*the* **genetics**society

**ARTICLE**

# Inferring the demographic history of Japanese cedar, *Cryptomeria japonica*, using amplicon sequencing

Natsuki Moriguchi[1] · Kentaro Uchiyama[2] · Ryutaro Miyagi[3] · Etsuko Moritsuka[4] · Aya Takahashi [3] ·
Koichiro Tamura[3] · Yoshihiko Tsumura[5] · Kosuke M. Teshima[4] · Hidenori Tachida[4] · Junko Kusumi[6]

## Abstract

The evolution of a species depends on multiple forces, such as demography and natural selection. To understand the trajectory and driving forces of evolution of a target species, it is first necessary to uncover that species' population history, such as past and present population sizes, subdivision and gene flow, by using appropriate genetic markers. *Cryptomeria japonica* is a long-lived monoecious conifer species that is distributed in Japan. There are two main lines (omote-sugi and ura-sugi), which are distinguished by apparent differences in morphological traits that may have contributed to their local adaptation. The evolution of these morphological traits seems to be related to past climatic changes in East Asia, but no precise estimate is available for the divergence time of these two lines and the subsequent population dynamics in this species. Here, we analyzed the nucleotide variations at 120 nuclear genes in 94 individuals by using amplicon sequencing in combination with high-throughput sequencing technologies. Our analysis indicated that the population on Yakushima Island, the southern distribution limit of *C. japonica* in Japan, diverged from the other populations 0.85 million years ago (MYA). The divergence time of the other populations on mainland Japan was estimated to be 0.32 MYA suggesting that the divergence of omote-sugi and ura-sugi might have occurred before the last glacial maximum. Although we found modest levels of gene flow between the present populations, the long-term isolation and environmental heterogeneity caused by climatic changes might have contributed to the differentiation of the lines and their local adaptation.

✉ Junko Kusumi
jkusumi@scs.kyushu-u.ac.jp

[1] Graduate School of Systems Life Sciences, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan

[2] Forestry and Forest Products Research Institute, Matsunosato 1, Ibaraki, Tsukuba 305-8687, Japan

[3] Department of Biological Sciences, Tokyo Metropolitan University, Hachioji, Tokyo 192-0397, Japan

[4] Department of Biology, Faculty of Science, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan

[5] Faculty of Life and Environmental Sciences, University of Tsukuba, Ibaraki, Tsukuba 305-8572, Japan

[6] Department of Environmental Changes, Faculty of Social and Cultural Studies, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan

## Introduction

Knowledge of the demographic history of populations of a species is necessary to understand that species' evolution. For example, when we attempt to make inferences about selection based on genetic data, we first need to know the demographic history of populations because patterns of genetic diversity are determined by both the demographic history of and the types of selection acting on the populations (Nielsen 2005; Heuertz et al. 2006). The demographic history of populations of a species is determined by geographical events, environmental changes, and biological interactions (Begon et al. 2006). Geographical events include uplifts of land masses and the eruption of volcanoes. One of the most well-known environmental changes is the cycle of glacial and interglacial periods covering a period of approximately 100,000 years during the Quaternary period, which caused changes in the geographical distributions of organisms (Bennett 1997). In addition, the responses of populations to environmental changes differed from species to species depending on their life-history
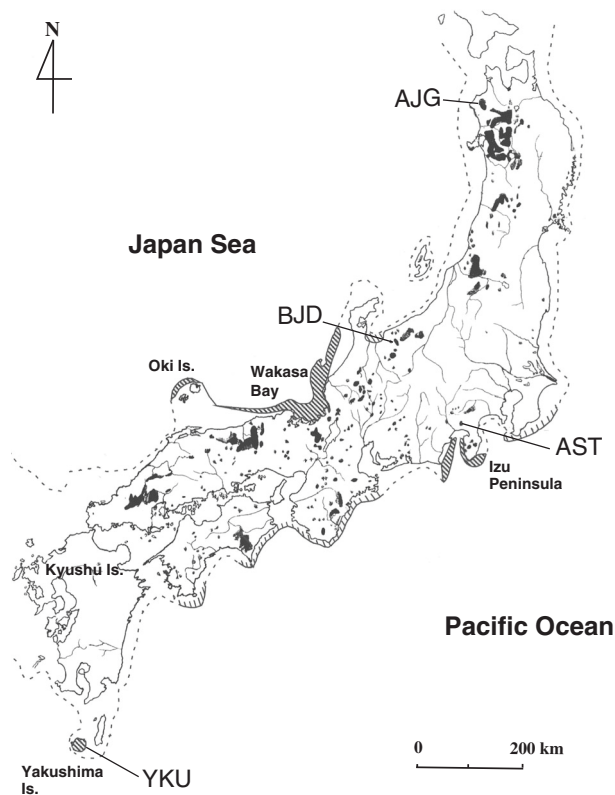
**Fig. 1** Natural distribution of *C. japonica* in Japan (filled-in black area) (Hayashi 1951) and the locations of the four populations surveyed in this study. The dotted line indicates the coastline ca. 18,000 years ago. Areas shaded in bold or within thin diagonal lines indicate established refugia (Izu Peninsula, Wakasa Bay, Oki Island, and Yakushima Island) and probable refugia, respectively, at that time (Tsukada 1986)

characteristics, such as generation time and the ability to disperse propagules. In this respect, how trees, being sedentary and having long generation time, responded to geographical and environmental changes and experienced demographic changes is an interesting problem.

*Cryptomeria japonica* (Thunb. ex L. f.) D. Don, the target species of the present study, is a monoecious conifer belonging to Cupressaceae and is an economically important species in Japan. It has a long generation time and a large genome size. *C. japonica* is an endemic species distributed widely in Japan (Yamazaki 1995, Farjon 2005) (Fig. 1). There are two main lines in Japan: omote-sugi, growing mostly on the Pacific Ocean side of the country including Yakushima Island, and ura-sugi, growing mostly on the Japan Sea side. These lines are distinguished by the morphological characteristics of the branchlets and needles (Yamazaki 1995), and adaptations to the climates of their habitats, namely, a dry climate in winter on the Pacific Ocean side and heavy snowfall during winter on the Japan Sea side, are responsible for these morphological differences. To clarify how such adaptive morphological differentiations have been formed along with the past

environmental changes, it is necessary to obtain accurate knowledge of the demographic history of this species.

Population genetic studies on *C. japonica* have been conducted using allozymes (Tsumura and Ohba 1992), DNA sequences of coding genes (Kado et al. 2003), cleaved amplified polymorphic sequence (CAPS) markers (Tsumura et al. 2007), nuclear microsatellite markers (Kimura et al. 2014; Takahashi et al. 2005), and single-nucleotide polymorphisms (SNPs) (Tsumura et al. 2012, 2014). Early genetic studies reported differentiation between omote-sugi and ura-sugi, but the resolution of the differentiation between the lines was low because only a small number of markers were examined. A recent study based on 3,930 SNP markers using samples from 14 natural populations across Japan (Tsumura et al. 2014) revealed that there were four genetic clusters of populations: two clusters in omote-sugi, comprising the population on Yakushima Island and the remaining populations on the Pacific Ocean side, and two clusters in ura-sugi, comprising the northern peripheral populations and the others on the Japan Sea side. This result was supported by a study based on 14 nuclear microsatellite markers (Kimura et al. 2014). In addition, Kimura et al. (2014) inferred the population history of these four clusters using Bayesian inference and concluded that the four clusters split simultaneously approximately 0.08 million years ago (MYA), suggesting a recent split of the two lines (omote-sugi and ura-sugi). However, the pollen data showed that climatic changes, which might have caused the trait adaptation to habitats found in omote-sugi and ura-sugi, occurred approximately 1.7 MYA and the peak of increase of *Cryptomeria* species occurred approximately 0.3 MYA (Igarashi et al. 2018), which is much earlier than the divergence time of these lines inferred based on microsatellite data (Kimura et al. 2014). This discrepancy between the scenarios inferred from the pollen data and the genetic data might be due to uncertainties of Bayesian inference using microsatellite data. For *C. japonica*, mutation rates and patterns of change in microsatellites have not been well characterized. Furthermore, Kimura et al. (2014) assumed simple demographic models without gene flow between populations despite the high dispersal ability of *C. japonica* pollen (wind pollination). Therefore, some uncertainties associated with the estimation of the divergence time remain. To test more complex scenarios and obtain a more reliable estimate of the divergence time of the genetic clusters, it would be more appropriate to use a large dataset of nucleotide variations for the inference of population parameters. Recent advances in high-throughput sequencing technologies and the development of statistical methods (see Nielsen and Beaumont 2009), accompanied by vast increases in computing power, enable us to infer the past population history of various species from genetic data with reasonable detail. Because of the considerable genome

size of *C. japonica* (~10 Gb; Hizume et al. 2001), whole-genome analysis is more difficult in this plant species than in others. A gene-targeting approach with high-throughput sequencing technologies would be suitable for surveying the genetic variation in this species. We used an amplicon-sequencing strategy (e.g., Bybee et al. 2011), because it is a highly targeted approach that enables us to deeply sequence PCR products for analyzing genetic variation. The deep sequencing would reduce missing data and errors in variant calls. In addition, we can get more accurate information on the mutation rate of nucleotide sequences compared with that of microsatellite.

In the present study, we analyzed nucleotide sequences of 120 nuclear genes in 94 individuals sampled from four representative populations of *C. japonica* by using an amplicon-sequencing method in combination with high-throughput sequencing technologies. Using the obtained data and applying a flexible composite likelihood method, we inferred the population history of the four representative populations.

## Materials and methods

### Investigated populations

We chose a representative population from each genetic cluster of populations identified by Tsumura et al. (2012, 2014): Yakushima (YKU), Ashitaka (AST), Ajigasawa (AJG), and Bijyodaira (BJD). The YKU and AST populations represent the two genetic clusters of omote-sugi: the population on Yakushima Island and the remaining populations on the Pacific Sea side, respectively (Fig. 1). The AJG and BJD populations represent the genetic clusters of ura-sugi: the populations in the northern peripheral part of the main island of Japan and the remaining populations on the Japan Sea side, respectively (Fig. 1). None of these representative populations showed deviations from Hardy–Weinberg equilibrium (HWE) and showed only small effects of admixture with the other genetic clusters (Tsumura et al. 2012, 2014). We used 24 individuals from each representative population; thus, the total number of individuals was 96. The samples used in this study were the same as those used in previous studies (Tsumura et al. 2012, 2014). The samples were collected from individuals several tens of metres or more apart from each other. We believe that this level of separation is sufficient for sampling non-related individuals, considering the seed dispersal distance and vegetative reproduction of *C. japonica* (Moriguchi et al. 2005; Takahashi et al. 2008). We used diploid DNA extracted from foliage samples. The genomic DNA samples of *C. japonica* used in the following analyses were part of those used in previous studies (e.g., Tsumura et al. 2012, 2014).

### PCR amplification and sequencing

The primers used for amplification were as follows: (1) those used to amplify 47 loci in the analysis of *Taxodium distichum*, a close relative of *C. japonica* (Ikezaki et al. 2016), and (2) 115 of those designed for amplification of expressed sequence tag (EST)-derived markers in *C. japonica* (Uchiyama et al. 2012). In total, we used 144 pairs of primers to amplify nuclear loci (Supplementary Table 1).

First, we determined the reference sequence for each of the targeted loci by using the Sanger method to map reads obtained from next-generation sequencing (NGS) of all of the samples. To measure error rates of the NGS, we planned to determine the sequences of five loci randomly chosen from the 144 target loci in eight individuals using the Sanger method: however, we could obtain only sequences of 1–7 individuals at each of the loci because of failures in obtaining clear DNA sequence traces in several cases. The error rates of NGS were calculated as per site differences between the sequences obtained by the Sanger method and those obtained by NGS.

The 144 target loci in the 96 individuals were amplified using an Access Array^TM System with a 48.48 Access Array IFC (Fluidigm, CA, USA). This equipment simultaneously performs multiplexed PCR of 48 different types of amplicons for each of 48 individual DNA samples. We ran this amplification six times to amplify 144 loci in 96 samples. After amplification, we measured the size and concentration of the PCR products for each sample using TapeStation D1000 screen tape with TapeStation D1000 reagents (Agilent Technologies, CA, USA).

We purified the PCR products using Agencourt AMPure XP PCR purification kits (Beckman Coulter, MA, USA), measured the concentrations of the purified PCR products using a Quantus Fluorometer and QuantiFluor dsDNA system (Promega, WI, USA), and adjusted the concentration in each sample to be suitable for library preparation. Then, we constructed an indexed pair-end library for each sample using a Nextera XT DNA sample preparation kit (Illumina, CA, USA). Finally, we mixed all of the indexed libraries and sequenced them using a MiSeq System (Illumina) with a MiSeq Reagent Kit v3, 600 cycles and the pair-end (300 + 300) option.

### Mapping of reads and SNP calling

We performed quality control of the raw reads obtained by the MiSeq System using FASTX Toolkit 0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit/). First, each read was trimmed based on its sequencing quality score; nucleotides with quality scores lower than 30 were trimmed from the end of the reads. After trimming, reads with a length shorter than 32 bases were discarded. Second, reads were discarded if

the percentage of sites with quality scores below 30 was >10%.

Then, we mapped the trimmed reads to the reference sequences using BWA-MEM with the paired-end mapping option and the other options set to default values (BWA ver. 0.7.12-r1039; Li et al. 2009) to make sam files, which were converted to bam files using SAMtools (Li et al. 2009). The program BCF tools (Li and Durbin 2009; Li et al. 2009) was used to make a vcf file containing information on the nucleotides at each site from the bam files of all samples. Then, sites with a depth <10 or a genotyping quality <100 were treated as missing data, and a fasta file for each locus was generated using a custom-made script. We discarded the sequences with half or more missing data sites in a locus. Then, two individuals having fewer than 90% of the loci and nine loci found in fewer than 90% of the individuals were discarded. Finally, to remove potential duplicated loci, we tested for HWE using a chi-squared test at each polymorphic site. Because 13 loci showed a significant deviation from HWE, we discarded them.

## Population genetic analyses

To determine the current population structure, we performed principal component analysis (PCA) using the R package SNPRelate (Zheng et al. 2012). Next, we performed Bayesian clustering analysis using STRUCTURE 2.3 (Hubisz et al. 2009). For the analysis using STRUCTURE, we randomly chose one SNP from each locus. The number of clusters, $K$, was set to 1–10. For each $K$, 20 independent runs, each consisting of 100,000 Markov chain Monte Carlo (MCMC) iterations after a burn-in of 10,000 iterations, were executed. We used CLUMPAK (Kopelman et al. 2015) to summarize the data for each $K$ and STRUCTURE HARVESTER (Earl and vonHoldt 2012) to calculate the mean log likelihood (Pritchard et al. 2000) and Evanno's ΔK (Evanno et al. 2005).

From the cleaned fasta files, we estimated the average number of nucleotide differences per site between pairs of sequences sampled from two populations ($\pi_b$) and sampled from the same population ($\pi_w$) (Hudson et al. 1992) and Tajima's $D$ (Tajima 1989) for each population using DnaSP v5 (Librado and Rozas 2009). Reading frames were inferred at 109 of the 120 loci, and the statistics were calculated separately for silent, replacement and all sites. We attached the subscript s, r, or a to the statistics at silent, replacement, or all sites, respectively. We used the HKA program developed by Jody Hey (available at https://bio.cst.temple.edu/~hey/software/#hka-div) for a multi-locus Tajima's $D$ test and the R package PopGenome (Pfeifer et al. 2014), to estimate the $F_{ST}$ between each pair of populations. We also performed an analysis of molecular variance (AMOVA) using Arlequin ver. 3.5.1.3 (Excoffier and Lischer 2010).

To find candidate genes under selection, we used BayeScan version 2.1 (Foll and Gaggiotti 2008). BayeScan identifies candidate loci under natural selection using differences in allele frequencies between populations, which are measured by $F_{ST}$, under a simple island model (Foll and Gaggiotti 2008). We used three sets of biallelic site frequency data: the first one consisted of the frequency data calculated for all four populations; the second one consisted of the data calculated for the YKU population and a group of the other populations (the YKU population vs. the others); and the third one consisted of the data calculated for the AST population and a group comprising the AJG and BJD populations (omote-sugi vs. ura-sugi). We excluded singleton sites from the datasets following the recommendation in the manual. We set the default values for the MCMC algorithm parameters in the BayeScan runs and the value of prior odds to 10. A site was considered an outlier when the posterior odds favouring a model with selection were above the threshold, which was computed by the software to set a false discovery rate (FDR) of < 5%.

To infer the order of population splits, we reconstructed a maximum likelihood (ML) tree of populations by TreeMix (Pickrell and Pritchard 2012) based on 77 loci for which the sequences in *Taxodium distichum* were available (Ikezaki et al. 2016). First, we reconstructed the ML tree without assuming migration between populations and using *T. distichum* as an outgroup to tentatively determine the order of splits among the four populations of *C. japonica*. Then, assuming this tentative relationship and introducing migration, we reconstructed the ML tree with migration among the four populations of *C. japonica*.

To infer the population history of *C. japonica*, we used the program *fastsimcoal2* (ver. 2.5.2.8) (Excoffier et al. 2013), which implements a model-based approach using composite likelihoods of two-dimensional joint site frequency spectra (SFS). *fastsimcoal2* can handle complex population models including migration between populations, historical events such as population size changes, population divisions, and admixture events, etc. Because *fastsimcoal2* does not allow unequal sample sizes among sites, the sample size was set to 15 individuals for each population, which is the maximum number of individuals with no missing data among the four populations. We randomly chose 15 individuals (30 alleles per locus) from the samples without missing data from each of the four populations. Thus, the total number of individuals used was 60. We considered two demographic models: Model 1 and Model 2 (see Fig. 2). Both models started with an ancestral population, ANC, of size $N_{ANC1}$. The size of this population changed to $N_{ANC2}$ at some point in the past and subsequently split into four descendant populations with the effective population sizes $N_{AJG}$, $N_{BJD}$, $N_{AST}$, and $N_{YKU}$. Symmetrical migration was allowed only between neighbouring populations (see Fig. 2)
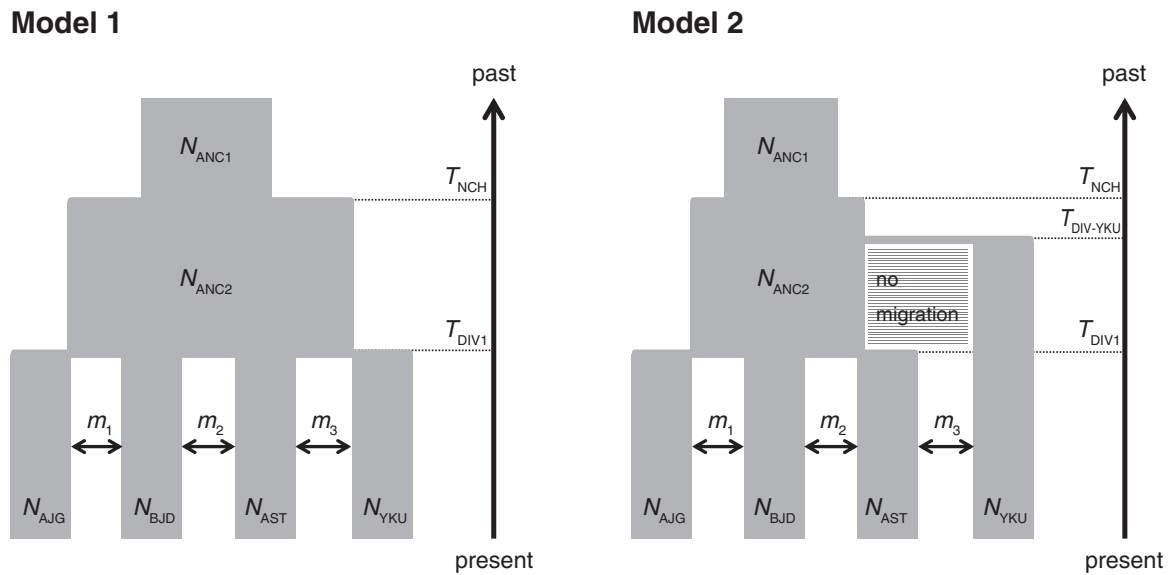
## Model 1

## Model 2



**Fig. 2** Two models used in the estimation of parameters by *fastsimcoal2*. The two models differ in the time when the YKU population split from the other populations

after the final split. The difference between the two models was the order of the population splits: in Model 1, the four descendant populations split simultaneously, corresponding to the best model suggested by Kimura et al. (2014), and in Model 2, the order of population splits was assumed based on our preliminary analysis using TreeMix (Pickrell and Pritchard 2012). Note that Model 1 was nested within Model 2, such that we could use a likelihood ratio test to choose the best model. We used all sites to calculate the SFS. ML estimates of the model parameters were obtained using sequential Markov coalescent simulations and an extension of the Expectation-Maximization algorithm, the Expectation/ Conditional Maximization (ECM) algorithm (see Excoffier et al. 2013 for details). In each maximization of the like-lihood, the minimum and maximum numbers of ECM cycles were set to 10 and 60, respectively, and this maximization process was repeated 50 times. The parameter set that gave the ML among the 50 replicates were used as estimates of the parameters. In the coalescent simulation, we regarded each locus as a linkage block. In addition, we assumed that the mutation rate per generation was $1.50 \times 10^{-9}$ (estimated from the mean divergence rate between *C. japonica* and *T. distichum* at the loci used for the analysis and the estimated divergence time between the two species, i.e., 90 MYA, by Leslie et al. (2012), and the recombination rate per generation was $1.27 \times 10^{-8}$ (Fujimoto et al. 2008). The generation time was set to 50 years. We carried out estimations of confidence intervals (CIs) for the parameters and likelihood ratio tests using parametric bootstrapping with replicated runs of 100, following the instructions in the *fastsimcoal2* manual. For parametric bootstrapping, DNA sequence data were generated by simulation using the estimates of the

parameters with the average length of the genes, and esti-mation was carried out for the generated data.

Finally, to examine the fit of the chosen model to the data, we ran simulations assuming the estimated parameters using *fastsimcoal2*; obtained the distributions of $\pi_W$, $F_{ST}$, and Tajima's $D$; and compared these distributions with the observed values.
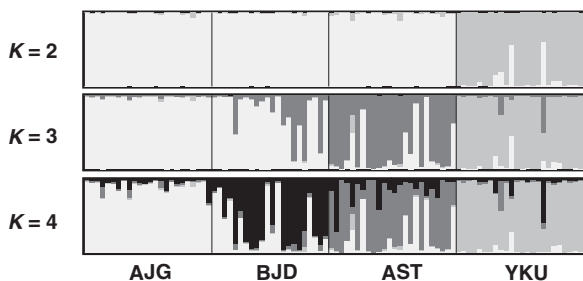
## Results

### NGS data processing

After quality filtering, we obtained nucleotide sequence data for 122 loci in 94 samples. The average error rate of the NGS data across the five loci was estimated to be 0.15%; however, we found that the error rate was excep-tionally high at one of the five loci, and the errors were mostly false positives. Thus, the NGS data for this locus showed that a site was heterozygous, whereas the Sanger method showed that the site was homozygous. If we removed this locus, then the error rate was reduced to 0.076%. We assumed that this error was caused by reads from duplicated loci, which would not have been detected by the HWE test. Therefore, we removed this locus from the subsequent analyses. In addition, we found more than two contigs in several individuals at another locus and thus removed this locus as a duplicated locus. Consequently, the data at 120 loci in 94 samples constituted the final dataset for the subsequent analyses. The total numbers of sites analyzed and polymorphic sites were 57,566 and 1,319, respectively.

## Population structure

We conducted a Bayesian clustering analysis by applying STRUCTURE to the data of 120 segregating sites, each of which was randomly chosen from a locus (Fig. 3 and Supplementary Fig. 1). The optimal number of clusters was $K = 2$ based on Evanno's $\Delta K$ and $K = 4$ based on the estimated log likelihood (Supplementary Fig. 1). The YKU population and the other three populations were clearly separated when $K = 2$ (Fig. 3). When $K = 3$, the AST population, located on the Pacific Ocean side (omote-sugi), was separated from the two populations that were located on the Japan Sea side (the AJG and BJD populations, ura-sugi) (Fig. 3). Four genetic clusters were distinct when $K = 4$, which was consistent with the results of previous studies (Tsumura et al. 2014; Kimura et al. 2014) (Fig. 3).

The PCA result agreed with that of STRUCTURE with $K = 3$: individuals from the YKU population were separated from those of the other populations by the PC1 axis, and those of the AST population and ura-sugi (the AJG and BJD populations) were separated by the PC2 axis (Supplementary Fig. 2).

Pairwise $F_{ST}$ values between populations are shown in Table 1. There was significant differentiation between every pair of the four populations ($P < 0.01$). Estimates of $F_{ST}$ between the YKU population and the other three populations were higher (0.0871–0.1081) than the estimates of $F_{ST}$ between pairs of the three populations. Furthermore, the estimate of $F_{ST}$ between the AJG and BJD populations was lower (0.0398) than that between the AST and AJD or BJD populations (0.0561 or 0.0554, respectively). The average

of $\pi_b$ across the 120 loci is also shown in Table 1. Because the values of $\pi_w$ were similar among populations, as described later, the levels of $\pi_b$ showed trends similar to those of $F_{ST}$. Based on the result of STRUCTURE with $K = 2$ and pairwise $F_{ST}$ values, we performed an AMOVA that assumed the divergence between the YKU population and the other three populations was the highest hierarchical level (Table 2). The proportion of variation between the YKU population and the other three populations was 5.38%, the proportion of variation among the three populations was 4.62%, and most of the variation was observed within populations (90.0%).

Finally, we inferred historical relationships between populations using TreeMix, assuming two models: one with an outgroup (*T. distichum*) and no migration and the other without an outgroup and with migration. The inferred order of splits was the same in both models. First, the YKU population split from the other populations; then, the AST population split from the AJG and BJD populations; and finally, the AJG and BJD populations split from each other (Supplementary Fig. 3). Surprisingly, migration from the AJG population to the AST population was detected, although no migration was detected between the remaining pairs of populations under the model with migration.

## Nucleotide diversity, neutrality test, and detection of $F_{ST}$ outliers

The average nucleotide diversity at all sites ($\pi_{wa}$) across the 120 loci ranged from $0.00309 \pm 0.00027$ (YKU population) to $0.00279 \pm 0.00030$ (BJD population) (Table 3). There were no significant differences in the average $\pi_{wa}$ between populations. The average nucleotide diversity at silent sites ($\pi_{ws}$) across the 109 loci ranged from $0.00469 \pm 0.00048$ (YKU population) to $0.00396 \pm 0.00041$ (AJG population). The differences between populations were not significant.

None of the Tajima's $D$ values of the 120 loci were significant with an FDR of 5% (Table 3). The mean value of Tajima's $D_a$ across the 120 loci was not significant in the YKU, AJD, nor BJD population but was significantly negative in the AST population ($P = 0.0048$). Similarly, significantly negative values for the averages of Tajima's $D_s$ and $D_r$ across the 109 loci were obtained in the AST



**Fig. 3** Distribution of genetic cluster memberships of 94 samples with $K = 2$, 3, and 4

**Table 1** Nucleotide diversity between populations ($\pi_b$) and $F_{ST}$

|  | AJG | BJD | AST | YKU |
|---|---|---|---|---|
| AJG | – | $0.00293 \pm 0.000296$ | $0.00310 \pm 0.000311$ | $0.00332 \pm 0.000294$ |
| BJD | $0.0398 \pm 0.00483$ | – | $0.00307 \pm 0.000310$ | $0.00326 \pm 0.000285$ |
| AST | $0.0561 \pm 0.00584$ | $0.0554 \pm 0.00649$ | – | $0.00334 \pm 0.000306$ |
| YKU | $0.1081 \pm 0.00991$ | $0.0964 \pm 0.00972$ | $0.0871 \pm 0.00928$ | – |

$\pi_b$ and $F_{ST}$ are shown above and below the diagonal, respectively. Standard errors are also shown

*AJG* Ajigasawa, *BJD* Bijyodaira, *AST* Ashitaka, *YKU* Yakushima

**Table 2** Results of analysis of molecular variance (AMOVA)

| Source of variation | d.f. | Sum of squares | Variance components | Percentage of variation |
|---|---|---|---|---|
| Among groups[a] | 1 | 737.98 | 5.71 | 5.38 |
| Among populations within groups | 2 | 648.93 | 4.91 | 4.62 |
| Within populations | 184 | 17,592.22 | 95.61 | 90.00 |
| Total | 187 | 18,979.13 | 106.23 | |

*d.f.* degrees of freedom

[a]The YKU population vs. the other three populations

**Table 3** Average and standard error of nucleotide diversity within populations ($\pi_w$) and Tajima's *D* in each population

| | AJG | BJD | AST | YKU |
|---|---|---|---|---|
| $\pi_{wa}$ (120 genes) | $0.00284 \pm 0.00029$ | $0.00279 \pm 0.00030$ | $0.00301 \pm 0.00032$ | $0.00309 \pm 0.00027$ |
| $\pi_{wa}$ (109 genes) | $0.00283 \pm 0.00031$ | $0.00283 \pm 0.00032$ | $0.00292 \pm 0.00033$ | $0.00308 \pm 0.00030$ |
| $\pi_{ws}$ (109 genes) | $0.00396 \pm 0.00041$ | $0.00401 \pm 0.00040$ | $0.00411 \pm 0.00044$ | $0.00469 \pm 0.00048$ |
| Tajima's $D_a$ (120 genes) | $0.113 \pm 0.102$ | $-0.153 \pm 0.100$ | $-0.299 \pm 0.091$* | $-0.041 \pm 0.101$ |
| Tajima's $D_a$ (109 genes) | $0.146 \pm 0.108$ | $-0.194 \pm 0.107$ | $-0.366 \pm 0.097$ | $-0.007 \pm 0.106$ |
| Tajima's $D_s$ (109 genes) | $0.213 \pm 0.111$ | $-0.088 \pm 0.115$ | $-0.291 \pm 0.094$* | $0.098 \pm 0.102$ |
| Tajima's $D_r$ (109 genes) | $0.017 \pm 0.098$ | $-0.194 \pm 0.108$ | $-0.315 \pm 0.100$* | $-0.250 \pm 0.099$ |

Because reading frames were determined for 109 genes, statistics were calculated separately at all sites ($\pi_{wa}$) and silent sites ($\pi_{ws}$) for those genes

*AJG* Ajigasawa, *BJD* Bijyodaira, *AST* Ashitaka, *YKU* Yakushima

*Significant at the 5% level

population ($P = 0.0050$ and 0.0090, respectively), but the values were not significant in the other populations.

To detect candidate loci under selection, we applied BayeScan to three datasets, namely, the YKU population vs. the other populations, omote-sugi vs. ura-sugi, and four populations of 923 SNPs excluding singleton sites. No significant sites were found with an FDR of 5%.

## Inferring the population history of *C. japonica*

Because the neutrality test and BayeScan detected no candidate loci under selection, all 120 loci were used for the inference of population history. Consequently, we estimated parameters of the population history by *fastsimcoal2* using the data of 57,566 sites in the 120 loci from 60 samples. Two demographic models, namely, Model 1 and Model 2, differing in the time when the YKU population split from the other populations, were used (see Fig. 2). Model 1, assuming simultaneous splits of four populations, corresponded to the best model suggested by Kimura et al. (2014). Model 2, assuming that the YKU population split earlier than the other three populations split, corresponded to our preliminary result obtained using TreeMix described above.

The likelihood ratio test showed that Model 2 fit significantly better than did Model 1 ($P = 0.01$). The point estimates of the parameters under these two models and CIs of the estimates under Model 2 obtained by parametric bootstrapping with 100 replicates are shown in Table 4.

In Model 2, the point estimate of the split of the YKU population from the other populations was 0.85 MYA, and that of the split of the remaining three populations was 0.32 MYA. Migration rates between populations were significantly >0 in all pairs of neighbouring populations; therefore, we can reject the hypothesis that there was no migration between the neighbouring populations. The effective population size including ancestral populations ranged from 35,000 to 61,000. Note that the CIs for the estimates were fairly large; for example, the CIs of population sizes for all pairs overlapped with each other, except for those of $N_{ANC1}$ and $N_{AJG}$.

Using the ML estimates of the parameters under Model 2, we conducted simulations and obtained the distributions of $\pi_w$, Tajima's *D*, and $F_{ST}$. The observed values of $\pi_w$ and Tajima's *D* in each population did not deviate from the predicted distributions (Fig. 4); however, the observed values of $F_{ST}$ between the AST and BJD populations and between the AST and YKU populations deviated from the predicted distributions ($P < 0.05$ and 0.01, respectively).

## Discussion

### Population structure and genetic diversity

Our analyses based on 120 nuclear genes showed that the four populations, representing the four clusters of natural populations of *C. japonica*, were separated into four

**Table 4** Estimates based on composite likelihood and their 95% confidence intervals (CIs) for demographic parameters obtained using *fastsimcoal2*

| Parameter[a] | Point estimation | | 95% CI (under Model 2) | |
|---|---|---|---|---|
| | Model 1 | Model 2 | 2.5% quantile | 97.5% quantile |
| $N_{ANC2}$ | 55,273 | 34,061 | 18,241 | 58,252 |
| $N_{ANC1}$ | 82,291 | 61,017 | 40,115 | 182,297 |
| $N_{AJG}$ | 26,708 | 35,427 | 22,279 | 39,243 |
| $N_{BJD}$ | 35,091 | 53,250 | 26,769 | 72,626 |
| $N_{AST}$ | 56,581 | 48,061 | 30,640 | 62,549 |
| $N_{YKU}$ | 36,356 | 36,695 | 29,692 | 46,155 |
| $m_1$ | 9.47E–05 | 4.85E–05 | 6.21E–06 | 1.41E–04 |
| $m_2$ | 7.56E–05 | 5.97E–05 | 1.53E–05 | 1.47E–04 |
| $m_3$ | 3.03E–05 | 6.26E–05 | 4.33E–05 | 9.74E–05 |
| $T_{DIV1}$ (year) | 609,150 | 317,300 | 199,079 | 719,053 |
| $T_{DIV-YKU}$ (year) | - | 854,600 | 666,306 | 1,775,345 |
| $T_{NCH}$ (year) | 4,164,000 | 479,350 | 394,488 | 8,696,396 |
| MaxEstLhood[b] | −25,267.637 | −25,256.526 | | |
| MaxObsLhood[c] | −24,148.982 | −24,148.982 | | |

*SFS* site frequency spectra

[a]See Fig. 2

[b]The estimated log likelihood

[c]The maximum log likelihood given the observed SFS

(STRUCTURE) or three genetic clusters (PCA) in our analyses. These results were compatible with the results reported in the study by Tsumura et al. (2014) based on genome-wide SNPs and that by Kimura et al. (2014) based on microsatellite makers.

The average $\pi_{ws}$ in each population was close to the average $\pi_{ws}$ of 12 nuclear loci (0.0044) reported by Fujimoto et al. (2008) obtained with the Sanger method (Table 3). When we estimated the error rate of NGS by comparing the sequences from NGS with those from the Sanger method at five loci, we found one error-prone locus. If such error-prone loci were abundant, then the silent nucleotide diversity estimated from NGS data would be overestimated and would differ from estimates obtained by the Sanger method. Because of the similarity between our estimate of the silent nucleotide diversity and the estimate by Fujimoto et al. (2008) and because our results for population structure were compatible with those of previous studies, we assumed that the error-prone locus was exceptional and that our NGS data were sufficiently accurate and could be used for population genetic analyses.

## Population history and change in the physical environment

Previous studies showed that the *C. japonica* population on Yakushima Island had a high degree of genetic differentiation from the other Japanese populations on the mainland (Kimura et al. 2014; Tsumura et al. 2014),

although it has been unclear when the Yakushima population diverged from the others. Our results indicated that the Yakushima population first split from the others approximately 0.85 MYA and the other three populations then split from each other approximately 0.32 MYA. These inferences of population history differed from those obtained by Kimura et al. (2014). A period of volcanic activity in the southern part of Kyushu Island, which is located just north of Yakushima Island, started ca. 0.9 MYA (Chapman et al. 2009). This volcanism might have induced the separation of the YKU population and the populations on Kyushu Island and consequently of those elsewhere in Japan. On the other hand, trees in Japanese warm-temperate evergreen oak forests seem to have been able to migrate between the Yakushima and Kyushu islands (e.g., see Yoshida et al. 2014) during the last 1 MY, and actually two islands were connected during several glacial maxima (see the coastline during the Last Glacial Maximum (LGM) in Fig. 1). Therefore, there were some isolation barriers between the Yakushima population and other populations of *C. japonica* since 0.85 MYA. For example, environmental conditions during the glacial maxima might not have allowed the migration of *C. japonica* between the two islands.

Our estimate also indicated that the remaining three populations (the AST, BJD, and AJG populations) separated from each other approximately 0.32 MYA. Fossil pollen of *Cryptomeria* is abundant from interglacial periods in the Quaternary and is dated to a maximum of approximately 0.35–0.38 MYA (Tsukada 1982, Igarashi et al.
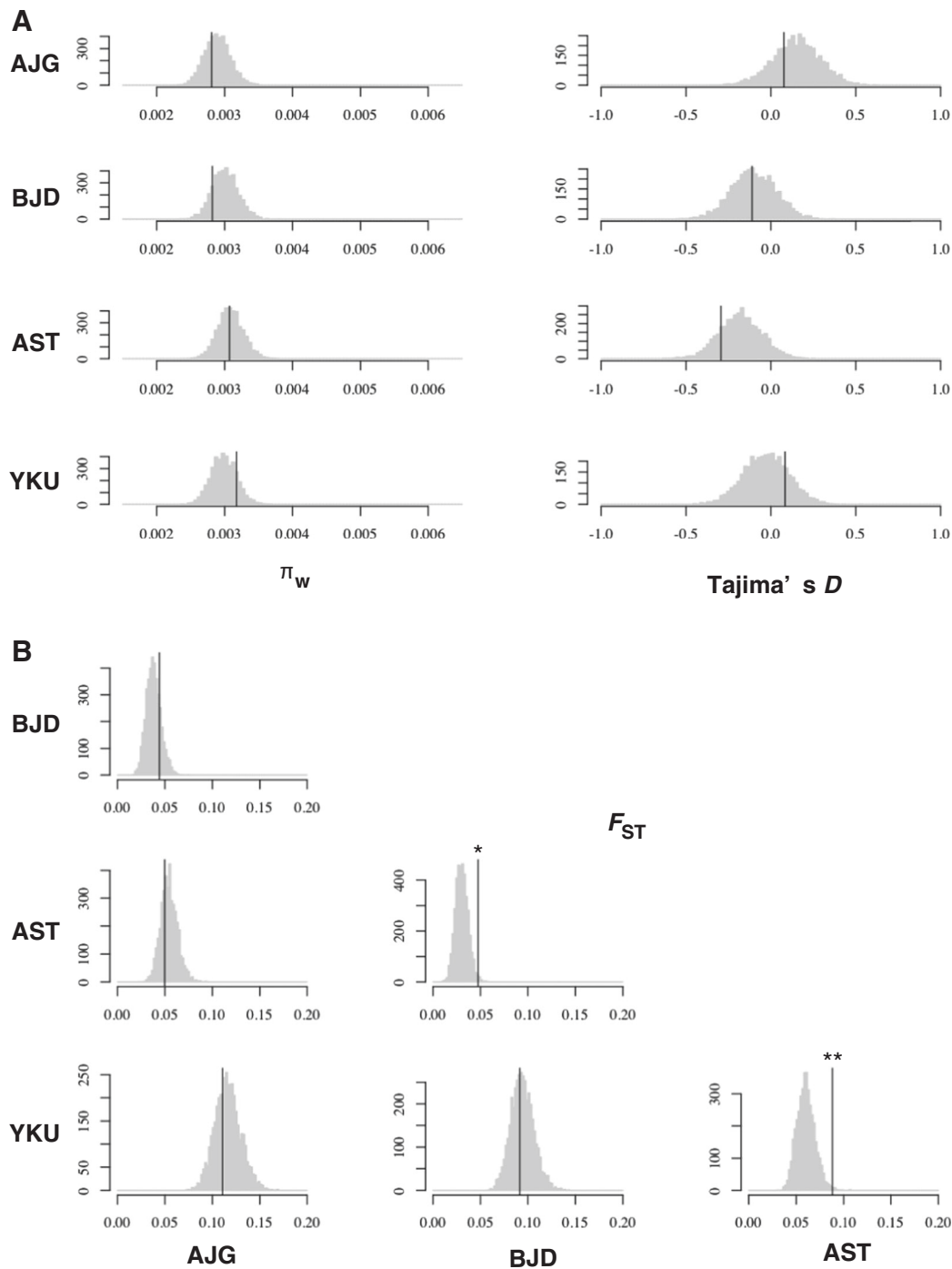
**Fig. 4** Distributions of $\pi_w$, Tajima's $D$, and $F_{ST}$ expected from the estimated parameters. Distributions of (**a**) $\pi_w$, Tajima's $D$, and (**b**) $F_{ST}$ obtained by simulation with 4000 replicates assuming the estimated parameters by *fastsimcaol2* are shown. The observed values are indicated by vertical lines. Observed values with asterisks significantly deviated from the predicted distributions at the 5% (*) and 1% (**) levels

2018). Our estimate suggested that some geographical or climatic events might have caused fragmentation of the large *C. japonica* population approximately 0.32 MYA. This estimated date of fragmentation is much earlier than the last interglacial period (0.14–0.12 MYA), and its 95% CI did not overlap with this period; instead, it corresponded to the third most recent interglacial period (Hansen et al. 2013). Igarashi et al. (2018) suggested that the increase in *Cryptomeria* pollen abundance in the sediments collected from the Japan Sea beginning 1.7 MYA may be related to

the enhanced East Asian winter monsoon and strong inflow of the Tsushima Current into the Japan Sea during inter-glacial periods (Gallagher et al. 2015), which may have promoted heavy snow along the Japan Sea side. Therefore, adaptation to heavy snowfalls may have occurred along the Japan Sea side long before the LGM. In addition, pollen analysis suggested that multiple refugia existed on the mainland of Japan during the LGM (Tsukada 1986). This suggestion supports our results, which indicated that the three populations were not connected in the last and perhaps second-to-last interglacial periods.

Previous results based on microsatellite data (Kimura et al. 2014), which suggested simultaneous splits of the four clusters represented by the four populations studied here in the early phase of the LGM (76,000 years ago), somewhat contradicted our results. Two differences in the methods of inference between our study and that of Kimura et al. (2014) may explain these discrepancies. First, the models employed by Kimura et al. (2014) did not incorporate migration. This may have led to underestimation of the times of splits between populations. Second, Kimura et al. (2014) used nuclear microsatellite markers for which the mutation rate, the estimate of which is necessary for time estimation, has not yet been well studied in *C. japonica*. For our inference, we used nucleotide sequences for which mutation rates could be estimated from the rate and time of divergence between related species (in our case, *T. distichum*). One may be concerned with the constancy of the mutation rate given that the evolutionary rates of *C. japonica* and *T. distichum*, which separated 90 MYA (Leslie et al. 2012), were used to estimate the mutation rate. However, the synonymous substitution rates in the lineages of the two species are similar ($6.7 \times 10^{-10}$ and $5.9 \times 10^{-10}$ per year, respectively; see Kusumi et al. 2015). Although any comparison of the accuracies of inferences among different genetic markers or different statistical methods may not be straightforward, especially when sample sizes are different, we believe that our estimation of the order and the ages of population splits is more reliable than that of the previous study.

## Genetic differentiation between the YKU population and the other populations

*C. japonica* has two main lines, omote-sugi, found on the Pacific Ocean side of Japan, and ura-sugi, found on the Japan Sea side. In our samples, the YKU and AST populations belong to omote-sugi, and the BJD and AJG populations belong to ura-sugi. These lines are distinguishable by their morphological characters (Yamazaki 1995) and seem to have adapted to the climate of their habitats. The genetic differentiation between the lines was recognized in

previous genetic studies (Tsumura et al. 2012, 2014) and in our study. Such phenotypic differentiation and genetic differentiation between the populations of the Japan Sea and Pacific sides has also been found in other tree species (e.g., *Fagus crenata*; see Hiraoka and Tomaru 2009). However, the YKU population showed a high degree of genetic differentiation from the other Japanese populations on the mainland, and our results suggested that the YKU population diverged earlier than did the other mainland populations. These results indicate that omote-sugi is not monophyletic. In addition to this genetic differentiation, previous studies suggested that *C. japonica* on Yakushima Island has distinct morphological and physiological characteristics, for example, higher levels of resin (Toda and Sato 1969) and short, open, and hard needles (Kimura et al. 2014). Considering the large genetic distance and morphological and physiological differences between the YKU population and the other three populations, it may be reasonable to classify the populations on Yakushima Island as a new variety.

## Adequacy of the model

Examination of the fit of the model used in the inference of demographic history showed that the observed values of $F_{ST}$ between the AST and BJD populations and between the AST and YKU populations were larger than those predicted from the parameters estimated by *fastsimcoal2* (Fig. 4). This discrepancy may have been caused by simplification of the model adopted for the inference. Although three populations, namely, the AST, BJD, and AJG populations, were assumed to have separated at the same time in Model 2, the results of TreeMix indicated that the split between the BJD and AJG populations occurred later than the other splits. In addition, the estimate of $F_{ST}$ between the BJD and AJG populations was smaller than the estimates of $F_{ST}$ between the AST and BJD populations and between the AST and AJG populations (Table 1). The lower predicted value of $F_{ST}$ between the AST and BJD populations might have been caused by the assumption of simultaneous splits of the three populations. The lower predicted value of $F_{ST}$ between the AST and YKU populations may have been caused by different factors, for example, a lack of migration between the ancestral population and the YKU population in Model 2 or a lack of change in population size after the splits of the BJD, AJG, and AST populations. Indeed, the average values of Tajima's $D$ in the AST population were significantly negative, suggesting that recent population expansion might have occurred in this population after the LGM. In addition, migration between non-neighbouring populations was not included in Model 2, although the result of the analysis by TreeMix indicated migration from

the AJG population to the AST population. These simplifications were made because we thought that the amount of data was insufficient to assume more complex models with a larger number of parameters.

## Effects of selection on inferences

In the inference of demographic history, we first tested whether the loci used had been influenced by selection by using Tajima's *D* and BayeScan; we did not detect any traces of selection at any of the 120 loci. Therefore, our inference was probably unaffected by positive selection. Recently, however, Ewing and Jensen (2016) noted that intermediate levels of background selection bias influenced estimates of past demographic history, especially estimates of effective population size. The bias is such that effective population size is further underestimated as the time moves towards the past. This might be true in our case; that is, the ancestral population size might have been much larger, which may agree with the data of fossil pollen showing maximum abundance approximately 0.35–0.38 MYA (Igarashi et al. 2018). However, magnitudes of negative selection have not been studied in *C. japonica*, except for those at a few nuclear loci examined by Fujimoto et al. (2008). Thus, there is a need for systematic studies of background selection in *C. japonica*, as was done for fruit flies (see Comeron 2014), to evaluate the effects of this background selection on such estimation. Because of the large genome size of *C. japonica* (Hizume et al. 2001; Tsumura 2011 for a review of the species), the possibility of carrying out this line of research in the near future seems remote. An alternative way of dealing with this effect is to use non-coding marker genes, which are considered to be less subject to selection, such as RAD markers (Miller et al. 2007) and MIG-seq markers (Suyama and Matsuki 2015).

## Conclusion

Our results suggested that the Yakushima population of *C. japonica* diverged first from the other populations 0.85 MYA and that the divergence between the two lines, namely, omote-sugi and ura-sugi, on mainland Japan occurred 0.32 MYA. These results imply that the Yakushima population has accumulated locally adaptive variations because of its long isolation and location at the southern edge of the species distribution and that the traits characterizing ura-sugi such as tolerance of snow might be derived. In addition, our study showed that the amplicon sequencing used here, with information available on the mutation rate at target genes and more sharing of sites among individuals, is a promising approach for inferring

demographic history. However, in terms of being realistic, the model we used for inference still has room for improvement, and we have to consider the effects of background selection. Further studies using a larger number of non-coding markers are also necessary to resolve these issues.

## Data archiving

The sequence data from this study have been submitted to the DDBJ Sequence Read Archive (https://www.ddbj.nig.ac.jp/dra/index.html) under accession no. DRA007815–DRA007910.

## Compliance with ethical standards

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

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## References

Bybee SM, Bracken-Grissom H, Haynes BD, Hermansen RA, Byers RL, Clement MJ et al. (2011) Targeted amplicon sequencing (TAS): a scalable next-gen approach to multi-locus, multi-taxa phylogenetics. Genome Biol Evol 3:1312–1323

Begon M, Townsend CR, Harper JL (2006) Ecology. Blackwell, Oxford

Bennett K (1997) Evolution and ecology: the pace of life. Cambridge University Press, Cambridge

Chapman N, Apted M, Beavan J, Berryman K, Cloos M, Connor C et al. (2009) Development of methodologies for the identification of volcanic and tectonic hazards to potential HLW repository sites in Japan: the Kyushu Case Study. Technical Reports of Nuclear Waste Management Organization of Japan (NUMO) NUMO-TR-09-02

Comeron JM (2014) Background selection as baseline for nucleotide variation across the drosophila genome. PLoS Genet 10: e1004434

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359–361

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

Ewing GB, Jensen JD (2016) The consequences of not accounting for background selection in demographic inference. Mol Ecol 25:135–141

Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M (2013) Robust demographic inference from genomic and SNP data. PLoS Genet 9:e1003905

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10:564–567

Farjon A (2005) A monograph of Cupressaceae and Sciadopitys. Royal Botanical Gardens, Kew, Richmond, UK

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics 180:977–993

Fujimoto A, Kado T, Yoshimaru H, Tsumura Y, Tachida H (2008) Adaptive and slightly deleterious evolution in a conifer. Cryptomeria Jpn J Mol Evol 67:201–210

Gallagher SJ, Kitamura A, Iryu Y, Itaki T, Koizumi I, Hoiles PW (2015) The Pliocene to recent history of the Kuroshio and Tsushima Currents: a multi-proxy approach. Progress in Earth and Planetary Science 2:17

Hansen J, Sato M, Russell G, Kharecha P (2013) Climate sensitivity, sea level and atmospheric carbon dioxide. Philos Trans A Math Phys Eng Sci 371:20120294

Hayashi Y (1951) The natural distribution of important trees, indigenous to Japan, Conifers report I. Bull Gov For Exp Station 48:1–240 (in Japanese)

Heuertz M, De Paoli E, Källman T, Larsson H, Jurman I, Morgante M et al. (2006) Multilocus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway spruce [*Picea abies* (L.) Karst]. Genetics 174:2095–2105

Hiraoka K, Tomaru N (2009) Genetic divergence in nuclear genomes between populations of *Fagus crenata* along the Japan Sea and Pacific sides of Japan. J Plant Res 122:269–282

Hizume M, Kondo T, Shibata F, Ishizuka R (2001) Flow cytometric determination of genome size in the Taxodiaceae, Cupressaceae sensu stricto and Sciadopityaceae. Cytologia 66:307–311

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. Mol Ecol Resour 9:1322–1332

Hudson RR, Slatkin M, Maddison WP (1992) Estimation of levels of gene flow from DNA sequence data. Genetics 132:583–589

Igarashi Y, Irino T, Sawada K, Song L, Furota S (2018) Fluctuations in the East Asian monsoon recorded by pollen assemblages in sediments from the Japan Sea off the southwestern coast of Hokkaido, Japan, from 4.3 Ma to the present. Glob Planet Change 163:1–9

Ikezaki Y, Suyama Y, Middleton BA, Tsumura Y, Teshima K, Tachida H et al. (2016) Inferences of population structure and demographic history for *Taxodium distichum*, a coniferous tree in North America, based on amplicon sequencing analysis. Am J Bot 103:1937–1949

Kado T, Yoshimaru H, Tsumura Y, Tachida H (2003) DNA variation in a conifer, *Cryptomeria japonica* (Cupressaceae sensu lato). Genetics 164:1547–1559

Kimura MK, Uchiyama K, Nakao K, Moriguchi Y, San Jose-Maldia L, Tsumura Y (2014) Evidence for cryptic northern refugia in the last glacial period in *Cryptomeria japonica*. Ann Bot 114:1687–1700

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Mol Ecol Resour 15:1179–1191

Kusumi J, Tsumura Y, Tachida H (2015) Evolutionary rate variation in two conifer species, *Taxodium distichum* (L.) Rich var. *distichum* (baldcypress) and *Cryptomeria japonica* (Thunb. ex L.f.) D.Don (Sugi, Japanese cedar). Genes Genet Syst 90:305–315

Leslie AB, Beaulieu JM, Rai HS, Crane PR, Donoghue MJ, Mathews S (2012) Hemisphere-scale differences in conifer evolutionary dynamics. Proc Natl Acad Sci USA 109:16217–16221

Li H, Durbin (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–60

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N et al. (2009) The sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079

Librado P, Rozas J (2009) DnaSPv5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452

Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Res 17:240–248

Moriguchi Y, Tani N, Itoo S, Kanehira F, Tanaka K, Yomogida H et al. (2005) Gene flow and mating system in five *Cryptomeria japonica* D. Don seed orchards as revealed by analysis of microsatellite markers. Tree Genet Genomes 1:174–183

Nielsen R (2005) Molecular signatures of natural selection. Annu Rev Ecol Syst 39:197–218

Nielsen R, Beaumont MA (2009) Statistical inferences in phylogeography. Mol Ecol 18:1034–1047

Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ (2014) PopGenome: an efficient swiss army knife for population genomic analyses in R. Mol Biol Evol 31:1929–1936

Pickrell JK, Pritchard JK (2012) Inference of population splits and mixtures from genome-wide allele frequency data (H Tang, ed.). PLoS Genet 8:e1002967

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

Suyama Y, Matsuki Y (2015) MIG-seq: an effective PCR-based method for genome-wide single-nucleotide polymorphism genotyping using the next-generation sequencing platform. Sci Rep 5:16963

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

Takahashi T, Tani N, Taira H, Tsumura Y (2005) Microsatellite markers reveal high allelic variation in natural populations of *Cryptomeria japonica* near refugial areas of the last glacial period. J Plant Res 118:83–90

Takahashi T, Tani N, Niiyama K, Yoshida S, Taira H, Tsumura Y (2008) Genetic succession and spatial genetic structure in a natural old growth *Cryptomeria japonica* forest revealed by nuclear and chloroplast microsatellite markers. Ecol Manag 255:2820–2828

Toda R, Sato T (1969) Variety, breeding and the expression of the characteristics of the variety of *Cryptomeria japonica*. In: Sakaguchi M (ed). Sugi no subete. Zenkoku Ringyo Fukyu Kyokai, Tokyo, p 64–70 (in Japanese)

Tsukada M (1982) *Cryptomeria Japonica*: glacial refugia and late-glacial and postglacial migration. Ecology 63:1091–1105

Tsukada M (1986) Altitudinal and latitudinal migration of *Cryptomeria japonica* for the past 20,000 years in Japan. Quat Res 26:135–152

Tsumura Y (2011) Cryptomeria. In: Kole C (ed) Wild crop relatives: genomic and breeding resources: forest trees. Springer-Verlag, Berlin Heidelberg, p 49–63

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

Tsumura Y, Ohba K (1992) Allozyme variation of five natural populations of *Cryptomeria japonica* in western Japan. Jpn J Genet 67:299–308

Tsumura Y, Uchiyama K, Moriguchi Y, Kimura MK, Ueno S, Ujino-Ihara T (2014) Genetic differentiation and evolutionary adaptation in *Cryptomeria japonica*. G3 4:2389–2402

Tsumura Y, Uchiyama K, Moriguchi Y, Ueno S, Ihara-Ujino T (2012) Genome scanning for detecting adaptive genes along environmental gradients in the Japanese conifer. Cryptomeria Jpn Hered 109:349–360

Uchiyama K, Ujino-Ihara T, Ueno S, Taguchi Y, Futamura N, Shinohara K et al. (2012) Single nucleotide polymorphisms in *Cryptomeria japonica*: their discovery and validation for genome

mapping and diversity studies. Tree Genet Genomes 8:1213–1222

Yamazaki T (1995) Cryptomeriaceae. In: Iwatsuki K, Yamazaki T, Boufford DE, Ohba H (eds) Flora of Japan, vol 1: Pteridophyta and Gymnospermae. Kodansya, Tokyo, p 264

Yoshida T, Tamekuni M, Yahara T, Inomata N, Tachida H (2014) Demographic history of a common pioneer tree, *Zanthoxylum ailanthoides*, reconstructed using isolation-with-migration model. Tree Genet Genomes 10:1213–1222

Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics 28:3326–3328