博士論文

Phylogeographic and taxonomic study of the Japanese *Geranium*

（日本産フウロソウ属の系統地理学的及び分類学的研究）

倉 田 正 観

Table of Contents

Chapter 1 General introduction

The Earth experienced global climate change during the Quaternary, which began 2.4 million years ago, and climate change has been amplified in the last megaannum. Cyclical glacial and interglacial periods, occurring every hundred thousand years, affected all living organisms and shaped the current day distributions of plants and animals (Hewitt 1996). Areas wherein species survive following periods of climate change are defined as refugia.

When biogeographic inquiry into refugia began, discussions relied upon fossil evidence and pollen records (e.g., Iversen 1942, 1954). Tralau (1963) summarized fossil data and compared them with the present-day distribution of European species (Birks 2008). The results indicated a downward and southward migration of arctic and alpine plants during glacial periods, and a reverse migration to their present-day ranges during warm interglacial periods (Birks 2008). Since these early inquiries, genetic markers (e.g., DNA sequencing, Amplified Fragment Length Polymorphism, Simple Sequence Repeat marker and Single Nucleotide Polymorphism [SNP] marker) have been applied; a recent survey of DNA phylogeographic data has supported and extended earlier work (Hewitt 1996).

During cold glacial periods in the temperate regions of the Northern Hemisphere the geographic ranges of most species were restricted to one or several southern refugia. However, during subsequent warming periods these ranges expanded, typically northward, into available space according to a given species' dispersal abilities and ecological requirements (Taberlet & Cheddadi 2002). Plants that were adapted to cold, dry climates (i.e., alpine and arctic plants) survived in ice-free areas, high elevation areas, and grassy low-elevation areas during glacial periods (Holderegger & Thiel-Egenter, 2009). Alpine species that had survived in grassy low-elevation areas during glacial periods subsequently migrated to grassy high-elevation areas in temperate regions during interglacial periods (e.g., Zhang et al. 2001; Ikeda & Setoguchi

2007). During the last glacial period, it is suggested that several peninsula, islands, and continents were connected due to low seawater levels. Therefore, the geographic distributions of many plant species expanded via land bridges (e.g., Milne 2006; Qi et al. 2014). For example, DeChaine (2008) proposed that population divergence among many tundra plant species was promoted by the Bering Land Bridge.

Broadly, the Japanese flora is composed of five zones; creeping pine and *Betula*, evergreen coniferous forest, mixed needleleaf and broadleaf forest, deciduous broad-leaved forest, and evergreen broad-leaved forest (Tabata 2000). In Japan, alpine plant species are distributed in high-elevation mountain areas and grassland plants are distributed in semi-natural grasslands. All flora in Japan are presumed to have immigrated to or emigrated from the Japanese archipelago at some time during glacial and interglacial periods. Presumably, these immigrations and emigrations occurred at different times, with various species surviving in different regions. For example, the temperate forest survived glacial periods in southern Japan and in the Chugoku region, and the distribution of this forest expanded during the late glacial period (Aoki et al. 2018). Japanese grassland flora is similar to that of northeastern China and Far East Russia, and it is suggested that grassland species emigrated from these regions to the Japanese archipelago during the glacial periods (Fig. 1-1a; Murata 1988). Japanese alpine flora is similar to that of eastern Siberia and western North America (i.e., Beringia), and it is believed that alpine species in Japan originated from Beringia (Fig. 1-1b; Koidzumi 1919; Hultén 1937; Tatewaki 1974; Shimizu 1982, 1983). During the glacial periods, land bridges connected the Islands of Kyushu and Honshu to the Korean Peninsula, and Hokkaido Island to Kuril Island and Sakhalin. Therefore, many plant species are assumed to have migrated between Japan and these surrounding regions. The directionality of these migrations, however, remains unclear.

High-elevation mountain areas in central Honshu Island are known as interglacial

refugia for alpine plant species. Recent molecular studies have indicated that populations in refugial regions are not genetically homogeneous but, instead, are often fragmented into several refugia with different genetic structures. This phenomenon is defined as "refugia within refugia", and several wild organisms are known to have retained such a phylogeographic structure (Gomez & Lunt 2007). Complex genetic structures have been documented in numerous alpine plant populations in the Alps; glacial refugia in this region are subdivided into several smaller refugia (e.g., Schönswetter et al. 2004, 2005). Although the demographic histories of certain Japanese alpine plant species have recently been assessed, understanding of the processes of population formation in southern refugia remains limited. Additionally, Japanese alpine species were believed to have immigrated to the Japanese archipelago from more northern areas (e.g., Sakhalin) during glacial periods (Koidzumi 1919; Hultén 1937), but recent work has suggested that some species emigrated from East Asia (including the Japanese archipelago) to Beringia (Ikeda et al. 2018). Further data are needed to understand the migratory history of Japanese alpine plant species. Currently, there is very little understanding of the population dynamics of Japanese grassland plant species, and these plants are under imminent threat of extinction due to reduction in grassland area. Thus, I aimed to elucidate the demographic history, genetic structure, and genetic diversity of Japanese grassland and alpine plants.

Thirteen *Geranium* species (Geraniaceae) are distributed in Japan. They are typically found in semi-natural grasslands, grassy mountain areas, the alpine zone, and along roadsides (Kadota 2016). Although two species are endemic, the remainder are distributed in areas outside of Japan (Xu & Aedo 2008; Kadota 2016). *Geranium* species supposedly originated around the Mediterranean and immigrated to East Asia five million years ago (Fiz et al. 2008). Therefore, *Geranium* species were likely to have immigrated to the Japanese archipelago more recently than five million years ago (Fiz et al. 2008; Marcussen & Meseguer 2017), and were therefore

affected by climate change during the Quaternary. This makes *Geranium* species a suitable model for elucidating the demographic history and genetic structure of Japanese grassland and alpine plants. *G. soboliferum* Kom is a grassland species that is found in northeastern China, Far East Russia and Japan (Northern Kyushu and central Honshu). This species has a fragmented distribution in Japan and is believed to have immigrated from northeastern China and/or Far East Russia to Japan during glacial periods (Fig. 1-2; Murata 1988). Therefore, *G. soboliferum* is a suitable species for assessing the geographic distribution transitions of species that migrated via the land bridge between the Japanese archipelago and the Korean Peninsula.

G. yesoense Franch. et Sav. is endemic to Japan. This species is found in several distinct habitats: alpine areas in central Honshu and grassy places near the coast in northern Japan (Tohoku and Hokkaido district). Given its distribution, it is suggested that the *G. yesoense* populations in central Honshu represent interglacial refugia (Fig. 1-2). By comparing genetic structures among refugial populations of *G. yesoense*, I can assess and discuss the presence of refugia within refugia in the Japanese archipelago. The alpine species *G. erianthum* DC. is distributed from northern Japan to western North America (Fig. 1-2; Shimizu 1982, 1983; Akiyama 2001). Given the breadth of its distribution, it has been suggested that *G. erianthum* either immigrated into East Asia, or emigrated from East Asia to Beringia. However, previous studies have not been able to assess the phylogenetic relationships of Japanese *Geranium* due to inadequate variety-level taxon sampling and low analysis resolution (Marcussen & Meseguer 2017; Wakasugi et al. 2017).

I employed polymorphic genotyping markers, genome-wide SNP genotyping, and next generation sequencing of chloroplast genomes to assess genetic structure, genetic differentiation among populations, phylogenetic relationships, and the phylogeographic history of each *Geranium* species. I specifically addressed the following: 1) The timing of divergence of

4

grassland plants in the Japanese archipelago and their genetic diversity; 2) refugia fragmentation of alpine plants in central Honshu; 3) the transitional distribution pattern of alpine plants during the last glacial period; and 4) the phylogenetic relationships among 13 Japanese *Geranium* species.
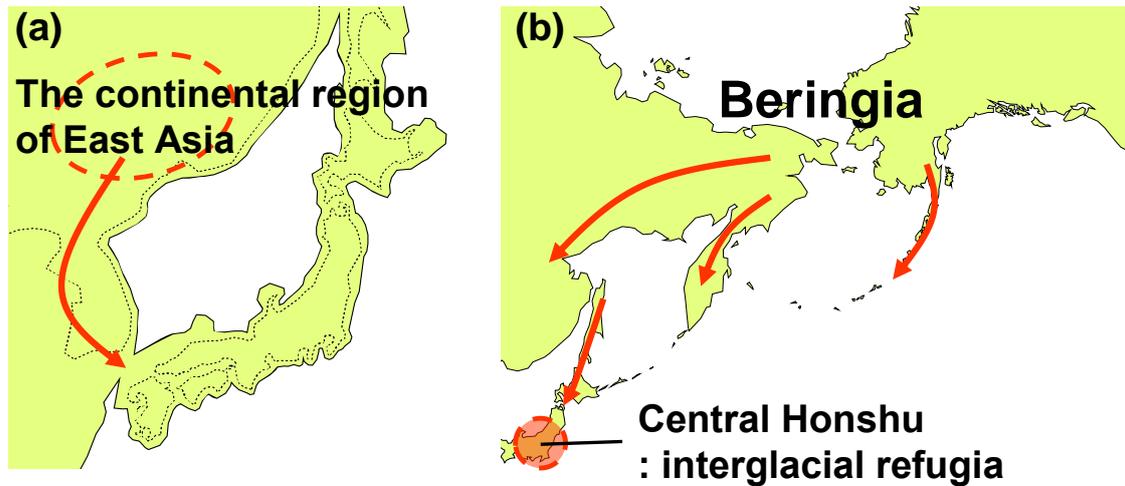
Fig. 1-1. Distributional transitions of Japanese flora. (a) Japanese grassland species emigrated from the continental region of East Asia; (b) Japanese alpine species originated in Beringia.
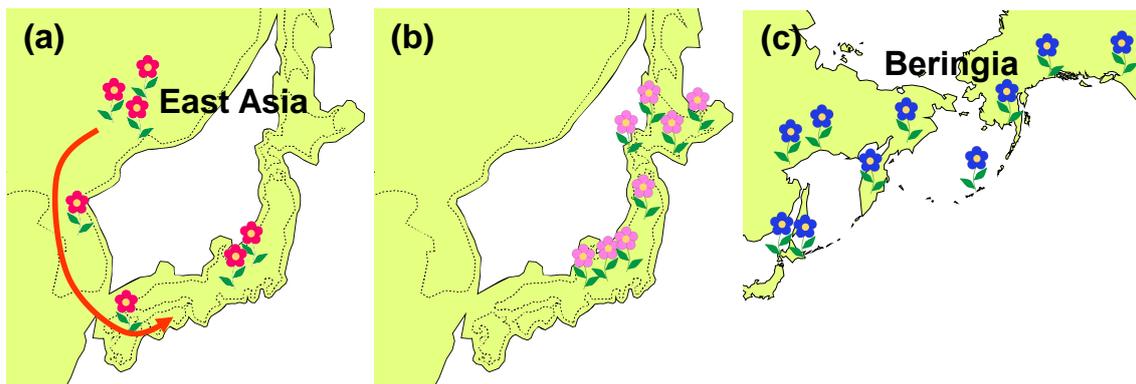
Fig. 1-2. Distributional transitions of Japanese *Geranium*. (a) It is likely that the grassland species *G. soboliferum* migrated from East Asia to Japan during glacial periods; (b) the alpine species of *G. yesoense* survived in high altitude refugia in central Honshu; (c) the alpine species *G. erianthum* originated in Beringia and immigrated to the Japanese archipelago during glacial periods.

Chapter 2 Phylogeographic and conservation genetic study of *G. soboliferum* Kom.

2-1. Introduction

Genetic diversity has long been recognized as fundamental to biodiversity, since it is necessary for maintaining healthy populations and drives biological adaptation. Species whose individuals and populations are relatively few in number tends to lose their genetic diversity (Frankham et al. 2002). When the population size declines, genetic diversity decreases at a higher rate due to increased inbreeding, demographic instability, and genetic drift (Frankham et al. 2002). This feedback between reduction of population size, loss of genetic diversity, and progression of inbreeding is known as the extinction vortex (Frankham et al. 2002). Once trapped in the vortex, species struggle to recover their former levels of genetic diversity. Smaller populations may incur increased short-term risk of extinction as a result of reduced reproduction, while long-term risk may also be increased since they are less capable of responding to environmental changes (Kery et al. 2000). Therefore, it is imperative that the genetic characteristics (e.g., effective population size, historical demography, presence of inbreeding) of endangered species be evaluated to facilitate the development of effective management strategies.

Although the warm and wet climate conditions of the Holocene were conducive to the establishment of temperate forests in the lowlands of Japan, it has been suggested that vast grasslands once predominated in the era (Ogura 2006; Ushimaru et al. 2018). A major grassland type is semi-natural grassland, which is partly maintained by human activities, such as burning, mowing, and grazing. Recent shifts in land use, however, have caused rapid shrinkage of semi-natural grasslands, and Japan's grassland area now constitutes only around 1% of its territory (Ministry of Agriculture, Forestry and Fisheries 2005; Ogura 2006). However, semi-natural grasslands with rich grassland flora can be still found in the Aso-Kuju region of

7

Kyushu Island, and in the central region of Honshu Island (Fig. 2-1). In the former region, volcanic activity and prolonged human managements have shaped a vast semi-natural grassland terrain (area: 220 km^2) in Aso in Kumamoto Prefecture, Kyushu Island (Fig. 2-1). The grassland spans over 22,000 ha, and includes various landscape and vegetation types, including scattered wetlands, drier Poaceae-dominated grassland, and partially shaded grassland mixed with shrubby species. More than 600 wild grassland plants grow in this area, of which 67 species are on the Red List (Ministry of the Environment 2012, 2015, 2017; Takahashi 2013); for example, *Viola orientalis*, *Echinops setifer*, *Silene sieboldii* and *Campanula glomerata* are believed to be relict plants that migrated from the Eurasian continent during glacial period(s) (Okubo 2002; Takahashi 2009, 2013). Since these herbaceous plants inhabit grassland ecosystems, they disappeared from most areas of the Japanese islands, which were densely forested following the last glacial period (Okubo 2002; Takahashi 2009, 2013). In this way, the semi-natural grassland in the Aso region, which has long been maintained by human activities, is particularly valuable in that it provides habitats for these endangered grassland plant species. During the last century, however, grassland in the Aso region has rapidly declined, decreasing in area by around half since 1905 (Matsunaga 2005). Accordingly, the number of animals and plants living in this grassland is also decreasing, as a result of the reduction in grassland area and deterioration in the quality of microhabitats, which in turn increases their risk of extinction (Sei et al. 2015). Thus far, population genetic analyses of several grassland plants have been conducted (e.g., Kaneko et al. 2009, Yamasaki et al. 2013; Yokogawa et al. 2013, Fujii et al. 2016). While these studies focused on the population structures and bottlenecks of species inhabiting typical *Miscanthus*-dominated grasslands, the important wetland habitat has never been evaluated in terms of plant population genetics.

Recently, as a result of the impacts of land development and climate change, wetlands

have declined in the Aso-Kuju region, bringing many wetland plants to the brink of extinction (Ministry of the Environment 2015). *Geranium soboliferum* Kom. var. *kiusianum* (Koidz.) H. Hara is a perennial herb found in wetlands of high-altitude grassland areas in the Aso-Kuju region. Currently, several hundred individuals grow in the wild, and the probability of extinction is estimated at 25% over the coming 100 years (Ministry of the Environment 2015). Therefore, this plant has been designated "Vulnerable (VU)" in the Red Data Book (Ministry of the Environment 2015). It is a variety of a widespread species of *Geranium soboliferum* Kom., which is widely distributed in Northeastern China, Far East Russia and Japan (central Honshu) (Xu and Aedo 2008). The *G. soboliferum* populations in central Honshu are distinguished as *Geranium soboliferum* Kom. var. *hakusanense* (Matsum.) Kitag (Iwatsuki et al. 2003; Ohashi et al. 2016), and are also designated as "Near Threatened (NT)" (Ministry of the Environment 2015). Compared with *G. soboliferum* var. *kiusianum*, *G. soboliferum* var. *hakusanense* has broader habitat preferences (wet and grassy areas on mountainsides) and is more widely distributed (Fig. 2-1). As in the case of Aso in Kyushu, some volcanos in central Honshu were active during the Holocene in central Honshu, and relatively large grassland areas remain there as refugia for various grassland animals and plants (Suka 2008). Therefore, the grassland of central Honshu is considered an important region for biodiversity in Japan's semi-natural grasslands. As mentioned above, however, the semi-natural grassland area is in decline, and the habitat of *G. soboliferum* var. *hakusanense* is therefore under threat; thus, a conservation genetic study of *G. soboliferum* var. *hakusanense* in addition to var. *kiusianum* is necessary. Reveling the evolutionarily significant units and phylogeographic histories of these varieties will provide baseline information for the management of the remaining populations, in the face of rapid changes in grassland ecosystems. Phytogeographic analysis has suggested that the current distribution of *G. soboliferum* in Japan is likely to have been shaped by migration(s) from

9

Northeastern China and/or Far East Russia during glacial periods, when the Japanese islands were almost connected to the Asian continent (Murata 1988). Later, the Aso-Kuju populations were isolated from other *G. soboliferum* (var. *hakusanense* and var. *soboliferum*) populations, and accumulated morphological differences, with *G. soboliferum* var. *kiusianum* being pubescent on the entire underside surface of the leaf while *G. soboliferum* var. *hakusanense* has appressed trichomes only on the vein of the leaf's underside (Kadota et al. 2016). However, no phylogeographic/population genetic analysis has been conducted hitherto to explicitly investigate the establishment history of these wetland plants in the Japanese archipelago.

It is therefore critical to evaluate the genetic diversity status of the remaining populations, and to direct conservational efforts toward these varieties, where the appropriate conservational strategies vary among threatened species and populations (Frankham et al. 2002). Population genetic analysis of the wetland *Geranium* species will offer insight into the roles of wetlands in the Aso-Kuju region as reservoirs of populations, and the migration source/sink relationships among spatially isolated habitats. Furthermore, as an isolated regional group in the Aso-Kuju region, population genetic demographic analysis should be conducted with the aim of understanding the historical relationship between other regional populations and the time frame within which the lineage divergence of *G. soboliferum* occurred. In this study, I specifically aim to evaluate the following: (1) the genetic diversity and genetic structure of residual *G. soboliferum* var. *kiusianum* and other *G. soboliferum* populations, (2) the population dynamics of *G. soboliferum* (including *G. soboliferum* var. *kiusianum*, via genetic demographic analysis), and (3) the peculiarity of *G. soboliferum* var. *kiusianum* as a wetland plant in the Aso-Kuju region, which exhibits differential habitat preferences to those of other herbaceous plants investigated previously.

2-2. Materials and methods

Morphological differences among three varieties

The three varieties are morphologically distinguishable from one another. *G. soboliferum* var. *kiusianum* has a lower leaf surface with appressed trichomes throughout; *G. soboliferum* var. *hakusanense* has a lower leaf surface with appressed hairs on the veins; *G. soboliferum* var. *soboliferum* has smaller flowers, at less than 2 cm across (Akiyama et al. 2001; Kadota et al. 2016). These entities are geographically isolated on Kyushu Island, Honshu Island, and the Asian continent. Therefore, in this study, I adopted the regional population grouping used in previous taxonomic treatments, which was almost in agreement with the results of my genetic analysis, with the exception that one of the *G. soboliferum* var. *soboliferum* varieties had undergone less genetic divergence (see results).

Plant sampling and microsatellite analysis

From September 2013 to September 2014, I collected mature leaves of *G. soboliferum* var. *kiusianum* and *G. soboliferum hakusanense* from 13 populations (6 populations from *G. soboliferum* var. *kiusianum* and 7 populations from *G. soboliferum* var. *hakusanense*). Additionally, 9 samples of *G. soboliferum* var. *soboliferum* were collected from China (Fig. 2-1). Sampling was performed randomly throughout each population. A total of 508 individuals were collected, and were immediately dried and preserved at room temperature using silica gel.

Genomic DNA was extracted using a modified cetyl trimethylammonium bromide (CTAB) method (Milligan 1992). Each individual's genotypes were characterized at 10 polymorphic markers, which were developed by Kurata et al. (2017): VKGP_120298, VKGP_15780, VKGP_4763, VKGP_29306, VKGP_76222, VKGP_20310, VKGP_22221, VKGP_28824, VKGP_105296 and VKGP_16051. Polymerase chain reaction (PCR) was

11

conducted following a standard protocol using the Qiagen Multiplex PCR Kit (Qiagen, Hilden, Germany), in a final volume of 10 µL, which contained approximately 5 ng of DNA, 5 µL of 2× Multiplex PCR Master Mix, and 0.01 µmol/L of forward primer, 0.2 µmol/L reverse primer, and 0.1 µmol/L of M13 primer (fluorescently labeled using Beckman Dye; Beckman Coulter, Brea, CA, USA). The PCR thermal profile involved denaturation at 95 °C for 3 min, followed by 35 cycles of 95 °C for 30 s, 60 °C for 3 min, 68 °C for 1 min, and a final 20 min extension step at 68 °C. PCR products were loaded onto an auto sequencer (GenomeLab^{TM} GeXP; Beckman Coulter) to assess fragment lengths using Fragment Analysis Software (ver. 8.0; Beckman Coulter).

Population genetic and structural analysis

To characterize each population, the following five summary statistics were calculated using FSTAT2.9.3 (Goudet 1995) and GenAlEx v.6.501 (Peakall and Smouse 2006): number of alleles (A), allelic richness (Ar), expected heterozygosity (H_e), observed heterozygosity (H_o), and fixation index (F_{IS}). Moreover, statistically significant differences in genetic diversity (i.e., in Ar and H_e) among Japanese population groups were tested across 2,000 permutations using FSTAT2.9.3; in these analyses, I aggregated the regional populations into two varieties (*G. soboliferum* var. *kiusianum* and *G. soboliferum* var. *hakusanense*), and compared their Ar and H_e values.

The F_{ST} index (Weir and Cockerham 1984) was calculated as a proxy of genetic differentiation among populations, and tested across 1,000 permutations using FSTAT2.9.3 to determine whether it differed significantly from zero. To ascertain the genetic structures of these populations, I used the model-based Bayesian clustering program STRUCTURE 2.3.4 (Pritchard et al. 2000); STRUCTURE analysis determines the number of clusters and assigns

12

individuals to K clusters. The population structure was simulated using the sequential number of genetic clusters for each taxon and for the overall population: $K = 1$–14 (*G. soboliferum* Kom.), $K = 1$–6 (*G. soboliferum* var. *kiusianum*) and $K = 1$–8 (*G. soboliferum* var. *hakusanense* and *G. soboliferum* var. *soboliferum*) under an admixture and allele frequencies-correlated model (*F* model) (Falush et al. 2003). Simulation runs consisted of 1,000,000 Markov chain Monte Carlo (MCMC) simulations, following a burn-in period of 100,000 iterations; however, for the analysis of *G. soboliferum* Kom., a longer MCMC run of 1,500,000 steps was adopted to ensure good parameter convergence. Twenty replicates were performed for each value of K. All replicates were processed without prior information of the sampling locations. To determine the optimal value of K, I adopted the ΔK method (Evanno et al. 2005) and the value of Ln P(D). I sought the maximum ΔK or the value of Ln P(D). If Ln P(D) increased according to the number of clusters, the value of K could be determined as that just before the value of Ln P(D) reached the plateau. While STRUCTURE assumes Hardy-Weinberg equilibrium (HWE) within inferred clusters, my high F_{IS} suggested *G. soboliferum* Kom. is a partially selfing plant, or is experiencing population inbreeding. InStruct is a Bayesian clustering program that infers selfing (Gao et al. 2007); therefore, I alternately used InStruct to detect genetic structures in the varieties. Each run consisted of 1,000,000 MCMC after a burn-in period of 500,000 iterations, and 20 replicates were performed for each value of K. All replicates were processed without prior information of sampling locations. I selected mode 2, which infers population structure and population selfing rates and ran the program taking selfing into consideration. Setting the value of K to the same as that for the STRUCTURE analysis, I adopted a deviance information criterion (DIC) to decide the optimal value of K, along with the value of LnP(D).

Furthermore, I performed principal coordinate analysis (PCoA) to elucidate interindividual genetic variance , and also conducted analysis of molecular variance (AMOVA)

to partition genetic variation between regions (F_{RT}, differentiation between *G. soboliferum* var. *kiusianum* and *G. soboliferum* var. *hakusanense*), among populations within regions (F_{SR}), and among all populations (F_{ST}) using GenAlEx v.6.501 (Peakall and Smouse 2006). Additionally, I evaluated the isolation by distance (IBD) of *G. soboliferum* var. *kiusianum* and *G. soboliferum* var. *hakusanense* to assess the relationship between geographic and genetic distance. The geographic distance of each population was calculated from their latitude/longitude, and F_{ST} / (1-F_{ST}) was adopted as their genetic distance (Rousset 1997); the significance of correlations was then verified by the Mantel test using GenAlEx v.6.501. The Tochigi population was eliminated from this analysis, since this population is located in a wild flower garden and therefore may have been transplanted from another population.

To estimate the recent inter-population gene flow within each region, BayesAss3.0 (Wilson and Rannala 2003) was applied to the population groups of the Aso-Kuju region and central Japan. Each run consisted of 10,000,000 MCMC following a burn-in period of 1,000,000 iterations, and every 1,000 iterations were sampled from the remaining 9 million iterations.

Demographic analysis

Approximate Bayesian computation (ABC) is a powerful and flexible approach to estimating demographic and historical parameters and quantitatively comparing alternative scenarios (Bertorelle et al. 2010). To determine the most probable demographic scenario among alternatives with regard to *G. soboliferum*, I adopted the ABC method implemented in DIYABC v2.0 (Cornuet et al. 2014), which has been widely used in genetic demographic studies (e.g., Bodare et al. 2013, Semerikov et al. 2013, Sakaguchi et al. 2013).

G. soboliferum is distributed across three isolated regions: the Eurasian continent (*G.*

14

soboliferum var. *soboliferum*), central Honshu (*G. soboliferum* var. *hakusanense*), and northern Kyushu (*G. soboliferum* var. *kiusianum*). As mentioned above, *G. soboliferum* var. *kiusianum* and *G. soboliferum* var. *hakusanense* are believed to be relicts that migrated southward from the Eurasian continent to the Japanese archipelago during the last glacial period (Murata 1988). Therefore, I considered the possibility that the distributional transition of *G. soboliferum* was influenced by the expansion and/or reduction of grassland areas, processes associated with past climate changes. Although the details of their distributional transition are unknown, due to the scarcity of data relating to relict herbaceous plants, several hypotheses may be proposed with regard to their distributional changes: (1) The current regional populations fragmented simultaneously, reflecting global climate change; (2) Populations of the East Asian continent expanded their distribution to the Japanese archipelago, and thereafter the ancestral population of *G. soboliferum* var. *kiusianum* and *G. soboliferum* var. *hakusanense* were divided across two regions, resulting in a more recent population split within the archipelago; (3) Alternatively, it may be assumed that the ancestral populations of *G. soboliferum* var. *hakusanense* first diverged from continental populations and, subsequently, the ancestral populations of *G. soboliferum* var. *kiusianum* invaded the Japanese archipelago via a land bridge from the Korean Peninsula; (4) The ancestral populations of *G. soboliferum* var. *hakusanense* may have arrived in the central region of Honshu Island from the continent via the northern land bridge, located between Sakhalin and Hokkaido Island. This scenario predicts that var. *hakusanense* and var. *soboliferum* would have the closest genetic relationship, while the southerly isolated var. *kiusianum* was not involved in the formation of var. *hakusanense*; and (5) The intermediate geographic position of var. *kiusianum* may indicate that the populations were generated by a past admixture of var. *hakusanense* and var. *soboliferum*, which expanded the distributions in response to cooler and drier climates during glacial periods.

15

To model and compare the alternative population demographic models, I first defined three populations: *soboliferum* (Jilin, Eurasian continent populations, *G. soboliferum* var. *soboliferum*), *hakusanense* (Nobeyama, Kirigamine, Saku, Nasu, Kiyosato, Fujinomiya, central Honshu populations, *G. soboliferum* var. *hakusanense*) and *kiusianum* (Tadewara_A and B, Ide Farm_A and B, Mt. Daikanbou_A and B, northern Kyushu, *G. soboliferum* var. *kiusianum*). Note that the Tochigi population in central Honshu was excluded from this analysis. Five candidate demographic scenarios were then visualized (Fig. 2-2). The prior distributions of each parameter are noted in the Supplementary Materials. Additionally, I estimated effective population size for each *G. soboliferum* var. *kiusianum* population by applying a constant population-size model and size-change model using DIYABC, to compare the census and effective population size of var. *kiusianum*. In these simulations, I constructed two candidate demographic scenarios for each population (Tadewara, Ide Farm and Mt. Daikanbou): the scenarios operated on the premise that each population had or had not experienced declining population size. Each scenario was visualized in Fig. S2-2, and the prior distribution of each parameter is summarized in the Supplementary Materials.

Next, I established a series of summary statistics: the mean *A* across loci, the mean gene diversity across loci (Nei 1987), and the mean allele size variance across loci were selected as single-sample statistics. Additionally, the F_{ST} between both samples, the mean index of classification and the $(\delta\mu)^2$ distance between both samples were selected as two sample statistics. One million simulations were performed for each scenario, and the most appropriate scenario was determined by comparing posterior probabilities. I also assessed the effects of the number of simulated data points closest to those observed, which were used to evaluate posterior probabilities using the varying portion of the closest data point (i.e. every 0.1% in 0.1–1.0% of the closest simulations). Moreover, the divergence time in relation to absolute time was

16

calculated by multiplying the estimated time parameter in the number of generations by 3–5 years, which is appropriate as the generation time for this herbaceous species (The Alpine Garden Society of Tokyo 2004). For the analysis based on the single-population model, I set a series of summary statistics; the mean *A* across loci, the mean gene diversity across loci (Nei 1987), and the mean allele size variance across loci were selected as single-sample statistics, and 1,000,000 simulations were performed for each scenario.

2-3. Results

Genetic diversity

Using 10 polymorphic markers, I detected 147 alleles among 508 individuals of *G. soboliferum* complex (*G. soboliferum* Kom.). The *A* per locus ranged from 5 (VKGP_29306 and VKGP_22221) to 31 (VKGP_15780). The summary statistics for each population are presented in Table 2-1. For the six populations of *G. soboliferum* var. *kiusianum* in the Aso-Kuju region in northern Kyushu, I detected 96 alleles among 266 individuals; the *Ar* of each population ranged from 2.499 (Ide farm_B) to 4.863 (Tadewara_A), while H_e and H_o varied from 0.298 (Ide farm_B) to 0.547 (Tadewara_A) and 0.216 (Ide farm_B) to 0.552 (Tadewara_A), respectively. The F_{IS} ranged from -0.009 (Mt. Daikanbou_B) to 0.288 (Ide farm_B). Regarding the two populations of Ide Farm_A and B, significant deviations from zero were detected. For the seven populations of *G. soboliferum* var. *hakusanense* in central Honshu, I detected 118 alleles among 233 individuals. The *A* per locus ranged from 2 (VKGP_29306) to 25 (VKGP_15780). The *Ar* of each population ranged from 2.680 (Fujinomiya) to 5.068 (Nobeyama). The H_e and H_o of each population ranged from 0.338 (Tochigi) to 0.659 (Nobeyama) and 0.218 (Tochigi) to 0.580 (Nobeyama), respectively. The F_{IS} ranged from 0.130 (Nasu) to 0.386 (Tochigi), and significant deviations from zero were detected in all populations. Regarding the single population of *G. soboliferum* var. *soboliferum* in China, 45 alleles were detected among nine individuals; *Ar* was 4.418, while H_e and H_o were 0.581 and 0.485, respectively. The F_{IS} was 0.224, and no significant deviations from zero were detected at any loci. Per-locus information is presented in Supplementary Table S2-1; see also Kurata et al. (2017) for further details of the genetic markers.

Genetic diversity of the grouped *G. soboliferum* var. *kiusianum* and *G. soboliferum* var. *hakusanense* populations was as follows: *Ar* = 2.349, H_e = 0.398; and *Ar* = 3.224, H_e = 0.561,

respectively. The results of statistical analysis indicated that the genetic diversity of *G. soboliferum* var. *kiusianum* was slightly, but significantly, lower than that of *G. soboliferum* var. *hakusanense* ($P < 0.01$). The differences in genetic diversity between the populations are summarized in Table S2-2.

Genetic differentiation and structure analysis

I performed PCoA to visualize the genetic relationships between individuals. Although *G. soboliferum* var. *kiusianum* and *G. soboliferum* var. *hakusanense* were genetically differentiated along the PCoA axis 1 (10.22% variance explained), *G. soboliferum* var. *soboliferum* overlapped with both *G. soboliferum* var. *kiusianum* and *G. soboliferum* var. *hakusanense*. The *G. soboliferum* var. *kiusianum* populations had a wider range on both axes, which was comparable with that of the more widely distributed *G. soboliferum* var. *hakusanense* and *G. soboliferum* var. *soboliferum* (Fig. 2-5). As with the results of the STRUCTURE analysis, Ide Farm and other two populations were divided into distinct groups on the second axis. The population-based PCoA results are presented in the Supplementary Materials (Fig. S2-3). Individuals were mostly clustered together with other individuals of the same population; however, the Ide Farm population was genetically differentiated from other *G. soboliferum* var. *kiusianum* populations and partially overlapped with var. *hakusanense* populations.

Overall, the STRUCTURE analysis of the samples revealed high levels of genetic heterogeneity in the Japanese and Chinese populations, evinced by multiple peaks in the ΔK statistic and clustering patterns (Fig. 2-3a, 3d). At $K = 2$, the individuals of both Japanese varieties were clustered separately, and the Chinese population (Jilin) was assigned to the light blue cluster together with the central Honshu populations (i.e., *G. soboliferum* var. *hakusanense*). Following this, northern Kyushu populations (i.e., *G. soboliferum* var. *kiusianum*) were

19

separated into two clusters at $K = 3$, and the central Honshu and Chinese populations were split off at $K = 6$, but the Nasu population was assigned to the red cluster along with the Chinese population. Additionally, the central Honshu populations were separated into three clusters. InStruct clustering of all samples revealed higher levels of genetic heterogeneity than did the STRUCTURE analysis. The most appropriate K value was 7 or 11. At $K = 11$, almost all individuals from different populations were clustered separately, except for those pairs that were geographically close (i.e., Mt. Daikanbou_A + B, Tochigi + Fujinomiya, and Nobeyama + Kiyosato pairs) (Fig. 2-4a, 4d). To explore the genetic differentiation patterns further, I analyzed the microsatellite genotypes for each variety, recognized as the uppermost genetic cluster unit in the above analysis.

(1) *G. soboliferum* var. *kiusianum* in northern Kyushu. In the STRUCURE analysis, the ΔK method indicated that the most appropriate K value was 3, at which a sharp single peak was detected and the log-likelihood value plateaued (Fig. 2-3b). The populations of Tadewara_A, B and Ide Farm_A, B and Mt. Daikanbou_A, B were assigned to different clusters, corresponding to their geographic locations (Tadewara, Ide Farm and Mt. Daikanbou) (Fig. 2-3e). The InStruct result of *G. soboliferum* var. *kiusianum* was almost equivalent to that achieved with the STRUCTURE analysis, but $K = 6$ was deemed the most appropriate K value based on the DIC result. At $K = 6$, with the exception of Mt. Daikanbou, the populations were assigned to different clusters (Fig. 2-4b, 4e). Although the *G. soboliferum* var. *kiusianum* populations were adjacent to one another (i.e. up to 20 km), Bayesian clustering analyses indicated clear genetic differentiation among the populations. Pairwise F_{ST} values ranged from 0.046 to 0.499 (the maximum differentiation value was obtained between Ide Farm_B and Mt. Daikanbou_B) (Table S2-3a), and the mean pairwise F_{ST}, achieved using the jackknife method, was 0.368. Genetic differentiation (pairwise F_{ST} value) was significant among all populations (*P*

< 0.01).

(2) *G. soboliferum* populations in central Honshu and China. The ΔK method indicated that the most appropriate K value was two for STRUCTURE analysis; the populations of Nasu, Saku, Kirigamine, Nobeyama, Kiyosato and Jilin were assigned to the light blue cluster, and Tochigi and Fujinomiya were assigned to the dark green cluster (Fig. 2-3c, 3f). The result of InStruct clustering of *G. soboliferum* var. *hakusanense* and *G. soboliferum* var. *soboliferum* revealed higher levels of genetic heterogeneity than did the STRUCTURE analysis. The LnP(D) value showed that the most appropriate K value was 5, but the DIC result indicated that the most appropriate K value was 8 (Fig. 2-4c, 4f). At $K = 5$, almost all populations were clustered separately, but Tochigi and Fujinomiya, and Nobeyama and Kiyosato, had the same genetic structures. Additionally, at $K = 8$, Jilin was assigned to the orange cluster. Pairwise F_{ST} values ranged from 0.037 to 0.375 (Table S2-3b), and the mean pairwise F_{ST} value, achieved using the jackknife method, was 0.184. The deviations of genetic differentiation (pairwise F_{ST} value) from zero were significant among all populations ($P < 0.01$).

No significant correlations were observed between the geographic and genetic distances of both varieties (var. *kiusianum* vs. var. *hakusanense*): $r^2 = 0.9878$, $P = 0.151$ and $r^2 = 0.3825$, $P = 0.079$, respectively; (Fig. 2-6). Additionally, genetic divergences between and within the two varieties found in Japan were detected by AMOVA. A total genetic differentiation of 65% was found among all populations, and a 25% differentiation was found between populations within regions. A 10% differentiation was found between regions (Table S2-4).

The estimated migration rate within each *G. soboliferum* var. *kiusianum* population ranged from 0.0046 (Ide farm_A -> Tadewara_B) to 0.2963 (Mt. Daikanbou_A -> Mt. Daikanbou_B); within *G. soboliferum* var. *hakusanense* populations, it ranged from 0.0066 (Fujinomiya -> Saku) to 0.2484 (Nobeyama -> Kiyosato) (Table 2-2a and 2b; these values

indicate the migration rates of recent generations). Overall, the migration rates among almost all populations were very low, with the exception of two closely adjacent paired populations (Mt. Daikanbou_A-B in Kyushu, and Kiyosato-Nobeyama in central Honshu).

Demographic analysis

Scenario 2, in which *G. soboliferum* var. *soboliferum* and the Japanese populations branched off first, followed by the northern Kyushu and central Honshu populations, had the highest posterior probability among the wide range of closest simulations (Fig. S2-1). Based on the selected scenario, the effective population size parameters were estimated as the median values (95% highest posterior densities) of the effective population sizes, being 8,290 (5,110–9,860), 7,770 (5,280–9,410) and 6,240 (3,420–8,590) for N1 (*soboliferum*), N2 (*hakusanense*) and N3 (*kiusianum*), respectively, (Table 2-3). The median values of the divergence times were 6,400 (3,130–9,460) and 1,710 (604–3,370) generations ago for t_a and t_b, respectively. If I assume a generation time of 3–5 years, the divergence times were 9,390– 7,300 years ago for divergence time t_a and 1,812–16,850 years ago for divergence time t_b. Moreover, the effective population sizes of *G. soboliferum* var. *kiusianum* from Tadewara, Ide Farm, and Mt. Daikanbou were 2,390 (1,080–5,720), 1,500 (682–4,820) and 814 (390–2,650), respectively (Table S2-5). The total effective population size of the three populations was 4,704 (i.e., 2,390 + 1,500 + 814).

2-4. Discussion

Genetic diversity and genetic structure of the three varieties

My assessments showed that the genetic diversity of *G. soboliferum* var. *kiusianum* ($Ar = 2.349$, $H_e = 0.398$) was slightly but significantly lower than that of *G. soboliferum* var. *hakusanense*, ($Ar = 3.224$, $H_e = 0.561$) ($P < 0.01$). Comparison of the distribution patterns of these two varieties revealed that *G. soboliferum* var. *kiusianum* is particularly associated with wetlands and has a more limited and discontinuous distribution (Akiyama et al. 2001). Therefore, *G. soboliferum* var. *kiusianum* may have experienced a more rapid decrease of genetic diversity owing to its limited suitable habitats. *G. soboliferum* var. *hakusanense*, however, has a wider distribution, with even more remnant populations in central Honshu, which may have allowed the taxon to retain more genetic diversity. Although the number of residual individuals of *G. soboliferum* var. *kiusianum* was estimated to be in the hundreds (Ministry of the Environment 2014), their effective population sizes were genetically estimated at several thousand in this study. This discrepancy between census and effective population size may stem from the underestimation of census size, as non-flowering individuals are less detectable in field surveys but can nonetheless contribute to effective population size. There is also some uncertainty with regard to the assumed mutation rate of microsatellite markers (Schlötterer et al. 1998; Vigouroux et al. 2002; Schlötterer 2004), which could have a direct impact on the estimation of effective population size. Furthermore, my complemental analysis of effective population size indicated that size values can vary to some extent, depending on the demographic models applied; based on a single-population model, the three populations in the Aso-Kuju region have a total size of 4,704 (median value of each population), which was more than 1,000 times smaller than the estimate from the branching model (6,240). Thus, it should be noted that further efforts are required to bridge the gap among estimates of population size, and

that the values obtained in this study would be useful only for the purpose of relative comparison between populations.

G. soboliferum var. *soboliferum* was shown to have a high level of genetic diversity (Table 2-1), although my analysis included only nine individuals. Grasslands in China cover an area of approximately 4 million km^2, equivalent to 40% of the total land area of Japan (Akiyama and Kawamura 2007; Zhou et al. 2017). Generally, genetic variation is positively correlated with population size (Frankham 1996). Comparison of the effective population sizes of *G. soboliferum* var. *soboliferum* and both Japanese varieties revealed no significant differences in this study, but this should be re-evaluated using larger sample sets from *G. soboliferum* var. *soboliferum* populations in future studies.

The PCoA and STRUCTURE results suggested that, while *G. soboliferum* var. *kiusianum* and *G. soboliferum* var. *hakusanense* were clustered into different genetic groups, *G. soboliferum* var. *soboliferum* was somewhat intermediate between both Japanese groups. This genetic position is apparently inconsistent with the inference from demographic analysis that var. *soboliferum* split first from the other two Japanese varieties, which may have caused var. *soboliferum* to exhibit a higher level of divergence from other regional populations. A possible explanation for this inconsistency is that the actual population size of continental var. *soboliferum* is large, has experienced stability, has not experienced significant genetic drifts, and therefore retains higher ancestral allelic frequency. However, the distributions of Japanese varieties are sporadic and small, and each population is subject to severe drifts, as exemplified in var. *kiusianum* populations (Fig. 2-6). Therefore, the PCoA, an algorithm that extracts synthetic variables reflecting the directions of total variance within the dataset, would have positioned the two Japanese varieties at either end of the first PCoA axis, with continental var. *soboliferum* occupying the intermediate position.

24

Genetic comparison with other grassland plant species in the Aso-Kuju region

Several grassland plants face extinction in the Aso-Kuju region, and population genetic analyses have been conducted for such plants using microsatellite markers to obtain genetic insights for the purpose of conservation (e.g. Kaneko et al. 2009; Yamasaki et al. 2013; Yokogawa et al. 2013; Fujii et al. 2016). With regard to inbreeding coefficients, which can be directly compared among species, the F_{IS} estimates varied greatly among the study systems, with F_{IS} = -0.052 to 0.150 in *Silene kiusiana* (Caryophyllaceae) (Yamasaki et al. 2013), F_{IS} = -0.168 to 0.185 in *Veronicastrum sibiricum* var. *zuccarinii* (Plantaginaceae) (Fujii et al. 2016), F_{IS} = -0.07 to 0.26 in *Polemonium kiushianum* (Polemoniaceae) (Yokogawa et al. 2013), and F_{IS} = -0.078 to 0.124 in *Echinops setifer* (Asteraceae) (Kaneko et al. 2009). Compared with these systems, my estimate of F_{IS} = -0.009 to 0.288 in *G. soboliferum* var. *kiusianum* appears to be at the upper level. While most of the other systems' populations yielded no evidence of inbreeding (F_{IS} values not significantly deviated from zero), I detected significantly high F_{IS} values over 0.25 for particular populations in *G. soboliferum* var. *kiusianum* (Table 2-1). Although the mating system of *G. soboliferum* var. *kiusianum* remains obscure, in view of the partial self-compatibility of the related species *G. thunbergii* (Kandori 2002), I suggest that the lower levels of genetic diversity and associated higher levels of F_{IS} in *G. soboliferum* var. *kiusianum* may be partly attributed to its reproductive ecology, which involves mating with related individuals and/or selfing. Additionally, since its habitat is exclusive to wetlands, the inhibition of effective gene flow may have further decreased the genetic diversity of *G. soboliferum* var. *kiusianum*.

Genetic differentiation and gene flow among populations

While *G. soboliferum* var. *kiusianum* is locally distributed in northern Kyushu, *G. soboliferum* var. *hakusanense* is widely distributed throughout central Honshu (Akiyama et al. 2001). Generally, with longer geographic distances, genetic differentiation among populations increases due to the intensified effects of genetic drift (Wright 1943). Contrary to my expectation that the wide-ranging *G. soboliferum* var. *hakusanense* would exhibit higher genetic differentiation than *G. soboliferum* var. *kiusianum*, my analysis yielded paradoxical results. The mean F_{ST} achieved using the jackknife method showed that the genetic differentiation of *G. soboliferum* var. *kiusianum* was significantly higher than that of *G. soboliferum* var. *hakusanense*. Furthermore, the STRUCTURE and InStruct results revealed that *G. soboliferum* var. *kiusianum* had remarkable genetic structuring even within its narrow distributional range, which was further supported by the exceptionally high F_{ST} values among local populations (Fig. 2-6). These apparently contradictory findings can be interpreted by considering the exclusively wetland habitat of *G. soboliferum* var. *kiusianum*. Wetlands typically occur in discrete patches in a matrix of upland habitats, such that most local wetland species populations are small and isolated and, thus, vulnerable to extinction (Møller and Rørdam 1985; Dodd 1990; Sjögren 1991; Gibbs 2000). Likewise, *G. soboliferum* var. *kiusianum* grows in scattered populations throughout disparate wetlands, and habitat discontinuity would likely have restricted the movement of seeds and pollens between isolated patches, as evinced by the low levels of interpopulational gene flow (Table 2-2). Therefore, population isolation would have triggered accelerated genetic differentiation in *G. soboliferum* var. *kiusianum* (Table S2-3a). I further argue that microhabitat quality may correlate with the degree and patterns of gene flow among individual plants within a population. For example, the common reed (*Phragmites australis*), a wetland plant that exhibits high competitiveness, is distributed with *G. soboliferum* var. *kiusianum* only in the Ide Farm population. The common reed's long stems cause each *G.*

soboliferum var. *kiusianum* individual to be physically isolated within the population, a phenomenon that I suggest may have led to gene exchange between adjacent plants, or even selfing (see the exceptionally high F_{IS} values estimated for both populations; Table 2-1).

Demographic history of G. soboliferum Kom.

The results of my demographic analysis revealed that Japanese populations branched off from the continental population of *G. soboliferum* var. *soboliferum* between 9,390 and 47,300 years ago. While I acknowledge that my demographic estimation, achieved with ABC simulation, may be biased due to the unequal sample sizes of population groups, and to clustering patterns that were apparently inconsistent with the assumed demographic units, the estimated divergence time includes the Last Glacial Maximum (LGM) period, and thereafter Japanese populations branched off into two populations at the beginning of the Holocene epoch. In the drier and cooler climate of the LGM, the East Asian continent's lowland areas (corresponding to the current Korean Peninsula and Northern China and Far East Russia) are believed to have harbored steppe and boreal coniferous woodland vegetation (e.g., Harrison et al. 2001), the most likely habitats for *Geranium soboliferum*. At that time, the Japanese islands were almost connected to the peninsula, and the herbaceous plants would have extended their distributions via a land bridge into the Kyushu and Honshu areas (Tabata 1997; Suka 2008). The once-continuous distributions would have gradually become isolated as the climate normalized after glacial periods, separating the Japanese varieties from *G. soboliferum* var. *soboliferum*. Simultaneously, the Holocene climate allowed the denser temperate forests to dominate the Japanese islands (Takahara 2009). Changes in vegetation would have likely led to geographic isolation among regional *Geranium* populations, dated to 1,812–16,850 years ago, based on my parameter estimate (Table 2-3).

The current distributions of grassland plants, including relict species, are now maintained by human management (e.g. grazing, mowing, and burning activities) (Murata 1988), with large populations in northern Kyushu (Aso-Kuju region) and central Honshu (e.g. Nagano Prefecture). It is conceivable that the semi-natural grasslands in northern Kyushu and central Honshu have long served as suitable habitats for *G. soboliferum* regional populations, which evolved independently from each other over more than five millennia (Table 2-3). Particularly in the case of *G. soboliferum* var. *kiusianum*, the genetic uniqueness, as well as the probable ecological specialization to wetland habitats, merit conservation. Since their habitats are confined to a few wetlands in semi-natural grasslands, without human management vegetation succession will cause the grassland area to decline in this region. Therefore, appropriate human activities (e.g., burning, mowing, and grazing) is critical for maintaining semi-natural grasslands in the Aso-Kuju region (Yasunaka et al. 2015), which will allow the residual *G. soboliferum* var. *kiusianum* populations to survive. Similarly, although *G. soboliferum* var. *hakusanense* is more widely distributed, significant F_{IS} values were detected for almost all *G. soboliferum* var. *hakusanense* populations. This finding suggests that inbreeding is widespread within this taxon's population, which may possibly be attributed to the ongoing shrinkage of its habitat patches. Further studies of local gene flow patterns and landscape genetics, with the aim of determining the key factors with regard to the genetic statistics of *G. soboliferum* populations, will be beneficial for the efficient conservation of Japanese *G. soboliferum* Kom. habitats.

Table 2-1. Genetic diversity of the 14 populations of *Geranium soboliferum* var. *kiusianum*, *G. soboliferum* var. *hakusanense* and *G. soboliferum* var. *soboliferum*.

Population name	Abbreviation	Variety	Prefecture, Country	Altitude (m)	N	A	Ar	Pr	H_e	H_o	F_{IS}
Tadewara_A	Td_A	GSvk	Oita, Japan	1,030	50	5.5	4.863	6	0.547	0.552	0.002
Tadewara_B	Td_B	GSvk	Oita, Japan	1,030	32	2.7	2.634	0	0.421	0.424	0.010
Ide farm_A	Id_A	GSvk	Kumamoto, Japan	800	66	4.5	3.580	6	0.353	0.276	0.226[*]
Ide farm_B	Id_B	GSvk	Kumamoto, Japan	800	40	2.7	2.499	0	0.298	0.216	0.288[*]
Mt. Daikanbou_A	Dk_A	GSvk	Kumamoto, Japan	880	47	3.8	3.458	2	0.381	0.381	0.014
Mt. Daikanbou_B	Dk_B	GSvk	Kumamoto, Japan	880	31	2.7	2.693	1	0.393	0.366	-0.009
Average				-		3.7 (96)[*1]	3.288 (2.349)[*2]	(23)[*3]	0.399 (0.398)[*2]	0.369	0.089
Nasu	Ns	GSvh	Tochigi, Japan	500	34	5.6	4.255	3	0.604	0.535	0.130[*]
Tochigi (cultivated)	Tc	GSvh	Tochigi, Japan	100	16	2.2	2.138	0	0.338	0.218	0.386[*]
Saku	Sk	GSvh	Nagano, Japan	1,230	36	5.4	4.019	1	0.612	0.539	0.134[*]
Kirigamine	Kr	GSvh	Nagano, Japan	1,630	37	5.9	3.952	8	0.564	0.489	0.150[*]
Nobeyama	Nb	GSvh	Nagano, Japan	1,340	35	7.6	5.068	1	0.659	0.580	0.137[*]
Kiyosato	Ky	GSvh	Yamanashi, Japan	1,620	32	6.7	4.702	6	0.643	0.569	0.132[*]
Fujinomiya	Fj	GSvh	Shizuoka, Japan	690	43	3.5	2.680	1	0.403	0.323	0.209[*]
Average				-		5.3 (118)[*1]	3.830 (3.224)[*2]	(45)[*3]	0.546 (0.561)[*2]	0.465	0.183
Jilin	Jl	GSvs	Jilin, China	530	9	4.5	4.418		0.581	0.485	0.224

Note : Gsvk = *Geranium soboliferum* var. *kiusianum*; Gsvh = *Geranium soboliferum* var. *hakusanense*; Gsvs = *G. soboliferum* var. *soboliferum*; N = number of samples; A = number of alleles per locus; Ar = allelic richness; Pr = private allele; H_e = expected heterozygosity; H_o = observed heterozygosity; F_{IS} = fixation index. Asterisk denotes significant departure from Hardy–Weinberg equilibrium ([*] $P < 0.05$). [*1], total number of alleles of each variety; [*2,] the value of grouped populations of each variety; [*3], the number of private alleles of each variety.

Table 2-2a. Migration rates among each *G. soboliferum* var. *kiusianum* population. Italics indicate the self-recruitment rates for each population.

In/From	Tadewara_A	Tadewara_B	Ide farm_A	Ide farm_B	Mt. Daikanbou_A	Mt. Daikanbou_B
Tadewara_A	*0.9665*	0.0090	0.0061	0.0060	0.0064	0.0059
Tadewara_B	0.0090	*0.9560*	0.0087	0.0088	0.0088	0.0088
Ide farm_A	0.0049	0.0046	*0.9517*	0.0171	0.0170	0.0047
Ide farm_B	0.0073	0.0072	0.0123	*0.9588*	0.0072	0.0073
Mt. Daikanbou_A	0.0060	0.0057	0.0103	0.0063	*0.6754*	0.2963
Mt. Daikanbou_B	0.0090	0.0089	0.0090	0.0090	0.0091	*0.9550*

Table 2-2b. Migration rates among each *G. soboliferum* var. *hakusanense* population. A population from Tochigi was excluded since it was a cultivated population. Italics indicate the self-recruitment rates for each population.

In/From	Nasu	Saku	Kirigamine	Nobeyama	Kiyosato	Fujinomiya
Nasu	*0.9499*	0.0087	0.0084	0.008	0.0084	0.0083
Saku	0.0118	*0.9346*	0.012	0.0146	0.0101	0.0091
Kirigamine	0.0076	0.0098	*0.9052*	0.0518	0.009	0.0088
Nobeyama	0.0131	0.0136	0.0078	*0.7018*	0.2484	0.0075
Kiyosato	0.0098	0.0269	0.0121	0.009	*0.922*	0.0117
Fujinomiya	0.0068	0.0066	0.0069	0.0067	0.0068	*0.9596*

Table 2-3. Estimates of demographic parameters in Scenario 2 in DIYABC analysis. N1, N2, N3: effective population sizes. t_a and t_b; time parameters in the number of generations. Mean, median, mode and 95% quantiles of posterior probability are reported.

Parameter	mean	**median**	mode	q050	q950
N1 (var. *soboliferum*)	8.02E+03	**8.29E+03**	9.75E+03	5.11E+03	9.86E+03
N2 (var. *hakusanense*)	7.63E+03	**7.77E+03**	7.71E+03	5.28E+03	9.41E+03
N3 (var. *kiusianum*)	6.12E+03	**6.24E+03**	6.35E+03	3.42E+03	8.59E+03
t_a	6.38E+03	**6.40E+03**	6.40E+03	3.13E+03	9.46E+03
t_b	1.83E+03	**1.71E+03**	1.52E+03	6.04E+02	3.37E+03
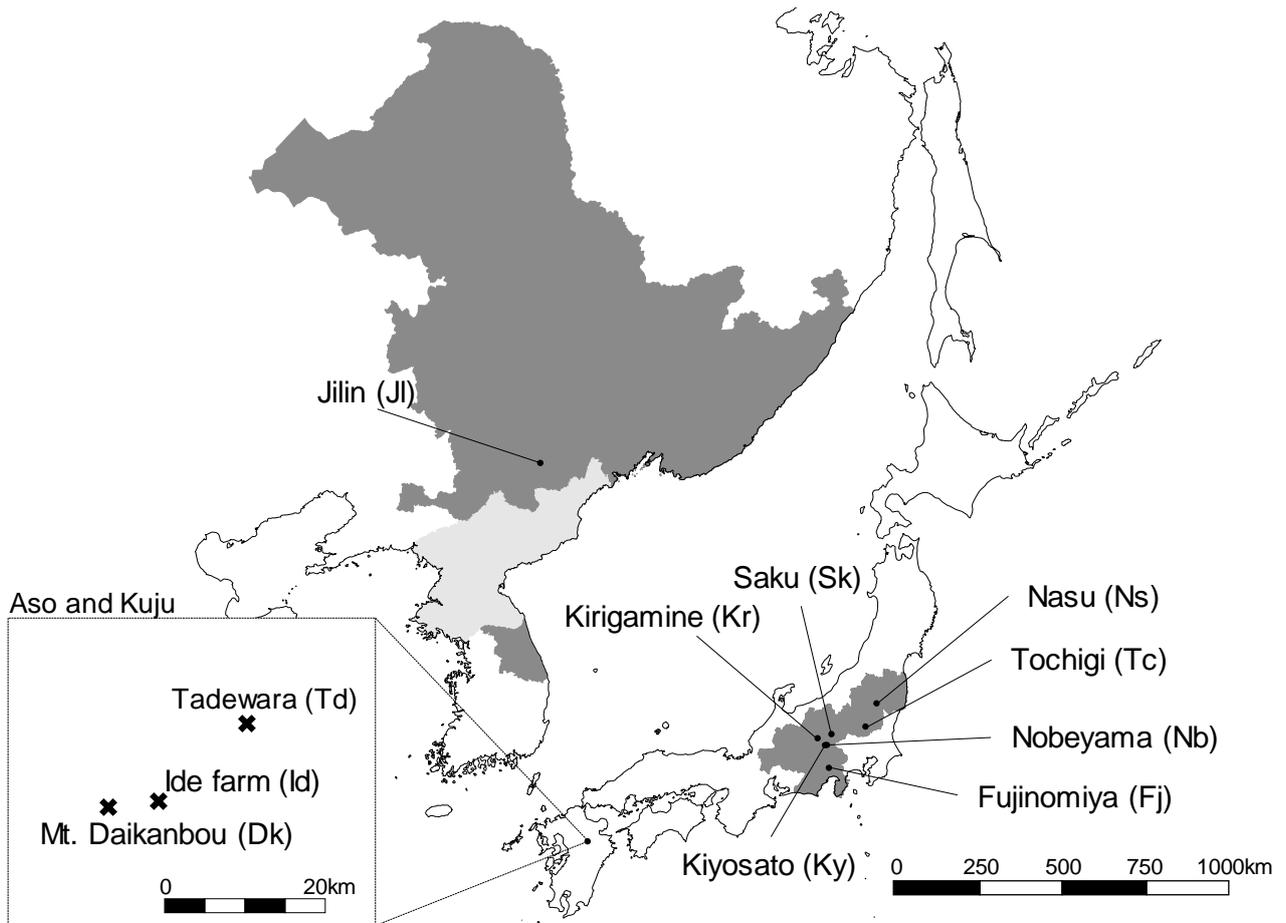
Fig. 2-1. Location of the study site. The right part of the figure is a wide-area map of the study site, while left is an enlarged view of the Aso-Kuju region. Each population from the Aso-Kuju region is divided into two: Tadewara_A and _B, Ide Farm_A and _B, and Mt. Daikanbou_A and _B. The characters within parentheses are the abbreviated population names. Dark gray blob shows distribution area and light gray blob shows uncertain distribution area.
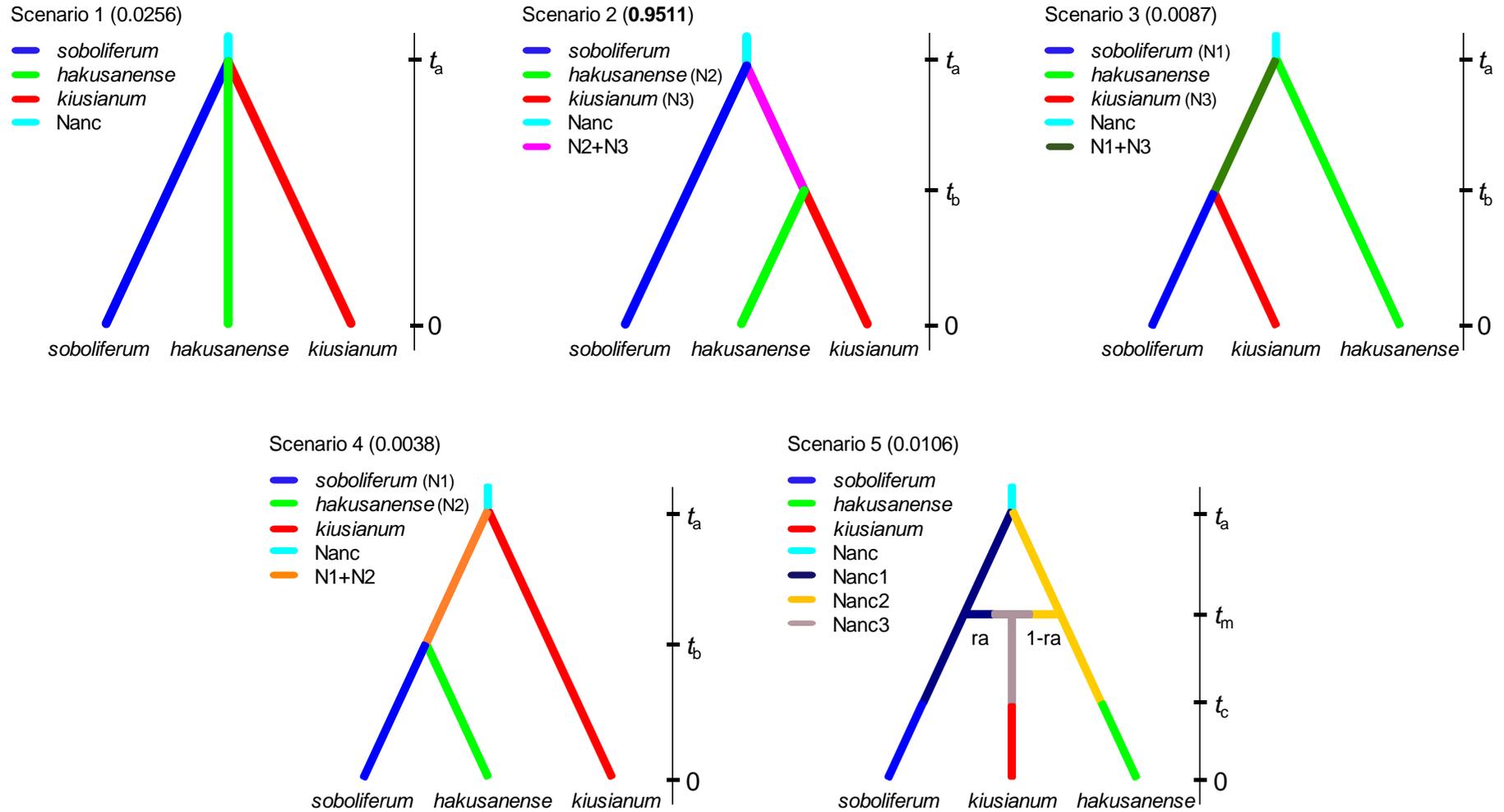
Fig. 2-2. Five demographic scenarios considered in the genetic demographic analysis. Pop1 is the Jilin population, Pop2 is *G. soboliferum* var. *hakusanense* from central Honshu, and Pop3 is *G. soboliferum* var. *kiusianum* from northern Kyushu. Nx is the effective population size ("x" is the arbitrary number of populations or ancestral populations). t_x is the branching time for each population or the mergence time of two populations, and "0" is the present time. The numbers within parentheses are the posterior probabilities of the simulations.
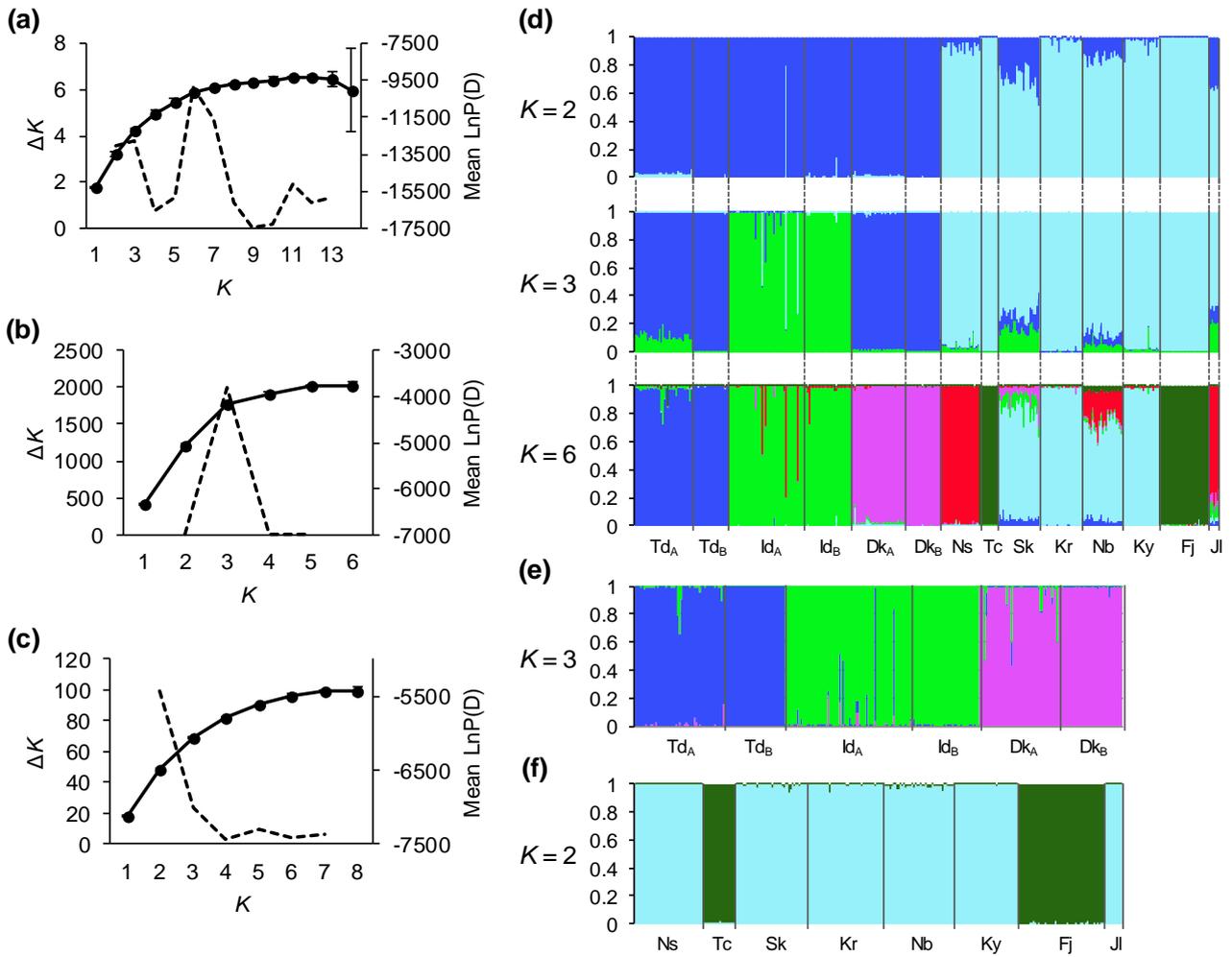
Fig. 2-3. Results of STRUCTURE analysis. (a), (b), (c) ΔK (dotted line) and Mean est. LnP(D) (solid line). (d), (e), (f) Bar plots for the appropriate numbers of clusters. (a) and (d) are the results for all populations in this study. (b) and (e) are the results of *G. soboliferum* var. *kiusianum*. (c) and (f) are the results of *G. soboliferum* var. *hakusanense* and var. *soboliferum*. Each bar plot color represents an individual's genetic cluster. The Y-axis shows the probability of each individual's genetic cluster. Td$_A$, Tadewara_A; Td$_B$, Tadewara_B; IdA, Ide Farm_A; IdB, Ide farm_B; Dk$_A$, Mt. Daikanbou_A; Dk$_B$, Mt. Daikanbou_B, Ns, Nasu; Tc, Tochigi; Sk, Saku; Kr, Kirigamine; Nb, Nobeyama; Ky, Kiyosato; Fj, Fujinomiya; Jl, Jilin.

34
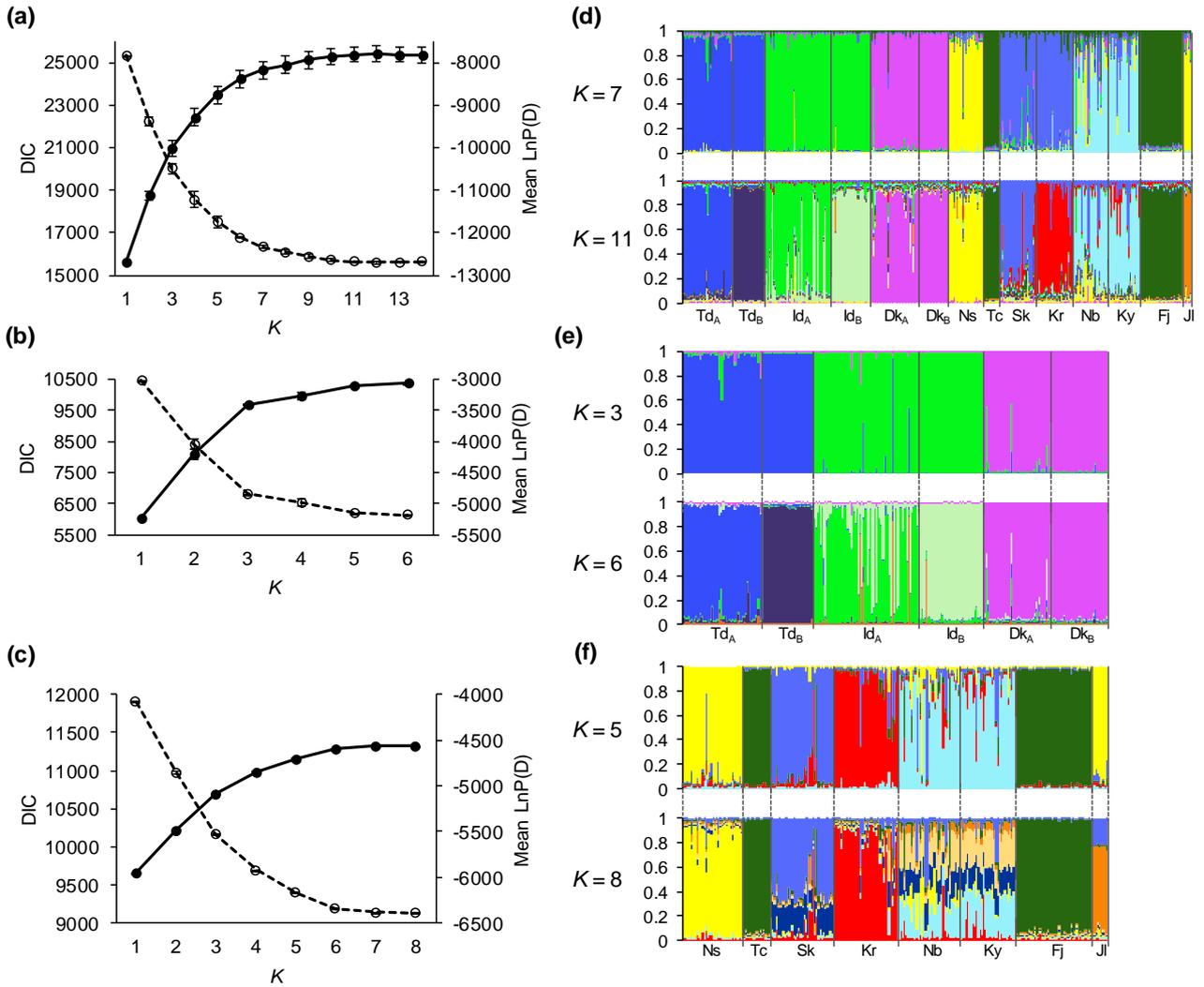
Fig. 2-4. Results of InStruct. (a), (b), (c) Deviance information criterion (DIC; dotted line) and Mean est. LnP(D) (solid line). (d), (e), (f) Bar plots for the appropriate numbers of clusters. (a) and (d) are the results of all populations in this study. (b) and (e) are the results of *G. soboliferum* var. *kiusianum*. (c) and (f) are the results of *G. soboliferum* var. *hakusanense* and var. *soboliferum*. Each bar plot color represents the individual's genetic cluster. The Y-axis shows the probability of each individual's genetic cluster. Td_A, Tadewara_A; Td_B, Tadewara_B; Id_A, Ide Farm_A; Id_B, Ide Farm_B; Dk_A, Mt. Daikanbou_A; Dk_B, Mt. Daikanbou_B, Ns, Nasu; Tc, Tochigi; Sk, Saku; Kr, Kirigamine; Nb, Nobeyama; Ky, Kiyosato; Fj, Fujinomiya; Jl, Jilin.
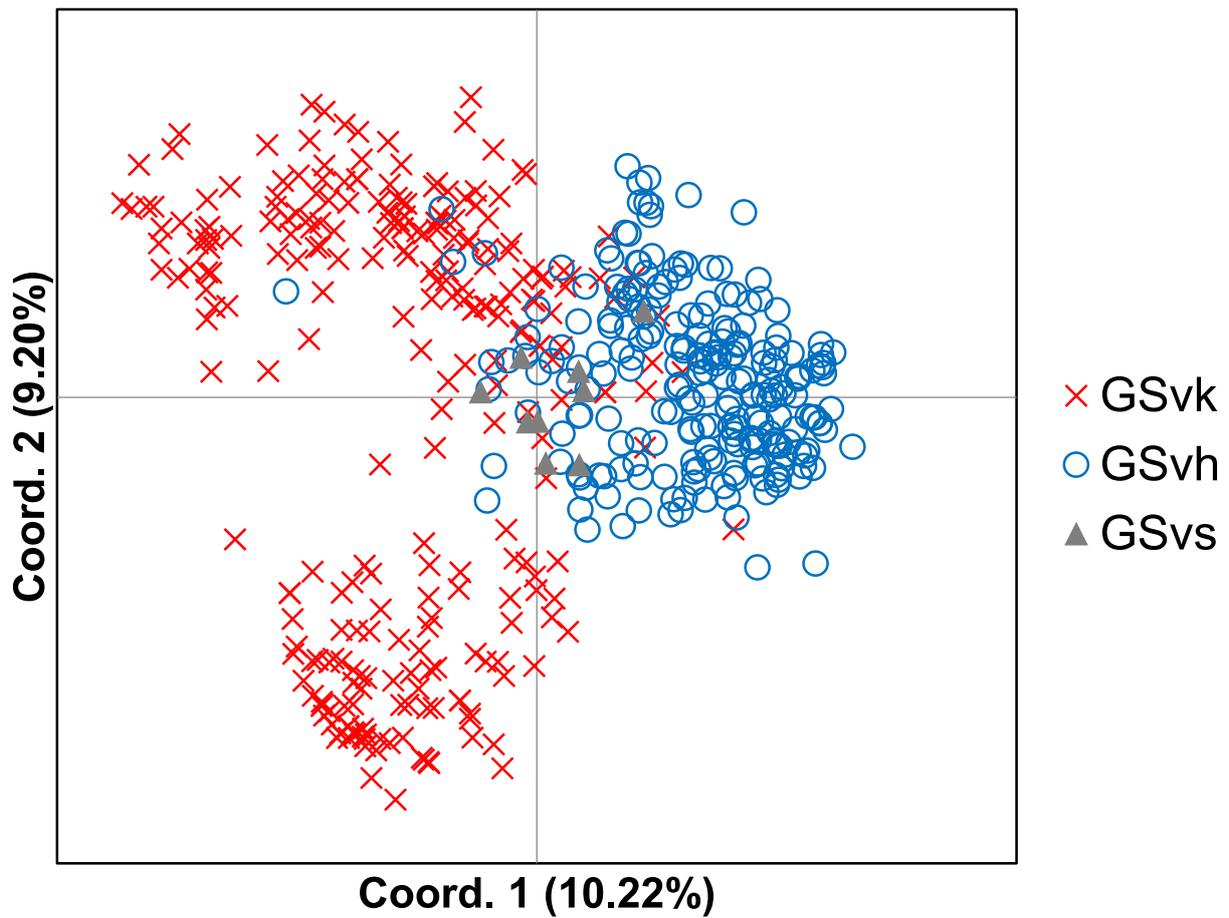
Fig. 2-5. Principal coordinate analysis (PCoA) of the interpopulation of *G. soboliferum* var. *kiusianum*, *G. soboliferum* var. *hakusanense* and *G. soboliferum* var. *soboliferum*. GSvk *G. soboliferum* var. *kiusianum*; GSvh, *G. soboliferum* var. *hakusanense*; and GSvs *G. soboliferum* var. *soboliferum*. Variability (%) is shown by the first three axes, namely 10.22%, 9.20% and 7.16%.
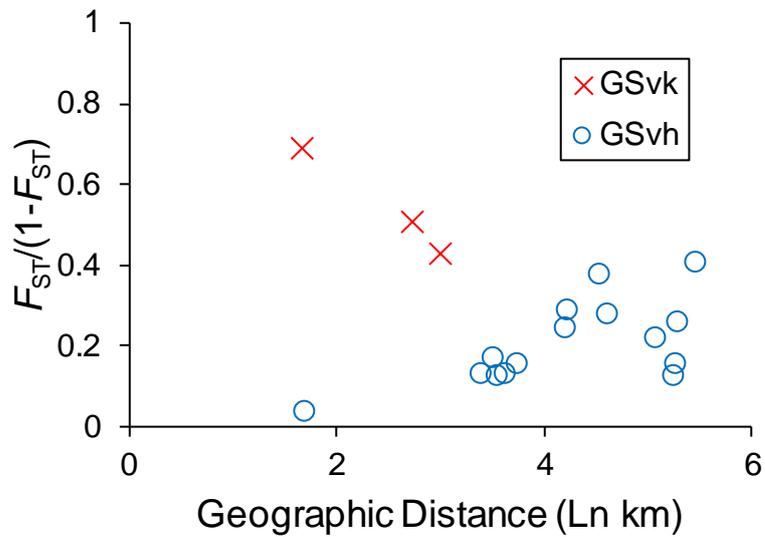
Fig. 2-6. Isolation by distance of *G. soboliferum* var. *kiusianum* and *G. soboliferum* var. *hakusanense*. GSvk; *G. soboliferum* var. *kiusianum*, GSvh; *G. soboliferum* var. *hakusanense*. The horizontal axis represents geographic distance and the vertical axis represents genetic distance; $F_{ST} / (1 - F_{ST})$. A broken line is an approximately straight line and a numerical expression. r^2 is the coefficient of determination. The Mantel test did not detect any significant correlation between the geographic and genetic distances of either variety ($P > 0.05$). Approximately straight lines of GSvk and GSvh are y = -0.1909x + 1.0115 and y = 0.0624x – 0.0551, respectively.

Chapter 3 Phylogeographic study of species complex *G. yesoense* Franch. et Sav.

3-1. Introduction

During the Quaternary, which began ca. 2.4 million years ago, Earth experienced global cooling and climate fluctuations, which intensified beginning 1 million years ago (Hewitt 2000; Davis and Shaw 2001). Since then, the climate has cycled between glacial and interglacial periods every hundred thousand years (100 kyr), affecting the distributions of living organisms (Hewitt 1996) including alpine plants. During glacial periods, alpine plants persisted in ice-free nunataks and peripheral areas at high altitudes and in grassy areas at low altitudes; these areas functioned as glacial refugia (Holderegger & Thiel-Egenter 2009). Conversely, alpine plants persisted during interglacial periods in high-altitude grassy areas in temperate regions; these comprise interglacial refugia (e.g., Zhang et al. 2001; Ikeda & Setoguchi 2007). Various phylogenetic and population genetic studies have identified the locations of glacial and interglacial alpine refugia worldwide. For example, in the European Alps, which have been repeatedly glaciated during the Quaternary (Cox & Moore 2000), high mountain taxa originated from lowland, postglacial immigrants from peripheral refugia (Stehlik 2002; Comes & Kadereit 2003). In North America, large ice-free areas existed in the vicinity of Beringia, providing a major refugium for alpine plants (Hultén 1937), with some species originating in this area (e.g., Tremblay & Schoen 1999; Abbott et al. 2000; Abbott & Comes 2004; Alsos et al. 2005; Eidesen et al. 2007; Skrede et al. 2009; Allen et al. 2012).

In recent years, molecular studies have found that populations in refugial regions are not genetically homogeneous, but are fragmented instead, with populations from different refugia differing in their genetic structures. This phenomenon is known as "refugia within refugia," and many wild species retain this sort of phylogeographic structure (Gomez & Lunt 2007). Complex genetic structures have been documented in numerous alpine plant populations

in the Alps, and glacial refugia in this region are subdivided into several smaller refugia (e.g., Schönswetter et al. 2004, 2005; Ehrich et al. 2007; Ronikier et al. 2008; Alvarez et al. 2009). Some western North American alpine plants may have originated in different refugia, and cryptic refugia might have contributed to their persistence during the last glacial maximum (e.g., Marr et al. 2008, 2013; Allen et al. 2012). In the East African mountains, *Arabis alpina* populations were subdivided into different population groups (Ehrich et al. 2007). It is thus apparent that fragmentation can generate refugia within refugia in both the Northern and Southern Hemispheres. In addition, the phenomenon of recurrent immigration may result in a high degree of genetic variation within refugial areas. The Arctic-alpine species *Saxifraga oppositifolia* is characterized by high genetic diversity, which is attributable in part to recurrent immigration (Abbott et al. 2000). Colonization of some Arctic-alpine plants in Svalbard has probably occurred numerous times (Gabrielsen et al. 1997; Tollefsrud et al. 1998; Skrede et al. 2006), and diverse lineages of *Oxyria digyna* in North America have resulted from repeated distribution shifts over multiple glacial cycles (Allen et al. 2012). In the Japanese Archipelago, recurrent immigration is thought to have occurred in high altitude mountain areas where southern and northern linages of *Pedicularis* are in secondary contact (Fujii et al. 1997), but there are few studies demonstrating that recurrent immigration has increased lineage diversity in southern refugial areas.

The Japanese alpine flora is believed to have originated from, and is similar to, that of eastern Siberia and the Bering district (Koidzumi 1919). Hultén (1937) has discussed local endemism among Japanese alpine plants of Arctic origin. Recent phylogenetic studies have demonstrated that the distributions of alpine plants in Japan had shifted numerous times in the past million years (Fujii et al. 1997, 1999; Senni et al. 2005). According to Koidzumi (1919) and Hultén (1937), Japanese alpine plants migrated to the Japanese Archipelago from the north

via Sakhalin, Kuril Island, and Kamchatka during glacial periods, and their distribution was restricted to alpine areas during interglacial periods. During interglacial periods, the high mountains (up to 3000 m asl) of central Honshu have provided refugia for many alpine plants (Koidzumi 1919; Fujii et al. 1997, 1999; Senni et al. 2005; Ikeda & Setoguchi 2007, 2013a; Ikeda et al. 2008a, 2008b, 2008c, 2009a, 2009b, 2014b). Other studies have explored the demographic histories of northern and southern populations of Japanese alpine plants. The northern and southern Japanese populations of *Cardamine nipponica* (Brassicaceae) and *Cassiope lycopodioides* (Ericaceae), which are distributed from central Honshu to Alaska, diverged from each other ca. 110,000–200,000 years ago (Ikeda et al. 2009a; Ikeda et al. 2014b). On the other hand, endemic plants of *Phyllodoce nipponica* (Ericaceae) diverged from other related species ca. 800,000 years ago (Ikeda et al. 2014a), and northern and southern Japanese populations diverged earlier (180,000–250,000 years ago; Ikeda & Setoguchi 2013a). Similar patterns of genetic differentiation between central Honshu and northern Japan have been demonstrated (Ikeda et al. 2009a, 2014b; Ikeda & Setoguchi 2013b), along with differences in divergence timing, suggesting that alpine plants have had multiple opportunities to migrate into the mountains of central Honshu. Cryptic or fragmented refugia may have emerged under a scenario of multiple immigration/emigration events, as has been demonstrated elsewhere (Schönswetter et al. 2004, 2005; Ehrich et al. 2007; Marr et al. 2008, 2013; Ronikier et al. 2008; Alvarez et al. 2009; Allen et al. 2012). However, my understanding of the processes driving the formation of alpine plant populations in southern refugia remain limited.

Geranium yesoense Franch. et Sav. (Geraniaceae) is an herbaceous plant found in the Japanese Archipelago (Ohwi 1965; Akiyama 2001). It is classified into three varieties based on leaf incision and the density of sepal pubescence (*G. yesoense* var. *yesoense*, var. *nipponicum* Nakai, and var. *pseudopratense* Nakai) (Akiyama 2001; Ikeda et al. 2015; Kadota 2016).

Among these, only var. *nipponicum* is adapted to high-altitude areas, whereas the other two varieties occur in low-altitude coastal habitats in northern regions. Hara (1948) indicated that individuals growing on Mt. Ibuki (Maibara City, Shiga Pref.) had densely hairy sepals and were similar to var. *yesoense* (Ohwi 1965); Akiyama (2001) later identified this population as *G. yesoense* var. *yesoense*. Hara (1948) observed that individuals around Mt. Asama (Tsumagoi Village, Gunma Pref.) were also morphologically similar to var. *yesoense* based on sepal hairiness. Both mountains are located at the southern edge of the high mountain region of central Honshu. Because var. *yesoense* is distributed in low-altitude coastal habitats in more northerly regions, individuals growing on Mt. Ibuki and Mt. Asama may represent disjunct populations that originated from past immigration events. Alternatively, if sepal hairiness represents an environmental adaption, the "var. *yesoense*" populations on these two southern mountains, which are distinguished solely by sepal hairiness, may simply represent an ecotypic variation of the surrounding var. *nipponicum* populations. Exploring these two alternative hypotheses can help with elucidating the biogeographic process that generated the morphological diversity observed in southern interglacial refugia (i.e., recurrent immigration vs. one-time immigration and subsequent parallel evolution of densely hairy sepals).

Wakasugi et al. (2017) quantified morphological traits and performed phylogenetic analysis for this species complex; the results indicated that the Mt. Ibuki population had densely hairy sepals, but that other morphological traits did not differ among the three varieties. They found no clear genetic differences among the taxa and recommended that the varieties be synonymized with var. *yesoense* (Wakasugi et al. 2017). However, they used only two genetic markers (a chloroplast fragment and an internal transcribed spacer region) in their phylogenetic analysis. Because both genetic markers are suitable for DNA barcoding studies, which is intended for species-level identification (Kress et al. 2005; Li et al. 2011), they may lack the

power required for comparing intraspecific taxa. Recently, obtaining genome-wide genetic polymorphisms of non-model organisms has become easier due to advancements in sequencing technology. In particular, genotyping-by-sequencing (e.g. restriction site-associated DNA markers [Baird et al. 2008] and multiplexed inter-simple sequence repeat [ISSR] genotyping-by-sequencing [MIG-seq; Suyama & Matsuki 2015]) enables us to obtain large volumes (hundreds to hundreds of thousands) of single nucleotide polymorphism (SNP) data, and has been used in population genetic studies (e.g. Takahashi et al. 2016; Watanabe et al. 2018; Hirano et al. 2019).

My objective was to explore the genetic structure and phylogenetic relationships among *G. yesoense* varieties and populations using genome-wide SNP genotyping and next-generation sequencing of whole chloroplast genomes. Specifically, I tested alternative biogeographic scenarios that might have generated the morphologically distinct populations found in southern alpine areas (Mt. Ibuki and Mt. Asama), as a means of exploring the potential presence of refugia within refugia in the Japanese Archipelago. To this end, I evaluated the following: (1) phylogenetic relationships among *G. yesoense* varieties and the genetic structure of *G. yesoense* populations using genome-wide SNP markers, and (2) the demographic and phylogeographic history of *G. yesoense* varieties, using coalescent simulations based on the approximate Bayesian inference method (i.e., divergence time and presence of admixture events across varieties/populations).

3-2. Materials and methods

Morphological variation in G. yesoense Franch. et Sav.

Akiyama (2001) and Kadota (2016), who are the current taxonomic authorities for Japanese *Geranium* species, recognize three morphologically and ecologically distinct varieties of *G. yesoense*. Both *G. yesoense* var. *yesoense* and var. *pseudopratense* have pedicels with spreading hairs (Akiyama 2001; Kadota 2016); the former also has dense, spreading hairs on the sepals and deeply incised leaves, whereas the latter has sparsely hairy sepals and shallowly incised leaves (Akiyama 2001). *Geranium yesoense* var. *nipponicum* has retrorse or sparse spreading hairs on the pedicels and appressed or spreading hairs on the sepals (Akiyama 2001; Kadota 2016). Both var. *yesoense* and var. *pseudopratense* are distributed in coastal grasslands, whereas var. *nipponicum* is found in high mountain areas (Ohwi 1965; Shimizu 1982, 1983). Distinguishing characteristics of the three varieties are summarized in Table 3-1.

Sample collection

Between July 2017 and August 2018, I collected mature leaves from 42 populations of var. *yesoense*, var. *nipponicum*, and var. *pseudopratense* (Fig. 3-1, Table 3-1, Table 3-3)—15 populations of var. *yesoense*, 16 populations of var. *nipponicum*, and 11 populations of var. *pseudopratense*. In addition, I collected 10 samples from the Mt. Ibuki population, and eight samples from the Mt. Asama populations (Fig. 3-1, Table 3-1, Table 3-3). A total of 168 leaf samples were collected from 45 populations, dried immediately using silica gel, and preserved at room temperature in a dark room.

Phylogenetic analysis of complete chloroplast genome sequences

One specimen of *G. yesoense* var. *yesoense*, collected from Akkeshi Town (Hokkaido

43

Pref.), one specimen of var. *nipponicum* from Mt. Gassan (Tsuruoka City, Yamagata Pref.), one specimen of var. *pseudopratense* from Hakodate City (Hokkaido Pref.), and one specimen from Mt. Ibuki (Maibara City, Shiga Pref.) were chosen to represent their respective varieties. I also included *G. soboliferum* var. *hakusanense* as an outgroup.

Prior to DNA extraction, I washed the fine leaf powder samples using a buffer consisting of 3% HEPES buffer (pH 8.0), 1% polyvinylpyrrolidone, 0.9% ascorbic acid, and 2% 2-mercaptoethanol. Genomic DNA was extracted using a modified cetyltrimethylammonium bromide method (Milligan 1992), and total DNA was measured using a Qubit 2.0 fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). Paired-end sequencing was performed using the BGISEQ-500 next-generation sequencing platform (BGI, Shenzhen, China). Low-quality reads were removed using Trimmomatic-0.39 (Bolger et al. 2014). When the average quality per four sequences dropped below 15, all subsequent sequences were removed (SLIDINGWINDOW: 4:15). Sequences were also removed at the beginning or end of reads with quality < 20 (LEADING: 20 and TRAILING: 20). In addition, reads that were shorter than 30 bp were removed (MINLEN: 30). The trimmed sequence reads were assembled *de novo* using NOVOPlasty (Dierckxsens et al. 2017); the *G. incanum* (NC_030045.1) plastid sequence was used as a reference sequence, and each locus (*ccsA*, *cemA*, *rbcL*, *ndhJ*, *psbK*, *rpoA*, and *psbZ*) of *G. incanum* was designated as a seed sequence in the assembly analyses. Because the assembly did not yield continuous sequences of sample chloroplasts, contigs around the above loci were used in the phylogenetic analysis. Contigs were aligned using MAFFT 7.310-1 (Katoh et al. 2002, 2013). All gaps originating from inversions and/or deletions were eliminated, and phylogenetic analysis was performed based on nucleotide sequence using RAxML 8.2.11 (Stamatakis 2014). I used the nucleotide substitution model GTRGAMMA, and the number of bootstraps was set to 1000. I later calculated bootstrap values of the best tree using the

pgsumtree option in Phylogears2 (Tanabe 2008). The resulting maximum likelihood (ML) tree was visualized and edited in FigTree 1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/).

Divergence time between var. *nipponicum* and other varieties was calculated using the formula $T = k/2v$, where T = divergence time, k = evolutionary distance, and v = evolutionary rate. Branch length, which was estimated using RAxML, was used as the evolutionary distance (k), and $1.0–3.0 \times 10^{-9}$ was used as the evolutionary rate (Wolfe et al. 1987).

MIG-seq experiment

MIG-seq with a next-generation sequencing platform was used to identify genome-wide SNPs (Suyama & Matsuki 2015). Using this method, the sequence library comprises the ISSR loci, which is amplified by multiplex polymerase chain reaction (PCR). After library size selection, next-generation sequencing was performed (Suyama & Matsuki 2015). During the first PCR step, I amplified ISSRs using the following primers: $(ACT)_4TG$, $(CTA)_4TG$, $(TTG)_4AC$, $(GTT)_4CC$, $(GTT)_4TC$, $(GTG)_4AC$, $(GT)_6TC$, $(TG)_6AC$, $(ACT)_4TG$, $(CTA)_4TG$, $(TTG)_4AC$, $(GTT)_4CC$, $(GTT)_4TC$, $(GTG)_4AC$, $(GT)_6TC$, and $(TG)_6AC$. At the second PCR step, common forward and reverse primers were used, including nine- and five-base indices, respectively, and an adapter sequence for Illumina sequencing. PCR products were then pooled, and size selection was carried out using the magnetic bead method (SPRIselect; Beckman Coulter, Brea, CA, USA), with a target size of 350–800 bp. Finally, the library was sequenced using the Illumina MiSeq platform (Illumina, San Diego, CA, USA) and a MiSeq Reagent kit v3 (Illumina).

SNP detection

The first 14 read-2 (reverse-read) bases and low-quality sequences of raw reads were

removed using the FASTX-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/), with the minimum quality score set to 30 (-q 30). Reads comprising > 40% bases that failed to meet the minimum quality score were removed (-p 40). In addition, the primer sequence region of each read was removed using TagDust (Lassmann et al. 2009). After removal of unnecessary reads, the remaining reads were assembled with Stacks 1.48 (Catchen et al. 2013), and SNPs were called using the same program. In *ustacks*, the maximum distance (in nucleotides) option was set to 2 (-M 2) and the minimum depth option for creating the stack was set to 3 (-m 3), while the maximum distance option (-N) was set to 2. Catalogs were created using *cstacks*, and the genotypes of individuals were determined using *sstacks*. SNPs for phylogenetic and population analysis were exported using *Populations*, with the following settings: genotyping rate, which is the minimum percentage of individuals in a population, was set to 0.1 (-r 0.1); minor alleles (< 1% within population) were removed (--min-maf 0.01); and loci with unreasonable heterozygosity were removed (--max-obs-het 0.5). A total of 10,223 SNPs were detected. I filtered these using PLINK 1.9 (Purcell et al., 2007), removing individuals which had 80% missing data, and loci that were held by < 40% of individuals (--mind 0.8, --geno 0.4). Minor alleles (< 3%) were also removed (--maf 0.03). Finally, individuals that had more than 45% missing data were removed. These SNPs were used in phylogenetic and STRUCTURE analyses (Pritchard et al., 2000), and for calculating summary statistics. For demographic analysis, because at least one individual must host a genotype within a population, the locus of any genotype that was absent in all individuals was removed from the analysis.

Population analysis

I used GenAIEx 6.501 (Peakall & Smouse 2006) and hierfstat (Goudet 2005) to calculate the following four summary statistics to characterize each population: allelic richness

(Ar), expected heterozygosity (H_e), observed heterozygosity (H_o), and fixation index (F_{IS}). I also

assessed the correlation between Ar and latitude using R 3.6.1 (R Development Core Team,

2019).

The F_{ST} index (Weir & Cockerham 1984) between populations was calculated using

GENEPOP 4.7.2 (Raymond & Rousset 1995). We used the model-based Bayesian clustering

program STRUCTURE 2.3.4 (Pritchard et al. 2000), which assigns individuals to K clusters, to

infer genetic structures associated with the samples. Population structure was estimated for the

overall sample, as well as for each taxon separately, with $K = 1–10$ under an admixture and

correlated allele frequencies model (F model) (Falush et al. 2003). Simulation runs consisted of

100000 Markov chain Monte Carlo simulations, following a burn-in period of 100000 iterations.

For each value of K, 20 runs were performed, and I used the ΔK method (Evanno et al. 2005)

and the value of ln P(D) to determine the optimal value of K by maximizing ΔK or ln P(D). If ln

P(D) increased with the number of clusters, the value of K is defined as the value immediately

before the value of ln P(D) levelled out.

Additionally, I evaluated the isolation by distance for each variety to assess the

relationship between geographic and genetic distance. The geographic distance associated with

each population was calculated based on latitude and longitude coordinates, and $F_{ST}/(1 - F_{ST})$

was used to represent genetic distance (Rousset 1997). The significance of correlations was then

examined with a Mantel test using GenAlEx 6.501.

Phylogenetic analysis was performed using the concatenated SNP matrix and RAxML

8.2.11; I adopted the GTRGAMMA nucleotide substitution model and the JC69 model of DNA

evolution (Jukes & Cantor 1969). I set the number of bootstraps at 1000 and later calculated

bootstrap values of the best tree using the pgsumtree option in Phylogears2. The resulting ML

tree was visualized and edited in FigTree 1.4.3.

Demographic analysis

Approximate Bayesian computation (ABC) is a powerful and flexible approach for estimating demographic and historical parameters, and for quantitative comparison of alternative scenarios (Bertorelle et al. 2010). I used the ABC method, implemented in DIYABC 2.1 (Cornuet et al. 2014), to determine the most probable demographic scenario with respect to *G. yesoense* varieties.

To define population units in my ABC analyses, *G. yesoense* was divided into three varieties according to morphological delimitations. In addition, because phylogenetic and STRUCTURE analyses suggested that the Mt. Ibuki and Mt. Asama populations, which are located in central Honshu, were genetically close to those of var. *pseudopratense* (see Results), I included them as separate populations.

I used a two-step process for the demographic analysis of the five populations (defined as *G. yesoense* var. *yesoense*, var. *nipponicum*, var. *pseudopratense*, Mt. Ibuki, and Mt. Asama). First, I performed ABC analysis to select the most probable divergence scenario for the three varieties. Next, I conducted the actual demographic analysis of all five populations. Divergence scenarios (hereafter referred to as Scenarios 1-1 to 1-5) for the three varieties were defined as follows: (1-1) the three varieties diverged simultaneously, reflecting global climate change; (1-2) the northern (Hokkaido and northern Tohoku district) and southern (central Honshu) populations diverged; followed by a subsequent divergence of northern populations along an east–west axis; (1-3) the west–east divergence occurred in northern Japan, after which western populations in northern Japan expanded their distribution to central Honshu. Ancestral populations of var. *nipponicum* and var. *pseudopratense* were then isolated in two separate regions during interglacial periods; (1-4) ancestral populations of var. *pseudopratense* first diverged from ancestral populations of var. *yesoense*, and eastern populations in northern Japan

48

later expanded their distribution to central Honshu, after which the ancestral population of var. *yesoense* and var. *nipponicum* was isolated in two regions during interglacial periods; and (1-5) the intermediate phylogenetic position of var. *pseudopratense* may indicate that populations were generated by past admixture of var. *yesoense* and var. *nipponicum*, which expanded their distributions in response to cooler and drier climates.

I found that Scenario 1-2 was the best supported divergence scenario (Fig. S3-1); Scenario 1-5 was also supported by a moderate posterior probability (Fig. S3-1). Although Scenario 1-1 was also associated with moderate posterior probability, I did not factor this scenario into the second step of the ABC analysis because Scenarios 1-1 and 1-2 were nearly identical aside from the timing of first divergence. Following these branching patterns, I established nine possible demographic scenarios for the five populations (hereafter referred to as Scenarios 2-1 to 2-9). In Scenario 2-1, northern populations expanded their distribution to central Honshu, where they persisted in high altitude areas during interglacial periods. The west–east (west coast vs. east coast) divergence occurred at a later date in northern Japan, after which west coast populations expanded their range to central Honshu along the coast, while the Mt. Ibuki and Mt. Asama populations diverged gradually during interglacial periods. Scenario 2-2 assumes an identical divergence history to Scenario 2-1 up to the divergence in northern Japan, but the Mt. Asama population is assumed to have originated from ancestral populations of var. *nipponicum*. Scenario 2-3 also shares a divergence history with Scenario 2-1 up to the divergence event in northern Japan but assumes that the populations on Mt. Ibuki and Mt. Asama originated from secondary contact between var. *nipponicum* and var. *pseudopratense* during glacial periods. Scenario 2-4 is similar to Scenario 2-3 but differs with respect to contact timing. Scenario 2-5 is similar to Scenarios 2-3 and 2-4 but assumes that only the Mt. Asama populations originated from secondary contact during glacial periods, whereas the Mt. Ibuki

population diverged from ancestral populations of var. *pseudopratense* during interglacial periods. Under Scenario 2-6, ancestral populations of var. *yesoense* and var. *nipponicum* diverged during interglacial periods, and hybrid populations later evolved into var. *pseudopratense* through secondary contact in northern Japan, after which ancestral populations of var. *pseudopratense* migrated southward during glacial periods, persisting on Mt. Ibuki and Mt. Asama during interglacial periods. Scenario 2-7 is similar to Scenario 2-6 but differs with respect to the timing of divergence of Mt. Ibuki and Mt. Asama populations from ancestral var. *pseudopratense*. Scenario 2-8 is similar to Scenarios 2-6 and 2-7 up to the point of secondary contact between var. *yesoense* and var. *nipponicum*. Thereafter, ancestral populations of var. *nipponicum* and var. *pseudopratense* are assumed to have had secondary contact in central Honshu, and these hybrid populations persisted on Mt. Asama. Still later, ancestral populations of var. *pseudopratense* migrated southward to Mt. Ibuki along the west coast, where they persisted during interglacial periods. Finally, Scenario 2-9 is similar to Scenario 2-8 but the timing of divergence between var. *pseudopratense* and the Mt. Ibuki population, as well as the timing of secondary contact between the two varieties (var. *nipponicum* and var. *pseudopratense*) are reversed. These nine scenarios are shown in Fig. 2. The historical distributions of each demographic parameter are summarized in supplementary text.

Next, I established a set of six summary statistics. These included the mean genetic diversity, F_{ST} distance, and Nei's distance (excluding zeros) (Nei 1972, 1987), plus their associated variances, as well as the mean and variance of admixture estimates (excluding zeros), which were used to represent admixture among groups (Choisy et al. 2004). Five million simulations were performed for each scenario, and the most appropriate scenario was determined by comparing posterior probabilities. I also assessed the effects of the number of simulated data points closest to the observed data points, which were used to evaluate posterior

probabilities by varying the proportional distance to the closest data point (i.e. every 0.1% in 0.1–1.0% of the closest simulations). Divergence time in relation to absolute time was calculated by multiplying the estimated time parameter with the number of generations, with 3–5 years estimated to be the generation time for the study species (Alpine Garden Society of Tokyo 2004).

Observation and analysis of the density of spreading hairs on the sepals

I followed the methods of Wakasugi et al. (2017) for analyzing the density of sepal hairs. I arbitrarily chose one flower bud from each specimen and counted the number of spreading hairs along the central vein of the sepal (Wakasugi et al. 2017) using a stereomicroscope (MZ95; Leica, Wetzlar, Germany). I classified hair density into three categories as follows: dense ($\geq$ 15 spreading hairs), moderate (10–14 hairs), and sparse ($\leq$ 9 hairs) (Wakasugi et al. 2017). I converted these three classes into numerical scores, whereby "dense" = 3 points, "moderate" = 2 points, and "sparse" = 1 point, and produced a ternary diagram to assess the relationship between genetic differentiation (Q-value from STRUCTURE analysis) and pubescence scores. Observations were made at the National Museum of Nature and Science in Tokyo. A total of 714 individuals from 30 populations were included in my assessment. This included data from 443 individuals, representing nine populations, that were assessed by Wakasugi et al. (2017), and an additional 271 individuals from 21 populations that I assessed for this study.

3-3. Results

Chloroplast genome phylogeny

As the result of *de novo* assembly, seven relatively long, stable contigs were obtained from the regions surrounding the *cemA*, *ccsA*, *rpoA*, *ndhJ*, *psbZ*, *rbcL*, and *psbK* genes in each taxon. The lengths of contigs surrounding each gene were as follows: *cemA*, 17,287–18,839 bp; *ccsA*, 24,191–24,642 bp; *rpoA*, 10,195–11,827 bp; *psbK*, 8248–10,429 bp; *psbZ*, 7113–7786 bp; *ndhJ*, 17,282–17,302 bp; and *rbcL*, 6359–19,573 bp (Table 3-2). Following alignment, contigs from the four taxa were concatenated, and the resulting concatenated sequence was used in phylogenetic analysis (total length, 82,430 bp).

The ML tree indicated that var. *nipponicum* branched off from other *G. yesoense* varieties at the base of the *G. yesoense* clade (Fig. 3-3). A monophyletic clade comprising var. *yesoense*, var. *pseudopratense*, and one individual from Mt. Ibuki was supported with a high bootstrap value (= 100) (Fig. 3-3). The evolutionary distance between var. *nipponicum* and the other varieties was 0.0015 (0.0008 + 0.0007). Calculation of divergence time assuming a chloroplast genome evolutionary rate of $1.0–3.0 \times 10^{-9}$/site/year indicated that the two groups diverged ca. 0.25–0.75 million years ago.

Although I obtained a total of 5 GB of read data in this study, I was unable to obtain a complete chloroplast genome sequence. The chloroplast genome of Geraniaceae is characterized by a high frequency of large repeats (> 100 bp) relative to those of other rosids, and the number of large repeats (> 200 bp) in *Geranium* is the highest among the genera in Geraniaceae (Guisinger et al. 2011). The length of the sequence reads was 100 bp, and that of the insert sequence was 300 bp. If large repeats (> 100 bp) occurred in the *G. yesoense* chloroplast genome, the assembler would not be able to form contigs from those regions.

Nuclear genome phylogeny based on MIG-seq

The total and average number of reads obtained from MIG-seq were 16,734,570 and 98,438 reads, respectively. Reads were processed via the Stacks pipeline, which yielded 10,223 SNPs from 168 samples. After filtering, 280 SNPs and 158 samples remained in the data matrix. Summary statistics for each variety and population are shown in Table 3-3. The H_e and H_o of var. *yesoense* were 0.077 and 0.056, respectively. The *Ar* varied from 1.023 (Ogawara) to 1.090 (Onbetsu). The H_e and H_o for var. *nipponicum* were 0.169 and 0.071, respectively, and its *Ar* ranged from 1.055 (Mt. Sannomine) to 1.131 (Mt. Tate). The H_e and H_o of var. *pseudopratense* were 0.138 and 0.070, respectively, and its *Ar* ranged from 1.053 (Rishiri) to 1.104 (Shiribetsu). The H_e and H_o of the Mt. Ibuki plants were 0.114 and 0.065, respectively, and their *Ar* was 1.134. The H_e and H_o of the Mt. Asama plants were 0.112 and 0.088, respectively, and their *Ar* were 1.099 and 1.126. There was a significant correlation between *Ar* and latitude (correlation coefficient = −0.475, P = 0.001); however, there were no significant correlations within varieties (Fig. 3-4).

Based on 280 SNPs and 158 individuals, the phylogenetic ML tree separated var. *yesoense* and var. *nipponicum* into two major clades (Clades A and B), whereas var. *pseudopratense* was located at an intermediate position between these two varieties with relatively high support (Fig. 3-5). Clade A (bootstrap value = 96) included var. *yesoense*, which is distributed along the eastern side of Hokkaido and in northern Honshu. Although var. *pseudopratense* populations were polytomic, they were clearly separated from the other two varieties with high bootstrap support (bootstrap value = 96 and 84). In addition, the *G. yesoense* population on Mt. Ibuki, which is distributed to the far south on the mountain, was grouped with var. *pseudopratense*. The other populations of var. *nipponicum*, which occur in high altitude areas on Honshu, were included in Clade B with full bootstrap support, except for the

populations from the Mt. Asama area, which were located at an intermediate position between var. *pseudopratense* and Clade B (var. *nipponicum*).

Genetic differentiation among populations and phylogenetic structure based on MIG-seq

Pairwise F_{ST} values were 0.639 for var. *yesoense* and var. *nipponicum*, 0.383 for var. *yesoense* and var. *pseudopratense*, and 0.513 for var. *nipponicum* and var. *pseudopratense*. Significant correlations were observed between geographic and genetic distances for var. *nipponicum* and var. *pseudopratense* ($R^2 = 0.111$, $P = 0.001$ and $R^2 = 0.331$, $P = 0.001$, respectively; Fig. S3-2), whereas no significant correlation was observed for var. *yesoense* ($R^2 = 0.077$ and $P = 0.063$; Fig. S3-2).

I assessed population structure with STRUCTURE analysis, using the same SNP dataset. The ΔK method indicated that the most appropriate K value was 2 (Fig. 3-6A). At $K = 2$, var. *yesoense*, var. *pseudopratense*, and the Mt. Ibuki and Mt. Asama populations were all assigned to the blue cluster, whereas populations of var. *nipponicum*, excluding the Mt. Asama populations, were assigned to the red cluster (Fig. 6B). The ΔK method indicated a second peak at $K = 3$, and the log-likelihood value plateaued at $K = 3$ when plotted against sequential K; thus, I also considered $K = 3$ to be an appropriate K value (Fig. 3-6A). At $K = 3$, populations of var. *yesoense* were assigned to the blue cluster, and var. *nipponicum* populations, with the exception of Mt. Asama plants, were assigned to the red cluster. Populations of var. *pseudopratense*, along with a large portion of the Mt. Ibuki and Mt. Asama populations, were assigned to the green cluster (Fig. 3-6B). STRUCTURE analysis indicated that the three varieties (var. *yesoense*, var. *nipponicum*, and var. *pseudopratense*) diverged from one another, whereas the Mt. Ibuki and Mt. Asama populations may have originated from admixture between var. *pseudopratense* and var. *nipponicum*.

Demographic analysis

Model selection for the first step of the ABC analysis indicated that Scenario 1-2 was associated with the highest posterior probability (0.586) of the five candidate scenarios. The branching pattern of Scenario 1-2 was consistent with that inferred from the chloroplast DNA phylogenetic tree: during the first split, var. *nipponicum* was derived from ancestral populations, which then gave rise to var. *yesoense* and var. *pseudopratense*. Scenario 1-5 was also associated with a moderate posterior probability of 0.229; this scenario assumed that var. *pseudopratense* originated from hybridization between var. *yesoense* and var. *nipponicum*. The second step of the analysis, which focused on five populations, was performed based on the results of the first-step simulations. Of the nine divergence scenarios, Scenario 2-4, which stemmed from Scenario 1-2, was associated with the highest posterior probability (0.3390; Fig. 3-2); this scenario posits that the Mt. Ibuki and Mt. Asama populations resulted from independent admixture events between var. *nipponicum* and var. *pseudopratense*. In addition, Scenarios 2-3, 2-8, and 2-9, which were similar to Scenario 2-4, were associated with equivalent posterior probabilities (0.2832, 0.1680, and 0.1736, respectively). Scenario 2-3 differs from Scenario 2-4 with respect to the timing of admixture between var. *nipponicum* and var. *pseudopratense*. Scenarios 2-8 and 2-9 assume that both var. *pseudopratense* and the Mt. Ibuki populations originated from hybridization between var. *yesoense* and var. *nipponicum*, and that the Mt. Asama population was derived from hybridization between var. *nipponicum* and var. *pseudopratense*. Based on Scenario 2-4, which was associated with the highest posterior probability, effective population size parameters were estimated based on median values (95% highest posterior densities) of the effective population size. These were 43,200 (20,800–80,400) for var. *yesoense*, 36,400 (16,100–71,000) for var. *nipponicum*, 37,600 (18,800–63,800) for var.

pseudopratense, 5560 (1610–14,600) for the Mt. Ibuki population, and 11,300 (4700–23,500) for the Mt. Asama population (Table 3-4). The median divergence and admixture times were 3160 (890–7820), 6900 (2850–15,200), 52,400 (27,100–87,600) and 220,000 (159,000–294,000) generations ago for $t1$, $t2$, $t3$ and $t4$. respectively (Table 3-4). Assuming a generation time of 3–5 years, the divergence times of $t4$ and $t3$ corresponded to 660,000–1,100,000 and 157,200–262,000 years ago, respectively. Admixture times of $t2$ and $t1$ were 20,700–34,500 and 9480–15,800 years ago, respectively.

Morphological observations

Comparisons of genetic differentiation and sepal pubescence implied that there is a strong relationship between genetic group and pubescence (Fig. 3-7). Sparsely pubescent populations were assigned to Cluster 2 (Cl2), which is the var. *nipponicum* cluster. Populations with moderate pubescence were assigned to Cluster 3 (Cl3), which is the var. *pseudopratense* cluster, and populations exhibiting dense sepal pubescence were assigned to Cluster 1 (Cl1), which is the var. *yesoense* cluster. Sepal hair density was moderate in the Mt. Ibuki and Mt. Asama populations (morphological score: 1.5–1.8), and their genetic structure most closely fit that of Cluster 3.

3-4. Discussion

Phylogenetic history of G. yesoense *varieties*

In contrast to Wakasugi et al. (2017), who found no evidence of genetic differences among *G. yesoense* varieties, my genome-wide markers consistently indicated that the three varieties studied here are distinct and independent evolutionary units. In particular, var. *nipponicum*, which is adapted to high-altitude mountain environments, diverged early in the evolution of the species complex, and its branch length was longer than those of clades containing other varieties (Fig. 3-3). On the other hand, whereas the phylogenetic tree based on genome-wide SNPs indicated that the other varieties (var. *yesoense* and var. *pseudopratense*) diverged into different phylogenetic groups, the chloroplast phylogenetic tree showed that they are grouped within a subclade with almost no genetic substitutions between them. Generally speaking, mutation rates are higher in the nuclear genome than in the chloroplast genome (Wolfe et al. 1987); thus, I conclude that the divergence between var. *yesoense* and var. *pseudopratense* was too recent for substitutions to have accumulated independently in their chloroplast genomes. Divergence between var. *nipponicum* and the other varieties was estimated to have occurred 0.25–0.75 million years ago. Other studies have demonstrated that most Japanese alpine plants diverged into southern and northern lineages in central Honshu and in regions including north-central Honshu and Hokkaido (Fujii et al. 1997, 1999; Ikeda et al. 2006, 2008a, 2008b, 2009a, 2014b; Ikeda & Setoguchi 2013a, 2013b). The divergence times for *Cardamine nipponica* and *Cassiope lycopodioides* were estimated to be ca. 0.11–0.2 million years ago (Ikeda et al. 2009a, 2014b), and those of other species to be during the late Pleistocene (Ikeda & Setoguchi 2013a; Fujii et al. 2008). The reported time estimates are roughly concordant with the time of divergence between *G. yesoense* var. *nipponicum* and the other varieties.

Model selection in the ABC analysis supported the following histories of these relict populations: 1) past north–south fragmentation was followed by *in situ* persistence in refugia, and 2) re-immigration from northern regions into southern refugial areas. As with other Japanese alpine plants, *G. yesoense* diverged into two major lineages in the middle Pleistocene, with the northern lineage represented by var. *yesoense* and var. *pseudopratense*, and the southern lineage by var. *nipponicum* (Fig. 3-2 and Fig. S3-1). The divergence time estimated from the ABC analysis coincided with the timing of the Mid-Brunhes Event, ca. 0.4 million years ago, which was the warmest interglacial period in the last 0.8 million years (Jansen et al. 1986; Yin & Berger 2010; Tarasov et al., 2011). Therefore, if I assume that climate warming drove the isolation of northern and southern populations, it follows that the ancestral population of var. *nipponicum* migrated southward to central Honshu during one or more of the previous glacial periods, where it persisted in high altitude mountain areas and their periphery.

Multiple waves of immigration shaped the complex genetic structure in the southern refugial region

The Bayesian clustering analysis of genome-wide markers indicated that the central Honshu populations exhibit complex genetic structures; i.e., that some populations are genetically closer to var. *pseudopratense*, which is primarily distributed in northern Japan, than to the surrounding var. *nipponicum* populations. The populations on Mt. Ibuki and Mt. Asama were noted to have sepals that are more densely hairy than is typical for var. *nipponicum* (Hara 1948); thus, they resemble var. *pseudopratense* more closely. However, hair density can be affected by local environmental conditions (Tsukaya & Tsuge 2001) and may not be reflective of phylogenetic origin if parallel variations in this trait were observed in different evolutionary clades (Cosacov et al. 2009). In this study, molecular and morphological data supported the

hypothesis that both the Mt. Ibuki and Mt. Asama populations are relics of the northern var. *pseudopratense* (Fig. 3-7).

Where refugial populations are divided into several smaller populations, long-term isolation can lead to genetic differentiation resulting from drift (Bennett & Provan 2008; Marr et al., 2013). My data, however, did not support the hypothesis that refugial populations of the southern var. *nipponicum* experienced genetic drift, because the genetic structure of central Honshu plants was homogeneous aside from the Mt. Ibuki and Mt. Asama populations (Fig. 3-6B). In addition, the central Honshu populations exhibited higher genetic diversity compared to the northern varieties (Fig. 3-4). Thus, I conclude that the complex genetic structure of the populations in central Honshu is unlikely to have arisen due to range fragmentation in refugial areas, and that the entire mountain area in central Honshu comprises a single interglacial refugium.

Furthermore, my data suggest that multiple waves of immigration and subsequent admixture increased the genetic complexity among refugial populations (Gabrielsen et al. 1997; Tollefsrud et al. 1998; Abbott et al. 2000; Skrede et al. 2006; Allen et al. 2012). The ABC analysis suggested that the Mt. Asama and Mt. Ibuki populations originated from the admixture of var. *nipponicum* and var. *pseudopratense* during the last glacial period (Fig. 3-2, Table 3-4). The phylogenetic tree, based on genome-wide SNPs, indicate that the Mt. Ibuki and Mt. Asama plants were similar to those of var. *pseudopratense*. In particular, the Mt. Asama populations were intermediately located between the var. *nipponicum* and var. *pseudopratense* clades (Fig. 3-5), suggesting that these populations originated from hybridization between var. *nipponiucm* and var. *pseudopratense*. During the last glacial maximum, boreal vegetation dominated central and northern Honshu as a result of significant temperature decreases (Takahara & Kitagawa 2000; Harrison et al. 2001). It is thus likely that *G. yesoense* (i.e., var. *pseudopratense*) migrated

to central Honshu from areas further north, and that secondary contact occurred between refugial populations of var. *nipponicum* and the re-immigrating populations of var. *pseudopratense*. Conversely, the Mt. Ibuki population was grouped with var. *pseudopratense*, and the third-most supported scenario according to ABC analysis suggested that this population diverged directly from var. *pseudopratense* during the last glacial period (Fig. 3-2). An abrupt decrease in temperate Japanese cedar (*Cryptomeria japonica*) was observed around Mt. Ibuki at 28,000 years BP, along with an increase in boreal coniferous trees (e.g. *Tsuga*, *Pinus*, and *Betula*), implying climate cooling (Takahara & Kitagawa 2000). The glacial climate around Mt. Ibuki is assumed to have been suitable for var. *pseudopratense* during the last glacial period, thus var. *pseudopratense* had suitable habitats in this area. Although the climate would also have been suitable for alpine populations of var. *nipponicum*, Mt. Ibuki's geographic isolation from other peaks may have hindered the migration of this variety from other locations, resulting in higher genetic purity in the rear-edge population of var. *pseudopratense*. There are two processes by which secondary contact between var. *pseudopratense* and var. *nipponicum* might have occurred: 1) southward migration of var. *pseudopratense* to central Honshu and the persistence of var. *nipponicum* at low-altitude areas during glacial periods, resulting in contact on the west coast of Japan, and 2) southward migration of var. *pseudopratense* to central Honshu during glacial periods and migration from the coastline to high-altitude mountain areas during interglacial periods, resulting in secondary contact between var. *nipponicum* and var. *pseudopratense* in mountain or alpine areas. However, I cannot be certain which of these hypotheses is correct.

My demographic result was based on only ABC analysis. Because ABC analysis originate in just approximate likelihood, the estimations may be underestimated. Demographic history of Japanese alpine plants was elucidated by using Isolation with Migration Model (e.g.

Hata et al. 2017; Ikeda et al. 2018). This method originates in the Bayesian Markov chain Monte Carlo (MCMC), and divergence time, gene flow and effective population size are estimated by using this model (Hey & Nielsen 2007). More detailed demographic analysis of *G. yesoense* using Isolation with Migration Model need for elucidating its demographic history. In addition, although *G. yesoense* species complex are morphologically distinguishable from one another with two identification keys (dense of hairs on the sepals and deepness of incised leaves), the later was not verified in this study. Wakasugi et al. (2017) suggested that identification keys of *G. yesoense* overlapped interspecies. Therefore, not only hairs on the sepals but also other identification (i.e. deepness of incised leaves) need to be verified and compared with genetic differentiation (Q-value from STRUCTURE analysis) in the future.

Conclusion

Other studies have explored historical factors that have shaped the complex genetic structure of Japanese alpine plants occurring in southern refugia. I found that the Mt. Ibuki and Mt. Asama populations, which are morphologically similar to the northern lineage of var. *pseudopratense*, do not represent independent evolution or environmental adaptation. Rather, they are likely true relict populations of var. *pseudopratense*. My results demonstrate that re-immigration into refugial populations in central Honshu, along with subsequent hybridization, added substantial complexity to the genetic structure of refugial populations. Although previous phylogeographic studies of alpine plants have found that the southern high mountains in central Honshu are stable long-term refugia, my genetic analysis suggested that re-immigration by a northern lineage further increased the phylogenetic diversity of *G. yesoense* in this region. Studies of other alpine species should be undertaken to determine whether multiple immigrations to central Honshu have occurred throughout the glacial-interglacial climate cycles

of the late Pleistocene.

Table 3-1. Distinguishing characters of the three *Geranium yesoense* varieties, and number of populations and samples used in genetic analyses.

Variety	Pedicel hairs	Sepal hairs	Leaf incision	Habitat	Populations[*1]	Samples[*2]	Samples[*3]
var. *yesoense*	spreading	dense	deep	Coastal grasslands	15	1	54
var. *nipponicum*	retrorse/sparse spreading	appressed/spreading	shallow	High-altitude areas	16	1	46
var. *pseudopratense*	spreading	sparse	-	Coastal grasslands	11	1	40

Note: [*1]The number of populations used in multiplexed inter-simple sequence repeat genotyping-by-sequencing (MIG-seq) analysis; [*2]The number of samples used in complete chloroplast genome sequencing; [*3]The number of samples used in MIG-seq analysis

Table 3-2. Contig lengths (number of base pairs).

Taxon name/population	Locus						
	cemA	*ccsA*	*rpoA*	*psbK*	*psbZ*	*ndhJ*	*rbcL*
var. *yesoense*	17,287	24,206	10,647	8248	7113	17,302	10,448
var. *nipponicum*	18,339	25,195	10,195	10,429	7786	17,302	19,573
var. *pseudopratense*	17,312	24,191	10,865	8765	7113	17,108	6359
Mt. Ibuki	17,328	24,642	11,827	8253	7113	17,282	10,475
G. soboliferum var. *hakusanense*	15,567	27,315	10,420	12,329	7422	18,012	7103

Table 3-3. Genetic diversity of *G. yesoense* varieties.

Variety	No.	Population	Temp.	Precip.	Alt.	N	H_o	H_e	F_{IS}	Ar
	1	Yubetsu	6.3	855	2	4	0.033	0.020	−0.620	1.026
	2	Abashiri	6.3	856	3	4	0.059	0.052	−0.146	1.063
	3	Syari	6.0	912	14	4	0.071	0.053	−0.310	1.065
	4	Asahikawa	6.3	1149	170	4	0.054	0.044	−0.180	1.051
	5	Nakashibetsu	5.0	1056	136	4	0.065	0.045	−0.385	1.055
	6	Ochiishi	6.0	1082	30	4	0.083	0.067	−0.208	1.082
	7	Oboro	5.8	1106	28	4	0.043	0.034	−0.211	1.041
var. *yesoense*	8	Shiranuka	5.9	1044	6	4	0.060	0.045	−0.263	1.055
	9	Onbetsu	5.6	1010	38	2	0.077	0.057	−0.352	1.090
	10	Toyokoro	5.6	1027	15	3	0.059	0.053	−0.099	1.071
	11	Tomakomai	7.3	1102	3	4	0.068	0.064	−0.073	1.079
	12	Bansei	6.0	1130	10	3	0.033	0.033	0.018	1.049
	13	Taiki	6.2	1172	11	2	0.050	0.042	−0.205	1.072
	14	Ogawara	9.9	1198	7	3	0.023	0.018	−0.286	1.023
	15	Tanesashi	10.0	1136	48	5	0.045	0.039	−0.117	1.049
	16	Mt. Mahiru	7.4	1669	677	3	0.059	0.056	−0.056	1.090
	17	Mt. Yakeishi	5.6	1715	1045	4	0.071	0.096	0.239	1.084
	18	Mt. Gassan	3.8	1974	1483	4	0.062	0.069	0.095	1.095
	19	Mt. Akakuzure	7.6	1741	918	1	0.039	0.020	−1.000	1.065
	20	Mt. Amakazari	5.5	1737	1441	4	0.101	0.090	−0.137	1.124
	21	Mt. Korenge	0.7	2178	2266	4	0.058	0.070	0.132	1.123
	22	Senjogahara	5.8	2183	1417	2	0.064	0.073	0.091	1.090
	23	Nozori	5.7	1735	1418	4	0.084	0.069	−0.196	1.108
var. *nipponicum*	24	Mt. Tate	0.0	2276	2399	2	0.063	0.065	0.009	1.131
	25	Mt. Kasagatake	4.1	1956	1752	1	0.039	0.020	−1.000	1.059
	26	Mt. Nishihodaka	3.4	2028	1871	2	0.032	0.029	−0.148	1.056
	27	Yahimagahara	4.8	1850	1683	3	0.045	0.062	0.202	1.102
	28	Mt. Sannomine	6.2	2202	1359	2	0.057	0.054	−0.071	1.055
	29	Mugikusa	3.6	2006	1894	4	0.095	0.082	−0.162	1.121
	30	Mt. Kita	3.3	2199	2001	2	0.039	0.043	0.024	1.083
	31	Mt. Hijiri	2.2	2460	2232	4	0.042	0.045	0.048	1.071
var. *pseudopratense*	32	Rishiri	6.5	1145	11	4	0.034	0.038	0.094	1.053
	33	Kamui	7.7	1281	135	3	0.079	0.073	−0.118	1.103

No.	Name	Temp.	Precip.	Alt.	N	H_o	H_e	F_{IS}	Ar
34	Shiribetsu	8.4	1241	4	4	0.076	0.077	−0.015	1.104
35	Kariba	7.0	1254	319	4	0.076	0.073	−0.034	1.099
36	Setana	8.9	1194	49	4	0.067	0.078	0.107	1.102
37	Esashi	9.1	1250	58	4	0.076	0.064	−0.186	1.080
38	Hakodate	9.2	1171	10	4	0.083	0.075	−0.119	1.095
39	Aomori	10.6	1461	54	3	0.074	0.089	0.131	1.083
40	Nyudozaki	11.0	1647	26	4	0.066	0.067	−0.004	1.090
41	Toga	11.0	1647	72	3	0.051	0.042	−0.203	1.068
42	Mt. Kanpu	10.8	1652	102	3	0.063	0.053	−0.185	1.081
43	Mt. Ibuki	10.9	2118	660	10	0.065	0.114	0.391	1.134
44	Mt. Asama_1	4.3	1831	1696	4	0.106	0.126	0.132	1.126
45	Mt. Asama_2	4.3	1831	1696	4	0.065	0.077	0.105	1.099

Note: No., population number in Fig. 1; Temp., mean annual temperature (°C); Precip., mean annual precipitation (mm); Alt., altitude of each population (m); N, number of individuals in each population minus the individuals excluded by filtering; H_o, observed heterozygosity; H_e, expected heterozygosity; F_{IS}, fixation index; Ar, allelic richness.

Table 3-4. Estimates of demographic parameters associated with Scenario 2-4 from demographic analysis.

	Mean	Median	Mode	q050	q950
N1 (var. *yesoense*)	45,600	43,200	40,400	20,800	80,400
N2 (var. *nipponicum*)	39,100	36,400	28,600	16,100	71,000
N3 (var. *pseudopratens*e)	38,800	37,600	35,700	18,800	63,800
N4 (Mt. Ibuki)	6470	5560	4390	1610	14,600
N5 (Mt. Asama)	12,400	11,300	9230	4700	23,500
t_1	3590	3160	2450	890	7820
t_2	7720	6900	5400	2850	15,200
t_3	53,900	52,400	47,200	27,100	87,600
t_4	223,000	220,000	219,000	159,000	294,000
rb	0.245	0.236	0.222	0.0954	0.424
rc	0.157	0.152	0.146	0.0648	0.266

Note: Nx, effective populations size for each variety or population x; t_X, divergence or admixture time, with generation time as the unit. t_4 is the divergence time between var. *nipponicum* (N2) and other varieties, t_3 is the divergence time between var. *yesoesne* and var. *pseudopratense*, t_2 and t_1 are the admixture time between var. *nipponicum* and var. *pseudopratense*; rb and rc, admixture rates during admixture events; q050 and q095 show 95% highest posterior densities.
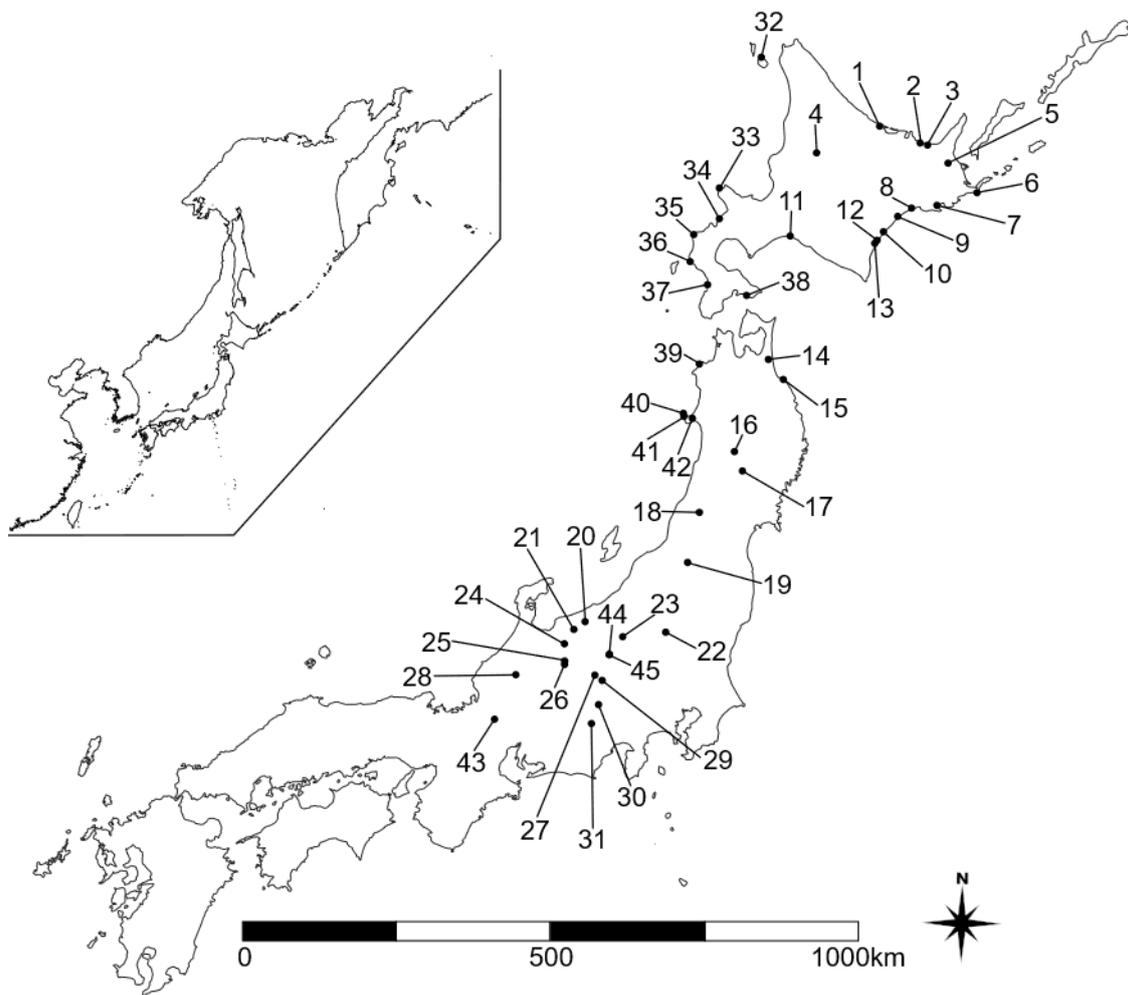
Fig. 3-1. Geographic locations of the sample populations. Numbers refer to sample populations, as displayed in Table 3-3.
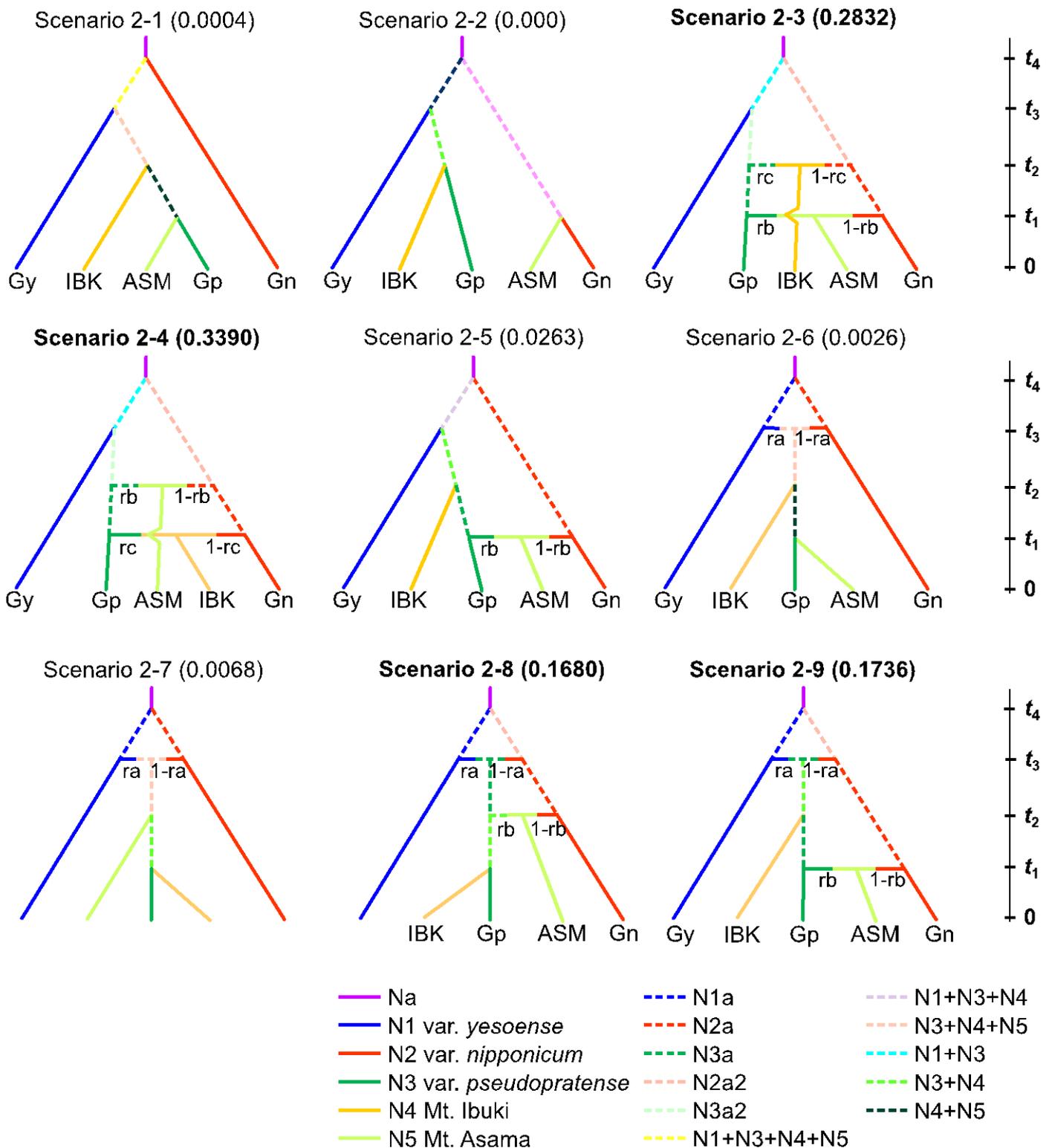
Fig. 3-2. Nine divergence scenarios considered at the second step of the approximate Bayesian computation analysis. Nx, effective population size (x, arbitrary number of populations or ancestral populations); t_x, branching time for each population or time of merging between two populations; "0," present time; ra, rb, and rc, admixture rates. The numbers in parentheses represent posterior probabilities from the simulations. Gy, var. *yesoense*; Gn, var. *nipponicum*; Gp, var. *pseudopratense*; IBK, Mt. Ibuki; ASM, Mt. Asama.
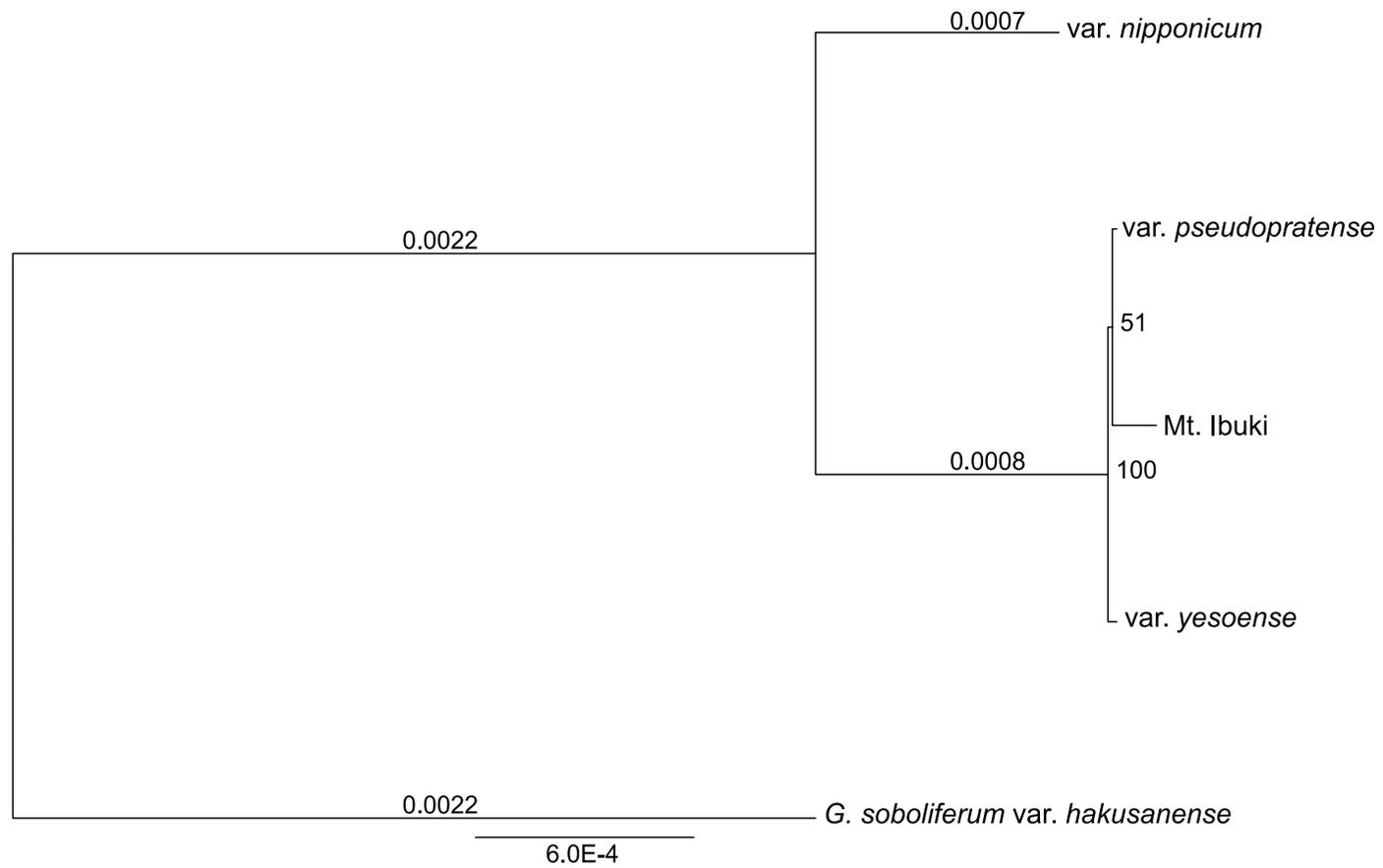
Fig. 3-3. Maximum likelihood (ML) phylogenetic tree based on the chloroplast genome. Branch length is shown along the branches, and bootstrap values are displayed at the nodes.
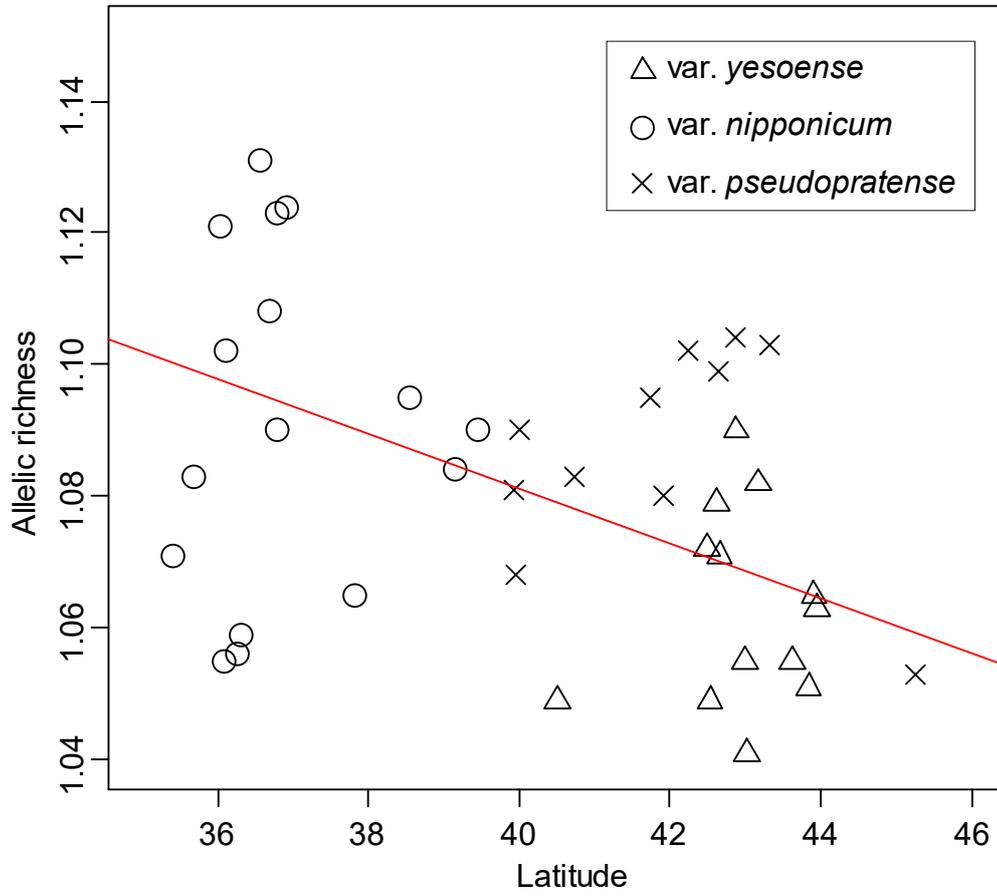
Fig. 3-4. Relationships between allelic richness (*Ar*) and latitude in *G. yesoense* complex. The x-axis represents latitude and the y-axis represents *Ar*. Regression line among all varieties is shown with solid line.
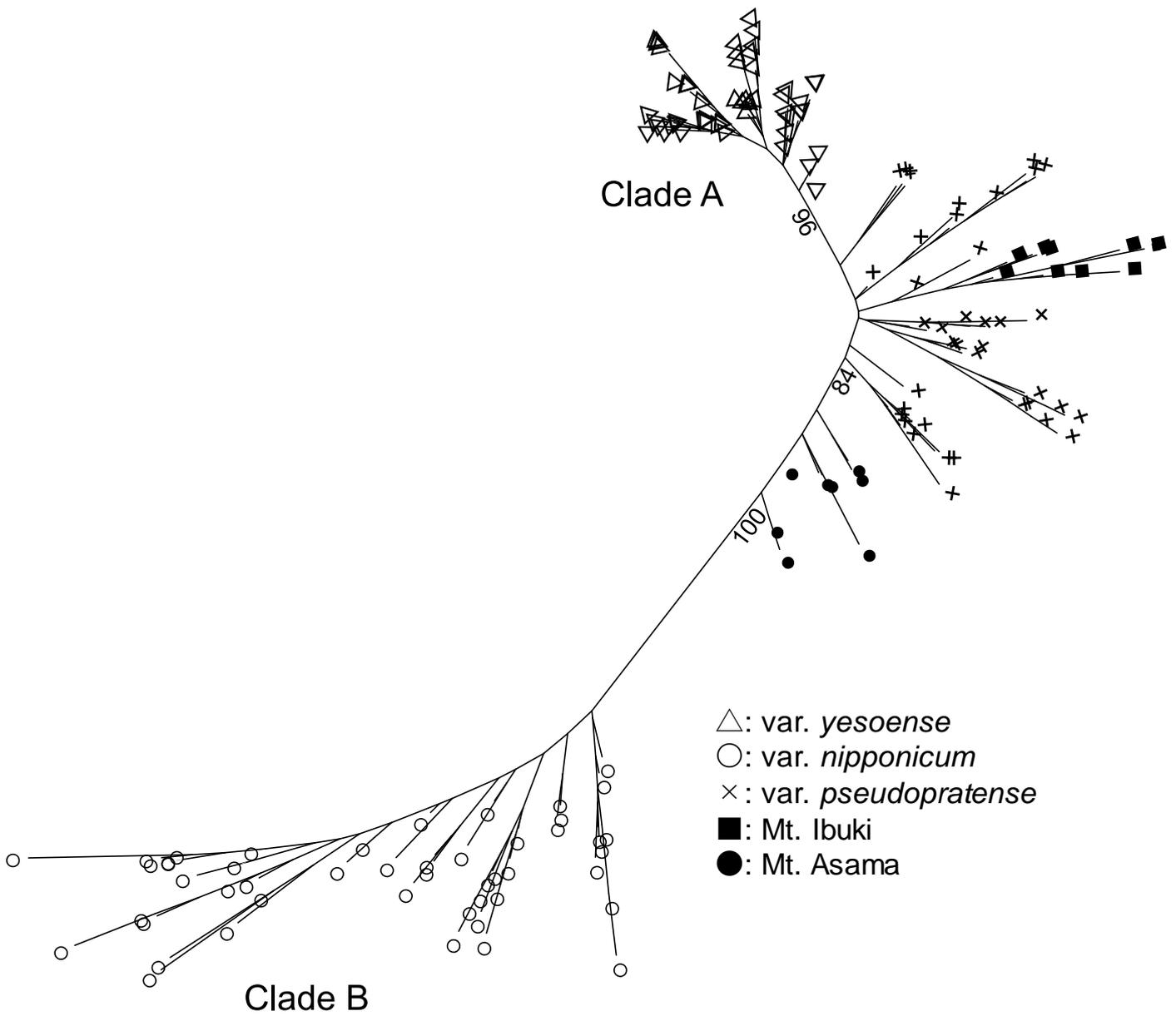
Fig. 3-5. ML phylogenetic tree based on genome-wide single nucleotide polymorphisms (SNPs) obtained using multiplexed inter-simple sequence repeat genotyping-by-sequencing. Numbers at the major nodes signify bootstrap values.
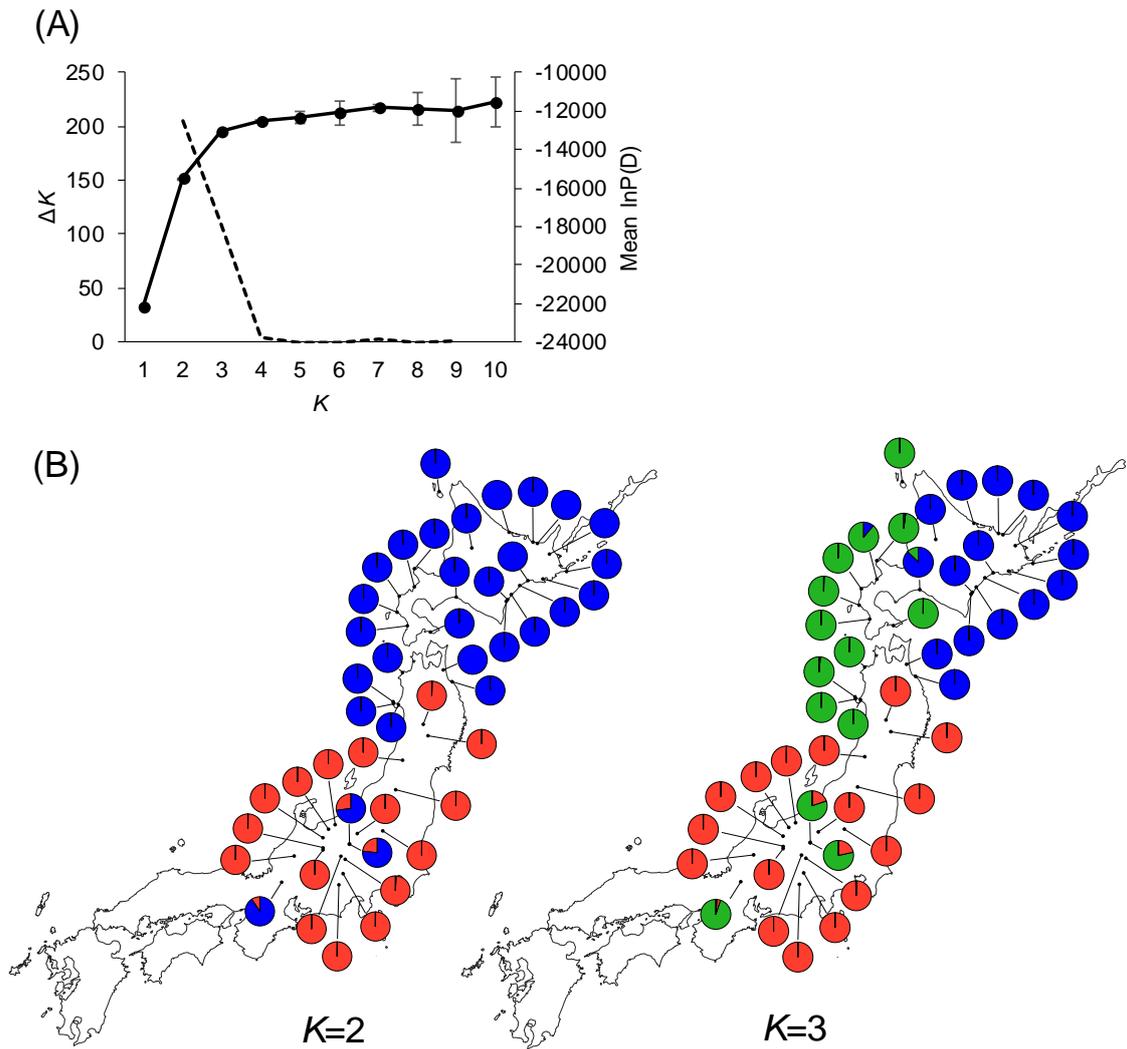
Fig. 3-6. Results of STRUCTURE analysis. (a) ΔK (dotted line) and mean estimated ln P(D) (solid line). (b) Pie charts for *K* = 2 and *K* = 3.
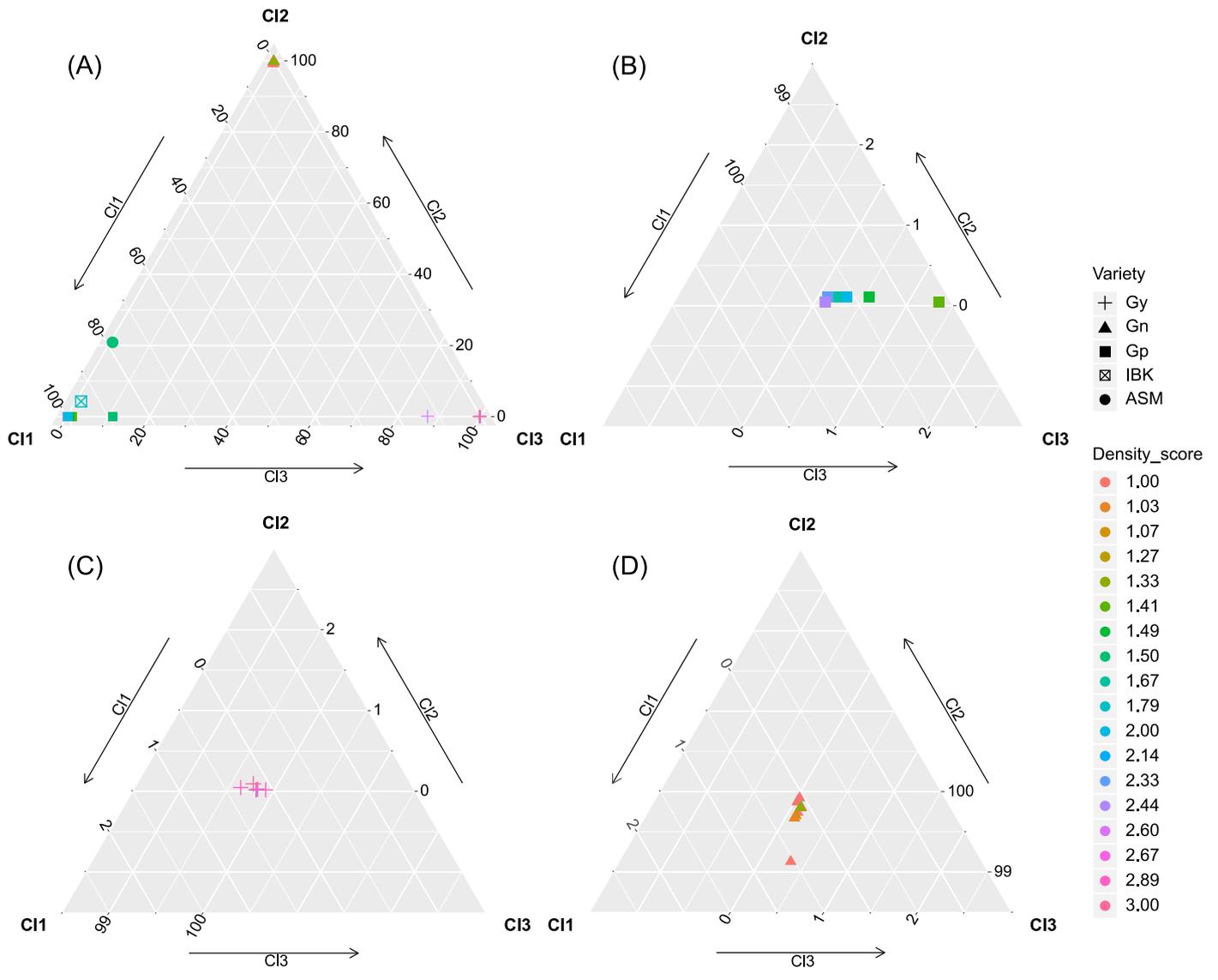
Fig. 3-7. Relationship between genetic differentiation and density of sepal pubescence. (a) Ternary plot for all populations; (b) enlargement of the left-hand corner of (a); (c) enlargement of the top of the (a); (d) enlargement of the right-hand corner of (a). Clx, cluster name from STRUCTURE analysis (x, arbitrary number of clusters). Axes represent Q-values derived from STRUCTURE analysis, and "Density_score" indicates pubescence scores. Gy, var. *yesoense*; Gn, var. *nipponicum*; Gp, var. *pseudopratense*; IBK, Mt. Ibuki; ASM, Mt. Asama.

Chapter 4 Phylogeographic study of alpine plant *G. erianthum* DC.

4-1. Introduction

During the glacial periods in late Pleistocene, global seawater level depression connected the Japanese archipelago to the Korean Peninsula, Sakhalin and Kuril Islands through the land bridges (Ota & Machida 1987; Tsukada 1988; Dobson 1994; Korotky et al. 1997; Razjigaeva et al. 2008; McKay 2012; Ye et al. 2015). Alpine plant species are thought to have migrated via these northern bridges between the Japan and the surrounding regions, which would have made Japanese alpine flora more diverse. Currently, the Japanese alpine plants are classified into six elements: Cosmopolite element, Circumpolar element, Asian element, Pacific Ocean element, Low altitude mountain element and Endemic element (Hultén 1937; Tatewaki 1974; Shimizu 1982, 1983). Pacific Ocean element is further classified into two sub-elements of North Pacific Ocean element and both sides Pacific coast element (Shimizu 1982, 1983). The former is distributed to encircle the North Pacific Ocean, while the distributions of the both sides Pacific coast elements are fragmented between Eastern Siberia and North America, lacking around Bering Strait. The former, North Pacific Ocean element is composed of 71 species which are distributed from Northeastern Asia to Western North America (Shimizu 1982, 1983), e.g., *Phyllodoce aleutica* (Ericaceae), *Lagotis glauca* (Plantaginaceae) and *Pedicularis japonica* (Orobanchaceae). The center of their distributions is thought to be located around Beringia (e.g. the Alaska Peninsula and Eastern Siberia), therefore the Japanese archipelago is located at one southern edge of the elements' distributions. In glacial periods, alpine plants are generally thought to have expanded their distributions to southern regions or persisted at ice-free zones in high-latitude, known as nunatak and its peripheral area, while high-altitude mountains in southern regions were to the refugia for them during interglacial periods (Holderegger & Thiel-Egenter 2009; Stewart et al. 2010). Following this global model of

distribution changes in response to climate change, the present distribution of Pacific Ocean element is thought to be formed by migrations from northern regions (e.g. the Alaska Peninsula) into southern area during glacial periods, and later isolated at high-altitude mountain areas in southern regions (e.g. Central Honshu in the Japanese archipelago) during interglacial periods or post-glacial period.

Molecular studies have revealed the distribution dynamics of the alpine plant elements in Japan. Fujii et al. (1997, 1999) revealed that Pacific Ocean element of *Pedicularis japonica* (Orobanchaceae) and *Primula cuneifolia* (Primulaceae) immigrated into Japanese archipelago at least two times, and the high altitudinal mountains of Central Honshu (reaching 3,000m a.s.l) have been an important interglacial refugium. Hata et al. (2017) revealed that a Pacific Ocean element of *Therorhodion camtschaticum* (Ericaceae) harboured genetic similarity between the Aleutian Islands and southern Kamchatka, and suggested that *T. camtschaticum* had once a wider range during Last Glacial Maximum (LGM), and the Aleutian Islands were the source of colonization into southern Kamchatka. More recently, Ikeda et al. (2018) found genetic evidence against the long-believed notion that the vast ice-free area was spreading in Beringia as one of the glacial refugia (Hultén 1937), from which many Pacific Ocean element plants have expanded to Japan (Koidzumi 1919). In their study, Japanese populations of *Phyllodoce aleutica* (Ericaceae) retained significantly high genetic diversity compared to eastern Beringia populations, and the Beringia populations were inferred to have originated through post-glacial colonization from more southerly located East Asia (Ikeda et al. 2018). Although this population dynamics scenario is contrary to the scenario of Beringia populations as the source, it is plausible as living organisms might have changed their distribution in different ways and multiple times during repeated glacial/interglacial periods, which can result in the similar distribution pattern of Pacific Ocean elements.

Geranium erianthum DC. (Geraniaceae) is an herbaceous plant and distributed in Eastern Asia, Eastern Siberia, Kurils, Alaska and western North America (British Columbia) (Akiyama 2001; Xu & Aedo 2008). It is classified as North Pacific Ocean element in Japan (Shimizu 1982, 1983). Because of its wide distribution range and poor seed dispersal ability via barochory (Aedo 2001), *G. erianthum* is a suitable case to elucidate the historical population dynamics of the Pacific Ocean elements. The previous studies suggested that the Pacific Ocean elements have two major refugia as the source of enlarging their distributions. First candidate area is Beringia (e.g. eastern Siberia, Kamchatka, the Aleutian Islands and the Alaska Peninsula). During glacial periods, it is thought that many plants survived at scatter ice-free areas (e.g. Abbott et al. 2000; Skrede et al. 2009). As with these plants, *G. erianthum* may have survived at ice-free area in Beringia during glacial periods, and later, gradually immigrated into western region (e.g. Siberia and East Asia) and eastern region (Norther America). Second candidate area is East Asia (e.g. continental East Asia and the Japanese archipelago). During interglacial periods, alpine plants which had migrated southward during glacials were thought to have survived at high-altitude area in temperate zone and peripheral grassy places (e.g. Winkworth et al. 2005; Birks 2008; Holderegger & Thiel-Egenter 2009). If *G. erianthum* survived the suitable areas in East Asia during interglacial periods, as with *P. aleutica*, *G. erianthum* would have also been expanded its distributions to Beringia and eastern areas. In addition to these candidates, it is also known that cryptic refugia existed in British Columbia, and its refugial plant populations repeatedly migrated from different source regions over multiple glacial cycles (Allen el al. 2012). Likewise, *G. erianthum* might have persisted at cryptic refugia in British Columbia during glacial periods, and later, migrated to other regions during the climate cycles, then to form current widespread distribution. In this way, there are at least three candidate regions for the origin and refugia of *G. erianthum*.

In order to reveal which regions (Beringia, East Asia and British Columbia) had been important refugia or distributional source of *G. erianthum*, phylogenetic and population genetic studies are required. Thus far, only one molecular study of *G. erianthum* was reported by Wakasugi et al. (2017), but their phylogenetic analysis failed to detect genetic differences even between *G. erianthum* and its related species of *G. onoei* Franch. et Sav., due to lack of genetic polymorphisms in their genetic markers. Recently, it is becoming much easier to obtain genome-wide genetic polymorphisms of non-model organisms owing to implement of the sequence technology. Especially, genotyping-by-sequencing (restriction site-associated DNA (RAD) markers: Baird et al. 2008, multiplexed ISSR genotyping by sequencing: Suyama & Matsuki 2015) enables me to get hundreds to hundreds of thousands of single nucleotide polymorphisms (SNP) data, and have been applied to population genetic studies (e.g. Takahashi et al. 2016; Watanabe et al. 2018; Hirano et al. 2019). When combined with model-based inference, such as Approximate Bayesian Computation, these SNPs data sets becomes of high power to reconstruct the population dynamics in relation to historical distributional changes (e.g. Jeffries et al. 2016; Sun et al. 2016; Clark et al. 2019).

Overall, the aim of this study is to reveal which regions had been important refugia or harboured source populations for subsequent range expansion of the widespread Pacific Ocean element of *G. erianthum*. Comparing phylogenetic relationships, genetic differentiation, genetic structure and population dynamics among three candidate refugial/source regions (Beringia, East Asia and British Columbia), I aimed to determine the source and/or refugia of *G. erianthum*, to understand the historical range dynamics of alpine plants which are now widely distributed in northern Pacific coasts. To this end, I evaluated the followings in this study: (1) species' diversity of *G. erianthum* and its related species and their phylogenetic relationships, (2) characterizing diversity and differentiation of *G. erianthum* regional populations to detect the

refugial populations which are expected to retain high genetic diversity, and with longer geographic distances, genetic differentiation among populations increases due to the intensified effects of genetic drift (Wright 1943), and (3) inferring demographic scenarios of *G. erianthum* using coalescent simulations based on Approximate Bayesian Computation, to objectively discuss its historical distribution changes.

4-2. Materials and Methods

Plant sampling

From July 2017 to August 2018, I collected mature leaves of *G. erianthum* and *G. erianthum* f. *glabriusculum* Horie from 15 populations (55 samples). Additionally, 15 samples from Beringia populations (Kamchatka, the Aleutian Islands, the Alaska Peninsula) and British Columbia population were collected. A total of 70 individuals were collected, and they were immediately dried and preserved at room temperature using silica gel and preserved at dark room.

Phylogenetic analysis of chloroplast whole genome sequences

G. erianthum from Obihiro city (Hokkaido Pref.), Kamchatka (Russia), British Columbia (North America) and f. *glabriusculum* from Asahikawa city (Hokkaido Pref.) were chosen as the samples for chloroplast genome sequencing. In addition, the related taxonomic groups of *G. onoei* var. *onoei* f. *onoei*, *G. onoei* var. *onoei* f. *alpinum* Yonek. and *G. onoei* var. *onoei* f. *yezoense* (H.Hara) Yonek. were included the phylogenetic analysis and *G. yesoense* var. *yesoense* was included as outgroup.

Prior to DNA extraction, I washed the fine leaf powder samples using a buffer consisting of 3% HEPES buffer (pH 8.0), 1% polyvinylpyrrolidone, 0.9% ascorbic acid, and 2% 2-mercaptoethanol. Genomic DNA was extracted using a modified cetyltrimethylammonium bromide method (Milligan 1992), and total DNA was measured using a Qubit 2.0 fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). Paired-end sequencing was performed using the BGISEQ-500 next-generation sequencing platform (BGI, Shenzhen, China). Low-quality reads were removed using Trimmomatic-0.39 (Bolger et al. 2014). When the average quality per four sequences dropped below 15, all subsequent sequences were removed

80

(SLIDINGWINDOW: 4:15). Sequences were also removed at the beginning or end of reads with quality < 20 (LEADING: 20 and TRAILING: 20). In addition, reads that were shorter than 30 bp were removed (MINLEN: 30). The trimmed sequence reads were assembled *de novo* using NOVOPlasty (Dierckxsens et al. 2017); the *G. incanum* (NC_030045.1) plastid sequence was used as a reference sequence, and each locus (*ccsA*, *cemA*, *rbcL*, *ndhJ*, *psbK*, *rpoA*, and *psbZ*) of *G. incanum* was designated as a seed sequence in the assembly analyses. Because the assembly did not yield continuous sequence of the chloroplasts of the samples, the contigs around the above locus were used in the phylogenetic analysis. Each coding sequence was aligned by MAFFT 7.310-1 (Katoh et al. 2002, 2013). Every gaps originated from inversions and/or deletions were eliminated and phylogenetic analysis based on nucleotide sequence was performed by RAxML 8.2.11 (Alexandros 2014); I adopted the nucleotide substitution model of GTRGAMMA, and the number of bootstrap was set 1,000, and later, calculate the bootstrap values of the best tree by using pgsumtree option of Phylogears2 (Tanabe 2008). The acquired Maximum Likelihood (ML) tree was visualized edited by using FigTree v1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/).

Divergence times among the three species were calculated using the formula: $T = k/2v$, where T = divergence time, k = evolutionary distance, and v = evolutionary rate. Branch length, which was estimated using RAxML, was used as the evolutionary distance (k), and $1.0–3.0 \times 10^{-9}$ was used as the evolutionary rate (Wolfe et al. 1987).

MIG-seq experiment

MIG-seq with a next-generation sequencing platform was used to identify genome-wide SNPs (Suyama & Matsuki 2015). Using this method, the sequence library comprises the ISSR loci, which is amplified by multiplex polymerase chain reaction (PCR).

After library size selection, next-generation sequencing was performed (Suyama & Matsuki 2015). During the first PCR step, I amplified ISSRs using the following primers: $(ACT)_4TG$, $(CTA)_4TG$, $(TTG)_4AC$, $(GTT)_4CC$, $(GTT)_4TC$, $(GTG)_4AC$, $(GT)_6TC$, $(TG)_6AC$, $(ACT)_4TG$, $(CTA)_4TG$, $(TTG)_4AC$, $(GTT)_4CC$, $(GTT)_4TC$, $(GTG)_4AC$, $(GT)_6TC$, and $(TG)_6AC$. At the second PCR step, common forward and reverse primers were used, including nine- and five-base indices, respectively, and an adapter sequence for Illumina sequencing. PCR products were then pooled, and size selection was carried out using the magnetic bead method (SPRIselect; Beckman Coulter, Brea, CA, USA), with a target size of 350–800 bp. Finally, the library was sequenced using the Illumina MiSeq platform (Illumina, San Diego, CA, USA) and a MiSeq Reagent kit v3 (Illumina).

SNP detection

The first 14 read-2 (reverse-read) bases and low-quality sequences of raw reads were removed using the FASTX-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/), with the minimum quality score set to 30 (-q 30). Reads comprising > 40% bases that failed to meet the minimum quality score were removed (-p 40). In addition, the primer sequence region of each read was removed using TagDust (Lassmann et al. 2009). After removal of unnecessary reads, the remaining reads were assembled with Stacks 1.48 (Catchen et al. 2013), and SNPs were called using the same program. In *ustacks*, the maximum distance (in nucleotides) option was set to 2 (-M 2) and the minimum depth option for creating the stack was set to 3 (-m 3), while the maximum distance option (-N) was set to 2. Catalogs were created using *cstacks*, and the genotypes of individuals were determined using *sstacks*. SNPs for phylogenetic and population analysis were exported using *Populations*, with the following settings: genotyping rate, which is the minimum percentage of individuals in a population, was set to 0.1 (-r 0.1); minor alleles (<

1% within population) were removed (--min-maf 0.01); and loci with unreasonable heterozygosity were removed (--max-obs-het 0.5). A total of 20,403 SNPs were detected. I filtered these using PLINK 1.9 (Purcell et al., 2007), removing individuals which had 80% missing data, and loci that were held by < 40% of individuals (--mind 0.8, --geno 0.4). Minor alleles (< 3%) were also removed (--maf 0.03). Finally, individuals that had more than 50% missing data were removed. These SNPs were used in phylogenetic and STRUCTURE analyses (Pritchard et al. 2000), and for calculating summary statistics. For demographic analysis, because at least one individual must host a genotype within a population, the locus of any genotype that was absent in all individuals was removed from the analysis.

Population analysis

To characterize each population, the following four summary statistics were calculated using GenAlex v.6.501 (Peakall & Smouse 2006) and hierfstat in R 3.6.1 (Goudet 2005; R Development Core Team, 2019): allelic richness (Ar), expected heterozygosity (H_e), observed heterozygosity (H_o) and fixation index (F_{IS}).

The F_{ST} index (Weir & Cockerham 1984) between populations were calculated using GENEPOP 4.7.2 (Raymond & Rousset 1995). I used the model-based Bayesian clustering program STRUCTURE 2.3.4 (Pritchard et al. 2000), which assigns individuals to K clusters, to infer genetic structures associated with the samples. Population structure was estimated for the overall sample, as well as for each taxon separately, with $K = 1–12$ under an admixture and correlated allele frequencies model (F model) (Falush et al. 2003). Simulation runs consisted of 100,000 Markov chain Monte Carlo simulations, following a burn-in period of 100,000 iterations. For each value of K, 20 runs were performed, and I used the ΔK method (Evanno et al. 2005) and the value of ln P(D) to determine the optimal value of K by maximizing ΔK or ln

P(D). If ln P(D) increased with the number of clusters, the value of K is defined as the value immediately before the value of ln P(D) levelled out.

Additionally, I evaluated the isolation by distance (IBD) of each variety to assess the relationship between geographic and genetic distance. The geographic distance of each population was calculated from their latitude/longitude, and $F_{ST}/(1 - F_{ST})$ was adopted as their genetic distance (Rousset 1997); the significance of correlations was then verified by the Mantel test using GenAlEx v.6.501.

Demographic analysis

Approximate Bayesian computation (ABC) is a powerful and flexible approach to estimating demographic and historical parameters and quantitatively comparing alternative scenarios (Bertorelle et al. 2010). To determine the most probable demographic scenario among alternatives with regard to *G.erianthum*, I adopted the ABC method implemented in DIYABC v2.1 (Cornuet et al. 2014).

To set up the population units of my ABC analysis, *G. erianthum* was divided into three geographic groups (Hokkaido, Rishiri–Rrebun islands and Beringia [Kamchatka, Aleutian Islands, Alaska and British Columbia]). However, the populations *G. erianthum* f. *glabriusculum* was excluded this analysis because STRUCTURE analysis suggested that these populations had unique genetic structure and differed from that of other Hokkaido populations.

I performed ABC analysis to select the most probable divergence scenario of *G. erianthum* which is distributed from Japanese archipelago to Beringia. The assumed divergence scenarios were set as follows: (1) Three populations diverged simultaneously, reflecting global climate change; (2) Hokkaido and Beringia population diverged at first, and later, Rishiri–Rebun and Hokkaido population diverged; (3) Hokkaido and Beringia population diverged at first, and

later, Rishiri–Rebun population and Beringia population diverged; (4) Hokkaido and Beringia population diverged at first, and later, they encountered secondly and Risiri–Rebun population originated in the hybrid between Hokkaido and Beringia population; (5) Rishiri–Rebun and Beringia population diverged at first, and later, they encountered secondly and Hokkaido population originated in the hybrid between Rishiri–Rebun and Beringia population. The five alternative scenarios were visualized in Fig. 4-1.

Next, I established a series of summary statistics: I adopted six summary statistics, which was mean of non-zero values and variance of non-zero values of Genetic diversity, F_{ST} distance and Nei's distance, respectively (Nei 1972, 1987). In addition, mean of non-zero values and variance of non-zero values of admixture estimates were adopted as admixture summary statistics among groups (Choisy et al. 2004). Five million simulations were performed for each scenario, and the most appropriate scenario was determined by comparing posterior probabilities. I also assessed the effects of the number of simulated data points closest to those observed, which were used to evaluate posterior probabilities using the varying portion of the closest data point (i.e. every 0.1% in 0.1–1.0% of the closest simulations). Moreover, the divergence time in relation to absolute time was calculated by multiplying the estimated time parameter in the number of generations by 3–5 years, which is appropriate as the generation time for this herbaceous species (The Alpine Garden Society of Tokyo 2004).

4-3. Results

Chloroplast genome phylogeny

As the result of *de novo* assembly, seven comparatively and stably long contigs were obtained at surrounding regions of the genes *cemA*, *ccsA*, *rpoA*, *ndhJ*, *psbZ*, *rbcL* and *psbK* for each species. The length of contigs surrounding at the gene *cemA* ranged 16,198–24,365 bp, *ccsA* 32,703–51,956 bp, *rpoA* 9,851–10,274 bp, *ndhJ* 16,492–18,991 bp, *psbZ* 7,749–8,860 bp, *rbcL* 6,669–7,825 bp and *psbK* 9,904–19,521 bp (Table 4-1). After aligning each contigs, each aligned contig was concatenated and phylogenetic analysis was performed using concatenated sequence (total length 66,179 bp).

The maximum likelihood (ML) tree suggested that *G. erianthum* branched off from the related species *Geranium onoei* Franch. et Sav. (Fig. 4-2). Because the evolutionary distance was 0.0002 (0.0001 + 0.0001), the divergence time between *G. erianthum* and *G. onoei* was estimated about 100,000 years ago. On the other hand, *G. onoei* var. *onoei* f. *yezoense* (H.Hara) Yonek. formed a clade with *G. erianthum*. The clade branched into two subclades. Subclade A comprised *G. erianthum* from Hokkaido, Kamchatka and British Columbia and was also supported high bootstrap value (=100) (Fig. 4-2). Subclade B comprised two forms of *G. erianthum* f. *glabriusculum* and *G. onoei* var. *onoei* f. *yezoense*.

Although the total read data in this study was 5Gb, I could not get chloroplast whole genome sequence. Geraniaceae chloroplast genome contains a high frequency of large repeats (>100bp) relative to other rosids, and the number of large repeats (>200bp) of the genus *Geranium* was largest in Geraniaceae (Guisinger et al. 2011). The length of sequence reads was 100bp and insert sequence was 300bp. If the large repeats (>100bp) occurred in *G. erianthum* chloroplast genome, the assembler could not form contigs which located at large repeat regions.

Nuclear genome phylogeny based on MIG-seq

The total and average number of reads obtained from MIG-seq were 19,755,164 and 135,309 reads respectively. The reads were processed by Stacks pipeline, which yielded 14,623 SNPs for 70 samples. After the filtering procedure by PLINK 1.9, 341 SNPs and 60 samples remained in the data matrix. Summary statistics for each variety and population were calculated and presented in Table 4-2. Population genetic diversity (H_e and H_o) of Hokkaido district was 0.145 and 0.054, respectively. The Ar varied from 1.056 to 1.109. The H_e and H_o of Rishiri and Rebun Island was 0.111 and 0.058, respectively. The Ar varied from 1.075 and 1.099. The H_e and H_o of Beringia and British Columbia was 0.144 and 0.054, respectively. The Ar varied from 1.054 to 1.129. There was no significant difference between Hokkaido and Beringia for the Ar values.

Genetic differentiation among populations and phylogenetic structure based on MIG-seq

Pairwise F_{ST} values of each population were summarized (Table 4-3), and the value of within Hokkaido varied from -0.026 (Takanegahara–Tokachi) to 0.548 (Sapporo–Hikatsu), within Beringia and British Columbia varied from -0.178 (British Columbia–Kachemak Bay) to 0.759 (Crater Lake–Aleutian). The F_{ST} value between Rishiri and Rebun island was 0.211. Significant correlations were observed between the geographic and genetic distances in Hokkaido district: $R^2 = 0.0965$, $P = 0.023$; (Fig. 4-3a), while no significant correlation was observed for Beringia: $R^2 = 0.0099$ and $P = 0.140$ (Fig. 4-3b).

Using the same SNP data set, STRUCTURE analysis was performed to detect population structure within the samples. The ΔK method indicated that the most appropriate K value was 2 (Fig. 4-4a). At $K = 2$, the populations of British Columbia and Alaska were assigned to the blue cluster, and the populations of Hokkaido were almost assigned to the red cluster (Fig.

4-4b). Rishiri, Rrebun, Kamchatka and Aleutian populations were admixed two clusters (Fig. 4-4b). At *K* = 3, the populations of British Columbia and Alaska were assigned to the blue cluster, and Rishiri and Rebun population were assigned to the green cluster (Fig. 4-4b). Kamchatka and Aleutian population were admixed by two clusters (Blue and Green cluster) (Fig. 4-4b). Hokkaido populations were also admixed by the two clusters (Red and Green cluster) (Fig. 4-4b).

Demographic analysis

Model selection for ABC analysis showed that scenario 1 had the highest posterior probability among the wide range of closest simulations (Fig. 4-1). Scenario 1 supposed the following: the ancestral populations of Hokkaido, Rishiri–Rebun island and Beringia diverged from same ancestral populations at the same timing (Fig. 4-1). In addition to scenario 1, scenario 3 had moderate posterior probability. Scenario 3 supposed that Hokkaido and Beringia populations diverged at first, and later, Rishiri–Rebun and Beringia population diverged each other. Based on selected scenario (scenario 1), the effective population size parameters were estimated as the median values (95% highest posterior densities) of effective population sizes, being 47,000 (20,200–78,100), 15,300 (6,460–37,600) and 36,800 (18,200–66,600) for N1 (Hokkaido), N2 (Rishiri–Rebun) and N3 (Beringia), respectively (Table 4-4a). The median value of the divergence time was 14,300 (7,590–22,000) generations ago for *t2* (Table 4-4a). If I assume a generation time of 3–5 years, the divergence time of *t2* was 42,900–71,500 years ago. On the other hand, based on scenario 3, the effective population size parameters were estimated as the median values (95% highest posterior densities) of effective population sizes, being 77,500 (52,600–93,800), 9,560 (2,880–34,100) and 31,500 (15,100–59,400) for N1 (Hokkaido), N2 (Rishiri–Rebun) and N3 (Beringia), respectively (Table 4-4b). The median values of the

divergence times were 15,200 (5,590–30,200) and 22,300 (12,400–34,900) generations ago for $t1$ and $t2$, respectively (Table 4-4b). If I assume a generation time of 3–5 years, the divergence time of $t2$ and $t1$ were 66,900–111,500 and 45, 600–76,000 years ago, respectively. Comparing scenario 1 and scenario 3, the difference among them is divergence timing of Rishiri–Rebun population. Although it is plausible that Rishiri–Rebun population diverged from Beringia population at different timing because Rihiri and Rebun Island located between Asian continent and Hokkaido, I will discuss based on scenario 1 due to its highest posterior probability.

4-4. Discussion

There are three candidate regions for the important refugia or harboured source populations for subsequent range expansion of the widespread Pacific Ocean element *G. erianthum*. The candidates are as follows; 1) traditional alpine plants' origin of Beringia, 2) British Columbia in which several cryptic refugial populations exist, 3) East Asia that was exposed its high genetic diversity and the possibility as the source of alpine plants recently.

Although genetic diversity (i.e. *Ar*) was not different between Hokkaido and Beringia, the genetic differentiation among populations was significantly different between Hokkaido and Beringia. Generally, with longer geographic distances, genetic differentiation among populations increases due to the intensified effects of genetic drift (Wright 1943). There was no IBD between Beringia and British Columbia populations (Fig. 4-3b, Table 4-3b). If the F_{ST} values showed low level and there was no correlation with geographical distance, the populations would be suggested to originate from a single glacial refugia (Stevens & Hogg 2003). On the other hand, if the F_{ST} values showed moderate level and there was a strong correlation with geographic distance, the populations would suppose to originate from several refugia, which might have an older demographic history (Stevens & Hogg 2003). In addition, because sufficient time is needed for reaching equilibrium between gene flow and genetic drift (Slatkin 1993), it is suggested that Beringia and British Columbia populations had not have sufficient time after their distribution expansion. The STRUCTURE analysis suggested that although Aleutian Islands population retained diversified genetic structure, that of Alaska–British Columbia populations were homogeneous. If Alaska–British Columbia populations had originated in a few ancestral populations, limited alleles might be fixed by founder effect. Therefore, Alaska–British Columbia populations might originate in emigrated population from peripheral areas. Phylogenetic tree based on chloroplast genome suggested that

there was no genetic difference between Kamchatka and British Columbia (Fig. 4-2). These results suggest that genetic homogeneous populations are distributed in Beringia thorough 1,500 km. If Alaska and British Columbia are important refugia or harboured source populations for subsequent range expansion, these populations should retain high genetic diversity or complex genetic structure. However, I couldn't observe those phenomena. Therefore, I concluded that Beringia and British Columbia populations were not important refugia and source of expansion. Thus, the candidate populations as refugia or harboured source of 1) and 2) are excluded.

Significantly isolation by distance (IBD) was observed among Hokkaido populations (Fig. 4-3, Table 4-3). Therefore, Hokkaido populations were affected by genetic drift and had have sufficient time after their distribution expansion. In addition, phylogenetic analysis based on chloroplast genome indicated that intraspecific taxa of *G. erianthum* which is distributed in Hokkaido was polyphyly, and cryptic lineages existed in Hokkaido (Fig. 4-2). It is suggested that these results show high genetic diversity of Hokkaido populations and the possibility as cryptic refugia of some Hokkaido populations during interglacial periods.

Comparing genetic structure, Rishiri–Rebun populations were admixture between Hokkaido and Alaska–British Columbia and similar to Kamchatka population (Fig. 4-4). This result suggested that Rishiri–Rebun populations and Kamchatka population had been connected genetically during the glacial periods through East Asia and eastern Siberia populations. The previous study couldn't reveal phylogenetic relationship between *G. erianthum* and the related species *G. onoei* due to inadequate analysis resolution (Wakasugi et al. 2017). Using high resolution genome data, I elucidated that *G. eiranthum* significantly belonged to different phylogenetic clade with *G. onoei* (Fig. 4-2). Comparing the other Japanese *Geranium* species, *G. erianthum* and *G. onoei* have not accumulated genetic differences. This result indicated that their speciation was occurred recently. *G. erianthum* is distributed from northern Japan (Tohoku

district), northeastern China and eastern Siberia to northwestern North America (Jones & Jones 1943; Shimizu 1982, 1983; Akiyama 2001; Xu & Aedo 2008). On the other hand, *G. onoei* is distributed from northeastern China, eastern Siberia and Korean Peninsula to central Japanese archipelago (central Honshu), and not distributed in northern Japan (i.e. Tohoku and Hokkaido district). In addition, the related species *G. platyanthum* Duthie is also distributed in northeastern China and Far Easte Russia (Xu & Aedo 2008). In the western North America, *G. erianthum* is grouped within the taxonomic section of sect. *Sylvatica*, and *G. maculatum*, *G. pretense*, *G. oreganum*, *G. viscosissimum*, *G. richardsonii*, *G. californicum*, and *G. attenuilobum* are also grouped within sect. *Sylvatica* (Jones & Jones 1943). Only *G. richardsonii* overlaps its distribution with *G. erianthum*, but the centre of *G. richardsonii* distribution is western America (Aedo 2001). Therefore, *G. erianthum* is the only *Geranium* species in northwestern North America. Hata et al. (2017) revealed that genetic similarity between Aleutian Islands and Kamchatka populations and concluded that Aleutian Islands populations were the important source of colonization into southern Kamchatka. My study also suggested that the genetic structure of Aleutian Islands population was similar to that of Kamchatka, and their genetic structure were admixture type between Alaska – British Columbia and Hokkaido. However, only *G. erianthum* is distributed in Alaska and Aleutian Islands (Tatewaki & Konayashi 1934; Byrd 1984; Talbot & Talbot 1994; Boucher & Mead 2006; Talbot et al. 2010). The rate of diversification is strongly related to species number (Emerson & Kolm 2005). Therefore, eastern Siberia, Far East Russia and northeastern China are the hotspots of *G. erianthum* and the related *Geranium* species. Although East Asia and eastern Siberia populations were not able to include into present study, it is suggested that *G. erianthum* and the related species were promoted their speciation around East Asia, and later only *G. erianthum* immigrated into northern Japan and Alaska via the land bridge.

Demographic analysis revealed that Hokkaido, Rishiri–Rebun and Beringia (Kamchatka, Aleutian Islands, Alaska and British Columbia) populations diverged at same timing, and the estimated its divergence time was ca. 50,000 (42,900–71,500) years ago (Fig. 4-1, Table 4-4). My demographic estimation may be biased due to the unequal sample size of population group, but my estimation was thought to be approximately correct because the divergence time between *G. erianthum* and the related species *G. onoei* based on chloroplast genome was 100,000 years ago. The estimated divergence time corresponds to the last glacial period. During the last glacial period, it is suggested that between Hokkaido and Sakhalin, Bering strait and Aleutian Islands were connected through the land bridges due to sea level depression. During 90,000 to 80,000 yr B.P., because Hokkaido was slightly warmer, drier summers than in the subsequent glacial period and conifer and deciduous broad-leaved tree were distributed (Heusser & Morley 1985), several high-altitude areas in Hokkaido were thought to be refugial populations for *G. erianthum* which emigrated from East Asia ancient times as cryptic refugia. Thereafter, generally cool, dry environments 60,000–25,000 yr B.P. were preceded by cool moist climates from ~80,000–60,000 yr B.P. (Heusser & Morley 1985). Because it is suggested that boreal conifer forests and forest–tundora transition existed (Heusser & Morley 1985), *G. erianthum* which survived in East Asia immigrated into Hokkaido and till British Columbia through the land bridges.

In conclusion, although *G. erianthum* had been distributed in East Asia (including Hokkaido) during the older glacial periods, and later, it was slightly warmer. As the result, *G. erianthum* survived at East Asia and a part of population had survived at limited area in Hokkaido as an interglacial cryptic refugia during the warmer climate. Thereafter, during the last glacial period, *G. errianthum* individuals which were distributed in East Asia immigrated into Beringia, British Columbia and Hokkaido immediately. In this way, the present distribution

93

was formed.

Table 4-1. The length of each contig. First line is composed locus name. The unit is base pair (bp).

Species/Population	*cemA*	*ccsA*	*rpoA*	*psbK*	*psbZ*	*ndhJ*	*rbcL*
G. erianthum (Hokkaido)	23,294	51,957	10,274	19,521	8,015	18,991	7,350
G. erianthum (Kamchatka)	17,566	32,673	9,851	13,371	8,860	16,758	7,350
G. erianthum (British Columbia)	22,978	32,703	9,851	13,549	8,306	16,492	6,669
G. erianthum var. *glabriusculum*	16,198	33,859	10,274	13,383	7,749	16,498	7,350
G. onoei var. *onoei* f. *yezoense*	24,365	33,868	10,274	9,904	8,303	16,500	7,825

Note: the results of *G. onoei* var. *onoei* f. *onoei* and *G. onoei* var. *onoei* f. *alpinum* were omitted.

Table 4-2. Genetic diversity of the *G. erianthum* populations

	Population	Abbreviation	N	H_o	H_e	F_{IS}	Ar
	Furano1	frn1	2	0.038	0.035	-0.103	1.071
	Furano2	frn2	3	0.061	0.068	0.069	1.098
	Furano3	frn3	4	0.065	0.09	0.208	1.109
	Hikatsu	hik	2	0.021	0.025	0.148	1.048
	Mt. Hirayama	hra	2	0.037	0.031	-0.188	1.075
	Tokachi	tkc	4	0.051	0.061	0.116	1.079
Hokkaido	Mt. Taisetsu	tst	2	0.041	0.048	0.093	1.084
	Mt. Mashike	msk	2	0.051	0.06	0.106	1.1
	Mt. Yubari	ybr	3	0.048	0.078	0.33	1.102
	Mt. Yotei	ytz	4	0.053	0.087	0.32	1.108
	Mt. Maruyama	mar	4	0.061	0.075	0.157	1.09
	Teshio	tes	3	0.054	0.081	0.272	1.108
	Sapporo	sap	4	0.035	0.042	0.141	1.056
Rishiri–Rebun	Rishiri	ris	4	0.056	0.074	0.198	1.099
	Rebun	reb	2	0.041	0.043	-0.01	1.075
	Aleutian	alt	4	0.074	0.102	0.222	1.129
	Boundary Range	br	1	0.041	0.021	-1	1.056
Beringia	Crater Lake	cl	2	0.035	0.035	-0.037	1.054
	Kachemak Bay	kb	3	0.052	0.07	0.208	1.094
	Kamchatka	kmt	4	0.048	0.086	0.373	1.108
	British Columbia	bc	1	0.041	0.021	-1	1.055

Table 4-3a. Pairwise populations F_{ST} values of *G. erianthum* (Hokkaido, Rishiri and Rebun)

	frn1	frn2	frn3	hik	hra	ris	reb	tkc	tst	msk	ybr	ytz	mar	tes	sap
frn1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
frn2	0.164	-	-	-	-	-	-	-	-	-	-	-	-	-	-
frn3	0.126	0.173	-	-	-	-	-	-	-	-	-	-	-	-	-
hik	0.210	0.386	0.342	-	-	-	-	-	-	-	-	-	-	-	-
hra	0.204	0.180	0.233	0.464	-	-	-	-	-	-	-	-	-	-	-
ris	0.340	0.364	0.388	0.462	0.376	-	-	-	-	-	-	-	-	-	-
reb	0.272	0.399	0.337	0.383	0.259	0.211	-	-	-	-	-	-	-	-	-
tkc	0.258	0.359	0.223	0.198	0.268	0.424	0.379	-	-	-	-	-	-	-	-
tst	0.080	0.194	0.116	0.070	0.152	0.333	0.236	-0.026	-	-	-	-	-	-	-
msk	0.065	0.200	0.154	0.284	0.211	0.315	0.264	0.282	0.064	-	-	-	-	-	-
ybr	0.119	0.072	0.033	0.256	0.237	0.332	0.288	0.290	0.150	0.049	-	-	-	-	-
ytz	0.103	0.248	0.224	0.264	0.154	0.287	0.169	0.186	0.057	0.114	0.150	-	-	-	-
mar	0.281	0.354	0.347	0.493	0.400	0.420	0.412	0.412	0.393	0.241	0.252	0.342	-	-	-
tes	0.141	0.202	0.214	0.334	0.195	0.229	0.266	0.312	0.162	0.102	0.213	0.196	0.279	-	-
sap	0.372	0.374	0.407	0.548	0.465	0.446	0.464	0.521	0.477	0.455	0.421	0.381	0.501	0.378	-

Table 4-3b. Pairwise populations F_{ST} values of *G. erianthum* (Beringia and British Columbia)

	alt	br	cl	kb	bc	kmt
alt	-	-	-	-	-	-
br	0.270	-	-	-	-	-
cl	0.322	0.213	-	-	-	-
kb	0.232	-0.060	0.003	-	-	-
bc	0.197	0.692	0.187	-0.217	-	-
kmt	0.090	0.134	0.205	0.111	0.073	-

Table 4-4a. Estimates of demographic parameters in Scenario 1 in DIYABC analysis

	Mean	Median	Mode	q050	q950
N1 (Hokkaido)	47,600	47,000	48,900	20,200	78,100
N2 (Beringia)	17,700	15,300	14,500	6,460	37,600
N3 (Rishiri–Rebun)	38,600	36,800	33,200	18,200	66,600
t_2	14,500	14,300	14,600	7,590	22,000

Note: N1, N2, N3: effective population sizes. t_2: time parameters in the number of generations. Mean, median, mode and 95% quantiles of posterior probability are reported

Table 4-4b. Estimates of demographic parameters in Scenario 3 in DIYABC analysis

	Mean	Median	Mode	q050	q950
N1 (Hokkaido)	76,100	77,500	80,800	52,600	93,800
N2 (Beringia)	12,600	9,560	6,380	2,880	34,100
N3 (Rishiri–Rebun)	33,400	31,500	30,500	15,100	59,400
t_1	16,300	15,200	12,500	5,590	30,200
t_2	22,800	22,300	19,700	12,400	34,900

Note: N1, N2, N3: effective population sizes. t_1 and t_2: time parameters in the number of generations. Mean, median, mode and 95% quantiles of posterior probability are reported
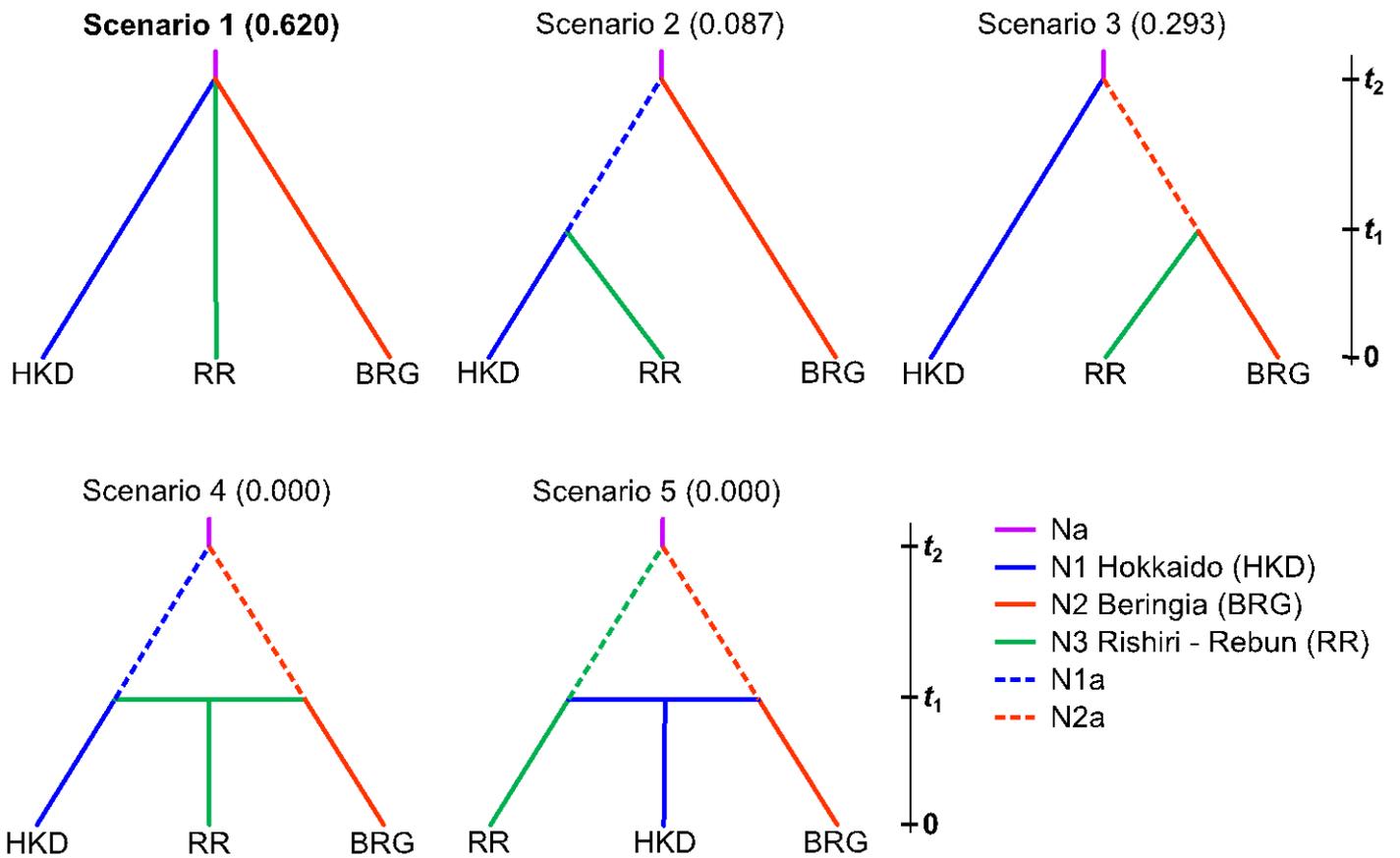
Fig. 4-1. Five divergence scenarios considered in ABC analysis. HKD: Hokkaido, BRG: Beringia, RR: Rishiri–Rebun. Nx is the effective population size ("x" is the arbitrary number of populations or ancestral populations). t_x is the branching time for each population or the mergence time of two populations, and "0" is the present time. The numbers within parentheses are the posterior probabilities of the simulations.
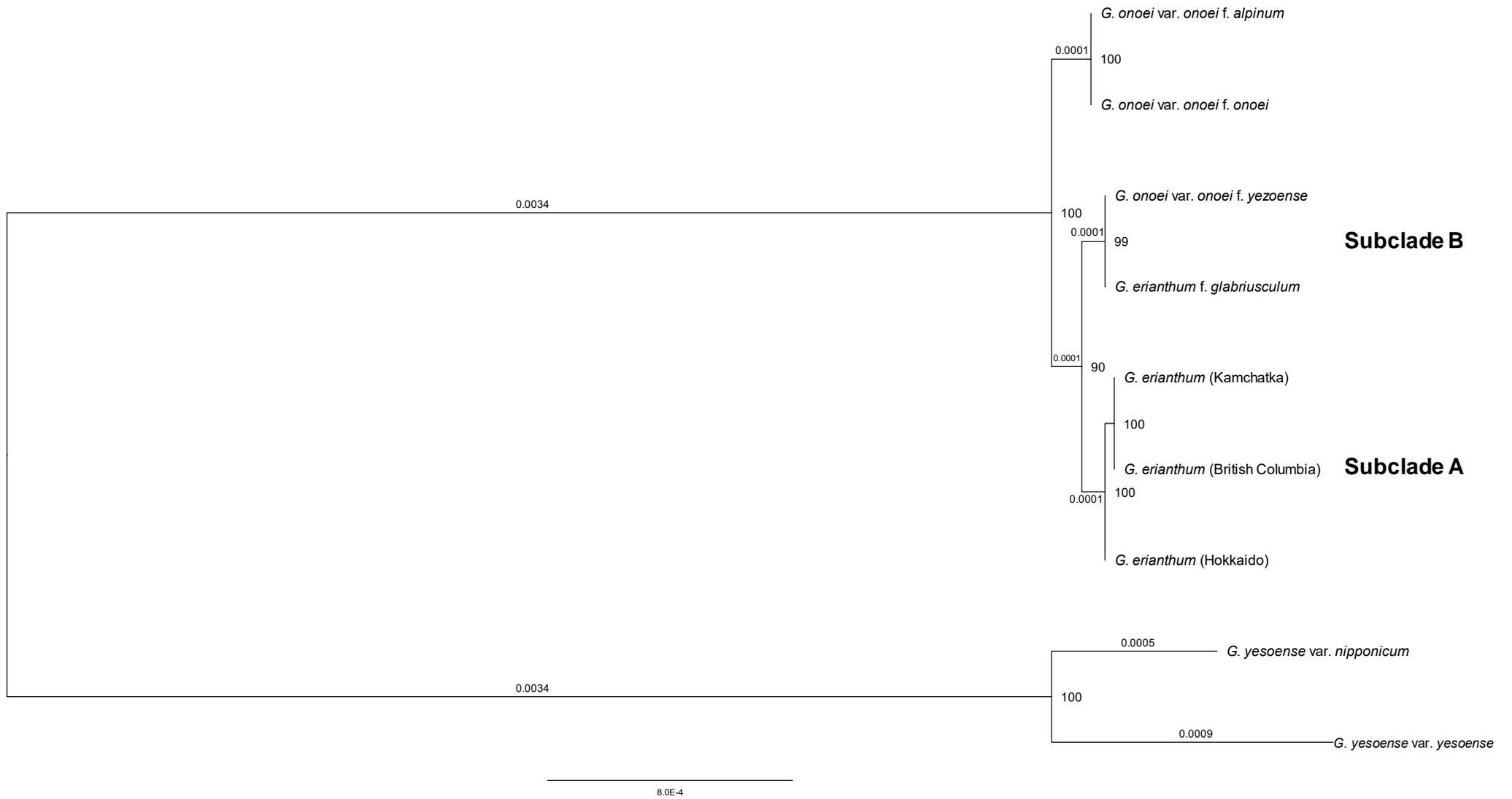
Fig. 4-2. Chloroplast genome phylogenetic ML tree. The numbers by the branch are blanch length. The numbers at the node are bootstrap values.

100
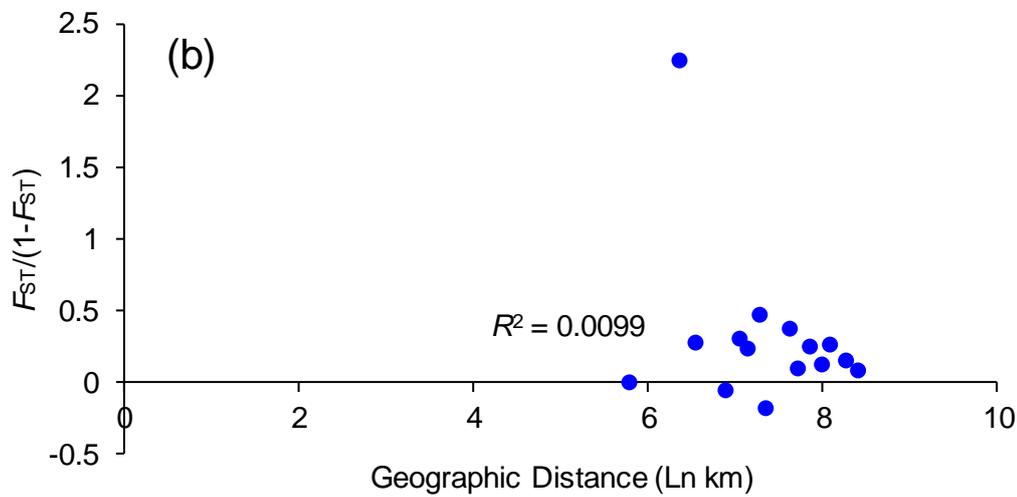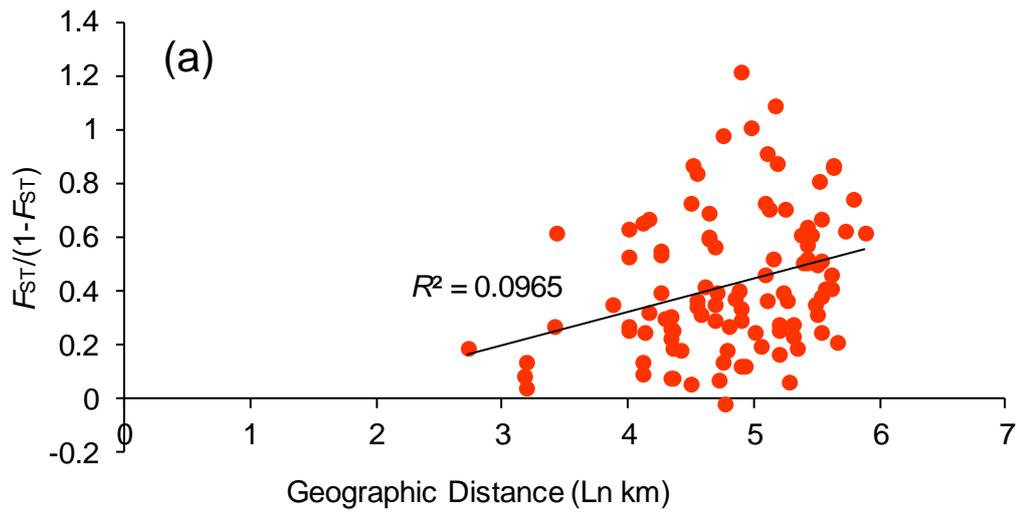
Fig. 4-3. Isolation by distance of *G. erianthum* populations. (a) among Hokkaido populations, (b) among Beringia populations. The horizontal axis represents geographic distance and the vertical axis represents genetic distance; $F_{ST}/(1 - F_{ST})$.
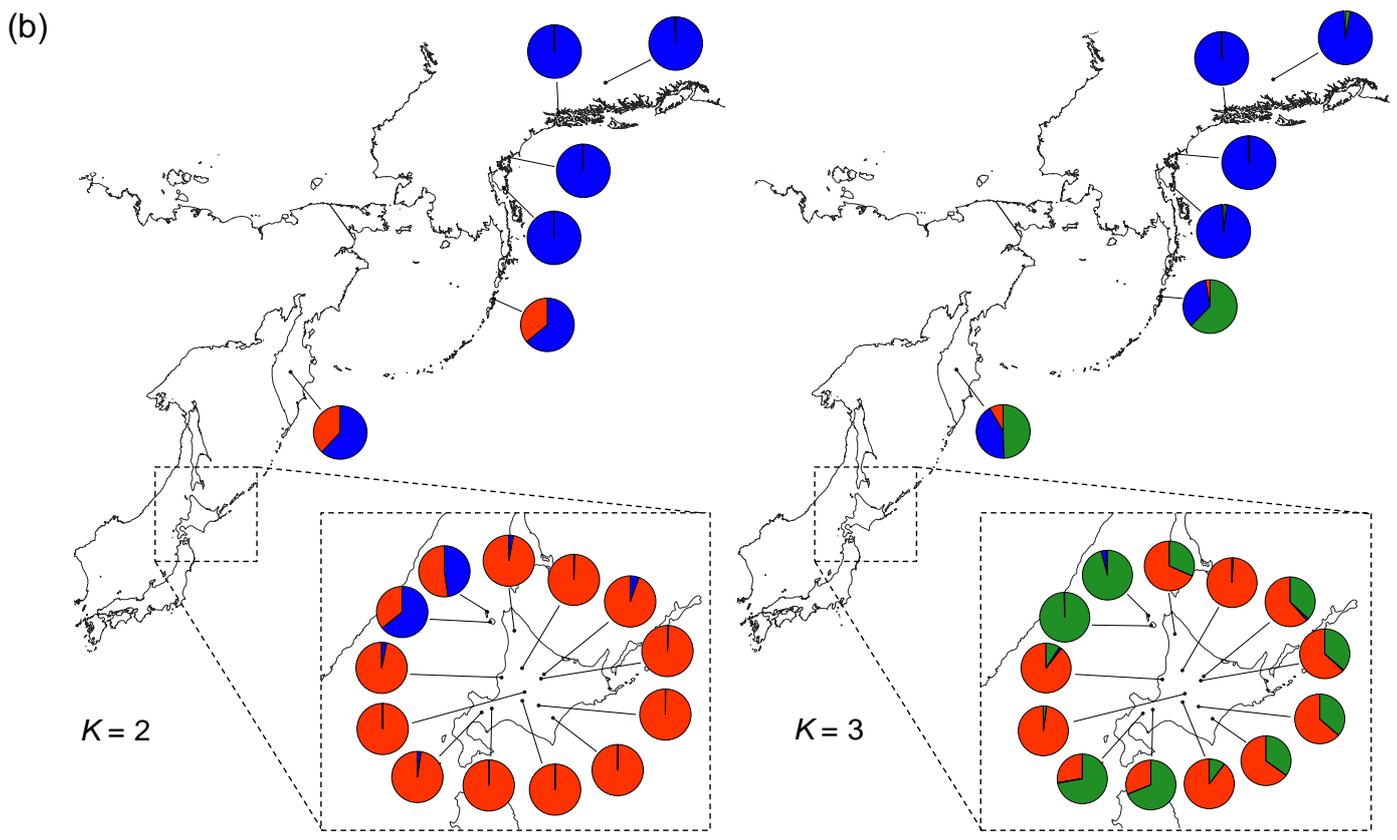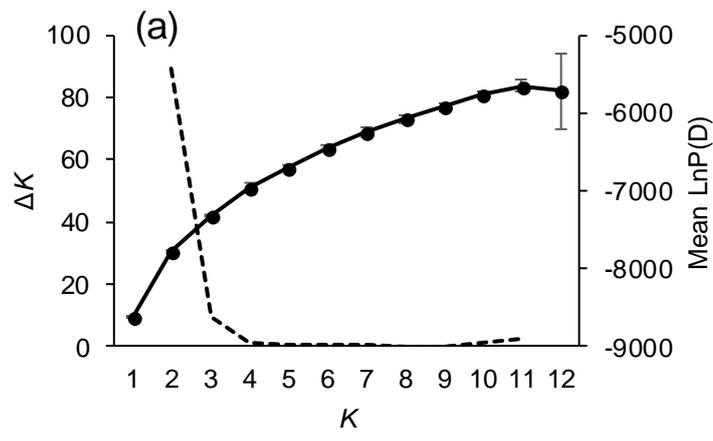
Fig. 4-4. Results of STRUCTU RE analysis. (a) Δ*K* (dotted line) and Mean est. LnP(D) (solid line). (b) Pie chart for the appropriate numbers of clusters.

Chapter 5 A phylogenetic and taxonomic study of Japanese *Geranium*

5-1. Introduction

The genus *Geranium* was described by Carl von Linnaeus in 1753. Knuth (1912) recognized ca. 260 species, and 428 species are currently accepted (The Plant List 2013, version 1.1.; http://www.theplantlist.org, accessed November 2019). Knuth (1912) based the classification on mericarp characters, a system that has subsequently been used in several taxonomic revisions (e.g., Warburg 1938; Carolin 1964; Bortenschlager 1967; Robertson 1972; Tokarski 1972; Yeo 1973; 1984). Yeo (1984) classified *Geranium* into three different subgenera (subg. *Geranium*, *Robertium*, and *Erodioidea*). The subg. *Geranium* is widely distributed globally, excluding low-latitude areas, deserts, and the polar regions (Aedo et al. 1998). The subg. *Robertium* is distributed from Europe to East Asia (Aedo et al. 1998), whereas the subg. *Erodioidea* is distributed around the Mediterranean, central eastern Africa, and central eastern South America (Aedo et al. 1998).

Thirteen *Geranium* species are distributed across the Japanese Archipelago (Kadota 2016). It is generally accepted that Japanese *Geranium* are represented by two subgenera (subg. *Geranium* and *Robertium*) based on their mericarp projections. Of these two, subg. *Robertium* comprises a single species (*G. robertianum* L.), and the remaining 12 species are members of subg. *Geranium*. *Geranium robertianum* is grouped within Section *Robertiana* (Kadota 2016), and the remaining 12 species are grouped within five sections (Sects. *Columbina*, *Sylvatica*, *Plaustria*, *Sibirica*, and *Mexicana*; Kadota 2016). Additionally, nine varieties and 14 forms are reported among these 13 species (Kadota 2016).

Taxonomic systems should reflect phyletic relationships, and phylogenetic analyses are the best tools by which to elucidate these relationships. Fiz et al. (2008) summarized the phylogenetic relationships within Geraniaceae using chloroplast fragments (*trnL-F* and *rbcL*)

and further roughly estimated phylogenetic relationships within the genus *Geranium*. The genera *Geranium* and *Erodium* are sister taxa, and subg. *Robertium* diverged at the base of the *Geranium* clade (Fiz et al. 2008). Although the phylogenetic relationships among some species within Geraniaceae were elucidated by Fiz et al. (2008), relationships among Japanese *Geranium* species have remained unclear due to inadequate sampling. Marcussen & Mesenguer (2017) also performed phylogenetic analyses of *Geranium* using chloroplast fragments (*trnL-F* and *rbcL*) and nuclear sequences (internal transcribed spacer [ITS] region), but only one taxon from Japan was included in these analyses (*G. yesoense*).

Phylogenetic analysis of Japanese samples were only available after Wakasugi et al. (2017), when 14 Japanese *Geranium* species samples were compared using chloroplast fragments (*trnL-F*) and nuclear sequences (ITS). Although the results of those analyses clearly show the phylogenetic relationships among the investigated species, infra-species relationships (i.e., within varieties and forms) were not enough assessed due to low resolution. Therefore, variety- and form-level relationships among Japanese *Geranium* taxa remain unclear and need to be assessed using high-resolution genetic markers. Recent advances in the implementation of sequence technology have made genome-wide genetic polymorphisms of non-model organisms easier to obtain. In this study, I report the variety- and form-level phylogenetic relationships within Japanese *Geranium* based on chloroplast genome data and nuclear genome-wide single-nucleotide polymorphisms (SNPs). I also provide further details on the taxonomy of Japanese *Geranium* based on the results of my phylogenetic analyses.

5-2. Materials and Methods

Plant sampling

Mature leaves of 31 *Geranium* taxa were obtained between July 2017 and October 2018 (Table 5-1), including 23 samples from Korea, China, Taiwan, and Russia. Leaf samples were immediately dried, preserved at room temperature using silica gel, and stored in a dark room. Voucher specimens were registered at Herbarium of the University of Tokyo (TI), Japan.

Nuclear genome-wide SNP analyses of sect. *Sylvatica* were performed using seven taxa represented by 121 individuals from 37 populations (Table 5-2). Additionally, nuclear genome-wide SNP analyses of sect. *Mexicana* and sect. *Sibirica* were performed using six taxa represented by 78 individuals from 23 populations (Table 5-3).

Phylogenetic analyses based on chloroplast whole-genome sequences

Prior to DNA extraction, finely-powdered leaf samples were washed using a buffer solution (3% HEPES buffer [pH 8.0] comprising 1% polyvinylpyrrolidone, 0.9% ascorbic acid, and 2% 2-mercaptoethanol). Genomic DNA was extracted using a modified cetyl trimethylammonium bromide method (Milligan 1992), and total DNA was measured using a Qubit 2.0 fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). Paired-end sequencing was performed using a BGISEQ-500 next-generation sequencing platform (BGI, Shenzhen, China). The sequence reads were assembled *de novo* using NOVOPlasty (Dierckxsens et al. 2017), with the *G. incanum* (NC_030045.1) plastid sequence used as a reference sequence, and each locus (*ccsA*, *cemA*, *rbcL*, *ndhJ*, *psbK*, *rpoA*, and *psbZ*) of *G. incanum* designated as a seed sequence in the assembly analyses. Because the assembly did not yield continuous chloroplast sequences from the samples, contigs around the above loci were used in the phylogenetic analyses. Contigs were aligned using MAFFT version 7.310-1 (Katoh et al. 2002, 2013). All

gaps originating from inversions and/or deletions were eliminated, and phylogenetic analyses were performed based on nucleotide sequences using RAxML version 8.2.11 (Stamatakis 2006), using the nucleotide substitution model GTRGAMMA and 1000 bootstraps. Bootstrap values of the best tree were calculated using the pgsumtree option in Phylogears2 (Tanabe 2008). The resulting maximum likelihood (ML) tree was visualized and edited by FigTree version 1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/).

Phylogenetic analysis based on nuclear genome-wide SNPs

Phyletic estimations based on the chloroplast genome indicated that the phylogeny of *G. onoei* var. *onoei* f. *yezoense* (sect. *Sylvatica*) and *G. tripartitum* var. *hastatum* (sect. *Mexicana*) was not consistent between the present taxonomic system and estimated phylogenetic relationships (Fig. 5-1; Yamazaki 1993, Yonekura 2005). Therefore, a phyletic estimation of sect. *Sylvatica* and sect. *Mexicana* based on nuclear genome-wide SNPs was performed. In addition, *G. sibiricum* (sect. *Sibirica*) was included in the analysis of sect. *Mexicana* because this species e is closely related to sect. *Mexicana* (Wakasugi et al. 2017). Multiplexed inter-simple sequence repeat genotyping by sequencing (MIG-seq; Suyama & Matsuki 2015) was performed to obtain nuclear genome-wide SNPs.

The first 14 bases of read 2 (reverse-read) sequences and low-quality sequences of raw reads were removed using the FASTX-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/), with the minimum quality score set to 30 (-q 30). Reads comprising >40% bases that failed to meet the minimum quality score were removed (-p 40). In addition, the primer sequence region of each read was removed using TagDust (Lassmann et al. 2009). After the removal of unnecessary reads, the remaining reads were assembled with Stacks version 1.48 (Catchen et al. 2013), and SNPs were called using the same program. In *ustacks*, the maximum distance (in nucleotides)

was set to 2 (-M 2) and the minimum depth for creating the stack was set to 3 (-m 3). The maximum distance (-N) was set to 2. Catalogs were created using *cstacks*, and the genotypes of individuals were decided using *sstacks*. SNPs were exported using *Populations*, with the following settings: genotyping rate, which is the minimum percentage of individuals in a population, was set to 0.1 (-r 0.1); minor alleles (<1% within population) were removed (--min-maf 0.01); and loci with unreasonable heterozygosity were removed (--max-obs-het 0.95). A total of 15,179 SNPs for sect. *Sylvatica* and 12,153 SNPs for sect. *Mexicana* and sect. *Sibirica* were obtained. These SNPs were filtered using PLINK version 1.9 (Purcell et al. 2007) by removing individuals with 80% missing data, and loci that were held by <40 % of individuals (--mind 0.8, --geno 0.4). Minor alleles (<3%) were also removed (--maf 0.03). Phylogenetic analyses were then performed using the concatenated SNP matrix and RAxML version 8.2.11 (Stamatakis 2014); the GTRGAMMA nucleotide substitution model and the JC69 model of DNA evolution (Jukes & Cantor 1969) were adopted. The number of bootstraps was set at 1000, and bootstrap values of the best tree were later calculated using the pgsumtree option in Phylogears2. The resulting ML tree was visualized and edited in FigTree version 1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/).

5-3. Results

Phylogenetic analyses based on the chloroplast genome

Phylogenetic analyses based on the chloroplast genome (total length, 61,578 bp) revealed that species classified in the same Section were grouped in the same clades (Fig. 5-1), thus indicating consistency with the current *Geranium* classification. Nearly all varieties and forms were closely related to their respective fundamental varieties; i.e., *G. wilfordii* var. *yezoense* was closely related to *G. wilfordii* var. *wilfordii*, *G. onoei* var. *onoei* f. *alpinum* was closely related to G. *onoei* var. *onoei* f. *onoei*, *G. erianthum* f. *glabriusculum* was closely related to *G. erianthum,* and the *G. shikokianum* species complex formed a subclade (Fig. 5-1).

Geranium onoei var. *onoei* f. *yezoense* is described as a form of *G. onoei* var. *onoei* (Hara 1948; Yonekura 2005). Wakasugi et al. (2017) suggested that *G. onoei* var. *onoei* f. *yezoense* is closely related to *G. onoei* var. *onoei* and *G. erianthum*, but did not find evidence of genetic differences among these taxa. By contrast, my results indicate that *G. onoei* var. *onoei* and *G. erianthum* belong to different clades, and *G. onoei* var. *onoei* f. *yezoense* belongs to the *G. erianthum* clade (Fig. 5-1).

Each species belonging to sect. *Mexicana* (*G. thunbergii*, *G. wilfordii*, and *G. tripartitum*) formed a distinct clade (Fig. 5-1). Because *G. tripartitum* var. *hastatum* has never been assessed in phylogenetic analyses, its phyletic position has never been defined. The ML tree based on the chloroplast genome indicates that *G. tripartitum* var. *hastatum*, which is described as a variety of *G. tripartitum*, is closely related to *G. thunbergii* (Fig. 5-1).

Although a total of 5 GB of read data was obtained in this study, I was unable to obtain a complete chloroplast genome sequence. The chloroplast genome in Geraniaceae is characterized by a high frequency of large repeats (>100 bp) relative to those of other rosids, and the number of large repeats (>200 bp) is highest in *Geranium* among all genera in

Geraniaceae (Guisinger et al. 2011). The length of the sequence reads was 100 bp and that of the insert sequence was 300 bp. It is estimated that the assemble here was not able to reconstruct contigs for regions with large repeats (>100 bp) occurring in the *G. yesoense* chloroplast genome.

Nuclear genome phylogeny based on MIG-seq of sect. Sylvatica

The total and average number of reads obtained from MIG-seq were 19,755,164 and 135,309, respectively. After completing the filtering procedure using PLINK version 1.9, 262 SNPs and 117 samples remained in the data matrix. In the phylogenetic ML tree, as with the phylogenetic analyses based on the chloroplast genome, the *G. onoei* species complex and the *G. erianthum* species complex formed independent clades (Fig. 5-2). In addition, *G. onoei* var. *onoei* f. *yezoense* appeared to be interspersed with *G. erianthum* samples (Fig. 5-2).

Nuclear genome phylogeny based on MIG-seq of sect. Mexicana *and* sect. Sibirica

The total and average number of reads obtained from MIG-seq were 25,306,440 and 160,167, respectively. After filtering using PLINK version 1.9, 220 SNPs and 72 samples remained in the data matrix. In the phylogenetic ML tree, as with the phylogenetic analyses based on the chloroplast genome, sects. *Mexicana* and *Sibirica* formed distinct clades (Fig. 5-3). *Geranium tripartitum* was placed in a paraphyletic group with *G. tripartitum* var. *hastatum*, which was sister to other taxa, including samples of *G. tripartitum* and *G. thunbergii* (Fig. 5-3).

5-4. Discussion

The results of the phylogenetic analyses based on the chloroplast genome support the present taxonomic system used for *Geranium* in Japan. Specifically, the present taxonomy accurately reflects phyletic relationships because all Sections emerged as monophyletic and nearly all varieties were closely related to the fundamental varieties. However, these analyses were based on only one sample per taxon, and thus more detailed phylogenetic analyses are required in the future. Interestingly, *G. yesoense* specimens exhibited genetic structures reflective of the varieties described within the species complex. Based on these results, the three varieties of the *G. yesoense* species complex (var. *yesoense*, var. *pseudopratense* and var. *nipponicum*) may be considered independent taxonomic groups. However, the morphological features of these varieties, in particular the diagnostic traits, tend to overlap with one another (Wakasugi et al. 2017). In addition, Hakodate in Hokkaido Prefecture is the type locality of both var. *yesoense* and var. *pseudopratense*, which exhibit different morphological characteristics—var. *yesoense* has dense hairs on the sepals and var. *pseudopratense* has moderately dense hairs on the sepals. Therefore, the taxonomy of the *G. yesoense* species complex could not be resolved. A further detailed taxonomic study is required to resolve the taxonomy of this species complex.

There was discordance between the phylogenetic relationships elucidated from the analyses here and the current taxonomy of Japanese *Geranium. Geranium onoei* var. *onoei* f. *yezoense* belonged to the clade with *G. erianthum* (Figs. 5-1, 5-2), and *G. tripartitum* var. *hastatum* was grouped with *G. thunbergii* in the phylogenetic tree based on the chloroplast genome, but this variety diverged from *G. tripartitum* and *G. thunbergii* at the base of the nuclear genome-based phylogenetic tree (Figs. 5-1, 5-3). Therefore, alternative phylogenies for *G. onoei* var. *onoei* f. *yezoense* and *G. tripartitum* var. *hastatum* should be proposed. First, *G.*

onoei var. *onoei* f. *yezoense* should be treated as a variety of *G. erianthum* due to its phylogenetic proximity with *G. erianthum*. These two species are morphologically distinguishable by *G. onoei* var. *onoei* f. *yezoense* having spreading hairs on the stem and petioles with spreading or glandular hairs. Next, *G. tripartitum* var. *hastatum* should be treated as a distinct species due to its phylogenetic independence from *G. thunbergii* and *G. tripartitum*. *Geranium tripartitum* var. *hastatum* is distinguishable from *G. thunbergii* by having more deeply lobed leaves and is also distinguishable from *G. tripartitum* morphologically.

Geranium tripartitum var. *hastatum* was initially described as a full species—*G. hastatum* Nakai (Nakai 1909). Later, Hara (1948) proposed that *G. hastatum* should be considered a variety of *G. wilfordii*. Thereafter, Yamazaki (1993) proposed that *G. wilfordii* var. *hastatum* should be considered a new variety of *G. tripartitum*, which has been adopted in the current taxonomic system. Results of the chloroplast-genome phylogenetic analyses in this study indicate that *G. tripartitum* var. *hastatum* is phylogenetically close to *G. thunbergii*. Therefore, *G. tripartitum* var. *hastatum* may be closely related to all species of sect. *Mexicana*. Although hybridization between *G. tripartitum* and closely related species has never been reported, the closely related *G. thunbergii* and *G. wilfordii* can hybridize (Nishida et al. 2020). The phylogenetic analyses revealed that there is a discordance between the chloroplast and nuclear genomes. Phyletic discordance between chloroplast and nuclear genomes may originate from past hybridization events (e.g., Pan et al. 2007, Liu et al. 2008). However, the probability of previous hybridization was not assessed in this study as sample and population numbers were limited. *Geranium tripartitum* var. *hastatum* and *G. thunbergii* were found growing together at Nikko, Japan, but only *G. tripartitum* var. *hastatum* was collected at this site. With more detailed sampling, interspecies gene flow could be assessed using SNP or simple sequence repeat markers. Finally, the possibility of introgressive hybridization between *G. tripartitum* var.

hastatum and *G. thunbergii* could be investigated using chloroplast fragments (e.g., *trnL-F*, *matK*, and *rbcL*). Increasing the number of sampling locations of *G. tripartitum* and *G. tripartitum* var. *hastatum* would also contribute greatly to resolving their phylogenetic relationship.

Taxonomic treatment

Geranium erianthum var. *yezoense* (Hara) Kurata, **comb. nov.**

Basionym: *G. eriostemon* Fisch. ex DC. var. *reinii* (Franch. et Sav.) Maxim. f. *yezoense* Hara J. J. B. 22: 166 & 171. 1948. *G. onoei* Franch. et Sav. var. *onoei* f. *yezoense* (H.Hara) Yonek. J. J. B. 80: 325. 2005.— Type: Japan, Hokkaido, Sapporo, 21 June 1892, *K. Miyabe* (holo- TI!).

Perennial herbs; stem with spreading hairs, apical cauline leaves sessile, opposite, petiole with spreading or glandular hairs, leaves 7–9-lobed deeply, 5–20 cm wide, 2–2-flowered, flowers June to July, petals darker.

Distribution. Endemic to Hokkaido, Japan.

Habitat. Forests and grassy places.

Japanese name. Ezogunai-furo

Notes. *Geranium erianthum* is identified from *G. onoei* based on presence of spreading hairs on the stems and petioles. *G. erianthum* has no spreading hairs, and *G. onoei* has spreading hairs on the stems and petioles. Although *G. onoei* var. *onoei* f. *yezoense* has spreading hairs on the stems and petioles (Fig. 5-4), which these features accord with *G. onoei*, phylogenetic analyses show that *G. onoei* var. *onoei* f. *yezoense* is close to *G. erianthum*. Furthermore, *G. eriostemon*

112

Fisch. ex DC. var. *reinii* (Franch. et Sav.) Maxim. is treated as synonym of *G. onoei* var. *onoei* f. *onoei*.

Representative specimens examined. **Japan.** Hokkaido: Sapporo, 17 June 1943, M. Mizushima (TI). Sapporo, 20 June 1962, K. Hiyama (MAK).

Geranium hastatum Nakai in Bot. Mag. Tokyo 23(268): 100. 1909.

≡ *G. wilfordii* Maxim. var. *hastatum* H.Hara in J. Jap. Bot. 22: 170 & 172. 1948. ≡ *G. tripartitum* R.Knuth var. *hastatum* (H.Hara) T.Yamaz. in J. Jap. Bot. 68: 239. 1993. — Type: Japan, Tochigi, Nikko, 29 September 1879, *collector unknown s.n.* [lecto- TI! designated by Hara (1948) as holo].

Sepals with spreading hairs on outer surface, pedicels with retrorse and appressed hairs, leaf blade hastate, flowers July to October, petals white.

Distribution. Endemic to Tochigi and Gunma, Japan.

Habitat. Grassy places in summer green forest (Kadota 2016).

Japanese name. Hokogata-furo

Notes. G. hastatum is identified from *G. wilfordii* based on the presence of long hairs on the sepals (i.e. *G. wilfordii* has no hairs on the sepals). *G. hastatum* retains phylogenetic independence from *G. thunbergii* and *G. tripartitum*. In addition, unlike other related species (i.e. *G. wilfordii*, *G. thunbergii* and *G. tripartitum*), *G. hastatum* has hastate leave (Fig. 5-5).

Representative specimens examined. **Japan.** Gunma: Marunuma, July 1930, H. Asuyama (TNS41118). Tochigi: Nikko, June 1887, H. Sakurai (TNS7585).

Table 5-1. Species of *Geranium* sampled and sampling locations

Section	Species	Locality	N
	G. erianthum	Obihiro, Hokkaido Pref. Japan	1
	G. erianthum f. *glabrisculum*	Asahikawa, Hokkaido Pref. Japan	1
	G. erianthum	Kamchatka, Russia	1
Sylvatica	*G. erianthum*	British Columbia, Canada	1
	*G. onoe*i var. *onoei* f. *onoei*	Ueda, Nagano Pref. Japan	1
	G. onoei var. *onoei* f. *alpinum*	Mt. Kitadake, Yamanashi Pref. Japan	1
	G. onoei var. *onoei* f. *yezoense*	Sapporo, Hokkaido Pref. Japan	1
	G. krameri	Minamimaki, Nagano Pref. Japan	1
	G. krameri cf.	Far East Russia	1
	G. yoshinoi	Kitahiroshima, Hiroshima Pref. Japan	1
	G. shikokianum var. *shikokianum*	Oda, Shimane Pref. japan	1
	G. shikokianum var. *yamatense*	Tenkawa, Nara Pref. Japan	1
	G. shikokianum var. *kaimontanum*	Fujikawaguchiko, Yamanashi Pref. Japan	1
Palustria	*G. shikokianum* var. *yoshiianum*	Yakushima, Kagoshima Pref. Japan	1
	G. yesoense var. *yesoense*	Akkeshi, Hokkaido Pref. Japan	1
	G. yesoense var. *pseudopratense*	Hakodate, Hokkaido Pref. Japan	1
	G. yesoense var. *nippponicum*	Yuza, Yamagata Pref. Japan	1
	G. yesoense cf.	Maibara, Shiga Pref. Japan	1
	G. soboliferum var. *hakusanense*	Minamimaki, Nagano Pref. Japan	1
Sibirica	*G. sibiricum*	Kitami, Hokkaido Pref. Japan	1
	G. thunbergii	Okutama, Tokyo Pref. Japan	1
	G. thunbergii	Chongqing, China	1
	G. wilfordii var. *wilfordii*	Okutama, Tokyo Pref. Japan	1
Mexicana	*G. wilfordii* var. *wilfordii*	Sichuan, China	1
	G. wilfordii var. *yezoense*	Kitami, Hokkaido Pref. Japan	1
	G. tripartitum	Fujikawa, Yamanashi Pref. Japan	1
	G. tripartitum var. *hastatum*	Nikko, Tochigi Pref. Japan	1
Columbina	*G. carolinianum*	Zhejiang, China	1
-	*G. koreanum*	Korea	1
-	*G. hayatanum*	Taiwan	1
-	*G. rectum*	Xinjiang Uyghur Autonomous Region	1
-	*G. saxatile*	Tajikistan	1
-	*G. dahuricum*	Xinjiang Uyghur Autonomous Region	1

-	*G. rostornii*	Hubei, China	1
-	*Geranium* sp.	Halbin, China	1
-	*Geranium* sp.	Russia	1
-	*Geranium* sp.	Inner Mongolia	1

N: number of samples; " - " : samples that could not be classified into an appropriate Section.

Table 5-2. *Geranium* species sampled and sampling locations of *Geranium* sect. *Sylvatica*

Species	Population/Locality	N
	Rebun, Hokkaido Pref. Japan	3
	Rishirifuji, Hokkaido Pref. Japan	4
	Mt. Mashike, Mashike, Hokkaido Pref. Japan	2
	Takanegahara, Biei, Hokkaido Pref. Japan	4
	Mt. Hirayama, Kamikawa, Hokkaido Pref. Japan	4
	Furano, Hokkaido Pref. Japan	3
	Furano, Hokkaido Pref. Japan	3
	Furano, Hokkaido Pref. Japan	4
	Mt. Yubari, Minamifurano, Hokkaido Pref. Japan	4
G. erianthum	Hidaka, Hokkaido Pref. Japan	2
	Obihiro, Hokkaido Pref. Japan	4
	Mt. Yotei, Makkari, Hokkaido Pref. Japan	4
	Esso, Kamchatka, Russia	4
	Avacha, Kamchatka, Russia	1
	Aleutian Islands, USA	4
	Boundary Range, USA	2
	Crater Lake, USA	2
	Kachemak Bay, USA	3
	British Columbia, USA	1
G. erianthum f. *glabrisculum*	Horonobe, Hokkaido Pref. Japan	3
	Asahikawa, Hokkaido Pref. Japan	4
	Mt. Amakazari, Otari, Nagano Pref. Japan	1
	Mt. Sannomine, Hakusan, Ishikawa Pref. Japan	4
	Tsumagoi, Gunma Pref. Japan	4
	Ueda, Nagano Pref. Japan	4
	Minamimaki, Nagano Pref. Japan	1
*G. onoe*i var. *onoei* f. *onoei*	Mt. Korenge, Hakuba, Nagano Pref. Japan	4
	Mt. Nishihodaka, Matsumoto, Nagano Pref. Japan	2
	Mt. Norikura, Matsumoto, Nagano Pref. Japan	4
	Matsumoto, Nagano Pref. Japan	4
	Mt. Akaishi, Shizuoka, Shizuoka Pref. Japan	4
	Mt. Kushigata, Fujikawa, Yamanashi Pref. Japan	4
	Mt. Ibuki, Maibara, Shiga Pref. Japan	4

G. onoei var. *onoei* f. *alpinum*	Mt. Kitadake, Yamanashi Pref. Japan	4
	Mt. Kisokomagatake, Komagane, Nagano Pref. Japan	2
G. onoei var. *onoei* f. *albiflorum*	Mt. Bandai, Inawashiro, Fukushima Pref. Japan	2
	Mt. Kitadake, Yamanashi Pref. Japan	4
G. onoei var. onoei f. *yezoense*	Sapporo, Hokkaido Pref. Japan	4

N: number of samples

Table 5-3. *Geranium* species sampled and sampling locations of *Geranium* sect. *Mexicana* and sect. *Sibirica*

Section	Species	Population/Locality	*N*
Mexicana	*G. thunbergii*	Esashi, Hokkaido Pref. Japan	3
		Saku, Nagano Pref. Japan	3
		Okayama, Okayama Pref. Japan	3
		Amakusa, Kumamoto Pref. Japan	2
		Yakushima, Kagoshima Pref. Japan	3
		Sichuan, China	1
		Chongqing, China	2
		Guizhou, China	3
	G. wilfordii var. *wilfordii*	Matsumoto, Nagano Pref. Japan	4
		Okutama, Tokyo Pref. Japan	4
		Saku, Nagano Pref. Japan	4
		Sichuan, China	1
	G. wilfordii var. *yezoense*	Kitami, Hokkaido Pref. Japan	4
	G. tripartitum	Fujikawa, Yamanashi Pref. Japan	1
	G. tripartitum var. *hastatum*	Nikko, Tochigi Pref. Japan	1
Sibirica	*G. sibiricum*	Rishirifuji, Hokkaido Pref. Japan	4
		Biei, Hokkaido Pref. Japan	4
		Kitami, Hokkaido Pref. Japan	4
		Nemuro, Hokkaido Pref. Japan	4
		Sapporpo, Hokkaido Pref. Japan	3
		Tomakomai, Hokkaido Pref. Japan	4
		Primorskii, Russia	3
		Kangwon Province, Korea	3

N: number of samples
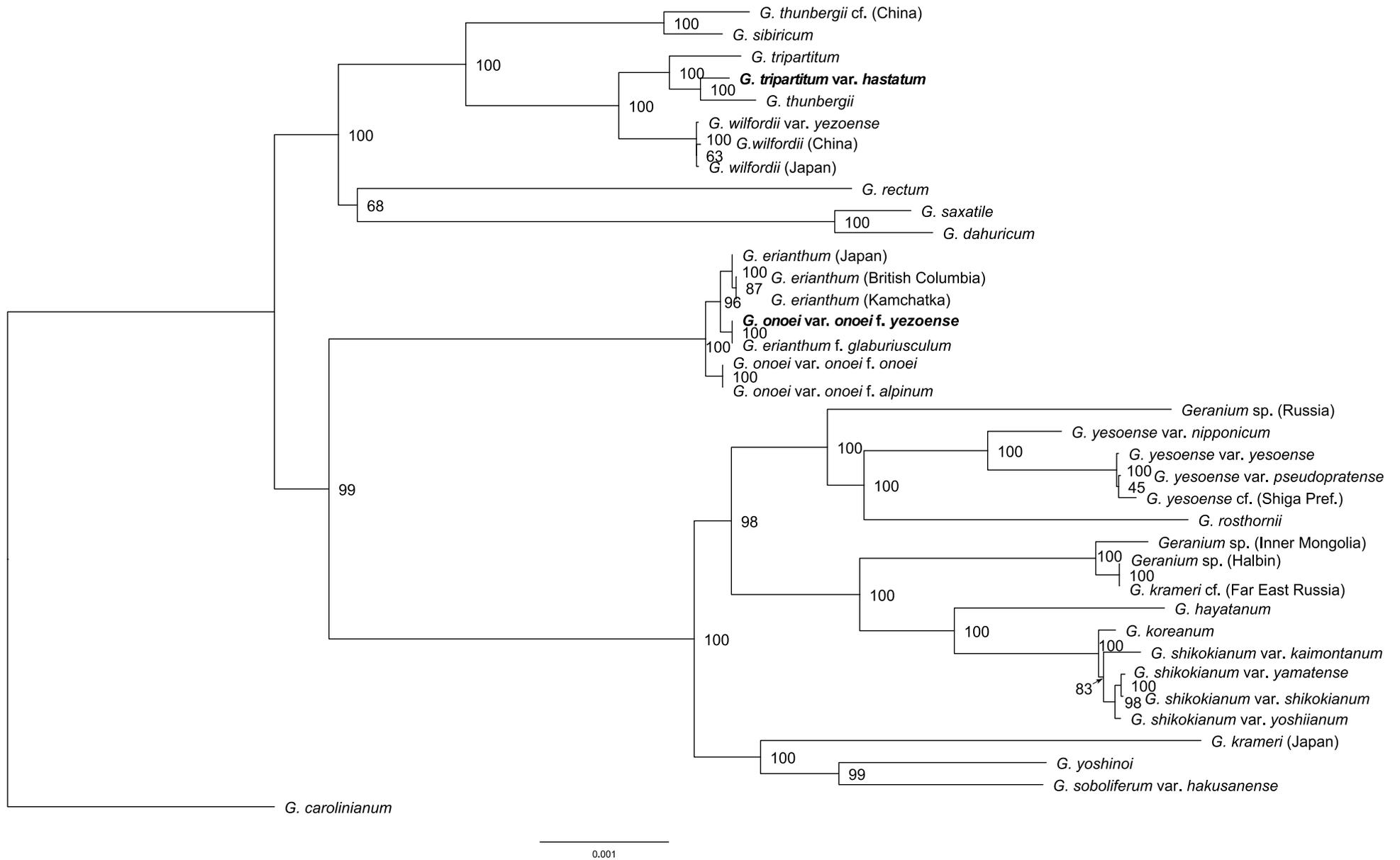
Fig. 5-1. Maximum likelihood tree of *Geranium* species. Samples that could not be identified to species are listed as *Geranium* sp.
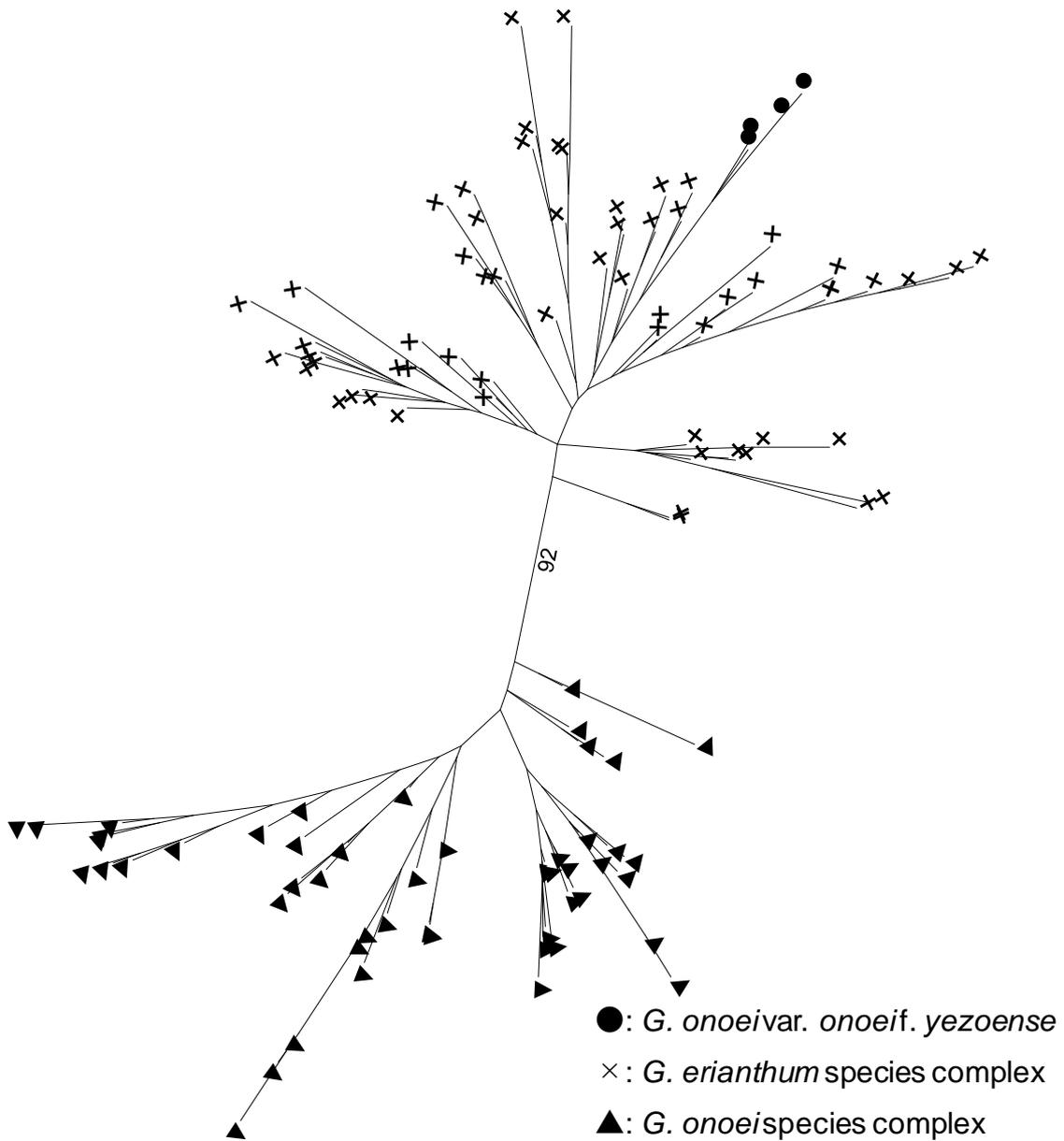
Fig. 5-2. Maximum likelihood tree of sect. *Sylvatica* based on multiplexed inter-simple sequence repeat genotyping by sequencing (MIG-seq) analysis of the *Geranium* nuclear genome. The number represents the bootstrap value associated with the node.
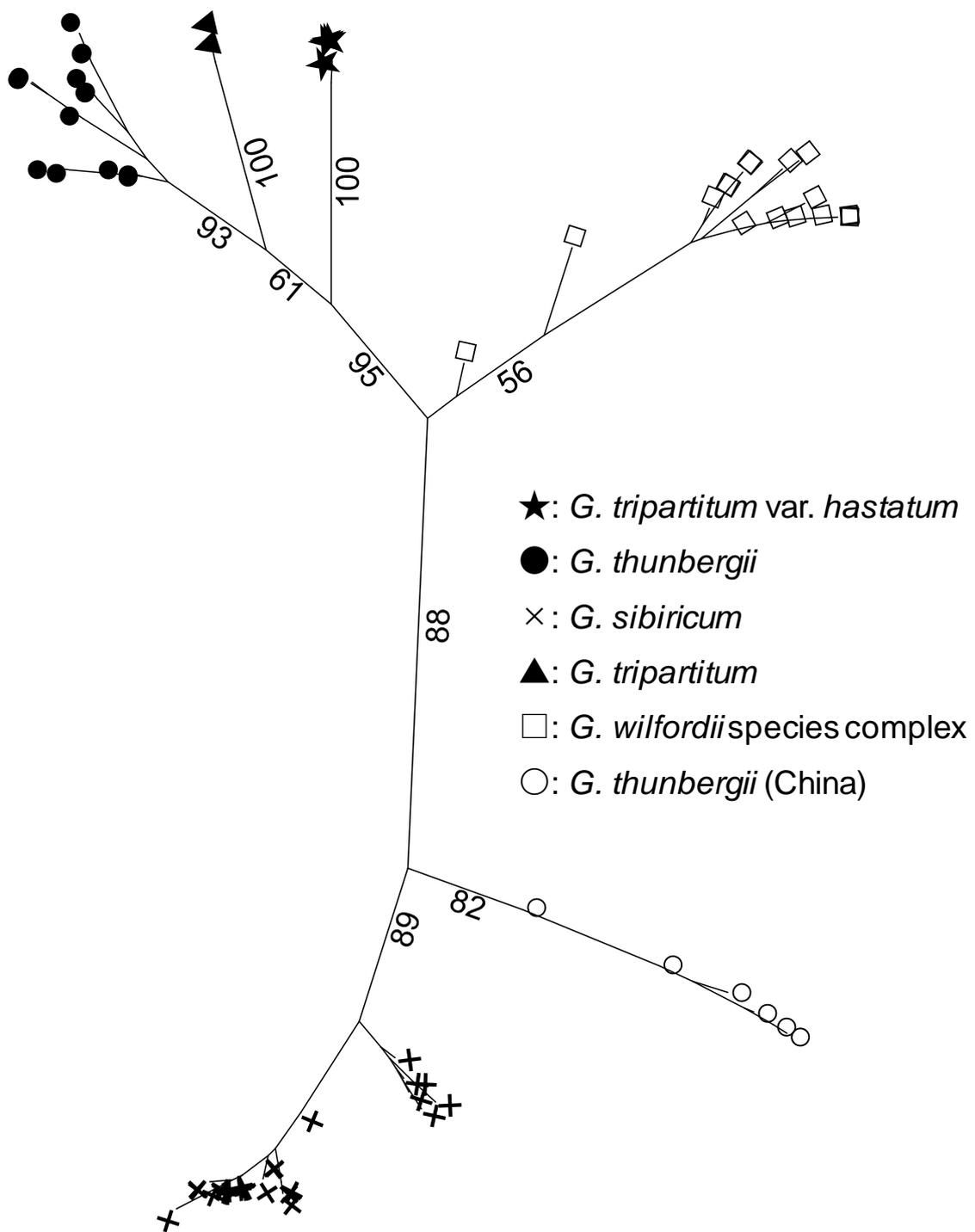
Fig. 5-3. Maximum likelihood tree of sect. *Mexicana* and sect. *Sibirica* based on MIG-seq analysis of the *Geranium* nuclear genome. Numbers represent the bootstrap values associated with the major nodes.

Fig. 5-4. *Geranium erianthum* var. *yezoense*. A: Petiole (holo- TI, TI00011256). B: Stem (Japan, Hokkaido pref., Sapporo, 17 June 1943, collected by M. Mizushima, registered in TI).
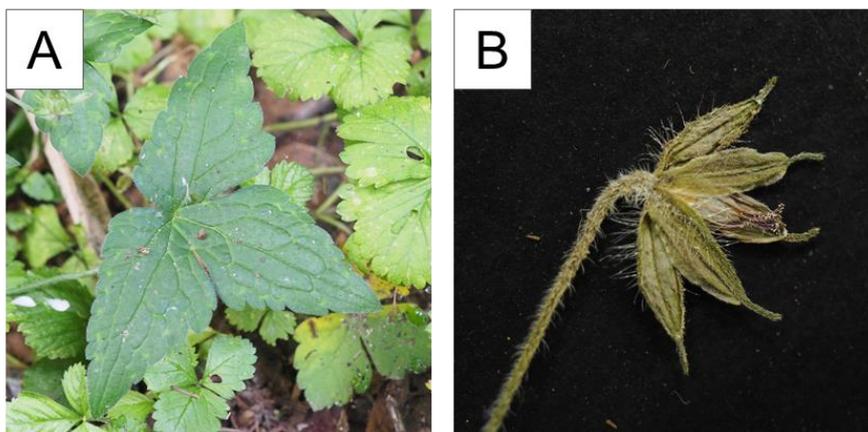
Fig. 5-5. *Geranium hastatum*. A: Leaf. B: Pedicel. (Japan, Tochigi Pref., Nikko, 14 August 2018).

Chapter 6 General Discussion

Demographic histories of Japanese Geranium *species*

      Although grassland species in the Japanese archipelago are believed to have emigrated from Northeastern China and Far East Russia, scientific studies have been too scarce to confirm this pattern from a genetic basis. My population-level genetic study revealed that grassland plants of *G. soboliferum* emigrated to the Japanese Archipelago from continental East Asia during the Last Glacial Maximum. My study is a novel phylogenetic assessment of Japanese grassland species and constitutes the first study elucidating the demographic history of these plants.

      Similar to *G. soboliferum*, the grassland species *Pulsatilla cernua* (Ranunculaceae) apparently originated from continental East Asia and immigrated to the Japanese Archipelago during the last glacial period (Takaishi et al. 2019). *Miscanthus sinensis* (Poaceae), which grows in grasslands, emigrated from mainland China to the Ryukyu Islands and from the Korean Peninsula to Kyushu (Shimono et al. 2013). Takaishi et al. (2019) suggested that populations of *M. sinensis*, which are distributed on the western side of northern Japan, were genetically similar to populations found in Far East Russia. In fact, many alpine species that are currently distributed in central Honshu originated from more northern continental areas (Koidzumi 1919; Hultén 1937). For example, Okaura et al. (2007) suggested that the deciduous broad-leaf tree *Quercus crispula* (Fagaceae) immigrated to the Japanese archipelago from the north (e.g., Sakhalin) and south (e.g., the Korean Peninsula and mainland China). Therefore, the possibility that Japanese grassland plants also migrated from northern areas via the land bridge cannot be ruled out. To assess the likelihood of plants having migrated along this route, a more detailed study using an additional grassland taxon that is distributed from Hokkaido to Kyushu would be necessary.

Japanese alpine species are believed to have originated in Beringia (Koidzumi 1919; Hultén 1937). However, Ikeda et al. (2018) indicated that the alpine species *Phyllodoce aleutica* (Ericaceae) originated in East Asia, rather than Beringia, which contradicts this conventional hypothesis. My study suggests that the alpine species *G. erianthum*, which is distributed around the North Pacific Ocean, underwent a distribution expansion from East Asia to Beringia, and that the current-day Japanese populations originated from continental East Asia. In addition, my results suggest that the alpine endemic *G. yesoense* migrated from more northern areas (e.g., Sakhalin) to central Honshu during an ancient glacial period. The phylogenetic maximum-likelihood (ML) tree also suggested that *G. yesoense* is genetically close to *G. rosthornii*, which is distributed in central and southeastern China. Based on these findings, it is probable that these and other alpine species originated in continental East Asia.

It has been suggested that *G. yesoense* immigrated to central Honshu multiple times, and that re-immigration led to a complex genetic population structure. Recent phylogenetic studies revealed that Japanese alpine plants diverged into two major lineages (i.e., northern and southern lineages) and their distributions fluctuated throughout the past megaannum; the northern and southern populations of *Cardamine nipponica* (Brassicaceae), *Cassiope lycopodioides* (Ericaceae), *Phyllodoce nipponica* (Ericaceae) and *Pedicularis japonica* (Orobanchaceae) diverged from each other approximately 110, 200, 250 and 600–3,800 thousand years ago, respectively (Ikeda et al. 2009a; 2014b; Fujii 2008; Ikeda & Setoguchi 2013a). Therefore, alpine species in Japan have experienced recurrent fluctuations in their geographic distributions. However, empirical studies of the genetic structures of refugial populations remain insufficient. My study is the first to verify the complexity of the genetic structure in a refugial population (i.e., to determine a "refugia within refugia"). A previous study had been unable to provide a detailed demographic history of *G. yesoense* because the authors

applied conventional genetic methods (Wakasugi et al. 2017). Using chloroplast genome and genome-wide SNP analyses, I was able to elucidate phylogenetic relationships, genetic structure, and demographic history within this species. These results highlight the limitations of conventional genetic analyses and the importance of performing a least some analyses using genome data. However, my study focused on one alpine plant species. Many other alpine species are distributed in problematic regions of Mt. Ibuki and Mt. Asama (e.g., *Aletris foliate*, *Leontopodium japonicum*, and *Thymus quinquecostatus*). Broader conclusions on refugia within refugia in the Japanese archipelago can only be achieved through additional phylogeographic studies targeting other alpine taxa.

It is likely that immigration and emigration patterns and timing would be highly species-specific. For example, *Kirengeshoma palmata* (Hydrangeaceae), which is distributed in East China and southern Japan, immigrated from the Japanese archipelago to mainland China via the East China Sea basin during the early-to-mid Pleistocene (Qiu et al. 2009), and this study indicated that some populations, which are now distributed in East Asia, originated in the Japanese archipelago. To understand the demographic history of Japanese flora, more detailed studies need to be conducted across a range of taxa. However, my findings indicate that, at minimum, Japanese *Geranium* species emigrated to the Japanese archipelago from continental East Asia via the Korean Peninsula and Sakhalin during the last glacial period.

Demographic transitions in Geranium

The genus *Geranium* originated from peripheral regions of the Mediterranean Sea; it is believed that species immigrated to North America 20 million years ago (Fiz et al. 2008). It is speculated that *Geranium* species then immigrated into East Asia from the Mediterranean Sea region 5 million years ago (Fiz et al. 2008; Marcussen & Meseguer 2017). My study indicated

that *Geranium* species distributed in East Asia experienced a distribution shift beginning in the mid-Pleistocene (ca. 1 million years ago). There is further evidence that some plant species (including *Geranium*) emigrated from East Asia into North America during the last glacial period (Ikeda et al. 2018). Therefore, present day populations in North America are likely to have originated from two regions (i.e., the Mediterranean Sea and East Asia). Currently, 18 *Geranium* species occur in Canada and the northern United States (Jones & Jones 1943). Considering the demographic history of *Geranium* in North America, it is likely that these species diverged into two lineages (i.e., an older and a newer lineage). My work has contributed to my understanding of the demographic history of *Geranium* species distributed in the Northern Hemisphere.

Conclusion

Many species were believed to have emigrated to the Japanese archipelago from the surrounding regions via land bridges that were accessible during glacial periods. However, recent studies have indicated that East Asia (including the Japanese archipelago), rather than Beringia, was a critical source of refugial populations of alpine plant species, and that refugial populations were fragmented into several complex refugia. Aside from alpine species, Japanese grassland plant species, which are under threat of habitat destruction, are poorly understood in terms of their demographic history. My genetic analyses suggested that the grassland species *G. soboliferum* emigrated from northeastern China and Far East Russia to the Japanese archipelago during the last glacial maximum. My results also provide support for previous evidence that the alpine species *G. erianthum* emigrated from East Asia to Beringia and western North America. Finally, the genetic structure I observed in representative interglacial refugial populations in central Honshu was indicative of fragmentation due to re-immigration from northern areas.

Therefore, some plant species migrated from East Asia, or into the Japanese archipelago from peripheral regions, during glacial periods. Some alpine species migrated southward to the central Honshu area multiple times. It is important to note that both my study and previous research have been limited to a small number of taxa. Therefore, I would aim to perform phylogeographic studies using a greater number of taxa to understand clearly the demographic history of Japanese flora.

References

Abbott RJ, Comes HP (2004) Evolution in the Arctic: a phylogeographic analysis of the circumarctic plant, Saxifraga oppositifolia (Purple saxifrage). *New Phytologist*, 161, 211-224.

Abbott RJ, Smith LC, Milne RI, Crawford RMM, Wolff K, Balfour J (2000) Molecular analysis of plant migration and refugia in the Arctic. *Science*, 289, 1343-1346.

Aedo C (2001) The genus *Geranium* L.(geraniaceae) in North America II. Annual species (parte A). *Anales del Jardín Botánico de Madrid*, 59(1), 3-65.

Aedo C (1998) World checklist of *Geranium* L. (Geraniaceae). *Anales del Jardín Botánico de Madrid*, 56(2).

Aoki K, Tamaki I, Nakao K, Ueno S, Kamijo T, Setoguchi H, Murakami N, Kato M, Tsumura Y (2018). Approximate Bayesian computation analysis of EST-associated microsatellites indicates that the broadleaved evergreen tree *Castanopsis sieboldii* survived the Last Glacial Maximum in multiple refugia in Japan. *Heredity*, 122, 326-340.

Akiyama, S (2001) Geraniaceae. *In* Iwatsuki K, Boufford DE, Ohba H [eds.], Flora of Japan, vol. IIb, 289. Kodansha Ltd., Tokyo, Japan.

Akiyama T, Kawamura K (2007) Grassland degradation in China: Methods of monitoring, management and restoration. *Grassland Science*, 53, 1-17.

Allen GA, Marr KL, McCormick LJ, Hebda RJ (2012) The impact of Pleistocene climate change on an ancient arctic-alpine plant: multiple lineages of disparate history in *Oxyria digyna*. *Ecology and Evolution*, 2, 649-665.

Alpine Garden Society of Tokyo (2004). *Tane kara tanoshimu sanyasou (How to grow wild flower and alpine plants seed)*. Tokyo, JP: Nousan gyoson bunka kyoukai (Rural Culture Association Japan) (in Japanese).

Alsos IG, Engelskjon T, Gielly L, Taberlet P, Brochmann C (2005) Impact of ice ages on circumpolar molecular diversity: insights from an ecological key species. *Molecular Ecology*, 14, 2739-2753.

Alvarez N, Thiel-Egenter C, Tribsch A, Holderegger R, Manel S, Schonswetter P, Taberlet P, Brodbeck S, Gaudeul M, Gielly L, Kupfer P, Mansion G, Negrini R, Paun O, Pellecchia M, Rioux D, Schupfer F, Van Loo M, Winkler M, Gugerli F, IntraBioDiv C (2009) History or ecology? Substrate type as a major driver of patial genetic structure in Alpine plants. *Ecology Letters*, 12, 632-640.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA (2008) Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *Plos One*, 3.

Bakker FT, Vassiliades DD, Morton C, Savolainen V (1998) Phylogenetic relationships of Biebersteinia Stephan (Geraniaceae) inferred from *rbcL* and *atpB* sequence comparisons. *Botanical Journal of the Linnean Society*, 127, 149-158.

Bennet KD, Provan J (2008) What do we mean by 'refugia'? *Quaternary Science Reviews*, 27, 2449-2455.

Bertorelle G, Benazzo A, Mona S (2010) ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Molecular Ecology*, 19, 2609-2625.

Billings WD, Mooney HA (1968) Ecology of arctic and alpine plants. *Biological Reviews of the Cambridge Philosophical Society*, 43, 481-529.

Birks HJB, Willis KJ (2008) Alpines, trees, and refugia in Europe. *Plant Ecology & Diversity*, 1, 147-160.

Bodare S, Tsuda Y, Ravikanth G, Shaanker RU, Lascoux M (2013) Genetic structure and demographic history of the endangered tree species *Dysoxylum malabaricum* (Meliaceae) in Western Ghats, India: implications for conservation in a biodiversity hotspot. *Ecology and Evolution*, 3, 3233-3248.

Bolger A M, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114-2120.

Bortenschlager, S. (1967) Vorläufige Mitteilungen zur Pollenmorphologie in der Familie der Geraniaceen und IHRE Systematische Bedeutung. *Grana Palynologica* 7: 400-468.

Boucher TV, Mead BR (2006) Vegetation change and forest regeneration on the Kenai Peninsula, Alaska following a spruce beetle outbreak, 1987-2000. *Forest Ecology and Management*, 227, 233-246.

Byrd GV (1984) Vascular vegetation of Buldir Island, Aleutian Islands, Alaska, compared to another Aleutian Island. *Arctic*, 37, 37-48.

Carolin RC (1964) The genus *Geranium* L. in the south western Pacific area. *Proceedings of the Linnean Society of New South Wales*, 89: 326−361

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) *Stacks*: an analysis tool set for population genomics. *Molecular Ecology*, 22, 3124-3140.

Chakraborty R, Kimmel M, Stivers DN, Davison LJ, Deka R (1997) Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. *Proceedings of the National Academy of Sciences of the United States of America*, 94, 1041-1046.

Chapin FS, Korner C (1994) Arctic and alpine biodiversity – patterns, causes and ecosystem consequences. *Trends in Ecology & Evolution*, 9, 45-47.

Chen D, Ahlford A, Schnorrer F, Kalchhauser I, Fellner M, Viragh E, Kiss I, Syvanen AC, Dickson BJ (2008) High-resolution, high-throughput SNP mapping in Drosophila melanogaster. *Nature Methods*, 5, 323-329.

Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch SR, Park WD, Ayres N, Cartinhour S (2000) Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). *Theoretical and Applied Genetics*, 100, 713-722.

Choisy M, Franck P, Cornuet JM (2004) Estimating admixture proportions with microsatellites: comparison of methods based on simulated data. *Molecular Ecology*, 13, 955-968.

Clark L V, Jin X, Petersen K K, Anzoua K G, Bagmet L, Chebukin P, Deuter M, Dzyubenko E, Dzyubenko N, Heo K, Johnson D A, Jørgensen U, Kjeldsen J B, Nagano H, Peng J, Sabitov A, Yamada T, Yoo J H, Yu C Y, Long S P, Sacks E J (2019) Population structure of *Miscanthus sacchariflorus* reveals two major polyploidization events, tetraploid-mediated unidirectional

introgression from diploid *M. sinensis*, and diversity centred around the Yellow Sea. *Annals of Botany*, 124(4), 731-748.

Comes HP, Kadereit JW (2003) Spatial and temporal patterns in the evolution of the flora of the European Alpine System. *Taxon*, 52, 451-462.

Conrad DF, Jakobsson M, Coop G, Wen XQ, Wall JD, Rosenberg NA, Pritchard JK (2006) A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. *Nature Genetics*, 38, 1251-1260.

Cornuet JM, Pudlo P, Veyssier J, Dehne-Garcia A, Gautier M, Leblois R, Marin JM, Estoup A (2014) DIYABC 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.

Cosacov A, Sérsic Alicia N, Sosa V, J. Arturo De-Nova Arturo J, Nylinder S, Cocucci Andrea A. New insights into the phylogenetic relationships, character evolution, and phytogeographic patterns of *Calceolaria* (Calceolariaceae). America Journal of Botany, 96, 2240-2255.

Cox, CB, Moore, PD (2000) Biogeography; an Evolutionary and Ecological Approach (6th ed.). Oxford, UK: Blackwell Science.

Davis BM, Shaw GR (2001) Range Shifts and Adaptive Responses to Quaternary Climate Change. *Science*, 292, 673-679.

DeChaine EG (2008). A bridge or a barrier? Beringia's influence on the distribution and diversity of tundra plants. *Plant Ecology & Diversity*, 1(2), 197-207.

Dierckxsens N, Mardulyn P, Smits G (2017) NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Research*, 45.

Dobson M (1994) Patterns of distribution in Japanese land mammals. *Mammal Review*, 24, 91-111.

Dodd CK (1990) Effects of habitat fragmentation on a stream-dwelling species, the flattened musk turtle *Sternotherus depressus*. *Biological Conservation*, 54, 33-45.

Ehrich D, Gaudeul M, Assefa A, Koch MA, Mummenhoff K, Nemomissa S, Consortium I, Brochmann C (2007) Genetic consequences of Pleistocene range shifts: contrast between the Arctic, the Alps and the East African mountains. *Molecular Ecology*, 16, 2542-2559.

Eidesen PB, Alsos IG, Popp M, Stensrud O, Suda J, Brochmann C (2007) Nuclear vs. plastid data: complex Pleistocene history of a circumpolar key species. *Molecular Ecology*, 16, 3902-3925.

Emerson BC, Kolm N (2005) Species diversity can drive speciation. *Nature*, 434, 1015-1017.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, 14, 2611-2620.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, 164, 1567-1587.

Fiz O, Vargas P, Alarcon M, Aedo C, Garcia JL, Aldasoro JJ (2008) Phylogeny and historical biogeography of Geraniaceae in relation to climate changes and pollination ecology. *Systematic Botany*, 33, 326-342.

Frankham R (1996) Relationship of genetic variation to population size in wildlife. *Conservation Biology*, 10, 1500-1508.

Frankham R, Ballou JD, Briscoe DA (2002) Introduction to Conservation Genetics. Introduction to Conservation Genetics, i-xxi, 1-617.

Fujii N (2008) Phylogeographic importance of the mountainous region of central Honshu for Japanese alpine plants. *Bunrui* 8(1): 5-14. (in Japanese)

Fujii N, Takasawa S, Yokogawa M, Yamasaki T, Harada K, Kaneko S, Isagi Y (2016) Novel Nuclear Microsatellite Markers Reveal Genetic Diversity and Structure of *Veronicastrum sibiricum* var. *zuccarinii* (Plantaginaceae) in the Aso Region, Kyushu, Japan. *Acta Phytotaxonomica et Geobotanica*, 67, 69-82.

Fujii N, Ueda K, Watano Y, Shimizu T (1997) Intraspecific sequence variation of chloroplast DNA in *Pedicularis chamissonis* Steven (Scrophulariaceae) and geographic structuring of the Japanese "Alpine" plants. *Journal of Plant Research*, 110, 195-207.

Fujii N, Ueda K, Watano Y, Shimizu T (1999) Further analysis of intraspecific sequence variation of chloroplast DNA in *Primula cuneifolia* Ledeb. (Primulaceae): Implications for biogeography of the Japanese alpine flora. *Journal of Plant Research*, 112, 87-95.

Gabrielsen TM, Bachmann K, Jakobsen KS, Brochmann C (1997) Glacial survival does not matter: RAPD phylogeography of Nordic *Saxifraga oppositifolia*. *Molecular Ecology*, 6, 831-842.

Gao H, Williamson S, Bustamante CD (2007) A Markov chain Monte Carlo approach for joint inference of population structure and inbreeding rates from multilocus genotype data. *Genetics*, 176, 1635-1651.

Gibbs JP (2000) Wetland loss and biodiversity conservation. *Conservation Biology*, 14, 314-317.

Gibbs RA, Taylor JF, Van Tassell CP, Barendse W, Eversoie KA, Gill CA, Green RD, Hamernik DL, Kappes SM, Lien S, Matukumalli LK, McEwan JC, Nazareth LV, Schnabel RD, Weinstock GM, Wheeler DA, Ajmone-Marsan P, Boettcher PJ, Caetano AR, Garcia JF, Hanotte O, Mariani P, Skow LC, Williams JL, Diallo B, Hailemariam L, Martinez ML, Morris CA, Silva LOC, Spelman RJ, Mulatu W, Zhao KY, Abbey CA, Agaba M, Araujo FR, Bunch RJ, Burton J, Gorni C, Olivier H, Harrison BE, Luff B, Machado MA, Mwakaya J, Plastow G, Sim W, Smith T, Sonstegard TS, Thomas MB, Valentini A, Williams P, Womack J, Wooliams JA, Liu Y, Qin X, Worley KC, Gao C, Jiang HY, Moore SS, Ren YR, Song XZ, Bustamante CD, Hernandez RD, Muzny DM, Patil S, Lucas AS, Fu Q, Kent MP, Vega R, Matukumalli A, McWilliam S, Sclep G, Bryc K, Choi J, Gao H, Grefenstette JJ, Murdoch B, Stella A, Villa-Angulo R, Wright M, Aerts J, Jann O, Negrini R, Goddard ME, Hayes BJ, Bradley DG, da Silva MB, Lau LPL, Liu GE, Lynn DJ, Panzitta F, Dodds KG (2009) Genome-Wide Survey of SNP Variation Uncovers the Genetic Structure of Cattle Breeds. *Science*, 324, 528-532.

Gomez A, Lunt DH (2007) Refugia within refugia: patterns of phylogeographic concordance in the Iberian Peninsula. Phylogeography of Southern European Refugia: Evoluationary Perspective on the Origins and Conservation of European Biodiversity, 155-188.

Gotanda K, Nakagawa T, Tarasov P, Kitagawa J, Inoue Y, Yasuda Y (2002) Biome classification from Japanese pollen data: application to modern-day and late Quaternary samples. *Quaternary Science Reviews*, 21, 647-657.

Goudet J (1995) FSTAT (Version 1.2): A computer program to calculate F-statistics. *Journal of Heredity*, 86, 485-486.

Goudet J (2005) HIERFSTAT, a package for R to compute and test hierarchical *F*-statistics. *Molecular Ecology Notes*, 5, 184-186.

Guisinger MM, Kuehl JV, Boore JL, Jansen RK (2011) Extreme Reconfiguration of Plastid Genomes in the Angiosperm Family Geraniaceae: Rearrangements, Repeats, and Codon Usage. *Molecular Biology and Evolution*, 28, 1543-1543.

Hara H (1948) Notes on the Japanese species of *Geranium* (in Japanese). *Journal of Japanese Botany*, 22: 21–28

Harrison SP, Yu G, Takahara H, Prentice IC (2001) Palaeovegetation - Diversity of temperate plants in east Asia. *Nature*, 413, 129-130.

Hata D, Higashi H, Yakubov V, Barkalov V, Ikeda H, Setoguchi H (2017) Phylogeographical insight into the Aleutian flora inferred from the historical range shifts of the alpine shrub *Therorhodion camtschaticum* (Pall.) Small (Ericaceae). *Journal of Biogeography*, 44, 283-293.

Heusser LE, Morley JJ (1985) Pollen and radiolarian records from deep-sea core RC-14-103 - climate reconstructions of northeast Japan and northwest pacific for the last 90,000 years. *Quaternary Research*, 24, 60-72.

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, 58, 247-276.

Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. *Nature*, 405, 907-913.

Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *PNAS*, 104(8), 2785-2790.

Hirano T, Saito T, Tsunamoto Y, Koseki J, Ye B, Do VT, Miura O, Suyama Y, Chiba S (2019) Enigmatic incongruence between mtDNA and nDNA revealed by multi-locus phylogenomic analyses in freshwater snails. *Scientific Reports*, 9.

Holderegger R, Thiel-Egenter C (2009) A discussion of different types of glacial refugia used in mountain biogeography and phylogeography. *Journal of Biogeography*, 36, 476-480.

Hultén E (1937) Outline of the history of arctic and boreal biota during the Quaternary period: their evolution during and after the glacial period as indicated by the equiformal progressive areas of present plant species. Stockholm: Thule.

Ikeda H, Fujii N, Setoguchi H (2009a) Application of the Isolation with Migration Model Demonstrates the Pleistocene Origin of Geographic Differentiation in *Cardamine nipponica* (Brassicaceae), an Endemic Japanese Alpine Plant. *Molecular Biology and Evolution*, 26, 2207-2216.

Ikeda H, Yakubov V, Barkalov V, Setoguchi H (2014a) Molecular evidence for ancient relicts of arctic-alpine plants in East Asia. *New Phytologist* 203: 980-988.

Ikeda H, Higashi H, Yakubov V, Barkalov V, Setoguchi H (2014b) Phylogeographical study of the alpine plant *Cassiope lycopodioides* (Ericaceae) suggests a range connection between the Japanese archipelago and Beringia during the Pleistocene. *Biological Journal of the Linnean Society*, 113, 497-509.

Ikeda H, Senni K, Fujii N, Setoguchi H (2008a) Consistent geographic structure among multiple nuclear sequences and cpDNA polymorphisms of *Cardamine nipponica* Franch. et Savat. (Brassicaceae). *Molecular Ecology*, 17, 3178-3188.

Ikeda H, Senni K, Fujii N, Setoguchi H (2008b) Post-glacial range fragmentation is responsible for the current distribution of *Potentilla matsumurae* Th. Wolf (Rosaceae) in the Japanese archipelago. *Journal of Biogeography*, 35, 791-800.

Ikeda H, Senni K, Fujii N, Setoguchi H (2008c) Survival and genetic divergence of an arctic-alpine plant, *Diapensia lapponica* subsp *obovata* (Fr. Schm.) Hulten (Diapensiaceae), in the high mountains of central Japan during climatic oscillations. *Plant Systematics and Evolution*, 272, 197-210.

Ikeda H, Senni K, Fujii N, Setoguchi H (2009b) High mountains of the Japanese archipelago as refugia for arctic-alpine plants: phylogeography of *Loiseleuria procumbens* (L.) Desvaux (Ericaceae). *Biological Journal of the Linnean Society*, 97, 403-412.

Ikeda H, Setoguchi H (2007) Phylogeography and refugia of the Japanese endemic alpine plant, *Phyllodoce nipponica* Makino (Ericaceae). *Journal of Biogeography*, 34, 169-176.

Ikeda H, Setoguchi H (2013a) A multilocus sequencing approach reveals the cryptic phylogeographical history of *Phyllodoce nipponica* Makino (Ericaceae). *Biological Journal of the Linnean Society*, 110, 214-226.

Ikeda, H., and Setoguchi, H. (2013b) Phylogeographic study of *Phyllodoce aleutica* (Ericaceae) in the Japanese Archipelago. *Bulletin of the National Museum of Nature and Science Series B*, 39: 87-94.

Ikeda H, Shimizu A, Aedo C (2015) Nomenclature and typification of *Geranium yesoense* var. *pseudo-palustre* (Geraniaceae). *Journal of Japanese Botany*, 90: 281–284.

Ikeda H, Yakubov V, Barkalov V, Setoguchi H (2018) Post-glacial East Asian origin of the alpine shrub *Phyllodoce aleutica* (Ericaceae) in Beringia. *Journal of Biogeography*, 45, 1261-1274.

Iversen, J. (1942) En pollenanalytisk tidsfæstelse af ferskvandslagene ved Nørre Lyngby. *Meddelelser fra Dansk Geologisk Forening*, 10: 130–151.

Iversen, J (1954) The Late-glacial flora of Denmark and its relation to climate and soil. *Danmarks Geologiske Undersøgelse II Række*, 80: 87–119.

Jakobsson M, Scholz SW, Scheet P, Gibbs JR, VanLiere JM, Fung HC, Szpiech ZA, Degnan JH, Wang K, Guerreiro R, Bras JM, Schymick JC, Hernandez DG, Traynor BJ, Simon-Sanchez J, Matarin M, Britton A, van de Leemput J, Rafferty I, Bucan M, Cann HM, Hardy JA, Rosenberg NA, Singleton AB (2008) Genotype, haplotype and copy-number variation in worldwide human populations. *Nature*, 451, 998-1003.

Jansen JHF, Kuijpers A, Troelstra SR (1986) A mid-Brunthes climatic event long term changes in global atmosphere and ocean circulation. *Science*, 232, 619-622.

Jeffries DL, Copp GH, Handley LL, Olsen KH, Sayer CD, Hanfling B (2016) Comparing RADseq and microsatellites to infer complex phylogeographic patterns, an empirical perspective in the Crucian carp, *Carassius carassius*, L. *Molecular Ecology*, 25, 2997-3018.

Jones Neville G, Jones Freeman F (1943) A revision of the perennial species of *Geranium* of the United States and Canada. *Rhodora*, 45, 5-26

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Mammalian Protein Metabolism (ed. Munro HN), pp. 21-132. Academic Press, New York.

Kadota Y (2016) Geraniaceae. *In* Wild flowers of Japan, Vol.3 (in Japanese). Heibonsha, Tokyo, Japan.

Kandori I (2002) Diverse visitors with various pollinator importance andtemporal change in the important pollinators of *Geranium thunbergii* (Geraniaceae). *Ecological Research*, 17, 283-294.

Kaneko S, Sei S, Takahashi Y, Isagi Y (2009) The current status of *Echinops setifer* Iljin (Asteraceae) on Aso Mountain and Chugoku Mountain regions (in Japanese). *Japanese journal of conservation ecology* 14: 125-130.

Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, 30, 3059-3066.

Katoh K, Standley DM (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, 30, 772-780.

Kery M, Matthies D, Spillmann HH (2000) Reduced fecundity and offspring performance in small populations of the declining grassland plants *Primula veris* and *Gentiana lutea*. *Journal of Ecology*, 88, 17-30.

Knuth R (1912) *Geranium* L. in Das Pflanzenreich IV.129 (Heft 53) ed. A. Engler. Leipzig: Engelmann, pp 43–221, 575–583

Koidzumi G. 1919. Genetic and floristic phytogeographyof the alpine flora of Japan. *Botanical Magazine*, 33, 193–222.

Korotky A, Grebennikova T, Razjigaeva N, Volkov V, Mokhova L, Ganzey L, Bazarova V (1997) Marine terraces of western Sakhalin island. *Catena*, 30, 61-81.

Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH (2005) Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 8369-8374.

Kurata S, Sakaguchi S, Ito M (2017) Development of nuclear microsatellite markers for the threatened wetland plant *Geranium soboliferum* var. *kiusianum* (Geraniaceae). *Plant Species Biology*, 32, 466-470.

Lassmann T, Hayashizaki Y, Daub CO (2009) TagDust-a program to eliminate artifacts from next generation sequencing data. *Bioinformatics*, 25, 2839-2840.

Li DZ, Gao LM, Li HT, Wang H, Ge XJ, Liu JQ, Chen ZD, Zhou SL, Chen SL, Yang JB, Fu CX, Zeng CX, Yan HF, Zhu YJ, Sun YS, Chen SY, Zhao L, Wang K, Yang T, Duan GW, China Plant BOLG (2011) Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants.

Proceedings of the National Academy of Sciences of the United States of America, 108, 19641-19646.

Liu Z, Chen Z, Pan J, Li X, Su M, Wang L, Li H, Liu G (2008) Phylogenetic relationships in *Leymus* (Poaceae: Triticeae) revealed by the nuclear ribosomal internal transcribed spacer and chloroplast *trnL-F* sequences. *Molecular Phylogenetics and Evolution*, 46, 278-289.

Marcussen T, Mesenguer Andrea S (2017) Species-level phylogeny, fruit evolution and diversification history of *Geranium* (Geraniaceae). *Molecular Phylogenetics and Evolution*, 110, 134-149

Marr KL, Allen GA, Hebda RJ (2008) Refugia in the Cordilleran ice sheet of western North America: chloroplast DNA diversity in the Arctic-alpine plant *Oxyria digyna*. *Journal of Biogeography*, 35, 1323-1334.

Marr KL, Allen GA, Hebda RJ, McCormick LJ (2013) Phylogeographical patterns in the widespread arctic-alpine plant *Bistorta vivipara* (Polygonaceae) with emphasis on western North America. *Journal of Biogeography*, 40, 847-856.

Matsunaga Y (2005) Decrease of grassland during the past 100 years in Aso, Kyushu (in Japanese). *Proceedings of the General Meeting of the Association of Japanese Geographers*, 67, 199.

McKay BD (2012) A new timeframe for the diversification of Japan's mammals. *Journal of Biogeography*, 39, 1134-1143.

Milligan, B. (1992) Plant DNA isolation. In: Hoelzel AR (ed) Molecular Genetic Analysis of Populations: a Practical Approach. IRL Press, Oxford, pp 59-88

Milne RI (2006) Northern hemisphere plant disjunctions: A window on tertiary land bridges and climate change? *Annals of Botany*, 98, 465-472.

Ministry of the Environment of japan. (2012) Red List of threatened wild life of japan, Vascular Plants, revised edition. Ministry of the Environment of Japan (in Japanese). https://ikilog.biodic.go.jp/Rdb/booklist. Accessed 22 February 2018

Ministry of the Environment of japan. (2015) Red List of threatened wild life of japan, Vascular Plants, revised edition. Ministry of the Environment of Japan (in Japanese). https://ikilog.biodic.go.jp/Rdb/booklist. Accessed 22 February 2018

Ministry of the Environment of japan. (2017) Red List of threatened wild life of japan, Vascular Plants, revised edition. Ministry of the Environment of Japan (in Japanese). https://ikilog.biodic.go.jp/Rdb/booklist. Accessed 22 February 2018

Ministry of the Environment of Japan (2015) Red data book 2014. Vascular Plants, third edition (in Japanese). Gyousei, Tokyo.

Mitchell AD, Heenan PB, Paterson AM (2009) Phylogenetic relationships of Geranium species indigenous to New Zealand. *New Zealand Journal of Botany*, 47, 21-31.

Møller TR, Rørdam CP (1985) Species numbers of vascular plants in relation to area, isolation and age of ponds in Denmark. *Oikos*, 45, 8-16.

Murata G (1988) Flora and its characteristic in Japan 17: The distribution on the continental elements in Japan (in Japanese). *Nihon No Seibutsu*, 2, 21-25.

Nakai T (1909) Aliquot Nova Plantae ex Asia orientale. *The botanical magazine*, 23(268), 99-108.

Nei M (1972) Genetic distance between populations. *American Naturalist*, 106.

Nei M (1987) Molecular evolutionary genetics. *Molecular evolutionary genetics*, i-x, 1-512.

Nishida S, Takakura K, Naiki A, Nishida T (2020) Habitat partitioning in native *Geranium* species through reproductive interference. *Annals of Botany*, 125(4), 651-661.

Ogura J (2006) The transition of grassland area in Japan (in Japanese). J. Kyoto Seika Univ., 30, 160-172.

Ohwi J (1965) Flora of Japan Shibundo, Tokyo, Japan

Okaura T, Harada K (2002) Phylogeographical structure revealed by chloroplast DNA variation in Japanese Beech (Fagus crenata Blume). *Heredity*, 88, 322-329.

Okaura T, Quang DN, Ubukata M, Harada K. (2007) Phylogeographic structure and late Quaternary population history of the Japanese oak *Quercus mongolica* var. *crispula* and related species revealed by chloroplast DNA variation. *Genes & Genetic Systems*, 82, 465-477

Okubo K. (2002) The Present State in the Study of Biological Diversity on Semi-natural Grassland in Japan (Activities of Biodiversity Conservation on Semi-natural Grassland in Japan) (in Japanese). *Japanese Journal of Grassland Science*, 48: 268-276.

Okuda M, Nakazato H, Miyoshi N, Nakagawa T, Okazaki H, Saito S, Taira A (2006) MIS11-19 pollen stratigraphy from the 250-m Choshi core, northeast Boso Peninsula, central Japan: Implications for the early/mid-Brunhes (400-780 ka) climate signals. *Island Arc*, 15, 338-354.

Ota, Y, Machida, H (1987) Quaternary sea-level changes in Japan. In: Sea-Level Changes (ed. by M.J. Tooley & 1. Shennan), The Institute of British Geographers Special Publications Series, Basil Blackwell,Oxford, No. 20, 182-224.

Pan J, Zhang D, Sang T (2007) Molecular phylogenetic evidence for the origin of a diploid hybrid of *Paeonia* (Paeoniaceae). *American Journal of Botany*, 94(3), 400-408.

Pax DL, Price RA, Michaels HJ (1997) Phylogenetic position of the Hawaiian geraniums based on *rbcL* sequences. *American Journal of Botany*, 84, 72-78.

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, 6, 288-295.

Price RA, Palmer JD (1993) Phylogenetic relationships of the Generaniales from *rbcL* sequence comparisons. *Annals of the Missouri Botanical Garden*, 80, 661-671.

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

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, 81, 559-575.

Qi XS, Yuan N, Comes HP, Sakaguchi S, Qiu YX (2014) A strong 'filter' effect of the East China Sea land bridge for East Asia's temperate plant species: inferences from molecular

phylogeography and ecological niche modelling of *Platycrater arguta* (Hydrangeaceae). *BMC Evolutionary Biology*, 14.

Qiu XY, Sun Y, Zhang X, Lee J, Fu C, Comes PH (2009) Molecular phylogeography of East Asian *Kirengeshoma* (Hydrangeaceae) in relation to Quaternary climate change. *New Phytologist* 183, 480-495 and landbridge configurations

Raymond M, Rousset F (1995) Genpop (version-1.2) – population-genetics software for exact tests and ecumenicism. *Journal of Heredity*, 86, 248-249.

Razjigaeva NG, Ganzey LA, Belyanina NI, Grebennikova TA, Ganzey KS (2008) Paleoenvironments and landscape history of the Minor Kuril Islands since the Late Glacial. *Quaternary International*, 179, 83-89.

Rind D, Peteet D (1985) Terrestrial conditions at the last glacial maximum and climap sea-surface temperature estimates: Are they consistent. *Quaternary Research*, 24, 1-22.

Robertson (1972) The genera of the Geraniaceae in the southeastern United States. *Journal of the Arnold Arboretum*, 53: 182-201.

Ronikier M, Cieslak E, Korbecka G (2008) High genetic differentiation in the alpine plant *Campanula alpina* Jacq. (Campanulaceae): evidence for glacial survival in several Carpathian regions and long-term isolation between the Carpathians and the Alps. *Molecular Ecology*, 17, 1763-1775.

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

Sakaguchi S, Bowman DMJS, Prior LD, Crisp MD, Linde CC, Tsumura Y, Isagi Y (2013) Climate, not Aboriginal landscape burning, controlled the historical demography and distribution of fire-sensitive conifer populations across Australia. *Proceedings of the Royal Society B-Biological Sciences*, 280.

Sei S, Takazawa S, Fujii N (2015) The current status of the grassland plants and restoration activity in the Aso region, Kyushu, Japan (in Japanese). *Bunrui* 15: 21-27.

Senni K, Fujii N, Takahashi H, Sugawara T, Wakabayashi M. (2005) Intraspecific Chloroplast DNA Variations of the Alpine Plants in Japan. *Acta Phytotaxonomica et Geobotanica*, 56(3), 265-275.

Schlötterer C (2004) The evolution of molecular markers - just a matter of fashion? *Nature Reviews Genetics*, 5, 63-69.

Schlötterer C, Ritter R, Harr B, Brem G (1998) High mutation rate of a long microsatellite allele in Drosophila melanogaster provides evidence for allele-specific mutation rates. *Molecular Biology and Evolution*, 15, 1269-1274.

Schönswetter P, Stehlik I, Holderegger R, Tribsch A (2005) Molecular evidence for glacial refugia of mountain plants in the European Alps. *Molecular Ecology*, 14, 3547-3555.

Schönswetter P, Tribsch A, Stehlik I, Niklfeld H (2004) Glacial history of high alpine *Ranunculus glacialis* (Ranunculaceae) in the European Alps in a comparative phylogeographical context. *Biological Journal of the Linnean Society*, 81, 183-195.

Semerikov VL, Semerikova SA, Polezhaeva MA, Kosintsev PA, Lascoux M (2013) Southern montane populations did not contribute to the recolonization of West Siberian Plain by Siberian larch (*Larix sibirica*): a range-wide analysis of cytoplasmic markers. *Molecular Ecology*, 22, 4958-4971.

Shimizu T (1982) The New Alpine Flora of Japan in Color, Vol. 1. Hoikusha, Osaka.

Shimizu T (1983) The New Alpine Flora of Japan in Color, Vol. 2. Hoikusha, Osaka.

Shimono Y, Kurokawa S, Nishida T, Ikeda H, Futagami N (2013) Phylogeography based on intraspecific sequence variation in chloroplast. *Botany*, 91, 449-456

DNA of Miscanthus sinensis (Poaceae), a native pioneer grass in Japan

Sjögren P (1991) Extinction and local isolation Gradients in metapopulations: The case of the pool frog (Ranalessonae). *Biological Journal of the Linnean Society*, 42, 135-147.

Skrede I, Borgen L, Brochmann C (2009) Genetic structuring in three closely related circumpolar plant species: AFLP versus microsatellite markers and high-arctic versus arctic-alpine distributions. *Heredity*, 102, 293-302.

Skrede I, Eidesen PB, Portela RP, Brochmann C (2006) Refugia, differentiation and postglacial migration in arctic-alpine Eurasia, exemplified by the mountain avens (*Dryas octopetala* L.). *Molecular Ecology*, 15, 1827-1840.

Slatkin M (1993) Isolation by distance in equilibrium and nonequilibrium populations. *Evolution*, 47, 264-279.

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30, 1312-1313.

Stehlik I (2002) Glacial history of the alpine herb *Rumex nivalis* (Polygonaceae): A comparison of common phylogeographic methods with nested clade analysis. American Journal of Botany, 89, 2007-2016.

Stehlik I, Blattner FR, Holderegger R, Bachmann K (2002) Nunatak survival of the high Alpine plant *Eritrichium nanum* (L.) Gaudin in the central Alps during the ice ages. *Molecular Ecology*, 11, 2027-2036.

Stevens MI, Hogg ID (2003) Long-term isolation and recent range expansion from glacial refugia revealed for the endemic springtail *Gomphiocephalus hodgsoni* from Victoria Land, Antarctica. *Molecular Ecology*, 12, 2357-2369.

Stewart JR, Lister AM, Barnes I, Dalen L (2010) Refugia revisited: individualistic responses of species in space and time. *Proceedings of the Royal Society B-Biological Sciences*, 277, 661-671.

Suka T (2008) Historical changes of semi-natural grasslands in the central mountainous area of Japan and their implications for conservation of grassland species (in Japanese). *Bulletin of Nagano Environmental Conservation Research Institute*, 4, 17-31.

Sun Y, Surget-Groba Y, Gao SX (2016) Divergence maintained by climatic selection despite recurrent gene flow: a case study of *Castanopsis carlesii* (Fagaceae). *Molecular Ecology*, 25, 4580-4592.

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

Tabata H (1997) Satoyama and its conservation (in Japanese). Hoikusha, Osaka

Tabata (2000) The Japanese vegetation zone is wrong. *Kagaku*, 70(5), 421-430

Taberlet P, Cheddadi R (2002) Quaternary Refugia and Persistence of Biodiversity. *Science*, 297(5589), 2009-2010.

Takahara H (2009) Fire and vegetation history since the last glacial period in Japan (in Japanese). *Shinrin Kagaku*, 55, 10-13.

Takahara H, Kitagawa H (2000) Vegetation and climate history since the last interglacial in Kurota Lowland, western Japan. Palaeogeography Palaeoclimatology Palaeoecology, 155, 123-134.

Takahashi Y (2009) Management and restoration of grassland landscape for species conservation: a case of Aso Grassland (in Japanese). *The Japanese Institute of Landscape Architecture*, 72, 394-398.

Takahashi Y (2013) Collaborative Activities for Conservation and Restoration of Aso Grassland (in Japanese). *Ohara Institute for Social Research*, 655, 3-18.

Takahashi Y, Suyama Y, Matsuki Y, Funayama R, Nakayama K, Kawata M (2016) Lack of genetic variation prevents adaptation at the geographic range margin in a damselfly. *Molecular Ecology*, 25, 4450-4460.

Takaishi A, Andrey E. Kozhevnikov AE, Kozhevnikova ZV, Ikeda H, Fujii F, Soejima A (2019) Phylogeography of *Pulsatilla cernua* (Ranunculaceae) a grassland species in Japan. *Ecology and Evolution*, 9(12), 7262-7272.

Talbot SS, Schofield WB, Talbot SL, Daniels FJA (2010) Vegetation of eastern Unalaska Island, Aleutian Islands, Alaska. *Botany-Botanique*, 88, 366-388.

Talbot SS, Talbot SL (1994) Numerical classification of the coastal vegetation of Attu island, Aleutian Islands, Alaska. *Journal of Vegetation Science*, 5, 867-876.

Tanabe AS (2008). Phylogears. Ver. 2.0. Available at: https://www.fifthdimension.jp/products/

Tarasov PE, Nakagawa T, Demske D, Osterle H, Igarashi Y, Kitagawa J, Mokhova L, Bazarova V, Okuda M, Gotanda K, Miyoshi N, Fujiki T, Takemura K, Yonenobu H, Fleck A (2011) Progress in the reconstruction of Quaternary climate dynamics in the Northwest Pacific: A new modern analogue reference dataset and its application to the 430-kyr pollen record from Lake Biwa. *Earth-Science Reviews*, 108, 64-79.

Tatewaki M (1974) The Geobotanical Relationship between Beringia and Northern Japan, with Special Reference to the Arctic-Alpine Flora of the Latter. *Journal of the Faculty of Agriculture, Hokkaido University*, 57(3), 340-348

Tatewaki M, Kobayashi Y (1934) A contribution to the flora of the Aleutian Islands. *Journal of the Faculty of Agriculture, Hokkaido Imperial University*, 36(1), 1-119

Tokarski M (1972) Morphological and taxonomical analysis of fruits and seeds of the European and Caucasian species of the genus *Geranium* L. *Monographiae Botanicae*, 36.

Tollefsrud MM, Bachmann K, Jakobsen KS, Brochmann C (1998) Glacial survival does not matter - II: RAPD phylogeography of Nordic *Saxifraga cespitosa*. *Molecular Ecology*, 7, 1217-1232.

Tralau, H (1963) The recent and fossil distribution of some boreal and arctic montane plants in Europe. *Arkiv für Botanik Series 2*, 5, 533–582.

Tremblay NO, Schoen DJ (1999) Molecular phylogeography of *Dryas integrifolia*: glacial refugia and postglacial recolonization. *Molecular Ecology*, 8, 1187-1198.

Tsukada M (1988) Glacial and Holocene vegetation history of Japan. Vegetation History, In Huntley Handbook of Vegetation Science 7, Kluwer, Dordrecht, 459-518.

Tsukaya H, Tsuge T (2001) Morphological adaptation of inflorescences in plants that develop at low temperatures in early spring: The convergent evolution of "downy plants". *Plant Biology*, 3, 536-543.

Ushimaru A, Uchida K, Suka T (2018) Grassland Biodiversity in Japan: Threats, Management and Conservation. Grasslands of the World, 1st edn. CRC Press, Boca Raton, 197-220.

Vargas P (2003) Molecular evidence for multiple diversification patterns of alpine plants in Mediterranean Europe. *Taxon*, 52, 463-476.

Vigouroux Y, Jaqueth JS, Matsuoka Y, Smith OS, Beavis WF, Smith JSC, Doebley J (2002) Rate and pattern of mutation at microsatellite loci in maize. *Molecular Biology and Evolution*, 19, 1251-1260.

Wakasugi Y, Azuma H, Naiki A, Nishida S (2017) Morphological and Molecular Phylogenetic Analyses of *Geranium yesoense* (Geraniaceae) in Japan. *Acta Phytotaxonomica Et Geobotanica*, 68, 129-144.

Warburg E. F. (1938) Taxonomy and Relationship in the Geraniales in the Light of their Cytology. *New Phytologist*, 37, 130-159

Watanabe Y, Kawamata I, Matsuki Y, Suyama Y, Uehara K, Ito M (2018) Phylogeographic analysis suggests two origins for the riparian azalea *Rhododendron indicum* (L.) Sweet. *Heredity*, 121, 594-604.

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

Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics*, 163, 1177-1191.

Winkworth RC, Wagstaff SJ, Glenny D, Lockhart PJ (2005) Evolution of the New Zealand mountain flora: Origins, diversification and dispersal. *Organisms Diversity & Evolution*, 5, 237-247.

Wolfe KH, Li WH, Sharp PM (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences of the United States of America*, 84, 9054-9058.

Wright S (1943) Isolation by distance. *Genetics*, 28, 114-138.

Xu L, Aedo M (2008) Flora of China, Vol.11. Science Press, Beijing, and Missouri Botanical Garden Press, St. Louis

Yamasaki T, Ozeki K, Fujii N, Takehara M, Yokogawa M, Kaneko S, Isagi Y (2013) Genetic Diversity and Structure of *Silene kiusiana* (Caryophyllaceae) in the Aso Region, Kyushu, Japan, Revealed by Novel Nuclear Microsatellite Markers. *Acta Phytotaxon Geobot* 63: 107-120.

Yamazaki T (1993) On *Geranium hastatum* Nakai and G. wilfordii Maxim. var. *chinense* (Migo) Hara (in Japanese). *Joumal of Japanese Botany*, 68(4), 237-239.

Yasunaka Y, Oishi K, Anzai H, Miwa M, Kumagai H, Hiraoka H, Ieiri S. (2015) Assessing the effect of controlled burning and grazing on vegetation change in the grasslands of Aso region using satellite image analyses. *Japanese Agricultural Systems Society* 31, 117-125.

Ye JW, Guo XD, Wang SH, Bai WN, Bao L, Wang HF, Ge JP (2015) Molecular evidence reveals a closer relationship between Japanese and mainland subtropical specimens of a widespread tree species, Acer mono. *Biochemical Systematics and Ecology*, 60, 143-149.

Yeo PF (1973) The biology and systematics of *Geranium*, sections *Anemonifolia* Knuth and *Ruberta* Dum. *Botanical Journal of the Linnean Society*, 67, 285-346.

Yeo PF (1984) Fruit-discharge-type in *Geranium* (Geraniaceae): its in classification and its evolutionary implications. *Botanical Journal of the Linnean Society*, 89, 1-36.

Yin QZ, Berger A (2010) Insolation and CO_2 contribution to the interglacial climate before and after the Mid-Brunhes Event. *Nature Geoscience*, 3, 243-246.

Yokogawa M, Kaneko S, Takahashi Y, Isagi Y (2013) Genetic consequences of rapid population decline and restoration of the critically endangered herb *Polemonium kiushianum*. *Biological Conservation*, 157, 401-408.

Yonekura K (2005) Taxonomic Notes on Vascular Plants in Japan and Its Adjacent Regions (1). New Combinations and New Names of Japanese Plants. *Journal of Japanese Botany*, 80, 323-333.

Zhang LB, Comes HP, Kadereit JW (2001) Phylogeny and quaternary history of the European montane/alpine endemic Soldanella (Primulaceae) based on ITS and AFLP variation. *American Journal of Botany*, 88, 2331-2345.

Zhou W, Yang H, Huang L, Chen C, Lin X, Hu Z, Li J (2017) Grassland degradation remote sensing monitoring and driving factors quantitative assessment in China from 1982 to 2010. *Ecological Indicators*, 83, 303-313.

Supplementary (Chapter 2)

The prior distributions of each parameter in ABC analysis

The prior distribution of each parameter was set as follows; N1 (soboliferum), N2 (hakusanense), N3 (kiusianum), N_{anc1}, N_{anc2} and N_{anc3} of effective population size and t_a, t_b, t_c and tm of time parameters were assumed to have a uniform distribution [10, 10,000]. We assumed the parameter of the admixture rate (ra) in scenario 5 to also have a uniform distribution [0.001, 0.999]. Furthermore, we assumed the relationships between population size parameters, as N_{anc1} ≥ N_{anc2}, N_{anc3} and N_{anc1} ≥ N1 and N_{anc2} ≥ N2 and N_{anc3} ≥ N3 and N1 ≥ N2, N3. Following Cornuet et al. (2014), when allelic length exceeded 40 motifs we took (Maximum_allele_size – Minimum_allele_size) / motif size + ~10 additional motifs to re-define the allelic range of the locus (e.g., (200 nu – 100 nu) / 2 + 10 = 50 + 10 = 60). Therefore, we defined the allele range of three loci (VKGP_15780, VKGP_28824, VKGP_16051) as > 40 motifs: 38, 38 and 34, respectively. The allelic ranges for the other seven loci were defined as 40, sine their allelic range is < 40 motifs. We used EST-SSR markers in this study (ESTs; expressed sequence tags), which are believed to have lower mutation rates than anonymous SSRs, since they are located at coding regions. Additionally, a difference in mutation rate between di- and trinucleotide SSR motifs was observed in several species; namely, the mutation rate of dinucleotide is higher than that of trinucleotide (Chakraborty et al. 1997; Cho et al. 1999). Therefore, we set a different mutation rate between di- and trinucleotide repeat motifs: uniform distributions [1.0×10^{-5}, 1.0×10^{-3}] and [1.0×10^{-6}, 1.0×10^{-3}], respectively. The individual locus mutation rate was taken as the gamma distribution [1.0×10^{-6}, 1.0×10^{-3}] and [1.0×10^{-7}, 1.0×10^{-3}] respectively. The mean coefficient P was assumed to have gamma prior distribution [1.0×10^{-1}, 3.0×10^{-1}], while the individual locus coefficient P was assumed to have gamma distribution [1.0×10^{-2}, 9.0×10^{-1}] as priors. The mean single nucleotide insertion/deletion (SNI/D) rate was postulated as log-uniform distribution [1.0×10^{-8}, 1.0×10^{-5}], and the individual locus SNI/D rate was postulated as gamma distribution [1.0×10^{-9}, 1.0×10^{-4}].

The prior distribution of each parameter for the G. soboliferum var. kiusianum populations was almost equivalent to the above explanation, but N1 represented current population size while Na was the ancestral population size of each population.

Table S2-1a. Per-locus information for *G. soboliferum* var. *kiusianum*

| | *G. soboliferum* var. *kiusianum* | | | | | | | | | | | |
| | Tadewara_A | | | | Tadewara_B | | | | Ide farm_A | | | |
Locus	A	Ar	H_e	H_o	A	Ar	H_e	H_o	A	Ar	H_e	H_o
VKGP_120298	5	4.173	0.482	0.209	5	4.857	0.503	0.393	5	4.441	0.662	0.519
VKGP_15780	15	11.384	0.842	0.857	3	3.000	0.425	0.438	7	6.764	0.754	0.482
VKGP_4763	5	4.982	0.729	0.780	2	2.000	0.383	0.323	3	2.922	0.422	0.262
VKGP_29306	3	2.997	0.518	0.592	2	2.000	0.492	0.438	2	2.000	0.229	0.264
VKGP_76222	4	3.999	0.660	0.776	3	2.987	0.543	0.688	2	1.852	0.061	0.063
VKGP_20310	3	2.752	0.239	0.271	1	1.000	0.000	0.000	3	1.814	0.033	0.017
VKGP_22221	3	2.913	0.177	0.095	2	2.000	0.467	0.613	2	1.863	0.062	0.000
VKGP_28824	5	4.853	0.490	0.510	3	2.750	0.500	0.500	7	5.097	0.682	0.596
VKGP_105296	6	5.474	0.717	0.694	2	2.000	0.350	0.194	7	4.735	0.247	0.206
VKGP_16051	6	5.103	0.616	0.735	4	3.747	0.550	0.656	7	4.316	0.376	0.354

| | Ide farm_B | | | | Mt. Daikanbou_A | | | | Mt. Daikanbou_B | | | |
| | | | | | *G. soboliferum* var. *kiusianum* | | | | | | | |
Locus	A	Ar	H_e	H_o	A	Ar	H_e	H_o	A	Ar	H_e	H_o
VKGP_120298	3	2.855	0.519	0.000	3	2.664	0.155	0.111	1	1.000	0.000	0.000
VKGP_15780	3	2.686	0.497	0.486	5	4.584	0.399	0.368	2	2.000	0.392	0.393
VKGP_4763	2	1.855	0.050	0.051	3	2.917	0.526	0.500	4	3.982	0.589	0.500
VKGP_29306	2	1.963	0.064	0.000	3	2.297	0.053	0.027	1	1.000	0.000	0.000
VKGP_76222	2	1.615	0.025	0.026	2	2.000	0.440	0.522	3	2.952	0.518	0.677
VKGP_20310	4	3.547	0.180	0.065	4	3.559	0.345	0.417	2	2.000	0.148	0.161
VKGP_22221	1	1.000	0.000	0.000	1	1.000	0.000	0.000	2	2.000	0.109	0.038
VKGP_28824	4	3.615	0.577	0.718	5	4.638	0.497	0.476	3	3.000	0.476	0.414
VKGP_105296	3	2.977	0.541	0.300	6	5.331	0.676	0.674	5	4.994	0.718	0.767
VKGP_16051	3	2.880	0.526	0.514	6	5.590	0.720	0.714	4	4.000	0.621	0.714

Note : A; number of allele, Ar; allelic rishness, H_e; expected heterozygosity; H_o; observed heterozygosity.

Table S2-1b. Per-locus information for *G. soboliferum* var. *hakusanense* and var. *soboliferum*

	G. soboliferum var. *hakusanense*											
	Nasu				Tochigi				Saku			
Locus	A	*Ar*	H_e	H_o	A	*Ar*	H_e	H_o	A	*Ar*	H_e	H_o
VKGP_120298	5	4.420	0.596	0.516	1	1.000	0.000	0.000	4	3.687	0.447	0.412
VKGP_15780	8	6.050	0.738	0.581	3	3.000	0.607	0.231	9	6.332	0.762	0.771
VKGP_4763	3	2.382	0.513	0.500	2	2.000	0.260	0.154	4	3.753	0.396	0.382
VKGP_29306	2	1.942	0.148	0.161	2	2.000	0.204	0.231	2	2.000	0.459	0.286
VKGP_76222	4	3.430	0.574	0.467	1	1.000	0.000	0.000	4	3.334	0.499	0.500
VKGP_20310	6	5.570	0.782	0.735	3	2.813	0.506	0.375	4	3.371	0.639	0.486
VKGP_22221	2	2.000	0.390	0.094	1	1.000	0.000	0.000	4	3.406	0.646	0.406
VKGP_28824	8	5.905	0.587	0.618	3	3.000	0.636	0.308	9	7.788	0.781	0.727
VKGP_105296	7	6.857	0.835	0.735	3	3.000	0.561	0.313	7	5.689	0.736	0.741
VKGP_16051	11	9.366	0.876	0.941	3	3.000	0.602	0.571	7	6.140	0.751	0.676

Locus	Kirigamine				Nobeyama				Kiyosato			
	A	*Ar*	H_e	H_o	*A*	*Ar*	H_e	H_o	*A*	*Ar*	H_e	H_o
VKGP_120298	4	3.414	0.525	0.548	4	3.968	0.718	0.516	5	4.786	0.706	0.677
VKGP_15780	12	9.139	0.782	0.571	16	11.230	0.889	0.774	12	7.740	0.757	0.548
VKGP_4763	6	5.324	0.732	0.567	6	4.809	0.737	0.688	6	4.773	0.654	0.600
VKGP_29306	2	1.5	0.038	0.038	2	2.000	0.366	0.276	2	1.905	0.124	0.133
VKGP_76222	4	3.086	0.404	0.387	5	3.888	0.305	0.281	5	3.790	0.452	0.406
VKGP_20310	4	3.65	0.625	0.656	5	4.680	0.735	0.714	4	3.987	0.684	0.577
VKGP_22221	3	2.667	0.531	0.290	4	3.590	0.569	0.417	3	3.000	0.638	0.433
VKGP_28824	10	6.429	0.642	0.464	12	9.953	0.889	0.906	12	9.237	0.833	0.875
VKGP_105296	6	5.667	0.713	0.706	11	7.885	0.699	0.500	9	7.611	0.758	0.677
VKGP_16051	8	5.455	0.649	0.656	11	8.308	0.685	0.724	9	7.836	0.825	0.759

G. soboliferum var. *hakusanense*

| | G. soboliferum var. hakusanense | | | | G. soboliferum var. soboliferum | | | |
| | Fujinomiya | | | | Jilin | | | |
Locus	A	Ar	H_e	H_o	A	Ar	H_e	H_o
VKGP_120298	2	1.934	0.153	0.119	5	4.980	0.741	0.778
VKGP_15780	7	5.052	0.711	0.763	6	5.876	0.765	0.778
VKGP_4763	2	1.547	0.049	0.050	3	3.000	0.586	0.222
VKGP_29306	2	2.000	0.499	0.452	5	4.882	0.747	0.444
VKGP_76222	3	2.647	0.253	0.238	2	2.000	0.117	0.125
VKGP_20310	2	1.992	0.242	0.282	4	3.889	0.667	0.222
VKGP_22221	2	2.000	0.496	0.206	1	1.000	0.000	0.000
VKGP_28824	5	4.953	0.782	0.214	7	6.556	0.735	0.778
VKGP_105296	3	2.665	0.419	0.465	5	5.000	0.664	0.625
VKGP_16051	7	4.778	0.421	0.444	7	7.000	0.789	0.875

Note : A; number of allele, Ar; allelic rishness, H_e; expected heterozygosity; H_o; observed heterozygosity.

Table 2-1c. Determination values (r^2) of correlation analysis about allelic richness (*Ar*) of interpopulation.

var. *kiusianum*	Tadewara_A	Tadewara_B	Ide farm_A	Ide farm_B	Mt. Daikanbou_A	Mt. Daikanbou_B
Tadewara_A	-	0.0645	0.6974	0.0379	0.258	0.0204
Tadewara_B	-	-	0.2276	0.0087	0.0164	0.0249
Ide farm_A	-	-	-	0.2702	0.5156	0.0411
Ide farm_B	-	-	-	-	0.5794	0.0177
Mt. Daikanbou_A	-	-	-	-	-	0.3285
Mt. Daikanbou_B	-	-	-	-	-	-

var. *hakusanense*	Nasu	Tochigi	Saku	Kirigamine	Nobeyama	Kiyosato	Fujinomiya
Nasu	-	0.5231	0.5407	0.3410	0.5379	0.6330	0.4968
Tochigi	-	-	0.4503	0.5644	0.4503	0.4628	0.3920
Saku	-	-	-	0.6930	0.9025	0.9624	0.7522
Kirigamine	-	-	-	-	0.8896	0.7192	0.5531
Nobeyama	-	-	-	-	-	0.8885	0.7806
Kiyosato	-	-	-	-	-	-	0.6700
Fujinomiya	-	-	-	-	-	-	-

Table S2-2. The difference of genetic diversity among populations.

allelic richness	Nasu	Tochigi	Saku	Kirigamine	Nobeyama	Kiyosato	Fujinomiya
Tadewara_A	ns	ns	ns	ns	ns	ns	ns
Tadewara_B	ns	ns	ns	ns	**	*	ns
Ide farm_A	ns	ns	ns	ns	**	*	ns
Ide farm_B	*	ns	**	**	***	**	ns
Mt. Daikanbou_A	ns	ns	ns	ns	*	ns	ns
Mt. Daikanbou_B	ns	ns	ns	ns	**	*	ns

expected heterozygosity	Nasu	Tochigi	Saku	Kirigamine	Nobeyama	Kiyosato	Fujinomiya
Tadewara_A	ns	ns	ns	ns	ns	ns	ns
Tadewara_B	ns	ns	ns	ns	ns	ns	ns
Ide farm_A	*	ns	*	ns	**	*	ns
Ide farm_B	**	ns	**	*	***	**	ns
Mt. Daikanbou_A	ns	ns	ns	ns	*	ns	ns
Mt. Daikanbou_B	ns	ns	ns	ns	*	*	ns

Note : ns; not significant, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

Table S2-3a. Pairwise F_{ST} values (upper right half) and geographic distance (km: below left half) of *G. soboliferum* var. *kiusianum*.

Population	Tadewara_A	Tadewara_B	Ide farm_A	Ide farm_B	Mt. Daikanbou_A	Mt. Daikanbou_B
Tadewara_A	-	0.219[*]	0.349[*]	0.386[*]	0.312[*]	0.327[*]
Tadewara_B	0.175	-	0.461[*]	0.482[*]	0.417[*]	0.412[*]
Ide farm_A	15.46	15.42	-	0.169[*]	0.400[*]	0.451[*]
Ide farm_B	15.45	15.41	0.06	-	0.443[*]	0.499[*]
Mt. Daikanbou_A	19.95	19.88	5.35	5.32	-	0.046[*]
Mt. Daikanbou_B	19.92	19.85	5.30	5.27	0.05	-

[*] $P < 0.001$

Table S2-3b. Pairwise F_{ST} values (upper right half) and geographic distance (km: below left half) of *G. soboliferum* var. *hakusanense* and var. *soboliferum*.

Population	Nasu	Tochigi	Saku	Kirigamine	Nobeyama	Kiyosato	Fujinomiya	Jilin
Nasu	-	0.331[*]	0.181[*]	0.206[*]	0.111[*]	0.135[*]	0.290[*]	0.196[*]
Tochigi	75	-	0.234[*]	0.277[*]	0.237[*]	0.242[*]	0.167[*]	0.375[*]
Saku	158	36	-	0.136[*]	0.112[*]	0.116[*]	0.219[*]	0.223[*]
Kirigamine	199	36	42	-	0.148[*]	0.115[*]	0.276[*]	0.270[*]
Nobeyama	188	124	34	33	-	0.037[*]	0.197[*]	0.193[*]
Kiyosato	192	129	37	29	5	-	0.226[*]	0.209[*]
Fujinomiya	234	160	99	93	67	68	-	0.364[*]
Jilin	1193	1209	1149	1129	1162	1158	1214	-

[*] $P < 0.001$

Table S2-4. Analysis of molecular variance table of two Japanese varieties (*G. soboliferum* var. *kiusianum* and var. *hakusanense*).

Source	df	SS	MS	Variance (%)	
Regions	1	266.25	266.25	10%	$F_{rt} = 0.103^*$
Populations within regions	11	794.16	72.20	25%	$F_{sr} = 0.279^*$
Among all populations	985	2351.46	2.39	65%	$F_{st} = 0.353^*$
Total	997	3411.87		100%	

Note : Frt; differentiation between *G. soboliferum* var. *kiusianum* and *G. soboliferum* var. *hakusanense*, Fst; differentiation between populations within *G. soboliferum* var. *kiusianum* and *G. soboliferum* var. *hakusanense* respectively, Fst; differentiation among all populations. * $P < 0.001$.
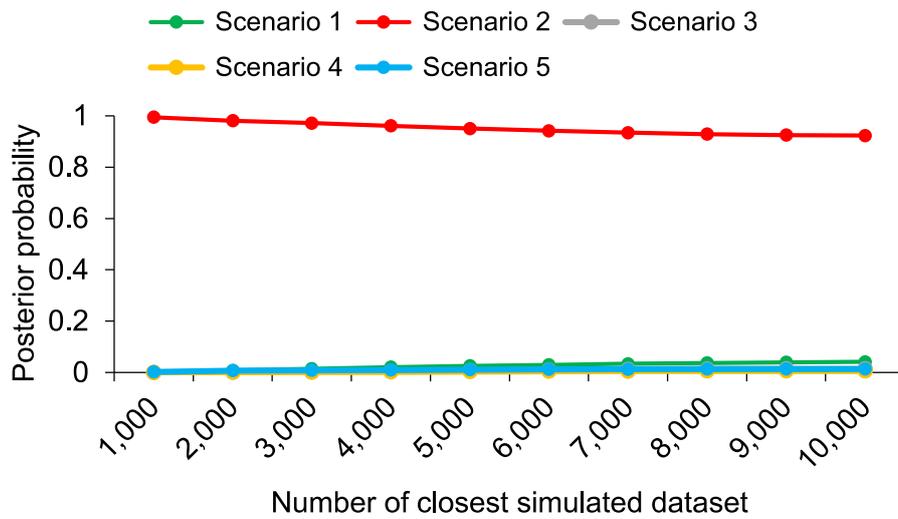
Fig. S2-1. Posterior probability of each demographic scenario in DIYABC analysis.
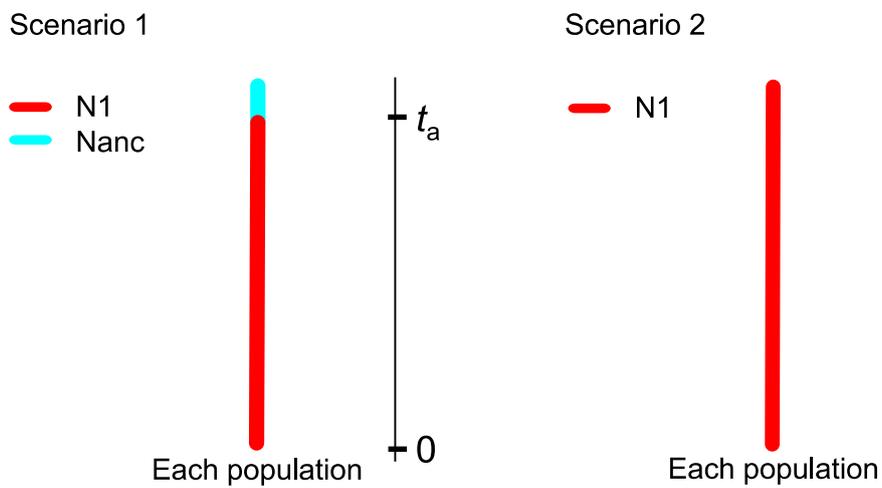
Fig. S2-2. The demographic scenario based on a single-population model in DIYABC analysis. The populations are those of Tadewara, Ide Farm, and Mt. Daikanbou.
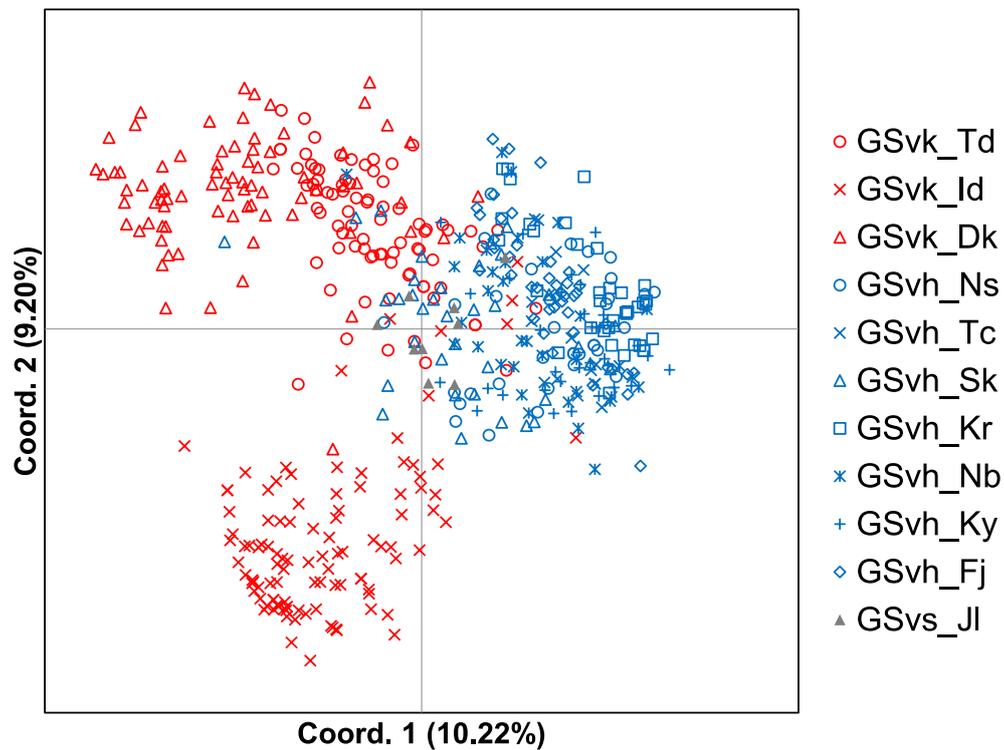
Fig. S2-3. Population-based PCoA. GSvk_Td, *G. soboliferum* var. *kiusianum* (Tadewara); GSvk_Id, *G. soboliferum* var. *kiusianum* (Ide Farm); GSvk_Dk, *G.soboliferum* var. *kiusianum* (Mt. Daikanbou); GSvh_Ns, *G. soboliferum* var. *hakusanense* (Nasu); GSvh_Tc, *G. soboliferum* var. *hakusanense* (Tochigi); GSvh_Sk, *G. soboliferum* var. *hakusanense* (Saku); GSvh_Kr, *G. soboliferum* var. *hakusanense* (Kirigamine); GSvh_Nb, *G. soboliferum* var. *hakusanense* (Nobeyama); GSvh_Ky, *G. soboliferum* var. *hakusanense* (Kiyosato); GSvh_Fj, *G. soboliferum* var. *hakusanense* (Fujinomiya); GSvs_Ji, *G. soboliferum* var. *soboliferum* (Jilin).

Supplementary (Chapter 3)

The histrical distributions of each demographic parameter on DIYABC

1st step ABC analysis

Prior distribution of each parameter was set as follows; N1 (var. *yesoense*), N2 (var. *nipponicum*), N3 (var. *pseudopratense*), N_{1a} and N_{2a} of effective population size and t_1 of time parameters was assumed to have uniform distribution [10, 100,000]. Effective population size of ancestral population (Na) and the first divergence time of t_2 were assumed to have uniform distribution [10, 500,000]. We assumed the parameter of admixture rate (ra) to have uniform distribution [0.001, 0.999]. Furthermore, we assumed the relationship between population size parameters, as Na ≥ N1, N1a and Na > N1, N2, N3, N1a, N2a. In addition, we assumed the relationship between time parameters, as $t_2 > t_1$.

2nd step ABC analysis

Prior distribution of each parameter was set as follows; N1 (var. *yesoense*), N2 (var. *nipponicum*), N3 (var. *pseudopratense*), N4 (Mt. Ibuki), N5, (Mt. Asama), N_{1a}, N_{2a}, N_{3a}, N_{2a2} and N_{3a2} of effective population size and t_1, t_2 and t_3 of time parameters were assumed to have uniform distribution [10, 100,000]. Effective population size of ancestral population (Na) and the first divergence time of t_4 were assumed to have uniform distribution [10, 500,000]. We assumed the parameter of admixture rate (ra, rb and rc) to have uniform distribution [0.001, 0.999]. Furthermore, we assumed the relationship between population size parameters, as Na ≥ N1, N1a and Na > N2, N3, N4, N5, N2a, N3a2 and N2a2 ≥ N2, N2a and N3a ≥ N3 and N3a2 ≥ N3, N3a and N1 > N4, N5 and N2 > N4, N5 and N3 > N4, N5. In addition, we assumed the relationship between time parameters, as $t_4 > t_1$, t_2, t_3 and $t_3 > t_1$, t_2 and $t_2 > t_1$.
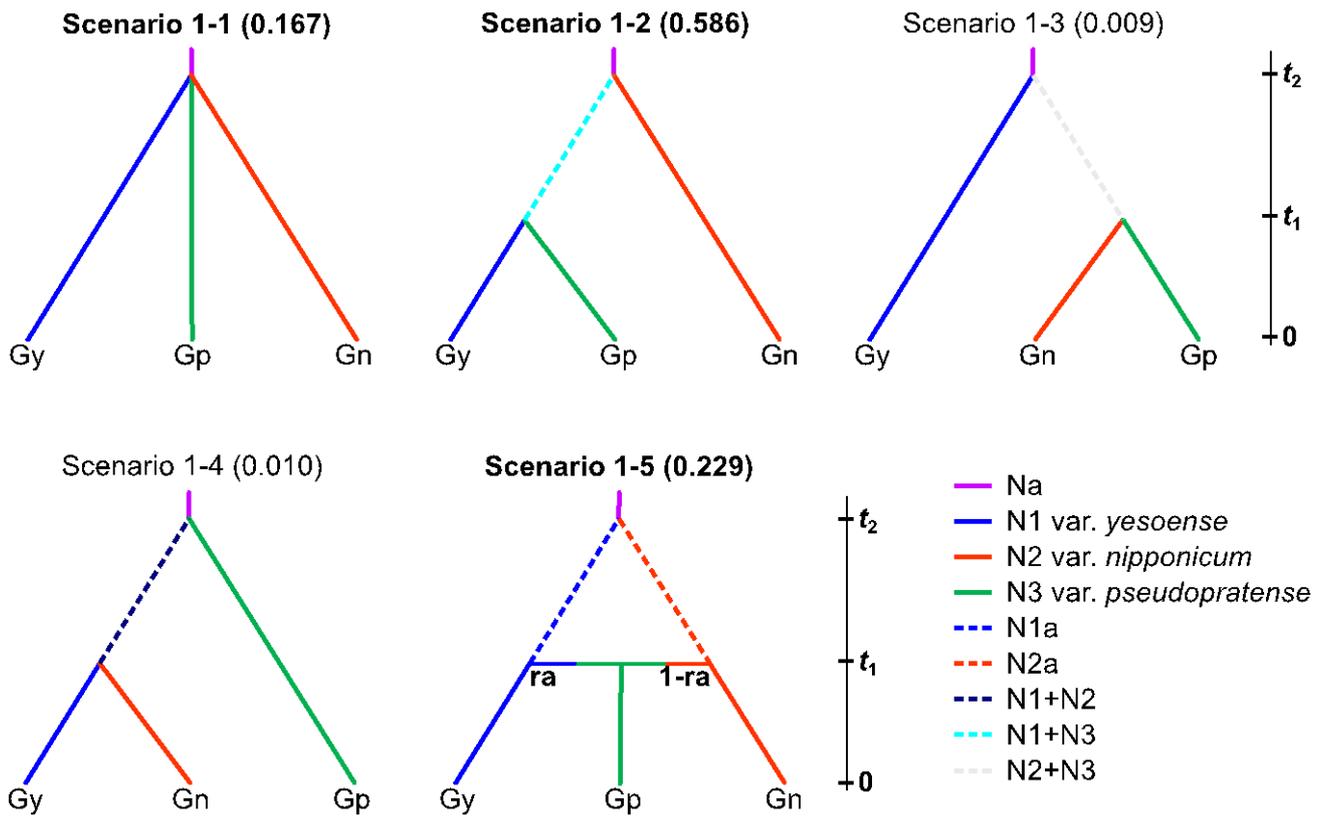
Fig. S3-1. Five divergence scenarios considered at the first step of the approximate Bayesian computation analysis. Nx, effective population size (x, arbitrary number of populations or ancestral populations); t_x, branching time of each population, or time of merging between populations; "0," present time; ra, admixture rate. The numbers in parentheses represent the posterior probabilities from the simulations. Gy, var. *yesoense*; Gn, var. *nipponicum*; Gp, var. *pseudopratense*.
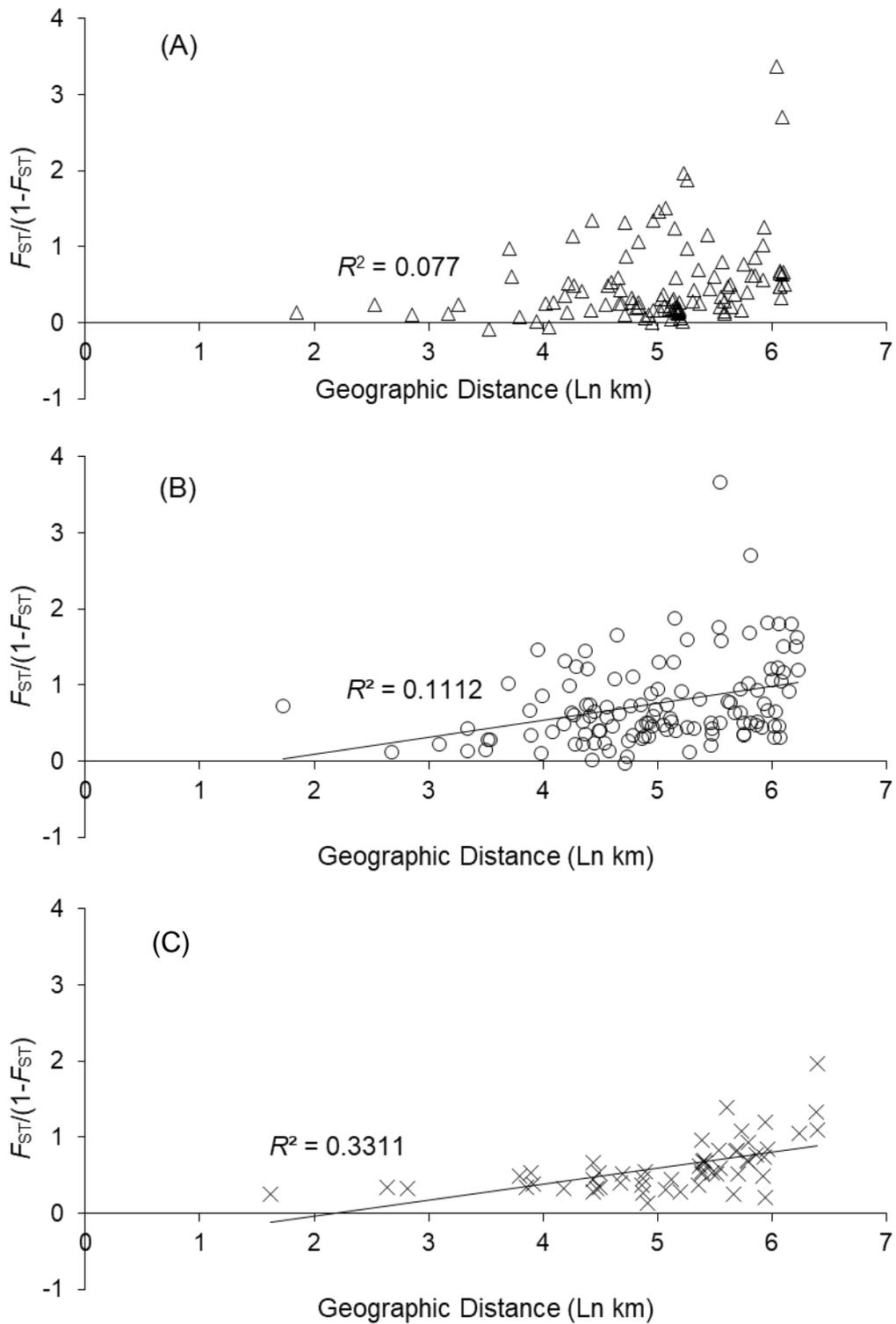
Fig. S3-2. Isolation by distance for *G. yesoense* (A) var. *yesoense*, (B) var. *nipponicum*, and (C) var. *pseudopratense*. The horizontal axis represents geographic distance and the vertical axis represents genetic distance; i.e., $F_{ST}/(1 - F_{ST})$.