Tree Genetics & Genomes (2017) 13:121
https://doi.org/10.1007/s11295-017-1202-4

ORIGINAL ARTICLE

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

Lerma San Jose-Maldia[1,2] · Asako Matsumoto[2] · Saneyoshi Ueno[2] · Ayako Kanazashi[2] · Munetake Kanno[3] · Kanji Namikawa[4] · Hiroshi Yoshimaru[2] · Yoshihiko Tsumura[2,5]

Received: 21 November 2016 / Revised: 30 August 2017 / Accepted: 8 October 2017

Abstract In this study, we assessed geographic patterns of genetic variations in nuclear and chloroplast genomes of two related native oaks in Japan, *Quercus aliena* and *Q. serrata*, in order to facilitate development of genetic guidelines for transfer of planting stocks for each species. A total of 12 populations of *Q. aliena* and 44 populations of *Q. serrata* were analyzed in this study. Genotyping of nuclear microsatellites in *Q. aliena* was done with only nine populations ($n = 212$) due to limited numbers of individuals in two populations, while all 12 populations ($n = 89$) were used in sequencing chloroplast DNA (cpDNA). In *Q. serrata*, 43 populations ($n = 1032$) were genotyped by nuclear microsatellite markers, while cpDNA of 44 populations ($n = 350$) was sequenced. As anticipated, geographic patterns detected in the variations of *Q. aliena*'s nuclear genome and its chloroplast haplotype distribution clearly distinguished northern and southern groups of populations. However, those of *Q. serrata* were inconsistent. The geographic distribution of its chloroplast haplotypes tends to show the predicted differentiation between northern and southern lineages, but geographic signals in the genetic structure of its nuclear microsatellites are weak. Therefore, treating northern and southern regions of Japan as genetically distinct transferrable zones for planting stocks is highly warranted for *Q. aliena*. For *Q. serrata*, the strong NE-SW geographic structure of cpDNA should be considered.

Keywords *Quercus serrata* · *Quercus aliena* · Chloroplast DNA · Microsatellites · Phylogeographic structure · Transfer zone

Communicated by A. Kremer

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11295-017-1202-4) contains supplementary material, which is available to authorized users.

✉ Yoshihiko Tsumura
tsumura.yoshihiko.ke@u.tsukuba.ac.jp

[1] College of Forestry and Natural Resources, University of the Philippines Los Baños, 4031 Laguna, Philippines

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

[3] IDEA Consultants, Inc. Tohoku Branch, Natural Environment Conservation, Sendai, Japan

[4] Hokkaido University of Education Sapporo, 5-3-1 Ainosato, Kita-ku, Sapporo 002-8502, Japan

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

Introduction

Phylogeographic studies have provided significant insights into historical changes in the distribution and genetic structures of extant populations. In recent decades, there has also been increasing interest in inferring biogeographic histories from comparisons of phylogeographic patterns of genetic variations in related and co-distributed taxa (Avise 2000; Arbogast and Kenagy 2001). Such analyses have shown (inter alia) that the past and current patterns depend on the major factors involved. Some groups of related taxa display strong, common phylogeographic and population genetic structure because they have been largely affected by similar past geological events and climatic conditions (reviewed in Arbogast and Kenagy 2001). However, structures of other groups are highly complex because the taxa have been affected by biogeographic or ecological factors at least as strongly as by common historical events (e.g., Maliouchenko et al. 2007;

Published online: 31 October 2017

Toyama and Yahara 2009; reviewed in Ohsawa and Ide 2011; Guicking et al. 2011; Iwasaki et al. 2012).

The genetic structure of forest trees, especially the geographic component, is particularly relevant for the management and conservation of genetic resources (Petit et al. 1993). In addition, molecular phylogenies have potential utility not merely for accurately defining target entities for conservation but also for predicting trends of long-term population processes, which may be related to translocation issues (Moritz 1995; reviewed in Hufford and Mazer 2003).

Previous studies suggest that introducing foreign genotypes into new environments could result in loss of local adaptation and lead to outbreeding depression, genetic swamping (Templeton 1986; reviewed in Hufford and Mazer 2003), mortality (Campbell 1979; Hamann et al. 2011; McKay et al. 2005), and/or threats to regional genetic and species diversity (Vander Mijnsbrugge et al. 2010). Thus, efforts to offset potential risks of translocation require careful consideration of seed sources and the establishment of "seed transfer zones" or regions, within which planting stocks (seeds and seedlings) can be moved with minimal detrimental effects on the average fitness of populations (Hufford and Mazer 2003). Most guidelines for seed zoning and transfer have been based on ecological classifications, climatic data, ecophysiological information, and/or results of provenance tests (Gomory et al. 1998). However, following advances in population genetic analyses, there has been growing awareness of the utility of geographic patterns of genetic variations in neutral molecular markers in the delineation of robust seed zones for genetic conservation and restoration activities (e.g., Gomory et al. 1998; Weels et al. 2003; Krauss and Koch 2004; Moncada et al. 2007; Keir et al. 2011; De Souza et al. 2012). Recently, genetic guidelines for transferring planting stocks have been proposed for a number of widely distributed broadleaf species in Japan based on results of several studies using molecular markers (Forestry and Forest Products Research Institute 2011, ISBN:978-4-902606-75-1; Tsumura and Suyama 2015). The guidelines include recommendations that transfer zones should be strongly delineated (i.e., transfers into or out of identified zones should be strongly avoided) when a species displays significant congruent geographic patterns of genetic differentiation in both nuclear and chloroplast DNA, while provisional boundaries between zones should be delineated when the geographic pattern is significant in only one genome. The latter does not necessarily entail strict transfer of planting stocks within distinct regions, but encourages consideration of the observed geographic structure. Broadleaf trees are mostly used for restoration activities, greening programs, and establishment of forest parks and green recreational areas in Japan, and such genetic guidelines would be particularly useful for planning and designing strategies for seed and seedling transfer.

There are larger volumes of oaks (*Quercus* spp.) in Japan than any other genera of broadleaf trees, and they are considered one of the most important tree groups in the country (Taoda 2005; Tani and Kawawata 2008). Their importance in Japan rests more on ecological, social, and esthetic values than on economic revenues, except in certain parts of the country where harvesting of some oaks in natural forests for commercial use is being promoted (Ohsawa et al. 2011). Some *Quercus* species comprise the main component of deciduous temperate forests of Japan (Kanno et al. 2004; Taoda 2005), which include four naturally occurring species of section *Prinus*; *Quercus serrata*, *Q. crispula*, *Q. aliena*, and *Q. dentata* (Kitamura and Horikawa 1951). These species all have differing distribution patterns and habitat preferences (Kanno et al. 2004; Okaura et al. 2007). *Q. serrata* and *Q. crispula* have wide distributions, while *Q. aliena* and *Q. dentata* have highly fragmented populations in Japan, and *Q. serrata* and *Q. aliena* are better adapted to warm regions, while *Q. crispula* and *Q. dentata* prefer cool climates.

According to phylogeographic structures of the four related species previously inferred from analyses of chloroplast DNA (cpDNA) variation, they all have two distinct lineages, separated in northeast and southwest regions of central Japan by major mountain chains (Kanno et al. 2004; Okaura et al. 2007). However, the studies were not based on sufficient genetic data for all congeneric species to support the generally inferred phylogeographic structure of these oaks, due to a lack of either sufficient polymorphism or sufficient numbers of surveyed populations. Moreover, genetic markers with multiple modes of inheritance should ideally be examined to elucidate the phylogeographic structure of a species fully, due to associated variations in levels of gene flow. Many previous studies have already shown that a species may have varying geographic structure in its genomes. Therefore, it is essential to consider the genetic structures of different genomes of focal species, especially when developing genetic guidelines for designating seed and seedling transfer zones, such as those of the Forestry and Forest Products Research Institute (FFPRI) mentioned above. Thus, in this study we assessed the geographical patterns of genetic variations in *Q. aliena* and *Q. serrata*, by analyzing both cpDNA sequences and nuclear microsatellite markers from expressed sequence tag sites (EST-SSRs). CpDNA has been used in many phylogeographic studies because it retains strong signatures of population history, due to its slow evolution and lack of recombination (Hillis and Moritz 1990). A major strength (and weakness) is that it is maternally inherited, thus it reflects the genetic structure and history of single lineages. In contrast, the analysis of nuclear microsatellite markers can reveal the genetic structure and population history at bi-parentally inherited loci. The high variability of these markers enables higher resolution analysis of genetic structure, with the power to detect loss of genetic (allelic) diversity, which is useful for predicting

potential long-term population genetic risks associated with translocation. The main purpose of this study was to obtain genetic information to facilitate the development of genetic guidelines for transfer of planting stocks of the two species.

Materials and methods

Characteristics of the study species

Quercus aliena and *Q. serrata* co-occur in secondary forests in temperate zones of Japan (Kanno et al. 2004; Iwabuchi et al. 2006; Okaura et al. 2007). *Q. serrata* is generally more abundant, but they have different patterns of geographic distribution across Japan. *Q. aliena* is more common in southwestern Japan, but mostly occurs in the Chugoku region (Fig. 5c), and is sporadically distributed (mostly in forests on alluvial plains around lakes and rivers) in northeastern Japan (Nozaki et al. 2001; Kanno et al. 2004). *Q. serrata* is widely distributed from the Ishikari lowlands (43° 32′ N, 141° 55′ E) in the southern tip of the northern island Hokkaido to Kyushu (31° 34′ N, 130° 48′ E) in southern Japan, except for populations at the northern limit, which are relatively small and fragmented (Kitamura and Horikawa 1951; Kanno et al. 2004). Outside of Japan, the two species are naturally distributed and co-occur in secondary forests in Korea (You et al. 1995; Lee and You 2012) and China (Kitamura and Horikawa 1951; Tani and Kawawata 2008).

We analyzed a total of 12 populations of *Q. aliena* and 44 populations of *Q. serrata* distributed across almost the entire ranges of the species in Japan (see Table 1, Fig. 5a–c) and for detailed information about the populations, Online Resource 1). Due to limited numbers of individuals in some populations of *Q. aliena*, only nine populations (n = 212) were used for genotyping nuclear microsatellite markers, while all 12 populations (n = 89) were used for sequencing cpDNA. For analysis of *Q. serrata*, 43 populations (n = 1032) were genotyped by nuclear microsatellite markers, while cpDNA of 44 populations (n = 350) was sequenced.

DNA extraction, microsatellite genotyping, and sequencing of chloroplast DNA regions

DNA was extracted using a DNeasy Plant Mini Kit (Qiagen, Hilden, Germany). Each individual was genotyped at 30 simple sequence repeats of expressed sequence-tagged sites (EST-SSR) loci. All the loci except three (see Online Resource 1) were amplified using primers developed from *Quercus mongolica* var. *crispula* (Ueno et al. 2008; Ueno and Tsumura 2008; and this study); *Castanopsis sieboldii* var. *sieboldii* (Ueno et al. 2009a); and *Fagus crenata* (Ueno et al. 2009b) and a multiplex PCR Kit (Qiagen, Hilden, Germany). The (7-μL) reaction mixtures contained 3.0 μL Qiagen Master Mix, 0.2 μM of each forward and reverse primer, and 10 ng/μL of genomic DNA. The temperature program, provided by a GeneAmp 9700 cycler (Applied Biosystems), consisted of 15 min initial denaturation at 95 °C, followed by 35 cycles of denaturation at 94 °C for 30 s, annealing at 58 °C for 1 min and 30 s, extension at 72 °C for 1 min, and a final extension at 60 °C for 30 min. The products from each locus were separated using a 3100-Avant Genetic Analyzer (Applied Biosystems), and their sizes were scored using Genotyper 3.2 (Applied Biosystems).

Four polymorphic and informative non-coding regions (Online Resource 3a) screened from a total of 43 cpDNA regions (Online Resource 3b) in the preliminary analysis were sequenced and analyzed in samples of both species. Previously reported primers were used to amplify three of the analyzed regions: the *trn*T (UGU) –*trn*L (UAA) 5′exon spacer (Taberlet et al. 1991), *rps*16 intron (Shaw et al. 2005), and *rp*L32-*trn*L (Shaw et al. 2007). The sequences of forward and reverse primers used to amplify the 3′ to_*rps*2 fragment (5′-GTCATATATTTGATCCCGCC-3′ and 5′-AACC GGAACTAGTCGGATG-3′, respectively) have not been previously published. The PCR conditions were the same as those used for the EST-SSR amplification, except that each region was amplified in 10-μL reaction mixtures, containing 5 μL Qiagen Master Mix, and the annealing temperature was reduced to 57 °C. The products were purified using ExoSAP-IT™ (GE Healthcare Limited) and sequenced in both directions using a BigDye Terminator Sequencing Kit (PE Biosystems). The sequencing products were then purified by ethanol precipitation and analyzed in a 3100 Genetic Analyzer (Applied Biosystems). Sequence data were assembled and manually edited using Sequencher 10.4.1 (Gene Codes Corporation).

Analyses of nuclear microsatellite data

The genetic diversity of microsatellite loci at each locus and over all loci, across populations in each species, was estimated in terms of the total number of alleles (A) and Nei's unbiased estimate of gene diversity within populations statistic (H_S; Nei 1987). Across loci, the genetic diversity within populations in each species was assessed in terms of the unbiased expected heterozygosity (H_E; Nei 1987) and allelic richness (AR; El Moussadik and Petit 1996; Petit et al. 1998). The F_{IS} value of each population was calculated to measure the degree of inbreeding, and significant deviations of F_{IS} values from 0 were determined by permutation tests, with Bonferroni correction (Rice 1989). FSTAT 2.9.3.2 (Goudet 2002) was used for calculating all these statistics and significance tests.

The genetic differentiation at each locus and over all loci, across populations in each species, was estimated by the fixation index, F_{ST} (Wright 1951; Weir and Cockerham 1984). Population differentiation was tested, without assuming

Table 1 Estimates of genetic diversity, inbreeding coefficients, and results of tests for bottlenecks in populations of *Q. aliena* and *Q. serrata* based on 30 EST-SSR markers

Species/population code	Number of analyzed trees		*AR*	H_E	F_{IS}[a]	Bottleneck[b]	Hd
	EST-SSR	cpDNA					
Quercus aliena							
AKT	20	8	5.0	0.565	0.085	0.04598	0
IWT		8					0.429
MYG	24	8	4.6	0.530	0.058	0.00280	0
FKS	24	7	5.2	0.572	0.082	0.03495	0
NGT	24	7	5.2	0.583	0.059	0.01013	0
MIE		4					0.833
SHG	24	8	4.9	0.586	0.090*	0.00278	0.250
OSK	24	8	5.1	0.567	0.033	0.02746	0
OKY	24	8	5.1	0.545	0.022	0.21398	0
HRS	24	8	5.0	0.539	− 0.007	0.16937	0
FKK	24	8	5.8	0.615	0.017	0.00565	0.250
KMM		7					0.286
Average	23.4	7.4	6.1	0.571			
Quercus serrata							
TKK	24	8	5.7	0.592	0.077	0.16423	0
HID	24	8	5.3	0.595	0.139***	0.00310	0.679
HKD	24	8	4.5	0.538	0.001	0.41840	0.571
AMR	24	8	5.6	0.588	0.093***	0.06737	0
MOR	24	7	5.9	0.604	0.078	0.19653	0
SND	24	7	5.5	0.578	0.027	0.15426	0
AKT	24	8	5.5	0.585	0.090	0.03028	0
YGT	24	8	5.8	0.590	0.023	0.07163	0.536
NAK	24	8	5.1	0.554	0.079	0.20225	0.250
TUK	24	8	5.6	0.593	0.087	0.06468	0.429
KNG	24	8	5.6	0.562	0.066	0.24492	0
CHB	24	8	6.1	0.599	0.030	0.24550	0.571
TCE	–	8	–	–	–	–	0.250
NIK	24	8	5.9	0.609	0.079	0.10647	0
GUN	24	8	5.4	0.585	0.047	0.02363	0
YAM	24	8	5.7	0.600	0.030	0.04184	0
YNS	24	8	5.9	0.594	0.086	0.20225	0.679
IYA	24	8	5.8	0.589	0.040	0.11032	0
IDA	24	8	5.3	0.594	0.119***	0.03028	0.250
NIG	24	8	5.6	0.561	0.087	0.23227	0
TYM	24	8	5.5	0.593	0.128***	0.01636	0
AIC	24	8	5.8	0.603	0.058	0.02401	0
AON	24	8	5.8	0.582	0.046	0.23855	0.571
SZK	24	8	5.5	0.579	0.064	0.09904	0
GFU	24	8	5.5	0.592	0.040	0.00347	0.250
KOM	24	8	5.9	0.587	0.018	0.14942	0.429
MIE	24	8	5.6	0.581	0.058	0.26455	0
KYO	24	8	5.6	0.602	0.044	0.04203	0
NAR	24	8	5.7	0.595	0.091***	0.01636	0
WAK	24	8	5.9	0.595	0.068	0.45161	0
WKY	24	8	5.5	0.591	− 0.008	0.13106	0.429
TSU	24	8	5.6	0.588	− 0.009	0.06737	0.571
OKA	24	8	6.0	0.624	0.083	0.03333	0

Table 1 (continued)

Species/population code	Number of analyzed trees		*AR*	H_E	F_{IS}[a]	Bottleneck[b]	Hd
	EST-SSR	cpDNA					
SHM	24	8	5.5	0.574	0.023	0.22608	0
YGC	24	8	5.9	0.603	0.028	0.08209	0.250
TOK	24	8	5.4	0.571	0.059	0.16787	0.821
KGW	24	8	6.1	0.609	0.044	0.06207	0
EHI	24	8	5.6	0.597	− 0.017	0.04597	0.571
ODO	24	8	5.5	0.600	0.030	0.01042	0
FKK	24	8	5.6	0.623	0.042	0.01042	0.821
TAT	24	8	5.6	0.620	0.010	0.01311	0
OIT	24	8	5.8	0.589	0.024	0.24492	0
MYZ	24	8	5.7	0.588	0.077	0.14468	0
KAG	24	8	5.5	0.594	0.053	0.09546	0.429
Average	24.0	8.0	5.6	0.590			0.838

AR average allelic richness computed based on minimum sample size of 19 and 18 individuals of *Q. aliena* and *Q. serrata*, respectively, H_E unbiased expected heterozygosity, F_{IS} inbreeding coefficient, *Hd* haplotype diversity (values in parentheses indicate estimates from pooled individuals)

Significant at $^*P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$

[a] *P* values calculated from proportions of permutations that gave larger than observed values of F_{IS} were used to test if F_{IS} within populations significantly differed from zero and were assessed against the adjusted 5% nominal level for multiple comparisons within *Q. aliena* and *Q. serrata*

[b] Bottleneck tests were conducted under infinite allele model (IAM) and stepwise mutation model (SMM) assumptions, but only *P* values associated with tests under the IAM are shown since no significant bottleneck was detected under the SMM

random mating, by calculating log-likelihood G statistics through randomizations of genotypes among samples (Goudet et al. 1996). The degree of genetic differentiation was also evaluated by the standardized G_{ST}, G'_{ST} (Hedrick 2005).

To examine geographical trends in genetic diversity, the relationships between genetic diversity (*AR* and H_E) and latitude of populations were assessed by correlation analyses and one-tailed Student's *t* tests implemented in Microsoft Excel (2010). Theoretically, genetic bottlenecks cause losses of both alleles and heterozygosity (Nei et al. 1975; Maruyama and Fuerst 1985; Allendorf and Leary 1986). However, if a population has been subjected to a recent bottleneck, its allelic diversity will theoretically have declined more than its heterozygosity, which will be larger than values expected at mutation-drift equilibrium (H_{EQ}) since H_{EQ} is calculated from the observed number of alleles at focal loci and population size (Nei et al. 1975; Maruyama and Fuerst 1985; Cornuet and Luikart 1996). To detect signatures of bottlenecks in studied populations, we used the program Bottleneck 1.2 (Piry et al. 1999) and applied the one-tailed Wilcoxon's sign-rank test (Luikart et al. 1998) to assess excess heterozygosity in each population. *P* values associated with bottleneck tests were calculated by performing 10,000 permutations under both the infinite allele model (IAM) and stepwise mutation model (SMM).

According to the key assumption underlying the isolation by distance model, genetic differentiation should be stronger between geographically distant populations than between proximal populations (Rousset 1997). To test this assumption, we assessed the relationship between populations' genetic similarity, expressed as $F_{ST}/(1 - F_{ST})$ (Rousset 1997), and the natural logarithm of linear distances between them (converted from pairwise geographic coordinates by the Geographic Distance Matrix Generator 1.2.3; Ersts 2006), using the GENEPOP 4.0.10 web platform (Raymond and Rousset 1995; Rousset 2008). The degree and significance of this relationship, for both species, were tested using Mantel tests (Mantel 1967) with 10,000 permutations implemented in GENEPOP.

Genetic structure at the individual level was inferred by model-based clustering analysis implemented in Structure 2.3.3 (Pritchard et al. 2000), first using data pertaining to both species, and then separately for each species to detect genetic structure at species and population levels. The model assumes the existence of *K* ancestral genetic clusters, to which individual multilocus genotypes are probabilistically assigned through Markov chain Monte Carlo simulations and the *K* is sought that optimizes the log probabilities of the data (LnPD), and thus provides the best model of observed genetic structure. We analyzed our data following the admixture model of ancestry, with prior information on the location of populations

(LOCPRIOR; Hubisz et al. 2009), and correlated allele frequencies to improve the clustering for closely related populations (Falush et al. 2003). Twenty replications were performed for each value of K between 1 and 9, with 100,000 iterations after 50,000 burn-in iterations. However, no clear indication of the optimal K for each species was obtained as LnPD progressively increased as K increased (data not shown). Thus, we calculated ΔK values (Evanno et al. 2005), using Structure Harvester (Earl and vonHoldt 2011), and assumed that the K with the highest ΔK value provided the best indication of clustering for each species. The genetic relationship between populations in each species was explored by generating neighbor-joining (NJ) trees (Saitou and Nei 1987) based on D_A distances (Nei et al. 1983) between populations, using Populations 1.2.30 (Langella 1999). The robustness of the relationships was assessed by bootstrap percentages from 1000 replicates.

Analyses of cpDNA data

Haplotype classification was solely based on nucleotide substitutions because there were numerous gaps due to large indels in one of the cpDNA regions (*rpL32-trn*L).The haplotype frequencies and diversities within populations (Nei 1987) were computed using DnaSP 5.10 (Librado and Rozas 2009). The total and mean diversities (H_T and H_S) and relative differentiation ($G_{ST} = D_{ST}/H_T$, where D_{ST} is the absolute differentiation computed as $H_T - H_S$; Nei 1973) in each species were estimated, following assumptions of Pons and Petit (1995), by CONTRIB 1.02 (Petit et al. 1998).

To examine the phylogenetic relationships of haplotypes, minimum spanning networks were calculated, using Prim's (1957) algorithm as implemented in Arlequin 3.5 (Excoffier and Lischer 2010) and illustrated using HapStar 0.5 (Excoffier et al. 2005; Prim 1957).

Inference of historical and demographic history by coalescence simulation using microsatellite genotype data

To infer the evolutionary history underlying the observed significant genetic differentiation at nuclear microsatellites between northeast (NE) and southwest (SW) populations of *Q. aliena* (see "Results"), we used the coalescence simulation implemented in DIYABC 1.0 (Cornuet et al. 2010). DIYABC can analyze more complex evolutionary scenarios, involving both recent and ancient historical events (Cornuet et al. 2008, 2010), than STRUCTURE, which is more reflective of recent admixture (Cornille et al. 2012). The data subset for this analysis was drawn from eight individuals of each population in each cluster, except the SHG population due to moderate admixture and population substructure observed in the STRUCTURE analysis. We evaluated four scenarios (Fig. 6) that could explain the most probable divergence history of the SW and NE populations of *Q. aliena*, including time of divergence and associated changes in demography brought by bottleneck events. Assumptions on time and associated bottlenecks were inferred from published records and/or reviews about pollen fossil records of oaks, as well as glacial and post-glacial re-colonization routes, with the assumed northward expansion of broadleaved trees in Japan, including oaks (reviewed in Okaura et al. 2007), and the associated reduction in population sizes during periods of retreat and expansion of populations in response to changes in climate.

The four scenarios differ in the presumed time of divergence of the SW and NE populations, that is, the presumed divergence time (td) in scenarios 1 and 2 was older than that in scenarios 3 and 4. In terms of associated bottlenecks, all scenarios had considered the occurrence of bottleneck (whether before or after split) in the history of the present SW and NE populations, with presumed bigger population sizes. Scenarios 1 and 3 had one bottleneck event, but the bottleneck in scenario 1 is assumed to be longer than in scenario 3. On one hand, scenarios 2 and 4 had two bottleneck events, with the more recent one (t1 to tbk and td1 to tbk, respectively) being more severe than the earlier, leading to further decrease in effective population sizes of SW and NE (i.e., N1bk and N2bk, respectively). Between scenarios 2 and 4, scenario 4 assumes that the second bottleneck from time t to tbk, which is assumed to be more severe than the earlier one, happened after the population split into SW and NE populations, resulting to population sizes N1bk and N2bk, respectively.

Specifically, the following assumptions for each scenario were inferred. In scenario 1, the SW and NE populations experienced a period of one historically long bottleneck (from time td2 to tbk) after the split of ancestral population, NA. Each population then recovered into its present population sizes, N1 and N2, respectively. In scenario 2, the SW and NE populations experienced two major bottleneck events since split of ancestral population (NA) in which the most recent (occurred from t1 to tbk) was more extreme than the initial bottleneck (occurred from td2 to t1), causing further decrease of population size after the split into SW and NE populations (population sizes are N1a and N2a, respectively), and each continue to experience further bottleneck (population sizes are N1bk and N2bk, respectively). In scenario 3, the SW and NE populations first experienced bottleneck (from time t to td1) before their split. At time td1, each population recovered into its present population sizes (N1 and N2, respectively). In scenario 4, however, the more recent bottleneck causing further decrease in population sizes is due to the split of the ancestral population that underwent bottleneck (Na).

To determine which mutation model and prior distribution of parameters (i.e., effective population size and times of split and bottleneck) provide the best fits to the observed data, we ran simulations under both generalized stepwise mutation

(GSM) and stepwise mutation model (SMM) assumptions, with uniform and log-uniform priors. The GSM model considers two mutation parameters: the mutation rate (μ) and geometric distribution (P) of the number of repeated motifs involved in mutation. Uniform prior distribution was assigned for the mean μ ($\overline{\mu}$) and mean P, and values for each locus were drawn from a gamma distribution. Single-nucleotide insertion-deletion mutation (SNI), with a mean mutation rate of μSNI, was included to take into account alleles that have not followed strict stepwise mutations in motif size, again using values drawn from a gamma distribution (Online Resource 3). For SMM-based simulations, the mean P and locus-specific values were set to 0 since the repeating unit involved in mutation under SMM is always 1. We generated 10^6 simulated data for each scenario by calculating the following summary statistics: mean number of alleles (A) across loci, mean genetic diversity (H_E) across loci, mean size variance (V) across loci, mean Garza-Williamson M index, classification index (LIK), and F_{ST} between samples. Following the results of pre-evaluation of scenario-prior combinations under the two mutation models and subsequent model-checking (see "Results" and Online Resources 5, 6, 7, and 8), we computed the historical demographic parameters based on scenario 4 under the GSM model, with uniform priors, using 10^3 simulated data closest to the observed dataset, after *logit* transformation of parameter values. Measures to assess bias and precision of the analysis and obtained parameters were also calculated.

Results

Genetic diversity and differentiation at nuclear microsatellite loci

Estimates of microsatellite diversity within populations of each species in terms of average AR and H_E were 6.1 and 0.571, respectively, in *Q. aliena* and 5.6 and 0.590, respectively, in *Q. serrata* (Table 1). There were significant deviations of F_{IS} from 0 in one population of *Q. aliena* and five populations of *Q. serrata* (Table 1), but they were not significant ($P > 0.05$) after Bonferroni correction. The 30 nuclear microsatellite loci were polymorphic across populations in both studied species (Online Resource 2). The average number of alleles (A) and gene diversity (H_S) in *Q. aliena* were 9.2 and 0.567, respectively, and 12.0 and 0.591, respectively, in *Q. serrata*. Over all loci, estimates of genetic differentiation among populations (F_{ST} and G'_{ST}) varied among loci in each species (Online Resource 2). F_{ST} values obtained for *Q. aliena* and *Q. serrata* were 0.113 and 0.014, respectively, and G'_{ST} values were 0.369 and 0.055, respectively.

There were small but significant negative correlations between genetic diversity (AR and H_E) in *Q. aliena* and *Q. serrata* populations and their latitude (Fig. 1), and in both species H_E declined slightly less with increases in latitude than AR, as expected due to the weaker sensitivity of H_E to changes in population size in a short time (Maruyama and Fuerst 1985; Cornuet and Luikart 1996).

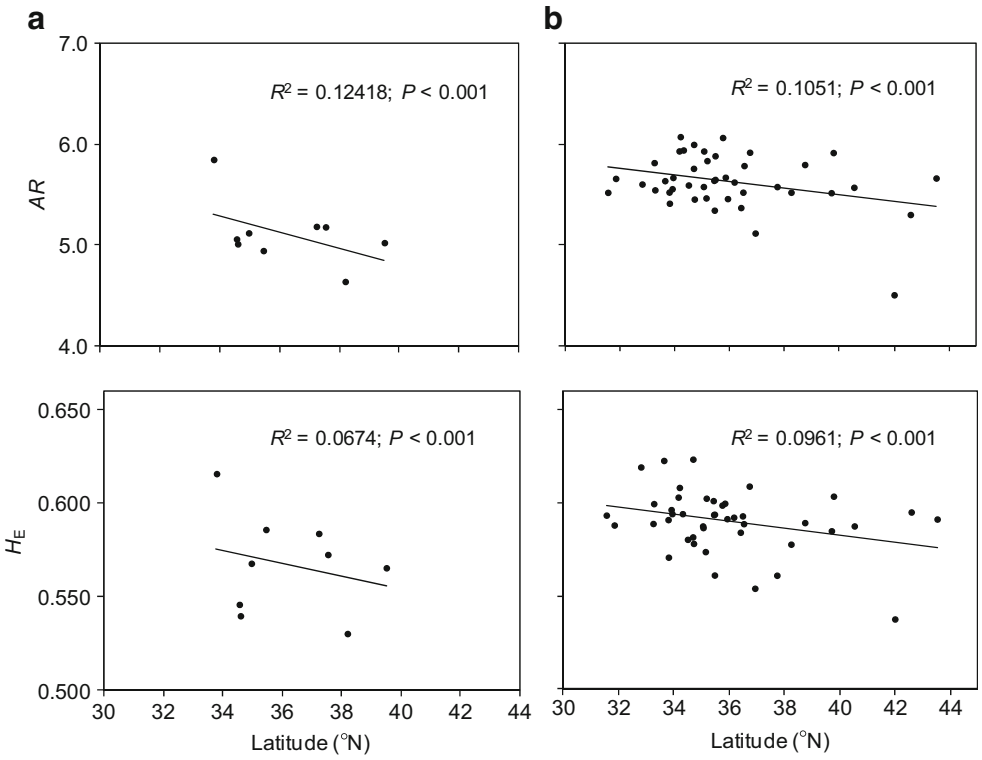

Fig. 1 Relationship between genetic diversity, in terms of AR and H_E, and latitude of populations of **a** *Quercus aliena* and **b** *Q. serrata* based on 30 EST-SSR markers

Evidence of significant recent population bottlenecks was detected in seven populations of *Q. aliena* and 15 populations of *Q. serrata* (Table 1), but associated probabilities were mostly weak and not significant after Bonferroni correction ($P > 0.05$), except for three populations of *Q. aliena* (MYG, SHG, and FKK).

Genetic structure revealed by nuclear microsatellite analysis

Genetic differentiation in *Q. aliena* and *Q. serrata* displayed significant IBD patterns (Fig. 2), but the degree of IBD was stronger in *Q. aliena* (Fig. 2a) than in *Q. serrata* (Fig. 2b). F_{ST} values between the isolated HKD population of *Q. serrata* and the other populations were substantially higher than the values for all other pairs of populations, but excluding HKD from the analysis did not lessen the IBD ($R^2 = 0.1756$, $P < 0.001$). The STRUCTURE analyses of the two species (separately or together) indicated that 2 was the optimal number of ancestral genetic clusters, since ΔK was highest at this value (data not shown). This seems biologically reasonable given the reported differentiation between northern and southern lineages of oaks (Kanno et al. 2004; Okaura et al. 2007; FFPRI 2011). At $K = 2$, there were strong signals of the geographic genetic structure in *Q. aliena*, clearly separating NE and SW populations (Fig. 3B (*a*)). Average F_{ST} values of the genetic clusters representing the NE and SW populations were 0.092 and 0.110, respectively. Further subdivision was observed at $K = 3$, with the SHG population mainly forming a different cluster (Fig. 3B (*b*)), but the NE-SW trend remained strong. In contrast, the genetic structure of *Q. serrata* at $K = 2$ did not reveal distinct clustering of populations in accordance with the predicted differentiation of NE and SW groups. Instead, there was a rather smooth cline in fractions of membership of individuals to each of the two ancestral genetic clusters across the species' distribution (Fig. 3B (*c*)). At $K = 3$, a moderate genetic substructure was only observed in populations in Kyushu (Fig. 3B (*d*)). In addition, the HKD population tended to occupy a separate position in both inferred clusters (at $K = 2$). In accordance with the STRUCTURE analysis, the NJ trees indicated strong genetic relationships among the NW and SE clusters of *Q. aliena* populations (Fig. 4a), and consistently weak relationships within these groups of *Q. serrata* populations (Fig. 4b).

Estimation of historical and demographic parameters

Pre-evaluation of scenario-prior combinations showed that microsatellite datasets of *Q. aliena* similar to observed datasets could be generated with uniform priors, under both

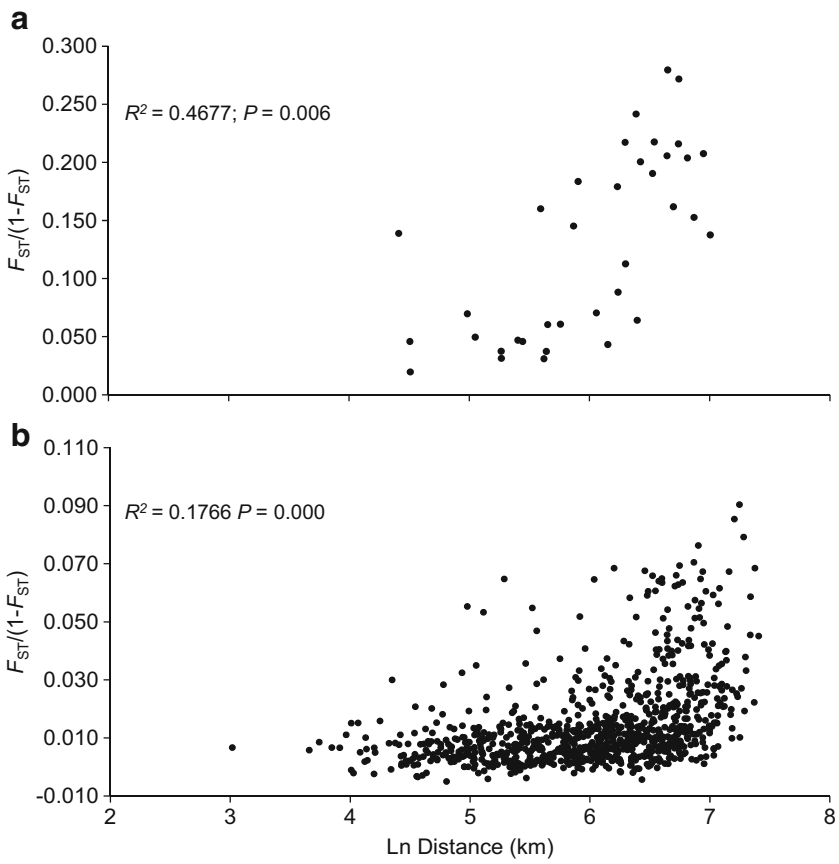

Fig. 2 Relationship of genetic similarity expressed as $F_{ST}/(1 - F_{ST})$ and the natural logarithm of geographic distances between populations of **a** *Q. aliena* and **b** *Q. serrata* based on 30 EST-SSR markers

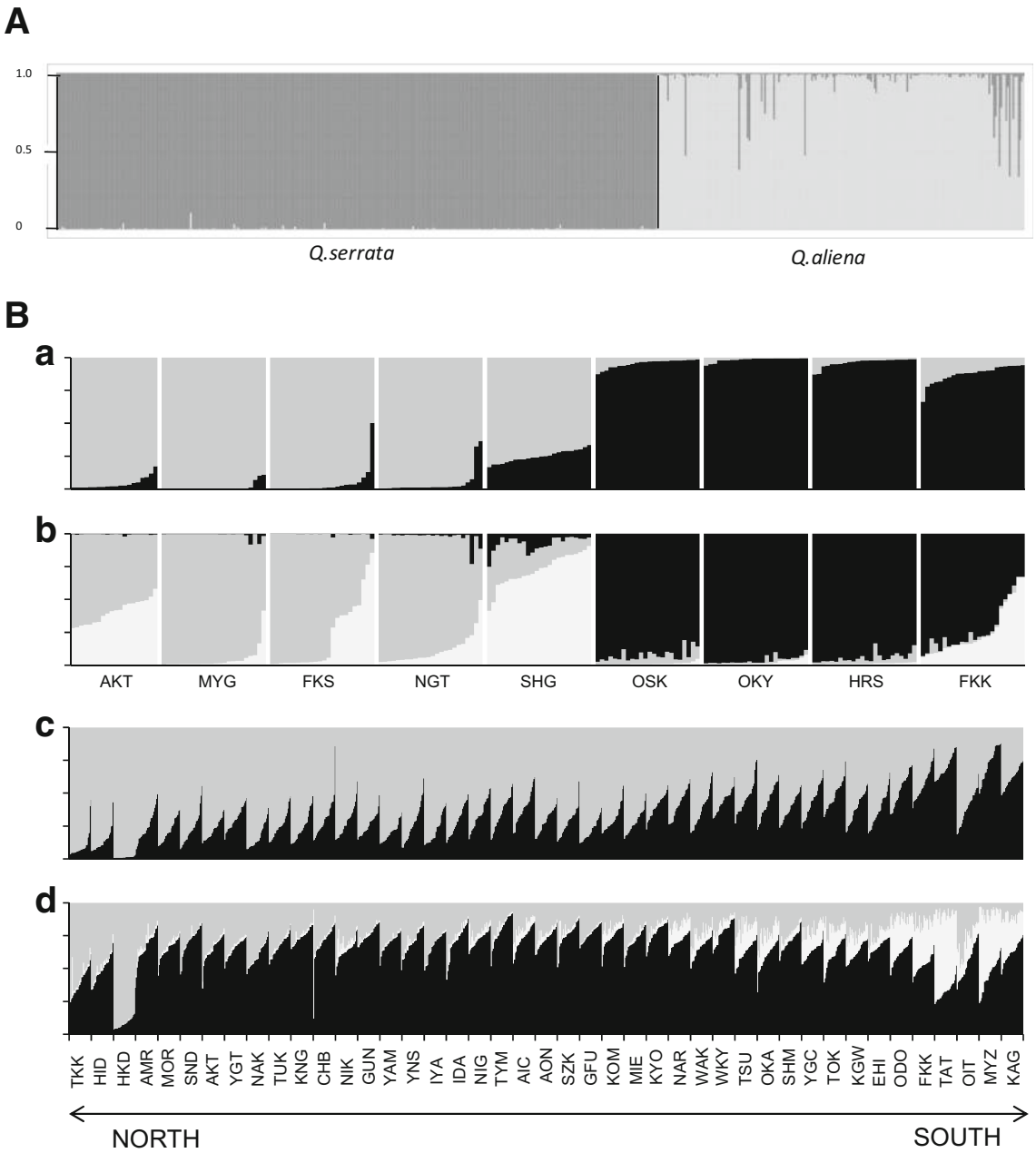

Fig. 3 Results of STRUCTURE analysis based on 30 EST-SSR markers using data for both *Q. aliena* and *Q. serrata* (**A**) and each species separately (**B**). Proportion of membership coefficients of *Q. aliena* (*a*, *b*) and *Q. serrata* (*c*, *d*) individuals for the inferred clusters with $K = 2$, the optimal number of clusters according to ΔK values (*c*, *d*), and $K = 3$ (*b*, *d*). Each vertical bar shows the average obtained from 20 replicates. Population names are indicated in Table 1

GSM and SMM models, for scenarios 3 and 4, but not for scenarios 1 and 2 (Online Resources 6a and 6b). However, scenario 4 was consistently supported with considerably higher posterior probability than scenario 3 under both models (Online Resources 7a and 7b), although confidence in scenario 4 was blurred due to higher frequencies of type I and II errors than under scenario 3 (Online Resource 7b). However, in the model-checking (Online Resource 8a and 8b) only scenario 4 under the GSM model did not have mis-specified summary statistics (Online Resource 8b), indicating a better fit of this scenario. This suggests that the SW and NE populations separated after the predicted bottleneck. Under this scenario, the estimated time of divergence (td) of the nuclear genome of NE and SW populations of the species was 785 (median) generations ago and their divergence is associated with severe bottleneck (Table 2 and Fig. 5). A very similar estimate (median $td = 781$) was obtained after removing the six loci with imperfect repeat motifs, but allele distributions are consistent with a stepwise mutation pattern (data not shown).

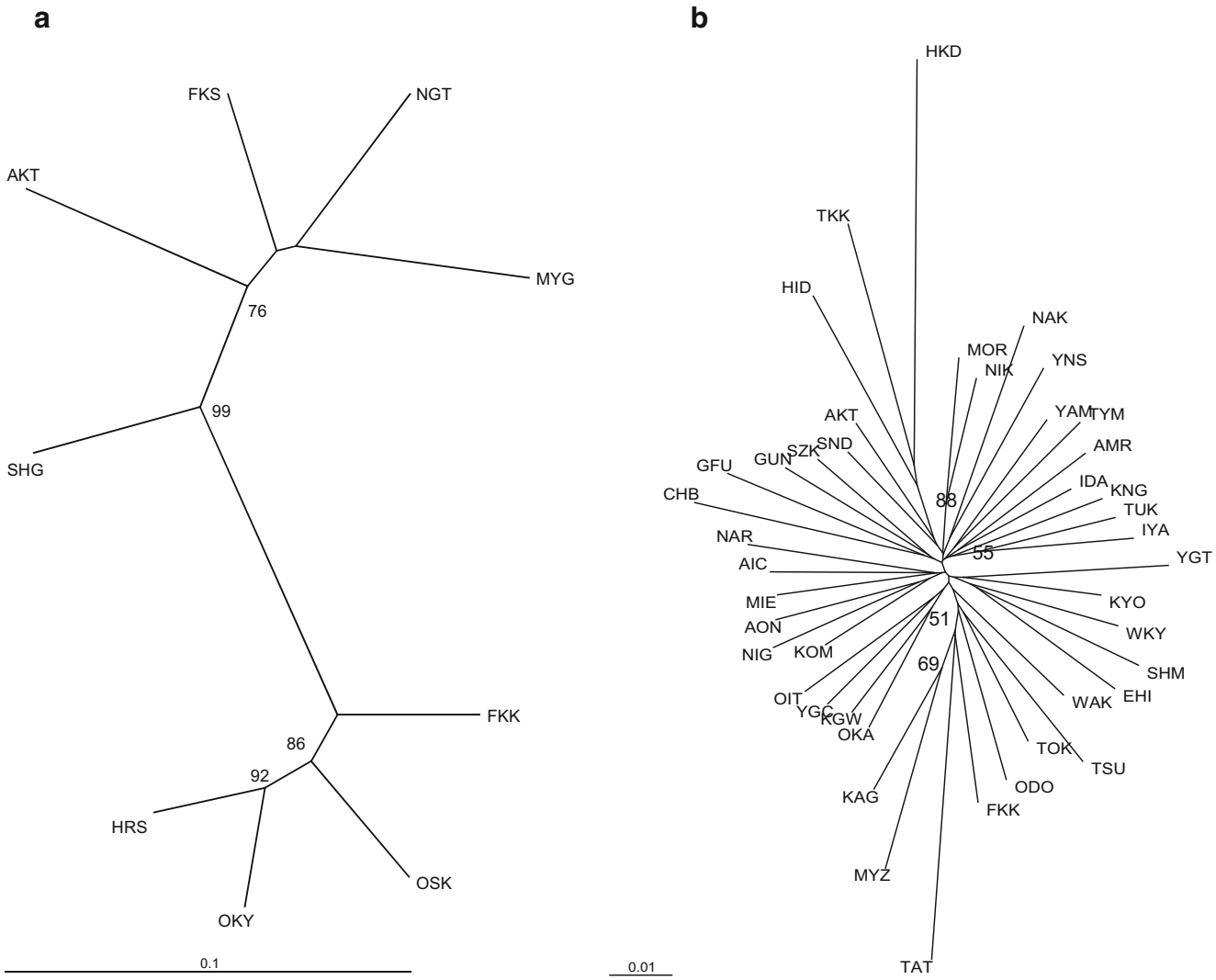

Fig. 4 Neighbor-joining tree based on D_A distances among populations of **a** *Q. aliena* and **b** *Q. serrata* using 30 EST-SSR markers. Nodes supported by greater than 50% bootstrap values from 1000 replicates are shown

Genetic variation and differentiation of cpDNA

The combined length of the four examined cpDNA regions in the two species was 3340 bp, with a total of 15 nucleotide substitution sites (Online Resource 3). The *rp*L32-*trn*L region was the most polymorphic, with nine substitution sites. Nucleotide diversity (π) of the four regions was 0.00068 in *Q. aliena* and 0.00075 in *Q. serrata*. The sequences have been deposited in the DNA Data Bank of Japan (Online Resource 4).

In total, 20 haplotypes were detected in the two species (Fig. 5d, e): eight in *Q. aliena*, 19 in *Q. serrata*, and 7 common to both species. In *Q. aliena*, haplotype 2 was the most common (42%) and the dominant haplotype among northeastern populations, while the least common (1%) was haplotype 20, a private haplotype of the MIE population and the only *Q. aliena* haplotype not found in *Q. serrata* (Fig. 5d and Online Resource 3). Similarly, haplotype 2 was the most common (29.1%) haplotype of *Q. serrata* and the dominant haplotype among northeastern populations of the species, while haplotypes 10 and 19 were the least common (0.3%) and only detected in southwest Japan. Eleven haplotypes were private to specific *Q. serrata* populations, but one of them (haplotype 15) was also present (and more frequent) in *Q. aliena*. Within-population haplotype diversities of *Q. aliena* and *Q. serrata* ranged from 0.000 to 0.833 and 0.000 to 0.821, respectively (Table 1). However, geographic locations of haplotypes in both species were not clearly reflected in the haplotype networks (Fig. 5f, g). Average values for diversity within populations (H_S) and in the total population (H_T) of *Q. aliena* were 0.171 and 0.816, respectively, and the relative genetic differentiation (G_{ST}) was 0.791. Corresponding estimates for *Q. serrata* were 0.213, 0.849, and 0.750, respectively.

Table 2 Estimates of demographic parameters based on 30 EST-SSR markers obtained using coalescence simulation based on the most probable evolutionary scenario (scenario 4) underlying the divergence of the nuclear genome of NE and SW populations of *Q. aliena* under the GSM model, with uniform prior distribution of parameters

Parameter	Point estimate		
Original	Median	$Q_{0.05}$	$Q_{0.95}$
N1	8100	4730	9840
N2	5950	3410	8920
N1bk	789	190	2570
N2bk	615	173	1680
Na	1580	613	3880
NA	4940	1820	9300
tbk	525	116	882
td1	785	413	981
t	6540	1880	9690

Priors and relationships of parameters are indicated and illustrated in Online Resource 4 and Fig. 6

N1 effective population size of present SW population, *N2* effective population size of present NE populations, *N1bk* bottlenecked condition of N1, *N2bk* bottlenecked condition of N2, *Na* bottlenecked condition of ancestral population before splitting into SW and NE populations, *NA* ancestral population (see Fig. 6 for assumptions), *tbk* time of bottleneck after split of populations, *td* time of divergence, *t* assumed as pre-LGM period

$Q_{0.05}$ 5% quantile value of the posterior distribution, $Q_{0.95}$ 95% quantile value of the posterior distribution

Discussion

Genetic diversity and structure of *Q. aliena* and *Q. serrata*

Previous observations on *Q. crispula* in Japan have detected significant, but moderate, latitudinal reductions in allelic diversity of populations from south to north (Ohsawa et al. 2011; A. Matsumoto et al. in preparation). We observed the same geographical cline in *Q. aliena* and *Q. serrata* in this study (Fig. 2), but it appears to be weaker than in *Q. crispula*. However, the clines are considerably weaker in all three of these oaks than in several other deciduous broadleaf species examined, such as *F. crenata* (Hiraoka and Tomaru 2009a), which co-occurs with oaks in cool temperate forests. This clinal reduction of within-population genetic diversity in species has been associated with the northward expansion of deciduous forests in Japan following the LGM. However, not much substantial evidence of recent bottlenecks was detected either in this study or the cited study on *Q. crispula.* The weak clinal reduction in *AR* towards the north may be related to the smaller spatio-temporal distributions of refugial populations of oaks during the LGM than those of other taxa, including *F. crenata*, which reportedly had multiple refugia along both the Pacific Ocean and Japan Sea Side of Japan during the LGM (Tsukada 1982a, b; Hiraoka and Tomaru 2009a). However, it is surprising that the loss of allelic diversity in *Q. aliena* was not more pronounced, given the strong fragmentation of populations, especially in northeastern Japan. This may reflect a small difference in effective population sizes of "refugial" and "founding" populations of *Q. aliena* during the LGM. In accordance with this hypothesis, our coalescence simulations indicate that the effective sizes of the assumed "founder populations" of NE and SW populations, N1bk and N2bk, respectively, were of similar magnitude and both probably recovered at considerable rates after their split. This is evident from the large changes from N1bk (founder of *Q. serrata*) to N1 (present population), and from N2bk (founder of *Q. aliena*) to N2 (present population) (Table 2). However, the prior-posterior distribution for N1 generally overlapped and skewed towards the upper boundary (data not shown), suggesting that our observed data lacked sufficient information to estimate this parameter as precisely as other parameters.

The sharp genetic differentiation between NE and SW populations of *Q. aliena* at nuclear microsatellites and the cpDNA distribution observed in the present study is generally consistent with findings of previous phylogeographic studies on Japanese oaks, namely, *Quercus crispula* and *Q. serrata* (Kanno et al. 2004; Okaura et al. 2007). This differentiation is also congruent with the reported geographic structure of variation in the length and epidermis of the NE and SW populations' acorns, which are characterized as "large and glabrous type" and "small and hairy type," respectively (Kanno and Suzuki 2005). In addition, acorn characteristics are reportedly intermediate in populations of Shiga Prefecture, along the boundary of the two groups, which also have moderate admixture of the two ancestral genetic clusters of nuclear DNA. Similarly, geographic patterns of variations at neutral markers and both morphological and chemical properties are congruent in populations of several other Japanese tree species, e.g., *Cryptomeria japonica* and *F. crenata* (Yasue et al. 1987; Tsumura et al. 2014; Hagiwara 1977; Tomaru et al. 1997, 1998). However, an intra-specific sub-division of *Q. aliena* into two varieties (vars. *aliena* and *pellucida*) by Ohba (2006), based on density of stellate hairs on abaxial leaf surfaces does not follow the geographic pattern (Kanno and Suzuki 2005). Therefore, the discrepancy between genetic backgrounds of populations revealed by this study and this morphological variation challenges the suitability of the current intra-specific classification of *Q. aliena*.

In contrast to the consistent patterns in *Q. aliena*, geographic patterns of variations at nuclear and chloroplast markers were strikingly inconsistent in *Q. serrata*. However, the geographic distribution of its cpDNA haplotypes displays a strong north-south trend and is very similar to the pattern reported by Okaura et al. (2007), including the suggested boundary between NE and SW clusters along the Itoigawa-Shizuoka

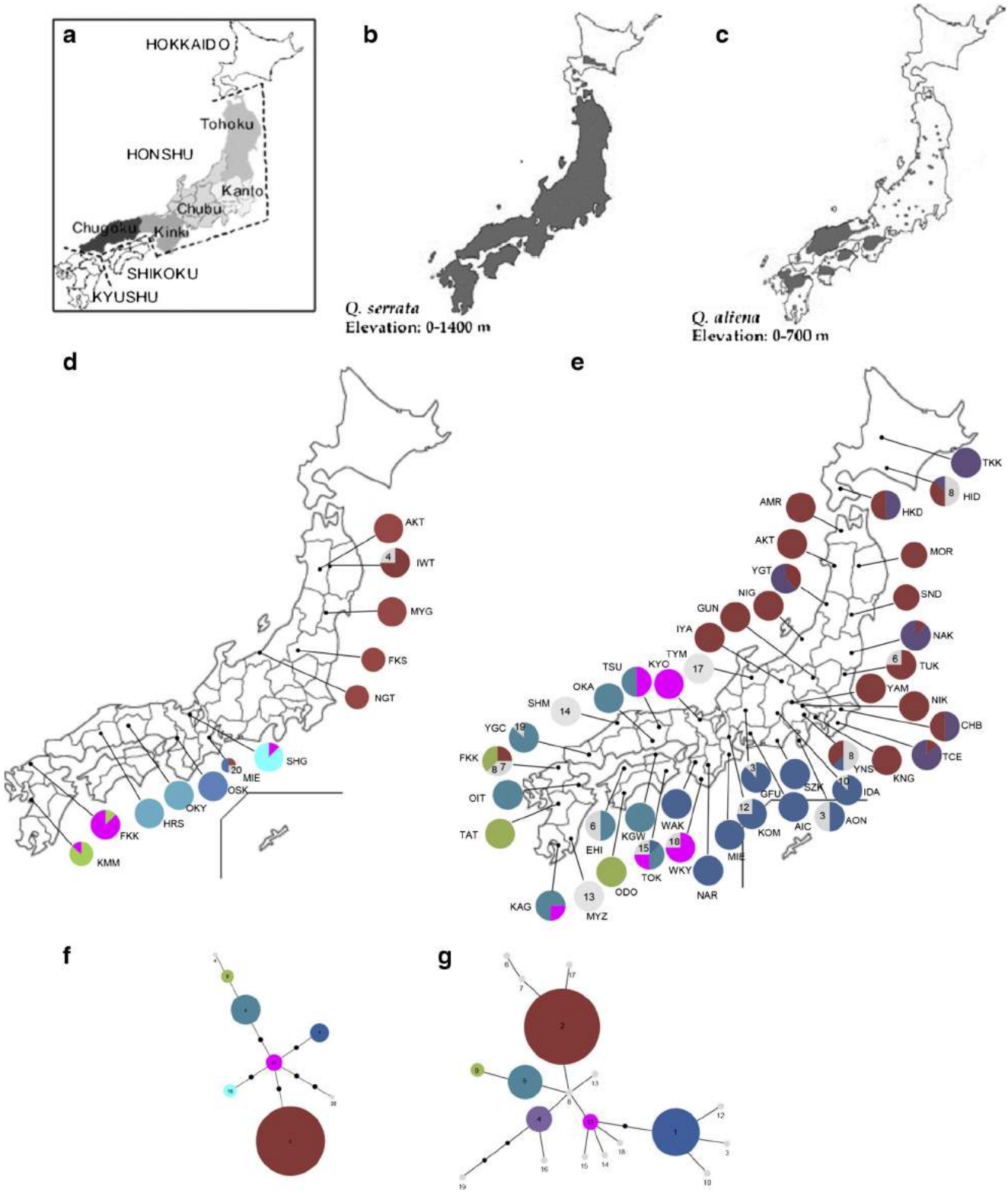

Fig. 5 Map of Japan **a** showing the four main islands (Hokkaido, Honshu, Shikoku, and Kyushu) and five regions (Tohoku, Kanto, Chubu, Kinki, and Chugoku). The ranges of natural distribution of *Quercus aliena* and *Q. serrata* are shown as gray areas on **b**, **c**, respectively. The geographic distributions of chloroplast haplotypes detected in **d** *Q. aliena* and **e** *Q. serrata*. The sizes of pie graphs are proportional to the number of analyzed individuals in each population as indicated in Table 1. Minimum spanning networks for chloroplast haplotypes in *Q. aliena* and *Q. serrata* are shown in **f**, **g**, respectively. The sizes of circles are proportional to the frequency of observed haplotypes in each species, except for gray-shaded circles (less than 5%). Small black circles indicate vectors for missing haplotypes

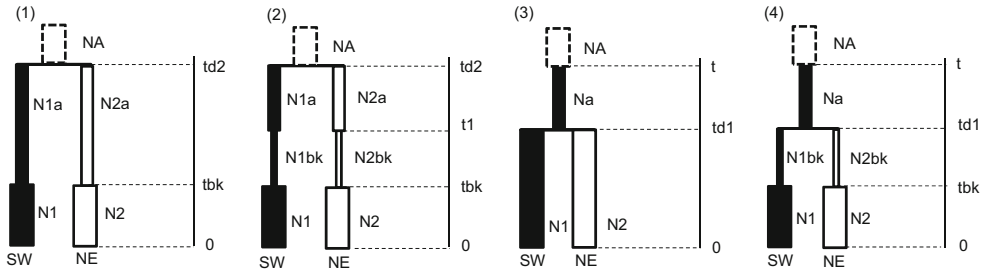

Fig. 6 Tested evolutionary models underlying the divergence of NE and SW populations of *Q. aliena* based on 30 EST-SSR markers. Scenarios 1 and 2 assume that populations split before the predicted bottleneck, while scenarios 3 and 4 assume the split occurred after the predicted bottleneck. Scenarios 2 and 4 differ from 1 and 3, respectively, by further assuming that each of the populations experienced a bottleneck following the LGM (as explained in the "Materials and methods" section). Parameters are defined in Table 2 and their relationships, including prior distributions, are indicated in Online Resource 7a and 7b. NA is bounded by broken lines to represent an unknown relationship with the present populations

Tectonic Line (ISTL). According to a previous study, a barrier created by volcanic activity of a mountain chain in the middle of the western side of the Fossa Magna (a back-arc rift basin of Tertiary sediments, which mirrors the ISTL) caused the initial separation of populations currently located at each side of the ISTL (reviewed in Ohsawa and Ide 2011). Effects of geographic barriers on patterns and discontinuities of genetic structure or haplotypes have also been documented in other tree species (e.g., Ferris et al. 1993; Rossetto et al. 2007). Recent evidence indicates that genomes of very few Japanese plant species, except *F. japonica* (Hiraoka and Tomaru 2009a) and *Quercus* spp., are differentiated across the ISTL. However, Ohsawa and Ide (2011) suggest that differentiation or phylogeographic patterns created by physical barriers formed by factors such as the ISTL tend to disappear rapidly in evolutionary terms, especially at nuclear loci.

Inconsistencies in geographic patterns of genetic structure of nuclear genomes have also been reported in several related and co-distributed species (e.g., *Elaeocarpus largiflorens* and *E. angustifolia*, Rossetto et al. 2007; *F. crenata* and *F. japonica*, Hiraoka and Tomaru 2009a, b). The stark contrast in genetic structures of nuclear genomes between *Q. aliena* and *Q. serrata* likely stemmed from varying distributions of their populations. Strong fragmentation of *Q. aliena*, especially in northern Japan, may have been the main cause of the stronger differentiation of its nuclear genome. Theoretical expectations regarding effects of fragmentation include reductions in genetic diversity and increases in inter-population divergence (Ohsawa et al. 2006), both of which were observed in *Q. aliena* relative to *Q. serrata*. Genetic differentiation (G'_{ST}) of the nuclear genome is considerably higher in *Q. aliena* (0.369) than in *Q. serrata* (0.055) and other widely distributed deciduous broadleaf species in Japan, such as *F. crenata* (based on F_{ST} = 0.027, an analogous of G_{ST}; Hiraoka and Tomaru 2009a) and *Q. crispula* (G'_{ST} = 0.090; Ohsawa et al. 2011). IBD patterns are also more pronounced in *Q. aliena* than in *Q. serrata*, indicating that genetic drift could be strongly relative to gene flow in *Q. aliena*. Lower levels of IBD in *Q. serrata* are most likely due to the greater density of mother trees associated with its broader distribution and dominance in secondary forests, especially over *Q. aliena*, which can be attributed to its wider ranges of light and moisture tolerance (Lee and You 2012).

Inferences on phylogeographic structure and historical demography of *Q. aliena* and *Q. serrata*

Phylogeographic studies on *Quercus* species based on cpDNA as earlier mentioned (Kanno et al. 2004; Okaura et al. 2007) revealed two distinct lineages (NE and SW) for these oaks and that they have contrasting evolutionary histories, which is suggested to have been separated into northeast and southwest regions by major mountain chains of central Japan. According to Okaura et al. (2007), local populations of the four related oak species (section *Prinus*) mentioned in this study often share cpDNA haplotypes by chloroplast capture or introgression through natural hybridization among the related oak species. This observation is very clear in our two studied oaks, with almost completely shared cpDNA haplotypes. Even among the genus *Quercus*, with more than 300 species, natural hybridization and introgression between closely related species have been described but morphological and ecological boundaries between species are clear (e.g., Whittemore and Schaal 1991; Dumolin-Lapegue et al. 1997; Ferris et al. 1998; Belahbib et al. 2001 reviewed in Okaura et al. 2007). Thus, oaks are suggested to be "large sets of broadly sympatric species exchanging genes" (Van Valen 1976 as reviewed in Kremer and Petit 1993) but manages to remain morphologically distinct. These are likely the same mechanisms shaping the structure of cpDNA of the four mentioned native oaks. Molecular markers such as AFLP (Matsumoto et al. 2009) had identified the distinct genetic differences of these species, although hybrid individuals had been detected. In our separate study using a fairly similar set of nuclear microsatellites as used in this study, we observed a very clear genetic differentiation among these species and found that *Q. aliena* and *Q. serrata* are more closely related than

any of the two other native oaks (unpublished data but preparation for publication is underway).

However, Kanno et al. (2004) postulated that the NE populations were established by northward expansion of SW populations through a refugium in Kanto region following the LGM. In contrast, Okaura et al. (2007) concluded that the haplotypes in northeast Japan probably originated from populations in the Eurasian continent that migrated through a northern land bridge connecting northern Hokkaido. The haplotype network in the present study is consistent with establishment of the northern populations from a refugium around central Honshu. More specifically, the YNS population of *Q. serrata*, containing haplotype 8, from which most haplotypes of NE and SW populations likely originated, could be in part of that refugium. However, the presence of haplotype 2 in the MIE population of *Q. aliena*, which originated from the inferred ancestral haplotype 11 (Fig. 5a), suggests that this haplotype did not evolve independently from southern populations. This haplotype could be rare in southern parts of Japan and have established in northern Japan through a small number of founding individuals.

In forest tree species, the generation time is always a big question in natural forest (Petit and Hampe 2006). Some consider generation as the time of the start of first flowering (or first reproduction), while others use a range of values based on observed differences in maturity and reproduction of individuals of a species. For example, Okumura and Oki (1992) reported a range of minimum age of reproductive maturity of several oaks, varying from 20 to 40 years. Most of the studies used their mutation rate to estimate the generation time (Cavender-Bares et al. 2011) and sometimes used a specific formula (Yang et al. 2016) or ecological observation data (Tsuda et al. 2015). There is still no concrete method to estimate the generation time in natural forest trees. However, even in the same genus, several generation times were reported for oaks, which varied from 30 to 220 years. Bognoli et al. (2016) used 30 and 50 years for the generation time in Mediterranean oak; Ortego et al. (2015) used 50 years for scrub oak; Yang et al. (2016) hired 80 years for Chinese oaks; and Cavender-Bares et al. (2011) estimated the generation time for oaks (*Quercus* series *Virentes*) between 120 and 220 years. However, less than 40 years based on Okumura and Oki's reported minimum age of reproductive maturity for oaks for generation time is too short to use in our case. Thus, adapting the range of 50–220 years, for example, suggests that the divergence time (or td based on the median value of 785) of *Q. aliena* would be a number of times earlier than the presumed period of northward expansion of broadleaved trees in Japan following the LGM. This is if the equivalent of one generation is 50 to 220 calendar years from the present (or around 39,250 to 173,000 calendar years). Since generation time of studied oaks and related species has not been clearly estimated yet, there is one likely scenario supported by this study that *Q. aliena*, which might be true for the related oaks, established into SW and NE populations following sequential historical population bottlenecks.

Implications of geographic genetic structure of *Q. aliena* and *Q. serrata* for seed and seedling transfer

Although molecular markers generally reflect neutral variation, molecular marker-based assessment of genetic variation is reasonable for developing practical genetic guidelines for designating seed transfer zones for conservation and restoration activities, since methods for detecting locally adaptive traits, such as reciprocal transplant experiments, generally require considerable time and labor (Hufford and Mazer 2003). As early as the 1990s when isozymes and fragment length polymorphisms such as amplified fragment length polymorphism and restriction fragment length polymorphism (AFLP and RFLP) were the main genetic markers in analysis of population genetic structure, polymorphisms of these markers have already been suggested to delineate seed transfer zones for a given species (e.g., *Fagus sylvatica* in Czech Republic (Gomory et al. 1998); *Acacia rostellifera* and *A. cochlearis* in southwest Australia (Krauss and He 2006) and several other plant species in Australia mentioned in the paper; *Avenella flexuosa* and *Agrostis mertensii* in Norway (Jørgensen et al. 2016)). To date, with the advancement of genetic tools to elucidate even the most cryptic genetic structure of populations, a number of studies in many countries have strongly supported the power and practicality of molecular markers for seed zoning or delineating local provenances, with some of them already adapted in the current strategies for seed collection for restoration or conservation activities for the species.

In accordance with the FFPRI criterion for transferring species with strong and consistent genetic structure at two examined genomes, as described in the "Introduction" section, we show that the NE and SW populations of *Q. aliena* represent well-differentiated "gene pools" in both nuclear and cpDNA genomes; thus, the areas they cover should represent valid transfer zones for the species. In 2015, Tsumura and Suyama (2015) has published a book (in Japanese) that collates geographic genetic structures of various forest tree species in Japan for establishing guidelines for seed and seedling transfer, including broadleaf leaf, pines, and larch species in Japan. Our proposal on NE-SW seed transfer zones for the studied oaks as earlier reported in brief in Tsumura and Suyama (2015) is similar to the proposed genetic guidelines for transferring planting stocks of the closely related oak *Q. crispula* and several other deciduous broadleaf species as published in the said book. Although the geographic genetic structure of cpDNA and nuclear microsatellites of *Q. serrata* are inconsistent, we suggest that the strong NE-SW

geographic structure of cpDNA should be considered, in line with the FFPRI criteria for species with such inconsistency. A previous study reported that variation in acorn traits of *Q. serrata* tends to correlate with changes in geographic variables and environmental conditions (e.g., Iwabuchi et al. 2006). This may help (but not necessarily) to restrict the transfer of planting stocks in congruence with the cpDNA structure. Our proposal, however, does not in anyway discredit the use of climate, vegetation, topography, edaphic factors, and many other environmental variables as criteria for delineating seed zones in a conventional strategy. In the absence of current guidelines for native oaks in Japan, the proposed guidelines will be useful as logical framework for restoration of these ecologically important species.

Acknowledgements The authors would like to thank H. Nishikawa, K. Kitamura, H. Koyama, S. Kanetani, M. Saito, M. Aizawa, K. Ishida, Y. Tsuda, T. Kawahara, A. Tamura, T. Nagamitsu, and S. Fukushima for sampling of materials, and Y. Taguchi for assistance with laboratory work.

Funding information This research was supported by a grant for research on Genetic Guidelines for Restoration Programs using Genetic Diversity Information (2005–2009) from the Ministry of Environment, Japan.

Compliance with ethical standards

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

Data archiving statement The sequences were deposited in the DNA Data Bank of Japan. A full list of accession numbers for haplotypes of individuals is included in the supplementary information (Online Resource 3).

References

Allendorf FW, Leary RF (1986) Heterozygosity and fitness in natural populations of animals. In: Soulé M (ed) Conservation Biology. The science of scarcity and diversity. Sinauer Assoc., Sunderland, pp 57–76

Arbogast BS, Kenagy GJ (2001) Comparative phylogeography as an integrative approach to historical biogeography. J Biogeogr 28: 819–825

Avise JC (2000) Phylogeography: the history and formation of species. Harvard University Press, Cambridge

Belahbib N, Pemonge M-H, Ouassou A, Sbay H, Kremer A, Petit RJ (2001) Frequent cytoplasmic exchanges between oak species that are not closely related: and in Morocco. Mol Ecol 10(8):2003–2012

Bognoli F, Tsuda Y, Fineschi S, Bruschi P, Magri D, Zhelev P, Paule L, Simeone MC, Gonzalez-Martinez SC, Vendramin GG (2016) Combining molecular and fossil data to infer demographic history of *Quercus cerris*: insights on European eastern glacial refugia. J Biogeogr 43:679–690

Campbell RK (1979) Genecology of Douglas-fir in watershed in the Oregon cascades. Ecology 60:1036–1050

Cavender-Bares J, Gonzalez_Rodriguez A, Pahlich A, Koehler K, Nicholas D (2011) Phylogeography and climate niche evolution in live oaks (Quercus series Virentes) from the tropics to the temperate zone. J Biogeogr 38:962–981

Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics 144:2001–2014

Cornuet JM, Ravigné V, Estoup A (2010) Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). BMC Bioinformatics 11:401

Cornille A, Gladieux P, Smulders MJM, Roldán-Ruiz I, Laurens F, Le Cam B, Nersesyan A, Clavel J, Olonova M, Feugey L, Gabrielyan I, Zhang X-G, Tenaillon MI, Giraud T, Mauricio R (2012) New insight into the history of domesticated apple: secondary contribution of the European wild apple to the genome of cultivated varieties. PLoS Genet 8(5):e1002703

Cornuet J-M, Santos F, Beaumont MA, Robert CP, Marin J-M, Balding DJ, Guillemaud T, Estoup A (2008) Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. Bioinformatics 24:2713–2719

De Souza RF, Ziroldo BD, Rosetto EF, Cavalheiro AL, Torezan JMD, Vanzela ALL (2012) The use of genetic structure as a guide for seed gathering for forest restoration. Braz J Biosci 10:309–313

Dumolin-Lapegue S, Demesure B, Le Corre V, Fineschi S, Petit RJ (1997) Phylogeographic structure of white oaks throughout the European continent. Genetics 146:1475–1487

Earl DA, vonHoldt BM (2011) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Res. https://doi.org/10.1007/s12686-011-9548-7

El Moussadik A, Petit RJ (1996) High level of genetic differentiation for allelic richness among populations of the argan tree [(*Arganiaspinosa* (L.) Skeels)] endemic to Morocco. Theor Appl Genet 92:832–839

Ersts PJ (2006) [Internet] Geographic Distance Matrix Generator (version 1.2.3). American Museum of Natural History, Center for Biodiversity and Conservation, New York Available from http://biodiversityinformatics.amnh.org/open_source/gdmg.Accessed on 2012-11-22

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620

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

Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evol Bioinformatics Online 1:47–50

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

Ferris C, Oliver RP, Davy AJ, Hewitt GM (1993) Native oak chloroplasts reveal an ancient divide across Europe. Mol Ecol 2:337–343

Ferris C, King RA, Väinölä R, Hewitt GM (1998) Chloroplast DNA recognizes three refugial sources of European oaks and suggests independent eastern and western immigrations to Finland. Heredity 80(5):584–593

FFPRI (2011) Seedling transfer guideline of Japanese broadleaf tree species based on genetic data. ISBN:978-4-902606-75-1, https://www.ffpri.affrc.go.jp/pubs/chukiseika/documents/2nd-chukiseika20.pdf

Gomory D, Hynek V, Paule L (1998) Delineation of seed zones for European beech (*Fagus sylvatica* L.) in the Czech Republic based on isozyme gene markers. Ann Sci For 55:425–436

Goudet J (2002) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3.2). Available from: http://www2.unil.ch/popgen/softwares/fstat.htmGoudet et al. (2006)

Goudet J, Raymond M, de Meeüs T, Rousset F (1996) Testing differentiation in diploid populations. Genetics 144:1933–1940

Guicking D, Fiala B, Blattner F, Slik F, Mohamed M, Weising K (2011) Comparative chloroplast phylogeography of two tropical pioneer trees, *Macaranga gigantean* and *Macaranga pearsonii* (Euphorbiaceae). Tree Genet Genomes 7:573–585

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

Hamann A, Gylander T, Chen P (2011) Developing seed zones and transfer guidelines with multivariate regression trees. Tree Genet Genomes 7:399–408

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

Hillis DM, Moritz C (1990) Molecular taxonomy. Sinauer associates, Inc. Publishers, Massachusetts

Hiraoka K, Tomaru N (2009a) 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

Hiraoka K, Tomaru N (2009b) Genetic structure of *Fagus japonica* revealed by nuclear microsatellite markers. Int J Plant Sci 170: 748–758

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

Hufford KM, Mazer SJ (2003) Plant ecotypes: genetic differentiation in the age of ecological restoration. Trends Ecol Evol 18:147–155

Iwabuchi Y, Hoshino Y, Hukusima T (2006) Intraspecific variation of acorn traits of *Quercus serrata* Thunb. In Kanto region, central Japan. Veg Sci 23:81–88

Iwasaki T, Aoki K, Seo A, Murakami N (2012) Comparative phylogeography of four component species of deciduous broad-leaved forests in Japan based on chloroplast DNA variation. J Plant Res 125:207–221

Jørgensen MH, Elameen A, Hofman N, Klemsdal S, Malaval S, Fjellheim S (2016) What's the meaning of local? Using molecular markers to define seed transfer zones for ecological restoration in Norway. Evol Appl 9:673–684. https://doi.org/10.1111/eva.12378

Kanno M, Suzuki M (2005) Intraspecific variation of acorn morphology within *Quercus aliena* Blume in Japan and South Korea. Proceedings of conference OAK 2003 'Integration of Silviculture and Genetics in Creating and Sustaining of Oak Forests' held on 29 Sept. - 3 Oct. 2003 in Tsukuba and Nikko, Japan 153–158

Kanno M, Yokoyama J, Suyama Y, Ohyama M, Itoh T, Suzuki M (2004) Geographical distribution of two haplotypes of chloroplast DNA in four oak species (*Quercus*) in Japan. J Plant Res 117:311–317

Keir KR, Bemmels JB, Aitken SN (2011) Low genetic diversity, moderate local adaptation, and phylogeographic insights in *Cornus nuttallii* (Cornaceae). Am J Bot 98:1327–1336

Kitamura S, Horikawa (1951) On *Quercus* subgen. Lepidobalanus of Japan, Korea, and North China. Mem Coll Sci Univ Tokyo Ser B 20:20–25

Krauss SL, He TH (2006) Rapid genetic identification of local provenance seed collection zones for ecological restoration and biodiversity conservation. J Nat Conserv 14:190–199

Krauss SL, Koch JM (2004) Rapid genetic delineation of provenance for plant community restoration. J Appl Ecol 41:1162–1173

Kremer A, Petit RJ (1993) Gene diversity in natural populations of oak species. Ann Sci For 50(Suppl 1):186s–202s

Langella O (1999) Population 1.2.30 http://bioinformatics.org/~tryphon/populati-ns/

Lee SH, You YH (2012) Measurement of ecological niche of *Quercus aliena* and *Q. serrata* under environmental factors treatments and its meaning to ecological distribution. J Ecol Field Biol 35:227–234

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452

Luikart G, Allendorf FW, Cornuet JM, Sherwin WB (1998) Distortion of allele frequency distributions provides a test for recent population bottlenecks. J Hered 89:238–248

Maliouchenko O, Palmé AE, Buonamici A, Vendramin GG, Lascoux M (2007) Comparative phylogeography and population structure of European *Betula* species, with particular focus on *B. pendula* and *B. pubescens*. J Biogeogr 34:1601–1610

Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Res 27:209–220

Maruyama T, Fuerst TA (1985) Population bottlenecks and non-equilibrium models in population genetics. II. Number of alleles in a small population that was formed by a recent bottleneck. Genetics 111:675–689

Matsumoto A, Kawahara T, Kanazashi A, Yoshimaru H, Takahashi M, Tsumura Y (2009) Differentiation of three closely related Japanese oak species and detection of interspecific hybrids using AFLP markers. Botany 87:145–153

McKay JK, Christian E, Harisson S, Rice KJ (2005) "How local is local?" A review of practical and conceptual issues in the genetics of restoration. Restor Ecol 13:432–440

Moncada KM, Ehlke NJ, Muehlbauer GJ, Sheaffer CC, Wysw DL, DeHaan LR (2007) Genetic variation in three native plant species across the state of Minnesota. Crop Sci 47:2379–2389

Moritz C (1995) Uses of molecular phylogenies for conservation. Phil Trans R Soc Lond B349:113–118

Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci U S A 70:3321–3323

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. Evolution 29:1–10

Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. J Mol Evol 19:153–170

Nozaki R, Kurohara A, Kamei H (2001) A preliminary study on the *Quercus aliena* forest: a type of natural forest vegetation on the alluvial plain in Japan. Papers in commemoration of Prof. Dr. Shigetoshi Okuda's retirement: studies on vegetation of alluvial plains. 23–32 (in Japanese with English abstract)

Ohba H (2006) Fagaceae. In: Iwatsuki K, Boufford D, Ohba H (eds) Flora of Japan, volume IIa. Kodansha Ltd., Japan, pp 42–60

Ohsawa T, Ide Y (2011) Phylogeographic patterns of highland and lowland plant species in Japan. Alp Bot 121:49–61

Ohsawa T, Tsuda Y, Sawada H, Ide Y (2006) Genetic diversity and gene flow of *Quercus crispula* in a semi-fragmented forest together with neighboring forests. Silvae Genet 55:160–169\

Ohsawa T, Tsuda Y, Saito Y, Ide Y (2011) The genetic structure of *Quercus crispula* in northern Japan as revealed by nuclear simple sequence repeat loci. J Plant Res 124:645–654

Okaura T, Quang ND, Ubukata M, Harada K (2007) Phylogeographic structure and late Quaternary population history of the Japanese oak *Quercus mongolica* var. *crispula* and related species revealed by chloroplast DNA variation. Genes Genet Syst 82:465–477

Okumura T, Oki M (1992) Survey on flowering and fruiting of useful broadleaf tree species in deciduous broadleaf tree forest zone. Bulletin of Nagano Prefecture. For Res 6:1–15

Ortego J, Noguerales V, Gugger PF, Sork VL (2015) Evolutionary and demographic history of the Californian scrub white oak species complex: an integrative approach. Mol Ecol 24(24):6188–6208

Petit RJ, Hampe A (2006) Some evolutional consequences of being a tree. Ann Rev Ecol Evol Syst 37:187–214

Petit RJ, Kremer A, Wagner DB (1993) Geographic structure of chloroplast DNA polymorphisms in European oaks. Theor Appl Genet 87: 122–128

Petit RJ, El Mousadik A, Pons O (1998) Identifying populations for conservation on the basis of genetic markers. Conserv Biol 4:844–855

Piry S, Luikart G, Cornuet JM (1999) BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. J Hered 90:502–503

Pons O, Petit RJ (1995) Estimation, variance and optimal sampling of gene diversity. I: haploid locus. Theor Appl Genet 90:462–470

Prim RC (1957) Shortest connection networks and some generalizations. Bell Syst Tech J 36:1389–1401

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

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J Hered 86: 248–249

Rice WR (1989) Analyzing tables of statistical tests. Evolution 43:223–225

Rossetto M, Crayn D, Ford A, Ridgeway P, Rymer P (2007) The comparative study of range-wide genetic structure across related co-distributed rainforest trees reveals contrasting evolutionary histories. Aust J Bot 55:416–424

Rousset F (1997) Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. Genetics 145:1219–1228

Rousset F (2008) Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. Mol Ecol Res 8:103–106

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing the phylogenetic trees. Mol Biol Evol 4:406–425

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. Am J Bot 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. Am J Bot 94:275–288

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

Tani A, Kawawata (2008) Isoprene emission from the major native *Quercus* spp. in Japan. Atmos Environ 42:4540–4550

Taoda H (2005) Integration of silviculture and genetics in creating and sustaining of oak forests. Proceedings of OAK 2003 FFPRI Scientific Meeting Report 3 ISSN 1341–1969. Takeuchi, A., MD Co., Ltd., Tsukuba 26 September 2005

Templeton AR (1986) Coadaptation and outbreeding depression. In: Soule ME (ed) Conservation biology: the science of scarcity and diversity. Sinauer Associates, Sunderland, pp 105–116

Tomaru N, Mitsutsuji T, Takahashi M, Tsumura Y, Uchida K, Ohba K (1997) Genetic diversity in Japanese beech, *Fagus crenata*: influence of the distributional shift during the late-Quaternary. Heredity 78:241–251

Tomaru N, Takahashi M, Tsumura Y, Takahashi M, Ohba K (1998) Intraspecific variation and phylogeographic patterns of *Fagus crenata* (Fagaceae) mitochondrial DNA. Am J Bot 85:629–636

Toyama H, Yahara T (2009) Comparative phylogeography of two closely related *Viola* species occurring in contrasting habitats in Japanese archipelago. J Plant Res 122:389–401

Tsuda Y, Nakao K, Ide Y, Tsumura Y (2015) The population demography of Betula maximowicziana, a cool-temperate tree species in Japan, in relation to the last glacial period: its admixture-like genetic structure is the result of simple population splitting not admixing. Mol Ecol 24:1403–1418

Tsukada M (1982a) Late quaternary shift of Fagus distribution. Bot Mag Tokyo 95:203–217

Tsukada M (1982b) Late quaternary development of Fagus forest in the Japanese archipelago. Jpn J Ecol 32:113–118

Tsumura Y, Suyama Y (eds) (2015) Seedling transfer guideline of Japanese tree species. Bun-ichi Co Ltd, Tokyo, p 176 ISBN-10: 482996524X, (in Japanese)

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

Ueno S, Tsumura Y (2008) Development of microsatellite markers for *Quercus mongolica* var. *crispula* by database mining. Conserv Genet 9:1083–1085

Ueno S, Taguchi Y, Tsumura Y (2008) Microsatellite markers derived from *Quercus mongolica* var. *crispula* (Fagaceae) inner bark expressed sequence tags. Genes Genet Syst 83:179–187

Ueno S, Aoki K, Tsumura Y (2009a) Generation of expressed sequence tags and development of microsatellite markers for *Castanopsis sieboldii* var. *sieboldii* (Fagaceae). Ann For Sci 66:509. https://doi.org/10.1051/forest/2009037

Ueno S, Taguchi Y, Tomaru N, Tsumura Y (2009b) Development of EST-SSR markers from inner bark of cDNA library of *Fagus crenata* (Fagaceae). Conserv Genet 10:1477–1485

Van Valen L (1976) Ecological species, multispecies, and oaks. Taxon 25(2/3):233

Vander Mijnsbrugge K, Bischoff A, Smith B (2010) A question of origin: where and how to collect seed for ecological restoration. Basic Appl Ecol 11:300–311

Weels G, Zawko G, Rosetto M, Dixon K (2003) A molecular approach to provenance delineation for the restoration of hummock grasslands (*Triodia* spp.) in arid-tropical Australia. Ecol Manage Restor 4:S60–S68

Whittemore AT, Schaal BA (1991) Interspecific gene flow in sympatric oaks. Proc Natl Acad Sci 88(6):2540–2544

Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. Evolution 38:1358–1370

Wright S (1951) The genetical structure of populations. Ann Eugenics 15: 323–354

Yang J, Di X, Meng X, Fen L, Lui Z, Zhao G (2016) Phylogeography and evolution of two closely related oal species (*Quercus*) from north and northeast China. Tree Genet Genome 12:89

Yasue M, Ogiyama K, Suto S, Tsukahara H, Miyahara F, Ohba K (1987) Geographical differentiation of natural *Cryptomeria* stands analyzed by diterpene hydrocarbon constituents of individual trees. J Jpn For Soc 69:152–156

You YH, Gi KJ, Han D, Kwak YS, Kim JH (1995) Succession and heterogeneity of plant community in Mt. Yongam, Kwangnung experimental forest. Kor J Ecol 18:89–97