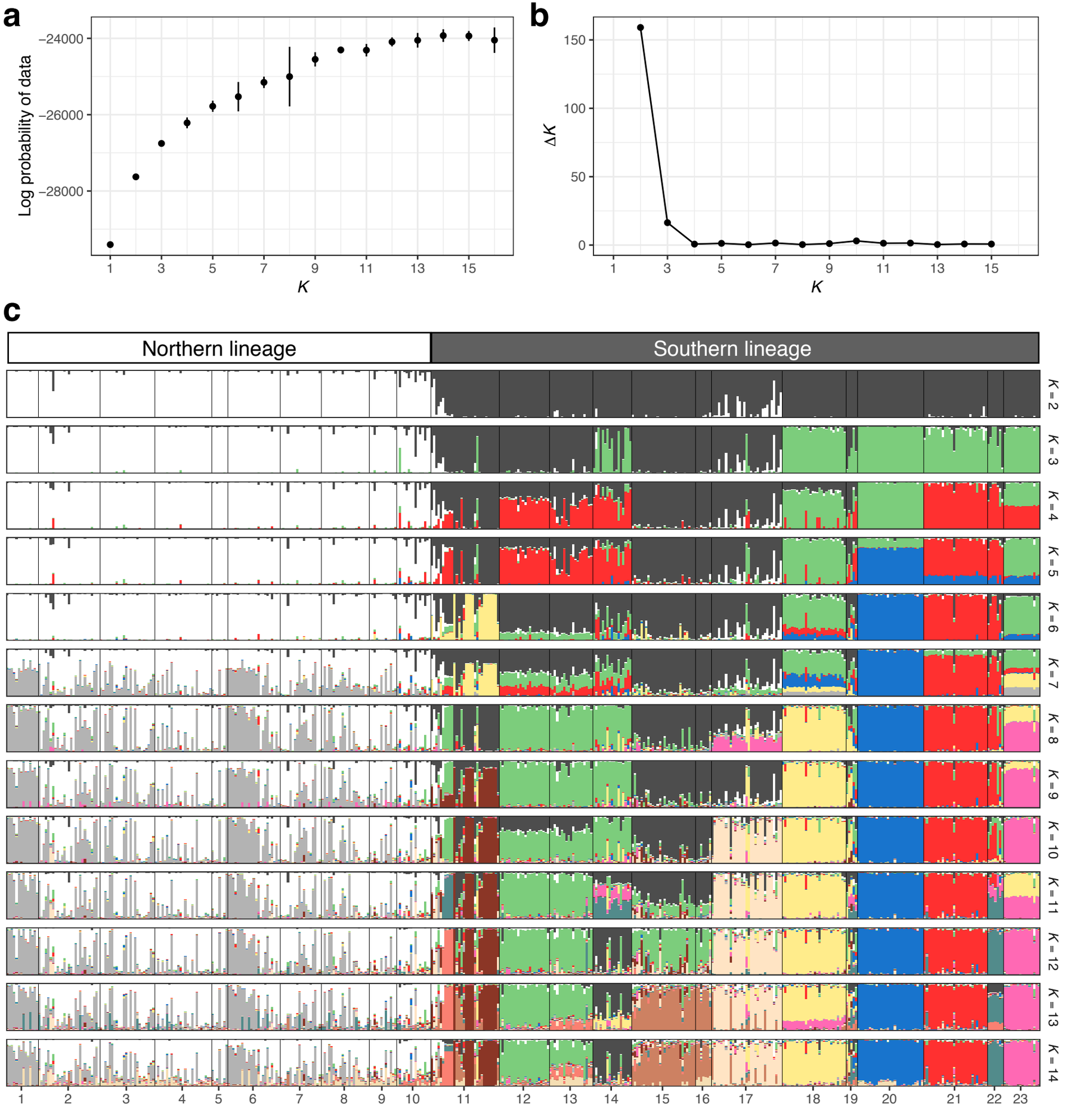
Nm_{SN}

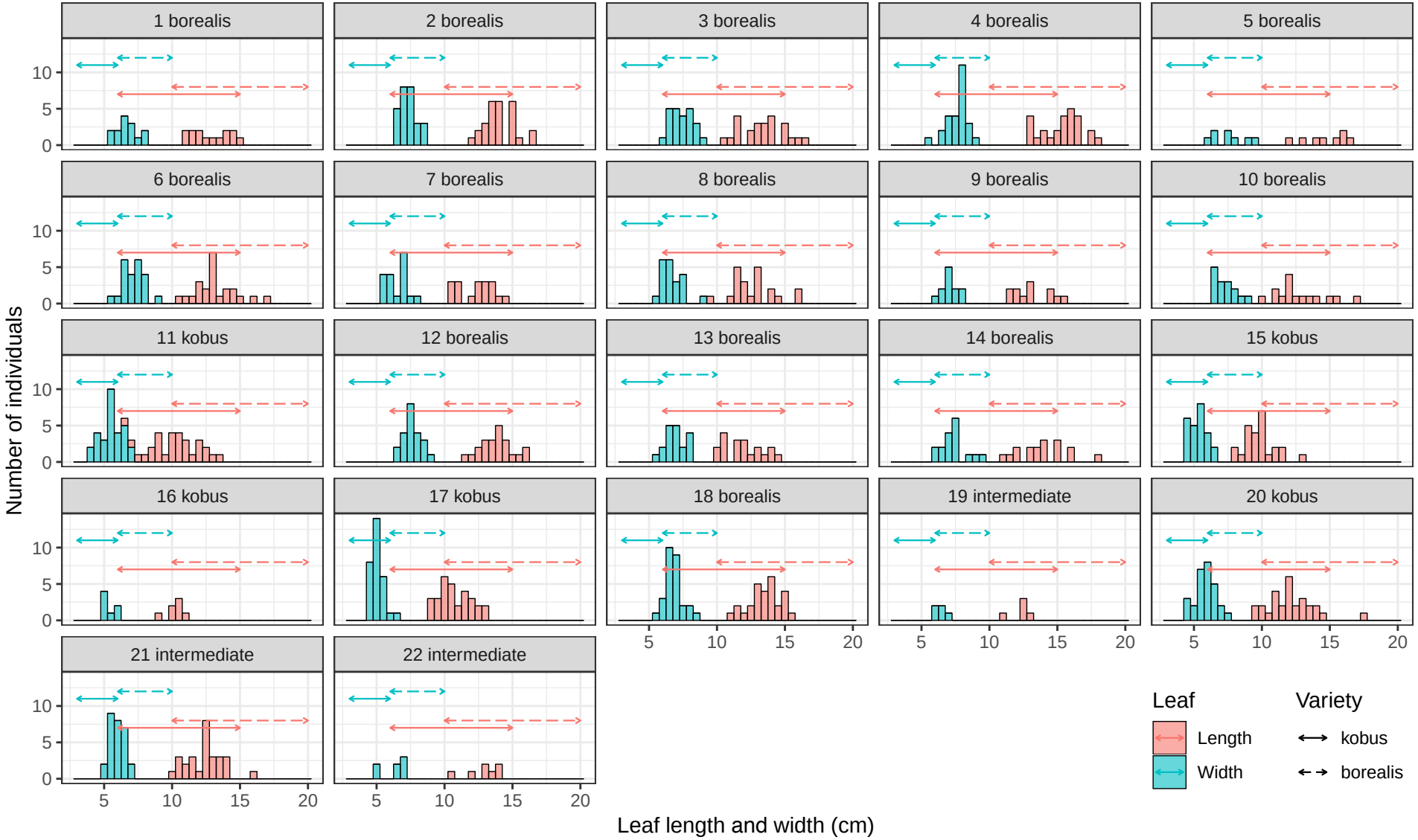
T_{DIV}

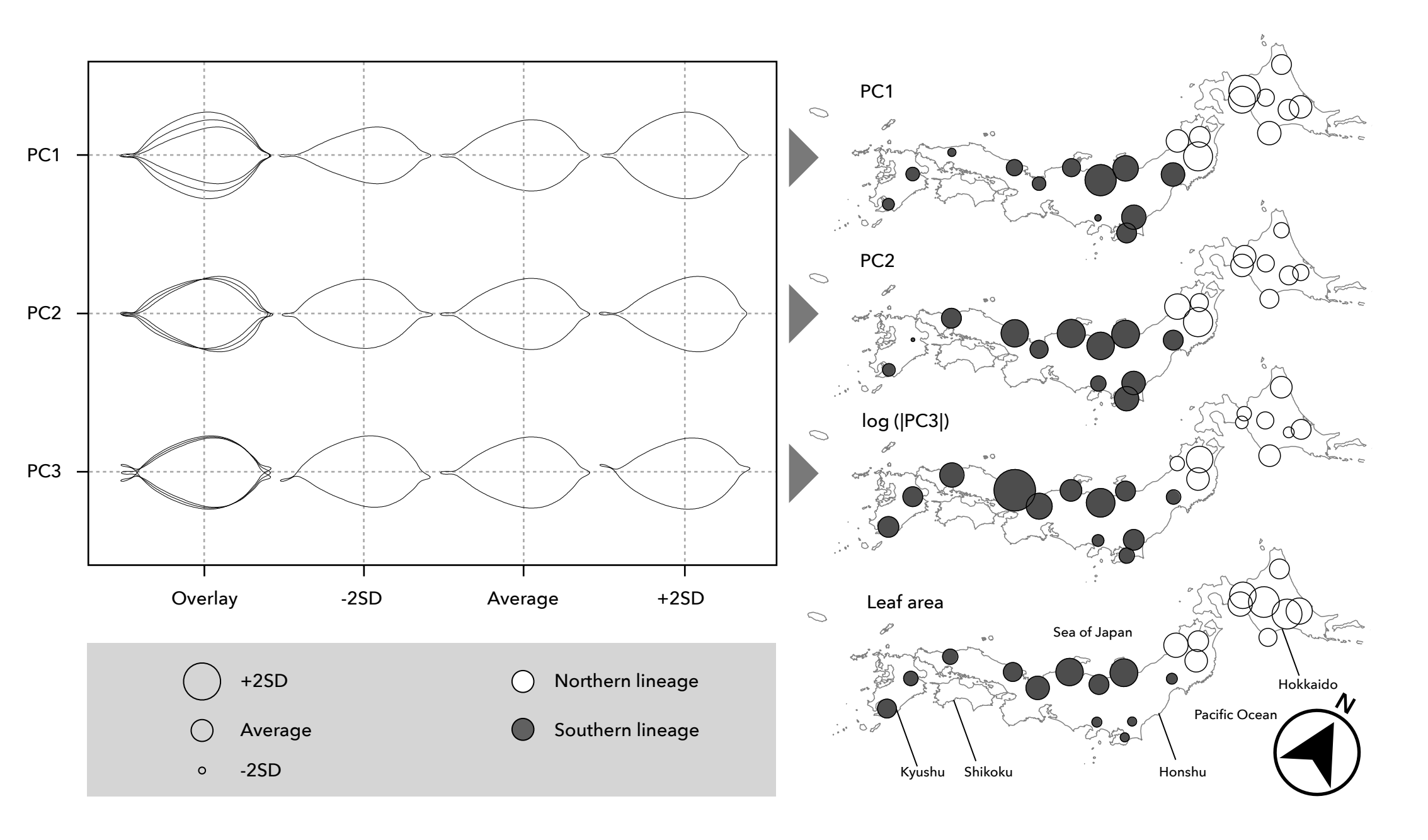
G

N_N

N_S


a
Log probability of data
−24000
−26000
−28000
1
3
5
7
9
11
13
15
K
b
ΔK
150
100
50
0
1
3
5
7
9
11
13
15
K
c
Northern lineage
Southern lineage
K = 2
K = 3
K = 4
K = 5
K = 6
K = 7
K = 8
K = 9
K = 10
K = 11
K = 12
K = 13
K = 14
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23


1 borealis
2 borealis
3 borealis
4 borealis
5 borealis
6 borealis
7 borealis
8 borealis
9 borealis
10 borealis
11 kobus
12 borealis
13 borealis
14 borealis
15 kobus
16 kobus
17 kobus
18 borealis
19 intermediate
20 kobus
21 intermediate
22 intermediate
Number of individuals
Leaf length and width (cm)
Leaf
Length
Width
Variety
kobus
borealis


PC1
PC2
PC3
Overlay
-2SD
Average
+2SD
PC1
PC2
log (|PC3|)
Leaf area
Sea of Japan
Pacific Ocean
Hokkaido
Honshu
Shikoku
Kyushu
N
+2SD
Average
-2SD
Northern lineage
Southern lineage

Present
LIG
LGM (MIROC)
LGM (CCSM)
Sea of Japan
Pacific Ocean
Hokkaido
Honshu
Shikoku
Kyushu
Jeju
45°N
40°N
35°N
130°E
135°E
140°E
145°E
0 ≤ *P* ≤ 0.3
0.3 < *P* ≤ 0.6
0.6 < *P* ≤ 1.0
Sea
Northern populations (10)
Southern populations (13)
Specimen

Electric supplementary materials

Title:

Population genetic structure and demography of *Magnolia kobus*: variety *borealis* is not supported genetically

Authors:

Ichiro Tamaki, Naomichi Kawashima, Kyohei Yukitoshi, Jung-Hyun Lee, Suzuki Setsuko, Akemi Itaya, and Nobuhiro Tomaru

Journal:

Journal of Plant Research

Corresponding author:

Nobuhiro Tomaru

Graduate School of Bioagricultural Sciences, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan

Tel: +81-52-789-4048

Fax: +81-52-789-5014

E-mail: tomaru@agr.nagoya-u.ac.jp

Contents:

Table S1 Prior distributions of parameters for population size change and population divergence analyses

Analysis	Parameter	Distribution
Population size change	N_{CUR}	Uniform (10^3, 1.5×10^5)
	G	Uniform (-0.001, 0)
	T	Uniform (1, 5×10^4)
	RN_{ANC} [a]	Uniform (1, 20)
Population divergence	N_N	Uniform (10^3, 1.5×10^5)
	N_S	Uniform (10^3, 1.5×10^5)
	G	Fixed to -2.04×10^{-4}
	T_{DIV}	Uniform (1, 5×10^4)
	Nm_{NS}	Uniform (1, 20)
	Nm_{SN}	Uniform (1, 20)
	β	Uniform (0, 1)
Common	mean μ	Log-uniform (10^{-5}, 10^{-3})
	shape	Uniform (0.5, 5)
	mean P_{GSM}	Uniform (0, 1)

[a] $N_{ANC} = N_{CUR} \times RN_{ANC}$

Table S2 Nucleotide sequence variation among three haplotypes (H to J) from *Magnolia kobus* and eight outgroup haplotypes consisting of seven haplotypes from *M. salicifolia* (A to G) and one haplotype from *M. denudata* in four chloroplast DNA regions

			*trn*S–*trn*G (684 bp)					*trn*T–*psb*D (1461 bp)										
Species	Haplotype	*N*	114	358	452	546	552	730	941	1026	1194	1232	1805	1825	1999	2041	2051	2089
M. salicifolia	A		-	T	G	G	T	A	G	G	T	C	T	G	G	G	A	A
M. salicifolia	B		-	•	•	•	•	•	•	•	•	•	•	A	•	•	•	•
M. salicifolia	C		-	•	•	•	•	•	•	•	•	•	•	A	•	•	•	•
M. salicifolia	D		-	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
M. salicifolia	E		-	•	•	•	•	•	•	•	C	•	•	•	•	•	•	•
M. salicifolia	F		-	•	•	•	•	•	•	T	•	•	•	•	•	•	•	G
M. salicifolia	G		I_1	•	•	•	•	•	•	T	•	•	•	•	•	A	•	G
M. kobus	H	35	-	•	•	T	G	G	•	T	•	•	•	•	•	•	•	G
M. kobus	I	11	-	•	•	T	G	•	•	T	•	•	•	•	•	•	•	G
M. kobus	J	2	-	C	•	T	G	•	•	T	•	T	•	•	•	•	•	G
M. denudata			-	•	A	•	•	•	A	T	•	•	C	•	A	•	G	G

Table S2 continued

			*trn*T–*trn*L (681 bp)														*rpl*36–*inf*A–*rps*8–*rpl*14 (1106 bp)			
Species	Haplotype	*N*	2161	2172	2173	2219	2263	2265	2296	2300	2357	2408	2636	2653	2736	2743	3249	3276	3449	3854
M. salicifolia	A		A	A	A	A	A	I_2	G	G	G	G	T	G	A	A	A	G	G	•
M. salicifolia	B		•	•	•	•	•	I_2	•	•	•	•	•	•	•	•	•	•	•	•
M. salicifolia	C		•	•	C	•	•	I_2	•	•	•	•	•	•	•	•	•	•	•	•
M. salicifolia	D		•	C	•	•	•	I_2	•	•	•	•	•	•	•	•	•	•	•	•
M. salicifolia	E		•	•	•	•	•	I_2	•	•	•	•	•	•	•	•	•	•	•	•
M. salicifolia	F		•	•	•	•	•	I_2	•	•	•	•	•	•	•	•	•	A	•	•
M. salicifolia	G		C	•	•	•	•	I_2	•	•	T	•	•	•	•	G	•	•	•	•
M. kobus	H	35	•	•	•	•	G	I_2	•	•	T	•	•	•	G	•	•	•	A	•
M. kobus	I	11	•	•	•	G	G	I_2	•	•	T	•	•	•	G	•	•	•	A	•
M. kobus	J	2	•	•	•	G	G	I_2	•	•	T	•	•	•	G	•	•	•	A	•
M. denudata			•	•	•	•	•	-	T	T	T	T	C	T	•	•	G	•	•	I_3

N, number of individuals; •, the same base as in haplotype A; -, deletion; I_1, insertion of TTATCTTTCTTTTCTTTATTCTAT; I_2, insertion of CTATAA; I_3, insertion of GAGAA. Sequence data for *M. salicifolia* and *M. denudata* are the same as those used in Tamaki et al. (2018). Gray columns indicate sites variable within *M. kobus*.

Table S3 Principal components (PCs) of leaf shape estimated by SHAPE

Principal component	Eigenvalue ($\times 10^{-3}$)	Contribution (%)	Cumulative contribution (%)
PC1	3.93	56.16	56.16
PC2	1.11	15.86	72.02
PC3	0.83	11.79	83.81
Overall	7.00		

Only the three PCs whose contribution to the overall variance was more than 5% are shown.

Table S4 Proportion of models correctly predicted and classification error rate estimated by random forest (RF)

Analysis	Lineage	Simulated model	Predicted model							Classification error rate
			SNM	PGM	SRM	ISM	IMM	IMM_{NS}	IMM_{SN}	
Population size change	Northern	SNM	**0.790**	0.108	0.103	–	–	–	–	0.210
		PGM	0.104	**0.891**	0.004	–	–	–	–	0.109
		SRM	0.234	0.026	**0.740**	–	–	–	–	0.260
	Southern	SNM	**0.801**	0.104	0.096	–	–	–	–	0.199
		PGM	0.111	**0.883**	0.006	–	–	–	–	0.110
		SRM	0.232	0.028	**0.740**	–	–	–	–	0.260
Population divergence		ISM	–	–	–	**0.790**	0.037	0.144	0.029	0.210
		IMM	–	–	–	0.005	**0.664**	0.181	0.149	0.336
		IMM_{NS}	–	–	–	0.068	0.227	**0.666**	0.039	0.334
		IMM_{SN}	–	–	–	0.048	0.262	0.104	**0.587**	0.413

Predictions by RF were composed of 1,000 trees based on a trained set of 10,000 simulated predictor variables (summary statistics). The response variable of RF was the demographic model. Proportions of correctly predicted demographic models are in bold face.

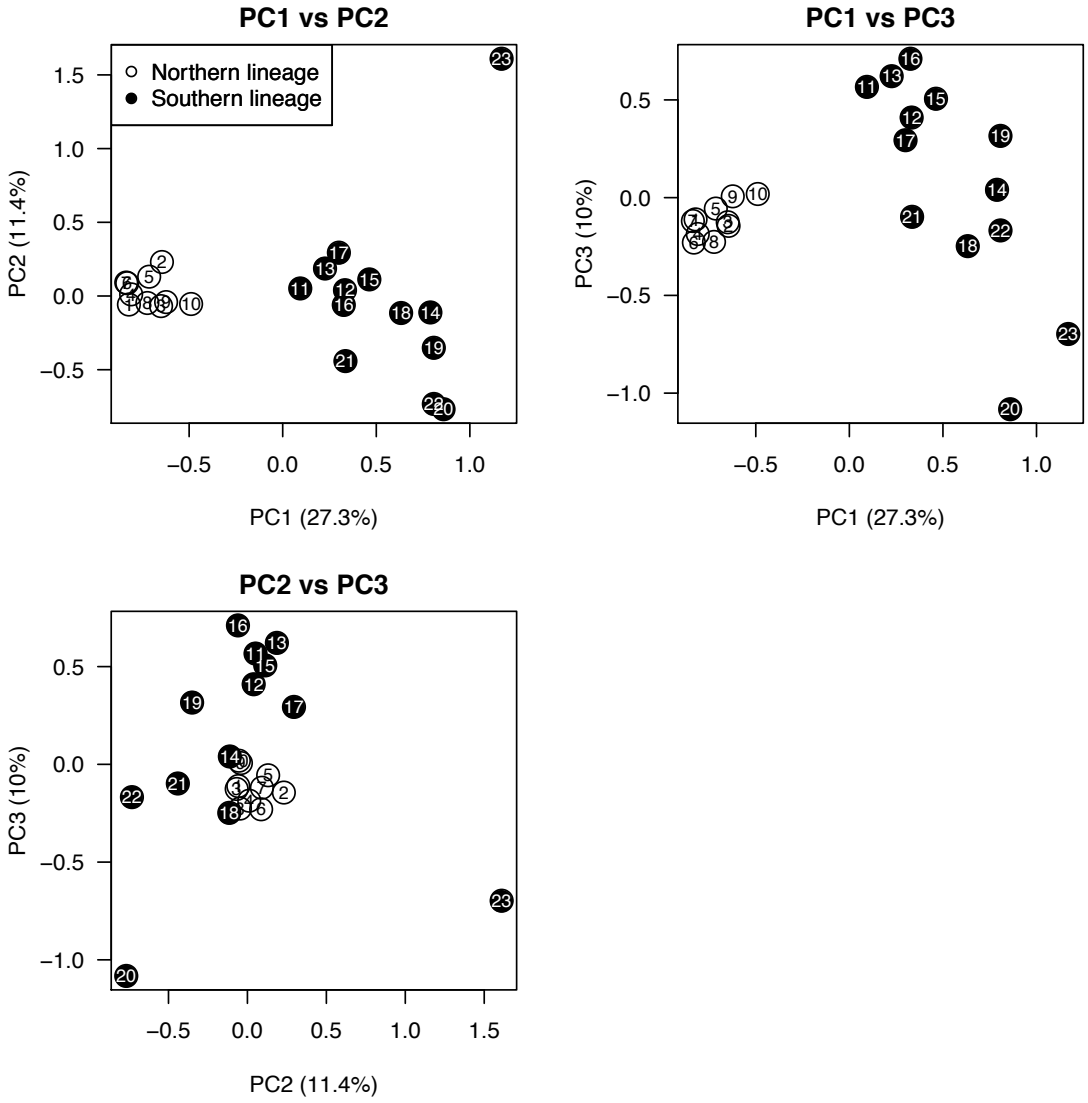


Fig. S1 Distributions of principal components estimated by allele frequencies for 23 *Magnolia kobus* populations. Numbers indicate populations listed in Table 1.

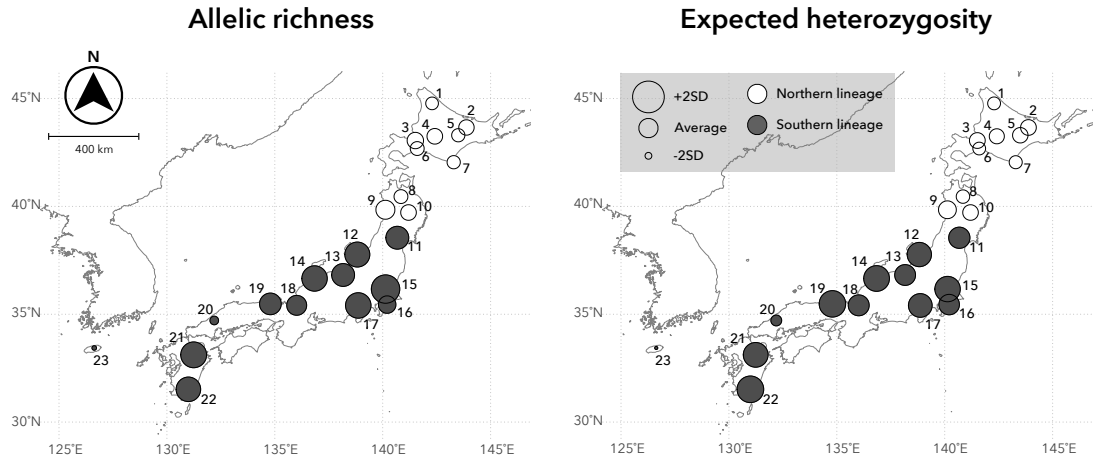


Fig. S2 Distributions of allelic richness and expected heterozygosity of 23 *Magnolia kobus* populations. Numbers indicate populations listed in Table 1.

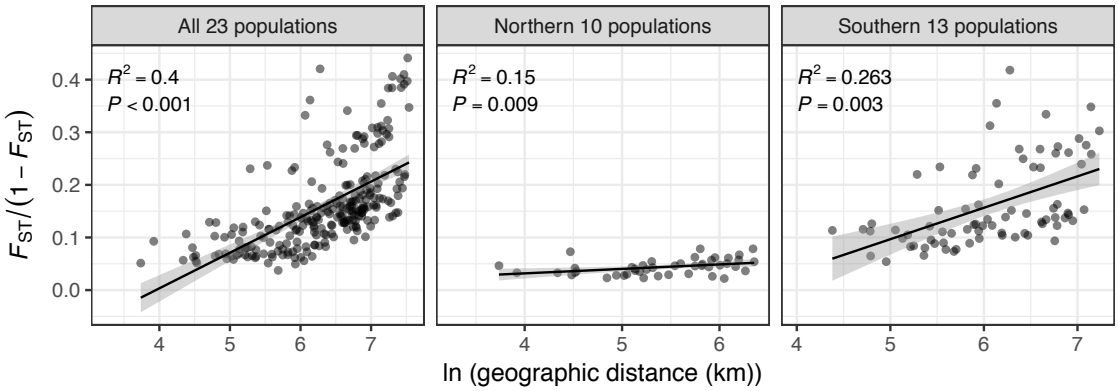


Fig. S3 Relationships between geographical and genetic distances for 23 *Magnolia kobus* populations.

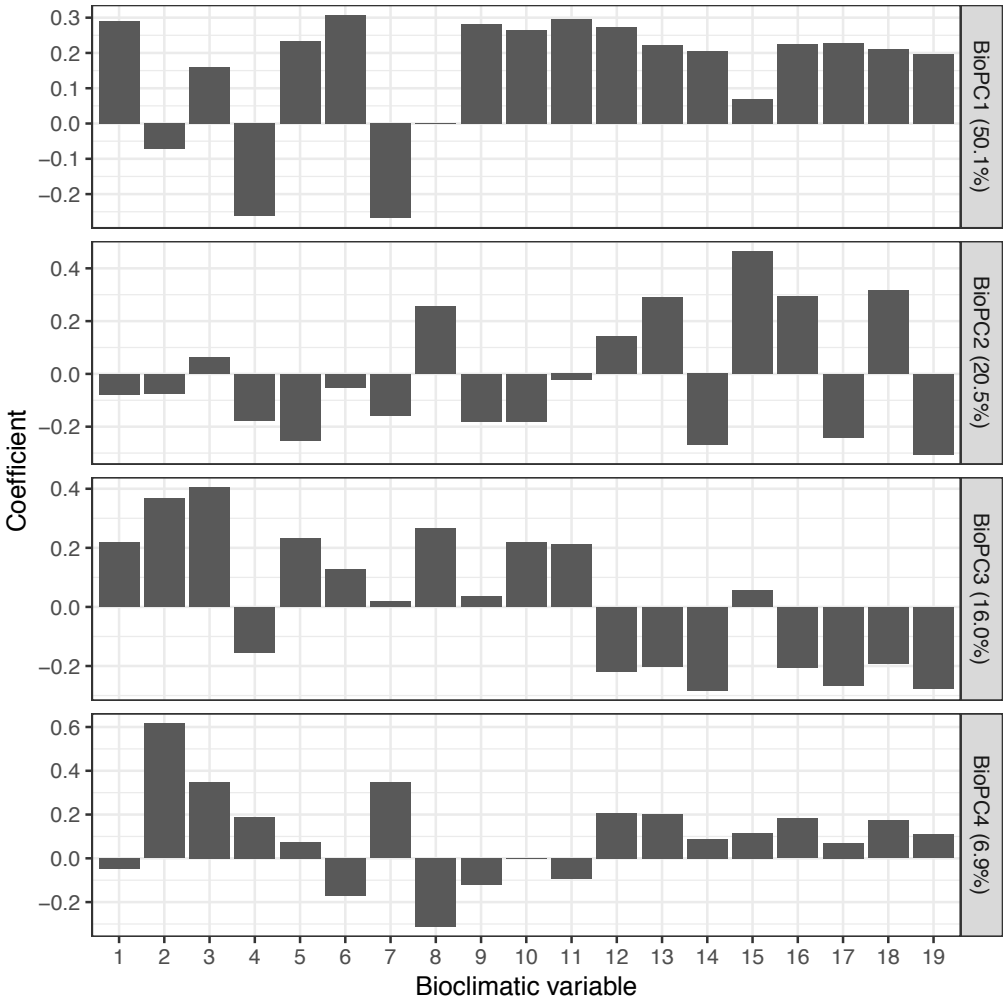


Fig. S4 Coefficients of bioclimatic variables for each principle component (BioPC). BioPCs whose contribution was more than 5% are shown. bio1, annual mean temperature; bio2, mean diurnal range [mean of monthly (max temp - min temp)]; bio3, isothermality (bio2 / bio7 × 100); bio4, temperature seasonality (standard deviation × 100); bio5, max temperature of warmest month; bio6, min temperature of coldest month; bio7, temperature annual range (bio5 - bio6); bio8, mean temperature of wettest quarter; bio9, mean temperature of driest quarter; bio10, mean temperature of warmest quarter; bio11, mean temperature of coldest quarter; bio12, annual precipitation; bio13, precipitation of wettest month; bio14, precipitation of driest month; bio15, precipitation Seasonality (coefficient of variation); bio16, precipitation of wettest quarter; bio17, precipitation of driest quarter; bio18, precipitation of warmest quarter; bio19, precipitation of coldest quarter.

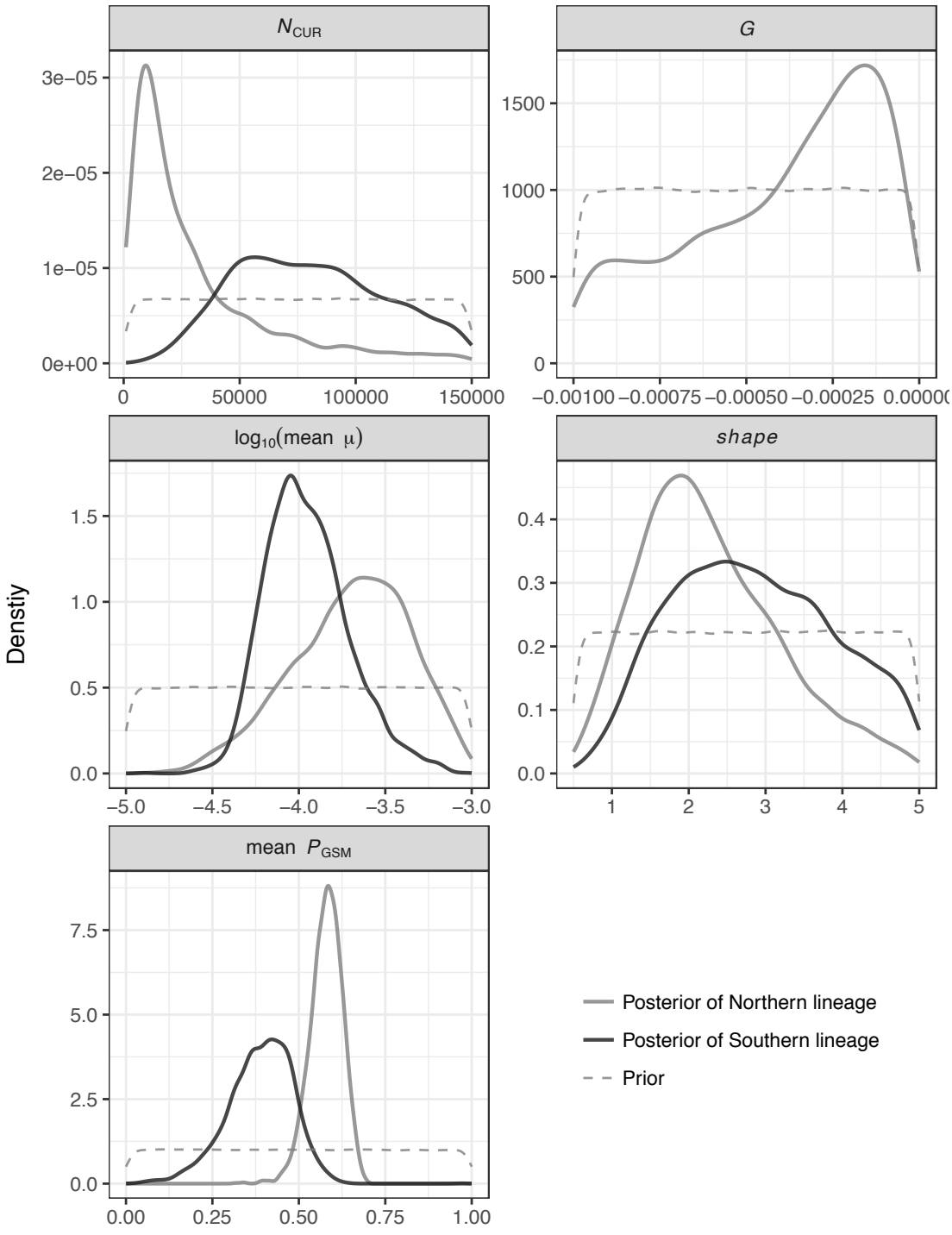


Fig. S5 Posterior and prior distributions of parameters in population size change analysis. N_{CUR}, current effective population size; G, population growth rate; mean μ, *shape* and mean P_{GSM}, parameters in generalized stepwise mutation model for nuclear microsatellites. Unit of effective population size is the number of diploid individuals.

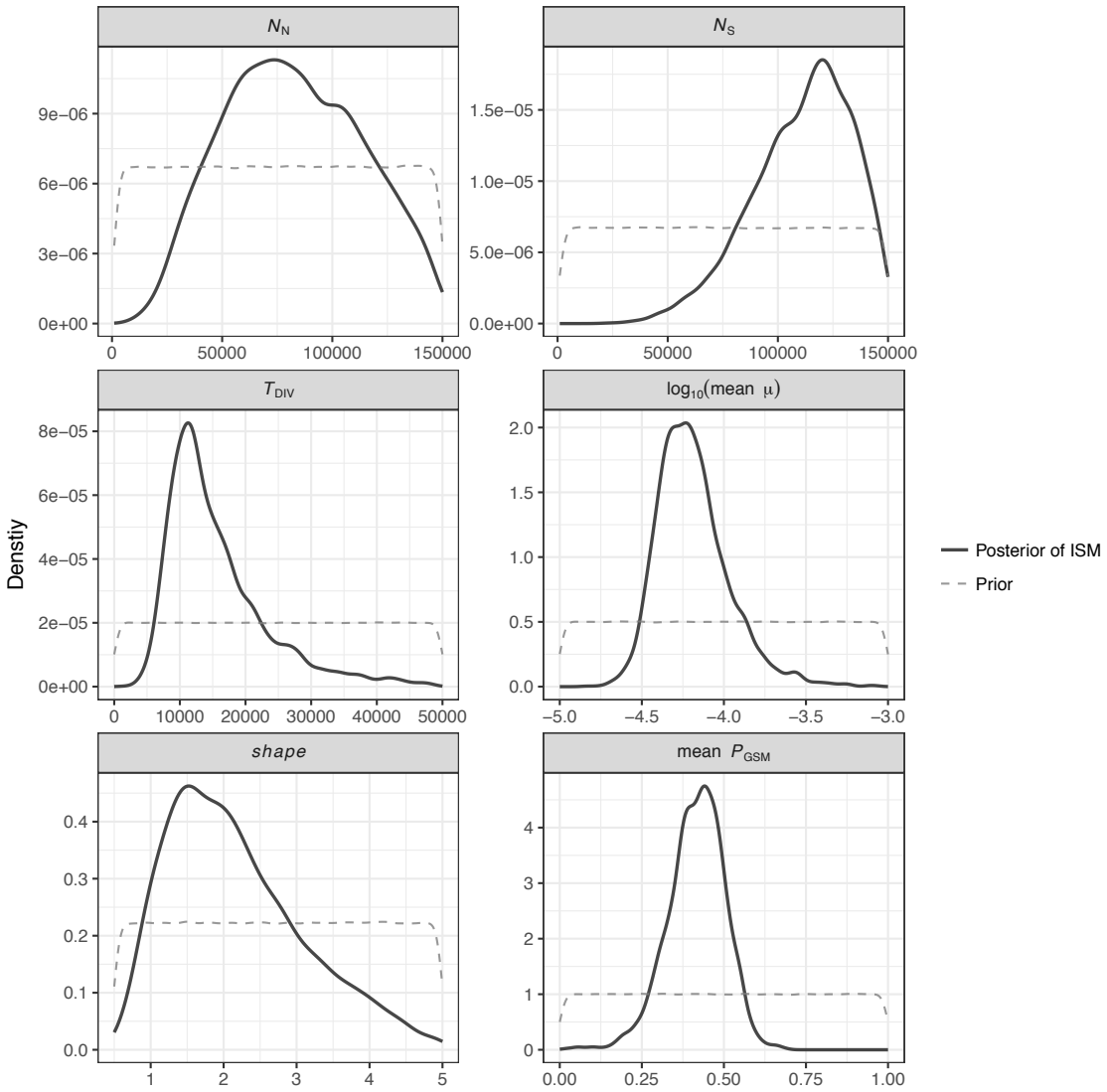
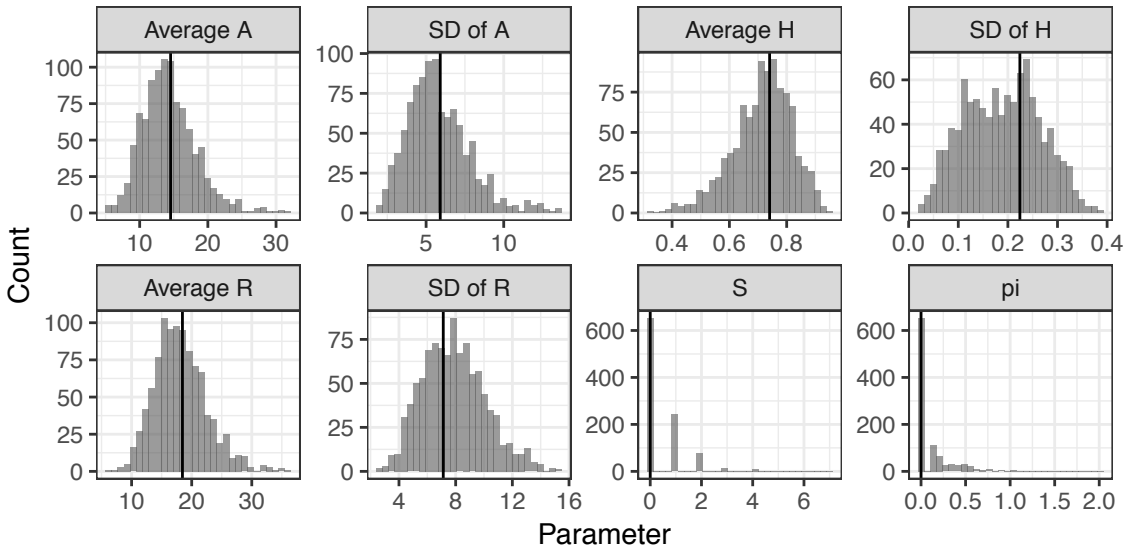


Fig. S6 Posterior and prior distributions of parameters in the best model, the isolation model. N_N and N_S, current effective population sizes of the northern and the southern lineages, respectively; T_{DIV}, divergence time; mean μ, *shape* and mean P_{GSM}, parameters in the generalized stepwise mutation model for nuclear microsatellites. Units of effective population size and divergence time are the numbers of diploid individuals and generations, respectively.

(A)



(B)

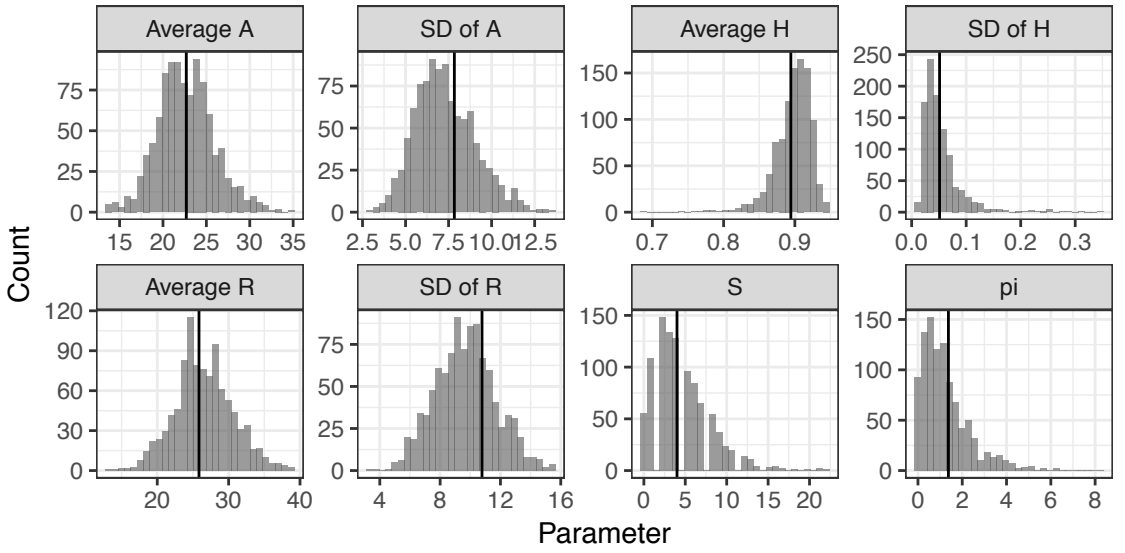


Fig. S7 Predicted and observed values (histogram and vertical bar, respectively) for the northern and the southern lineages in population size change analysis (A and B, respectively). Posterior predictive simulation was performed using the best model. *A*, number of alleles; *H*, expected heterozygosity; *R*, allele size range; *S*, number of polymorphic sites; *pi*, mean number of pairwise differences.

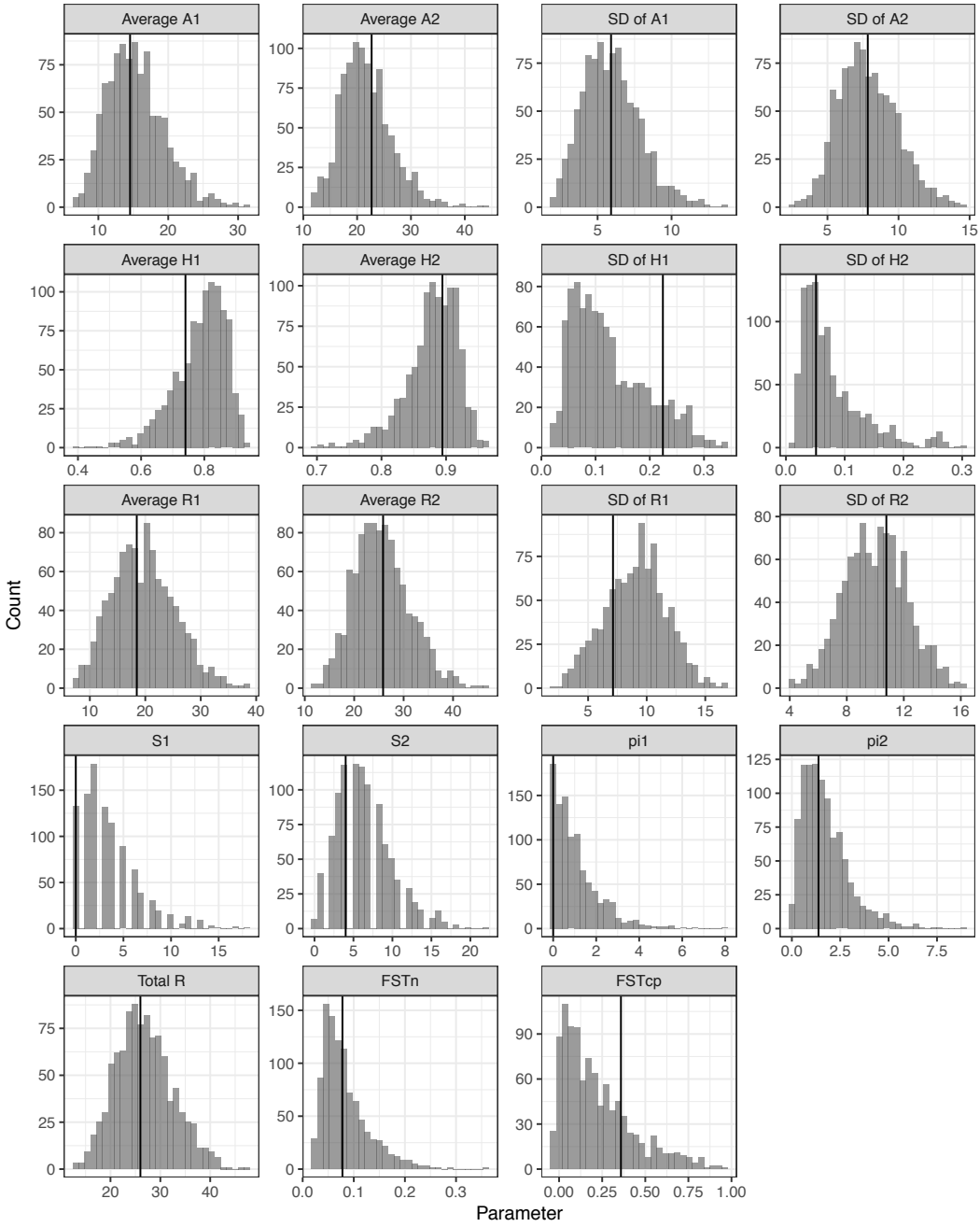


Fig. S8 Predicted and observed values (histogram and vertical bar, respectively) in population divergence analysis. Posterior predictive simulation was performed using the best model, the isolation model. *A*, number of alleles; *H*, expected heterozygosity; *R*, allele size range; *S*, number of polymorphic sites; *pi*, mean number of pairwise differences; Total R, allele size range for samples overall; *FSTn*, F_{ST} for over all loci of 13 nuclear microsatellites; *FSTcp*, F_{ST} for chloroplast DNA haplotypes. Diagrams for the northern and the southern lineages are indicated by 1 and 2, respectively.