Population connectivity and size reductions in the Anthropocene: the consequence of landscapes and historical bottlenecks in white forsythia fragmented habitats

Homervergel G. Ong[1], Eui-Kwon Jung[2], Yong-In Kim[3], Jung-Hoon Lee[3], Bo-Yun Kim[4], Dae-Hyun Kang[5], Jae-Seo Shin[2] and Young-Dong Kim[1,2]*

Abstract

Background White forsythia (*Abeliophyllum distichum*) is an endangered Korean Peninsula endemic that has been subjected to recent population genomics studies using SNPs via RAD sequencing. Here, we primarily employed the often underutilized haplotype information from RAD loci to further describe the species' previously uninvestigated haplotype-based genomic variation and structure, and genetic-geographic characteristics and gene flow patterns among its five earlier identified genetic groups. We also inferred the time of past events that may have impacted the effective population size of these groups, as well as the species' potential future distribution amidst the warming climate and anthropogenic threats.

Results Our findings emphasized the recognition of the species' regional patterns of genetic structure, and the role of topography and its associated gene flow patterns as some of the possible factors that may have influenced the species' present-day fragmented population distribution. The inferred bottleneck events during the Anthropocene, some of which aligned with the time of historical catastrophic events on the Peninsula (e.g., the Korean War), were revealed to have contributed to the generally low effective population size of its five lineages, particularly those with marginal distributional range. Future distribution under both optimistic and pessimistic climatic scenarios suggests unlikely suitable habitats for these populations to expand from their current range limits, at least in the next 80 years.

Conclusions The small effective population size and landscape-driven limited gene flow among white forsythia populations will remain a big challenge for the conservation management of the species' already fragmented population distribution. To help mitigate these impacts, the merging of various research approaches and the use of genomic data to their full potential is recommended to provide the optimized knowledge-based tools for the conservation of this endangered species, and other similar plants under pressure.

Keywords *Abeliophyllum distichum*, Climate change, Korean endemic, Landscape genetics, RAD markers

*Correspondence:
Young-Dong Kim
ydkim@hallym.ac.kr
Full list of author information is available at the end of the article

Ong *et al. BMC Ecology and Evolution*   (2024) 24:123

Page 2 of 21

Background

Naturally occurring fragmented plant distributions are not uncommon and can be widely spread around the world at varying limits and scales [1]. In this type of natural system, habitat suitability, landscape connectivity, and environmental selective pressures are just some of the factors that influence the patterns of gene flow, and therefore, contemporary population structure (e.g., [2–6]). Because of the stationary or less mobile life characteristics of plants, their habitat distribution suggests genetic structure that has been more generally impacted by time and geographic space [7, 8].

Habitat fragmentation, however, reduces genetic variation and among-population gene flow due to increased isolation and random genetic drift [9–11]. The conversion of natural environments into anthropogenic landscapes, for instance, contributes to the further isolation of species populations, resulting in interpopulation connectivity loss, and lowered fitness and capacity for adaptation, henceforth, their higher risks of extinction [11–13].

The increasing anthropogenic disturbance of the natural environment has reached such high magnitudes that scientists have proposed a new geological epoch coined the Anthropocene [14]. Whether this era began with industrialization in the 1800s [14–16] or a few thousand years ago with the emergence of agriculture [17, 18], experts seem to agree that this period is consistently characterized by a significant increase in human influence on biological and geochemical processes at a global scale. The Anthropocene has brought negative repercussions to species' habitats, ecosystems, and biodiversity; all of which are due to man's desire for socio-economic improvement (e.g., agriculture, industrialization, and urbanization) [17], and even geopolitical power (e.g., war, colonization) [19]. With increased anthropogenic pressures and global temperatures, small and fragmented plant populations face an elevated risk of extinction, especially if species are unable to shift their distributional range (e.g., to move higher in altitude and latitude), or if, albeit ironic, no man-made conservation measures are implemented [20, 21].

In this study, we investigated the effects of historical events on the change in population size of the endangered Korean flowering plant white forsythia. We also looked into the role of geography and gene flow to the species' patchy population distribution on the Korean Peninsula (KP) (Fig. 1), a geopolitical region that has undergone rapid changes within the past three-quarters of a century since its civil war (i.e., the Korean War 1950–1953).

White forsythia (*Abeliophyllum distichum* Nakai) is a rare Oleaceae shrub that naturally grows on well-drained, rocky mountain slopes in the central regions of the KP. Some of the species' wild populations are South Korea's natural monuments, which are small, protected habitats that carry important scientific and cultural value to the country [23, 24]. Although endemic to the Peninsula, white forsythia is also a known ornamental plant in America and Europe, where it was first introduced to the horticultural community after its discovery in 1919 [25, 26]. However, it was only much later that this early-spring blooming shrub began to attract the attention of local plant collectors, and unfortunately poachers, all happening at a time when the country was also entering its period of industrial revolution. Eventually, over-collection destroyed some of the species' wild habitats, including one that had been originally declared a natural monument [26, 27]. As such, white forsythia became one of the first Korean endemics to receive a formal recovery plan and became a model species for conservation in South Korea [25, 26]. The current natural state of existence of white forsythia in North Korea, however, may be in question, because the last time it was confirmed to naturally occur in the reclusive country was way back in the 1930s [26, 28]. At the latest, the species is either listed as 'endangered' by the Korean National Arboretum [29], or 'vulnerable' by the National Institute of Biological Resources [30] in South Korea.

Population genetic studies that were foundational in the molecular aspect of white forsythia conservation can be traced back to the late 1990s, with the use of allozyme markers (e.g., [31–33]). With the application of a more modern reduced-representation of genome method via restriction-site associated DNA (RAD) sequencing (e.g., GBS), range-wide population genomic and phylogeographic studies using single nucleotide polymorphisms (SNPs) were able to further elucidate the species' genetic structure and variation patterns [34] and define its five genetic lineages and their history of divergence [35].

In this study, we attempt to provide new insights into the demographic history of each of these five white forsythia lineages (with the use of SNPs), as well as the influence of physical landscapes on their distribution and genetic differentiation. Our primary intention, however, was to employ the often-untapped haplotype information from RAD loci, here defined as markers of short nucleotide sequence following RAD sequencing and related bioinformatics protocols (see Methods) to further investigate other measures of population genetic variation and structure not evaluated in earlier studies (i.e., [34, 35]). Using the physically linked variants/SNPs contained in RAD loci, we also aimed to examine the population connectivity patterns among the species' genetic groups. This type of marker has been effectively applied to some non-model species population genomic studies (e.g.,

Ong *et al. BMC Ecology and Evolution*　　　(2024) 24:123

Page 3 of 21

Fig. 1 Natural population distribution (circles) of white forsythia (inset) on the KP. White circles (labeled) represent the nine sampled sites from which genetic and geographic data for molecular analyses were obtained. Black circles represent 15 additional (of 24 total) occurrence data for ecological niche modeling. The polygon of broken gray line marks the calibration area for the latter analysis, while the solid light gray lines represent the chain and branches of the Baekdudaegan Mountain Range [22]

[36–40]) due to its potential to improve the estimation of population parameters (e.g., demography and gene flow) of relevance to conservation [41].

Specifically, our goals were: (i) to determine the species' haplotype diversity and fine-scale population genetic structure using RAD markers, (ii) to define population boundaries using the species' genetic and geographic information, (iii) to infer the effective population size in each of the five lineages and the time when the change(s) occurred, (iv) to infer migration patterns between/among the lineages, and (v) to predict the species' suitable habitats under different future climatic scenarios.

Results

Genomic variation and structure based on RAD loci

RAD-based haplotype variation in nine white forsythia populations revealed a generally low and narrow range of within-population pairwise nucleotide diversity pi (π), with a mean of 0.00018. The lowest was recorded from the NORTHERN YJ population ($\pi = 0.00014$), while

the highest was from the SOUTHERN BA-NM370a ($\pi = 0.00022$) (Table 1). All sampled sites showed the presence of private alleles, except those from the UPPER-CENTRAL lineage (GS-NM147 and GS-NM221). The number of private alleles was found highest for the marginal populations NORTHERN YJ and EASTERN AD, each with 43 and 41, respectively. As also shown in Table 1, Tajima's D (D) was positive for all sampled locations, with the two UPPER-CENTRAL populations (GS-NM147 and GS-NM221) recording the lowest values (D = 0.053 and D = 0.187, respectively). The highest Tajima's D was recorded from the NORTHERN YJ population (D = 1.123), which is nearly double that of the second-highest EASTERN AD (D = 0.727).

When between-population haplotype differentiation results are compared (Table 2), the highest pairwise Phi$_{ST}$ (Φ_{ST}) was recorded between the marginal population pairs NORTHERN YJ and EASTERN AD ($\Phi_{ST} = 0.393$), and between NORTHERN YJ and SOUTHERN BA-NM370b ($\Phi_{ST} = 0.394$). The lowest

Ong *et al. BMC Ecology and Evolution*   (2024) 24:123

Page 4 of 21

Table 1 Population information and RAD loci-based genetic variation

Population	Locality	Lineage	n	Sites	SNPs/total	PA	π	D
GS-NM147	Goesan	UPPER-CENTRAL	15	638,096	428/968	0	0.00017	0.053
GS-NM221	Goesan	UPPER-CENTRAL	15	638,101	383/968	0	0.00017	0.187
YD-NM364	Yeongdong	LOWER-CENTRAL	15	638,098	428/968	7	0.0002	0.292
YD	Yeongdong	LOWER-CENTRAL	15	638,097	335/968	14	0.00019	0.493
OC	Okcheon	LOWER-CENTRAL	14	638,108	339/968	9	0.00017	0.388
BA-NM370a	Buan	SOUTHERN	14	638,096	397/968	10	0.00022	0.607
BA-NM370b	Buan	SOUTHERN	13	638,088	317/968	2	0.00019	0.454
AD	Andong	EASTERN	15	638,101	293/968	41	0.00018	0.727
YJ	Yeoju	NORTHERN	15	638,092	210/968	43	0.00014	1.123

n number of samples, *PA* Private alleles, *π* pairwise nucleotide diversity pi, *D*- Tajima's D

Sites- number of sites (variant and invariant), SNPs/total- polymorphic SNPs over the total number of variants

Table 2 Pairwise population comparison of RAD-based Phi_{ST} (below diagonal) and D_{XY} (above diagonal). The lowest and highest values are in bold

	GS-NM147	GS-NM221	YD-NM364	YD	OC	BA-NM370a	BA-NM370b	AD	YJ
GS-NM147	0	**0.00143**	**0.00145**	0.00163	0.00149	0.00173	0.00183	0.00177	0.00193
GS-NM221	**0.054**	0	0.00155	0.00178	0.00164	0.00185	0.00198	0.0019	0.00214
YD-NM364	**0.117**	0.135	0	0.00181	0.00167	0.00186	0.00197	0.00192	0.00212
YD	0.166	0.183	0.149	0	0.00196	0.00213	0.00232	0.0023	0.00258
OC	0.136	0.156	0.149	0.206	0	0.00197	0.00216	0.00214	0.00242
BA-NM370a	0.183	0.208	0.205	0.259	0.233	0	0.00238	0.0023	0.00256
BA-NM370b	0.205	0.233	0.233	0.293	0.266	0.16	0	0.00253	**0.00284**
AD	0.195	0.215	0.224	0.293	0.258	0.299	0.342	0	**0.00287**
YJ	0.227	0.266	0.273	0.352	0.319	0.345	**0.394**	**0.393**	0

Phi_{ST}- between-populations differentiation for haplotype data based on the analysis of molecular variance, D_{XY}- absolute measure of between-populations locus differentiation

Phi_{ST} was revealed between the two UPPER-CENTRAL populations GS-NM147 and GS-NM221 ($\Phi_{ST}=0.054$). As also shown in Table 2, the absolute measure of between-population locus divergence (D_{XY}) was highest between the NORTHERN YJ and EASTERN AD populations ($D_{XY}=0.00287$), and between the former and the SOUTHERN BA-NM370b ($D_{XY}=0.00284$), while the lowest was found between the two UPPER-CENTRAL populations GS-NM147 and GS-NM221 ($D_{XY}=0.00143$).

Population assignment and genomic structure of samples in fineRADstructure (Fig. 2) generally grouped white forsythia samples according to their regional geographic distributions (i.e., into clusters of central and marginal population identities). When examined more locally, all individuals clustered based on their sampled locations (except those in UPPER-CENTRAL populations GS-NM147 and GS-NM221), despite not having been assigned to any a priori grouping. The plot showed particularly high levels of co-ancestry among respective members of the marginal EASTERN, NORTHERN, and SOUTHERN lineages by clearly grouping them into dark-colored blocks. Even within the SOUTHERN lineage, individuals from BA-NM370a and BA-NM370b populations displayed higher degrees of co-ancestry coefficients than average (Fig. 2). Substantial population sub-structuring was also visually evident among the three populations of the LOWER-CENTRAL lineage (YD-NM364, YD, and OC). The weakest genomic structure was reflected among the samples of the two UPPER-CENTRAL populations (GS-NM147 and GS-NM221), appearing to be indistinguishable from one another in one large block.

Genetic-geographic barriers to gene flow

Our Procrustes analyses of genotype information (i.e., genetic matrix in PC space) and sampling locations (i.e., geographic matrix in physical map) showed a very significantly positive correlation in a symmetric rotation
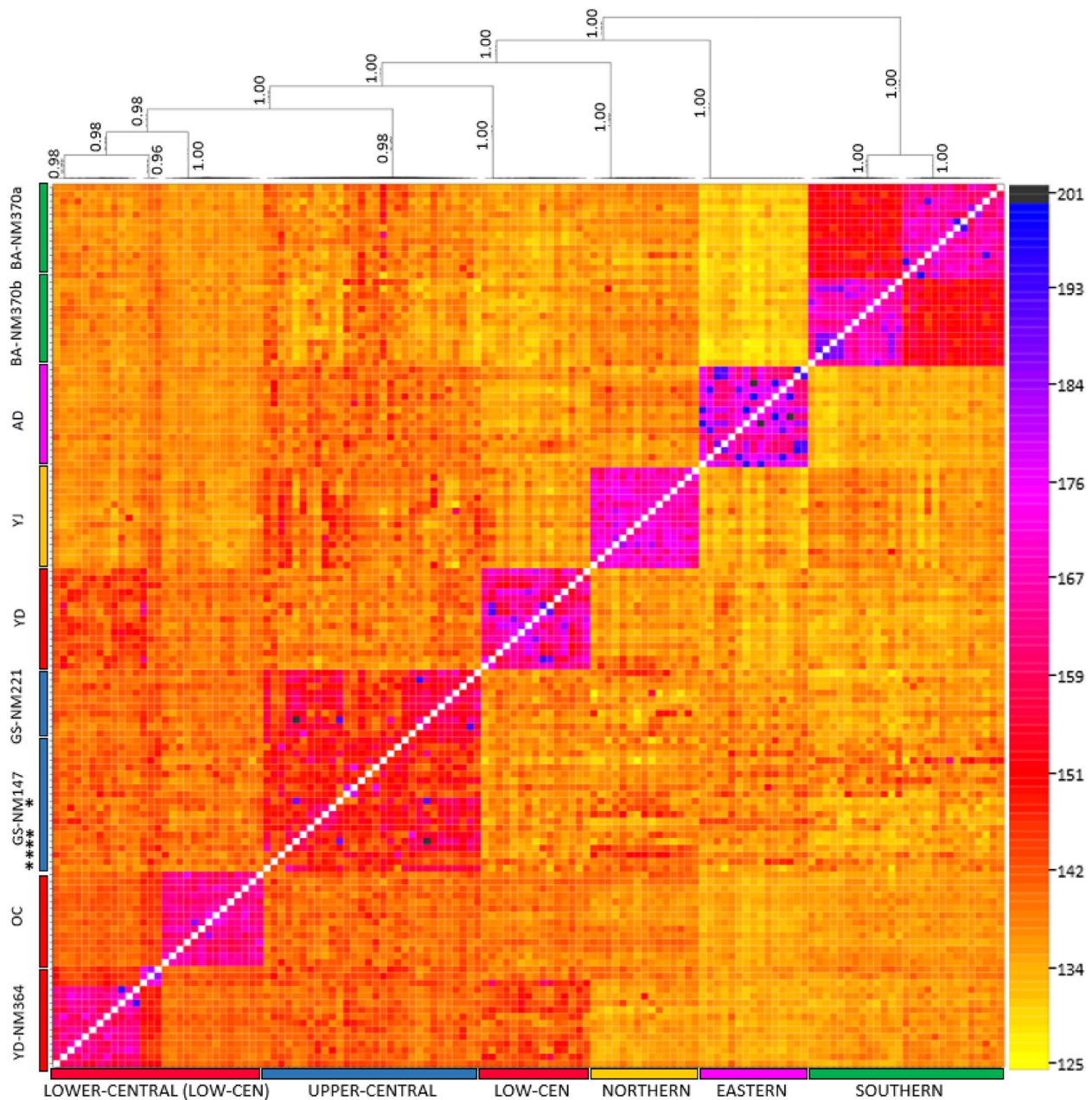
Fig. 2 Co-ancestry matrix between each pair (pixel) of 131 white forsythia samples in fineRADstructure, indicating membership assignment according to their color-coded sampled population (left) or lineage (below) and associated clades (above). The color scale (right) shows the strength of co-ancestry coefficients from low (yellow) to high (approaching black). The asterisks (*) indicate individuals assigned to population(s) other than their sampled location (i.e., GS-NM221 samples in GS-NM147 population block)

($t_0 = 0.891$, $p < 0.001$). Most genotypes, particularly those from the centrally distributed UPPER-CENTRAL (GS-NM147 and GS-NM221), and LOWER-CENTRAL (YD-NM364, YD, and OC) lineages were shown to have been plotted near their respective sampled sites, as indicated by the population-connecting arrows with relatively shorter shafts (Fig. 3). Only the genotypes from the NORTHERN YJ, and to some degree the EASTERN AD, appeared to have the strongest deviation from their respective geographic locations (Fig. 3). The Monmonier analysis using the Delaunay triangulation method (Fig. 4), on the other hand, detected a semi-radial genetic-geographic barrier that appears to have isolated the members of the core central distribution from the marginal ones.
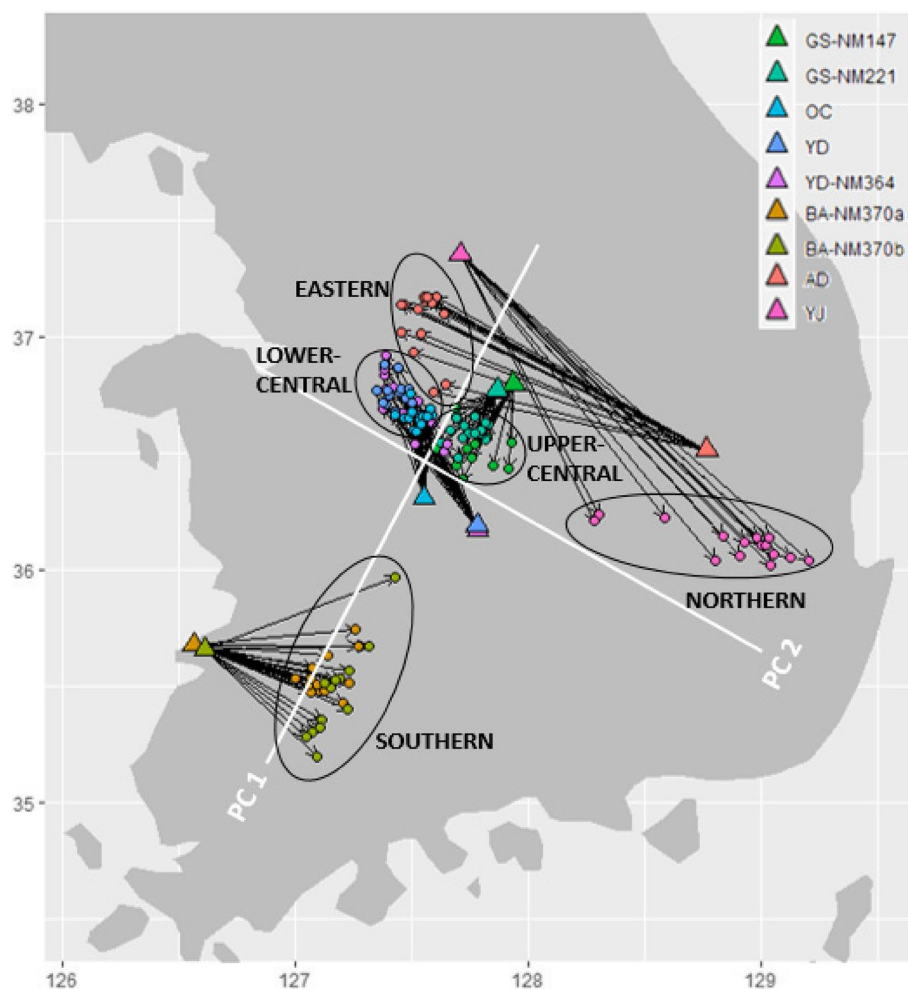
Ong *et al. BMC Ecology and Evolution*   (2024) 24:123

Page 6 of 21

Fig. 3 Procrustes adjustment plot showing the PC-rotated white forsythia genotypes (color-coded circles) against their sampled geographic locations (color-coded triangles). The length and direction of the arrows that connect the genotypes to their sampled locations suggest the degree of deviation/deformation of the former from the latter

Historical gene flow patterns

Among the seven proposed migration models, our MIGRATE-N inference singled out the Isolated-SOUTH-ERN Migration model as the most likely gene flow pattern for white forsythia (Table 3). The model proposes that although connectivity among all paired lineages is open, gene flow is absent between the SOUTHERN and each of the two remaining marginal lineages (NORTHERN and EASTERN). The selected scenario is supported by its Bezier approximation scores of marginal likelihood (-28438.13), which is the lowest among those of other migration models, and is almost 90 units lesser than that of the next best scenario (i.e., Isolated-EASTERN Migration Model, -28525.76), validating its selection as the more probable gene flow scenario of choice (Table 3).

Posterior distribution statistics (Additional file 1) indicated strong symmetrical gene flow patterns between the two core central lineages, as shown by their mutation-scaled migration rates (M), and non-overlapping to zero

confidence intervals (95% CIs): UPPER-CENTRAL to LOWER-CENTRAL (M = 29.7, 95% CI = 19.3–41.3), and LOWER-CENTRAL to UPPER-CENTRAL (M = 24.3, 95% CI = 12.7–35.3). High confidence in gene flow rates was also recorded for population pairs of either of the two core central lineages and any of the three marginal groups (except for that of the UPPER-CENTRAL to EASTERN migration). The weakest mutation-scaled migration rates were recorded from marginal NORTH-ERN to EASTERN (M = 11.7, 95% CI = 2.7–19.3), and vice versa (M = 3, 95% CI = 0–12).

The inferred number of migrants per generation (Nm) for each of the lineage pair combinations suggested generally low numbers of migrating individuals (Fig. 5). For instance, the highest number of migrants was recorded from the UPPER-CENTRAL to the LOWER-CENTRAL lineage (Nm = 0.24131), which is still only ca. one individual in the past four generations (see formula in Methods).

Fig. 4 Genetic barrier (blue) in white forsythia sampled populations (white circles) after an optimized Monmonier analysis via the Delaunay triangulation method. For reference, subsidiary mountain ranges in the map are labeled. Populations: GS-NM147 (1), GS-NM221 (2), YD-NM364 (3), YD (4), OC (5), BA-NM370a (6), BA-NM370b (7), AD (8), and YJ (9)

Table 3 Migration models tested in MIGRATE-N and ranked based on the log Bayes factor (LBF) after comparing their Bezier approximation scores. LBF of each model was compared to that of the default Full Migration model

Gene Flow Scenario	Bezier Score	LBF	Rank
Full Migration	-28598.89	0	4th
UPPER-CENTRAL Symmetrical Migration	-29549.87	950.98	6th
UPPER-CENTRAL Asymmetrical Migration	-30423.7	1824.81	7th
Non-connecting Marginals Migration	-28739.92	141.03	5th
Isolated-EASTERN Migration	-28525.76	-73.13	2nd
Isolated-NORTHERN Migration	-28579.01	-19.88	3rd
Isolated-SOUTHERN Migration	-28438.13	-160.76	1st

Changes in effective population size

After evaluating six scenarios of population size change for each of the five lineages, the best models were all inferred to show reductions of effective population size (i.e., either Scenario 1 or 3) (Table 4). For instance, the posterior probability (PP) estimates for marginal SOUTHERN (PP = 0.5312), EASTERN (PP = 0.5245), and NORTHERN (PP = 0.5444) lineages suggested Scenario 1 as their model of choice, implying that their effective population size (Ne) was a result of a population bottleneck in the ancestral population (Nanc) which occurred at t1 (Table 5). On the other hand, demographic inference for the UPPER-CENTRAL (PP = 1.0) and LOWER-CENTRAL (PP = 1.0) lineages, selected Scenario 3 as their best model, which suggests that their Ne was a

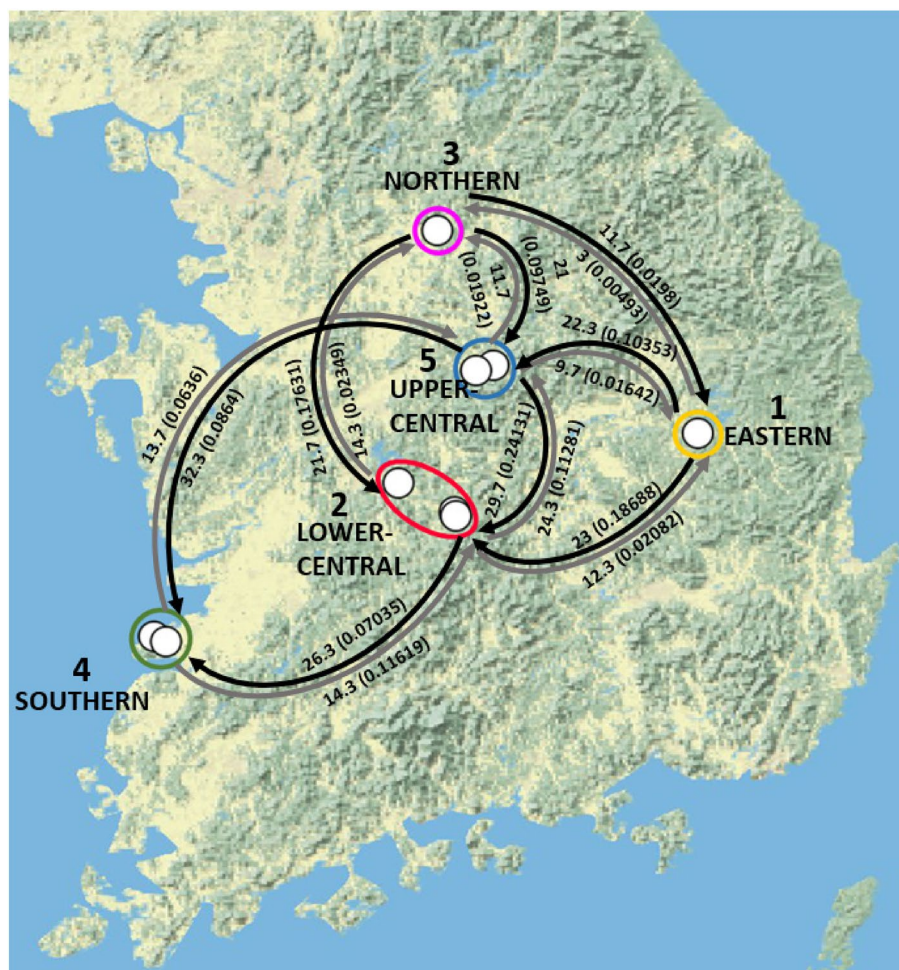Ong *et al. BMC Ecology and Evolution*   (2024) 24:123

Page 8 of 21

Fig. 5 Best-supported migration model (Isolated-SOUTHERN Migration) showing the mutation-scaled migration rate (M) and the per generation number of migrants (Nm) (in parenthesis) between lineage pairs (labeled/encircled/numbered). Black arrows indicate the stronger M as compared to its gray counterpart

Table 4 Per lineage most highly supported ABC demographic history model showing the posterior probability estimates computed via logistic regression method, with their corresponding 95% CI (in brackets). Effective population size (Ne) values are those of the mode based on SNP dataset

Lineage	n	SNP	Model	Logistic regression	Ne	Θ[a]
UPPER-CENTRAL	30	525	3	1.0000 [1.0000–1.0000]	51.8	0.01857
LOWER-CENTRAL	44	631	3	1.0000 [1.0000–1.0000]	11.9	0.0325
SOUTHERN	27	435	1	0.5312 [0.5207–0.5418]	10	0.0107
EASTERN	15	293	1	0.5245 [0.5139–0.5350]	3	0.00677
NORTHERN	15	210	1	0.5444 [0.5281–0.5607]	1	0.00657

n number of samples per lineage, *SNP* number of SNPs per lineage

[a] The last column shows the MIGRATE-N mutation-scaled effective population size theta (Θ) values based on RAD loci dataset

consequence of a population contraction that took place at t1, after an initial population size expansion (Nexp) of the ancestral population (Nanc) at a more distant t2 (Table 6).

Overall, our approximate Bayesian computation (ABC) analyses recorded generally low mode values of effective population size (Ne) among the marginal SOUTH-ERN (Ne = 10, 95% CI = 10–35), EASTERN (Ne = 3, 95%

Ong *et al. BMC Ecology and Evolution*   (2024) 24:123

Page 9 of 21

Table 5 Marginal populations' demographic history estimation statistics (mean, median, mode) of the posterior distribution for the parameters ancestral effective population size (Nanc), and the effective population size (Ne) after a bottleneck event at time 1 (t1), including their 95% CI under the best model (Scenario 1). Values interpreted and discussed in the paper are those of the mode (bold)

Lineage	Nanc	t1 (bottleneck)	Ne	Years bp[a]	Calendar years
Southern					
mean	12,000	17.5	16.1	140 (105–175)	1880 (1845–1915)
median	12,500	13.6	14	109 (82–136)	1911 (1884–1938)
mode	**17,700**	**8.89**	**10**	**71 (53–89)**	**1949 (1931–1967)**
95% CI	1,810–19,600	5.6–43.4	10–35.4	44.8–347.2	1648–1975
Eastern					
mean	6,020	5.73	6.6	46 (34–57)	1974 (1963–1986)
median	6,210	5.06	4.93	40 (30–51)	1980 (1969–1990)
mode	**9,730**	**5**	**3**	**40 (30–50)**	**1980 (1970–1990)**
95% CI	961–9,860	5–9.27	3–18.9	40–74	1946–1980
Northern					
mean	9,920	4.5	1.67	36 (27–45)	1984 (1975–1993)
median	9,770	2.9	1	23 (17–29)	1997 (1991–2003)
mode	**8,370**	**1**	**1**	**8 (6–10)**	**2012 (2010–2014)**
95% CI	952–19,500	1–16.8	1–4.48	8–134.4	1886–2012

[a] Absolute time (in years before present) was calculated by multiplying the mode of t1 with the generation time of the species set at 8 years [35] and/or 6–10 years (in parenthesis) according to [34]. Upper and lower CIs were multiplied by 8 years

Table 6 Central populations' demographic history estimation statistics (mean, median, mode) of the posterior distribution for the parameters ancestral effective population size (Nanc), effective population size after expansion (Nexp) at time 2 (t2), and the effective population size (Ne) after a bottleneck event at time 1 (t1), including their 95% CI under the best model (Scenario 3). Values reported and discussed in the paper are those of the mode (bold)

Lineage	Nanc	t2 (expansion)	Nexp	Years bp (t2)[a]	t1 (bottleneck)	Ne	Years bp (t1)[a]	Calendar years (t1)
Upper–Central								
mean	815	13,100	52,900	104,800 (78,600–131,000)	29.8	89.1	238 (178–289)	1782 (1731–1842)
median	587	13,700	54,700	109,600 (82,200–137,000)	30.6	85.4	244 (184–306)	1776 (1714–1836)
mode	**225**	**18,400**	**73,500**	**147,200 (110,400–184,000)**	**38.2**	**51.8**	**306 (229–382)**	**1714 (1638–1721)**
95% CI	37.7–2,890	3,540–19,700	21,100–74,100	28,320–157,600	7.18–48.8	20.8–182	57–390	1630–1963
Lower–Central								
mean	619	13,800	71,000	110,400 (82,800–138,000)	13.7	27.9	110 (82–137)	1910 (1883–1938)
median	432	14,800	76,100	118,400 (88,800–148,000)	11.9	20.1	95 (71–119)	1925 (1901–1949)
mode	**287**	**19,200**	**94,800**	**153,600 (115,200–192,000)**	**8.67**	**11.9**	**69 (52–87)**	**1951 (1933–1968)**
95% CI	111–2,350	3,730–19,700	25,700–98,800	29,840–157,600	5.15–31.7	10.8–81.5	41–254	1766–1979

[a] Absolute time (in years before present) was calculated by multiplying the mode of t1 with the generation time of the species set at 8 years [35] and/or 6–10 years (in parenthesis) according to [34]. Upper and lower CIs were multiplied by 8 years

CI=3–19), and NORTHERN (Ne=1, 95% CI=1–4) lineages than those from the UPPER-CENTRAL (Ne=52, 95% CI=23–182) and LOWER-CENTRAL (Ne=12, 95% CI=11–81) groups. The posterior probability distribution graph and the goodness-of-fit plots of these scenarios of choice are shown in Additional file 2.

After multiplying the posterior distribution of parameter t with the species' generation time of 8 and/or 6–10 years, the most probable events which may have

Ong *et al. BMC Ecology and Evolution*   (2024) 24:123

Page 10 of 21

significantly contributed to population contractions in marginal SOUTHERN, EASTERN, and NORTHERN lineages were inferred to have occurred ca. 1949 (1931–1967), ca. 1980 (1979–1990), and ca. 2012 (2010–2014), respectively (Table 5). On the other hand, population bottleneck events in the UPPER-CENTRAL and LOWER-CENTRAL lineages were inferred to have taken place ca. 1714 (1638–1721), and ca. 1951 (1933–1968), respectively (Table 6). All 95% CIs for each parameter (e.g., Ne, t) were not shown to overlap zero.

Habitat suitability and future distribution

Based on 24 white forsythia natural occurrence records, five layers of bioclimatic variables, and 24 replicate runs with optimized MAXENT settings, the species' inferred Present (1970–2000) habitat suitability was shown to have a higher predictive power of data discrimination than at random (AUC = 0.890, SD = 0.190). As projected in Fig. 6, the Present model appears to indicate a continuous suitable habitat in the central region of the KP, expanding beyond the areas of the confirmed occurrences, except the ones surrounding the southernmost occurrences in South Korea and those recorded in North Korea. Even when projected onto

the two differing SSP1-26 ('Sustainability') and SSP3-70 ('Regional Rivalry') future climatic scenarios, binary output maps revealed very little difference in the spread of potential habitat suitability for 2011–2040, 2041–2070, and 2071–2100, slightly favoring the conditions of SSP1-26 models (Fig. 6).

The response curves for all five environmental variables were shown to be unimodal and more or less normally distributed. The relative importance of these environmental predictors revealed that bio14 precipitation amount of the driest month and bio5 maximum temperature of the warmest month supplied the highest percent contribution and permutation importance to the model (Table 7). Response curves for the highest contributing variables bio14 and bio5 implied that the optimal suitability for the species peaked at ca. 24 mm, and at ca. 29 °C, respectively (Additional file 3).

Discussion

RAD loci-based genomic information

In the past few years, there has been a gradual increase in Korean non-model plant population genomic studies that employed RAD sequencing-related methods (e.g., [34, 42–46]). Many of such works, however, only utilized
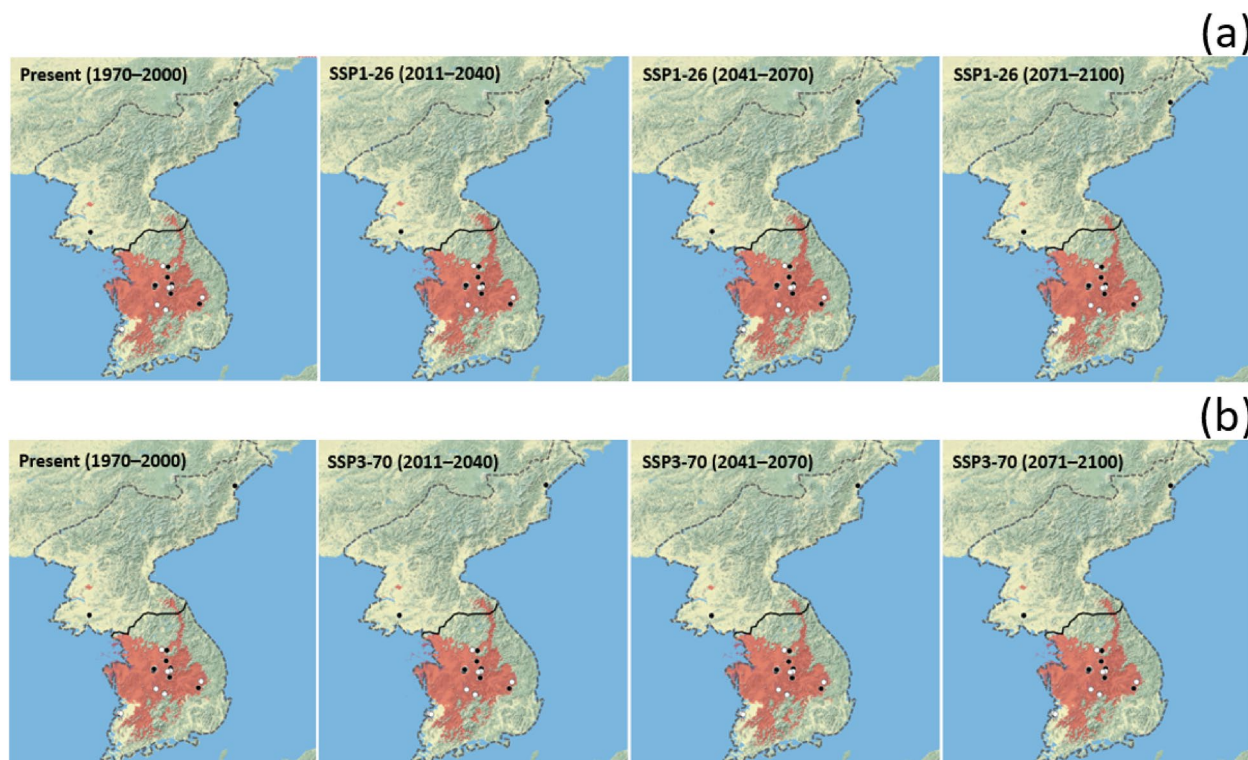

Fig. 6 Present (1970–2000) to future binary map projections of white forsythia suitable (red) and non-suitable (non-red) habitats modeled using the UKESM1-0-LL (**a**) optimistic SSP1-26, and (**b**) pessimistic SSP3-70 scenarios for up to about the next 20 to 80 years (i.e., 2011–2040, 2041–2070, and 2071–2100 time periods)

Table 7 Five environmental variables and their contribution and importance to the Present (1970–2000) niche model

Variable	Variable name (and unit)	Percent contribution	Permutation importance
bio14	precipitation amount of the driest month (mm)	66	63.6
bio5	maximum temperature of the warmest month (℃)	15.7	16.2
bio15	precipitation seasonality (mm)	14.6	6.6
bio4	temperature seasonality (℃, SD×100)	2.5	9.9
bio16	precipitation of the wettest quarter (mm)	1.3	3.7

SNPs, and have not fully exploited the linkage information contained in the short RAD sequence data. Outside the Korean region, the use of haplotype information for non-model organism population genomics has already been shown to help increase the accuracy of population structure inference and individual assignment (e.g., [36–38]), and even species delimitation (e.g.,[39, 40]; see [41] for review).

In this study, we took advantage of STACKS 2 features for estimating haplotype-based statistics (e.g., pi, Phi_{ST}, D_{XY}), and for outputting data format types which facilitated the use of RAD markers for further population genetic analyses (e.g., for Tajima's D, fineRADstructure, and MIGRATE-N). Using our RAD dataset, we showed that the genomic structure in our samples agreed with their geographically patchy population distribution, complementing the findings in previous white forsythia studies that used SNP markers. Values of haplotype-based Phi_{ST}, for example, were shown to be similar to the earlier reported AMOVA-based F_{ST} findings of moderate to high [34]. In this research, the relative between-population differentiation (Phi_{ST}) was also shown to mirror the results of absolute between-population differentiation (D_{XY}), highlighting a strong agreement in the structured distribution system in the species. These findings generally imply weak between-population genetic differentiation among sampled sites from the same genetic grouping (i.e., populations from the same central or marginal lineages), and conversely strong genetic structuration between populations from different genetic clusters. Specifically, the very high D_{XY} and Phi_{ST} values recorded when populations are paired with NORTHERN YJ suggest this population's strong isolation from the rest of the group, while the very low pairwise nucleotide diversity pi in this population may coincide with its small effective population size (see below geography-related genetic differentiation, and demographic history-related discussion, respectively).

Our analysis of haplotype linkage grouped our samples into eight to nine population blocks, which generally clustered together into five lineages, even in the absence of a priori population cluster setting (i.e., unlike the *K*-clustering usually set in popular programs like

STRUCTURE). The co-ancestry matrix in Fig. 2 plotted a strong regional geographic grouping, particularly among marginal lineages. The five populations forming the core central distribution were also shown to cluster in their clade. Within the UPPER-CENTRAL lineage, however, ca. 33% (5/15) of GS-NM221 samples appeared to associate with those of GS-NM147. This finer-scale population assignment indicates that some individuals in GS-NM147 may have been sourced from its neighboring population GS-NM221 (or vice versa) and is likely the reason for the lack of rare alleles in these populations (Table 1). These findings are in line with those of [34], recording no significant difference in pairwise F_{ST} estimates between the two UPPER-CENTRAL populations. Lee et al. [31], in their pioneering population genetic work on white forsythia a few decades earlier, cautioned on finding individuals from other populations to the former due to (unclearly documented) possibly multiple population augmentation work in the natural monument habitats found in the locality of Goesan. Our analysis also highlights the strong sub-structuring within the LOWER-CENTRAL lineage (YD-NM364, YD, and OC), consistent with those of earlier published work using SNP loci [34, 35]. See later discussion on the use of haplotype information at a within-population level and on demographic history inference.

Genetic-geographic barriers and migration patterns

Our analyses using the combined genetic and geographic data showed that the differences in the degree of between-population genetic structure in white forsythia is a consequence of spatial factors, but one that is not due to distance. The significantly positive correlation between the genotypes and their geographic positions in Procrustes analysis suggested that except for the samples from the NORTHERN YJ (and EASTERN AD), no other samples were shown to have significantly deviated from their sampled locations (Fig. 3), implying that population differentiation is more likely due to factors other than spatial extent. The results are in concordance with those reported by [34], particularly the absence of isolation by distance findings based on a Mantel test.

The above assumption is also supported by our Monmonier analysis, as it identified a continuous genetic barrier that distinctly separates the core central populations from nearby marginal ones, despite their proximity. As shown in Fig. 4, this genetic barrier appears to trace the rugged topography of the surrounding subsidiary branches ('sanmaek') of the Baekdudaegan Mountain Range (see also Fig. 1). For instance, the Sobaek Sanmaek and Charyeong Sanmaek, may have functioned as geographic impediments to gene flow between the core central lineages (UPPER-CENTRAL and LOWER-CENTRAL) and the two nearest non-central lineages (EASTERN and NORTHERN). At the same time, the semi-radial genetic-geographic barrier where members of the core central distribution are nested, may have served as a region where strong central population interconnectivity was maintained. The above findings are corroborated by the low population pairwise differentiation (e.g., low Phi_{ST} and D_{XY}) among the central range members, and the relatively strong symmetrical gene flow patterns (i.e., MIGRATE-N results) between the UPPER-CENTRAL and LOWER-CENTRAL lineages. Chung et al. [47], in their review of plant species distributed in the Baekdudaegan Mountain Range, highlighted the role of this rugged landscape both as a genetic corridor and a glacial refuge for Korean endemic flora.

The partial geographic blockade by the branches of the Noryeong Sanmaek, in addition to the wide flat plains that separate the populations in the SOUTHERN lineage (BA-NM370a and BA-NM370b) from the rest of the group (Fig. 4), supported the selection of our best gene flow scenario (i.e., Isolated-SOUTHERN Migration model). These findings underscore the position of this genetic subdivision as an early-diverging, rear-edge population that is the most distinct among all white forsythia genetic groups [35]. Still, we do not exclude the possibility that factors other than topography may have also affected gene flow patterns and population genetic structure in white forsythia, hence, evaluating the species' genetic-geographic relationships may require further assessments beyond simple barrier analyses. The revealed genetic structure (and gene flow dynamics), for instance, may have been also influenced by the species' assumed limited spread, wind-dispersed samaras [32, 34]. According to [48], species with low dispersal ability generally have high genetic structure (e.g., high Phi_{ST} values), and therefore, finer population differentiation resolution (e.g., in fineRADstructure). These characteristics have also likely complemented a highly local genetic-geographic methods like our Monmonier and Procrustes analyses. Future studies are recommended including the human-altered environment in highly developed South Korea, where the limited flat landscapes are often converted into agricultural areas or urban centers.

Population bottlenecks and historical events on the KP

Demographic history inference of endangered plants, such as the estimation of their past population divergence and size change, is fundamental to understanding their evolutionary potential and in defining appropriate conservation strategies (e.g., [49–52]). Although gradually increasing in the last few years, the number of Korean plant demographic history studies that employed the ABC method is still comparatively lacking. When available, published studies mainly focused on the time of weedy species invasion [53], population divergence in the distant past covering the Last Glacial Maximum (LGM) or earlier [54, 55], and in the case of white forsythia, expansion and divergence into five lineages post-LGM [35].

In this follow-up work to [35], our findings suggest that effective population size in these five genetic groups was most significantly impacted by rather recent bottleneck events (i.e., during the Anthropocene). Recalibration of t parameters suggests that population contractions in the SOUTHERN (t1 = 1949, 1931–1967) and LOWER-CENTRAL (t1 = 1951, 1933–1968) lineages took place during the time of the Korean War (1950–1953). Our deeper inquiry revealed that sampled populations in the SOUTHERN lineage (BA-NM370a and BA-NM370b) are within a 5-km radius of the areas (i.e., present-day Byeonsanbando National Park) where intense clashes between troops from the two Koreas took place [56]. Similarly, sampled habitats from the LOWER-CENTRAL lineage (e.g., YD-NM364 and YD) are situated at the center of the Town of Yeongdong/Yongdong, a locality so crucial at the beginning of the conflict that it was given its segment in the 3-year civil war (i.e., the Battle of Yongdong) [57, 58]. On the other hand, the population reduction in the NORTHERN lineage suggested the most recent and localized bottleneck event (t1 = 2012, 2010–2014), which can be attributed to the construction of several golf country clubs in its locality (Yeoju) in the early 2010s, as noted by Lee and colleagues [59] during their publication of this site's discovery. Our most recent inspection of the area (March 2024) revealed that more than 60% of the population reported ca. 10 years ago [59], had disappeared from this steep, stream bank-roadside habitat.

The possible bottleneck events in the remaining lineages (EASTERN and UPPER-CENTRAL), even though less accurately verifiable, may have also been impacted by several waves of human-related land use. The size reduction in the stream bank EASTERN population (t1 = 1980, 1979–1990), for example, may have been affected by the infrastructure development projects (e.g., farm roads and irrigation) in the 1970s and 1980s, including the completion of a dam in 1976 [60]. Even to this day, the locality of the EASTERN lineage (Andong) remains agricultural and

is purposively maintained as a cultural heritage region in the country [61].

On the other hand, the earliest Anthropocene population bottleneck in a white forsythia distribution may have occurred in the UPPER-CENTRAL lineage (t1 = 1714, 1638–1721). Although historically distant and rather broad in timescale interval, studies that investigated climate anomalies based on meteorological records from the 1600s to 1910 during the Joseon Dynasty (e.g., [62–64]) were able to pinpoint abnormally cold and extended winters, alluding to the causes and effects of the version of the Little Ice Age [65] on the KP. The climate anomalies and natural disasters resulted in several famines, forcing the locals to deplete forest resources for food, and especially for firewood (see [66] for the 250-year impacts of the 'ondol' heating system on Korean forest degradation). We believe that these unfortunate events may have caused the loss of the vegetation where white forsythia co-occurred (e.g., pine and oak forests), if not directly the destruction of the woody shrub colonies themselves. Situated in an old agricultural community, the UPPER-CENTRAL GS-NM221 population was the one most likely severely impacted by the phenomena in pre-modern Korean history.

Aside from findings of Anthropocene population bottlenecks, the best-fit demographic inference for the core central groups (Scenario 3) also revealed population expansion events in more distant periods in the past for the UPPER-CENTRAL (t2 = ca. 147 k ya, 110 k–184 k ya) and the LOWER-CENTRAL (t2 = ca. 154 k ya, 115 k–192 k ya) lineages (Table 6). The results aligned with [35] findings, which proposed that the current distribution of white forsythia is a result of divergence events from a core central distribution. Our demographic history inference in this work confirmed that the members of this major source of post-glacial expansion (i.e., populations from the core central lineages) are the closest descendants of the species' ancestral population [35], which, as suggested here, existed at least since the Last Interglacial. Future work on the demographic history of the ungrouped, individual populations in the core central range will more clearly explain a deeper evolutionary history of this lineage.

In terms of conservation, the particularly low effective population size of geographically marginal lineages suggests that these genetic groups suffer from the lack of gene flow (and suitable habitats; see the next section discussion). The revealed population size contraction patterns (i.e., higher Ne estimates in the central region than at peripheries) are supported by our MIGRATE-N inference of decreasing values of mutation-scaled effective population size theta as the distance from the central distribution increases (Table 4). The generally low

population size values (both Ne and theta) are also in agreement with the positive within-population Tajima's D (Table 1), which can either indicate population declines or balancing selection (e.g., [67, 68]). Here, we believe that our sampled populations have been more likely impacted by the former because balancing selection only often targets certain loci and not the entire genome. In particular, the lowest effective population size (Ne = 1) and theta (Θ = 0.00657) in the marginal NORTHERN YJ, are in good agreement with the highest Tajima's D (D = 1.123), and lowest genetic variation (π = 0.00014) findings recorded for the population. These indices suggest that NORTHERN YJ experienced a population decline that is not only recent but one that is also severe, emphasizing the immediate need for its conservation management.

The challenge of changing climate

Niche contraction–expansion and latitudinal shifts enabled white forsythia to survive the harsh environmental conditions after the LGM [35]. To take a peek at how this endangered species will respond to the ongoing and future climate change, we inferred its possible future distribution under two varying SSP climatic scenarios. Results showed that environmental gradients were optimal in areas surrounding the central distribution, but habitat suitability forecast for those in the marginal populations was poor, more so in areas in North Korea where sampled occurrences were last reported almost 100 years ago [26, 28, 69]. Even in the species' distributional range in South Korea, areas around the existing marginal limits appeared inhospitable, implying very little to no chance of range expansion. For instance, the further outward spread of the populations in the SOUTHERN and EASTERN lineages is highly unlikely, while the migration of the NORTHERN YJ from its current location to higher latitudes could be possible towards the northwest and northeast directions, the latter only feasible following what seems to be a long, narrow, and rugged corridor (Fig. 6).

On the other hand, the climatic variables that were found most useful are in good agreement with [35] findings (i.e., the same top two environmental predictors bio5 and bio14). Specifically, the highest contributing climatic predictor bio14 precipitation amount of the driest month (64%) is highly likely associated with the species' flowering and pollination mechanisms in late winter to early spring. Also, as mature, previous-year white forsythia samaras begin to fall off around winter, or even remain attached to the mother plant until spring (see Fig. 1), the southeastward winds that the winter East Asian monsoon brings imply that fruit dispersal is still at work even at this time of the year. This probable monsoonal

Ong *et al. BMC Ecology and Evolution*   (2024) 24:123

Page 14 of 21

dispersal mechanism appears to support our most possible migration model inference, which showed stronger between-lineage gene flow rates in a similar, more dominant southward direction (Fig. 5).

Overall, our ENM forecasts for white forsythia occurrences in South Korea suggest that the farther the distance from the central region, the less suitable environmental conditions become. This is because species abundance distributions are coupled with environmental gradients, and it is when this climatic gradient begins to deteriorate, is when populations start to decline [70]. In terms of conservation, our findings suggest the need for further regional identification of areas that will show topographical and latitudinal variation. Identifying geographic zones that are buffered from the warming climate, in conjunction with well-studied socio-economic development planning, will enable conservation managers to predict how plant populations may respond to future environmental disturbances both natural and anthropogenic [71].

Conclusions

The knowledge of white forsythia genetic variation and structure in connection with its landscape and gene flow patterns, demographic history, and future distribution is important for evaluating its evolutionary potential and survival in the face of changing climate and anthropogenic threats. In this research installment to the most recent population genomic and phylogeographic studies on the species [34, 35], we incorporated the use of haplotype and geographic information to more finely analyze the influence of genetic-landscape dynamics in this KP endemic.

Our findings emphasized the recognition of regional patterns of genetic structure and the role of topography as some of the possible factors that may have influenced the species' present-day fragmented population distribution. The bottleneck events in the Anthropocene, some of which aligning with the time of historical catastrophic events (e.g., the Korean War), were also revealed to have contributed to the low effective population size, particularly those of the marginal populations. Future distribution under both optimistic and pessimistic climatic scenarios suggests unlikely suitable habitats for these populations to expand from their current range limits, at least in the next 80 years. With our findings of generally weak historical gene flow and small effective population size, and the expected continuing anthropogenic land-use change in a highly developed country like South Korea, the poor population connectivity among the already fragmented white forsythia habitats will remain a big challenge for conservation. We hypothesize that population connectivity in the species has reached a level

at which habitat continuity is fractured, and that the current habitat fragmentation, following its definition ([72]; see [73] for review), signifies the final and likely irreversible stage of this process.

Regardless, we still would like to believe that the species' long-lived, perennial life form and its ability to reproduce vegetatively can help mitigate the effects of fragmentation. The small official reserve status in the form of natural monument habitats can help ensure the protection of these reservoirs of genetic diversity. In line with this, the existence of conservation-designed environments, such as *ex-situ* habitats in arboreta, or even individuals being cared for in home gardens, will play a crucial role in some ecological processes, such as in pollination, by providing niches for insect vectors [13]. To help achieve this conservation goal, the merging of various approaches in population genetics, landscape ecology, and plant-animal interactions is encouraged. With fast and powerful sequencing technologies being made more available for non-model organisms, the use of genomic data to their full potential, such as incorporating RAD loci haplotype-based analyses will provide much improved tools for the conservation of the endangered Korean endemic white forsythia, and other similar plants under pressure.

Methods

Sampling, library preparation, sequencing, and bioinformatics

We reanalyzed subsamples of data used in white forsythia population genomics study by [34]. These samples represented individuals collected from nine naturally occurring populations (Fig. 1), covering a maximum of 15 individuals per site. More detailed information about these locations is provided in [35]. Extracted DNAs from these samples were prepared for GBS library protocol using *ApeKI* restriction enzyme [74] and sequenced in high-throughput Illumina machines (USA). For more information about these procedures, see [34] and/or [35].

Unlike the above earlier studies, here, short-read Illumina sequences for bioinformatics were reprocessed using STACKS v.2.62 software (STACKS 2) to take advantage of its implemented programs and features, particularly the processing and outputting of data for RAD loci-related analyses [75]. Demultiplexing, trimming, and quality screening of paired-end reads were conducted using the 'process_radtags' program. Quality score of each read was checked using the default raw phred score (-q 10). All loci were de novo assembled by running the 'denovo_map.pl' program after performing parameter optimization using the r80 method [76, 77] as recommended in STACKS 2 protocol [78]. The program 'ustacks' was then used to assemble and pile reads/stacks

Ong *et al. BMC Ecology and Evolution*   (2024) 24:123

Page 15 of 21

with the default minimum depth of coverage (-m 3), and the r80 method-optimized parameters of maximum distance allowed between stacks (-M 5) and distance allowed between catalogued loci (-n 5). The products from 'ustacks' were then searched against the catalog in the program 'sstacks', while the catalog was built using the program 'cstacks'. After loci assembly, we manually removed samples with less than 8x depth of coverage.

Using the STACKS 2 'populations' program, we built a genotype matrix by applying the following filtering options: keeping SNPs present in all nine sampled populations (-p 9), removing SNPs with minor allele frequency less than 5% (-min-maf 0.05), retaining SNPs present in more than 90% of individuals across all populations (-R 0.9), retaining SNPs present in more than 50% of individuals in each population (-r 0.5), specifying a maximum observed heterozygosity to process a nucleotide (-max-obs-het 0.6), computing divergence from Hardy–Weinberg equilibrium (-hwe), and allowing the first SNP per RAD locus in data analysis (-write_single_snp).

Overall, a total of 968 SNPs, and 3,547 RAD loci comprising ca. 638,000 sites were detected across 131 individuals with 14x average depth of coverage. The above SNP (variant sites) and RAD loci (variant and invariant sites) datasets (in vcf files) were used as the main ('neutral') markers in all downstream analyses, unless other datasets with different quality filters (and file format) are specified (e.g., for analysis in fineRADstructure and MIGRATE-N). The raw sequence data (fastq files) used in this research can be found in the Dryad repository: https://doi.org/10.5061/dryad.69p8cz9b2.

Haplotype-based genomic variation and structure

Via the 'populations' program in STACKS 2, we used the haplotype-based statistics feature to estimate the within-population pairwise nucleotide diversity pi (π). We also ran the between-populations analyses of absolute measure of locus differentiation or nucleotide divergence D_{XY} [79], and the AMOVA-based measure [80] pairwise Phi_{ST} (Φ_{ST}). The number of private alleles per population was similarly computed using the same program. In the software VCFtools [81], we employed our RAD loci dataset once again to calculate another haplotype-based statistic Tajima's D (D) in each population [82] after setting a window size of 10 kb (-TajimaD 10000).

To run a population structure analysis specifically designed for RAD loci, we used the software fineRADstructure [83]. The program uses information on haplotype linkage among individuals to produce a summary of nearest-neighbor haplotype relationships in the form of a co-ancestry matrix coefficient without prior assumptions of populations. Individual co-ancestry at each locus is equally divided among all other individuals with identical haplotypes, and the rare haplotypes defined by rare SNPs contribute the most to the co-ancestry index, providing a measure that emphasizes recent co-ancestry relationships [e.g., 38, 39]. The program uses a Bayesian approach to select the most probable configuration based on likelihood ratios between MCMC samples [83].

The input file for fineRADstructure was separately produced after running STACKS 2 'populations' program option -radpainter by allowing filtering parameters of 5% minor allele frequency (-maf 0.05), and the selection of variants present in at least 80% of samples in the population (-r 0.8). The generated file was subsequently converted using the python script Stacks2fineRAD.py [83]. Further filtering for the allowable number of SNPs per RAD locus and missing data across individuals (-n 2 -m 40) resulted in a matrix composed of 26,085 RAD loci (including 18,301 variants) across all our samples. The individual co-ancestry matrix to assign individuals to populations was run with the 100,000 burn-in iterations, 500,000 MCMC sampling iterations, and a thinning parameter of 1,000 (-$\times$100000 -y 500000 -z 1000), while the tree-building algorithm was run using the default parameters [84]. To visualize the results, we used the scripts fineradstructureplot.r and finestructurelibrary.r (available at https://github.com/millanek/fineRADstructure).

Genetic and geographic barriers

Given the non-linear, central-marginal distribution pattern earlier identified for white forsythia [34], a genetic barrier analysis using Monmonier's maximum difference algorithm [85, 86] was employed to assess possible genetic discontinuities (e.g., due to physical, geographic barriers) in the species' fragmented population distribution. First, we constructed a valuated graph via Delaunay triangulation of our population coordinates, with edge values reflecting (an earlier computed) between-population Phi_{ST} distances of RAD loci. We used the 'optimize.monmonier' function in the R package *adegenet* v2.1.1 [87] to identify and intersect the strongest genetic distances between neighboring population points using Euclidean distance, with the distance threshold between immediate neighbors being chosen as an abrupt decrease between the connected points (e.g., [88–90]). In this analysis, 10 different starting points were used to find the largest sum of local distances that explains the genetic distances among populations.

To further investigate the relationship between genetic and geographic data, we ran a Procrustes analysis, which is an optimal transformation test that maximizes the similarity between genetic diversity and geographic location (e.g., [6], [91, 92]). In this analysis, we applied the function 'procrustes' in the R package *vegan* v.2.6–4

Ong *et al. BMC Ecology and Evolution*   (2024) 24:123

Page 16 of 21

[91]. Here, two matrices were compared: the matrix for geographic coordinates (i.e., latitude and longitude) of sampled populations, and the matrix for principal components (i.e., PC1 and PC2) of our SNP dataset, performed using the function 'dudi.pca' in the R package *ade4* v1.7–22 [92]. The association/similarity statistic between the two matrices (t_0) was then computed using the formula $t_0 = \sqrt{1-D}$, where D is the minimum sum of the squared Euclidean distances between the two maps which were scaled to range between 0 and 1 [93, 94]. We assessed the statistical significance of this test by running 10,000 permutations using the 'protest' function in *vegan* to test the probability of observing a similarity statistic higher than the observed t_0 under the null hypothesis that no geographic pattern exists in the population structure [94].

Population connectivity patterns

The coalescent-based approach of the program MIGRATE-N v.4.4.3 [95] was used to estimate the mutation-scaled effective population size theta ($\Theta = 4Ne\mu$, where Ne represents the effective population size and µ the mutation rate per generation per locus), and the between populations mutation-scaled migration rate M ($M = m/\mu$, where m represents the migration rate per generation, and µ the mutation rate). Here, we employed our RAD dataset (i.e., sequence data) which better suits the evolutionary models of the program than the use of SNPs (variants-only sites) due to the possible ascertainment bias associated with the latter [96].

The input file was prepared by running the -fasta-samples command in STACKS 2 'populations' program to produce a per-locus, per-haplotype (fasta file) output, and by converting it using the Python script fasta2genotype.py found in [97]. To balance between population representation and computational demand, we clustered our samples into five genetic groups earlier identified in [34, 35], and randomly selected only 100 RAD loci (averaging ca. 180 bp each) from the total pool of our original dataset.

In all, we evaluated seven population connectivity models as follows: (i) Full Migration (default), which assumes migration paths for all lineages are open; (ii) UPPER-CENTRAL Symmetrical Migration, which assumes two-way gene flow between the UPPER-CENTRAL lineage and all groups, while migration paths for all other paired lineages are closed; (iii) UPPER-CENTRAL Asymmetrical Migration, which assumes one-way gene flow from the UPPER-CENTRAL lineage to all groups, while migration paths for all other paired lineages are closed; (iv) Non-connecting Marginals Migration, which assumes gene flow is absent between/among the three marginal lineages, while migration paths for all other

paired lineages are open; (v) Isolated-EASTERN Migration, which assumes gene flow is absent only between the EASTERN lineage and each of the two other marginal lineages, while migration paths for all other paired lineages are open; (vi) Isolated-NORTHERN Migration, which assumes gene flow is absent only between the NORTHERN lineage and each of the two other marginal lineages, while migration paths for all other paired lineages are open; and (vii) Isolated-SOUTHERN Migration, which assumes gene flow is only absent between the SOUTHERN lineage and each of the two other marginal lineages, while migration paths for all other paired lineages are open. See Additional file 4 for model diagrams and rationale for model assumptions.

For each migration model, the following parameters were set: Felsenstein 84 mutation model; default run values consisting of one long chain of 10,000 recorded steps, 100 increments, and a burn-in value of 1,000; default setting of uniform prior distribution parameters; setting of four replicates to sample a total of 4,000,000 parameter values; and a static heating scheme of four chains with default values 1, 1.5, 3, and 1,000,000. The number of migrants (Nm) entering one population from another per generation was also calculated using the formula $Nm = \Theta M/4$ [98].

To select the most probable gene flow scenario, the models were compared and ranked following the procedure described in the program protocol [99]. The method utilizes the Bezier approximation scores of marginal likelihood between different gene flow models and calculates their specific probability based on log Bayes factor (LBF), a very robust method even with incomplete population sampling and non-normality [98, 100].

Past population size changes inference

Using our SNP dataset, we clustered our populations into five genetic groups (as in our migration analysis) and inferred their possible historical population size reduction and/or expansion. We employed the approximate Bayesian computation method (ABC), a powerful and flexible approach for estimating demographic parameters by testing and comparing the most probable models/scenarios of change in population size parameters in units of effective population size (Ne) from that of the ancestral population (Nanc) to the effective population size(s) after bottleneck (Nbot) and/or expansion (Nexp) events at certain time parameters (e.g., t1, t2) in units of number of generations [101]. We used the program DIYABC v2.1 [102], specifically its function to calculate the population size distributions in single-population models by evaluating related summary statistics (sumstats) and predefined priors.

Ong *et al. BMC Ecology and Evolution*   (2024) 24:123

Page 17 of 21

For each lineage, we tested six demographic models, as shown in Fig. 7 (see Additional file 5 for prior distribution parameter settings). The input file was converted into the DIYABC format using the Python script vcf2DIYABC.py (link available at the program's website) and prepared using Hudson's simulation algorithm for SNP markers [103], which is equivalent to setting the minor allele frequency criterion to default [102]. We selected all single-sample gene diversity [79] sumstats available, as follows: (i) proportion of loci with null gene diversity (proportion of zero values), (ii) mean gene diversity across polymorphic loci (mean of non-zero values), (iii) variance of gene diversity across polymorphic loci (variance of non-zero values), and (iv) mean gene diversity across all loci (mean of complete distribution). Overall, we employed 11 historical parameters (Ne1, Ne2, Ne3, Ne4, Ne5, Ne6, Nanc, Nbot, Nexp, t1, and t2) and the four sumstats combinations to generate a reference table based on 6×10^6 simulated datasets (ca. one million runs per scenario).

To determine the most highly supported demographic model for each lineage, posterior probabilities were computed via the logistic regression method based on the 1% of simulated datasets closest to the observed data after replacing the original sumstats with discriminant scores [102, 104]. To evaluate how well the most highly supported scenario and its prior and posterior parameters fit the data (i.e., the goodness-of-fit), we ran the DIYABC 'model checking' option. Finally, after selecting the 'parameter estimation' option, the posterior distribution parameters for the scenario of choice were computed by selecting the logit transformation on the 1% of the closest simulated datasets. The mean posterior distribution

for parameter t was recalibrated to the absolute time (i.e., into calendar years) by multiplying it with the generation time of the species, which was set to 8 [35] and/or 6–10 years [34].

Present and future habitat suitability

We employed ecological niche modeling (ENM; also called species distribution modeling) to infer the species' suitable habitats in the present and predict its probable distribution in the future. Occurrence data from South Korea (22 points) were taken from [35], while reports of occurrences from North Korea (2 points) were obtained from Global Diversity Information Facility (GBIF), as specimen voucher-verified by [69, 105]. These 24 occurrences were then filtered at ca. 1 km radius using the R package *spThin* v0.2.0 [106], while the calibration area was set to the geographic extent of the KP (i.e., mainland geopolitical areas of South and North Koreas; see Fig. 1). Still, we acknowledge that the study region occupies a relatively larger area than the species' known distribution, hence, our output maps should be interpreted with caution.

A total of 19 Present (1970–2000) bioclimatic variables (e.g., temperature and precipitation predictors) were downloaded from WorldClim database v1.4 [107]. The variables represented the highest spatial resolution of 30 arcsec or ca. 1×1 km and comprised sumstats at different temporal resolutions interpolated from WorldClim weather station data. To retrieve the environmental layers of the study region, we used QGIS v3.22.5 geographic information system software [108]. Spatial resolutions of
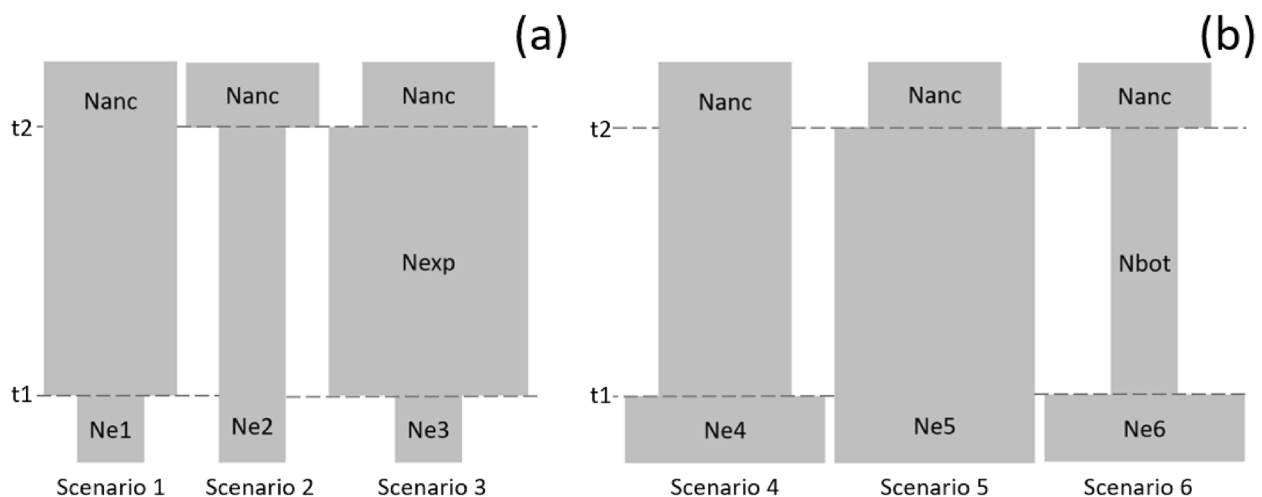

Fig. 7 Six ABC demographic scenarios inferred for each lineage, categorized into (**a**) Reduction (1–3), and (**b**) Expansion (4–6) models, showing the effective population size in each group (e.g., Ne1, Ne2, Ne3, Ne4, Ne5, Ne6) due to change to of the ancestral effective population size (Nanc) after expansion (Nexp) and/or contraction (Nbot) at one and/or two different times of events (e.g., t2, t1) in the past

Ong *et al. BMC Ecology and Evolution*   (2024) 24:123

Page 18 of 21

these layers were uniformly adjusted using the R package *raster* v3.6.26 [109].

Using *ENMTools* v1.1.2 [110] and *usdm* v4.3.2 [111] R packages, we ran several distribution models to eliminate variables that showed combinations of high correlations (r $\geq$ 0.8) and those with very high variance inflation factor (VIF > 10). We also removed the ones with the least percent contribution (< 1%) to the overall model by initially screening them in MAXENT v3.4.4 based on a stepwise removal using the jackknife test [112]. Below are the final five environmental variables that showed the highest contribution to the model prediction: temperature seasonality (bio4), maximum temperature of the warmest month (bio5), precipitation amount of the driest month (bio14), precipitation seasonality (bio15), and precipitation of the wettest quarter (bio16).

Before the final maximum entropy runs, we first optimized candidate niche models using the R package *ENMeval* v2.0.4 [113] to avoid model over-prediction, which is one of the downsides often cautioned with the use of MAXENT default settings [114]. Over-fitting and complexity of the models were evaluated by varying the regularization multiplier (RM) (e.g., 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4) and by testing different feature class combinations of linear (L), quadratic (Q), hinge (H), and product (P) constraints (e.g., L, Q, H, P, LQ, LH, LP, QH, QP, HP, LQH, QHP, LQHP). We selected the jackknife partition settings to 24 (corresponding to our 24 occurrence points), resulting in a total of 104 niche model candidates and the outputting of a sampling bias file to be used for the final MAXENT run. The best niche model parameters were selected by choosing the delta.AICc with zero value.

As recommended by MAXENT program authors [112], we selected the complementary log–log output (cloglog) setting to calculate the species' probability of presence. We also selected the following *ENMeval* optimized parameter combinations: RM of 1, HP feature class combinations, and a maximum number of 10,000 background points. The following options were additionally selected: initial random seed for each iteration, jackknife to measure variable importance, cross-validation method of replication by running 24 replicates, a maximum iteration of 1,000, default prevalence of 0.5, threshold rule of 10% training presence, and the use of the *ENMeval*-produced sampling bias file. All other parameters were set to default. We referred to the area under the curve (AUC) of the Receiver Operating Characteristic (ROC) to evaluate the discriminatory capacity of the predicted model.

The model that was calibrated onto the Present (1970–2000) conditions was then projected to future climatic settings using the same five environmental predictors downloaded from CHELSA future climatologies database [115], specifically those produced by Coupled Model Intercomparison Project (CMIP6) from UKESM1-0-LL. In this paper, we modeled the species' probable distribution under two differing shared socioeconomic pathways [116]: the optimistic SSP1-26 ('Sustainability'), and rather pessimistic SSP3-70 ('Regional Rivalry') scenarios for up to about the next 20 to 80 years (i.e., 2011–2040, 2041–2070, and 2071–2100 time periods). Following the threshold value for each temporal model, the resulting continuous maps were then converted to binary output maps to project the suitable and unsuitable areas (or presence and absence of distribution) from the present to the different climatic scenarios in the future. All ENM maps were prepared using the QGIS software [108].

Supplementary Information

Availability of data and materials
The raw genotype data (fastq files) is available on Dryad repository (https://doi.org/10.5061/dryad.69p8cz9b2), while some supporting information are within the article as Additional files.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Ong *et al. BMC Ecology and Evolution*   (2024) 24:123

Page 19 of 21

Author details
[1]Multidisciplinary Genome Institute, Hallym University, Chuncheon 24252, South Korea. [2]Department of Life Science, Hallym University, Chuncheon 24252, South Korea. [3]On Biological Resource Research Institute (OBRRI), Chuncheon 24252, South Korea. [4]National Institute of Biological Resources (NIBR), Incheon 22689, South Korea. [5]Ecosystem Research Division, Korea National Park Research Institute, Wonju 26441, South Korea.

References

1. Watson DM. A conceptual framework for studying species composition in fragments, islands and other patchy ecosystems. J Biogeogr. 2002;29(5–6):823–34.
2. Fields PD, Reisser C, Dukić M, Haag CR, Ebert D. Genes mirror geography in *Daphnia magna*. Mol Ecol. 2015;24(17):4521–36.
3. De Vriendt L, Lemay MA, Jean M, Renaut S, Pellerin S, Joly S, et al. Population isolation shapes plant genetics, phenotype and germination in naturally patchy ecosystems. J Plant Ecol. 2017;10(4):649–59.
4. Kay KM, Woolhouse S, Smith BA, Pope NS, Rajakaruna N. Sympatric serpentine endemic *Monardella* (Lamiaceae) species maintain habitat differences despite hybridization. Mol Ecol. 2018;27(9):2302–16.
5. Gutiérrez-Ortega JS, Salinas-Rodríguez MM, Ito T, Pérez-Farrera MA, Vovides AP, Martínez JF, et al. Niche conservatism promotes speciation in cycads: the case of *Dioon merolae* (Zamiaceae) in Mexico. New Phytol. 2020;227(6):1872–84.
6. Ren G, Mateo RG, Conti E, Salamin N. Population genetic structure and demographic history of *Primula fasciculata* in Southwest China. Front Plant Sci. 2020;11:551329.
7. Wright S. Isolation by distance. Genetics. 1943;28(2):114–38.
8. Loveless MD, Hamrick JL. Ecological determinants of genetic structure in plant populations. Annu Rev Ecol Syst. 1984;15:65–95.
9. Young A, Boyle T, Brown T. The population genetic consequences of habitat fragmentation for plants. Trends Ecol Evol. 1996;11(10):413–8.
10. Sork VL, Nason J, Campbell DR, Fernandez JF. Landscape approaches to historical and contemporary gene flow in plants. Trends Ecol Evol. 1999;14(6):219–24.
11. Aguilar R, Quesada M, Ashworth L, Herrerias-Diego Y, Lobo J. Genetic consequences of habitat fragmentation in plant populations: susceptible signals in plant traits and methodological approaches. Mol Ecol. 2008;17(24):5177–88.
12. Saunders DA, Hobbs RJ, Margules CR. Biological consequences of ecosystem fragmentation: a review. Conserv Biol. 1991;5(1):18–32.
13. Newman BJ, Ladd P, Brundrett M, Dixon KW. Effects of habitat fragmentation on plant reproductive success and population viability at the landscape and habitat scale. Biol Conserv. 2013;159:16–23.
14. Crutzen PJ, Brauch HG. Paul J. Crutzen: a pioneer on atmospheric chemistry and climate change in the Anthropocene. Vol 50. Springer; 2016.
15. Steffen W, Crutzen PJ, McNeill JR. The Anthropocene: are humans now overwhelming the great forces of nature. Ambio. 2007;36(8):614–21.
16. Steffen W, Grinevald J, Crutzen P, McNeill J. The Anthropocene: conceptual and historical perspectives. Philos Trans R Soc A. 1938;2011(369):842–67.
17. Ellis EC, Kaplan JO, Fuller DQ, Vavrus S, Klein Goldewijk K, Verburg PH. Used planet: a global history. Proc Natl Acad Sci USA. 2013;110(20):7978–85.
18. Ruddiman WF, Ellis EC, Kaplan JO, Fuller DQ. Defining the epoch we live in. Science. 2015;348(6230):38–9.
19. Lawrence MJ, Stemberger HL, Zolderdo AJ, Struthers DP, Cooke SJ. The effects of modern war and military activities on biodiversity and the environment. Environ Rev. 2015;23(4):443–60.
20. Jump AS, Peñuelas J. Running to stand still: adaptation and the response of plants to rapid climate change. Ecol Lett. 2005;8(9):1010–20.
21. McKay JK, Christian CE, Harrison S, Rice KJ. "How local is local?"—a review of practical and conceptual issues in the genetics of restoration. Restor Ecol. 2005;13(3):432–40.
22. National Geographic Information Institute. The national atlas of Korea I; 2019. http://nationalatlas.ngii.go.kr/pages/page_1949.php. Accessed 06 January 2024.
23. Cultural Heritage Administration. Explore heritage, heritage search. 2006. http://english.cha.go. kr/ chaen/search/selectGeneralSearchList. do?mn=EN_02_02. Accessed 02 May 2021.
24. Korea Protected Areas Forum. Natural monuments and nature reserves of Korea. 2009. http://www.pafor um.or.kr/area/area.html?cate_idx=19. Accessed 06 Jan 2021.
25. Kim YS, Lee YM, Chun SH, Jeon JI, Kim SH. Necessity of recovery plan for the conservation of rare and endangered plants in Korea. Bull Seoul Natl Univ Arboretum. 1995;15:43–66 (In Korean).
26. Kim YS, Maunder M. Plants in peril, 24 *Abeliophyllum distichum*. Curtis's Bot Mag. 1998;15(2):141–6.
27. Lim DO, Hwang IC, Choi HW, Kim YS. Ecological characteristics and management proposal of *Abeliophyllum distichum* subpopulations in the Byeonsanbando National Park. Kor J Env Eco. 2009;23(2):116–26 (In Korean).
28. Son SW, Kim YS, Kim, H. *Abeliophyllum distichum*. IUCN Red List Threatened Species. 2016; e.T13188339A13189399. https://doi.org/10.2305/IUCN.UK.2016-1.RLTS.T13188339A13189399.en. Accessed 9 Sept 2021.
29. Korea National Arboretum. National red list of vascular plant in Korea. Pocheon: Korea National Arboretum; 2021. (In Korean).
30. National Institute of Biological Resources. Red data book of Republic of Korea: Vascular plants. 2nd ed., vol. 5. Incheon: National Institute of Biological Resources; 2021. (In Korean).
31. Lee SW, Kim SC, Lee HS. Allozyme variation in *Abeliophyllum distichum* Nakai, an endemic tree species of Korea. Silvae Genet. 1998;47(5):294–7.
32. Chung MG. Allozyme diversity in the endangered shrub *Abeliophyllum distichum* (Oleaceae): a monotypic Korean genus. Int J Plant Sci. 1999;160(3):553–9.
33. Kang U, Chang CS, Kim YS. Genetic structure and conservation considerations of rare endemic *Abeliophyllum distichum* Nakai (Oleaceae) in Korea. J Plant Res. 2000;113:127–38.
34. Lee JH, Ong HG, Kim BY, Kim YI, Jung EK, Chung MG, et al. Population genomics study for the conservation management of the endangered shrub *Abeliophyllum distichum*. Conserv Genet. 2022;23(4):683–97.
35. Ong HG, Kim YI, Lee JH, Kim BY, Kang DH, Jung EK, et al. Approximate Bayesian computation and ecological niche models elucidate the demographic history and current fragmented population distribution of a Korean endemic shrub. Ecol Evol. 2023;13(12):e10792.
36. McKinney GJ, Seeb JE, Seeb LW. Managing mixed-stock fisheries: genotyping multi-SNP haplotypes increases power for genetic stock identification. Can J Fish Aquat Sci. 2017;74(4):429–34.
37. Fedorov VB, Trucchi E, Goropashnaya AV, Waltari E, Whidden SE, Stenseth NC. Impact of past climate warming on genomic diversity and demographic history of collared lemmings across the Eurasian Arctic. Proc Natl Acad Sci USA. 2020;117(6):3026–33.
38. Bootsma ML, Miller L, Sass GG, Euclide PT, Larson WA. The ghosts of propagation past: haplotype information clarifies the relative influence of stocking history and phylogeographic processes on contemporary population structure of walleye (*Sander vitreus*). Evol Appl. 2021;14(4):1124–44.
39. Lavretsky P, Wilson RE, Talbot SL, Sonsthagen SA. Phylogenomics reveals ancient and contemporary gene flow contributing to the evolutionary history of sea ducks (Tribe Mergini). Mol Phylogenet Evol. 2021;161:107164.
40. Obiol JF, Herranz JM, Paris JR, Whiting JR, Rozas J, Riutort M, et al. Species delimitation using genomic data to resolve taxonomic uncertainties in a speciation continuum of pelagic seabirds. Mol Phylogenet Evol. 2023;179:107671.
41. Leitwein M, Duranton M, Rougemont Q, Gagnaire PA, Bernatchez L. Using haplotype information for conservation genomics. Trends Ecol Evol. 2020;35(3):245–58.
42. Cho MS, Takayama K, Yang J, Maki M, Kim SC. Genome-wide single nucleotide polymorphism analysis elucidates the evolution of *Prunus takesimensis* in Ulleung island: the genetic consequences of anagenetic speciation. Front Plant Sci. 2021;12:706195.
43. Lee SR, Choi TY, Jung SY. Genetic diversity on a rare terrestrial orchid, *Habenaria linearifolia* in South Korea: implications for conservation

offered by genome-wide single nucleotide polymorphisms. Front Plant Sci. 2022;13:772621.

44. Kang H, An SM, Park YJ, Lee YB, Lee JH, Cheon KS, et al. Population genomics study and implications for the conservation of *Zabelia tyaihyonii* based on genotyping-by-sequencing. Plants. 2022;12(1):171.

45. Hwang G, Jung UC, Kim ST. A survey of the genome-wide genetic variation of *Hibiscus hamabo* (Malvaceae). Korean J Pl Taxon. 2023;53(2):148–56.

46. Choi Y, Ha YH, Choi J. Genetic diversity assessment of a plant for forest restoration on the Korean Peninsula: a case study of Lespedeza cuneata G. Don (Fabaceae). J Asia-Pac Biodivers. 2024;17(1):35–42.

47. Chung MY, Son SW, Suh GU, Herrando-Moraira S, Lee CH, Lopez-Pujol J, et al. The Korean Baekdudaegan mountains: a glacial refugium and a biodiversity hotspot that needs to be conserved. Front Genet. 2018;9:395213.

48. Schierenbeck KA. Population-level genetic variation and climate change in a biodiversity hotspot. Ann Bot. 2017;119(2):215–28.

49. Bagnoli F, Della Rocca G, Spanu I, Fineschi S, Vendramin GG. The origin of the Afro-Mediterranean cypresses: evidence from genetic analysis. Perspect Pl Ecol Evol Syst. 2020;46:125564.

50. Hou H, Ye H, Wang Z, Wu J, Gao Y, Han W, et al. Demographic history and genetic differentiation of an endemic and endangered *Ulmus lamellosa* (*Ulmus*). BMC Plant Biol. 2020;20:1–14.

51. Chang JT, Chao CT, Nakamura K, Liu HL, Luo MX, Liao PC. Divergence with gene flow and contrasting population size blur the species boundary in Cycas sect. Asiorientales, as inferred from morphology and RAD-seq data. Front Plant Sci. 2022;13:824158.

52. Kou Y, Fan D, Cheng S, Yang Y, Wang M, Wang Y, et al. Peripatric speciation within *Torreya fargesii* (Taxaceae) in the Hengduan Mountains inferred from multi-loci phylogeography. BMC Ecol Evol. 2023;23(1):74.

53. Lee SR, Son DC. Genetic diversity pattern reveals the primary determinant of burcucumber (Sicyos angulatus L.) invasion in Korea. Front Plant Sci. 2022;13:997521.

54. Choi IS, Han EK, Wojciechowski MF, Heo TI, Park JS, Yang JC, et al. The genetic structure and demographic history of *Zabelia tyaihyonii*, endemic to Korean limestone karst forests, based on genome-wide SNP markers. Ecol Evol. 2023;13(7):e10252.

55. Han EK, Tamaki I, Oh SH, Park JS, Cho WB, Jin DP, et al. Genetic and demographic signatures accompanying the evolution of the selfing syndrome in *Daphne kiusiana*, an evergreen shrub. Ann Bot. 2023;131(5):751–67.

56. Moon K. ROK army operations in the Jirisan region during the Korean war: David Galula's counterinsurgency theory in action. Kansas: Fort Leavenworth; 2020. chromeextension://efaidnbmnnnibpcajpcglclefi ndmkaj/https://apps.dtic.mil/sti/pdfs/AD1159463.pdf. Accessed 6 Jan 2024.

57. Millett AR. The Korean war. Korea Institute of Military History. Volume 1. Nebraska: University of Nebraska Press; 2000.

58. Appleman RE. South to the Naktong, north to the Yalu. Washington: US Government Printing Office; 1961.

59. Lee HY, Kim TG, Oh CH. Recently augmented natural habitat of *Abeliophyllum distichum* Nakai in Yeoju-si, Gyunggi-do. Korea Korean J Environ Ecol. 2014;28:62–70 (In Korean).

60. Jeong JY, Kang YJ, Noh KW. Andong modern history: society, economy, vol. 3. Andong: Seongsim Book Publishing; 2010. (In Korean).

61. Lee S, Jung JH, Kwon D. Reconciling the conservation of cultural heritage with rural development: the "Andong" City as a case study. M/C Journal. 2022;25(3). https://doi.org/10.5204/mcj.2904.

62. Yi TJ. Meteor fallings and other natural phenomena between 1500–1750 as recorded in the annals of the Chosŏn dynasty (Korea). Celest Mech Dyn Astr. 1997;69(1):199–220.

63. Kim S. Successive volcanic eruptions (1809–1815) and two severe famines of Korea (1809–1810, 1814–1815) seen through historical records. Clim Change. 2023;176(1):1.

64. Lee BMS. Climate change, inequality, and vulnerabilities in pre-modern Korea: implications for mission after COVID-19. Transform Int J Holi. 2023;40(2):95–106.

65. Mann ME. Little ice age. In: MacCracken MC, Perry JS, editors. Encyclopedia of global environmental change. Volume 1. Chichester: John Wiley & Sons, Ltd; 2002. p. 504–509. http://www.meteo.psu.edu/holoc ene/public_html/shared/articles/littleiceage.pdf. Accessed 6 Jan 2024.

66. Bae JS, Kim YS. History lessons from the late Joseon Dynasty period of Korea: human technology (ondol), its impacts on forests and people, and the role of the government. Forests. 2020;11(12):1314.

67. Wu XY, Li TZ, Zheng F, Chen JB, Yan YH, Huang JX. Population diversity analysis provide insights into provenance identification of *Dendrobium catenatum*. Genes. 2022;13(11):2093.

68. Zhu X, Tang J, Jiang H, Yang Y, Chen Z, Zou R, et al. Genomic evidence reveals high genetic diversity in a narrowly distributed species and natural hybridization risk with a widespread species in the genus *Geodorum*. BMC Plant Biol. 2023;23(1):317.

69. Chang CS. Flora of the Korean peninsula phase II. Version 1.6. TB Lee herbarium. Occurrence dataset. 2021. https://doi.org/10.15468/0vcvsq. Accessed via GBIF.org 13 Feb 2024.

70. Sagarin RD, Gaines SD. The 'abundant centre' distribution: to what extent is it a biogeographical rule? Ecol Lett. 2002;5(1):137–47.

71. Meng H, Gao X, Song Y, Cao G, Li J. Biodiversity arks in the Anthropocene. Reg Sustain. 2021;2(2):109–15.

72. Wilcove DS, McLellan CH, Dobson AP. Habitat fragmentation in the temperate zone. In: Soule ME, editor. Conservation biology: the science of scarcity and diversity. Massachusetts: Sinauer; 1986. p. 237–56.

73. Fahrig L. Effects of habitat fragmentation on biodiversity. Annu Rev Ecol Evol Syst. 2003;34(1):487–515.

74. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE. 2011;6(5):e19379.

75. Rochette NC, Rivera-Colón AG, Catchen JM. Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. Mol Ecol. 2019;28(21):4737–54.

76. Paris JR, Stevens JR, Catchen JM. Lost in parameter space: a road map for stacks. Methods in Ecol Evol. 2017;8(10):1360–73.

77. Rochette NC, Catchen JM. Deriving genotypes from RAD-seq shortread data using Stacks. Nat Protoc. 2017;12(12):2640–59.

78. Rivera-Colón AG, Catchen J. Population genomics analysis with RAD, reprised: Stacks 2. In: Verde C, Giordano D, editors. Marine genomics: methods and protocols. New York: Springer; 2022. p. 99–149.

79. Nei M. Molecular evolutionary genetics. New York: Columbia University Press; 1987.

80. Excoffier L, Smouse PE, Quattro J. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics. 1992;131(2):479–91.

81. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011;27(15):2156–8.

82. Tajima F. The effect of change in population size on DNA polymorphism. Genetics. 1989;123(3):597–601.

83. Malinsky M, Trucchi E, Lawson DJ, Falush D. RADpainter and fineRADstructure: population inference from RADseq data. Mol Biol Evol. 2018;35(5):1284–90.

84. Lawson DJ, Hellenthal G, Myers S, Falush D. Inference of population structure using dense haplotype data. PLoS Genet. 2012;8(1):e1002453.

85. Monmonier MS. Maximum-difference barriers: an alternative numerical regionalization method. Geogr Anal. 1973;5(3):245–61.

86. Manni F, Guérard E, Heyer E. Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by using Monmonier's algorithm. Hum Biol. 2004;76(2):173–90.

87. Jombart T, Ahmed I. adegenet 1.3–1: new tools for the analysis of genome-wide SNP data. Bioinformatics. 2011;27(21):3070–1.

88. Li YS, Shih KM, Chang CT, Chung JD, Hwang SY. Testing the effect of mountain ranges as a physical barrier to current gene flow and environmentally dependent adaptive divergence in *Cunninghamia konishii* (Cupressaceae). Front Genet. 2019;10:742.

89. Gasca-Pineda J, Gutiérrez-Guerrero YT, Aguirre-Planter E, Eguiarte LE. The role of environment, local adaptation, and past climate fluctuation on the amount and distribution of genetic diversity in two subspecies of Mexican wild *Zea mays*. Am J Bot. 2020;107(11):1542–54.

90. Hodel RG, Knowles LL, McDaniel SF, Payton AC, Dunaway JF, Soltis PS, et al. Terrestrial species adapted to sea dispersal: differences in propagule dispersal of two Caribbean mangroves. Mol Ecol. 2018;27(22):4612–26.

Ong *et al. BMC Ecology and Evolution*   (2024) 24:123

Page 21 of 21

91. Oksanen J, Simpson GL, Blanchet FG, Kindt R, Legendre P, Minchin PR, et al. Package 'vegan': community ecology package. R package version 2.6–4 edn; 2022.

92. Dray S, Dufour AB. The ade4 package: implementing the duality diagram for ecologists. J Stat Softw. 2007;22:1–20.

93. Wang C, Szpiech ZA, Degnan JH, Jakobsson M, Pemberton TJ, Hardy JA, et al. Comparing spatial maps of human population-genetic variation using Procrustes analysis. Stat Appl Genet Mol Biol. 2010;9(1):13.

94. Wang C, Zöllner S, Rosenberg NA. A quantitative comparison of the similarity between genes and geography in worldwide human populations. PLoS Genet. 2012;8(8):e1002886.

95. Beerli P, Felsenstein J. Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. Proc Natl Acad Sci USA. 2001;98(8):4563–8.

96. Beerli P. Migrate documentation. Version 3.2.1. Florida: Florida State University; 2012.

97. Maier P, Vandergast AG, Ostoja SM, Aguilar A, Bohonak AJ. Pleistocene glacial cycles drove lineage diversification and fusion in the Yosemite toad (*Anaxyrus canorus*). Evolution. 2019;73(12):2476–96.

98. Beerli P. Effect of unsampled populations on the estimation of population sizes and migration rates between sampled populations. Mol Ecol. 2004;13(4):827–36.

99. Beerli P, Mashayekhi S, Sadeghi M, Khodaei M, Shaw K. Population genetic inference with MIGRATE. Curr Protoc Bioinformatics. 2019;68(1):e87.

100. Beerli P, Palczewski M. Unified framework to evaluate panmixia and migration direction among multiple sampling locations. Genetics. 2010;185(1):313–26.

101. Bertorelle G, Benazzo A, Mona S. ABC as a flexible framework to estimate demography over space and time: some cons, many pros. Mol Ecol. 2010;19(13):2609–25.

102. Cornuet JM, Pudlo P, Veyssier J, Dehne-Garcia A, Gautier M, Leblois R, et al. DIYABC v2 0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism DNA sequence and microsatellite data. Bioinformatics. 2014;30(8):1187–9.

103. Hudson RR. Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics. 2002;18(2):337–8.

104. Estoup A, Lombaert E, Marin JM, Guillemaud T, Pudlo P, Robert CP, et al. Estimation of demo-genetic model probabilities with Approximate Bayesian Computation using linear discriminant analysis on summary statistics. Mol Ecol Resour. 2012;12(5):846–55.

105. Chang CS. A checklist of North Korean vascular plants. Version 1.7. TB Lee herbarium. Checklist dataset. 2024. https://doi.org/10.15468/9t4tft. Accessed via GBIF.org 13 Feb 2024.

106. Aiello-Lammens ME, Boria RA, Radosavljevic A, Vilela B, Anderson RP. spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. Ecography. 2015;38(5):541–5.

107. Fick SE, Hijmans RJ. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int J Climatol. 2017;37(12):4302–15.

108. QGIS Development Team. QGIS geographic information system. Version 3.22.5. Open Source Geospatial Foundation; 2020. http://qgis.org. Accessed 9 Sept 2022.

109. Hijmans RJ, Van Etten J, Sumner M, Cheng J, Baston D, Bevan A, et al. Package 'raster'. R package version 3.6.26 edn; 2023.

110. Warren DL, Matzke NJ, Cardillo M, Baumgartner JB, Beaumont LJ, Turelli M, et al. ENMTools 10: an R package for comparative ecological biogeography. Ecography. 2021;44(4):504–11.

111. Naimi B, Hamm NA, Groen TA, Skidmore AK, Toxopeus AG. Where is positional uncertainty a problem for species distribution modelling? Ecography. 2014;37(2):191–203.

112. Phillips SJ, Anderson RP, Dudík M, Schapire RE, Blair ME. Opening the black box: an open-source release of Maxent. Ecography. 2017;40(7):887–93.

113. Muscarella R, Galante PJ, Soley-Guardia M, Boria RA, Kass JM, Uriarte M, et al. ENM eval: an R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. Methods in Ecol Evol. 2014;5(11):1198–205.

114. Merow C, Smith MJ, Silander JA Jr. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. Ecography. 2013;36(10):1058–69.

115. Karger DN, Conrad O, Böhner J, Kawohl T, Kreft H, Soria-Auza RW, Zimmermann NE, et al. Climatologies at high resolution for the Earth land surface areas. Sci Data. 2017;4:170122. https://doi.org/10.1038/sdata.2017.122.

116. Riahi K, Van Vuuren DP, Kriegler E, Edmonds J, O'neill BC, Fujimori S, et al. The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications an overview. Glob Environ Change. 2017;42:153–68.