



Assessments of fine-scale spatial patterns of SNPs in an old-growth beech forest

Masashi Tsukamoto¹ · Shinji Akada² · Shuichi Matsuda² · Hitomi Jouyu² · Hiromitsu Kisanuki¹ · Nobuhiro Tomaru³ · Takeshi Torimaru¹

Received: 24 January 2020 / Revised: 15 June 2020 / Accepted: 15 June 2020 / Published online: 30 June 2020
© The Author(s), under exclusive licence to The Genetics Society 2020

Abstract

The spatial patterns of non-neutral genetic variations at fine spatial scales and their possible associations with microenvironments have not been well-documented for tree populations. Based on 25–32 SNP markers, we examine whether non-neutral SNPs and their associations with microenvironments can be detected in *FcMYB1603*, a gene homologous to that encoding a protein induced by drought stress in *Arabidopsis thaliana* for the 166 adult trees in a 1-ha plot in a mature population of *Fagus crenata*. In the 83 individuals of a younger cohort of below canopy trees, the nonsynonymous SNP at locus *FcMYB1603_684* exhibited a spatial signature representing a departure from the expected spatial patterns of neutral genetic variation. Evaluations of non-neutrality for this locus were robust against the potential risks of false positives due to the low number of SNP loci, a low criterion set for minor allele frequency, and any edge effect on the trees' spatial structure. An older cohort exhibited no signal of the existence of non-neutral genetic variation, suggesting that temporal fluctuation in the microenvironmental conditions on the forest floor may have exposed different cohorts to different magnitudes of selection pressure. Although genotypes of the locus showed a spatial association with a microenvironmental variable potentially related to soil moisture, the present study was subject to a limitation due to the generally low polymorphism of nonsynonymous loci within the single plot, which suggests that it will be important to replicate the study design in order to carry out research on fine-scale non-neutral genetic variations.

Introduction

Spatial patterns of genetic variation provide information about the ecological processes through which tree populations establish and develop (Vekemans and Hardy 2004;

Wang and Bradburd 2014). Beside well-documented ecologically neutral processes (Vekemans and Hardy 2004), recent advance in sequencing technologies have enabled genome-wide scans of single nucleotide polymorphisms (hereafter SNP) (Peterson et al. 2012; Suyama and Matsuki 2015) and provided information about non-neutral genetic variations (Ahrens et al. 2018). In forest trees species, non-neutral SNPs associated with environments have been identified by comparing populations at regional or species range-wide scales where climates, elevations, or topographic conditions are markedly different from each other (Csillery et al. 2014; Tsumura et al. 2014). In contrast, few studies have attempted to identify non-neutral genetic variations at fine scales within local populations [but see Linhart and Grant (1996), compelling classical work using neutral markers]. Given that microenvironmental forces are strong enough to determine the success or failure of species-specific recruitment within forest communities even at scales of <1 ha (e.g., Akaji et al. 2017), such forces may also have the potential to drive survival selection against individuals even in the same species. Because

Associate editor: Pär K. Ingvarsson

Supplementary information The online version of this article (<https://doi.org/10.1038/s41437-020-0334-8>) contains supplementary material, which is available to authorized users.

✉ Takeshi Torimaru
torimaru@bio.mie-u.ac.jp

- ¹ Department of Sustainable Resource Sciences, Graduate School of Bioresources, Mie University, Tsu 514-8507, Japan
- ² Department of Applied Biology and Food Sciences, Faculty of Agriculture and Life Science, Hirosaki University, Hirosaki 036-8561, Japan
- ³ Graduate School of Bioagricultural Sciences, Nagoya University, Nagoya 464-8601, Japan

microenvironments are highly spatially structured (e.g., Harms et al. 2001), the spatial patterns of microenvironment-associated genes may differ from those of neutral genetic variations at fine scales. Examination of genome-microenvironment association based on SNPs will therefore help us to understand the ecological processes behind the fine-scale spatial distribution of genetic variations observed in tree populations.

Soil moisture is a major microenvironmental factor determining the success or failure of tree recruitment within local populations (Leck et al. 2008). Microtopography and soil conditions are often used as indicators of soil wetness, since microsites with concave topography and/or mature (i.e., less gravel) soils usually feature moist conditions (Barberis et al. 2002; Shin and Nakamura 2005). Such abiotic conditions are often reflected in the understory vegetation (e.g., Yamamoto et al. 1995). Individual plants are also exposed to competition with surrounding vegetation for soil water and if the amount of soil water available for individual plants is limited their growth will be reduced (Takahashi et al. 2003). Thus, assessing several indicators of soil wetness may better capture the microenvironmental signals to which plant genes respond. In addition, due to changes in abiotic and biotic microenvironmental conditions in forest floors over time (Kutnar et al. 2019; Torimaru et al. 2018), different cohorts can be exposed to different kinds and/or magnitudes of selection pressures at different times (e.g., Linhart and Grant 1996). In particular, because natural tree populations are generally composed of individuals from multiple generations, separate analysis of individual cohorts will increase the opportunity to detect non-neutral spatial patterns of environment-associated genes and to identify the factors that generate such patterns.

Transcription factors play substantial roles in the regulatory networks underlying plant responses to environmental stresses (Golldack et al. 2014). A major regulon in these responses to stresses is R2R3-MYB (myeloblastosis oncogen) transcription factors (Dubos et al. 2010). They are characterized by the presence of a conserved R2R3-DNA-binding domain in their N-terminal regions and of highly variable amino acids in their C-terminal parts (Paz-Ares et al. 1987). It has been reported that the *R2R3-MYB* gene family consists of 126 members of genes classified into 25 subgroups in *Arabidopsis thaliana* (Stracke et al. 2001) or 192 members classified into 48 clades in *Populus trichocarpa* (Wilkins et al. 2009). Some of these genes are involved in signaling pathways that are activated to cope with environmental stresses, including drought (Katiyar et al. 2012), cold weather (Agarwal et al. 2006), low nitrogen stress (Liang and He 2018), and herbivory attacks (Schafer et al. 2017). Thus, *R2R3-MYB* genes are promising genetic markers for accessing the non-neutral genetic variations associated with microenvironments.

Fagus crenata Blume (Fagaceae), which is a deciduous broadleaved tree, is the dominant canopy species in cool-temperate deciduous broadleaved forests (i.e., beech forests) in the Japanese archipelago. Spatially and temporally heterogeneous microenvironments on the forest floors are typical of these beech forests, which play a substantial role in the successful regeneration of tree species including *F. crenata* (Torimaru et al. 2018; Yamamoto et al. 1995). The species is monocious but exhibits self-incompatibility (Mukai 2008). Significant genetic structure has been reported among the trees within populations (Asuka et al. 2004b; Hanaoka et al. 2007), mainly due to the species' mode of gene dispersal; seed dispersal is by gravity and therefore restricted, and pollination is wind-mediated and distance-dependent (Inanaga et al. 2014; Oddou-Muratorio et al. 2010). Furthermore, a previous study identified 85 genes encoding the R2R3-DNA-binding domain (R2R3-DBD) among the MYB transcription factors in *F. crenata* (Matsuda et al. 2011). This large body of existing data makes the species an attractive model for exploring spatial signals of non-neutral genetic variations within tree populations.

The object of the study presented herein was to test two hypotheses: (i) whether spatial signals of non-neutral genetic variations could be detected, and if detected, (ii) whether such variations are associated with microenvironments in a local adult population of *F. crenata*. To this end, the spatial patterns of 19–25 genome-wide SNPs detected were examined in order to confirm the utility of those SNPs as a reference for neutral genetic variation. Six to seven SNPs in the *R2R3-MYB* gene of *F. crenata* (hereafter *FcMYB*) were targeted as candidate loci exhibiting non-neutral genetic variations. The population was divided into a younger and an older cohort, and those cohorts were tested to determine whether spatial outlier loci that departed from spatial patterns of neutral genetic variation could be detected. In addition, the associations of the spatial distribution of these SNPs with variables derived from microenvironments relating to soil wetness were investigated. After discussing the potential risks of erroneously identifying neutral loci as non-neutral, we propose further experimental designs in order to confirm the existence of genome-microenvironment association and to identify the selective agents and factors responsible for fine-scale spatial patterns of non-neutral genetic variation in different cohorts.

Materials and methods

Study site

The stand studied was situated in an old-growth beech forest near the Karikomiike Pond at the foot of Mt Gan-kyojisan [36°03'25" N, 136°44'23" E; summit, 1691 m

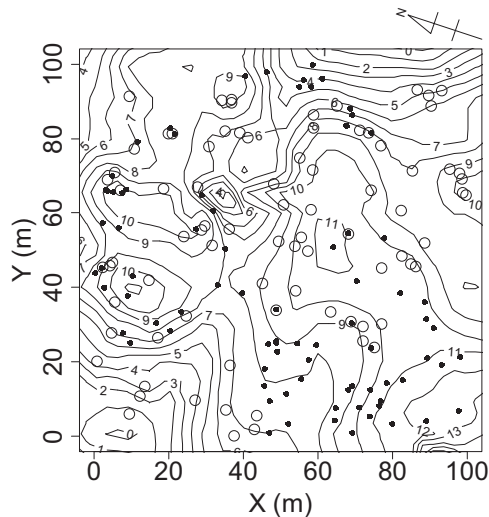


Fig. 1 Spatial distribution of individual stems of *Fagus crenata* adults. Large circles indicate stems in the upper layer; dots stems in the lower layer.

above sea level (a.s.l.)] in the southern part of the Hakusan Mountains, central Japan. The parental rock is volcanic (Ito and Shiratake 1983), and the dominant soil type is Dark Brown Forest Soil (Japan National Land Agency 1988). In 2016, we established a 1-ha (100 × 100 m) permanent plot at about 1100 m a.s.l. in the study stand and mapped all living adult stems [defined as woody stems ≥5.0 cm diameter at breast height] in three dimensions (i.e., on xyz axes) using a laser ranger (Makie et al. 2017). In 2018, there were 443 living stems representing 25 tree species in the 1-ha plot; *F. crenata* was dominant in terms of density (166 stems ha⁻¹) and widely distributed throughout the plot (Fig. 1). Dwarf bamboo [*Sasa kurilensis* (Rupr.) Makino et Shibata] and ferns (e.g., *Dryopteris crassirhizoma* Nakai, *Plagiogyria matsumurana* Makino) are the main species on the forest floor in this plot.

Field methods and microenvironmental analyses

Based on the categorization by structural layers that has been reported to be useful in inferring the ecological processes responsible for spatial patterns in different cohorts (Manabe et al. 2000; Yamamoto et al. 1995), the stems were assigned to one of the two groups according to their vertical position, crown position and height: 83 *F. crenata* individuals belonged to the canopy and subcanopy layers, with their height being ≥ca. 8 m [hereafter designated the upper layer (i.e., an older cohort)], and the remaining 83 to the understory layer [lower layer (i.e., a younger cohort)] (Fig. 1, Supplementary information Fig. S1). In June 2018, young leaves were collected from all 166 *F. crenata* individuals in the plot and stored at -25 °C until DNA was extracted.

Because we focused on loci that were expected to be associated with plant responses to drought stress (i.e., the *FcMYB1603* region; see below), four microenvironment parameters associated with soil wetness were measured as follows. First, we used the xyz coordinates of the census trees and 10 × 10 m grid points obtained from Makie et al. (2017) to estimate the elevation of each 5 × 5 m quadrat by linear interpolation. Then topographic wetness index (TWI) was calculated to quantify the pattern of soil water distribution that is affected by topography (Radula et al. 2018), with the aid of the following packages in R v.3.5.2 (R Development Core Team 2018): raster (Hijmans 2017) and dynatopmodel (Metcalf et al. 2018). The index is determined as follows:

$$TWI = \ln(\alpha/\tan\beta),$$

where α is the microtopographical upslope area draining through a certain point per unit contour length, which is equal to a certain grid cell width, and β is the local slope. A higher TWI represents a wetter microsite. Second, in addition to the proportions of rock and/or gravel (i.e., magnitude of soil immaturity), we used the amounts of fern and *Sasa* cover as biological indicators of the degree of soil maturity (Torimaru et al. 2018; Yamamoto et al. 1995). In the autumn of 2018, we estimated the proportion of cover of dwarf bamboo and that of ferns in each of 400 5 × 5 m contiguous quadrats within the plot; the proportion of the cover in each quadrat was visually inspected and quantified at 0.05 intervals except that in cases where cover was present but its proportion was <0.05, cover was set at 0.01 (e.g., Yamamoto et al. 1995). Similarly, the proportions of rocks and/or gravels (surface soil conditions) were visually inspected and quantified in each of those quadrats on the same scale as that used for the vegetation census above (e.g., Torimaru et al. 2018). Third, we considered *Sasa* cover to be an indicator of competition with beech trees for soil water, since trees competing with *Sasa* for soil water have reportedly exhibited reduced growth (Takahashi et al. 2003). However, *Sasa* cover is also expected to be influenced by the degree of soil maturity mentioned above and/or may be influenced by soil moisture gradients generated through microtopography, hence the raw values of *Sasa* cover may confer substantial bias when interpreting the effects of competition for soil water. To mitigate such biases, we performed non-metric multidimensional scaling (NMDS) analysis to summarize these four variables and converted them into more refined forms, using the “vegan” package (Oksanen et al. 2018) in R v.3.5.2 (R Development Core Team 2018). Before NMDS analyses, proportional variables were arcsin square-root transformed and TWI log-transformed to improve the normality of data distribution. The metaMDS procedure in vegan was used with default options, which include use of the Bray–Curtis dissimilarity

Table 1 Pearson's correlation coefficients among the four microenvironmental variables and the non-metric multidimensional scaling (NMDS) axes for 400 5 × 5 m quadrats.

NMDS axis	<i>Sasa</i> cover	Fern cover	Proportion of rock and/or gravel	TWI ^a
NMDS1 (Microtopographic effect)	0.448***	−0.234***	0.171**	−0.940***
NMDS2 (Effect of soil maturity)	−0.241***	0.719***	0.811***	−0.010 ^{n.s.}
NMDS3 (Effect of biological competition)	−0.729***	−0.207***	0.139 ^{n.s.}	−0.176**

Interpretation of each NMDS against the factors potentially affecting soil moisture is in parentheses.

Bonferroni correction was used to determine significance in multiple testing.

*** $P < 0.001$; ** $P < 0.01$; ^{n.s.} $P \geq 0.05$.

^aTopographic wetness index.

index and a maximum of 20 random starts in search of the stable solution, except that we tested a number of dimensions ranging between one and three. To evaluate the set of ordinations obtained, stress values (Kruskal 1964) were calculated and their significance was assessed by 1000 simulations with permutation of the microenvironmental variables among 400 quadrats (Sasaki et al. 2015), followed by Bonferroni correction.

Mig-seq experiment and SNP detection

Total genomic DNA was extracted using a modification of the hexadecyltrimethylammonium bromide method (Murray and Thompson 1980) as described in Asuka et al. (2004a). Genome-wide SNPs were detected using multiplexed ISSR genotyping by sequencing (Mig-seq) with a minor modification. In principle, this technique amplifies the loci between two ISSRs by PCR, and sequence analysis is carried out using a next-generation sequencer (Suyama and Matsuki 2015). The Mig-seq libraries were prepared following the protocol outlined in Suyama and Matsuki (2015) but we used the adapter sequences and barcode sequences (10–12 base) for the Ion Proton sequencing platform (Thermo Fisher Scientific) to identify each individual sample. The final PCR products for each individual were multiplexed in the size range 200–500 bp using Agencourt Ampure XP (Beckman Coulter Inc., California, USA) and sequenced on an Ion Proton platform using an Ion 318TM Chip v2 (Thermo Fisher Scientific) at the Center for Molecular Biology and Genetics in Mie University. We constructed three sets of data, representing all the adult individuals in the population ($n = 166$), and the individuals belonging to the lower ($n = 83$) or upper layer ($n = 83$). SNPs were called using Stacks 2.0 (Catchen et al. 2013) in which minor allele frequency (MAF) was set to 0.01 (see Appendices S1 and S2 in detail), and mapping of the reads to the reference sequence yielded, respectively, 35, 28, and 23 SNPs for the whole population, the lower layer and the upper layer. After applying the further filtering setting (Appendix S2), we detected, respectively, 25, 24, 19 SNPs with 91.4,

93.7, and 92.9% genotyping rates for the three categories (Tables 1, S2, and S3).

DNA sequencing of MYB region and SNP detection

We targeted the *R2R3-MYB* genes identified by Matsuda et al. (2011). Preliminary experiments demonstrated that expression of *FcMYB1603* increased strikingly after drought treatment both in the first leaves and in the roots of a month-old beech seedling (Appendix S3 and Figs. S2, S3), and the full genomic sequence of *FcMYB1603* was identified (Appendix S4). BLAST reported that the protein sequence showed high similarity (E value = 3×10^{-180}) to the transcription factor MYB102-like in *Quercus suber*, a member of the same family (Fagaceae). The expression of *MYB102* in *A. thaliana*, *AtMYB102*, was induced when plants were exposed to drought stress (Denekamp and Smeekens 2003) or to abscisic acid, which is a drought-induced hormone (Leonhardt et al. 2004). We designed forward (5'-GGAAAAAGCTGCCGACTTCG-3') and reverse (5'-AATGGTGTGGGCTCGATGT-3') primers from, respectively, exon 2 (R2R3-DBD region) and exon 3 in the *MYB* region. PCRs were carried out in 10- μ L volumes, each containing 1–10 ng of template DNA, 0.5 U of AmpliTaq Gold[®] 360 DNA Polymerase, 1 \times AmpliTaq Gold[®] 360 buffer (Thermo Fisher Scientific), 0.2 mM of each dNTP (New England BioLabs Inc., Massachusetts, USA), and 0.2 μ M of each primer pair. PCR was performed with an initial denaturation for 10 min at 95 °C, followed by 30 cycles of denaturation for 1 min at 94 °C, annealing for 1 min at 60 °C and extension for 1 min at 72 °C, with a final extension for 7 min at 72 °C. Sanger sequencing was performed at MacroGen Japan Inc. using a 3730xl DNA analyzer following the manufacturer's protocol (Thermo Fisher Scientific). DNA sequences were edited manually in ApE (Davis 2017), and sites that were heterozygous, polymorphic, or of low sequence quality were visually examined by checking electropherograms. Alignments were performed using MUSCLE implemented in MEGA 5 (Edgar 2004; Tamura et al. 2011). Sequences with a length of 868 bp were aligned for all of the 166 individuals, and there

were 13 polymorphic sites for each of the three categories (Fig. S4). After applying the filtering setting (Appendix S5), these were reduced to seven sites (=seven loci) for the whole population and lower layer, and six sites for the upper layer (Table 1). There were four and three sites with synonymous and nonsynonymous substitutions, respectively (Fig. S4).

In total, we used 32, 31, and 25 SNPs from Mig-seq and *FcMYB1603* loci in the subsequent data analyses for the whole population, the lower layer, and the upper layer, respectively.

Genetic variation and spatial genetic structure

The genetic diversity of *F. crenata* in the 1-ha plot was analyzed using standard population genetics parameters: observed (H_O) and expected (H_E) heterozygosity, and inbreeding coefficient F_{IS} (Weir and Cockerham 1984). Deviations from Hardy–Weinberg equilibrium at each locus were evaluated by the exact test using GENEPOP version 4.2 (Rousset 2008), with false discovery rate (FDR < 0.05) correction for multiple testing (Benjamini and Hochberg 1995).

To describe the spatial patterns of genetic variations of SNP markers, we calculated their kinship coefficient F_{ij} [coancestry; (Loiselle et al. 1995) in the plot using SPA-GeDi version 1.4c (Hardy and Vekemans 2002)]. The mean F_{ij} value was calculated for each of 10 continuous distance classes of 10 m intervals, from 0–10 to 90–100 m (the numbers of pairs of individuals per distance class ranged from 175 to 435 and from 117 to 525 in the lower and upper layers, respectively). The regression slope of F_{ij} against the logarithm of the distance between adult trees (hereafter b_F) was calculated following the procedure of Oddou-Muratorio et al. (2010). The significance of the mean F_{ij} and b_F values was assessed by 1000 simulations with permutation of the spatial distances between adult trees.

Spatial outlier detection based on the MSOD–MSR method

We used a two-step method of spatial outlier detection based on the power spectrum of the Moran eigenvector map (MEM) (Dray et al. 2006). The MEM power spectrum quantifies how the variation in a variable, such as the frequency of an allele at a SNP locus, is distributed across a range of spatial scales defined by MEM spatial eigenvectors (Wagner et al. 2017). The first step [Moran spectral outlier detection (MSOD)] uses genetic and spatial information to identify outlier loci by their unusual power spectrum. The second step uses Moran spectral randomization (MSR) to test the association between outlier loci and environmental predictors, accounting for spatial autocorrelation.

In the first step (MSOD), we identified the outlier loci that deviated from a spatial pattern of genotypes derived from an ecologically neutral process of gene dispersal (i.e., non-neutral loci). We firstly defined neighbors of each individual at n locations based on a Gabriel graph [a proximity graph that captures some concept of neighborliness; see Gabriel and Sokal (1969)], and the j th neighbors of the i th individual received a weight w_{ij} proportional to the inverse distance $1/d_{ij}$ between them. The weights w_{ij} of all neighbors of i th individual were then normalized so that they summed to 1. This spatial weight matrix W of size $n \times n$ was used to derive MEM by eigenanalysis of the symmetric matrix $W_s = 0.5 \times (W + W^T)$, where T denotes a matrix transposition (Dray et al. 2006). The eigenanalysis results in $n - 1$ orthogonal and uncorrelated eigenvectors V_k associated with the k th largest eigenvalue λ_k , while a single eigenvector with zero eigenvalue is dropped. The vectors of the correlation between each of m loci and the matrix V [of size $n \times (n - 1)$] were calculated, generating a matrix of correlations between each locus and each spatial eigenvector, which was denoted by $R.YV$ [of size $m \times (n - 1)$]. For the correlation between the l th locus and V_k in $R.YV$, denoted by $r.YV_{lk}$, the power spectrum $r.YV_{lk}^2$ indicated the proportion of variance of the l th locus explained by the k th spatial eigenvector, and satisfied the condition $\sum_k r.YV_{lk}^2 = 1$, which means that the $n - 1$ spatial eigenvectors together fully explain the variance in allele frequencies at the locus. The deviation of the spectrum $r.YV_{l.}$ of the l th locus from the mean spectrum S , denoted by D_l , was quantified as $D_l = \sum_k \left(\frac{r.YV_{lk}^2}{S_k} - 1 \right) \cdot b_k$, where S_k indicates the mean values of proportion of variance explained by the k th spatial eigenvector across all loci, and $b_k = 1$ if the term in brackets is negative and $b_k = 0$ if it is positive (Wagner et al. 2017). D_l values were scaled to obtain a z score [i.e., $z(D_l)$] for all loci. The $|z(D_l)|$ was compared with the cutoff value corresponding to the two-sided probability of 0.01 assuming normal distribution, $|z_{0.01}| = 2.58$, which is recommended because it offers the best balance between high power to detect true positives and low false positive rates (FPRs) (Wagner et al. 2017). In addition, to further assess the validity of significance for the loci with $|z(D_l)| > |z_{0.01}|$, we generated the empirical distribution of the z score by permuting the individual genotypes of one locus but keeping the other loci, and compared the observed $z(D_l)$ with the 99% confidence intervals of the empirical distribution.

One factor that can cause fluctuations in z scores is the edge effect, in which the limited number of neighbors of an individual located near the edge of the study plot increases its w_{ij} relative to that of one located distant from the edge even when the spatial patterns of neighbors are same between the two individuals. This may cause a tendency to detect a significant z score depending on the distance to the plot's edge from individual trees. To examine this potential

problem, we chose pairs of individuals separated by <5 m each other [we tentatively set this criterion based on the fact that the two individuals in which an outlier SNP genotype was detected were separated by ca. 5 m (see “Results”)]. We calculated the minimum distances to the edges of the plot from the centroids of the pairs. Then we applied the generalized linear model with a binomial distribution and logit link function. The response variable is one if the pairs of individuals exhibit $|z(D)| > |z_{0.01}|$ and the $z(D)$ departs from the 99% confidence intervals of the empirical distribution above, otherwise zero. The explanatory variable (i.e., distance to the nearest edge) was log-transformed to improve the normality of residual variances. The significance of the explanatory variable was tested by comparing the change in deviance with the χ^2 distribution (Bolker et al. 2009).

In the second step (MSR), because we focused on a *MYB* gene associated with drought stress, it was hypothesized that non-neutral spatial patterns of those SNPs detected are associated with some soil moisture indicators in the plot. To test this hypothesis, we firstly estimated values of NMDS axes for each individual’s location by interpolation and extrapolation using the *akima* package of R statistical environment (Akima and Gebhardt 2016). Then ordinary Pearson coefficients of correlations (r_{le}) were calculated between the l th locus and the e th NMDS variable. The r_{le} can be decomposed into two matrices representing the correlation of the spatial eigenvectors in matrix V_k with the locus ($r.YV_{lk}$) and the NMDS variable ($r.XV_{ek}$): $r_{le} = (\sum_k r.YV_{lk} \cdot r.XV_{ek})$ (Wagner et al. 2017). In MSR, an empirical distribution for r_{le} under the null hypothesis of no correlation that preserved the spatial structure of both variables (i.e., retaining the power spectra of both variables) was obtained by randomizing the NMDS variables with the “singleton” method (Wagner and Dray 2015), where the new correlation matrix between NMDS variable and spatial eigenvector ($r.XV_{ek,rand}$) was generated by randomizing the sign of each element in $r.XV_{ek}$. Then $r_{le,rand}$ was obtained by replacing the $r.XV_{ek,rand}$ with $r.XV_{ek}$ in the formula above. This procedure was repeated 5000 times and the P value was computed as the proportion of $|r_{le,rand}|$ larger than the $|r_{le}|$ observed (Wagner et al. 2017). Furthermore, to assess the validity of significance of associations of the NMDS variables with the loci that were identified as spatial outliers in MSOD, we permuted the individual genotypes of the locus while keeping the other loci, estimated the P values of the correlation coefficients based on the “singleton” method above, and determined the proportion of the P values that were <0.05. We performed the MSR analysis for all SNP loci, but any locus that was statistically significant in both the MSOD and the MSR analysis was considered to be substantially associated with the environment, as recommended by Wagner et al. (2017).

We performed all of the MSOD and MSR analyses using the source code from Wagner et al. (2017), with the aid of

the following packages in R v.3.5.2 (R Development Core Team 2018): *adespatial* (Dray et al. 2016) and *spdep* (Bivand and Piras 2015).

Results

Microenvironmental characteristics of the stand

The four microenvironmental variables were spatially heterogeneously distributed in the plot. Dwarf bamboo and ferns covered some of the forest floor; the proportion of *Sasa* cover per quadrat ranged from 0.00 to 0.95 with an average of 0.25, and that of fern cover ranged from 0.00 to 0.80 with an average of 0.12 (Fig. S5). The proportion of rocks and/or gravels ranged from 0.00 to 0.95 with an average of 0.10. Ground surfaces with high proportions of rocks and/or gravels were observed mostly at the eastern corner of the plot (Fig. S5). TWI ranged from 0.96 to 13.0 with an average of 4.70 and microsites with concave slopes tended to be wet (Fig. S5).

The NMDS analyses showed that the stress value was 0.302 when the four microenvironmental variables were summarized into one axis, whereas the values were 0.170 and 0.109 in the cases when they were summarized into two and three axes, respectively. There was a statistically significant stress value in the case of three axes ($P = 0.024$) but not in the remaining two cases (single axis; $P = 1.000$, two axes; $P = 0.132$). Thus we used the values of NMDS in the case of three axes for subsequent analyses. The first axis (hereafter NMDS1) was predominantly correlated with TWI (Table 1 and Fig. S6), and thus represented the magnitude of soil moisture associated with microtopography in the plot. The second axis (NMDS2) was predominantly and positively correlated with fern cover and the proportion of rock and/or gravels as well as negatively correlated with *Sasa* cover (Table 1 and Fig. S6). Given that these ferns prefer microsites with rocky soils and/or gravels (Torimaru et al. 2018) and that *Sasa* usually grows predominantly on microsites with mature soils (Yamamoto et al. 1995), the NMDS2 value represents the magnitude of soil maturity, and is likely to be associated with soil moisture. The third axis (NMDS3) was predominantly correlated with *Sasa* cover (Table 1 and Fig. S6), constituting an indicator potentially representing the magnitude of biological competition for soil water (Takahashi et al. 2003).

Genetic variation and spatial genetic structure of the *F. crenata* population

For the SNPs from Mig-seq experiments, the H_E values ranged from 0.021 to 0.271 (with an average of 0.075), 0.025 to 0.419 (0.103), and 0.024 to 0.461 (0.106) for the whole population, the lower and the upper layer, respectively (Tables 2 and S1–S3). For the SNPs from the

Table 2 Summary of SNPs from Mig-seq and *FcMYB1603* loci in the *Fagus crenata* trees in the 1-ha plot.

Genetic markers	No. of loci	All individuals				
		N	H_O	H_E	F_{IS}	b_F
Mig-seq SNPs	25	166.0	0.074	0.075	0.029 ^{n.s.}	-0.012 ^{***}
<i>FcMYB1603</i> SNPs	7	147.3	0.176	0.167	-0.022 ^{n.s.}	-0.012 ^{**}
All SNPs	32	151.4	0.096	0.095	0.018 ^{n.s.}	-0.012 ^{***}
Genetic markers	No. of loci	Lower layer				
		N	H_O	H_E	F_{IS}	b_F
Mig-seq SNPs	24	76.3	0.089	0.103	0.074 ^{n.s.}	-0.016 ^{**}
<i>FcMYB1603</i> SNPs	7	83.0	0.174	0.165	0.009 ^{n.s.}	-0.023 [*]
All SNPs	31	77.8	0.108	0.117	0.059 ^{n.s.}	-0.017 ^{***}
Genetic markers	No. of loci	Upper layer				
		N	H_O	H_E	F_{IS}	b_F
Mig-seq SNPs	19	75.3	0.093	0.106	0.068 ^{n.s.}	-0.035 ^{***}
<i>FcMYB1603</i> SNPs	6	83.0	0.181	0.193	-0.039 ^{n.s.}	0.004 ^{n.s.}
All SNPs	25	77.1	0.117	0.124	0.043 ^{n.s.}	-0.018 ^{**}

Bonferroni correction was used to determine significance in multiple testing for FIS.

Deviations of b_F from zero were assessed by 1000 simulations, permuting the spatial distances between adult trees.

N mean number of genotyped individuals per locus, H_O mean observed heterozygosity per locus, H_E mean expected heterozygosity per locus, F_{IS} mean fixation index per locus, b_F the estimates of slopes of regression for coancestry values (F_{ij}) against the logarithm of the spatial distances.

*** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$; n.s. $P \geq 0.05$.

FcMYB1603 region, the H_E values ranged from 0.030 to 0.501 (with an average of 0.167), 0.024 to 0.497 (0.165), and 0.024 to 0.566 (0.193) for the whole population, the lower layer and the upper layer, respectively. For each of the two markers, the mean values of H_E were similar between the layers, and the F_{IS} of every locus from each layer showed no significant deviation from zero ($P > 0.05$) (Tables 2 and S1–S3).

The correlograms of coancestry for the 166 individuals showed that mean values for the first distance class were significantly positive (Fig. 2). The values showed a general trend of decreasing with distance class, and such trends were also found in the individuals belonging to the lower layer irrespective of genetic markers and those in the upper layer for Mig-seq SNPs only (Fig. 2). Those correlogram patterns are consistent with the estimates of b_F , which exhibited significant negative values except for the SNPs including the *FcMYB1603* loci in the upper layer (Table 2).

Identification of outlier SNP loci from spatial signature and the association with microenvironmental variables

In the lower layer, *FcMYB1603_684* had a z score of -2.848, which was below the lower boundary of the

threshold, whereas all of the z scores of the Mig-seq SNPs were within the range of the threshold (Fig. 3). As a consequence, frequency distributions of D_l departed from normality in the lower layer (Shapiro–Wilk test, $W = 0.917$, $P = 0.02$). No such departure was found in the whole population ($W = 0.937$, $P > 0.05$) or upper layer ($W = 0.955$, $P > 0.05$). For *FcMYB1603_684* in the lower layer, since only the two minor heterozygotes and the remaining 81 major homozygotes were found (Fig. S7), we simulated all of the permutations (i.e. 3403 cases), computed the z score, and obtained the empirical distribution with a 99% confidence interval ranging between -2.787 and 0.253 (Fig. S8).

There were 58 cases in which the two individuals in a pair were separated from each other by <5 m. Nine cases showed z scores deviating from the 99% confidence interval of empirical distribution (Fig. S9), but we did not detect any effect of distance from the plot edge on the occurrence of a statistically significant z score (regression coefficient against the logarithm of distance to plot edge was -0.831, $\chi^2 = -3.058$, $P > 0.05$), indicating that outlier detection was not affected by the positions of individual trees relative to the plot edges.

Application of the MSR method identified three cases that demonstrated a statistically significant correlation of

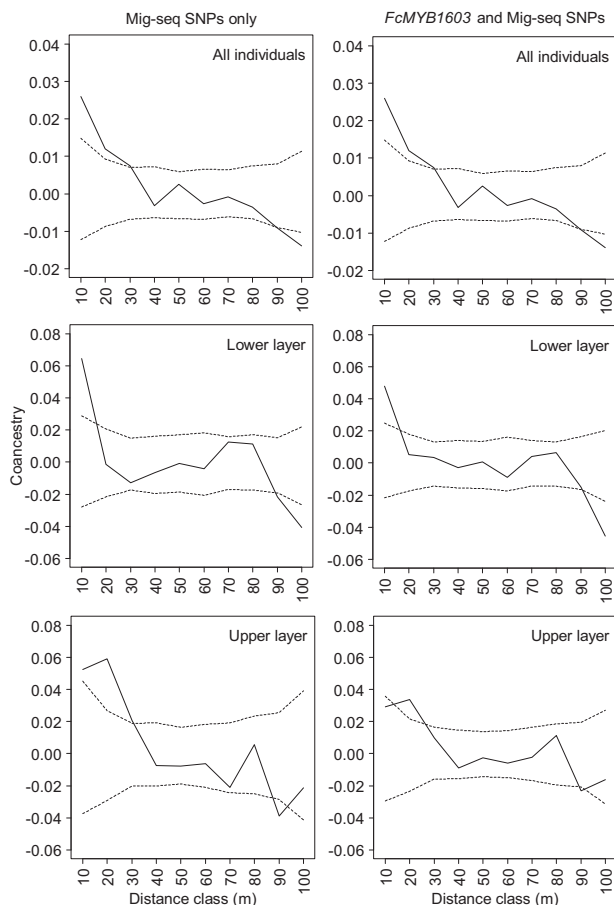


Fig. 2 Correlograms of mean coancestry values for Mig-seq SNPs only (left) and SNPs from Mig-seq and *FcMYB1603* (right) in *Fagus crenata* individuals in the 1-ha plot; the panels show all individuals (top, $n = 166$), lower layer (middle, $n = 83$), and upper layer (bottom, $n = 83$) in the populations. Distance classes were defined at continuous 10 m intervals from 0–10 to 90–100 m. Dashed lines indicate 95% confidence intervals of coancestry values based on 1000 simulations permuting the individual genotypes. Note that the numbers of Mig-seq SNPs were 25, 24, and 19 in the whole population, the lower layer, and the upper layer, respectively, and the numbers of *FcMYB1603* SNPs were seven for the whole population and the lower layer, and six for the upper layer.

SNPs with the NMDS variables in the whole population and the lower layer, whereas there were seven cases in the upper layer (Fig. 4). In the whole population and the upper layer, since all of the loci that showed significant association with NMDS variables could be identified as neutral based on MSOD analysis (Fig. 3), the associations were likely to be false positives (see Wagner et al. 2017 for interpreting the combined results of MSOD and MSR analyses in detail). In the lower layer, the SNPs whose z score in the MSOD method departed from the 99% confidence interval (i.e., *FcMYB1603_684*, Fig. 3) were associated with the NMDS3 variable (Fig. 4); the individuals that were heterozygous for the minor allele at the locus were located on sites with higher values of NMDS3 (i.e., microsites where there was

less competition with *Sasa* for soil water) (Fig. S7) and were close to each other (separated by 5.3 m). Simulation based on permuting the genotypes of *FcMYB1603_684* indicated that significant associations of the locus with NMDS3 variables were present in 180 cases out of all those analyzed; that is, the proportion was 0.053 ($=180/3403$). Among these cases, the observed pattern of spatial distribution of the genotypes was the only situation in which there was no spurious association with XY coordinates and MSOD detected the spatial outlier.

Discussion

The main findings arising from the present research are that (i) we detected a spatial signal of departure from neutrality for an SNP in *FcMYB1603*, a gene that is associated with drought stress in *F. crenata*, in the younger cohort but not in the older one, whereas genome-wide SNPs exhibited no departure from the spatial patterns expected from neutral genetic variation for any of the three categories, and (ii) the non-neutral locus identified, *FcMYB1603_684*, was spatially associated with a microenvironmental variable potentially related to soil moisture. Since false identification of neutral or non-neutral loci can lead to misunderstanding of the ecological processes behind the spatial patterns of genetic variations, we firstly focus on discussing whether the neutrality and non-neutrality derived from our statistical procedures are likely or not, and then infer the ecological phenomena responsible for the inter-generational difference in the detection of non-neutral genetic variation.

Evaluation of the spatial outlier detected by MSOD and MSR

The criterion for MAF set in our study (0.01) was lower than that in other published studies, where $MAF \geq 0.05$ was often used (Ahrens et al. 2018). Thus, there was some concern that such low levels of polymorphism may have been generated through processes other than gene dispersal, such as sequencing errors and/or mutations occurring randomly across the genome, which would also result in spatial patterns of genetic variation indistinguishable from those of neutral ones. However, our study detected a pattern of spatial genetic variation, that is a negative relationship between coancestry and spatial distances between the trees based on the Mig-seq SNP loci. This pattern has previously been reported to originate from gene dispersal by isolation by distance in tree populations (Vekemans and Hardy 2004). Because our preliminary study based on microsatellite markers also confirmed the pattern (Appendix S6, Fig. S10, and Tables S4, S5), the mode of isolation by

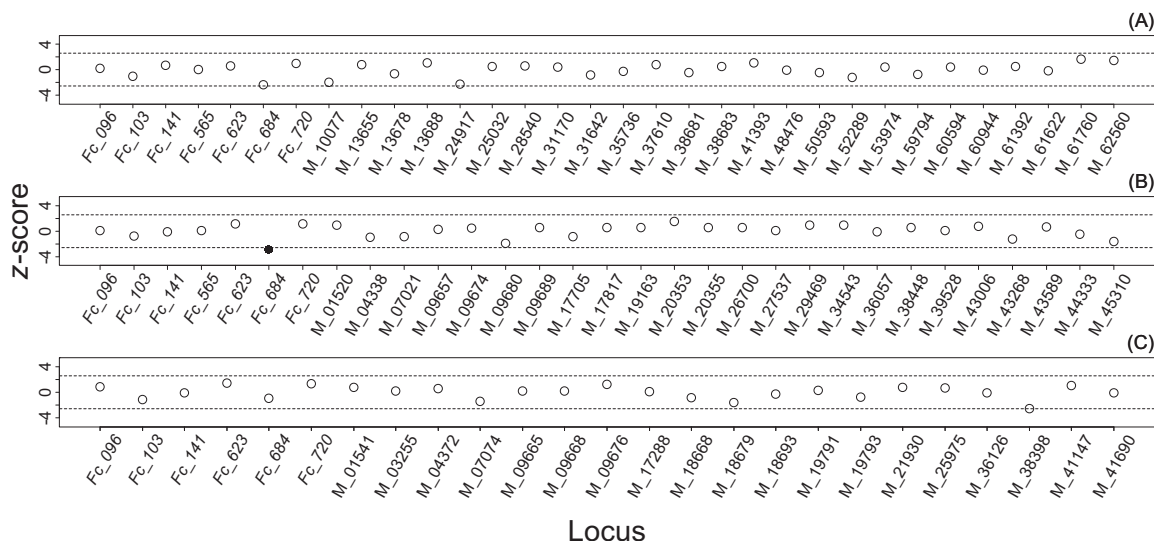


Fig. 3 Outlier detection with the Moran spectral outlier detection (MSOD) method in the *Fagus crenata* populations. All individuals [a $n = 166$, 32 SNPs], lower layer [b $n = 83$, 31 SNPs], and upper layer [c $n = 83$, 25 SNPs]. Each point shows the z score of a locus representing the degree of deviation from the mean spectrum of the

proportion of variance of the locus associated with each spatial eigenvector, plotted against the locus ID (“Fc” represents the FcMYB1603 locus, and “M” Mig-seq loci). Dashed lines indicate the probability levels of 0.01 used in MSOD. Open circles and dot indicate spatial non-outlier and outlier loci, respectively.

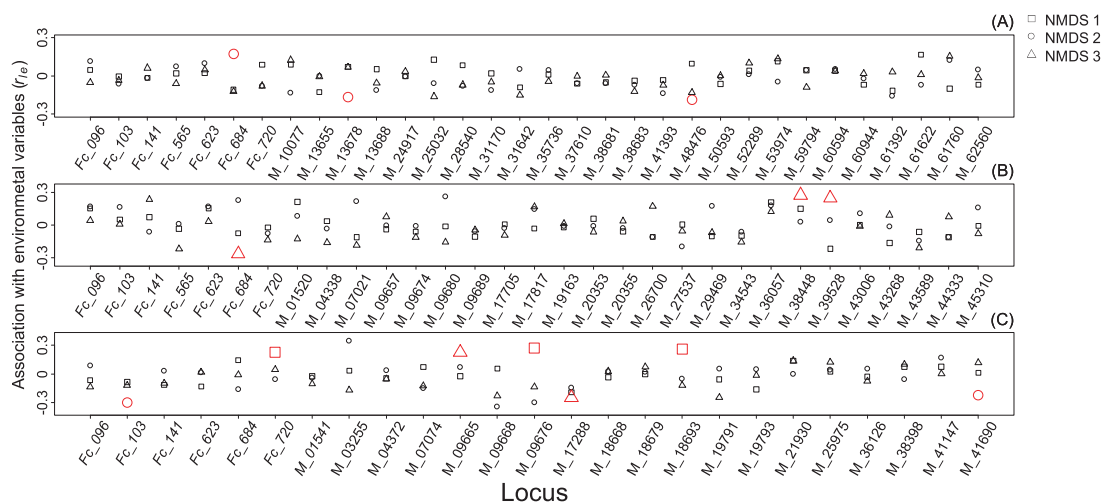


Fig. 4 Correlation coefficients of SNP genotypes with the three NMDS variables (r_{le}) in the *Fagus crenata* populations. All individuals [a $n = 166$, 32 SNPs], lower layer [b $n = 83$, 31 SNPs], and upper layer [c $n = 83$, 25 SNPs]. Positive values of r_{le} indicate that homozygotes with major SNP alleles tend to be located at sites with

higher values of NMDS variables. The correlation coefficients were tested based on Moran spectral randomization (MRS) with the singleton method, and significant correlation ($P < 0.05$) is shown by red large symbols.

distance evidently operates in the population studied. Thus the spatial patterns of the Mig-seq SNPs in our study were likely to have been generated through ecologically relevant and neutral processes.

In addition, a nominal level of sequencing error was obvious for the *FcMYB1603* region, since the Sanger method is the gold standard for DNA sequencing. Based on the synonymous mutation rate of 2.5×10^{-9} per site per year in *Populus* that is frequently used as the rate of mutation in angiosperm tree species (Ingvarsson 2008) and the mean stand age of 205 years

in the *F. crenata* populations in the region studied (Senno 1979), the probability of finding mutated sites across 868 bp of *FcMYB1603* per individual tree was estimated to be 4.4×10^{-4} ($= 1 - [(1 - 2.5 \times 10^{-9})^{868}]^{205}$), and the number of individuals with mutated sites in the plot studied was estimated to be 0.07 ($= 1 - (1 - 4.4 \times 10^{-4})^{166}$). Furthermore, considering that nonsynonymous substitutions, which occurred at some genomic sites in our study (Fig. S4), are usually less likely than synonymous ones (Hartl and Clark 2007), mutations were unlikely to have contributed to the genetic variation of

FcMYB1603 found in the present study. It has been reported that nonsynonymous SNPs, which could be also detected in the present study, are more likely to have a $MAF < 0.05$ (Cargill et al. 1999), and alleles associated with a specific environment can be significantly rarer than the average allele (Foumier-Level et al. 2011).

The distance of individuals from the plot edge had no effect on the z score for *FcMYB1603_684*, suggesting that the MSOD analysis suffered little from an edge effect for this locus. Because the present results showed that the two individuals with the minor genotype were close to each other (i.e., they formed a single spatial cluster), we could easily quantify the relationship of the plot edge to the features of the spatial structure of target individuals by using the distance of the centroid of the two individuals to the plot edge. However, in general the situation will be more complex, and it will not be an easy task to associate the plot edge with the spatial structure of the population where multiple individuals possessing the same genotypes can be divided into several spatial clusters. Thus, while our special case study allows us to investigate the edge effect on the z score for *FcMYB1603_684*, further studies are needed to establish a more general framework in which to evaluate the consequences of the edge effect for the detection of spatial outlier loci.

A more general concern is the small number of SNP loci used in our study, relative to other studies [i.e., generally >100 SNPs (Ahrens et al. 2018)]. To check the robustness of the MSOD and MSR approaches against a small number of SNPs, we borrowed the simulated dataset and the source code of the R program from the original studies introducing MSOD and MSR (Forester et al. 2016; Wagner et al. 2017). Our simulations confirmed that a reduction of SNP loci to 29 (equal to the mean number of loci across the three datasets in the present study) resulted in trends similar to the results from 100 SNP loci in terms of true positive rate (TPR; the proportion of selected loci correctly identified as outliers) and FPR (the proportion of neutral loci that were erroneously identified as outliers) (Fig. S11), as effectively as in the original studies (see Figs. S6 and S7 in Wagner et al. 2017). Despite some reductions in the power to detect loci under selection (i.e., lower TPR) in the case of 29 SNPs, there were nominal levels of FPR across the various situations of selection strength, dispersal and habitat configurations (Fig. S11). Furthermore, these negligible levels of FPR were maintained even if the assumption of normality of D_i was violated in MSOD (Fig. S11). Our results therefore suggest that the number of SNP loci used in the present study would be acceptable for MSOD and MSR analyses when exploring non-neutral loci at a fine spatial scale.

A concern specific to the present study arises from the results for the locus *FcMYB1603_684*, where the non-synonymous SNP was identified as non-neutral, and for

which the only two heterozygous individuals grew close to each other in a spot with a wetter microenvironment and all other trees were homozygotes for the major allele. There was thus concern that the two heterozygous individuals may have happened to grow in such a microenvironment. Our study evaluated the locus based on the two procedures (i.e., MSOD and MSR). In the first procedure (MSOD), all loci were used to quantify the mean values of proportion of variance explained by each spatial eigenvector across loci (i.e., S_k in “Materials and methods”) and to determine the mean and standard deviation of D_i values (see “Materials and methods”); hence, in order to calculate the z score of one locus, MSOD utilizes the polymorphisms and spatial patterns of the other loci analyzed, not solely those of the single locus targeted. Indeed, the mean values of z score of *FcMYB1603_684* decreased (and the proportion of significant z score increased) with the number of loci that were added to MSOD (Fig. 5), indicating the dependency of z score for one locus on the other loci. Thus, the low polymorphism of the locus *FcMYB1603_684* in the lower layer was likely to be compensated for by the information about the other 30 loci in the present study.

In contrast, caution should be exercised in interpreting the result of MSR analysis (i.e., the second procedure) showing the association of *FcMYB1603_684* with a microenvironmental variable potentially relating to soil moisture (i.e., NMDS3). While the result seems to be in line with those of other studies that have reported the associations of SNPs with environmental variables affecting soil moisture (Csillery et al. 2014; Krajmerova et al. 2017) at regional scales, such studies have adopted thresholds of MAF higher than in the present study. Because MSR used solely the polymorphism of the loci targeted when examining the association with each environmental variable, the low polymorphism of *FcMYB1603_684* could have inflated

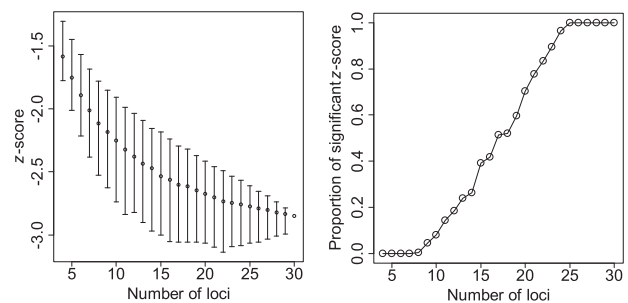


Fig. 5 Mean values (open circle) and standard deviations (error bars) of z score (left panel) and proportions of significant z score (right panel) for the locus *FcMYB1603_684* against the number of loci used (minimum was four loci) for MSOD analysis. Note that the numbers were those used to combine with the locus *FcMYB1603_684*. In cases where the number of loci ranged from 4 to 27, we randomly chose the loci, calculated the z scores for the locus *FcMYB1603_684*, and repeated the procedure 1000 times. Otherwise, we calculated the z scores for the locus *FcMYB1603_684* for all the combinations.

uncertainty when justifying the spatial association with the microenvironmental variable. Given the reported studies that showed generally low polymorphisms for nonsynonymous SNPs (Cargill et al. 1999; Fournier-Level et al. 2011), a dataset derived from a single plot as in our study will be inadequate, and studies applying MSR within local populations should a priori plan to utilize multiple populations in which to replicate the same design in order to justify the spatial association of non-neutral loci with microenvironments.

Inconsistency of detection of non-neutrality between the two layers

It is expected that separation of an adult population into different categories linked to cohorts would facilitate the identification of non-neutrality of SNPs in tree populations, since rapid changes in microenvironmental conditions can expose different cohorts within the populations to different kinds and/or magnitudes of selection pressure, thereby generating inter-generational differentiation in spatial patterns of non-neutral genetic variations (see “Introduction”). The present study supports this prediction; a signal of non-neutrality was detected in a younger cohort, whereas there were no signals of non-neutrality when pooling the trees belonging to different cohorts (i.e., whole population). Given that tree mortality occurs predominantly in the earlier stages of trees’ life history [especially at the time of seedling establishment; reviewed by Leck et al. (2008)], microenvironmental conditions at the timing of seedling establishment may have been different between the cohorts. Because our study did not succeed in relating the spatial patterns of non-neutral genetic variations to ecologically meaningful factors due to the inadequate dataset derived from a single plot, further studies combining the results of genome-microenvironment associations obtained from several *F. crenata* populations will be needed to identify the selective agents and factors that can explain the absence/presence of spatial patterns of non-neutral genetic variations among different cohorts.

Conclusions

No signals of departure from neutrality were detected among Mig-seq SNPs (i.e., genome-wide SNPs), whereas for the SNPs from the *FcMYB1603* region, one nonsynonymous SNP locus named *FcMYB1603_684* exhibited a spatial distribution that departed from those expected under the assumption of an ecologically neutral process of gene dispersal. Simulations suggested that the signal of statistical significance detected at the locus was robust against the potential risks of false positives that might have arisen due to the low number of SNP loci, a low criterion set for MAF, and any edge effect on the spatial structure of the trees, and

thus the locus could be considered to be a spatial outlier. Inconsistency of detecting non-neutrality between the two layers was found in the case of this locus, suggesting that temporal changes in microenvironmental conditions could expose those cohorts to different kinds and/or magnitudes of selection pressures. We concluded that the locus was at least in part affected by processes other than ecologically neutral processes of gene dispersal in the study plot. However, the present study was subject to several limitations, and although the locus exhibited a spatial association with a microenvironmental variable potentially related to soil moisture, the low level of polymorphism at the locus would have reduced the statistical reliability in the single plot used in the present study. The focus on a single population also meant that it was not possible to justify drawing conclusions about what kinds of ecological processes were behind the pattern observed for each cohort. These limitations indicate that further studies examining fine-scale non-neutral genetic variations should plan to utilize multiple plots to replicate the design in order to test for genome-microenvironment association and to identify the selective agents and factors responsible for the spatial patterns of non-neutral genetic variation in different cohorts.

Data availability

The sequences of *FcMYB1603* and the draft genome reported in this study were deposited in GenBank with accession numbers LC497503—LC497669, DRA008930, and BKZX01000001—BKZX01052213. Genotype data, spatial coordinates of individual trees and environmental variables are available from the Data Set file.

Acknowledgements The authors are grateful to the members of the Laboratory of Forest Conservation Ecology, Mie University, for assistance in the field, and to three anonymous reviewers for their useful comments. Support by Ann Sasabe is also grateful. This work was supported by JSPS KAKENHI Grant Number JP17K07838 and the Sumitomo Foundation Fiscal Grant for Environmental Research Projects (No. 173299). Some computations were performed on the NIG supercomputer at ROIS National Institute of Genetics.

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

Agarwal M, Hao YJ, Kapoor A, Dong CH, Fujii H, Zheng XW et al. (2006) A R2R3 type MYB transcription factor is involved in the

- cold regulation of CBF genes and in acquired freezing tolerance. *J Biol Chem* 281:37636–37645
- Ahrens CW, Rymer PD, Stow A, Bragg J, Dillon S, Umbers KDL et al. (2018) The search for loci under selection: trends, biases and progress. *Mol Ecol* 27:1342–1356
- Akaji Y, Hirobe M, Miyazaki Y, Makimoto T, Kinoshita S, Hattori I et al. (2017) Survival and growth of *Fagus crenata* seedlings in relation to biological and microtopographical factors in a cool temperate broadleaf forest. *J For Res* 22:294–302
- Akima H, Gebhardt A (2016) akima: interpolation of irregularly and regularly spaced data. R package version 0.6-2. <https://CRAN.R-project.org/package=akima>
- Asuka Y, Tani N, Tsumura Y, Tomaru N (2004a) Development and characterization of microsatellite markers for *Fagus crenata* Blume. *Mol Ecol Notes* 4:101–103
- Asuka Y, Tomaru N, Nisimura N, Tsumura Y, Yamamoto S (2004b) Heterogeneous genetic structure in a *Fagus crenata* population in an old-growth beech forest revealed by microsatellite markers. *Mol Ecol* 13:1241–1250
- Barberis IM, Batista WB, Pire EF, Lewis JP, Leon JC (2002) Woody population distribution and environmental heterogeneity in a Chaco forest, Argentina. *J Veg Sci* 13:607–614
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J R Stat Soc Ser B-Stat Methodol* 57:289–300
- Bivand R, Piras G (2015) Comparing implementations of estimation methods for spatial econometrics. *J Stat Softw* 63:1–36
- Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MHH et al. (2009) Generalized linear mixed models: a practical guide for ecology and evolution. *Trends Ecol Evol* 24:127–135
- Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N et al. (1999) Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet* 22:231–238
- Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Mol Ecol* 22:3124–3140
- Csillery K, Lalague H, Vendramin GG, Gonzalez-Martinez SC, Fady B, Oddou-Muratorio S (2014) Detecting short spatial scale local adaptation and epistatic selection in climate-related candidate genes in European beech (*Fagus sylvatica*) populations. *Mol Ecol* 23:4696–4708
- Davis MW (2017) ApE: a plasmid Editor v2.0.53. <http://biologylabs.utah.edu/Jorgensen/wayned/ape/>
- Denekamp M, Smeekens SC (2003) Integration of wounding and osmotic stress signals determines the expression of the AtMYB102 transcription factor gene. *Plant Physiol* 132:1415–1423
- Dray S, Bauman D, Blanchet G, Borcard D, Clappe S, Guenard G et al. (2016) adespatial: multivariate multiscale spatial analysis. R package version 0.2-0. <https://CRAN.R-project.org/package=adespatial>
- Dray S, Legendre P, Peres-Neto PR (2006) Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecol Model* 196:483–493
- Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L (2010) MYB transcription factors in *Arabidopsis*. *Trends Plant Sci* 15:573–581
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797
- Forester BR, Jones MR, Joost S, Landguth EL, Lasky JR (2016) Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. *Mol Ecol* 25:104–120
- Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schmitt J, Wilczek AM (2011) A map of local adaptation in *Arabidopsis thaliana*. *Science* 334:86–89
- Gabriel KR, Sokal RR (1969) A new statistical approach to geographic variation analysis. *Syst Biol* 18:259–278
- Golldack D, Li C, Mohan H, Probst N (2014) Tolerance to drought and salt stress in plants: unraveling the signaling networks. *Front Plant Sci* 5:151
- Hanaoka S, Yuzurihara J, Asuka Y, Tomaru N, Tsumura Y, Kakubari Y et al. (2007) Pollen-mediated gene flow in a small, fragmented natural population of *Fagus crenata*. *Can J Bot* 85:404–413
- Hardy OJ, Vekemans X (2002) SPAGEDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol Ecol Notes* 2:618–620
- Harms KE, Condit R, Hubbell SP, Foster RB (2001) Habitat associations of trees and shrubs in a 50-ha neotropical forest plot. *J Ecol* 89:947–959
- Hartl DL, Clark AG (2007) Principles of population genetics. Sinauer Associates, Inc., Sunderland, U.S.A
- Hijmans RJ (2017) raster: geographic data analysis and modeling. R package version 2.6-7. <https://CRAN.R-project.org/package=raster>
- Inanaga M, Nakanishi A, Torimaru T, Nishimura N, Tomaru N (2014) Distance-dependent but genetically random mating in a Japanese beech (*Fagus crenata*) population. *Botany* 92:795–803
- Ingvarsson PK (2008) Multilocus patterns of nucleotide polymorphism and the demographic history of *Populus tremula*. *Genetics* 180:329–340
- Ito M, Shiratake T (1983) Kaetsusannchi Gankyoushan syuhen no chikei chishitsu - Karikomiike heitanchi no keiseikatei to Kogankyoushiki no kousatsu [Topography and geology around Mt. Gankyoushi in Kaetsu Mountains. - Discussion on the process shaping the flat surface on Karikomiike pond and Kogankyoushiki pond -]. *Bull Fukui City Mus Nat Hist* 30:9–18 (in Japanese)
- Japan National Land Agency (1988) Soil map: Echizenkatsuyama - Hakusan (in Japanese)
- Katiyar A, Smita S, Lenka SK, Rajwanshi R, Chinnusamy V, Bansal KC (2012) Genome-wide classification and expression analysis of MYB transcription factor families in rice and *Arabidopsis*. *BMC Genom* 13:544
- Krajmerova D, Hrivnak M, Ditmarova L, Jamnicka G, Kmet J, Kurjak D et al. (2017) Nucleotide polymorphisms associated with climate, phenology and physiological traits in European beech (*Fagus sylvatica* L.). *New Forests* 48:463–477
- Kruskal JB (1964) Nonmetric multidimensional scaling: a numerical method. *Psychometrika* 29:115–129
- Kutnar L, Nagel TA, Kermavnar J (2019) Effects of disturbance on understory vegetation across Slovenian forest ecosystems. *Forests* 10:1048
- Leck MA, Simpson RL, Parker VT (2008) Why seedlings? In: Leck MA, Parker VT, Simpson RL (eds) Seedling ecology and evolution. Cambridge University Press, Cambridge, UK, p 3–12
- Leonhardt N, Kwak JM, Robert N, Waner D, Leonhardt G, Schroeder JI (2004) Microarray expression analyses of *Arabidopsis* guard cells and isolation of a recessive abscisic acid hypersensitive protein phosphatase 2C mutant. *Plant Cell* 16:596–615
- Liang J, He JX (2018) Protective role of anthocyanins in plants under low nitrogen stress. *Biochem Biophys Res Commun* 498:946–953
- Linhart YB, Grant MC (1996) Evolutionary significance of local genetic differentiation in plants. *Annu Rev Ecol Syst* 27:237–277
- Loiselle BA, Sork VL, Nason J, Graham C (1995) Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *Am J Bot* 82:1420–1425
- Makie G, Kisanuki H, Torimaru T (2017) Hakusan kokuritsu kouen Karikomiike ni okeru buna tennenrin no gunsyuukouzou [Community structure of a beech forest at Karikomiike pond in Hakusan national park]. *Chubu For Res* 65:25–27 (in Japanese)
- Manabe T, Nishimura N, Miura M, Yamamoto S (2000) Population structure and spatial patterns for trees in a temperate old-growth evergreen broad-leaved forest in Japan. *Plant Ecol* 151:181–197

- Matsuda S, Wakamatsu N, Jouyu H, Makita H, Akada S (2011) An extensive analysis of *R2R3-MYB* regulatory genes from *Fagus crenata*. *Tree Genet Genomes* 7:307–321
- Metcalfe P, Beven K, Freer J (2018) dynatopmodel: Implementation of the dynamic TOPMODEL hydrological model. R package version 1.2.1. <https://CRAN.R-project.org/package=dynatopmodel>
- Mukai Y (2008) Buna no jyuhun no bunshiseitaigaku: jikahuwagousei to kinkoujyakusei [Molecular ecology of beech pollination: self incompatibility and inbreeding depression]. In: Terazawa K, Koyama H (eds) Applied ecology for restruction of beech forests. Bun-ichi Sogo Shuppan Co., Tokyo (in Japanese)
- Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res* 8:4321–4325
- Oddou-Muratorio S, Bontemps A, Klein EK, Chybicki I, Vendramin GG, Suyama Y (2010) Comparison of direct and indirect genetic methods for estimating seed and pollen dispersal in *Fagus sylvatica* and *Fagus crenata*. *For Ecol Manag* 259:2151–2159
- Oksanen JF, Blanchet G, Friendly M, Kindt R, Legendre P, McGlinn D et al. (2018) vegan: Community Ecology Package. R package version 2.5-2. <https://CRAN.R-project.org/package=vegan>
- Paz-Ares J, Ghosal D, Wienand U, Peterson PA, Saedler H (1987) The regulatory c1 locus of *Zea mays* encodes a protein with homology to myb proto-oncogene products and with structural similarities to transcriptional activators. *Embo J* 6:3553–3558
- Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *Plos One* 7:e37135
- R Development Core Team (2018) R v.3.5.2: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
- Radula MW, Szymura TH, Szymura M (2018) Topographic wetness index explains soil moisture better than bioindication with Ellenberg's indicator values. *Ecol Indic* 85:172–179
- Rousset F (2008) GENEPOP '007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Mol Ecol Resour* 8:103–106
- Sasaki T, Koyama A, Koyanagi T, Furukawa T, Uchida K (2015) Data analysis of plant community structure and diversity, 1st edn. Kyoritsu Shuppan Co., Ltd, Tokyo, Japan (in Japanese)
- Schafer M, Brutting C, Xu SQ, Ling ZH, Steppuhn A, Baldwin IT et al. (2017) NaMYB8 regulates distinct, optimally distributed herbivore defense traits. *J Integr Plant Biol* 59:844–850
- Senno T (1979) On structure of beech (*Fagus crenata* Blume) stands in Mt.Hakusan. *Sci Rep Kyoto Prefect University, Agriculture* 31:125–136 (in Japanese with English summary)
- Shin N, Nakamura F (2005) Effects of fluvial geomorphology on riparian tree species in Rekifune River, northern Japan. *Plant Ecol* 178:15–28
- Stracke R, Werber M, Weisshaar B (2001) The *R2R3-MYB* gene family in *Arabidopsis thaliana*. *Curr Opin Plant Biol* 4:447–456
- 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* 5:12
- Takahashi K, Uemura S, Suzuki JI, Hara T (2003) Effects of understory dwarf bamboo on soil water and the growth of overstory trees in a dense secondary *Betula ermanii* forest, northern Japan. *Ecol Res* 18:767–774
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28:2731–2739
- Torimaru T, Akada S, Ishida K, Narita M, Higaki D (2018) Species habitat associations in an old-growth beech forest community organised by landslide disturbances. *J For Res* 23:98–104
- Tsumura Y, Uchiyama K, Moriguchi Y, Kimura MK, Ueno S, Ujino-Ihara T (2014) Genetic differentiation and evolutionary adaptation in *Cryptomeria japonica*. *G3-Genes Genomes Genet* 4:2389–2402
- Vekemans X, Hardy OJ (2004) New insights from fine-scale spatial genetic structure analyses in plant populations. *Mol Ecol* 13:921–935
- Wagner HH, Chavez-Pesqueira M, Forester BR (2017) Spatial detection of outlier loci with Moran eigenvector maps. *Mol Ecol Resour* 17:1122–1135
- Wagner HH, Dray S (2015) Generating spatially constrained null models for irregularly spaced data using Moran spectral randomization methods. *Methods Ecol Evol* 6:1169–1178
- Wang JJ, Bradburd GS (2014) Isolation by environment. *Mol Ecol* 23:5649–5662
- Weir B, Cockerham C (1984) Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358–1370
- Wilkins O, Nahal H, Foong J, Provart NJ, Campbell MM (2009) Expansion and diversification of the *Populus* R2R3-MYB family of transcription factors. *Plant Physiol* 149:981–993
- Yamamoto S, Nishimura N, Matsui K (1995) Natural disturbance and tree species coexistence in an old-growth beech - Dwarf bamboo forest, southwestern Japan. *J Veg Sci* 6:875–886