Molecular Phylogenetics and Evolution 126 (2018) 162–172

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Inferring spatial patterns and drivers of population divergence of *Neolitsea sericea* (Lauraceae), based on molecular phylogeography and landscape genomics

Ya-Nan Cao[a], Ian J. Wang[b], Lu-Yao Chen[a], Yan-Qian Ding[a], Lu-Xian Liu[a], Ying-Xiong Qiu[a,*]

[a] Key Laboratory of Conservation Biology for Endangered Wildlife of the Ministry of Education, and College of Life Sciences, Zhejiang University, Hangzhou, Zhejiang 310058, China
[b] Department of Environmental Science, Policy, and Management, University of California, Berkeley, CA 94720, USA

ARTICLE INFO

Keywords:
Biogeographical modelling
Isolation by distance
Isolation by environment
Land-bridge island flora
Landscape genomic approach
Restriction site-associated DNA sequencing (RADseq)

ABSTRACT

The relative roles of geography, climate and ecology in driving population divergence and (incipient) speciation has so far been largely neglected in studies addressing the evolution of East Asia's island flora. Here, we employed chloroplast and ribosomal DNA sequences and restriction site-associated DNA sequencing (RADseq) loci to investigate the phylogeography and drivers of population divergence of *Neolitsea sericea*. These data sets support the subdivision of *N. sericea* populations into the Southern and Northern lineages across the 'Tokara gap'. Two distinct sublineages were further identified for the Northern lineage of *N. sericea* from the RADseq data. RADseq was also used along with approximate Bayesian computation to show that the current distribution and differentiation of *N. sericea* populations resulted from a combination of relatively ancient migration and successive vicariant events that likely occurred during the mid to late Pleistocene. Landscape genomic analyses showed that, apart from geographic barriers, barrier, potentially local adaptation to different climatic conditions appears to be one of the major drivers for lineage diversification of *N. sericea*.

1. Introduction

Island systems are important models for evolutionary biology because they provide convenient, discrete biogeographic units of study (Demos et al., 2016). Since Darwin's time, research on islands has continued to advance the understanding of the evolutionary process (Losos and Ricklefs, 2009). Although islands harbor a high proportion of endemic species, they are generally considered as less diverse in species richness compared to adjacent mainlands (Whittaker and Fernández-Palacios, 2007). Of the 25 global biodiversity hotspots, nine encompass islands or archipelagos (Myers et al., 2000). Nevertheless, island biota is under severe threat due to habitat loss, climate change and biological invasions (Myers et al., 2000; Kreft et al., 2008). Knowledge about the evolutionary processes and drivers that lead to population divergence and speciation of the island's unique biota is particularly essential for the understanding, conservation, and management of island biodiversity (Futuyma, 2009; Höglund, 2009; Nosil, 2012). Compared with oceanic island systems (Le Roux et al., 2014; Spurgin et al., 2014), this information is currently lacking in land-bridge island systems (Comes et al., 2008; Nakamura et al., 2010; Barley et al., 2015).

The East Asia's land-bridge island system (including archipelagos and offshore islands of Japan, eastern China and south Korea) is one of the most threatened biodiversity hotspots in the world (Qian and Ricklefs, 2000; Kubota et al., 2015), yet we are far from understanding the evolutionary processes and forces driving population divergence and speciation of endemic biota in the East Asia's land-bridge islands. The high diversity in this region is generally thought to have resulted from allopatric speciation via genetic drift (coupled with restricted genetic exchange) across islands generated by tectonic movements and sea-level changes linked to the Plio-Pleistocene glacial/interglacial cycles (Nakamura et al., 2010; Qiu et al., 2011). Such an allopatric model of non-adaptive species diversification essentially implies that mutation and random genetic drift, rather than habitat-mediated selection, will be the primary factors causing divergence of populations occupying ecologically similar habitats (Comes et al., 2008). The primary evidence of allopatric divergence models has been based on the discontinuous distribution of closely related species or populations (Coyne and Orr, 2004); however, this can be misleading because it has long ignored adaptive divergence due to different environmental backgrounds in the face of gene flow (Dieckmann and Deobeli, 1999; Gavrilets, 2003). In fact, it has been recognized that floristic patterns of the Japan-Ryukyu-

* Corresponding author.
E-mail address: qyxhero@zju.edu.cn (Y.-X. Qiu).

https://doi.org/10.1016/j.ympev.2018.04.010
Received 29 October 2017; Received in revised form 2 March 2018; Accepted 9 April 2018
Available online 17 April 2018
1055-7903/

Taiwan islands are profoundly influenced by a latitudinal gradient of temperature of the coldest month (Schwind, 1967; Hübel, 1988). It is feasible, therefore, that, in addition to geographical isolation, spatially heterogeneous environments may potentially promote diversifying selection and local adaptation and, eventually, speciation.

Populations can be genetically differentiated both by geographic distance via spatially autocorrelated genetic drift through restricted gene flow and landscape connectivity, resulting in a correlation between genetic and geographic distances (IBD) (Wright, 1943), and by differences in their environments that reduce the rate of successful dispersal and gene flow (Bradburd et al., 2013; Wang and Bradburd, 2014), resulting in a correlation between genetic and environmental distances (IBE) (Wang, 2013). Traditionally, research on population divergence has focused on developing and applying approaches to disentangle the effects of geographic variables on gene flow (Richards-Zawacki, 2009; Spear and Storfer, 2010; Li et al., 2012); however, although several recent studies have proven that IBE plays a decisive role in spatial genetic divergence (Crispo et al., 2006; Nosil, 2012), patterns of IBE and analyses of how dispersal among populations inhabiting different environments is limited by local adaptation and divergent natural selection have received much less attention (Wang et al., 2013). Moreover, of the few investigations into IBE, the majority used only a few genotypic markers (tens to hundreds of microsatellites or AFLPs), precluding inference of the mechanisms driving selection (Manthey and Moyle, 2015), and only a handful of studies quantified IBE using massive datasets of genome-wide single-nucleotide polymorphisms (SNPs) (Lexer et al., 2014; Manthey and Moyle, 2015; Szulkin et al., 2016). Nevertheless, examining and quantitatively assessing the relative contributions of both geographic and ecological variables to the reduction of gene flow and differentiation among populations is of crucial importance for gaining a better understanding of how landscapes shape patterns of genetic isolation in nature (Crispo et al., 2006; Lee and Mitchell-Olds, 2012). Fortunately, the development of genome-scale genotyping approaches, such as RADseq, allows tens to hundreds of thousands of molecular genetic (DNA) markers to be studied in natural populations of non-model species (Baird et al., 2008; Wang et al., 2013). This method is providing unprecedented, genome-wide insights into the evolutionary drivers of species diversification at landscape scales (Nosil, 2012; Wang et al., 2013). Moreover, the rise of a series of modern spatial statistical frameworks and the increasing availability of high-resolution geographic and environmental data layers now makes it possible to precisely describe geographical and ecological landscapes and to synchronously assess the impacts of IBD and IBE on spatial patterns of genetic divergence (Wang et al., 2013). Certainly, other processes and variables beyond geographic and environmental distances can also influence rates of genetic divergence, particularly under scenarios with ongoing gene flow (Pinho and Hey, 2010), but disentangling the effects of IBD and IBE is a significant starting point for landscape genetic analyses and is the focus of this study.

Neolitsea sericea (Blume) Koidz. (Lauraceae) is a diploid, dioecious, and insect-pollinated tree (up to 15 m tall) of coastal, warm-temperate evergreen (WTE) forests, with red, one-seeded fruits (drupes) that are primarily dispersed by migratory birds (Yumoto, 1987; Arai and Kamitani, 2005). The species is widely distributed from Taiwan (Lanyu Island, China) across the Ryukyu Archipelago into northeastern Japan but also occurs disjunctively on offshore islands of eastern China (Zhoushan Archipelago) and in a few coastal localities of the southern Korean Peninsula (Yumoto, 1987). Based on the IMa analysis of our combined data set [10 nuclear microsatellites (nSSRs) and one chloroplast (cp) DNA sequence marker (*psb*A–*trn*H intergenic spacer)], we dated two distinct lineages located in areas north and south of the 'Tokara gap' (a long-standing deep tectonic strait) to the late Pleistocene (*c.* 0.07 Ma; 90% highest posterior density interval: 0.02–0.38 Ma). Moreover, we hypothesized that apart from geographical barriers, natural selection in divergent environments might have played a role in this north–south divergence in *N. sericea* (Honda et al., 2008; Zhai et al., 2012). However, estimation of the divergence time based on few genetic loci may be inaccurate. To infer the demographic history more precisely and in greater detail, it will be necessary to evaluate more complex demographic models using a larger number of loci and alternative methods, such as approximate Bayesian computation (Bertorelle et al., 2010). In addition, although this species represents an ideal model to elucidate the relative roles of geography, climate and ecology in the initial stages of species formation, it remains largely unknown how these factors contribute to genetic divergence through time.

In this study, we use a combination of Sanger sequencing of cpDNA and internal transcribed spacer (ITS) of nuclear ribosomal DNA regions and RADseq data to analyze range-wide phylogeographic structure and determine the potential drivers of spatial genetic variation of *N. sericea*. In particular, we addressed the question of whether or not RADseq data from many loci across the genome for few individuals can improve the resolution of phylogeographic studies based on cpDNA and ITS data from many individuals. Specifically, based on a subset of RAD-SNPs, we applied an approximate Bayesian computation (ABC) procedure to test competing scenarios of population divergence history that has potentially contributed to the contemporary distribution of the species. In addition, we used landscape genomic approaches to examine the relative contributions of geography and ecology to genetic divergence. Finally, we used a Bayesian approach to examine the genomic distribution of divergence for RADseq polymorphisms and identify divergent outlier loci that may be associated with local adaptation through divergent natural selection in the heterogeneous environments of the East Asian land-bridge islands.

2. Materials and methods

2.1. Plant materials

Leaf material was obtained from 61 populations (n = 605) of *N. sericea* throughout its range across the Zhoushan Archipelago (eastern China), Lanyu Island (Taiwan, China), the Korean Peninsula, the Ryukyu Archipelago and main islands of Japan (i.e. Kyushu, Shikoku, Honshu) (Table S1 and Fig. S1). Two related species [i.e. *N. aciculata* (Bl.) Koidz., *N. levinei* Merr.] were variously selected as outgroups for the analyses, based on previous molecular phylogenetic studies (Chanderbali et al., 2001; Li et al., 2007).

2.2. DNA isolation, Sanger sequencing and phylogeographic analyses

Total genomic DNA was extracted from the silica-dried leaf tissue. For the phylogeographic DNA survey, three intergenic spacer (IGS) regions (*psb*A–*trn*H, *psb*C–*trn*S, *trn*L–*trn*F) and the *rps*16 intron of cpDNA, were sequenced for all *N. sericea* samples. We also sequenced a subset of the samples (n = 265) for the entire ITS region. All of these regions were also sequenced for the outgroup (*N. levinei*). Sequences were edited, assembled, and aligned in GENEIOUS version 4.8.5 (Drummond et al., 2009). All cpDNA and ITS haplotype sequences identified in the present study were deposited in GenBank (see Table S3 for accession numbers).

For both cpDNA and ITS, we calculated the number of haplotypes, haplotype diversity (h), and nucleotide diversity (π) for each population using DNASP version 5.1 (Librado and Rozas, 2009). Genealogical relationships among haplotypes were reconstructed by a statistical parsimony network under the 95% criterion implemented in TCS version 1.2.1 (Clement et al., 2000). For cpDNA sequences, the network was rooted on *N. levinei*. However, for ITS sequences, an unrooted haplotype network was estimated because the sequence of outgroup (i.e. *N. levinei*) was too diverged (> 20 steps limitation of TCS network at the parsimony criterion of 95%) from those of ingroup to infer the most likely ancestral haplotypes (Cassens et al., 2003). To infer the phylogenetic relationships among cpDNA haplotypes, maximum-likelihood

(ML) analysis was conducted in RAXML-HPC (Miller et al., 2010) on the CIPRES cluster (http://www.phylo.org/), using the general time-reversible (GTR) model with gamma distributed rate heterogeneity. Relationships between either h or π of all populations and latitude were tested using single-variable regression implemented in R (R Core Team, 2013) with the 'ggplot2', 'ISwR' and 'Scales' packages (See more details in Appendix method S1).

2.3. RADseq data acquisition and analyses

Empirical comparisons between the results from microsatellites and RADseq for the phylogeography showed the RADseq approach with fewer samples but many more loci can infer range-wide genetic structure in agreement with the microsatellite data set, but has much greater power to detect fine-scale population structure (Jeffries et al., 2016). In this study, we thus selected 84 individuals from 27 populations (1–5 samples per population) for RADseq analysis (Fig. S1 and Table S1). These samples were chosen to represent a wide geographic range and all major phylogeographic clusters identified using the microsatellite data (Zhai et al., 2012). Data from *N. levinei* and *N. aciculata* were included as outgroups. Following Zhang et al. (2015), RADseq libraries were prepared for each DNA sample (in total 86 samples) and then sequenced by Novogene Bioinformatics Technology Co. Ltd. (Beijing, China) using the restriction enzyme *Eco*R1 and sample-specific barcodes. The individuals in the library were pooled and ran in a single lane on an Illumina Hiseq 2500 to generate paired-end reads.

Raw sequence data were processed in the software PYRAD version 3.0 (Eaton and Ree, 2013), which distinguishes sequencing errors within samples, identifies putative orthologous loci across samples, and assembles formatted data files. PYRAD can de-multiplex one or more Illumina sequence files in FASTQ format and assign sequence reads to each sample based on their sample-specific barcodes. Sequence files for analysis were prepared through three quality checks. The cut sites and barcodes were then trimmed from each sequence. *Neolitsea sericea* does not have a reference genome against which to align sequence reads; instead, we employed a multistep method to identify RAD loci within samples, assign a consensus sequence to each sample and assign consensus sequences across samples (See more details in Appendix method S2). To avoid the effect of missing data and ensure enough SNPs for analyses, we generated three "minimum-taxa" datasets, loci with a minimum ingroup sample coverage > 35, 50 and 70 were included in the final 'maximum', 'median' and 'minimum' datasets, respectively.

2.4. Population genomic and phylogenetic analyses

Bayesian clustering of individuals was conducted using the program STRUCTURE version 2.3.4 (Pritchard et al., 2000) for the three *N. sericea* datasets ('maximum', 'median' and 'minimum'). We utilized an admixture model with independent allele frequencies to determine the optimal number of clusters (K) from 1 through 27. For each K, five independent simulations were conducted with a burn-in of 20,000 followed by 20,000 replications. Following Evanno et al. (2005), the most probable value of K was determined by estimating the log probabilities (LnP) of the data (D) LnP(D) (Pritchard et al., 2000) and ΔK [i.e. the second-order rate of change of LnP(D) between successive K values]. In addition, based on the three datasets, population structure was also detected by discriminant analysis of principal component (DAPC) for *N. sericea*, using 'Adegenet' package (Jombart, 2008) in R. ML analyses were implemented in RAXML-HPC for the three concatenated "minimum-taxa" SNP supermatrices, with missing data coded as 'N's, using the GTR model.

2.5. Tests of population divergence history by ABC modelling

DIY-ABC version 2.1.0 (Cornuet et al., 2014) was used to gain further insights into the timing and divergence history of *N. sericea* based on 4,100 SNPs randomly selected from the 'minimum' dataset. We tested five plausible divergence scenarios on the basis of the genetic structure identified by the ML tree and STRUCTURE results (Figs. 2 and 3): the simultaneous divergence of the three groups (i.e. South, North1, North2) from a common ancestor (Fig. S2, scenario 1) against three alternative models, reflecting all possible relationships among these groups (see scenarios 2–4 in Fig. S2); and an admixture model, in which South and North1 groups diverged from an ancestral population at time t_0, and then an admixture event between them at time t_1 gave birth to North2 group with admixture rate ra (scenario 5 in Fig. S2) (See more details in Appendix method S3). $t_{\#}$ refers to divergence time scaled by generation time and $N_{\#}$ refers to effective population size of the corresponding populations. Estimation of generation time is still controversial in forest tree species, because of their longevity, overlapping generations, variable age of first flowering, and forest replacement dynamics (Petit and Hampe, 2006). To estimate the divergence times in millions of years ago, a conservative generation time of 10–30 years for *N. sericea* was assumed on the basis of field observations (G.M. Wang, personal communication).

2.6. Present and past distribution modelling

Ecological niche modelling (ENM) was carried out in MAXENT version 3.3.1 (Phillips et al., 2006) to predict suitable climate envelopes for *N. sericea* at the present and the Last Glacial Maximum (LGM: *c.* 21 Kya before present; BP), respectively. We designed the current species range (20–40°N and 115–145°E) as the study area. A current distribution model was developed using six out of 19 bioclimatic data layers (annual mean temperature, annual precipitation, precipitation of wettest, driest, warmest and coldest quarter) available from the WorldClim database (http://www.worldclim.org) at a 2.5 arc-min resolution for the present (1950–2000). The established model was then projected onto the set of climatic variables simulated by the Community Climate System Model (CCSM) version 3.0 (Collins et al., 2006) to infer the extent of suitable habitat during the LGM (See more details in Appendix method S4). In addition, the average logistic values between present and CCSM-LGM layers were also calculated to represent long-term suitable areas since the LGM for this species.

2.7. Generalized dissimilarity modelling of genomic, spatial and environmental data

Generalized Dissimilarity Modelling (GDM) uses a non-linear extension of matrix regression to model spatial patterns of pairwise biological dissimilarity (genetic composition) between sites against pairwise site differences in geographic and environmental variables (Fitzpatrick and Stephen, 2015). It uses a unique monotonic I-spline turnover function for each predictor (environmental and geographic variables) to quantify: (1) variation in the rate (the shape of each spline) of genetic turnover at different localities along each environmental and geographic gradient while keeping all other variables constant, and (2) the curvilinear relationships between genetic distance and ecological and geographic distances (Fitzpatrick and Stephen, 2015). The maximum height of each spline indicates the magnitude of total genetic change along that predictor, thus corresponding to the relative importance of that variable. GDM can assemble and evaluate any number of SNPs within a genome across many sampling locations and takes a site-by-SNP matrix and a corresponding site-by-environment predictor matrix as inputs (Fitzpatrick and Stephen, 2015) (See more details in Appendix method S5). Genetic differentiation between locations (grid cells) based on all putatively neutral SNPs from the 'minimum' dataset (i.e. 50,248 SNPs; see Results) was mapped by assigning the first three PC axes to an RGB colour palette. Greater differences between the colours in these grid cells indicate greater predicted genetic differentiation (Manion et al., 2016).

2.8. IBD and IBE analyses

Multiple Matrix Regression with Randomization (MMRR) framework (Wang, 2013) is a linear regression-based method, which was used to quantify the effects of multiple explanatory variables (here, environmental and geographic distances) on a single response variable, in this case, genetic distance (F_{ST}). We used pairwise genetic distances (F_{ST}) between populations based on 50,248 neutral SNPs from the 'minimum' dataset as calculated above. We extracted environmental variables from the 19 bioclimatic data layers to obtain an environmental dissimilarity matrix by calculating population pairwise Euclidean distances. Direct pairwise geographic distances between populations were calculated using the 'gdistance' package (Wang et al., 2013) in R. We then used the 'MMRR' function (Wang, 2013) in R to quantify the effects of ecology and geography on genetic divergence after standardizing all distance matrices. The simulations were run using 10,000 permutations.

2.9. Bayesian genome scan for potential outlier loci and associations between allele frequencies and environmental variables

Increasing amounts of missing data in the median' and 'maximum' SNP matrices would lead to biased estimates of F_{ST} (Hodel et al., 2017), hence, only 'minimum' dataset was included in landscape genomic analyses. We identified RAD-SNPs potentially under selection using BAYESCAN version 2.1 (Foll and Gaggiotti, 2008). As a fully Bayesian approach, BAYESCAN implements a reversible-jump MCMC algorithm to estimate posterior probabilities for a neutral model and a model with selection. We ran the software with default settings, except for a modification to increase the prior odds on the neutral model to 250 (with the neutral model being 250 times more likely than the model with selection), as recommended for data sets of this size (Foll and Gaggiotti, 2008). A locus was considered to be under divergent or purifying selection when the Bayes Factor (BF) comparing the two models was greater than 100. Greater differentiation than expected under neutrality indicated that a locus experienced divergent selection, while lower differentiation than expected under neutrality indicated purifying selection. The loci bearing signatures of selection were segregated into an outlier data set, and the remaining loci (with outliers removed) constituted the neutral data set. We then tested for associations between allele frequencies at loci putatively under divergent selection and environmental variables using two different methods: Multiple Linear Regression (MLR) analysis (Zulliger et al., 2013) in R and General Linear Mixed Model (GLMM) analysis as implemented in the 'spaMM' package (Rousset and Ferdy, 2014) (Spatial Mixed Model) in R (See more details in Appendix method S6).

3. Results

3.1. CpDNA and ITS phylogeography

The four concatenated cpDNA regions sequenced in all 605 individuals of *N. sericea* had a total length of 1745 bp, including 11 single-site mutations, 6 indels (i.e. insertions-deletions; 5–52 bp) and one inversion (8 bp; Table S4). Together, these 18 polymorphisms yielded 20 haplotypes ('chlorotypes', H1–20), of which 5 were specific to populations residing in areas south of the 'Tokara gap' (H1–5; hereafter 'Southern') and 14 to those dwelling in areas north of this gap (H7–20; hereafter 'Northern'), while H6 was shared among populations in both regions (Fig. 1a). The parsimony network (Fig. 1b) identified H7 and H11 from the northern region, and widespread H6 as the most likely ancestral ('interior') chlorotypes (Fig. 1a). However, due to insufficient phylogenetic signals available at the four cpDNA regions, the maximum likelihood analyses of 20 chlorotypes yielded a less-resolved tree with polytomy (Fig. S3). Regression analysis showed that h tended to significantly decrease with increasing latitude ($r = -0.05$, $P < 0.05$; Fig. S4a), but no significant correlation was found between π and latitude ($r = -0.009$, $P = 0.475$; Fig. S4b).

The total alignment of the ITS region surveyed across 265 individuals of *N. sericea* was 716 bp in length, including two single-site variants. These two polymorphisms identified 3 ribotypes (Ra-c; Table S5, Fig. 1a,c). Among the three ribotypes, Rb and Rc were specific to the southern and northern regions, respectively. However, the most frequent Ra was found in the northern region where it occurred in all populations except K1, and the southern region (R2, R3, T1). In the regression analysis, both h and π were independent of latitude ($P = 0.345$ and 0.346, respectively; Figs. S4c,d).

3.2. RADseq data sets

Following quality filtering and clustering (89% similarity), the Illumina sequencing-derived raw reads (averaged 6.20×10^6 per sample) were reduced to an average of 84,900 clusters (or "stacks") per sample, which ranged from 32,900 to 205,400 (Table S6). After quality filtering, mean, minimum and maximum number of sequencing coverage depth per sample were 17.78, 14.82 and 28.32, respectively. The number of consensus sequences called for each cluster was from 1.53×10^5 to 2.82×10^5 per sample with a mean value of 2.17×10^5 (Table S6). After clustering of consensus sequences across all 86 samples, the "minimum-taxa" data set yielded 62,978 loci, and a total of 126,017, 82,747 and 50,254 informative sites (unlinked SNPs) were identified for the 'maximum', 'median' and 'minimum' data sets, respectively.

3.3. Phylogenetic reconstruction and population structure

The topology of the ML tree based on 'median' data (Fig. 2a) was identical to that estimated from 'maximum' data, but was slightly different from that based on 'minimum' data (Fig. 2b). All topologies consistently supported the monophyly of *N. sericea* (BP = 100%). Within *N. sericea*, samples from areas north of the 'Tokara Gap', i.e. the Zhoushan Archipelago of eastern China, Korean Peninsula and main islands of Japan, were recovered as a well-supported Northern lineage (100%); the Northern lineage was further structured into two well-supported sublineages (each 100%), mainly comprised of samples from, respectively, the Zhoushan Archipelago, Korean Peninsula and Kyushu (N1 sublineage) and Shikoku/ Honshu (N2 sublineage; Fig. 2). The tree topology estimated from both 'median' and 'maximum' data sets (Fig. 2a) showed that the Northern lineage was nested within a paraphyletic grouping of samples from areas south of the 'Tokara Gap', i.e. the Lanyu Island of Taiwan (China) and the Ryukyu Archipelago. However, samples from Taiwan Island collapsed into a polytomy in the ML tree derived from 'minimum' data (Fig. 2b).

The STRUCTURE analysis provided the strongest support when samples were clustered into two groups ($K = 2$) and three groups ($K = 3$) based on 'minimum' dataset and other two "minimum-taxa" datasets, respectively, both when considering the probability of the LnP(D) and ΔK (Fig. S5). Based on the three datasets, at $K = 2$, the proportions of each individual in each population were assigned into two clusters (Northern vs Southern) with little admixture (Fig. 3a). These groups almost entirely corresponded to separate geographic regions, located either north or south of the 'Tokara gap'. At $K = 3$, samples from the Northern cluster were further subdivided into two 'gene pools', largely representing sublineages N1 vs. N2, excepting individuals from southwest Honshu, which showed evidence of extensive admixture (Fig. 3b). We validated Bayesian model-based clustering results from STRUCTURE analysis using DAPC method that is considered free of Hardy-Weinberg and linkage disequilibrium assumptions. According to the lowest associated BIC (Bayesian information criterion) (Fig. S6a), $K = 3$ represented the optimal number of clusters for all the three 'minimum-taxa' datasets. When applying $K = 3$, the results of the DAPC analysis were analogous to those of the STRUCTURE analysis with the exception of two populations

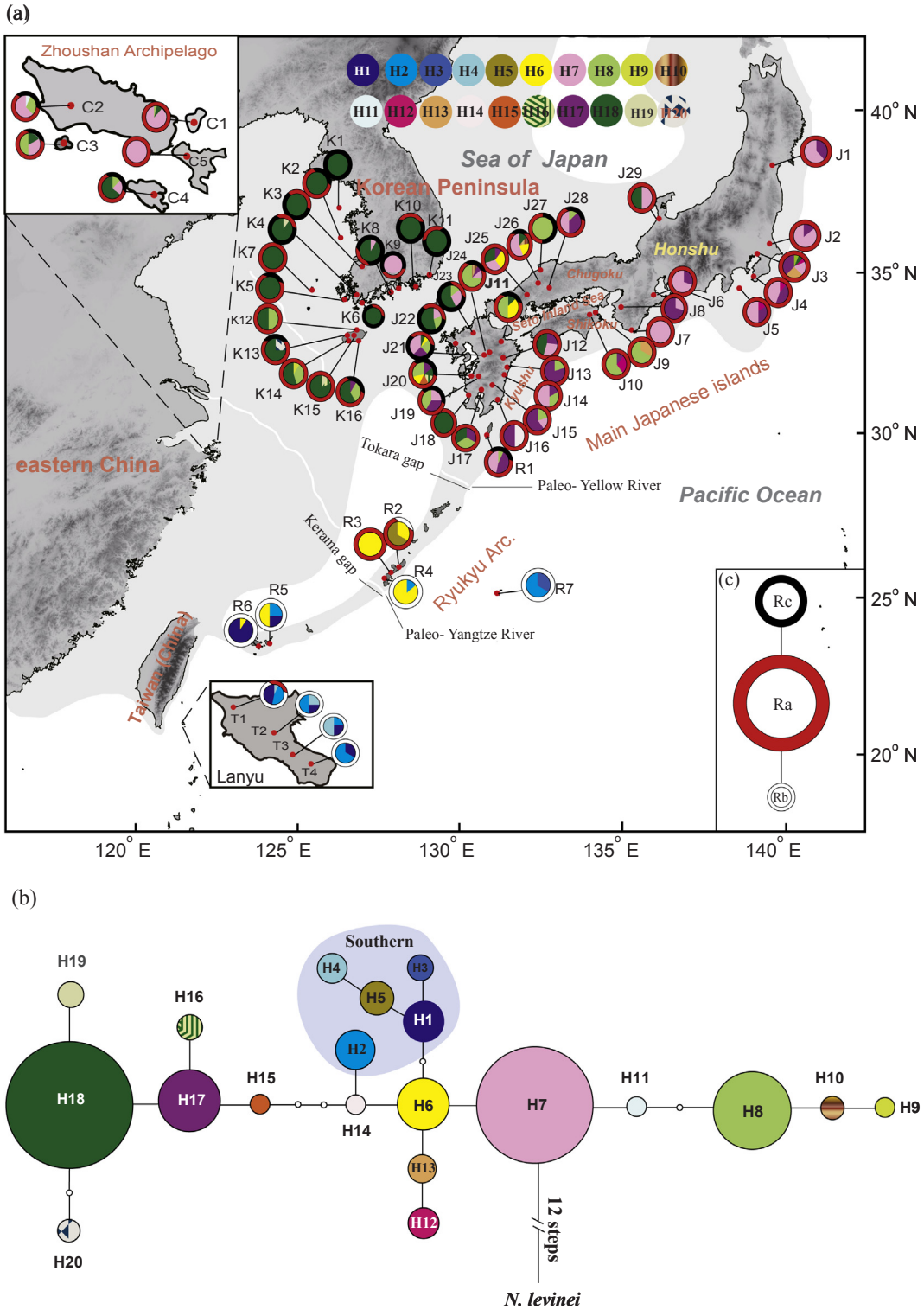

Fig. 1. (a) Outer and inner circles at sampling localities represent geographical distribution of 3 ribotypes and 20 chlorotypes across 61 populations of *N. sericea*, respectively. Dark-shaded areas indicate current mainland and island configurations, and light-shaded areas indicate palaeogeographic map of East Asia in the early Pleistocene (Kimura, 2000; Park et al., 2006; Osozawa et al., 2011). The black lines show the current locations of 'Tokara gap' and 'Kerama gap'. (b) TCS-derived network of 20 chlorotypes of *N. sericea* (H1–H20). The size of circles corresponds to the frequency of each chlorotype. The small open circle indicates an inferred intermediate chlorotypes not detected in this investigation. (c) TCS-derived network of 3 ribotypes of *N. sericea* (Ra-c). The size of circles corresponds to the frequency of each ribotype. The baseline map was created by us using ARCGIS version 10.2.2 (http://www.esri.com/software/arcgis/arcgis-for-desktop).

(J25 and J26), where individuals showed genetic admixture between the two sublineages in STRUCTURE (Fig. 3a,b), but were assigned to N1 group in DAPC (Fig. 3c, S6b). With $K = 2$, the results of the DAPC analysis were identical to those of the STRUCTURE analysis (Fig. S6c).

3.4. ABC-based inference of population divergence

In the ABC analysis, the most strongly supported scenario was scenario 5 with a posterior probability of 0.49 for the direct estimate

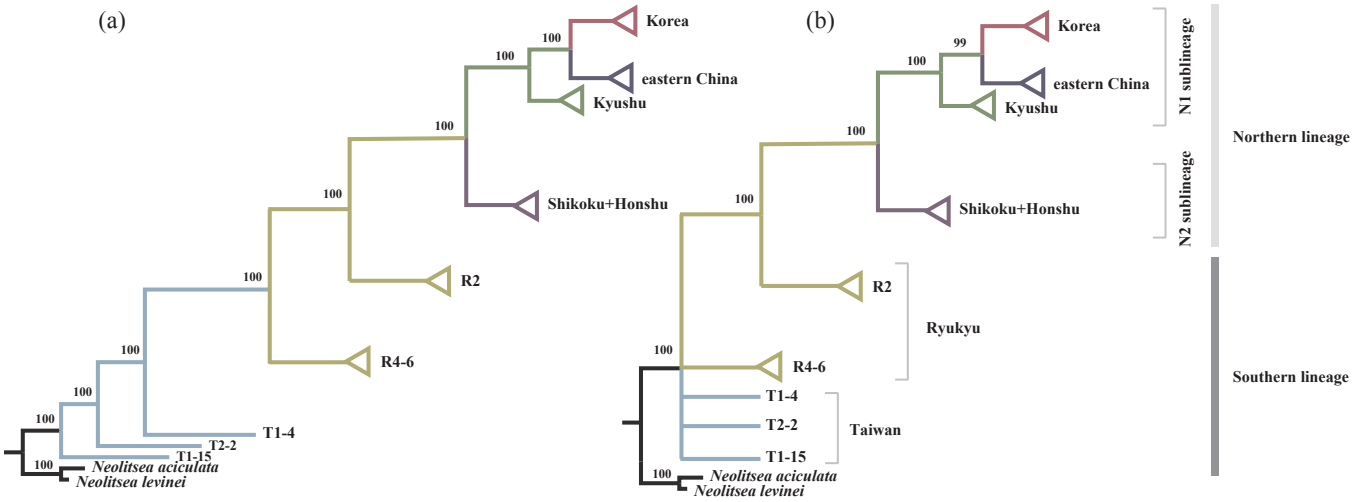

Fig. 2. Maximum likelihood (ML) phylogenetic tree inferred from the 'median' (a) and 'minimum' (b) RAD-SNP data for 84 *N. sericea* individuals and two outgroups (*N. aciculata* and *N. levinei*) using RAXML.

approach and 0.95 for the logistic regression approach (Fig. S7), in which N2 sublineage was assumed to result from admixture between the N1 sublineage and Southern lineage. However, the admixture rate (ra) was estimated to be 0 (Fig. S8), indicating that this admixture scenario is not compatible with our RAD-SNP data. For scenarios with population split without admixture, the DIY-ABC analysis supported scenario 2 as the most probable (posterior probability: 0.33 and 0.05 for the direct estimate approach and logistic regression approach, respectively; Fig. S7). Model testing analyses showed that the summary statistics from the observed data produced eigenvectors (PC1, PC2 and PC3) that were within the set of simulated data sets from the posterior predictive distribution based on scenario 2 (Fig. S9), indicating good model performance. According to this scenario, the Northern and Southern lineages split first at time t_0, followed by the divergence of the two Northern sublineages (N1 vs N2) at t_1. Assuming 10 years as generation time, divergence times were scaled to 0.145 Ma (95% CI: 0.069–0.236) and 0.065 Ma (95% CI: 0.036–0.09) (Table 1) for t_0 and t_1, respectively. Considering a longer generation time of 30 years, these values increased to 0.435 Ma (95% CI: 0.206–0.708) and 0.196 Ma (95% CI: 0.107–0.269) (Table 1).

3.5. Present and past distributions of N. sericea

MAXENT model had high predictive power [AUC = 0.928 ± 0.009 (mean ± SD)]. The current distribution of *N. sericea* (Fig. S1) is well captured by the climate-based envelope model (Fig. 4a), although there were also some predicted areas where this species does not occur at present, such as some lowland and hilly areas in Southeast China. Under the LGM (Last Glacial Maximum) climate, the species' distribution in the islands of the Ryukyu Archipelago changed little between the two periods (Fig. 4b), while shifted slightly southwestward along the Pacific Ocean side of the main islands of Japan. In addition, during the LGM, the most extensive stretches of hospitable habitat were predicted for the exposed East China Sea (ECS) basin, South Japan, the Ryukyu Archipelago, and Southeast China (Fig. 4b). The predicted long-term suitable areas for this species since the LGM were more or less similar to those of its present distribution, except for Southeast China where the species is not known to occur (Fig. 4c). Considering that this island endemic is a habitat specialist, failure to take local environmental factors into account when projecting distributions may result in overestimation of the potential climatic niche for *N. sericea*.

3.6. Modelling of genetic, spatial and climate niche data

GDM explained 62.5% of the deviance in spatial turnover of genetic composition, indicating a good fit of the model to the data. As for variable importance, the most prominent pattern was the strong relationship of genetic dissimilarity with geographic distance (GEO; importance value: 0.548), BIO3 [isothermality (BIO2/BIO7); importance value: 0.519] and BIO4 (temperature seasonality; importance value: 0.390). BIO8 (mean temperature of wettest quarter; importance value: 0.346) was the fourth most important variable, followed by BIO1 (annual mean temperature; importance value: 0.169) and BIO14 (precipitation of driest month; importance value: 0.166) (Fig. 5).

The I-spline for BIO14 had a spherical shape and indicated rapid change in allele frequencies for predictor dissimilarity value < 0.2 (standardized) and no turnover elsewhere along this gradient (i.e. increasing differences in BIO14 between sites led to increasing genetic differentiation only to a certain point). When mapped in space, the patterns of genetic composition of *N. sericea* predicted by GDM strongly revealed prominent genetic turnover in the central Ryukyu Archipelago across the 'Tokara gap' (Fig. 6). Moreover, progressive change in genetic composition from Kyushu to southwest Honshu and Shikoku and further northward to northeast Honshu along the coastal area of the Pacific Ocean was predicted in the main islands of Japan, while comparatively little genetic turnover was predicted elsewhere (Fig. 6).

3.7. IBD and IBE

The R^2 value for the MMRR simulation based on the neutral data was high (R^2 = 0.47, P = 0.0001), indicating good predictive model performance. This analysis suggested significant effects of both IBD and IBE on neutral SNP divergence in *N. sericea* with IBD explaining slightly more of the variation in genetic differentiation than IBE (β_D = 0.404, P = 0.0001; β_E = 0.382, P = 0.0001).

3.8. Detection of loci potentially under selection and environmental association analyses

A total of 85 SNPs were identified as potentially non-neutral outlier RAD loci with high posterior probabilities (> 99%), including 79 loci under purifying selection and 6 loci under divergent selection. Finally, the six loci putatively under selection (20610, 30713, 35238, 43150, 46,875 and 5422) were further confirmed as adaptive loci by the MLR analysis with R^2_{adj} > 0.5 (Table 2). When running linear regressions using each environmental variable individually, MLR and GLMM both

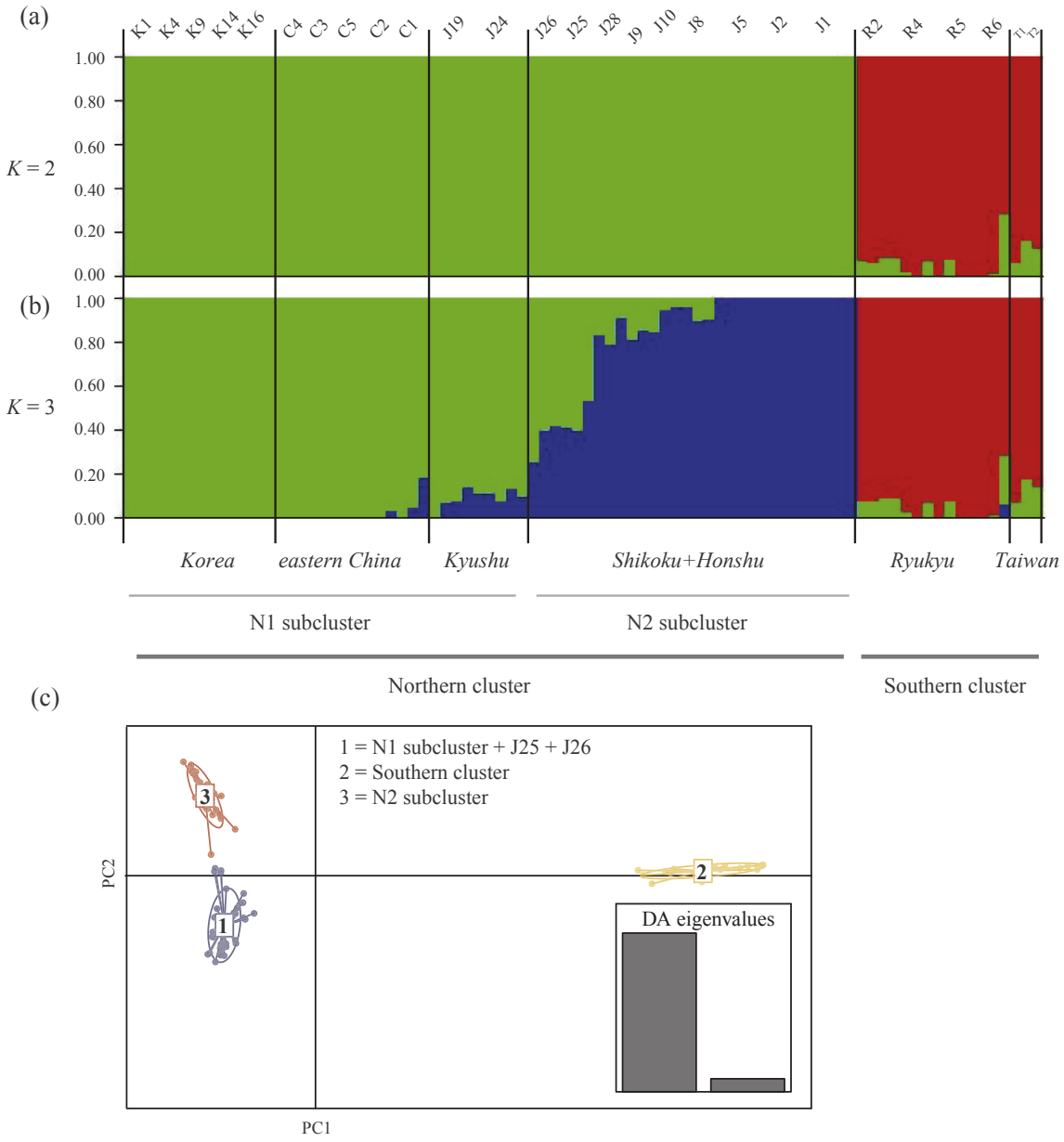

Fig. 3. Histogram of STRUCTURE analyses for the model with $K = 2$ (a) and $K = 3$ (b) (data only shown for the 'minimum' dataset). Each of the smallest vertical bars represents one individual. The assignment ratio of each individual into each of the clusters is shown along the y-axis. Each cluster is represented by a distinct colour. Note that STRUCTURE provided strongest support for $K = 2$ based on the 'minimum' data, and $K = 3$ based on the 'median' and 'maximum' data, both when considering the probability of the data [LnP(D)] and ΔK (Fig. S5). (c) Plots of the first two dimensions of discriminant analysis of principal component (DAPC) for *N. sericea* (data only shown for the 'minimum' dataset).

Table 1
Descriptions of prior settings and median estimates of posterior distributions for all parameters in the best-fitting scenario based on DIY-ABC.

	Parameters	Priors[a]	Posteriors		
			Median	95% lower bound	95% upper bound
Scenario 2	NA	(10–1,500,000)	148,000	88,800	291,000
	Na	(10–20,000)	9,320	3,040	17,900
	N(North1)	(10–20,000)	13,700	7,580	18,600
	N(North2)	(10–50,000)	33,600	18,800	46,200
	N(South)	(10–200,000)	131,000	75,700	183,000
	t_0 (in million years ago)	(10–40,000)	0.145–0.435	0.069–0.206	0.236–0.708
	t_1 (in million years ago)	(10–10,000)	0.065–0.196	0.036–0.107	0.09–0.269

[a] All priors are uniformly distributed. N(North1), N(North2), and N(South) denote the current effective population sizes of the N1 sublineage, N2 sublineage, and Southern lineage, respectively (see Fig. 2). NA is the effective population size of the common ancestor of the three groups. Na represents the effective population size of the common ancestor of the two Northern sublineages (N1 vs N2) from t_0 to t_1.

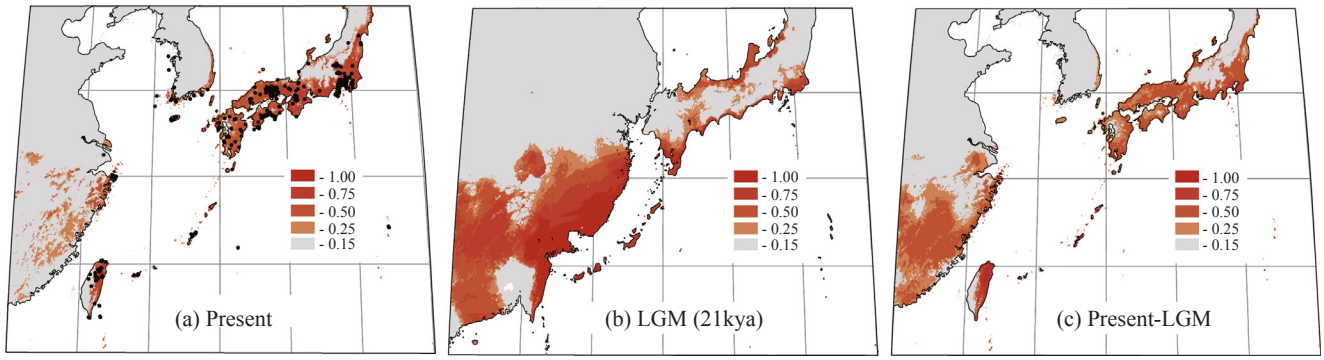

Fig. 4. Potential distributions as probability of occurrence for *N. sericea* in East Asia. (a) at present (1950–2000); (b) at the Last Glacial Maximum (LGM; *c.* 21 Kya BP) and (c) long-term suitable areas since the LGM. Black dots represent the extant occurrence points of the species. Predicted distribution probabilities (in logistic values) are shown in each 2.5 arc-min pixel. The baseline map was created by us using ARCGIS version 10.2.2.

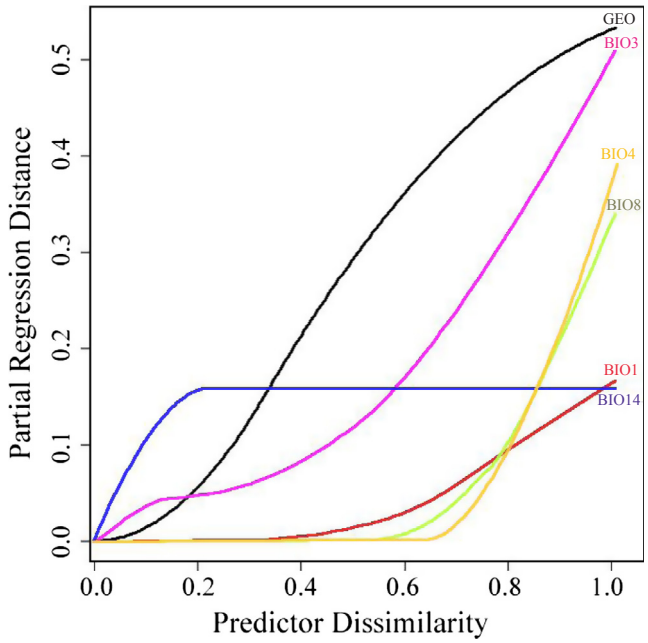

Fig. 5. GDM-fitted I-splines for each environmental predictor and geographic distance based on 50,248 neutral SNPs. The maximum height of each spline indicates the total amount of genetic turnover associated with that variable, while holding all other variables constant. The shape of each curve indicates how the rate of biological change in allele frequencies varies along the gradient. GEO, geographic distance.

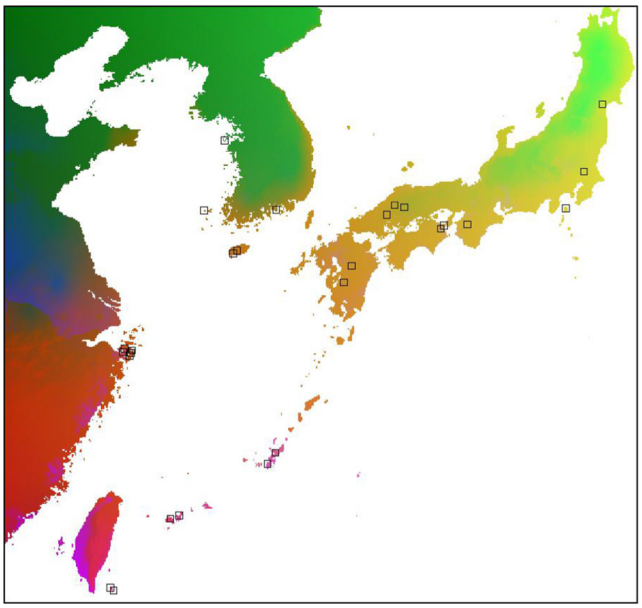

Fig. 6. Predicted spatial patterns of population-level genetic composition from GDM for 50,248 neutral SNPs. Black boxes indicate locations of the genotyped *N. sericea* populations. Colours represent genetic turnover gradients derived from transformed environmental predictors. Locations with similar colours are expected to contain populations with similar genetic composition. The baseline map was created by us using ARCGIS version 10.2.2.

revealed that all six loci were significantly ($P < 0.05$) correlated with BIO1 and BIO14. Additionally, MLR indicated that four loci (i.e. 20610, 30713, 35,238 and 43150) were also highly associated with BIO8 (Table 2).

Table 2
Results of outlier detection and environment-SNP association analyses on RADseq data of *N. sericea*.

Outlier ID	$R^2_{adj\ (MLR)}$	Significant environmental variables	
		MLR	GLMM
5422	0.702	BIO1**, BIO14*	BIO1***, BIO14**
20,610	0.921	BIO1***, BIO8*, BIO14***	BIO1***, BIO14***
30,713	0.921	BIO1***, BIO8*, BIO14***	BIO1***, BIO14***
35,238	0.921	BIO1***, BIO8*, BIO14***	BIO1***, BIO14***
43,150	0.921	BIO1***, BIO8*, BIO14***	BIO1***, BIO14***
46,875	0.852	BIO1***, BIO14**	BIO1**, BIO14*

* $P < 0.05$.
** $P < 0.01$.
*** $P < 0.001$.

4. Discussion

4.1. Phylogeography of N. Sericea

Consistent with previous results using 10 nuclear microsatellites (Zhai et al., 2012), our genomic phylogeography and landscape genomics support the subdivision of *N. sericea* populations into two major genetic groups (Southern vs Northern lineages), located south and north of the 'Tokara gap' (Fig. 2,3,6). However, the cpDNA/ITS sequences and microsatellite (Zhai et al., 2012) datasets found no evidence of any discernible population structure within each of the two major lineages of *N. sericea*, whereas two distinct sublineages (N1 vs N2) were identified for the Northern lineage of *N. sericea* from the RADseq data (Fig. 2,3,S5,S6). Therefore, our results lend support to previous empirical studies by showing that RAD sequencing is better able to detect fine-scale population structure than classical genetic

markers (Rašić et al., 2014; Jeffries et al., 2016), even with a substantially smaller sample size of individuals (14% of samples and 44% of populations, Table S1).

A hierarchical population split model (scenario 2) was found to be the most likely explanation for the observed data using DIY-ABC (Fig. S7), suggesting successive historical divergences of *N. sericea* populations. Based on the DIY-ABC analysis, the divergence between Northern and Southern lineages of *N. sericea* was dated from 0.145 Ma (generation time of 10 years) to 0.435 Ma (generation time of 30 years), i.e. falling into the mid to late Pleistocene. Although these estimates are a little earlier than the previous estimate (0.07 Ma; Zhai et al., 2012), they are still incompatible with the 'ancient sea-barrier hypothesis' of Ota (1998) for the Ryukyu Arc, where we would have expected much older divergences related to the initial formation of the Tokara and Kerama tectonic straits during the Pliocene (*c.* 5.3–1.8 Ma). Instead, being consistent with previous SSR data (Zhai et al., 2012), our RADseq data support that this north–south divergence in *N. sericea* probably resulted from land-bridge submergence in the Tokara region as a result of crustal movements and sea-level rise during the mid to late Pleistocene (Kimura, 2000).

Based on our cpDNA and ITS haplotype networks (Fig. 1b,c), both the Northern and Southern lineages across the 'Tokara gap' were found to harbour the ancestral haplotypes (e.g. H6, Ra), thus rendering it very difficult to infer ancestral distribution area of *N. sericea* in East Asia. By contrast, the tree topology recovered from ML analysis of the RAD-SNP data (Fig. 2) showed that samples from the Northern lineage held a nested and thus derived position relative to those from the south. This might reflect northward migration from areas south of the 'Tokara gap', possibly during glacial maxima via the exposed continental shelf (Fig. 1). In support of this, cytoplasmic diversity (*h*) showed a negative relationship with latitude (Fig. S4a), as would be expected under a scenario of south-to-north (re-)colonization. This also accords with palaeoecological/phylogeographic evidence (Kamei, 1981; Liew and Chung, 2001), ENM data (Fig. 4), and patterns of genetic composition of *N. sericea* predicted by GDM (Fig. 6), suggesting that during the last glacial(s), and possibly earlier cold periods, WTE forest retreated southward into coastal areas and/or the exposed continental shelf; the Ryukyu-Taiwan region likely acted as a major glacial refuge for components of WTE forest (including *N. sericea*) and source of northward colonization. According to our DIY-ABC divergence time analysis, a more recent divergence between the N1 and N2 sublineages around the Seto Inland Sea was estimated to be 0.065–0.196 Ma, assuming a generation time of 10 or 30 years, respectively. The present Japanese islands of Kyushu, Shikoku and Honshu are thought to have become separated only about 7,000 yr BP by the formation of the Seto Inland Sea (Ohshima, 1990). It thus appears unlikely that the genetic boundary between Kyushu (as part of the N1 sublineage) and Shikoku/(central-north) Honshu (N2) resulted from a strong landscape effect of the Seto Inland Sea as a dispersal barrier to gene flow in *N. sericea*. Rather, genetic divergence between these two sublineages of *N. sericea* would likely reflect refugial isolation in Kyushu/the exposed ECS basin and Shikoku/central Honshu during past glacial periods. However, for *N. sericea* populations from Zhoushan Archipelago, South Korea and Kyushu (N1 sublineage), the latest rise in sea-level of the ECS basin (< 16,000 yr bp) might separate a formerly widespread ancestral stock into three isolated population groups.

4.2. Disentangling IBD and IBE with landscape genomic tools

Reduced gene flow between populations due to geographic isolation (IBD) is commonly seen in empirical studies (Kuchta and Tan, 2005); however, the ubiquity of IBD has been long overemphasized (Wang et al., 2013; Wang and Bradburd, 2014). Recent landscape genetic studies have suggested that geography represents only one of the critical landscape components that can potentially influence genetic isolation and differentiation (Crispo et al., 2006; Wang et al., 2013; Wang and Bradburd, 2014; Lee and Mitchell-Olds, 2011), while environmental heterogeneity is also an important part of the landscape (Smith et al., 2005; Foll and Gaggiotti, 2006; Dudaniec et al., 2013; Bond et al., 2014). In our case, irrespective of their formation time and whether they were closed during late Pleistocene (Ota, 1998), the past formation of two gaps in Ryukyus have been considered to be the most likely barriers in the determination of genetic boundaries among island populations (Maekawa et al., 1999; Watanabe et al., 2006). But why then has a major genetic boundary of *N. sericea* occurred specifically across the 'Tokara gap' but not across the 'Kerama Gap' (Zhai et al., 2012)? In view of the fact that the two gaps formed at roughly the same time and the minimum geographical distance across the 'Kerama Gap' (*c.* 228 km) is currently and geohistorically far greater than the distance across the 'Tokara Gap' (*c.* 37 km wide, considering the nearest islands), this contrasting result observed in *N. sericea* cannot be explained by geographic distance alone. Some region-specific environmental factors (especially temperature) unique to the Ryukyus and main islands of Japan, especially across the Tokara region, have been hypothesized to have an additional role in the origin and maintenance of lineage divergence in *N. sericea* (Zhai et al., 2012), because the two major genetic clusters of *N. sericea* identified here occupied different climate conditions (Hübel, 1988).

The generalized dissimilarity models (GDM) originally conceived for species-level applications has been proved to have a great potential for modelling intraspecific diversity on the landscape and mapping ecological adaptation from genomic data (Fitzpatrick and Stephen, 2015). In this study, GDM revealed similar population structure in geographic and ecological space when compared to the tree-based ML and Bayesian structure results (Fig. 6). More importantly, GDM identified several pronounced gene-environment relationships in *N. sericea*, independent of the strong correlation between genetic divergence and geography (Fig. 5), many of which showed strong responses along temperature gradients (BIO3, BIO4, BIO8; Fig. 5). Therefore, our GDM analyses provide robust evidences for the local adaptation of *N. sericea* along a latitudinal temperature gradient, confirming our previous hypothesis of local adaptation of *N. sericea* in different environments (Zhai et al., 2012). Furthermore, our MMRR analysis, based on RADseq data, further revealed that IBD was the major contributor to genetic differentiation ($\beta_D = 0.404$, $P < 0.001$), but IBE played almost as large a role ($\beta_E = 0.382$, $P < 0.001$). Finally, to examine the possible processes generating the observed pattern of IBE, we identified 6 loci potentially under selection in our RADseq dataset, and tested for their associations with environmental variables. These loci were all under putative diversifying selection and associated with environmental variables that limit distributional ranges of *N. sericea*: BIO1 and BIO14 and BIO8 (except for loci 46,875 and 5422) (Table 2), suggesting that these regions of the genome are differentiating to a greater degree than the rest of the genome and that climate may play a role. Our results are highly consistent with other investigations into drivers of adaptive genetic divergence in plants and birds; namely, temperature and precipitation have been previously identified as the main drivers influencing allele frequencies at non-neutral outlier loci (Tsumura et al., 2012; Yoder et al., 2014; Manthey and Moyle, 2015). These climatic factors were also estimated as important driving forces of genome-wide neutral genetic divergence by GDM (particularly BIO8). Overall, our landscape genetic analyses using both GDM and MMRR show that geography and environment together played roles in shaping the genetic structure of *N. sericea*.

5. Conclusions

Our studies demonstrate that, compared to Sanger sequencing and microsatellites, RADseq approach with fewer samples but many more loci recovered finer population structure of *N. sericea*. In addition, RADseq was also used along with approximate Bayesian computation to show that the current distribution and differentiation of *N. sericea*

populations resulted from a combination of ancient migration and successive vicariant events. In summary, our findings, together with those from other studies in this region (Han et al., 2016), suggest that, apart from geographic barriers, barrier that potentially local adaptation to different climatic conditions appears to be one of the major drivers for lineage diversification of plants, endemic to East Asian land-bridge islands. This work highlights the utilities of RADseq based approaches for resolving complex spatial patterns and disentangling the effects of IBD and IBE on lineage diversification and speciation across heterogeneous ecological landscapes.

Acknowledgements

The authors thank Dr. Jung-Hyun Lee from Inha University for collecting plant materials in Korea Peninsula; Dr. Shota Sakaguchi from Kyoto University for collecting plant materials in Japan and running the ENMs; Hans-Peter Comes (Salzburg University, Austria) for insightful comments on an earlier version of this manuscript. This study was supported by the State Key Basic Research and Development Plan of China (Grant No. 2017YFA0605104), National Natural Science Foundation of China (Grant Nos. 31370241, 31570214) and the International Cooperation and Exchange of the National Natural Science Foundation of China (Grant Nos. 31511140095, 31561143015).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.ympev.2018.04.010.

References

Arai, N., Kamitani, T., 2005. Erratum: Seed rain and seedling establishment of the dioecious tree *Neolitsea sericea* (Lauraceae): effects of tree sex and density on invasion into a conifer plantation in central Japan. Botany 83, 1144–1150.
Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., Johnson, E.A., 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3, e3376.
Barley, A.J., Monnahan, P.J., Thomson, R.C., Grismer, L.L., Brown, R.M., 2015. Sun skink landscape genomics: assessing the roles of micro-evolutionary processes in shaping genetic and phenotypic diversity across a heterogeneous and fragmented landscape. Mol. Ecol. 24, 1696–1712.
Bertorelle, G., Benazzo, A., Mona, S., 2010. ABC as a flexible framework to estimate demography over space and time: some cons, many pros. Mol. Ecol. 19, 2609–2625.
Bond, M.H., Crane, P.A., Larson, W.A., Quinn, T.P., 2014. Is isolation by adaptation driving genetic divergence among proximate Dolly Varden char populations? Ecol. Evol. 4, 2515–2532.
Bradburd, G.S., Ralph, P.L., Coop, G.M., 2013. Disentangling the effects of geographic and ecological isolation on genetic differentiation. Evolution 67, 3258–3273.
Cassens, I., Waerebeek, K.V., Best, P.B., Crespo, E.A., Reyes, J., Milinkovitch, M.C., 2003. The phylogeography of dusky dolphins (*Lagenorhynchus obscurus*): a critical examination of network methods and rooting procedures. Mol. Ecol. 12, 1781–1792.
Chanderbali, A.S., Van, D.W.H., Renner, S.S., 2001. Phylogeny and historical biogeography of Lauraceae: evidence from the chloroplast and nuclear genomes. Ann. Misso. Bot. Gard. 88, 104–134.
Clement, M., Posada, D., Crandall, K.A., 2000. TCS: a computer program to estimate gene genealogies. Mol. Ecol. 9, 1657–1660.
Collins, W.D., Bitz, C.M., Blackmon, M.L., Bonan, G.B., Bretherton, C.S., Carton, J.A., Chang, P., Doney, S.C., Hack, J.J., Henderson, T.B., Kiehl, J.T., Large, W.G., McKenna, D.S., Santer, B.D., Smith, R.D., 2006. The Community Climate System Model version 3 (CCSM3). J. Clim. 19, 2122–2143.
Comes, H.P., Tribsch, A., Bittkau, C., 2008. Plant speciation in continental island floras as exemplified by *Nigella* in the Aegean Archipelago. Philos. T. R. Soc. B 363, 3083–3096.
Cornuet, J.M., Pudlo, P., Veyssier, J., Dehne-Garcia, A., Gautier, M., Leblois, R., Marin, J.M., Estoup, A., 2014. DIY-ABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. Bioinformatics 30, 1187–1189.
Coyne, J.A., Orr, H.A., 2004. Speciation. Sinauer Associates, Sunderland (MA).
Crispo, E., Bentzen, P., Reznick, D.N., Kinnison, M.T., Hendry, A.P., 2006. The relative influence of natural selection and geography on gene flow in guppies. Mol. Ecol. 15, 49–62.
Demos, T.C., Achmadi, A.S., Giarla, T.C., Handika, H., Maharadatunkamsi, Rowe, K.C., Esselstyn, J.A., 2016. Local endemism and within-island diversification of shrews illustrate the importance of speciation in building Sundaland mammal diversity. Mol. Ecol. 25, 5158–5173.
Dieckmann, U., Deobeli, M., 1999. On the origin of species by sympatric speciation. Nature 400, 354–357.
Drummond, A.J., Ashton, B., Cheung, M., Heled, J., Kearse, M., Moir, R., Stones-Havas, S., Thierer, T., Wilson, A., 2009. Geneious v4.8.5. Biomatters Ltd., Aukland, New Zealand.
Dudaniec, R.Y., Rhodes, J.R., Wilmer, J.W., Lyons, M., Lee, K.E., McAlpine, C.A., Carrick, F.N., 2013. Using multilevel models to identify drivers of landscape genetic structure among management areas. Mol. Ecol. 22, 3752–3765.
Eaton, D.A., Ree, R.H., 2013. Inferring phylogeny and introgression using RADseq data: an example from flowering plants (*Pedicularis*: Orobanchaceae). Syst. Biol. 62, 689–706.
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.
Fitzpatrick, M.C., Stephen, R.K., 2015. Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. Ecol. Lett. 18, 1–16.
Foll, M., Gaggiotti, O., 2006. Identifying the environmental factors that determine the genetic structure of populations. Genetics 174, 875–891.
Foll, M., Gaggiotti, O., 2008. A genome scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics 180, 977–993.
Futuyma, D.J., 2009. Evolution. Sinauer Associates, Sunderland, MA.
Gavrilets, S., 2003. Models of speciation: what have we learned in 40 years. Evolution 57, 2197–2215.
Han, Q.X., Higashi, H., Mitsui, Y., Setoguchi, H., 2016. Lineage isolation in the face of active gene flow in the coastal plant wild radish is reinforced by differentiated vernalisation responses. BMC Evol. Biol. 16, 1.
Hodel, R.G., Chen, S.C., Payton, A.C., Mcdaniel, S.F., Soltis, P., Soltis, D.E., 2017. Adding loci improves phylogeographic resolution in red mangroves despite increased missing data: comparing microsatellites and RAD-Seq and investigating loci filtering. Sci. Rep. 7, 17598.
Höglund, J., 2009. Evolutionary Conservation Genetics. Oxford University Press, Oxford.
Honda, M., Okamoto, T., Hikida, T., Ota, H., 2008. Molecular phylogeography of the endemic five-lined skink (*Plestiodon marginatus*) (Reptilia: Scincidae) of the Ryukyu Archipelago, Japan, with special reference to the relationship of a northern Tokara population. Pac. Sci. 62, 351–362.
Hübel, E., 1988. Die sommergrünen Wälder Japans und Westeurasiens, ein floristisch-klimatographischer Vergleich. Veröffentlichungen des Geobotanischen Instituts, ETH, Stiftung Rübel, Zürich, vol. 98, pp. 225–298.
Jeffries, D.L., Copp, G.H., Handley, L.L., Olsén, K.H., Sayer, C.D., Hänfling, B., 2016. Comparing RADseq and microsatellites to infer complex phylogeographic patterns, an empirical perspective in the Crucian carp, *Carassius carassius*, L. Mol. Ecol. 25, 2997–3018.
Jombart, T., 2008. Adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24, 1403–1405.
Kamei, T., 1981. Fauna and flora of the Japanese islands in the last glacial time. Quaternary Resour. 20, 191–205.
Kimura, M., 2000. Paleography of the Ryukyu Islands. Tropics 10, 5–24.
Kreft, H., Jetz, W., Mutke, J., Kier, G., Barthlott, W., 2008. Global diversity of island floras from a macroecological perspective. Ecol. Lett. 11, 116–127.
Kubota, Y., Shiono, T., Kusumoto, B., 2015. Role of climate and geohistorical factors in driving plant richness patterns and endemicity on the east Asian continental islands. Ecography 38, 639–648.
Kuchta, S.R., Tan, A.M., 2005. Isolation by distance and postglacial range expansion in the rough-skinned newt, *Taricha granulosa*. Mol. Ecol. 14, 225–244.
Le Roux, J.J., Strasberg, D., Rouget, M., Morden, C.W., Koordom, M., Richardson, D.M., 2014. Relatedness defies biogeography: the tale of two island endemics (*Acacia heterophylla* and *A. koa*). New Phytol. 204, 230–242.
Lee, C.R., Mitchell-Olds, T., 2012. Environmental adaptation contributes to gene polymorphism across the *Arabidopsis thaliana* genome. Mol. Biol. Evol. 29, 3721–3728.
Lee, C.R., Mitchell-Olds, T., 2011. Quantifying effects of environmental and geographical factors on patterns of genetic differentiation. Mol. Ecol. 20, 4631–4642.
Lexer, C., Wüest, R.O., Mangili, S., Heuertz, M., Stölting, K.N., Pearman, P.B., Forest, F., Salamin, N., Zimmermann, N.E., Bossolini, E., 2014. Genomics of the divergence continuum in an African plant biodiversity hotspot, I: drivers of population divergence in *Restio capensis* (Restionaceae). Mol. Ecol. 23, 4373–4386.
Li, Z., Zou, J.B., Mao, K.S., Lin, K., Li, H.P., Liu, J.Q., Källman, T., Lascoux, M., 2012. Population genetic evidence for complex evolutionary histories of four high altitude juniper species in the Qinghai-Tibetan Plateau. Evolution 66, 831–845.
Li, L., Li, J., Conran, J.G., Li, X.W., Li, H.W., 2007. Phylogeny of *Neolitsea* (Lauraceae) inferred from Bayesian analysis of nrDNA ITS and EST sequences. Plant Syst. Evol. 269, 203–221.
Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25, 1451–1452.
Liew, P.M., Chung, N.J., 2001. Vertical migration of forests during the last glacial period in subtropical Taiwan. Western Pac. Earth Sci. 1, 405–414.
Losos, J.B., Ricklefs, R.E., 2009. Adaptation and diversification on islands. Nature 457, 830–836.
Maekawa, K., Lo, N., Kitade, O., Miura, T., Matsumoto, T., 1999. Molecular phylogeny and geographic distribution of wood-feeding cockroaches in East Asian Islands. Mol. Phylogenet. Evol. 13, 360–376.
Manion, G., Ferrier, S., Lisk, M., Fitzpatrick, M.C., 2016. gdm: Functions for Generalised Dissimilarity Modelling. R package version 1, 13.
Manthey, J.D., Moyle, R.G., 2015. Isolation by environment in White-breasted Nuthatches (*Sitta carolinensis*) of the Madrean Archipelago sky islands: a landscape genomics approach. Mol. Ecol. 24, 3628–3638.

Miller, M.A., Pfeiffer, W., Schwartz T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Gateway Computing Environments Workshop (GCE) IEEE, pp. 1–8.
Myers, N., Mittermeier, R.A., Mittermeier, C.G., Da Fonseca, G.A.B., Kent, J., 2000. Biodiversity hotspots for conservation priorities. Nature 403, 853–858.
Nakamura, K., Denda, T., Kokubugata, G., Suwa, R., Yang, T.Y.A., Peng, C.-I., Yokota, M., 2010. Phylogeography of *Ophiorrhiza japonica* (Rubiaceae) in continental islands, the Ryukyu Archipelago, Japan. J. Biogeogr. 37, 1907–1918.
Nosil, P., 2012. Ecological Speciation. Oxford University Press, Oxford.
Ohshima, K., 1990. The history of straits around the Japanese Islands in the late-Quaternary. Quaternary Res. 29, 193–208.
Osozawa, S., Shinjo, R., Armid, A., Watanabe, Y., 2011. Palaeogeographic reconstruction of the 1.55 Ma synchronous isolation of the Ryukyu Islands, Japan, and Taiwan and inflow of the Kuroshio warm current. Int. Geol. Rev. 54, 1–20.
Ota, H., 1998. Geographic patterns of endemism and speciation in amphibians and reptiles of the Ryukyu Archipelago, Japan, with special reference to their paleogeographical implications. Res. Popul. Ecol. 40, 189–204.
Park, Y.C., Kitade, O., Schwarz, M., Kim, J.P., Kim, W., 2006. Intraspecific molecular phylogeny, genetic variation and phylogeography of *Reticulitermes speratus* (Isoptera: Rhinotermitidae). Mol. Cells 21, 89–103.
Petit, R.J., Hampe, A., 2006. Some evolutionary consequences of being a tree. Annu. Rev. Ecol. Evol. S. 37, 187–214.
Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. Ecol. Model. 190, 231–259.
Pinho, C., Hey, J., 2010. Divergence with gene flow: models and data. Ann. Rev. Ecol. Evol. Syst. 41, 215–230.
Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics 155, 945–959.
Qian, H., Ricklefs, R.E., 2000. Large-scale processes and the Asian bias in species diversity of temperate plants. Nature 407, 180–182.
Qiu, Y.X., Fu, C.X., Comes, H.P., 2011. Plant molecular phylogeography in China and adjacent regions: Tracing the genetic imprints of Quaternary climate and environmental change in the world's most diverse temperate flora. Mol. Phylogenet. Evol. 59, 225–244.
R Core Team, 2013. R: A Language and Environment for Statistical Computing. < http://www.R-project.org > .
Rašić, G., Filipović, I., Weeks, A.R., Hoffmann, A.A., 2014. Genome-wide SNPs lead to strong signals of geographic structure and relatedness patterns in the major arbovirus vector, Aedes aegypti. BMC Genom. 15, 275.
Richards-Zawacki, C.L., 2009. Effects of slope and riparian habitat connectivity on gene flow in an endangered Panamanian frog, *Atelopus varius*. Divers. Distrib. 15, 796–806.
Rousset, F., Ferdy, J.B., 2014. Testing environmental and genetic effects in the presence of spatial autocorrelation. Ecography 37, 781–790.
Schwind, M., 1967. Das Japanische Inselreich. Eine Landeskunde nach Studien und Reisen in 3 Bänden; Band 1: Die Naturlandschaft. Walter de Gruyter & Co, Berlin.
Smith, T.B., Calsbeek, R., Wayne, R.K., Holder, K.H., Pires, D., Bardeleben, C., 2005. Testing alternative mechanisms of evolutionary divergence in an African rain forest passerine bird. J. Evol. Biol. 18, 257–268.
Spear, S.F., Storfer, A., 2010. Anthropogenic and natural disturbance lead to differing patterns of gene flow in the Rocky Mountain tailed frog, *Ascaphus montanus*. Biol. Conserv. 143, 778–786.
Spurgin, L.G., Illera, J.C., Jorgensen, T.H., Dawson, D.A., Richardson, D.S., 2014. Genetic and phenotypic divergence in an island bird: isolation by distance, by colonization or by adaptation? Mol. Ecol. 23, 1028–1039.
Szulkin, M., Gagnaire, P.A., Bierne, N., Charmantier, A., 2016. Population genomic footprints of fine-scale differentiation between habitats in Mediterranean blue tits. Mol. Ecol. 25, 542–558.
Tsumura, Y., Uchiyama, K., Moriguchi, Y., Ueno, S., Ujino-Ihara, T., 2012. Genome scanning for detecting adaptive genes along environmental gradients in the Japanese conifer, *Cryptomeria japonica*. Heredity 109, 349–360.
Wang, I.J., 2013. Examining the full effects of landscape heterogeneity on spatial genetic variation: a multiple matrix regression approach for quantifying geographic and ecological isolation. Evolution 67, 3403–3411.
Wang, I.J., Glor, R.E., Losos, J.B., 2013. Quantifying the roles of ecology and geography in spatial genetic divergence. Ecol. Lett. 16, 175–182.
Wang, I.J., Bradburd, G.S., 2014. Isolation by environment. Mol. Ecol. 23, 5649–5662.
Watanabe, K., Kajita, T., Murata, J., 2006. Chloroplast DNA variation and geographical structure of the *Aristolochia kaempferi* group (Aristolochiaceae). Am. J. Bot. 93, 442–453.
Whittaker, R.J., Fernández-Palacios, J.M., 2007. Island Biogeography: Ecology, Evolution, and Conservation, second ed. Oxford University Press, Oxford.
Wright, S., 1943. Isolation by distance. Genetics 28, 114–138.
Yoder, J.B., Stanton-Geddes, J., Zhou, P., Briskine, R., Young, N.D., Tiffin, P., 2014. Genomic signature of adaptation to climate in *Medicago truncatula*. Genetics 196, 1263–1275.
Yumoto, T., 1987. Pollination systems of a warm temperate evergreen broad-leaved forest in Yakushima Island. Ecol. Res. 2, 133–145.
Zhai, S.N., Comes, H.P., Nakamura, K., Yan, H.F., Qiu, Y.X., 2012. Late Pleistocene lineage divergence among populations of *Neolitsea sericea* (Lauraceae) across a deep sea-barrier in the Ryukyu Islands. J. Biogeogr. 39, 1347–1360.
Zhang, B.D., Xue, D.X., Wang, J., Li, Y.L., Liu, B.J., Liu, J.Q., 2015. Development and preliminary evaluation of a genome wide single nucleotide polymorphisms resource generated by RAD-seq for the small yellow croaker (*Larimichthys polyactis*). Mol. Ecol. Resour. 16, 755–768.
Zulliger, D., Schnyder, E., Gugerli, F., 2013. Are adaptive loci transferable across genomes of related species? Outlier and environmental association analyses in Alpine Brassicaceae species. Mol. Ecol. 22, 1626–1639.