Theoretical and empirical studies on the roles of life history variation in maintaining genetic diversity in stage-structured perennial plant populations

Yoichi Tsuzuki

A Doctoral Dissertation

Graduate School of Environmental Science, Hokkaido University

2023

Contents

Related articles

Chapter 2

- Tsuzuki, Y., Takada, T., and Ohara, M. (2022). Modeling temporal dynamics of genetic diversity in stage-structured plant populations with reference to demographic genetic structure. *Theoretical Population Biology*, *148*, 76-85. doi: 10.1016/j.tpb.2022.11.001
- Tsuzuki, Y., Takada, T., and Ohara, M. (2022). Which life history strategy can maintain high genetic diversity in plants? *bioRxiv*. doi: 10.1101/2022.04.26.489613

Chapter 3

- Tsuzuki, Y., Sato, P. M., Matsuo, A., Suyama, Y., and Ohara, M. (2022). Genetic consequences of habitat fragmentation in a perennial plant *Trillium camschatcense* are subjected to its slow-paced life history. *Population Ecology*, *64*(1), 5-18. doi: 10.1002/1438-390X.12093

Chapter 5

- Tsuzuki, Y., and Ohara, M. (2022). Temporal skewness of pollination success in the spring ephemeral *Trillium camschatcense*. *The Science of Nature*, *109*, 35. doi: 10.1007/s00114-022-01807-8

Chapter 1

General Introduction

Human-induced environmental changes have been taking place around the globe (Ellis et al. 2010), which exposes biological populations to new selection pressures. Populations can persist if they rapidly adapt to new environments before becoming critically small (Carlson et al. 2014). Genetic diversity, or standing genetic variation, is one of the crucial components of rapid adaptation to environmental changes. Populations with low genetic diversity would have to wait for new beneficial mutations because of the scarcity of candidate adaptive genes, which usually takes a long time (Orr and Unckless 2014). On the other hand, populations with high genetic diversity possess a variety of alleles beforehand, which shortens the time for adaptation. The rapidity achieved by genetic diversity prevents populations from severe shrinkage and increases the chance of persistence under environmental changes (Agashe et al. 2011, Ramsayer et al. 2013).

There is an increasing number of evidence that evolution occurs on the contemporary time scale (Hairston et al. 2005, Schoener 2011), including species that rapidly adapt to anthropogenic environmental changes (Anderson et al. 2012, Fukano et al. 2020). Therefore, it could be said that rapid evolution from standing genetic variation is increasingly becoming relevant in today's biosphere. Understanding the determinant of genetic diversity is an important study topic to predict population viability under ongoing environmental changes.

Classical population genetics predicts that population size is the primary determinant of genetic diversity. This prediction is based on the theoretical analysis on Wright-Fisher model

(Fisher 1930, Wright 1931), which is an ideal population that satisfies constant population size, random mating, and non-overlapping generations. In Wright-Fisher model, because individuals randomly mate with one another and are replaced by the same number of offspring every generation, generation turnover is equivalent to random sampling of genes with replacement. Therefore, small population size leads to fewer sampling times, thus resulting in strong stochastic genetic drift and reduced genetic diversity (Crow and Kimura 1970).

This theoretical prediction is often applied to wild populations that have undergone habitat loss and fragmentation. In addition to the immediate effect of bottleneck, small fragmented populations should be experiencing intensified genetic drift and resultant loss of genetic diversity due to small population size. Although fragmentation surely decreased genetic diversity in general (Aguilar et al. 2008, González et al. 2020), some perennial plant species did not show a clear decline in genetic diversity with decreasing population size (Bacles and Jump 2011, Bezemer et al. 2019, Kramer et al. 2008), indicating that genetic diversity was not affected solely by population shrinkage. While the discrepancy was once regarded as "fragmentation paradox" at first (Kramer et al. 2008), it turned out that life history characteristics of the examined species played important roles in maintaining genetic diversity under small population size. Contrary to Wright-Fisher model, many plant species have overlapping generations due to their long lifespan, and populations consist of individuals with different life history stages and ages. Some of the life history stages consist of not only young but also old individuals and thus capture genetic variation of old generations. These key life history stages serve as genetic reservoirs from the past, contributing to the maintenance of genetic diversity (Aparicio et al. 2012, Cannon et al. 2022, Lowe et al. 2015).

The fact that genetic diversity is accumulated over the course of the life history (i.e., a

lifetime trajectory of an individual from birth to death) is not special to fragmented populations but a universal phenomenon. In the pioneering literature review (Loveless and Hamrick 1984), it was reported that long-lived plant species had generally higher genetic diversity than short-lived ones. Similarly, Hamrick and Godt (1996) analyzed the correlation between multiple life history traits and genetic diversity in plants, and showed that long generation time contributed to genetic diversity. Austerlitz and Garnier-Géré (2003) showed that long lifespan helps accumulate genetic variation during colonization to new habitats. In colonizing plant populations, long generation time requires a longer time for founder individuals to grow, reproduce, and dominate, thus suppressing the founder effect and allowing successive multiple introductions from the original range. This leads to higher genetic diversity despite the small population size at the frontier of the range expansion. Recent global assessment of genetic diversity in plant populations also showed that life form (annual vs perennial vs shrub vs tree) significantly affected genetic diversity, with stronger positive effects by long-lived life forms (De Kort et al. 2021).

While increasing evidence shows that long life history contributes to the accumulation of genetic diversity and sometimes obscures the impacts of small population size, preceding studies did not go so far as to examine on what occasions life history became crucial in maintaining genetic diversity under the influential effects of population size. This can be partly attributed to the fact that most previous studies focused on the interspecific comparison when assessing the impacts of life history. Because there has been much less attention paid to among-population comparison, it was difficult to jointly consider the impacts of population size, which is a population-level determinant of genetic diversity. In general, life history traits including longevity are different not only among species but also among populations (Ehrlén and Lehtilä 2002, Hughes 2017). It has been shown that the intraspecific variation of

life history has a non-negligible effect on the dynamics and the viability of populations. For example, life history diversity contributes to asynchronous birth, growth and reproduction among individuals and populations, which can stabilize population dynamics through portfolio effects (Moore et al. 2014, Schindler et al. 2010, Waddle et al. 2019). The crucial population-level consequences of life history variation might also apply to genetic diversity, influencing the level of genetic diversity independently of population size. Consequences of among-population life history variation on genetic diversity with comparison of that of population size is a frontier in understanding the contribution of life history cycle to genetic diversity.

In my doctoral dissertation, I studied how the interpopulation variation in the life history cycle and population size interact to drive the maintenance of genetic diversity in perennial plants both theoretically and empirically. In Chapter 2, I focused on demographic genetic structure, which is the within-population genetic composition structured by stage. I developed a mathematical model that describes the temporal dynamics of the stage-wise genetic diversity, and I formulated effective population size and the annual change rate of expected heterozygosity from the model. By analyzing the model, I examined on what occasions life history plays crucial roles in maintaining genetic diversity under the effects of population size, and confirmed that demographic genetic structure could be a useful tool to understand how genetic diversity is accumulated over the course of life history. In Chapter 3, I empirically tested the theoretical prediction of Chapter 2 using fragmented populations of a perennial herb *Trillium camschatcense* in the Tokachi region of Hokkaido, northern Japan. I conducted population genetic analysis using double digest restriction-site associated DNA sequencing (ddRAD-seq) to estimate demographic genetic structure and effective population size. Besides, I carried out a field census survey to estimate demographic rates (stage-specific survival rates and fecundity) and population size of each population, which were subsequently used to calculate the theoretical

expectation of demographic genetic structure and effective population size based on the model developed in Chapter 2. The resultant theoretical expectations were compared with observed values to examine whether the influence of life history on genetic diversity predicted by the theoretical model was in line with the observed data. In Chapter 4, I focused on the gap between the empirical genetic data and the theoretical prediction. I inversely estimated demographic rates and population size from the observed demographic genetic structure to depict how genes were transferred among life history stages along the life cycle. I suggested a concept of "genetically effective life history," which was different from the life history estimated from census survey and could bridge the gap between theoretical predictions and empirical systems. In Chapter 5, to grasp a possible scenario in which genetic diversity would help populations persist under environmental changes, I examined how the ongoing climate change is affecting the phenological overlap with pollinators in fragmented *T. camschatcense* populations in the Tokachi region. In two populations differing in the long-term trend of temperature rise, I tested whether pollinator visitation occurred during the entire flowering period or was temporally limited. In summarizing the overall studies, I discussed the genetic and evolutionary consequences of life history in Chapter 6.

Chapter 2

Theoretical prediction on the dynamics of genetic diversity and demographic genetic structure under a variety of plant life history strategies

Abstract

In stage-structured populations, especially in perennial plant species, genetic diversity is often compared among life history stages, such as seedlings, juveniles, and flowerings, using neutral genetic markers. The stage-wise genetic diversity is sometimes referred to as demographic genetic structure and has been regarded as a proxy of potential genetic changes, because individuals in mature stages will die and be replaced by those in more immature stages over the course of time. However, due to the lack of theoretical examination, it remained unclear whether the stage-wise genetic diversity genuinely represented temporal dynamics. To describe the dynamics and the inter-stage differences of genetic diversity in stage-structured plant populations, I developed a matrix model which was made up of difference equations of the probability of non-identical-by-descent of each life history stage. Using the model, I examined the relationships of demographic genetic structure with the annual change rate of expected heterozygosity (denoted by η), as well as with effective population size N_e (the determinants of diversity loss per generation time). Demographic genetic structure varied independently of N_e and η but was subjected to stable stage distribution: genetic diversity was higher in stages with more individuals. Especially, stages in which individuals were detained due to slow or retrogressive growth possessed high genetic diversity, probably because individuals with variable ages can be mixed and genetic variation from broader time span can be stored. η was high in slow-paced life histories (long generation time, higher age at maturity, long life expectancy), which could be also attributed to generation overlap because slow-paced life history would promote coexistence of differently-aged individuals within a population. These results indicate that demographic genetic structure does not reflect temporal trends in genetic diversity, and that slow-paced life history contribute to genetic diversity both at the stage- and at the population-level. The efficacy of demographic genetic structure lies in identifying key developmental stages that accumulate genetic variation from static and retrogressive life history processes to detect the contribution of life history to the maintenance of genetic diversity.

2.1 Introduction

Genetic diversity, or standing genetic variation, is a source of adaptive evolution (Barrett and Schluter 2008). Populations with high genetic diversity are more likely to adapt to environmental changes and to persist for a long period (Agashe et al. 2011, Ramsayer et al. 2013). Therefore, understanding how genetic diversity is maintained over time is important for assessing population viability (Mimura et al. 2017).

The rate of change in genetic diversity per generation time is primarily determined by the effective population size (N_e): the larger N_e, the weaker genetic drift, and the more likely genetic diversity is maintained (Crow and Kimura 1970). Although N_e was first theoretically proposed for populations without generation overlap, many species have overlapping generations and populations are made up of individuals differing in age or life history stage. For example, tree populations can roughly divided into three stages: seedlings, saplings, and adults. Theoretical studies have extended the concept of effective population size to populations structured by age (Felsenstein 1971, Hill 1972, 1979, Johnson 1977) or by stage (Orive 1993, Yonezawa et al. 2000) by formulating N_e with demographic rates (age- or stage-specific survival rates and fecundities). These formulations enable us to calculate N_e and to assess the temporal genetic dynamics in species with complex life histories (Waples et al. 2011, 2013).

Meanwhile, some empirical genetic studies do not examine N_e to predict future genetic diversity of stage-structured populations. Instead, genetic diversity is comparatively estimated for each life history stage at a single time point with neutral genetic markers (Aldrich et al. 1998, Ally and Ritland 2006, Kettle et al. 2007, Linhart et al. 1981, Murren 2003, Schmidt et al. 2018, Vranckx et al. 2014). The resultant stage-wise genetic diversity is referred to as demographic genetic structure (Aldrich et al. 1998) and is considered to reflect potential

genetic changes that accompany the turnover of constituent individuals. For example, if juvenile stage is less diverse than more mature stages, genetic diversity would decrease with the replacement of mature individuals to juveniles. Because species with stage structure are mostly long-lived and long-term genetic monitoring is impractical, demographic genetic structure has been considered as a rough but a convenient empirical approach to infer the temporal genetic dynamics (Mimura et al. 2017, Schmidt et al. 2018).

Despite its empirical usage, mathematical and theoretical basis of demographic genetic structure has been in its infancy. Unlike N_e, demographic genetic structure has not been formulated mathematically, and the lack of theoretical background draws concerns about the current interpretation. While analysis on demographic genetic structure implicitly assumes that individuals sequentially grow and die from juvenile to mature stage classes, this assumption is potentially invalid in perennial plants. In most perennial plant species, whose life histories depend on stage, not on age (Silvertown 1987), aging or passing of time does not necessarily promote growth and maturation. Some individuals might keep proceeding to more mature stages, while others remain in the same stage for a long period (stasis) or even reverse to more juvenile stages (retrogression). For example, long-lived woodland perennial herbs of the genus *Trillium* show stasis for more than ten years in juvenile stages as well as go back from a mature reproductive stage to a pre-reproductive one in response to resource exhaustion (Knight 2004, Ohara et al. 2001, Tomimatsu and Ohara 2010). The static and bidirectional flows in the life cycle complicate the order of individual turnover in a population, potentially obscuring the sign of temporal changes. It has not been theoretically confirmed if demographic genetic structure still serves as a genuine proxy for temporal changes. Mathematical formulation that encompasses demographic genetic structure, as well as the temporal change in genetic diversity, in stage-structured populations will provide integrative understandings on these

problems, but has never been achieved so far.

In this study, I develop a matrix model to describe the temporal dynamics of expected heterozygosity, which is used as a common metric of genetic diversity, for a neutral locus of a stage-structured perennial plant species. The model is constructed by deriving difference equations of the probability that two genes randomly sampled from a given life history stage are non-identical-by-descent. Based on the model, I formulate demographic genetic structure and N_e. Thus, our model allows integrative analysis on demographic genetic structure, temporal dynamics of genetic diversity, and their relationships. In the following sections, I first describe the derivation procedures (section 2.2.1) and the validation of our model (section 2.2.2). I then check whether demographic genetic structure reflects temporal dynamics (section 2.2.3). Lastly, I examine the determinants of demographic genetic structure, as well as effective population size, under a variety of plant life history strategies (section 2.2.4).

2.2 Materials and Methods

2.2.1 Model development

Overview

Felsenstein (1971) derived inbreeding effective population size for age-structured populations by formulating recurrence equations of the probability of non-identical-by-descent, which is also described in Charlesworth (1994). I partly follow mathematical formulation procedures in Felsenstein (1971) while adding necessary modifications to extend it to stage-structured populations. I develop difference equations of the probability of non-identical-by-descent at a neutral locus for a closed, stage-structured population, supposing a diploid perennial plant

species. I do not consider sex differences because most plants are hermaphrodite (Torices et al. 2011). I assume that mutations do not newly occur. Besides, as in Felsenstein (1971), I assume demographic equilibrium, where the census population size and its allocation to each stage (stage distribution) are constant over time. Census population size is set to N, which is divided into n life history stages ($N_1, N_2, \cdots, N_n$).

$$N = \sum_{i=1}^{n} N_i \tag{2.1}$$

The probability of transition (either growth, stasis, or retrogression) from stage j to stage i is t_{ij} per year. In each year, individuals randomly mate and f_{ij} newborns join stage i from a parent in stage j. a_{ij}, which denotes the sum of t_{ij} and f_{ij}, describes the total rate of flow from stage j to i between two successive years.

$$a_{ij} = t_{ij} + f_{ij}. \tag{2.2}$$

In age-structured life histories, flows of individuals among age classes are sparse: survival paths connect only adjacent ages in the direction from age i to $i+1$ (i.e., $t_{ij} = 0$ when $i \neq j+1$), and reproduction paths join only age class 1 (i.e., $f_{ij} = 0$ when $i \neq 1$). In plants, however, multiple survival paths come in and out from each stage by the combination of growth, stasis, and retrogression. Moreover, newborns do not always join the first stage, because newborn seeds either become dormant to join seed bank stage, or immediately germinate to join juvenile stages, resulting in multiple destinations (e.g., a perennial plant *Carduus nutans*, whose life cycle is shown in figure 1 of Shea and Kelly (1998)). Therefore, stage is not merely a pooling of successive age classes and stage-structured life histories are essentially different from age-structured ones. I need to consider all possible transition and reproduction

paths among stages, which is quite a distinct point compared to the age-structured model in Felsenstein (1971).

Population dynamics can be modeled by the following matrix population model.

$$\begin{pmatrix} N_{1,t} \\ \vdots \\ N_{i,t} \\ \vdots \\ N_{n,t} \end{pmatrix} = \begin{pmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\ \vdots & & \vdots & & \vdots \\ a_{i1} & \cdots & a_{ij} & \cdots & a_{in} \\ \vdots & & \vdots & & \vdots \\ a_{n1} & \cdots & a_{nj} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} N_{1,t-1} \\ \vdots \\ N_{i,t-1} \\ \vdots \\ N_{n,t-1} \end{pmatrix}. \tag{2.3}$$

$N_{i,t}$ denotes the number of individuals in stage i in year t, which is always equal to N_i for any t because I assume demographic equilibrium. Stable stage distribution, which is the relative number of individuals among stages in the equilibrium state, is proportional to the leading right eigenvector of the transition matrix (Caswell 2001).

I define $H_{ij,t}$ as the probability that two genes randomly sampled from stage i and j with replacement in year t are not identical-by-descent. Each gene has its own ancestry, and two-gene pairs that are (non-)identical-by-descent at $t = 0$ will remain the same for any t. Similarly, because I assume no mutations, two-gene pairs that are (non-)identical-by-state at $t = 0$ will also remain the same over time. This means that $H_{ij,t}$ behaves in the same manner as expected heterozygosity, which is the probability of non-identical-by-state and is commonly used as a proxy of genetic diversity. The goal of this study is to formulate $H_{ij,t}$ for all possible i and j, which enables us to obtain theoretical counterpart of demographic genetic structure, that is, stage-wise genetic diversity at a particular time point.

Here, I provide key derivation procedures, highlighting the differences with the preceding age-structured models in Felsenstein (1971). The complete derivation procedures are given

in Appendix 1.1.

Difference equations of $H_{ij,t}$

I begin with modeling the changes in $H_{ij,t}$ between two successive time points for all i and j, which are the stage-structured version of equations 2 to 5 in Felsenstein (1971). I separately consider two mutually exclusive situations: $i \neq j$ (case 1) and $i = j$ (case 2). Both cases can be further split into six situations. Firstly, two genes randomly sampled in year t were either in the same stage (say, stage m, case A) or in different stages (say, stage k and l, case B) in year $t-1$. Furthermore, genes can move among stages either by survival (grow, stasis, and retrogression) or by reproduction. Survival and reproduction are essentially different because reproduction allows one gene to be replicated and to move to multiple stages simultaneously and independently, while survival does not. There are three possibilities in how the two genes sampled were transferred from the previous year: both genes were transferred by survival (case α), one by survival and the other by reproduction (case β), and both by reproduction (case γ). Considering the combinations of where (case A and B) and how (case α, β, and γ) the two genes sampled came from, there are 6 mutually exclusive situations to be considered in both case 1 and 2 (Figure 2.1).

This classification scheme is original to my stage-structured model, and is not adopted in Felsenstein (1971). Compared to age-structured life histories, classes are more densely interconnected by survival and reproduction in stage-structured ones. It is necessary to consider as many as 12 situations to handle the complexity in plant life histories.

In case 1 (i.e., $i \neq j$), $H_{ij,t}$ can be decomposed as follows.

$$H_{ij,t} = H_{ij,t}|_{1 \cap A \cap \alpha} + H_{ij,t}|_{1 \cap A \cap \beta} + H_{ij,t}|_{1 \cap A \cap \gamma}$$

$$+ H_{ij,t}|_{1\cap B\cap\alpha} + H_{ij,t}|_{1\cap B\cap\beta} + H_{ij,t}|_{1\cap B\cap\gamma}, \tag{2.4}$$

where the cap symbol $\cap$ stands for the co-occurrence of multiple cases: $H_{ij,t}|_{1\cap Y\cap Z}$ stands for $H_{ij,t}$ that simultaneously satisfies case 1, Y, and Z ($Y = A, B$; $Z = \alpha, \beta, \gamma$). All six $H_{ij,t}|_{1\cap Y\cap Z}$ on the right side of equation 2.4 are formulated as follows (see Appendix A1.1.1 for details).

$$\begin{aligned}
H_{ij,t}|_{1\cap A\cap\alpha} &= \sum_{m=1}^{n}\left\{\frac{t_{im}t_{jm}N_m^2}{N_iN_j}\times\frac{1}{1-1/(2N_m)}H_{mm,t-1}\right\}\\
H_{ij,t}|_{1\cap A\cap\beta} &= \sum_{m=1}^{n}\left\{\frac{(t_{im}f_{jm}+f_{im}t_{jm})N_m^2}{N_iN_j}\times H_{mm,t-1}\right\}\\
H_{ij,t}|_{1\cap A\cap\gamma} &= \sum_{m=1}^{n}\left(\frac{f_{im}f_{jm}N_m^2}{N_iN_j}\times H_{mm,t-1}\right)\\
H_{ij,t}|_{1\cap B\cap\alpha} &= \sum_{k=1}^{n}\sum_{l>k}^{n}\left\{\frac{(t_{ik}t_{jl}+t_{il}t_{jk})N_kN_l}{N_iN_j}\times H_{kl,t-1}\right\}\\
H_{ij,t}|_{1\cap B\cap\beta} &= \sum_{k=1}^{n}\sum_{l>k}^{n}\left\{\frac{(t_{ik}f_{jl}+f_{ik}t_{jl}+t_{il}f_{jk}+f_{il}t_{jk})N_kN_l}{N_iN_j}\times H_{kl,t-1}\right\}\\
H_{ij,t}|_{1\cap B\cap\gamma} &= \sum_{k=1}^{n}\sum_{l>k}^{n}\left\{\frac{(f_{ik}f_{jl}+f_{il}f_{jk})N_kN_l}{N_iN_j}\times H_{kl,t-1}\right\}.
\end{aligned} \tag{2.5}$$

Each $H_{ij,t}|_{1\cap Y\cap Z}$ is shown as a summation of a multiplications of two terms. The first term is a conditional probability of case $1\cap Y\cap Z$ given case 1. For example, the first term of $H_{ij,t}|_{1\cap A\cap\alpha}$ can be rewritten as $(2t_{im}N_m)/(2N_i)\times(2t_{jm}N_m)/(2N_j)$, which is the number of two-gene pairs that fall into case 1, A, and α simultaneously under a specific m (i.e., $2t_{im}N_m\times 2t_{jm}N_m$) divided by the total number of pairs that satisfy case 1 (i.e., $2N_i\times 2N_j$). Here, the number of genes are twice the number of individuals because I assume diploid species. Similarly, the first term in the other five equations stand for the corresponding proportion of two-genes pairs. The second term stands for the probability of non-identical-

by-descent. Considering which stages the two genes sampled belonged to in year $t-1$, I replace the probability with either $H_{mm,t-1}$ or $H_{kl,t-1}$, except $H_{ij,t}|_{1\cap A\cap\alpha}$. In the case of $1\cap A\cap\alpha$, genes sampled from stage i must be mutually exclusive against those from stage j, because one gene in stage m in year $t-1$ could not move to both stage i and j simultaneously without being duplicated through reproduction. In other words, a gene that were in stage m in the previous year cannot be sampled twice, which violates the assumption of $H_{mm,t-1}$, that is, "sampling with replacement." Therefore, $H_{ij,t}|_{1\cap A\cap\alpha}$ inherits the probability that two genes randomly sampled from stage m "without" replacement in year $t-1$ were not identical-by-descent, which can be obtained by dividing $H_{mm,t-1}$ by the chance of not sampling the same gene twice ($= 1 - 1/(2N_m)$).

Substituting equations 2.5 to equation 2.4, $H_{ij,t}$ is formulated as follows.

$$\begin{aligned} H_{ij,t} = &\sum_{m=1}^{n} \frac{N_m^2}{N_i N_j} \left\{ \frac{t_{im}t_{jm}}{1-1/(2N_m)} + f_{im}t_{jm} + t_{im}f_{jm} + f_{im}f_{jm} \right\} H_{mm,t-1} \\ &+ \sum_{k=1}^{n}\sum_{l>k}^{n} \frac{N_k N_l}{N_i N_j}(a_{ik}a_{jl} + a_{il}a_{jk}) H_{kl,t-1}. \end{aligned} \tag{2.6}$$

As for case 2 (i.e., $i = j$), I decompose $H_{ii,t}$ into six conditional probabilities.

$$\begin{aligned} H_{ii,t} = &H_{ii,t}|_{2\cap A\cap\alpha} + H_{ii,t}|_{2\cap A\cap\beta} + H_{ii,t}|_{2\cap A\cap\gamma} \\ &+ H_{ii,t}|_{2\cap B\cap\alpha} + H_{ii,t}|_{2\cap B\cap\beta} + H_{ii,t}|_{2\cap B\cap\gamma}. \end{aligned} \tag{2.7}$$

The probabilities of non-identical-by-descent on the right side of equation 2.7 can be formulated with $H_{mm,t-1}$ and $H_{kl,t-1}$, as previously done for $H_{ij,t}$ in case 1 (see Appendix A1.1.2 for

details).

$$
\begin{aligned}
H_{ii,t}|_{2\cap A\cap\alpha} &= \sum_{m=1}^{n}\left\{\left(\frac{t_{im}N_m}{N_i}\right)^2\times\frac{1-1/(2t_{im}N_m)}{1-1/(2N_m)}H_{mm,t-1}\right\}\\
H_{ii,t}|_{2\cap A\cap\beta} &= \sum_{m=1}^{n}\left(\frac{2t_{im}f_{im}N_m^2}{N_i^2}\times H_{mm,t-1}\right)\\
H_{ii,t}|_{2\cap A\cap\gamma} &= \sum_{m=1}^{n}\left\{\left(\frac{f_{im}N_m}{N_i}\right)^2\times\left(1-\frac{1}{2f_{im}N_m}\right)H_{mm,t-1}\right\}\\
H_{ii,t}|_{2\cap B\cap\alpha} &= \sum_{k=1}^{n}\sum_{l>k}^{n}\left(\frac{2t_{ik}t_{il}N_kN_l}{N_i^2}\times H_{kl,t-1}\right)\\
H_{ii,t}|_{2\cap B\cap\beta} &= \sum_{k=1}^{n}\sum_{l>k}^{n}\left\{\frac{2(t_{ik}f_{il}+f_{ik}t_{il})N_kN_l}{N_i^2}\times H_{kl,t-1}\right\}\\
H_{ii,t}|_{2\cap B\cap\gamma} &= \sum_{k=1}^{n}\sum_{l>k}^{n}\left(\frac{2f_{ik}f_{il}N_kN_l}{N_i^2}\times H_{kl,t-1}\right).
\end{aligned}
\tag{2.8}
$$

Here, as with $H_{ij,t}|_{1\cap A\cap\alpha}$ in case 1, the second term of $H_{ii,t}|_{2\cap A\cap\alpha}$ and $H_{ii,t}|_{2\cap A\cap\gamma}$ are not exactly the same as $H_{mm,t-1}$. This is because the sources from which two genes are sampled cannot be replaced with stage m of the previous year. Case $2\cap A\cap\alpha$ and $2\cap A\cap\gamma$ are the same situations as the case of "$i=j>1$" and "$i=j=1$" of the age-structured model in Felsenstein (1971), respectively. Therefore, I followed Felsenstein (1971) to adjust $H_{mm,t-1}$ by multiplying $(1-1/(2t_{im}N_m))/(1-1/(2N_m))$ and $1-1/(2f_{im}N_m)$ in case $2\cap A\cap\alpha$ and $2\cap A\cap\gamma$. Improving the explanation of Felsenstein (1971) to fit to my stage-structured model, I give detailed procedures on the adjustment of $H_{mm,t-1}$ in Appendix A1.1.2.

Substituting equations 2.8 to equation 2.7, $H_{ii,t}$ is formulated as follows.

$$
\begin{aligned}
H_{ii,t} = \sum_{m=1}^{n}\Bigg\{&\left(\frac{t_{im}N_m}{N_i}\right)^2\frac{1-1/(2t_{im}N_m)}{1-1/(2N_m)}+\frac{2t_{im}f_{im}N_m^2}{N_i^2}\\
&+\left(\frac{f_{im}N_m}{N_i}\right)^2\left(1-\frac{1}{2f_{im}N_m}\right)\Bigg\}H_{mm,t-1}
\end{aligned}
$$

$$+\sum_{k=1}^{n}\sum_{l>k}^{n}\frac{2a_{ik}a_{il}N_kN_l}{N_i^2}H_{kl,t-1}. \tag{2.9}$$

Combining case 1 (equation 2.6) and 2 (equation 2.9), I construct a matrix equation.

$$\boldsymbol{h_t} = \boldsymbol{M}\boldsymbol{h_{t-1}}. \tag{2.10}$$

$\boldsymbol{h_t}$ and $\boldsymbol{h_{t-1}}$ are vectors, each of which consists of $H_{ij,t}$ and $H_{ij,t-1}$ for all possible pairs of i and j ($1 \leq i \leq n$, $1 \leq j \leq n$). As the number of two-stage pairs is $n(n+1)/2$, both $\boldsymbol{h_t}$ and $\boldsymbol{h_{t-1}}$ have $n(n+1)/2$ elements. $\boldsymbol{M}$ is a square matrix whose dimension is $n(n+1)/2$ and whose elements are equal to the corresponding coefficients of $H_{mm,t-1}$ and $H_{kl,t-1}$ in equations 2.6 and 2.9. The order of elements in $\boldsymbol{h_t}$ is arbitrary as long as it matches with that in $\boldsymbol{h_{t-1}}$ and $\boldsymbol{M}$.

In general, multiplying matrix $\boldsymbol{M}$ is asymptotically the same as multiplying the dominant eigenvalue of $\boldsymbol{M}$, while $\boldsymbol{h_t}$ converges to a scalar multiplication of the leading right eigenvector, for sufficiently large t.

$$\boldsymbol{h_t} = \eta\boldsymbol{h_{t-1}}, \tag{2.11}$$

$$\boldsymbol{h_t} \propto \boldsymbol{w}, \tag{2.12}$$

where η and $\boldsymbol{w}$ are the leading eigenvalue and its corresponding right eigenvector of matrix $\boldsymbol{M}$, respectively. I denote w_{ij} as the element of $\boldsymbol{w}$ that corresponds to $H_{ij,t}$ of $\boldsymbol{h_t}$.

$$H_{ij,t} \propto w_{ij}. \tag{2.13}$$

Equation 2.11 means that $H_{ij,t}$ changes with a constant rate η over the course of time

for all i and j. Here, I denote H_t as the probability of non-identical-by-descent of the whole population in time t. H_t can be formulated as the sum of $H_{ij,t}$ weighted by the number of individuals in stage i and j.

$$H_t = \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{N_i N_j}{N^2}H_{ij,t} \propto \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{N_i N_j}{N^2}w_{ij}. \tag{2.14}$$

Because I assume that population size (N) and the number of individuals in a given stage i (N_i) are constant, H_t changes with the same rate as $H_{ij,t}$, that is, η.

$$H_t = \eta H_{t-1}. \tag{2.15}$$

Felsenstein (1971) also reached an analogous conclusion in his age-structured model that $H_{ij,t}$ and the probability of non-identical-by-descent of the overall population changed at the same rate, which was the largest eigenvalue. However, the proportionality between the array of $H_{ij,t}$ and the leading right eigenvector $\boldsymbol{w}$, which is shown in equations 2.12 and 2.13, was not mentioned in Felsenstein (1971).

Demographic genetic structure

I use the logarithm of the ratio of $H_{ii,t}$ between different stages as a proxy of demographic genetic structure, that is, comparison of genetic diversity among stages. With regard to the comparison between stage i and j, the logarithmic ratio is formulated as follows, based on equation 2.13.

$$\log\left(\frac{H_{ii,t}}{H_{jj,t}}\right) = \log\left(\frac{w_{ii}}{w_{jj}}\right), \tag{2.16}$$

When $\log(H_{ii,t}/H_{jj,t})$ is positive, $H_{ii,t}$ is larger than $H_{jj,t}$ (genetic diversity is higher in stage i than in stage j), and when negative vice versa. It should be noted that $\log(H_{ii,t}/H_{jj,t})$ is time-invariant, although $H_{ii,t}$ and $H_{jj,t}$ themselves change with time.

Effective population size

As in Felsenstein (1971), I formulate effective population size N_e using the dominant eigenvalue η. The probability of non-identical-by-descent of the overall population decreases with the rate of $1/(2N_e)$ per generation time (Crow and Kimura 1970).

$$H_{t+T} = \left(1 - \frac{1}{2N_e}\right) H_t, \tag{2.17}$$

where T is generation time and is defined as the mean age of net fecundity in the cohort (Carey and Roach (2020), see Appendix 1.1.3 for details). Considering that H_t changes with the rate of η per year (equation 2.15), $1 - 1/(2N_e)$ should be equivalent to η^T. Therefore, I formulate N_e as follows.

$$N_e = \frac{1}{2(1 - \eta^T)} \tag{2.18}$$

To sum up, demographic genetic structure and effective population size are derived from the leading right eigenvector and from the dominant eigenvalue of matrix $\boldsymbol{M}$, respectively. Therefore, my matrix model integrates the two proxies of the temporal genetic dynamics, facilitating comprehensive understandings on demographic genetic structure.

2.2.2 Validation of the model

To ensure that my model was formulated adequately, I compared theoretically obtained η and demographic genetic structure with observed ones computed by stochastic simulation. I

arranged a set of life histories to be used for the comparison between theory and simulation. I considered perennial plants with two ($n = 2$: juvenile and adult) and three stages ($n = 3$: seed, juvenile, and adult; Figure 2.2). Equation 2.10 can be rewritten as follows.

$$\begin{pmatrix} H_{11,t} \\ H_{22,t} \\ H_{12,t} \end{pmatrix} = \boldsymbol{M_2} \begin{pmatrix} H_{11,t-1} \\ H_{22,t-1} \\ H_{12,t-1} \end{pmatrix}, \tag{2.19}$$

and

$$\begin{pmatrix} H_{11,t} \\ H_{22,t} \\ H_{33,t} \\ H_{12,t} \\ H_{23,t} \\ H_{13,t} \end{pmatrix} = \boldsymbol{M_3} \begin{pmatrix} H_{11,t-1} \\ H_{22,t-1} \\ H_{33,t-1} \\ H_{12,t-1} \\ H_{23,t-1} \\ H_{13,t-1} \end{pmatrix}. \tag{2.20}$$

Equation 2.19 and 2.20 correspond to the case of $n = 2$ and $n = 3$, respectively. The elements of $\boldsymbol{M_2}$ and $\boldsymbol{M_3}$ are functions of demographic rates (t_{ij}, f_{ij}) and the number of individuals in each stage (N_i). For each of the two- and the three-stage model, I randomly generated five hundreds life histories which differed in t_{ij}, f_{ij}, and N_j, covering a wide range of life history strategies (Figure A1). I indirectly determined parameter values of t_{ij}, f_{ij}, and N_j. Firstly, the total population size N was set to 100, and then N was randomly divided into all possible survival and reproduction paths (i.e., $t_{ij}N_j$ and $f_{ij}N_j$). In the case of the two-stage model, for example, 100 individuals were randomly split into five paths: stasis at juvenile, growth from juvenile to adult, retrogression from adult to juvenile, stasis at adult, and reproduction (Figure 2.2a). Next, N_i was calculated by $\sum_{i=1}^{n}(t_{ij}N_j + f_{ij}N_j)$, and

finally t_{ij} and f_{ij} were calculated by $t_{ij}N_j/N_j$ and $f_{ij}N_j/N_j$ respectively (see Appendix 1.2 for details). By determining $t_{ij}N_j$ and $f_{ij}N_j$ first, I could easily search the parameter space while keeping the number of individuals (i.e., N_i, $t_{ij}N_j$, and $f_{ij}N_j$ for all i and j) to be always integer. To consider the situation of $N = 500$ and $N = 1,000$, I multiplied N_1 and N_2 (when $n = 3$, N_3 as well) by 5 and 10 while keeping demographic rates unchanged. In total, I considered 1,500 sets of parameter values (500 sets of demographic rates $\times$ 3 sets of N) for each of the two- and the three-stage model.

For each parameter set, I simulated 200 years of temporal dynamics of expected heterozygosity at a neutral biallelic locus 100 times. I calculated the mean expected heterozygosity over the 100 replicates for the overall population and for all the two-stage pairs at every t, which were denoted by $\hat{H}_t$ and $\hat{H}_{ij,t}$, respectively. All simulations were initiated with maximum expected heterozygosity, in which two alleles share the gene pool half-and-half in all stages (i.e., $H_0 = H_{ij,0} = 0.5$ for all i and j). It should be noted that the initial state of equal gene frequencies among classes corresponds to a genetic equilibrium under no evolutionary forces (i.e., drift, selection, mutation and gene flow) (Charlesworth 1994). Therefore, it could be said that my simulation results reflected how genetic drift solely decreased genetic diversity in stage-structured populations.

I calculated the annual change rate of $\hat{H}_t$ by

$$r_t = \frac{\hat{H}_t}{\hat{H}_{t-1}}, \tag{2.21}$$

where $1 \leq t \leq 200$. I took logarithm of r_t and calculated its mean and standard error, which were subsequently compared to η. η is the theoretical counterpart r_t and was obtained as the dominant eigenvalue of matrix $\boldsymbol{M_2}$ or $\boldsymbol{M_3}$.

Using simulation results, I also calculated the mean of demographic genetic structure over the 200 years. As for the two-stage model, I calculated $\log(\hat{H}_{11,t}/\hat{H}_{22,t})$. I calculated $\log(\hat{H}_{11,t}/\hat{H}_{22,t})$, $\log(\hat{H}_{22,t}/\hat{H}_{33,t})$ and $\log(\hat{H}_{11,t}/\hat{H}_{33,t})$ in the case of the three-stage model. These four proxies of observed demographic genetic structures were compared to theoretical counterparts, that is, $\log(H_{11,t}/H_{22,t})$ for the two-stage model, as well as $\log(H_{11,t}/H_{22,t})$, $\log(H_{22,t}/H_{33,t})$ and $\log(H_{11,t}/H_{33,t})$ for the three-stage model. These four logarithmic ratios were obtained by solving the leading right eigenvector of $\boldsymbol{M_2}$ and $\boldsymbol{M_3}$ and substituting their elements to equation 2.16.

2.2.3 Analysis on demographic genetic structure

For the same 3,000 parameter sets as "Validation of the model" section, I analytically obtained η and N_e, which reflect the change rate of all $H_{ij,t}$ per year and per generation, respectively. η was obtained by solving the dominant eigenvalue of $\boldsymbol{M_2}$ and $\boldsymbol{M_3}$. Then, using η, I obtained N_e based on equation 2.18. I examined if η and N_e, both of which genuinely represent temporal dynamics of genetic diversity, were correlated with the four logarithmic ratios that stood for demographic genetic structure (i.e., $\log(H_{11,t}/H_{22,t})$ for the two-stage model, and $\log(H_{11,t}/H_{22,t})$, $\log(H_{22,t}/H_{33,t})$ and $\log(H_{11,t}/H_{33,t})$ for the three-stage model) to judge if demographic genetic structure could serve as a proxy for temporal dynamics of genetic diversity across a wide range of life history strategies.

2.2.4 Determinants of demographic genetic structure and N_e

To explore the determinants of demographic genetic structure, I analyzed the dependence of demographic genetic structure on total population size N and stable stage distribution using the 3,000 parameter sets. Stable stage distribution was quantified by the logarithm of the

ratio among N_1, N_2, and N_3 (i.e., log(N_1/N_2), log(N_2/N_3), and log(N_1/N_3)). Besides, to understand whether static and retrogressive flows in the life history, which are characteristic to stage-structured populations, affects demographic genetic structure, I divided the number of individuals in each stage into progressive and static inflows. Here, I denote the number of individuals that join stage i through progressive and static flows by $N_{i,p}$ and $N_{i,s}$, respectively, which can be obtained as follows.

$$N_{i,p} = \sum_{j=1}^{i-1} t_{ij} N_j + \sum_{j=1}^{n} f_{ij} N_j, \tag{2.22}$$

$$N_{i,s} = \sum_{j=i}^{n} t_{ij} N_j \tag{2.23}$$

Here, the progressive flow includes not only growth but also reproduction, because both growth and reproduction promotes generation turnover along the life cycle. I classified both stasis and retrogression into the static flow, following Silvertown et al. (1993) and Silvertown et al. (1996). I calculated the logarithmic ratios among stages (i.e., log($N_{i,p}/N_{j,p}$) and log($N_{i,s}/N_{j,s}$) for all i and j) and examined their correlation with demographic genetic structure log($H_{ii,t}/H_{jj,t}$).

I also analyzed the determinants of η and N_e with respect to total population size N and eight life history traits: generation time (T), survivorship curve type (H), age at sexual maturity (L_α), mean probability of progressive growth (γ) and retrogressive growth (ρ), the degree of iteroparity (S), mean reproductive rate (ϕ), and mature life expectancy (L_ω). These life history traits are known to explain the variation in life history strategies in plants (Capdevila et al. 2020, Salguero-Gómez et al. 2016). I calculated these traits from demographic rates following the methods of Caswell (2001) and Salguero-Gómez et al. (2016). Generation time is the mean age of net fecundity in the cohort, as explained before. Survivorship

curve type H is Keyfitz' entropy, which is a continuous variable that reflects the shape of survivorship curve from type I (high mortality at mature ages, $H < 1$), II (constant mortality throughout the lifespan, $H = 1$), to III (high mortality at young ages, $H > 1$). age at sexual maturity L_α is the average age to become sexually reproductive. Mean probability of progressive and retrogressive growth (γ and ρ) and mean reproductive rate (ϕ) are the number of individuals that grow, retrogress, and are born per year divided by the total population size, respectively. The degree of iteroparity S represents the temporal spread of reproductive events along lifespan and is quantified by Demetrius' entropy. Mature life expectancy L_ω is estimated by subtracting L_α from life expectancy of a newborn. I carried out principal component analysis (PCA) of the eight life history traits for each N and for each of the two- and the three-stage model, and examined how the resultant components are correlated with η and N_e.

2.3 Results

2.3.1 Validation of the model

The rate of change in expected heterozygosity of the overall populations (r_t), which was computed by simulation, took almost exactly the same value as the theoretical counterpart η for all 1,500 sets of parameter values in both the two- and the three-stage models (Figure 2.3, A2).

Comparison of demographic genetic structure between simulation and analytical results revealed that my theoretical model yielded almost equivalent logarithmic ratio of expected heterozygosity among stages to that of simulation (Figure 2.4, A3, A4).

To further confirm the validity of my model, I checked the temporal dynamics of $\hat{H}_{ij,t}$ and compared it with theoretical expectation, that is, the repeated multiplication of matrix

$\boldsymbol{M_2}$ or $\boldsymbol{M_3}$ to $\boldsymbol{h_t}$. I found that theoretical prediction fitted well to simulation results (Figure A5).

Thus, my model seems to describe the dynamics and the inter-stage ratio of expected heterozygosity validly across a wide range of parameter space.

2.3.2 Analysis on demographic genetic structure

All the four proxies of demographic genetic structure, which are theoretically obtained based on equation 2.16, have an apparent correlation neither with N_e nor with η regardless of N (Figure 2.5, A6, A7). Moreover, there is a strong positive correlation with stable stage distribution: expected heterozygosity is higher in stages with more individuals (Figure 2.7). When separating the number of individuals (N_i) into progressive and static flows ($N_{i,p}$ and $N_{i,r}$), the positive correlation with demographic genetic structure remains clearly in the logarithmic ratios of static inflows: expected heterozygosity is high in stages that contain larger number of static and retrogressive individuals. The correlation becomes weaker with increasing N, as logarithmic ratios converge to zero.

2.3.3 Analysis on effective population size and η

With increasing N, effective population size N_e increased and η converged to 1 (Figure 2.9), indicating that expected heterozygosity is not likely to decay under large census population size.

PCA yielded two axes that could jointly explain 75.5 % and 67.8 % of the total variation in the eight life history traits in the two- and the three-stage model, respectively (Figure 2.10). Judging from the factor loadings, the first axis PC1 was negatively correlated with mean reproductive rate ϕ and positively with generation time T, degree of iteroparity S, and

survivorship curve type H. In other words, PC1 reflected the gradient from long-lived (large T) and iteroparous (large S) life histories with low fecundity (low ϕ), to short-lived, highly reproductive, and semelparous ones with high mortality at mature ages. This correlation exactly corresponds to "the fast-slow continuum," which describes that the high mortality risk after sexual maturation is related to fast growth, high fecundity, and short lifespan.

When the total population size N was large, η was almost equivalent across the principal component space (Figure A8, A9). Under small N, η became divergent among life history strategies: η increased along the gradient from fast- to slow-paced life history (from large to small PC1 scores, Figure 2.10 (a, c)). In other words, η was high under low annual fecundity, long generation time, high degree of iteroparity, and survivorship curve type III (concentrated death at the beginning of the life cycle). PC2 did not show strong correlation with η (Figure 2.10 (a, c)). As for N_e, no apparent correlation was found with either PC1 or PC2 (Figure 2.10 (b, d)).

2.4 Discussion

2.4.1 Comparison with the age-structured model

In this study, I develop the matrix model that describes the dynamics of genetic diversity and demographic genetic structure in stage-structured populations. Although the procedures of model development are similar to the age-structured model in Felsenstein (1971), my model has a much wider applicability. First of all, because age-structured models, in which the probabilities of stasis and retrogression are zero, is a special case of stage-structured models, my model is more comprehensive. Besides, many plant species do not show demographic senescence (Jones et al. 2014), showing no age-dependent changes in demographic rates. Using stage-dependent demographic parameters would be more appropriate and predictive in

plant populations. These points support the novelty of my stage-structured model, especially in terms of expanding the applicability to many plant species.

2.4.2 Interpreting demographic genetic structure

A common interpretation on demographic genetic structure is that if juvenile stages are less diverse than mature stages, genetic diversity would decrease with time over the course of generation turnover (Aldrich et al. 1998, Ally and Ritland 2006, Kettle et al. 2007, Linhart et al. 1981, Murren 2003, Schmidt et al. 2018, Vranckx et al. 2014). However, my model shows that relative ratio of expected heterozygosity between stage classes does not correlate with either N_e or η: even though N_e and η are small, expected heterozygosity does not necessarily decline from mature to juvenile stages. Therefore, inferring temporal trends in genetic diversity solely from demographic genetic structure is potentially misleading. This study, to my knowledge, for the first time draws caution on the conventional use of demographic genetic structure.

Many previous empirical studies that analyzed demographic genetic structure found that genetic diversity did not decrease from the most mature to the most immature stages and took comparable values among stages (Aldrich et al. 1998, Ally and Ritland 2006, Kettle et al. 2007, Linhart et al. 1981, Murren 2003, Schmidt et al. 2018, Vranckx et al. 2014). My model shows that the logarithmic ratio of expected heterozygosity is distributed around zero, especially under large N, indicating that expected heterozygosity is basically almost equivalent to one another. Therefore, my model might be in line with previous empirical results.

While demographic genetic structure is irrelevant to temporal dynamics, it is tightly linked to stable stage distribution and especially the number of static and retrogressive individuals:

expected heterozygosity is relatively higher in stage with more static and retrogressive inflows (Figure 2.6). In general, small number of individuals intensifies stochastic genetic drift due to increased sampling bias in gene frequencies (Crow and Kimura 1970). When stage distribution is skewed, the degree of stochasticity will vary among stages. Stage with smaller number of individuals is made up of genes that were sampled fewer times from the gene pool of the previous year, thus suffering random perturbation in gene frequencies to a greater extent. The amplified stochasticity must have resulted in the lower genetic diversity. Importantly, the inter-stage difference is mainly driven by the amount of stasis and retrogression. Stasis and retrogression mixes differently-aged cohorts within and among stages: when retrogressive inflows are large, individuals old in age can remain in, or come back to, juvenile stages and coexist with younger ones. Overlap of cohorts could offset temporal random changes caused by genetic drift, accumulating genetic variation across a wide temporal breadth. This mechanism is similar with the storage effect, which was previously suggested by studies showing that generation overlap maintains genetic diversity under fluctuating selection pressures (Ellner and Hairston Jr 1994, Ellner 1996).

As the total population size N increases, inter-stage difference in genetic diversity disappears even when the stage distribution and the retrogressive inflows are skewed (Figure 2.7). This result indicates that the number of individuals of each stage is large enough to reduce stochasticity under large N, leading to comparable level of genetic diversity among stages.

To sum up, it can be said that genetic diversity becomes uneven among life history stages under small population size, and that the unevenness among stages reflects stable stage distribution and inflows of static and retrogressive individuals rather than the temporal dynamics of genetic diversity.

2.4.3 Temporal dynamics of genetic diversity under a variety of life history strategies

η became large with increasing PC1 scores (Figure 2.10), indicating that high η is realized in slow-paced life histories. The reason why slow-paced strategies maintain higher genetic diversity can be attributed to generation overlap. In populations with long generation time, old individuals remain alive and coexist with young ones, which facilitates generation overlap and the accumulation of genetic variation across broader time span (Ellner and Hairston Jr 1994). Although the contribution of longevity to genetic diversity had been suggested (Austerlitz and Garnier-Géré 2003, Hamrick and Godt 1996, Tsuzuki et al. 2022a, Aparicio et al. 2012), most studies handled only a limited number of life history strategies. Combined with the results on demographic genetic structure, my results successfully showed that slow and static growth along the life history contributes to the maintenance of genetic diversity at the level of life history stage and of the whole population in the overall life history spectrum.

On the other hand, the relation ship with the fast-slow continuum is less clear in N_e than in η. Considering the positive correlation between η and generation time, it might be that although life histories with high η do not lose genetic diversity very much annually, they spend much more time for generation turnover, thereby losing equivalent level of genetic diversity with fast-paced life histories per generation time.

Population size also contributes to the maintenance of genetic diversity, because N_e increases and η approaches to 1 with increasing N (Figure 2.9). Under large N, the variation in η among life history strategies is small, indicating that population size is large enough to diminish stochastic diversity loss in all types of life history strategies. N_e shows the contrasting pattern: the variation in N_e becomes large with increasing N. This is probably due to the mathematical relationships between N_e and η. Because the two statistics are

inversely related (Equation 2.18), large variation in one hand corresponds to small variation in the other.

Overall, the temporal dynamics of genetic diversity of the overall population are similarly subjected to life history and census population size as stage-wise genetic diversity. Large census population size diminishes stochastic diversity loss and obscures the impacts of generation overlap. Under small N, the contribution of generation overlap to genetic diversity becomes apparent, and life histories, or life history stages, with high degree of generation overlap possesses large amount of genetic variation.

2.4.4 Efficacy of theoretical prediction on demographic genetic structure

Providing theoretical background of demographic genetic structure, my model has some potential for application. One possibility is to compare raw demographic genetic structure, which is obtained by any neutral genetic markers, with the theoretical expectation calculated based on my model. Because our model can handle life histories with any number of classes and with any form of survival and reproduction flows among stages, a wide range of life histories including both inter- and intraspecific variation can be analyzed. The deviations of observed structure from expectation should reflect factors unexplored in the model. My model can work as a null model of demographic genetic structure. To make the most use of my model, it is necessary to monitor individuals from year to year to estimate demographic rates of each stage class. If long-term demographic monitoring is unavailable or impractical for some reasons, recording relative number of individuals among stage classes at a single time point would be at least desirable to consider stage distribution, which turned out to be a major determinant of demographic genetic structure.

While the conventional interpretation as a proxy of temporal trends seems to be invalid,

theoretical prediction on demographic genetic structure can also provide insights into how life history controls the maintenance of genetic diversity and adaptive evolution. It is known that a particular life history stage serves as a reservoir of genetic variation within a population and potentially contribute to adaptation to environmental changes (Yamamichi et al. 2019). For example, dormant seed bank stage possesses high standing genetic variation compared to vegetative stages (Nunney 2002). The accumulated genetic diversity in the seed bank contributes to rapid adaptation to above-ground environmental changes by supplying a variety of genotypes to vegetative stages through recursive germination (Agrawal et al. 2021, Orsini et al. 2013). In tree species, the above-ground adult stage consist of variety ages, thereby accumulating genetic diversity and increasing the adaptive capacity neutrally (Cannon et al. 2022). By estimating demographic genetic structure theoretically, we can understand which life history stages possess high genetic diversity, and can identify life history processes that are crucial in accumulating genetic variation and rapid adaptation. Considering that demographic genetic structure is variable under small N, the analysis on demographic genetic structure is especially applicable to small endangered populations, highlighting crucial life history stages where conservation actions are needed. The efficacy of predicting demographic genetic structure from my model should be further evaluated in wild populations.

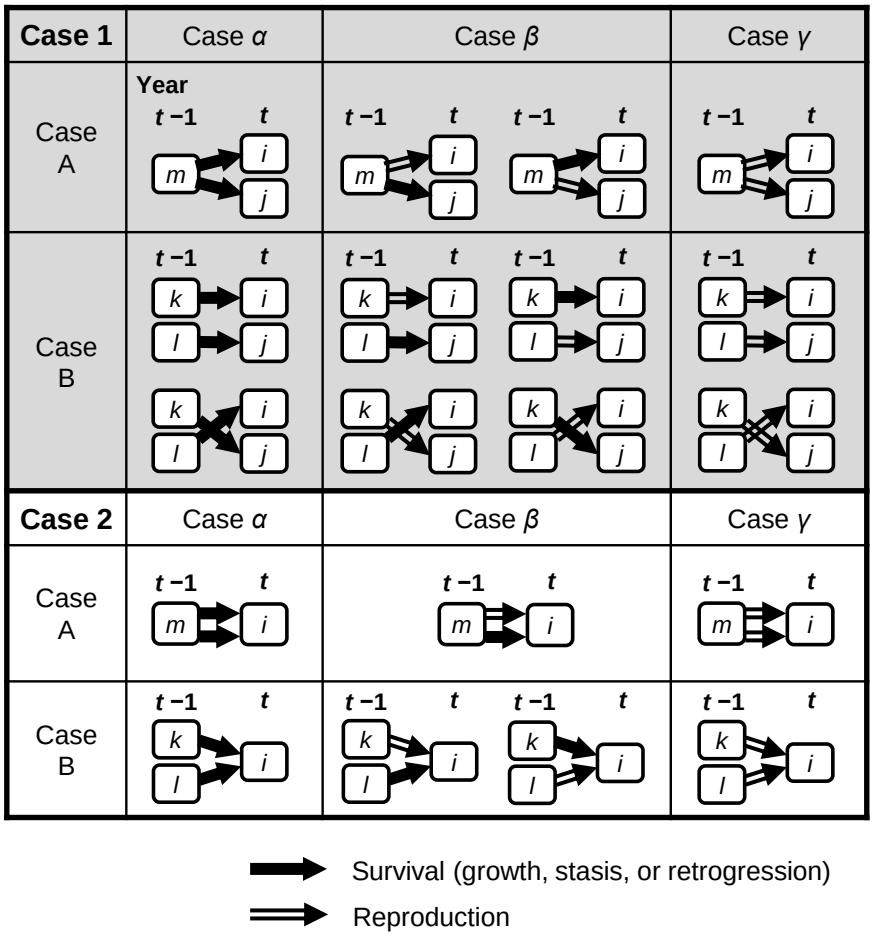


Figure 2.1: Temporal trajectories from time $t-1$ to t with regard to the two genes sampled in time t. Rounded rectangles stand for life history stages. Arrows stand for the temporal movements of genes either by survival (single line) or reproduction (double line). There are 12 mutually exclusive situations based on three criteria: (1) whether the destinations are different (case 1, shown on gray background) or not (case 2, shown on white); (2) whether the origins are the same (case A) or not (case B); (3) how the two genes were transferred (case α: survival; case β: survival and reproduction; case γ: reproduction)

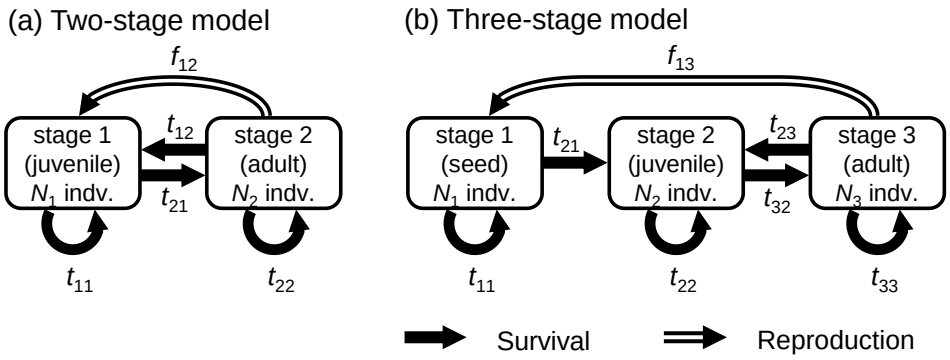


Figure 2.2: The two model used in analysis: (a) two-stage model and (b) three stage model. Arrows represent flow of individuals, or genes, either by survival (single line) or reproduction (double line)

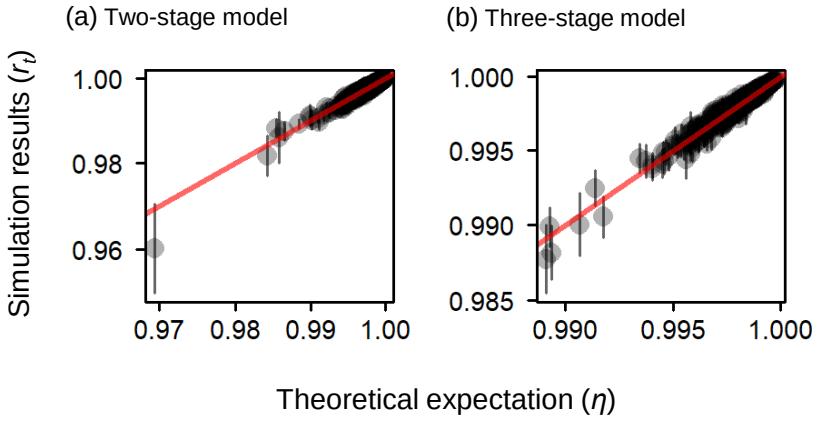


Figure 2.3: Comparison between the theoretical expectation of the annual change rate of the probability of non-identical-by-descent (η) and the simulation results of that of expected heterozygosity (r_t) for (a) the two-stage and (b) the three-stage model when $N = 100$. Each gray semi-transparent point corresponds to one of the 500 parameter sets. As for r_t, geometric mean over $1 \leq t \leq 200$ is shown with standard error (vertical bar). The red lines represent $\eta = r_t$

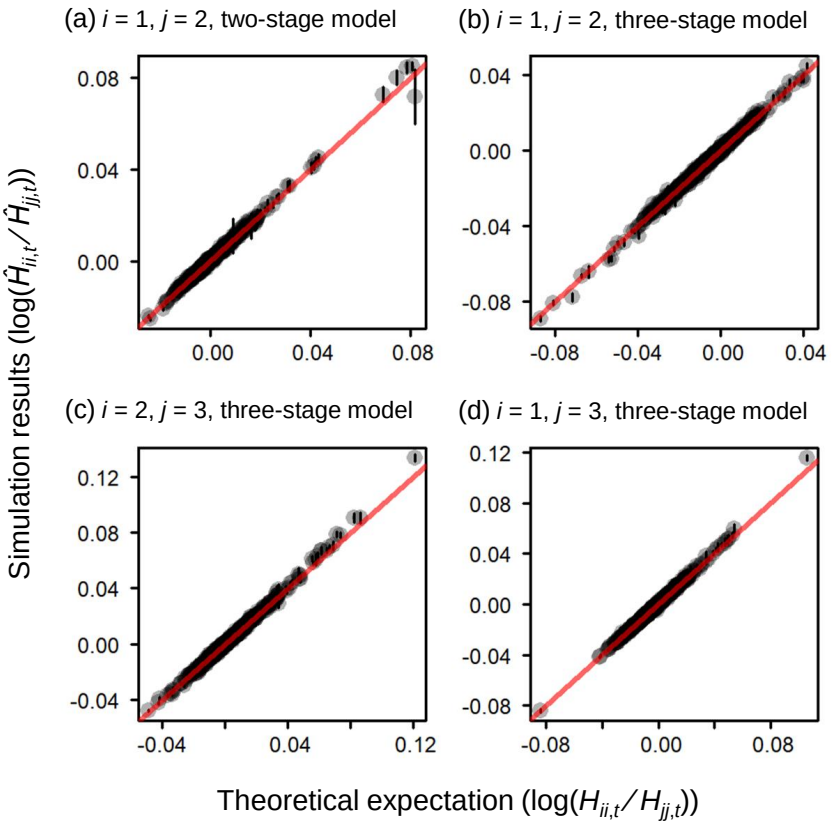


Figure 2.4: Comparison of demographic genetic structure between the theoretical expectations ($\log(H_{ii,t}/H_{jj,t})$) and the simulation results ($\log(\hat{H}_{ii,t}/\hat{H}_{jj,t})$) when $N = 100$. Each gray semi-transparent point corresponds to one of the 500 parameter sets. As for the simulation results, mean and standard error (vertical bar) over $1 \leq t \leq 200$ are shown. There is one proxy for the two-stage model (a: $i = 1$ and $j = 2$), while there are three proxies for the three-stage model (b: $i = 1$ and $j = 2$; c: $i = 2$ and $j = 3$; d: $i = 1$ and $j = 3$). The theoretical expectations exactly match with the simulation results when plotted on the red lines

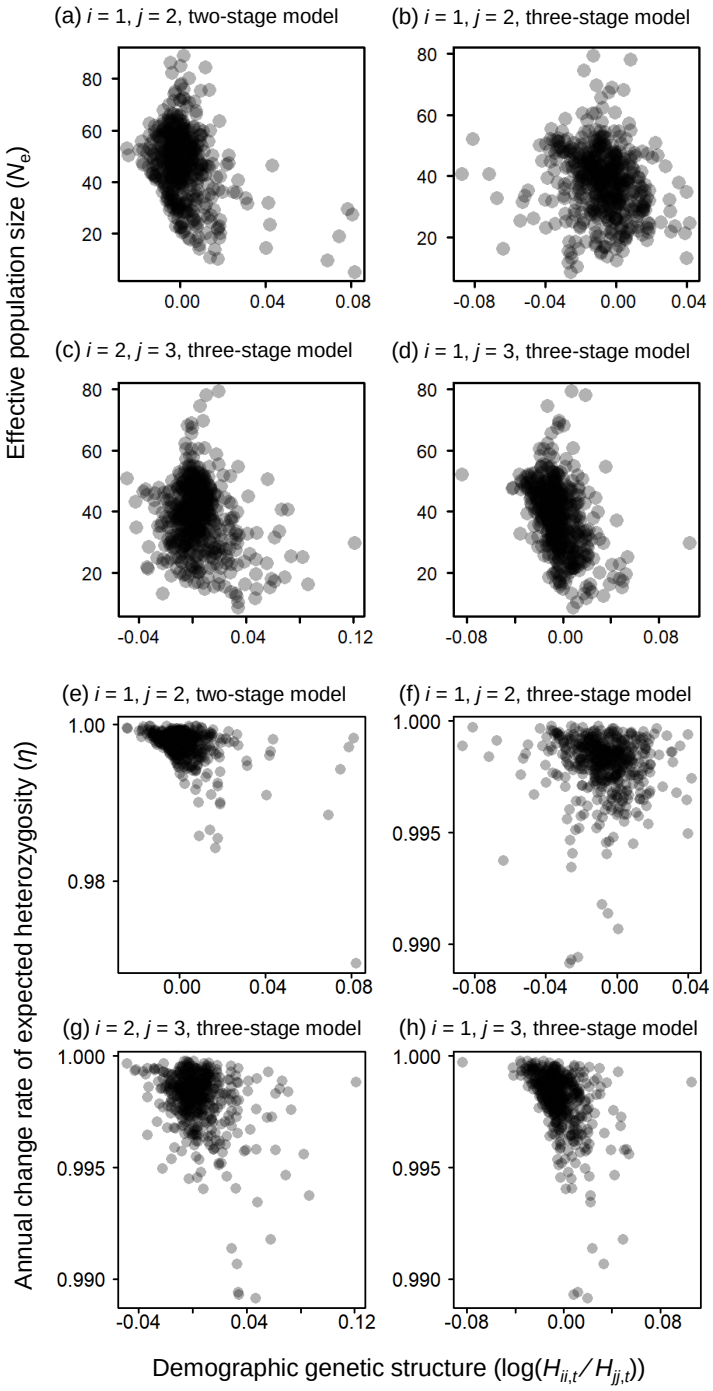


Figure 2.5: Comparison of demographic genetic structure ($\log(H_{ii,t}/H_{jj,t})$) with effective population size N_e (a-d) and the annual change rate of expected heterozygosity η (e-h) when $N = 100$. (a, e) $i = 1$ and $j = 2$ of the two-stage model, (b, f) $i = 1$ and $j = 2$, (c, g) $i = 2$ and $j = 3$, (d, h) $i = 1$ and $j = 3$ of the three-stage model

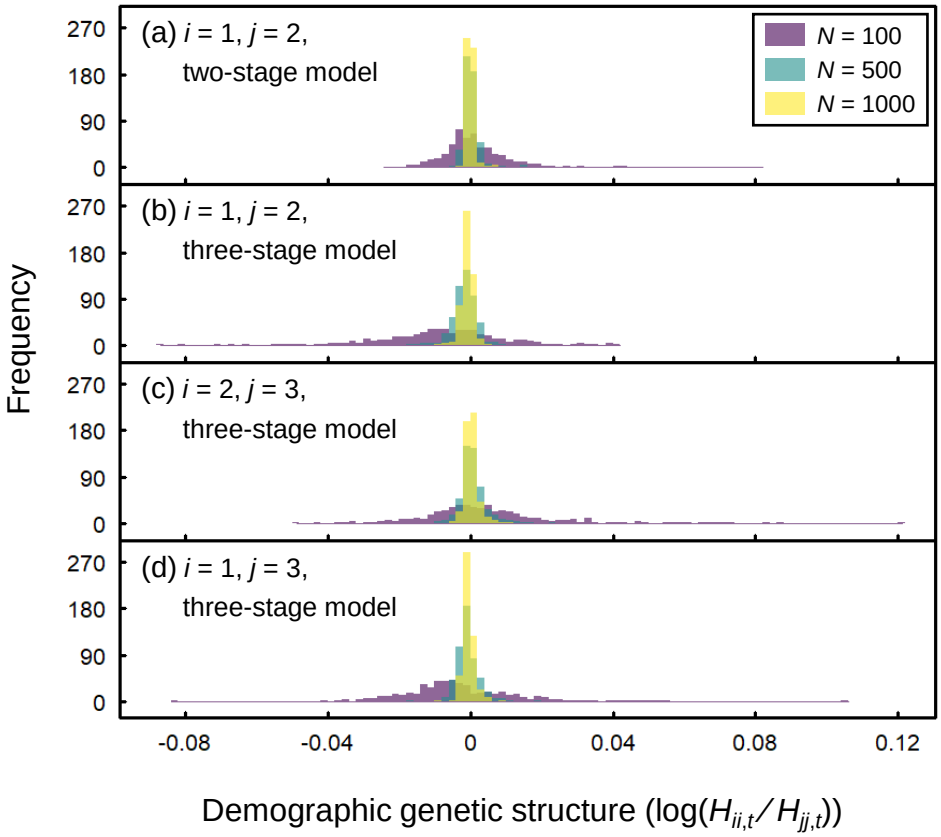


Figure 2.6: Histogram of demographic genetic structure ($\log(H_{ii,t}/H_{jj,t})$) with varying N. (a) $\log(H_{11,t}/H_{22,t})$ of the two-stage model, (b) $\log(H_{11,t}/H_{22,t})$, (c) $\log(H_{22,t}/H_{33,t})$, (d) $\log(H_{11,t}/H_{33,t})$ of the three-stage model

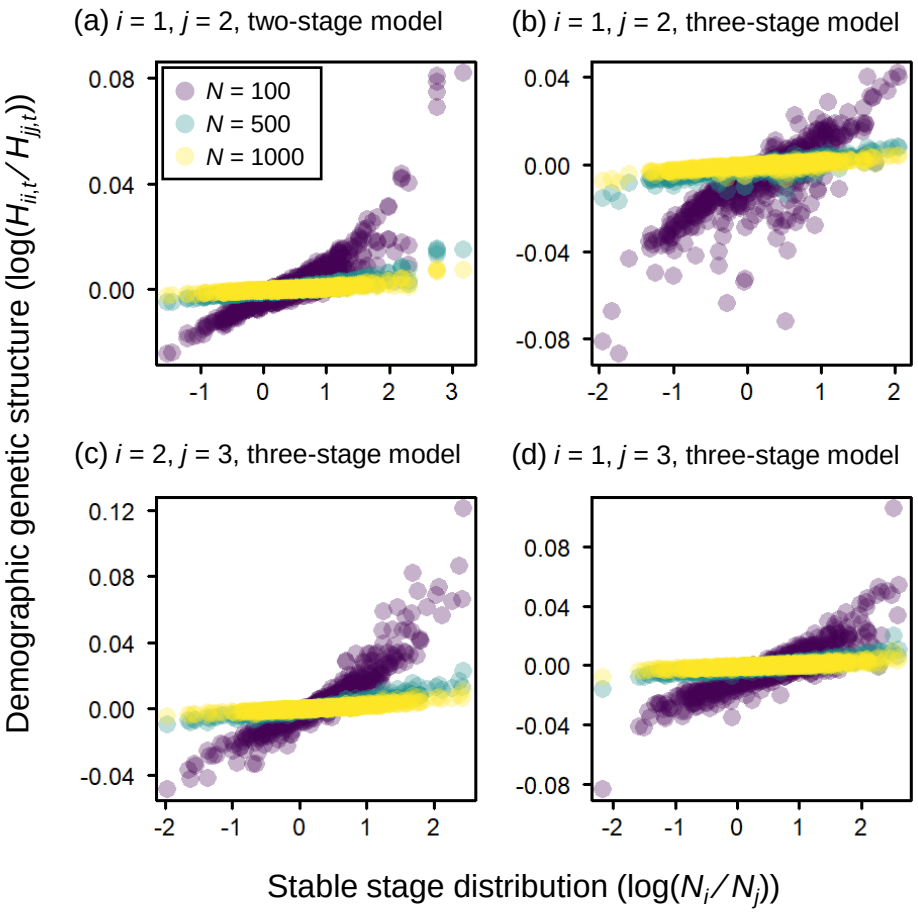


Figure 2.7: Relationships between stable stage distribution ($\log(N_i/N_j)$) and demographic genetic structure ($\log(H_{ii,t}/H_{jj,t})$) with varying N

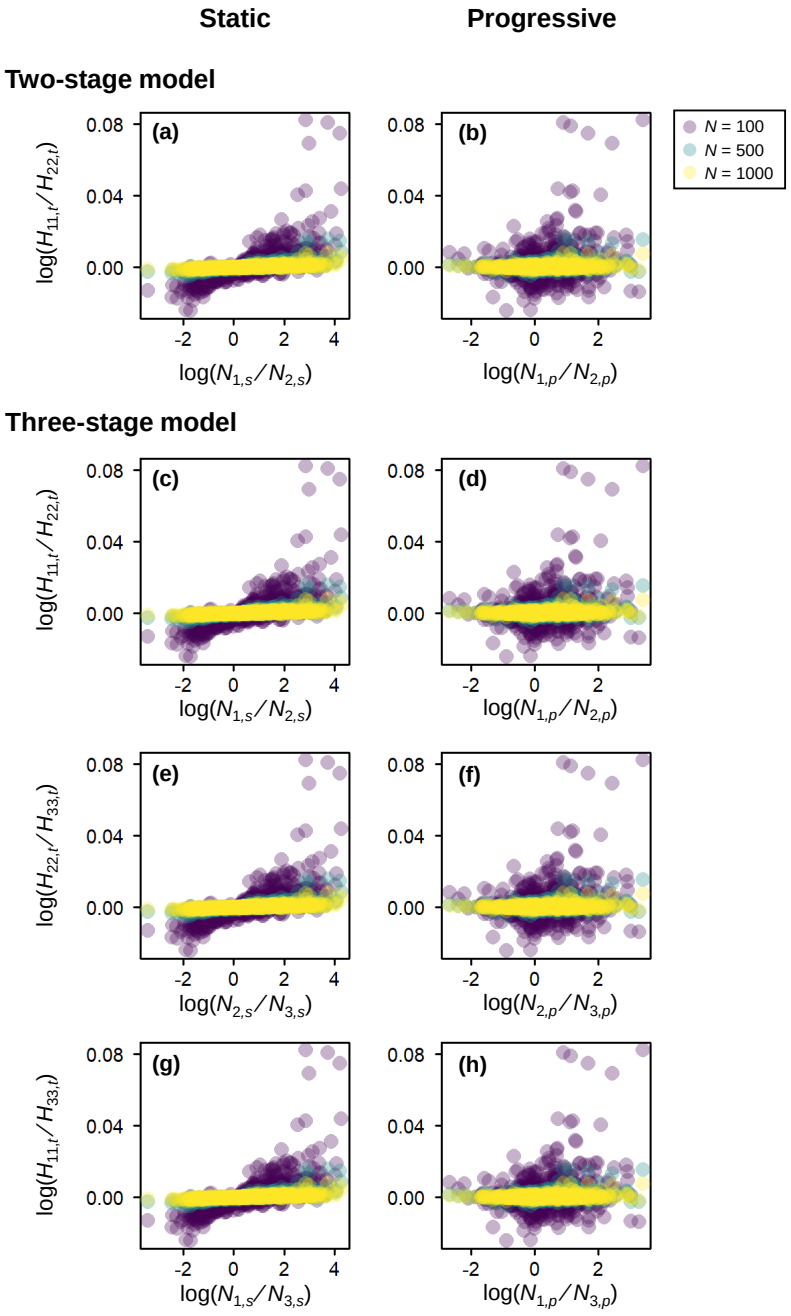


Figure 2.8: Relationships between logarithmic ratio of static and progressive flows ($\log(N_{i,s}/N_{j,s})$ and $\log(N_{i,p}/N_{j,p})$) and demographic genetic structure ($\log(H_{ii,t}/H_{jj,t})$) with varying N. (a, c, e, f) Static flows and (b, d, f, h) progressive flows. (a, b) $i = 1, j = 2$ of the two-stage model, (c, d) $i = 1, j = 2$ of the three-stage model, (e, f) $i = 2, j = 3$ of the three-stage model, (g, h) $i = 1, j = 3$ of the three-stage model

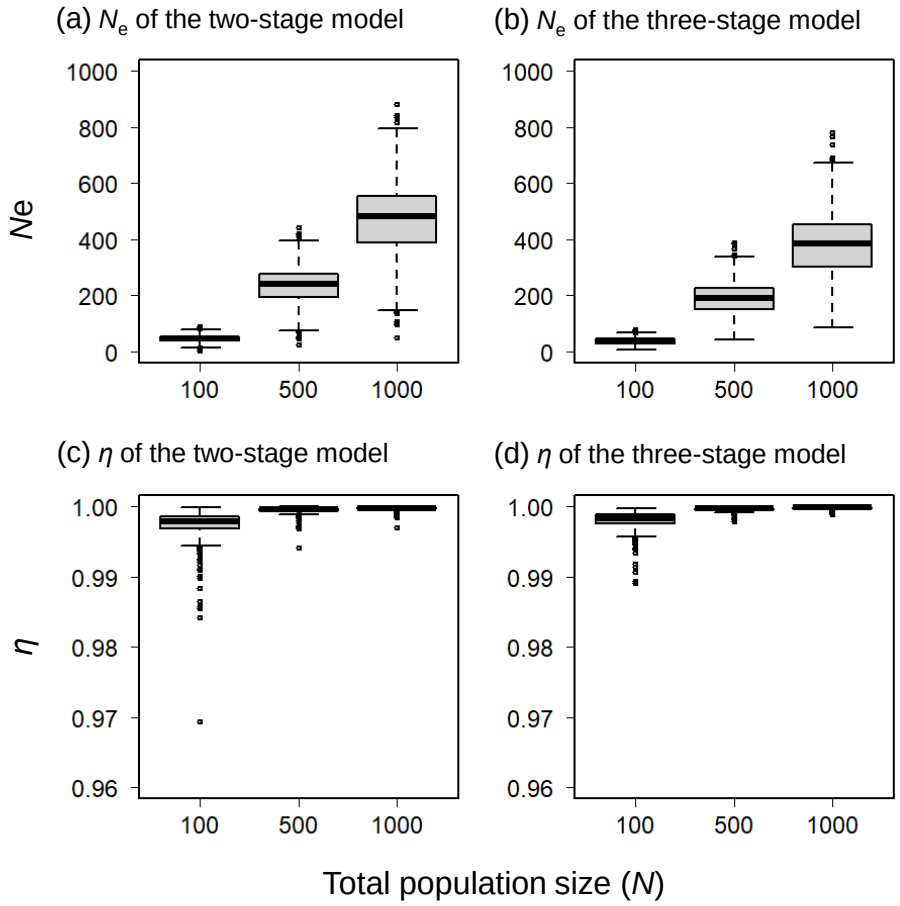


Figure 2.9: Effective population size (N_e) and the annual change rate of expected heterozygosity (η) of the 500 parameter sets for each of $N = 100, 500$, and 1000. (a) N_e of the two-stage model, (b) N_e of the three-stage model, (c) η of the two-stage model, (d) η of the three-stage model

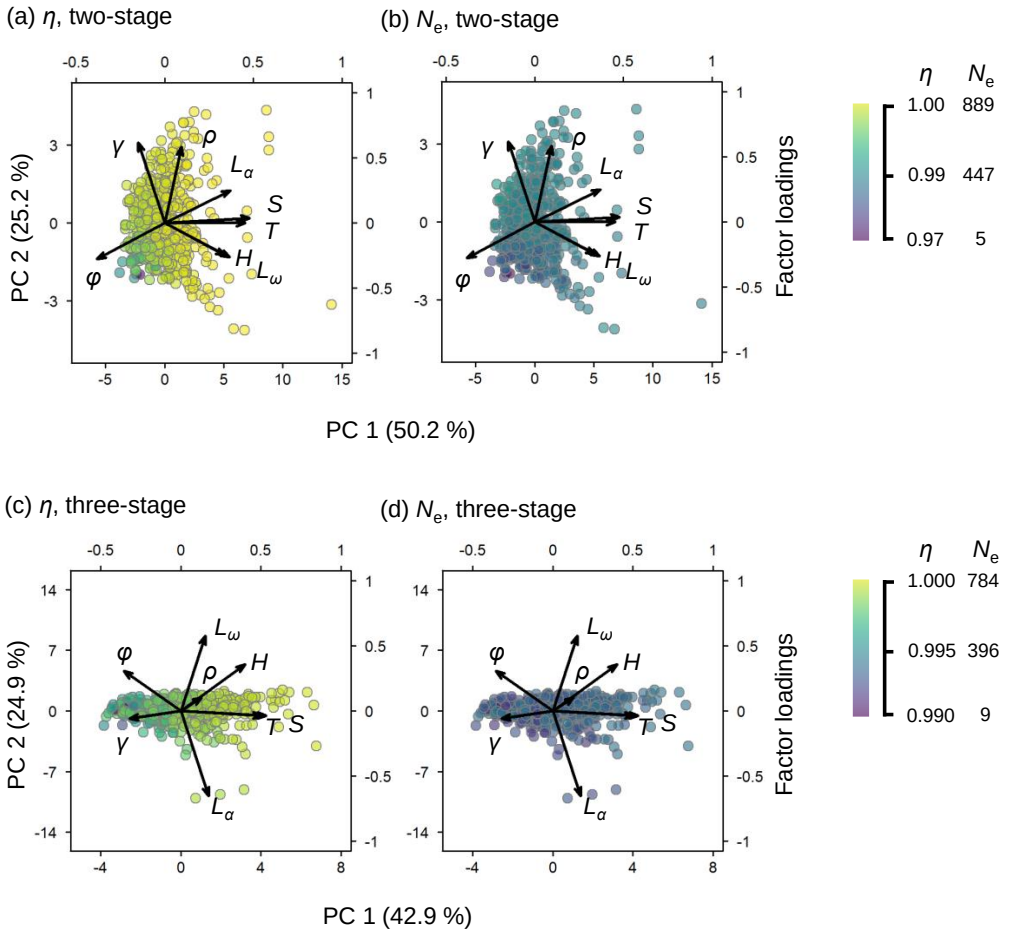


Figure 2.10: Principal components of the eight life history traits of the 500 parameter sets in the two- and the three stage model. Color denotes either annual change rate of expected heterozygosity (η) or effective population size (N_e) when $N = 100$. (a) η of the two-stage model, (b) N_e of the two-stage model, (c) η of the three-stage model, (d) N_e of the three-stage model

Chapter 3

Genetic impacts of intraspecific demographic variation in fragmented populations of a perennial herb *Trillium camschatcense*

Abstract

Generally, genetic diversity decreases over time in small populations due to genetic drift. While it has been shown that some life history characteristics, such as long lifetime and slow maturation, promote generation overlap and maintain genetic diversity despite small population size, the relative importance in determining the level of genetic diversity between population size and life history are not empirically clarified. In this study, I examined the determinants of genetic diversity by comparing theoretical expectation and empirical data of genetic diversity in 11 fragmented populations of a perennial herb *Trillium camschatcense* in Tokachi region, Hokkaido, northern Japan. The life history of *T. camschatcense* can be split into 5 developmental stages (seed, seedling, one-leaf, three-leaves, flowering), and I estimated stage-specific survival rates and fecundities and population size through four years of demographic field census. The estimated values were used to obtain theoretical expectation of demographic genetic structure, which is the stage-wise genetic diversity, and effective population size, using the model developed in Chapter 2. In parallel, I collected a total of 512 samples to empirically estimate demographic genetic structure and effective population size by double digest restriction site associated DNA sequencing (ddRAD-seq). Although genetic diversity was predicted to be very close among life history stages in the theoretical expectation, the observed genetic diversity was uneven among stages, which is the symptoms of small population size. As for effective population size, theoretical prediction was larger than the observation. These mismatches between the theoretical prediction and the observation can be explained by the possibility that not all individuals within a population survive and reproduce randomly in terms of genes and that individuals that effectively serve as 'population' would be spatially limited and be much more smaller than the total number of individuals. Because uneven demographic genetic structure indicates that a particular life history stage serves as a reservoir of genetic diversity in the life cycle, the empirical results suggest that the contribution of life history to maintaining genetic diversity could be larger than expected from the theory.

3.1 Introduction

Anthropogenic land-use change has been causing habitat loss and fragmentation worldwide (Fahrig 2003), which is predicted to decrease genetic diversity of remnant populations through bottleneck, genetic drift, restricted gene flow and inbreeding (Aguilar et al. 2008, Lowe et al. 2005, Young et al. 1996). Reduction in genetic diversity not only constrains the evolutionary potential (Bakker et al. 2010, Ramsayer et al. 2013), but could also suppress individual fitness and population growth (Reed and Frankham 2003, Williams 2001), thus threatening population viability. Hence, it is necessary to evaluate and mitigate the genetic consequences of habitat fragmentation on remnant populations.

Although many studies indicate that population shrinkage is a major driver of genetic diversity loss in fragmented landscapes (Leimu et al. 2006, Lowe et al. 2005), the susceptibility to population shrinkage differs among species depending on their life history characteristics (González et al. 2020, Kramer et al. 2008). Life history is a lifetime trajectory of individuals from birth to death, and because genetic information is stored in individuals, genetic dynamics is fundamentally subjected to how individuals survive, grow, and reproduce over the course of life history. In perennial plants, for example, slow growth and prolonged lifespan could allow individuals to persist for many years, resulting in slow generation turnover and delayed loss of genetic diversity (Aparicio et al. 2012, Martins et al. 2015). It is important, therefore, to account for life history cycle for assessing the genetic impacts of habitat fragmentation.

Meanwhile, it has been shown that habitat loss and fragmentation not only downsizes population but also alters survival and reproductive performance of individuals in remnant populations through environmental changes (Fahrig 2003). In general, environmental conditions are different between the core and the edge within a habitat, and small fragmented habitats

become more subjected to edge environments, which is known as edge effect. The edge effect influences survival and reproduction at various stages of life history. In some plant species, it is known that the survival of seedlings, as well as reproductive success of flowering stages, are restricted in small fragmented populations due to deteriorated environmental conditions (Aguilar et al. 2019, Benítez-Malvido et al. 2018, Tomimatsu and Ohara 2004). The changes in the demographic performance caused by edge effects can affect how genes are transferred and distributed among stage classes, potentially influencing the temporal dynamics of genetic diversity. However, it remained unclear how the changes in life history interact with population shrinkage and affect genetic diversity. This is mainly because of methodological limitation: the mathematical relationships between genetic diversity and stage-specific demographic parameters were theoretically unclear. Tsuzuki et al. (2022b), which is Chapter 2 of my dissertation, formulated genetic diversity of each life history stage (i.e., demographic genetic structure) and its temporal dynamics using stage-specific demographic rates and population size. The model allows us to predict how genetic diversity changes over time and which life history stage serves as a reservoir of genetic diversity under a given life history and a population size, and has paved the road to predict the genetic consequences of habitat fragmentation.

Here, I focused on fragmented populations of an understory perennial herb *Trillium camschatcense*. As with many other understory perennials, *T. camschatcense* grows slowly with limited sunlight, taking a long period of years to become sexually mature. The prolonged life cycle can be categorized into five distinct life history stages: seed (SE), seedling (SD), one-leaf (1L), three-leaves (3L), and flowering (FL) (Figure 3.1, Ohara and Kawano (2005)). Although *T. camschatcense* is widely distributed in northeast Asia, especially in Hokkaido, northern Japan, its habitats have been destroyed and fragmented due to agricultural and

residential land development. One of the most striking habitat fragmentation events took place in the Tokachi region (Figure 3.2a). According to the records in the early time of the settlement, this region had been almost completely forested, and understory plants including *T. camschatcense* were extensively distributed. However, intensive land development which started in the 1880s resulted in many small forest fragments within an agricultural matrix.

Previous studies showed that genetic diversity of fragmented *T. camschatcense* populations was not solely explained by population size. Although bottleneck was slightly detected using allozyme marker, the correlation between genetic diversity and population size was statistically marginal (Tomimatsu and Ohara 2003b). On the other hand, life history seemed to play an important role in maintaining genetic diversity. Tsuzuki et al. (2022a) analyzed stage-wise genetic diversity in two (large and small) populations using neutral single nucleotide polymorphisms (SNPs), and found that genetic diversity was higher in more mature stages in the examined two populations. The results indicated that stasis during juvenile and flowering stages (i.e., slow growth and long lifespan), which are the life history features of *T. camschatcense*, might have resulted in the incremental storage of genetic diversity along the life cycle. Due to the small number of population-level replicates and the absence of data on demographic rates, it is left unclear whether the supposed contribution of life history to genetic diversity is general and can be theoretically supported under the anthropogenically altered population size and demographic rates (stage-specific survival probabilities and fecundities).

In this study, I carried out demographic census survey to estimate demographic rates and population size in fragmented *Trillium camschatcense* populations in the Tokachi region. Then, using the model developed in chapter 2, I calculated demographic genetic structure (i.e., the among-stage ratio of genetic diversity) and effective population size N_e (i.e., the rate of diversity loss per generation) to predict stage-wise genetic diversity and its temporal

dynamics under the observed life history and population size. In parallel, I conducted population genetic analysis by double digest restriction-site associated DNA sequencing (ddRAD-seq) to obtain observed demographic genetic structure and N_e. Lastly, I checked if the theoretical expectation matched with observed counterparts to evaluate the contribution of life history to genetic diversity could be theoretically supported.

3.2 Materials and Methods

3.2.1 Study species and study sites

Trillium camschatcense is an understory spring ephemeral that sprouts and flowers around April and May in temperate deciduous forests. After the development of the canopy layer in summer, there is much less sunlight available for understory plants, and *T. camschatcense* becomes dormant underground during the rest of the year. Every spring, aboveground stems and leaves are regenerated and at that time the life history stage of each individual is updated. The life history of *T. camschatcense* consists of progressive, static, and retrogressive transitions among the five stage classes (Figure 3.1). Seeds (SE) germinate after two years of dormancy and grow as seedlings (SD) in the second year. From the third year on, individuals continue to grow as one leaf (1L) and subsequently as three-leaves (3L). It is not until they store enough assimilated resources that they become flowering (FL). Flowering individuals are long-lived and remain reproductive for years, and sometimes retrogress to three-leaves due to resource exhaustion. Retrogression occasionally occurs from three-leaves to one-leaf as well (Ohara and Kawano 2005).

In 2018, I randomly established two to five 1 m $\times$ 1 m quadrats in 11 remnant forests in the Tokachi region, in total 39 quadrats (Figure 3.2b, Table 3.1). While *Trillium camschatcense* was distributed on the entire forest floor, co-occurring species such as *Sasa chartacea* and

Carex cespitosa were heterogeneously distributed (Tomimatsu et al. 2011), and each forest could be split into several compartments based on the difference in understory vegetation. The quadrats were placed in different compartments to cover the heterogeneity of the understory vegetation within a population. I counted the number of individuals of each life history stage except underground seed in the quadrats. As a result, stage structure, or the relative abundance among stage classes, varied among populations (Figure 3.2c). Because stage structure is dependent on the balance of growth, stasis, retrogression, and reproduction of each stage (Caswell 2001), the result indicated the presence of interpopulation life history variation tentatively. Therefore, I selected the 11 populations as study sites.

3.2.2 Field census

From 2019 to 2022, I carried out (1) individual monitoring (capture-mark-recapture) and (2) fecundity survey to estimate demographic rates and population size.

As for (1) individual monitoring, I marked all individuals in the quadrats and recorded their two-dimensional coordinates, using tags made with a steel wire and a numbering tape. Every year except 2020, in which field survey was cancelled due to the COVID-19 pandemic, I monitored the presence/absence of previously marked individuals as well as tagged newly emerged ones until 2022 (Figure 3.3a). The monitoring was carried out in mid May, when foliage leaves are fully expanded and the flowering phenology is at its peak. Seedling and one-leaf individuals sometimes formed clumps and densely aggregated (Figure 3.3b) and it was hard to tell apart which individual was which when the aboveground stems and leaves were once withered and then regenerated in the next spring. Therefore, I tagged each clump as a whole, and monitored its composition (the number of individuals in each stage). As a result, I obtained the time-series state of solitary individuals and clumps, as well as that of

the total number of individuals in each stage (i.e., stage structure).

As for (2) fecundity survey, I recorded the number of flowers and fruits of each flowering individual in the quadrats in the same years as the individual monitoring survey (i.e., from 2019 to 2022 except 2020). I collected matured fruits in 2021 and 2022 in July, and counted the number of seeds within them (Figure 3.3c, d). I did not collected fruits in 2019, because seeds produced in 2019 would be recruited as seedlings in 2021 and had to be left intact to avoid interfering with the individual monitoring survey.

3.2.3 Integrated population model

I constructed an integrated population model (IPM) to estimate the probabilities of growth, stasis, and retrogression of each stage class, fecundity of the flowering stage, and population size that were best fitted to the field census data in the Bayesian framework. The model was separately constructed for each population. IPM is an integration of multiple sub-models that correspond to different aspects of population dynamics (Schaub and Kéry 2021). In my case, four sub-models were constructed in correspondence to the four types of data: time-series state of solitary individuals, time-series composition of clumps, time series of stage structure, and the number of seeds per flowering individual (Figure 3.4). In IPM, because parameter estimation is based on the overall likelihood across the sub-models, estimation is more precise compared to fitting each sub-model to each data set independently (Schaub and Kéry 2021).

I used a multi-state capture-recapture model for the sub-model of the time-series state of solitary individuals. I distinguished the observed state $y_{q,i,t}$ and the latent true state $z_{q,i,t}$ for individual i in quadrat q in year t. The year 2019, which was the first year of the individual monitoring, was treated as $t = 1$. The observed state was either 1 (SD), 2 (1L), 3 (3L),

4 (FL), or 5 (absence). The fifth state included the possibilities of oversight, vegetative dormancy, death, or pre-germination. Because individuals were alive in the case of oversight and vegetative dormancy, the latent state of the two situations had to be categorized as either of the four life history stages, while the other two situations (death and pre-germnation) as distinct latent states. Therefore, I arranged six latent states: 1 (SD), 2 (1L), 3 (3L), 4 (FL), 5 (dead), and 6 (pre-germination). The transition from $z_{q,i,t}$ to $z_{q,i,t+1}$ was modeled by a categorical distribution with state-specific probabilities of growth, stasis, and retrogression among states (Equation 3.1). The observed state $y_{q,i,t+1}$ was modeled by a categorical distribution with the state-dependent detection probabilities (Equation 3.2).

$$z_{q,i,t+1} \sim Categorical(\boldsymbol{T}_{q,z_{q,i,t}}), \tag{3.1}$$

$$y_{q,i,t} \sim Categorical(\boldsymbol{P}_{q,z_{q,i,t}}), \tag{3.2}$$

where $\boldsymbol{T}_{q,z_{q,i,t}}$ and $\boldsymbol{P}_{q,z_{q,i,t}}$ were the $z_{q,i,t}$-th column of the matrix $\boldsymbol{T_q}$ and $\boldsymbol{P_q}$, respectively.

$$\boldsymbol{T_q} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & b_{q,t} \\ g_{1,q} & s_{2,q} & r_{3,q} & 0 & 0 & 0 \\ 0 & g_{2,q} & s_{3,q} & r_{4,q} & 0 & 0 \\ 0 & 0 & g_{3,q} & s_{4,q} & 0 & 0 \\ m_{1,q} & m_{2,q} & m_{3,q} & m_{4,q} & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 - b_{q,t} \end{pmatrix}, \tag{3.3}$$

$$\boldsymbol{P_q} = \begin{pmatrix} d_{1,q} & 0 & 0 & 0 & 0 & 0 \\ 0 & d_{2,q} & 0 & 0 & 0 & 0 \\ 0 & 0 & d_{3,q} & 0 & 0 & 0 \\ 0 & 0 & 0 & d_{4,q} & 0 & 0 \\ 1-d_{1,q} & 1-d_{2,q} & 1-d_{3,q} & 1-d_{4,q} & 1 & 1 \end{pmatrix}. \tag{3.4}$$

Here, $m_{i,q}$ was the mortality of the i-th state in quadrat q.

$$m_{1,q} = 1 - g_{1,q} \tag{3.5}$$

$$m_{2,q} = 1 - s_{2,q} - g_{2,q} \tag{3.6}$$

$$m_{3,q} = 1 - r_{3,q} - s_{3,q} - g_{3,q} \tag{3.7}$$

$$m_{4,q} = 1 - r_{4,q} - s_{4,q} \tag{3.8}$$

$b_{q,t}$ was the entry probability at time t in quadrat q, which was the probability that a solitary individual that had not germinated and therefore had not been recorded before year t became seedling in year $t+1$. As for the observation process, individuals in the latent state i ($1 \leq i \leq 4$) in quadrat q were observed as state i with the detection probability $d_{i,q}$, otherwise they were observed as absent (i.e., observation state 5) due to either oversight or vegetative dormancy. Because the fifth and the sixth latent states (dead and pre-germination) must have been always observed as absent, the fifth elements of the corresponding columns of the detection matrix $\boldsymbol{P_q}$ were set to 1 while other elements were zero.

With regard to the sub-model of the time-series composition of clumps, $v_{q,i,j,t}$ denoted the latent true number of individuals of stage j ($1 \leq j \leq 4$) in clump i of quadrat q in year t, while $u_{q,i,j,t}$ was the corresponding observed number. I modeled the latent process

as follows.

$$\boldsymbol{v'}_{q,i,j,t} \sim Multinomial(v_{q,i,j,t}, \boldsymbol{T}_{q,j}), \quad (1 \leq j \leq 4) \tag{3.9}$$

$$v_{q,i,j,t+1} = \sum_{k=1}^{4} (j\text{-th elements of } \boldsymbol{v'}_{q,i,k,t}), \quad (2 \leq j \leq 4) \tag{3.10}$$

where $\boldsymbol{v'}_{q,i,j,t}$ was the vector whose k-th element represented the number of individuals that moved from latent state j to k in clump i of quadrat q between year t and $t+1$. Because the seedling stage always consists of new recruitment, not of the survival transition from the other stages, I used negative binomial distribution to model the latent number of seedlings $v_{q,i,1,j,t}$.

$$u'_{q,i,1,t} \sim NB(u_{q,i,1,t} + 1, d_{q,1}), \tag{3.11}$$

$$v_{q,i,1,t} = u_{q,i,1,t} + u'_{q,i,1,t}, \tag{3.12}$$

where $u'_{q,i,1,t}$ was the number of undetected seedlings. The observation process of the other three stages was modeled using binomial distribution.

$$u_{q,i,j,t} \sim Binomial(v_{q,i,j,t}, d_{q,j}), \quad (2 \leq j \leq 4) \tag{3.13}$$

The sub-model of stage structure was constructed as almost the same as that of clumps.

$$\boldsymbol{x'}_{q,i,t} \sim Multinomial(x_{q,i,t}, \boldsymbol{T}_{q,i}), \quad (1 \leq i \leq 4) \tag{3.14}$$

$$x_{q,i,t+1} = \sum_{j=1}^{4} \left(i\text{-th elements of } \boldsymbol{x'}_{q,j,t}\right), \quad (2 \leq i \leq 4) \tag{3.15}$$

$$x_{q,1,t} = Poisson(f_q g_{0,q} x_{q,4,t-2}), \tag{3.16}$$

$$w_{q,i,t} \sim Binomial(x_{q,i,t}, d_{q,i}), \tag{3.17}$$

where $x_{q,i,t}$ was the latent true number of individuals of stage i in quadrat q in year t, while $w_{q,i,t}$ was the observed counterpart. Here, Stage 1 to 4 stood for SD, 1L, 3L, and FL, respectively. $\boldsymbol{x}'_{q,i,t}$ was the vector of length 6 and represented the transition from stage i to the six latent states. Here, the true number of seedlings (i.e., $x_{q,1,t}$) was modeled by Poisson distribution using fecundity (f_q), germination rate ($g_{0,q}$), and the number of FL in two years before ($x_{q,4,t-2}$, Equation (3.16)). When $t \leq 2$, $x_{q,4,t-2}$ was substituted with $x_{q,4,1}$ in Equation (3.16).

Sub-model for fecundity was composed of three models, each of which handled the number of flowers, fruits, and seeds per flowering individual. Firstly, because the number of flowers per individual ranged from one to three, I used a categorical distribution as follows.

$$n_{\mathrm{fl},q,i,t} \sim Categorical(\boldsymbol{l_q}). \tag{3.18}$$

$n_{\mathrm{fl},q,i,t}$ was the number of flowers of individual i in quadrat q in year t while $\boldsymbol{l_q}$ was the vector with three elements which corresponds to the probability of having one, two, and three flowers, respectively.

$n_{\mathrm{fr},q,i,t}$ and $n_{\mathrm{se},q,i,t}$, which were the number of fruits and seeds of a flowering individual i in quadrat q in year t, were modeled by binomial and Poisson distribution, respectively.

$$n_{\mathrm{fr},q,i,t} \sim Binomial(n_{\mathrm{fl},q,i,t}, f_{\mathrm{f},q}) \tag{3.19}$$

$$n_{\mathrm{se},q,i,t} \sim Poisson(n_{\mathrm{fr},q,i,t} f_{\mathrm{s},q}) \tag{3.20}$$

$f_{\mathrm{f},q}$ and $f_{\mathrm{s},q}$ were fruit set and the number of seeds per fruit in quadrat q, respectively.

I defined a vector $\boldsymbol{a} = {}^t(1, 2, 3)$ where t mean transpose. Then, the mean number of flowers per individual was obtained by $\langle \boldsymbol{l_q}, \boldsymbol{a} \rangle$, that is, the inner product between $\boldsymbol{l_q}$ and $\boldsymbol{a}$. The net fecundity per flowering individual in quadrat q, which was denoted by f_q, was calculated by the multiplication of $\langle \boldsymbol{l_q}, \boldsymbol{a} \rangle$, $f_{\mathrm{f},q}$ and $f_{\mathrm{s},q}$.

$$f_q = \langle \boldsymbol{l_q}, \boldsymbol{a} \rangle \times f_{\mathrm{f},q} f_{\mathrm{s},q}. \tag{3.21}$$

I calculated the population mean of each parameter by taking the average across quadrats. The population mean parameters were denoted by the same character as the quadrat-level counterparts, while the subscript q was removed. For example, the number of individuals in stage i in year t was denoted by $x_{i,t}$. The mean numbers of SD, 1L, 3L, and FL individuals per year were denoted by x_1, x_2, x_3,and x_4, respectively, and was calculated as follows.

$$x_i = \sum_{t=1}^{4} x_{i,t}/4. \qquad (1 \leq i \leq 4) \tag{3.22}$$

Treating seed as stage 0, the expected number of underground seeds x_0 was calculated by the product of the mean net fecundity and the mean number of FL individuals.

$$x_0 = f x_4. \tag{3.23}$$

The summation of x_i ($0 \leq$ i ≤ 4) was equal to the population density per 1 m^2, which was denoted by D.

$$D = \sum_{i=0}^{4} x_i \tag{3.24}$$

Then, x_i ($0 \leq i \leq 4$) and D were multiplied by the area of the study sites A (Table 3.1) to

obtain the total number of individuals in each stage ($N_i, 0 \leq i \leq 4$) and population size N.

$$N_i = Ax_i, \quad (0 \leq i \leq 4) \tag{3.25}$$

$$N = AD = A\sum_{i=0}^{4} x_i. \tag{3.26}$$

The IPM was implemented in JAGS 4.3.0 using an R package "jagsUI" (Kellner 2021) in R 4.1.2 (R Core Team 2021). The number of Markov Chain Monte Carlo (MCMC) iterations was set to 1,000,000, including 100,000 steps for burn-in. The number of adaptation steps, chains, and thinning were 10,000, 3, and 10, respectively.

3.2.4 Theoretical prediction on demographic genetic structure and effective population size

A total of 2,700,000 sets of parameter values (= (1,000,000 iterations - 100,000 burn-in steps) $\times$ 3 chains) obtained by MCMC were used to calculate the posterior distribution of matrix $\boldsymbol{M}$ of Equation (2.10). In calculating $\boldsymbol{M}$, seed (SE), seedling (SD), one-leaf (1L), three-leaves (3L), and flowering (FL) were regarded as stage 0, 1, 2, 3, and 4, respectively. The detailed equations for each element of $\boldsymbol{M}$ are shown in Appendix 2. Then, the dominant eigenvalue η and the leading right eigenvector $\boldsymbol{w}$ of matrix $\boldsymbol{M}$ were calculated. Here, fecundity (f) was adjusted so that population growth rate became equal to unity. To obtain the theoretical expectation of demographic genetic structure, the elements of $\boldsymbol{w}$ that correspond to $H_{ii,t}$ $(1 \leq i \leq 4)$ were log transformed. By taking the logarithm, the ratio among w_{ii} was transformed to the difference among $\log w_{ii}$, which was equal to the difference of the corresponding

$\log H_{ii,t}$ based on equation (2.16).

$$\begin{aligned}\log H_{ii,t} - \log H_{jj,t} &= \log \frac{H_{ii,t}}{H_{jj,t}} \\ &= \log \frac{w_{ii}}{w_{jj}} \\ &= \log w_{ii} - \log w_{jj}. \end{aligned} \tag{3.27}$$

Therefore, the logarithm of w_{ii} represents the relative height of $H_{ii,t}$ among SD, 1L, 3L, and FL stages. I subtracted the mean of $\log w_{ii}$ $(1 \leq i \leq 4)$ from all $\log w_{ii}$ so that their distributional center was adjusted to 0.

To quantify the change rate of genetic diversity over time, I obtained theoretical expectation of effective population size based on Equation (2.18), which is mentioned as demographic N_e (N_e[DE]) hereafter. Effective population size represents the degree of diversity loss per generation: the larger N_e is, the more likely genetic diversity is maintained.

3.2.5 DNA extraction and SNP call

To confirm the theoretical expectation obtained in section 3.2.4 empirically, I carried out genetic analysis using genome-wide single nucleotide polymorphisms (SNPs). In May 2021, I randomly collected leaves of 16 flowering (FL) individuals from the 11 populations for estimating genetic diversity and effective population size. Among the 11 populations, I chose seven core populations (population A-G), in which 16 individuals were additionally randomly sampled from each of seedling (SD), one-leaf (1L), and three-leaves (3L) stages to estimate demographic genetic structure. The sampled leaves were first stored in the refrigerator and were subsequently dried using silica gel. Genome DNA was extracted from the dried leaves following the modified CTAB method (Murray and Thompson 1980).

The extracted DNA was used for double digest restriction-site associated DNA sequencing (ddRAD-seq) (Peterson et al. 2012, 2014). ddRAD-seq is one of the reduced-representation DNA sequencing techniques and detects genome-wide SNPs from DNA fragments cut out by two restriction enzymes, for which I used *PstI* and *MspI*. Library preparation and sequencing was performed in Kazusa DNA Research Institute using a DNBSEQ sequencer (MGI Tech Co., Ltd., Shenzhen, China, 150 bases, paired-end).

Primer regions, adapter sequences and low-quality reads were removed from the raw sequence data using Trimmomatic (Bolger et al. 2014). The remaining reads were used for *de novo* assembly. SNPs were called using Stacks ver. 2.6 (Rochette et al. 2019). The parameter values used in alignment and SNP identification (from ustacks to gstacks) were $m = 3$, $M = 2$, $N = 4$, and $n = 2$, which are the default of Stacks *de novo* assembly. SNPs that were successfully typed in at least 80% of all samples ($r = 0.8$, $p = 1$) were filtered with two additional criteria: minimum allele frequency must exceed 0.05 (min-maf=0.05) and observed heterozygosity must below or equal to 0.6 (max-obs-het=0.6). Among the extracted SNPs, one SNP was randomly chosen per locus to avoid strong linkage (write-random-snp). To delete SNPs that were potentially under selection, the output vcf files were reformatted to bed files by PLINK v2 (Chang et al. 2015) and were examined by PCAdapt package in R (Luu et al. 2017). SNPs whose adjusted P values were not below 0.1 were regarded as neutral and were used for the following analyses. The above procedures of SNP call were run separately for the eight sets of samples: (1) FL stage of all the 11 populations, and (2-7) the four stages (SD, 1L, 3L, and FL) of each of the 7 core populations (Table 3.2).

3.2.6 Population genetic analysis

Using SNPs called from FL individuals of the 11 populations (sample set 1 in Table 3.2), I carried out principle component analysis (PCA) to grasp the spatial genetic structure over the study region. Then, expected heterozygosity (H_e), observed heterozygosity (H_o), and allelic richness (A_r) were estimated and their correlation with population size (N) and density per 1 m^2 (D) were examined. Besides, effective population size were estimated for each population using NeEstimator2 (Do et al. 2014) with the linkage disequilibrium method. The estimated value was denoted by N_e[LD] hereafter.

Stage-wise expected heterozygosity ($H_{ii,t}$) was calculated for each of the 7 core populations, using SNPs called from the sample set 2 to 7, respectively. Because $H_{ii,t}$ is obtained for only year 2021 and there is no need to note t, I omit t from $H_{ii,t}$ hereafter for tidiness. For each SNP, I calculated H_{ii} as follows:

$$H_{ii} = 2p_i(1 - p_i), \tag{3.28}$$

where p_i is the frequency of one of the two alleles in stage i. I calculated the mean of H_{ii} over all SNPs, which was used as the observed H_{ii} in the following analyses. To test if the observed H_{ii} were significantly different among stages, I implemented permutation test with 1,000 iterations with Holm correction. Then, the observed H_{ii} were log transformed and centered as previously done for theoretically obtained w_{ii}.

3.3 Results

3.3.1 Interpopulation demographic variations

On average, 240.7 solitary individuals and 45.1 clumps were found per population during the census survey. A total of 136.6 flowering individuals were observed during the entire study period and were used for the sub-model of fecundity (Table 3.1).

Demographic rates (stage-specific probabilities of growth, stasis, and retrogression and fecundity) varied among the 11 populations (Figure 3.5, 3.6), yet sharing some common characteristics. The probabilities of growth (g_0, g_1, g_2, and g_3) and retrogression (r_3 and r_4) were generally lower than those of stasis (s_2, s_3, and s_4). Especially, g_0 (germination) and g_2 (growth from 1L to 3L) was consistently low in all populations, ranging from 0.02 to 0.21 and from 0.02 to 0.31, respectively. Fecundity (f) ranged from 32.19 (population J) to 103.06 (population H). Overall, posterior distributions of population J were broad for all parameters. This is probably due to the small sample size compared to the other populations (Table 3.1). There were no apparent correlations between demographic rates and habitat area (Figure 3.7). Population size (N) was above one hundred thousand in all populations, with the smallest in population J ($10^{5.62}$) and the largest in population I ($10^{8.37}$).

3.3.2 Genetic diversity and differentiation among populations

Using the 176 flowering individuals of the 11 populations, 4,823 SNPs were called by Stacks, from which 717 SNPs were removed by PCAdapt as potentially being under selection (Table 3.2). Based on the remaining 4,176 SNPs, the 11 populations were split into two clusters by PCA, which strongly reflected the geographical locations: three southern populations (population H, J, and K) and eight northern populations (population A-G, I, Figure 3.8).

Expected heterozygosity (H_e), observed heterozygosity (H_o) and allelic richness (A_r) did not have a significant positive correlation with population size (Figure 3.9a, c, e), indicating that the current population size is not the primary determinant of genetic diversity of the whole population. On the other hand, population density was significantly correlated with the three proxies of genetic diversity (Figure 3.9b, d, f). It should be noted that genetic diversity of the three southern populations (population H, J, and K) were generally lower than the northern cluster of populations (Figure 3.9), indicating the possibility of geographic/phylogenetic constraint on genetic diversity.

3.3.3 Theoretical and observed demographic genetic structure and effective population size

The centered logarithms of the theoretically obtained $H_{ii}(1 \leq i \leq 4)$ were distributed around zero in all populations, indicating that H_{ii} were almost equivalent to one another (Figure 3.10).

The *de novo* assembly of the 64 samples (= 16 individuals per stage $\times$ 4 stages) of each core population detected 6,455 SNPs on average. Subsequently, 480.4 SNPs were removed by PCAdapt as being potentially non-neutral, and the remaining 5974.6 SNPs were used for estimating H_{ii} ($1 \leq i \leq 4$). Unlike the theoretical expectation, the observed H_{ii} varied significantly among stages. Although no statistically significant difference was detected among stages in population A and B (Figure 3.7a, b), H_{33} were consistently significantly higher than H_{11} in the other five populations, indicating that the three-leaves stage was genetically more diverse than the seedling stage. H_{22} (one-leaf) was also higher than H_{11} (seedling) in population C, D, and G. There were always no significant difference between flowering and seedling in all the core populations.

Theoretical expectation of effective population size N_e[DE] was higher than observation N_e[LD] by an order of approximately 2 to 4 (Figure 3.11). There was no significant correlation between the two estimated values (correlation coefficient = 0.57, p = 0.07, Figure 3.11).

3.4 Discussion

3.4.1 Demography of the fragmented *T. camschatcense* populations

Demographic rates and population size N were different among populations (Figure 3.6). Overall, there were no apparent relationships between habitat area and demographic rates (Figure 3.7). Therefore, the life history variation among populations cannot be attributed to the difference in habitat area and the resultant edge effects, which is not in line with the previous study showing the signs of edge effects in fragmented *T. camschatcense* populations (Tomimatsu and Ohara 2004). It might be that the environmental conditions of remnant forests have changed from 2000s, during which the previous studies were carried out (Tomimatsu and Ohara 2002, 2004, 2010). For example, during the field work, I noticed that there were recently-formed gaps in forest crown in most populations due to fallen trees, regardless of the habitat area. The formation of crown gaps would lead to more intense sunlight and drier understory. Besides, agricultural land development is still ongoing in this region, and the construction of underground drainage is taking place around the study sites, which would make the remnant forests drier. The dry forests are more likely to be invaded by dwarf bamboo *Sasa chartacea*, which is a competitively strong species and decreases the abundance of understory herbaceous plants (Tomimatsu et al. 2011). Moreover, I also noticed that disturbance such as mowing by humans and herbivory by deer took place, which might also affect the demographic performance of *Trillium camschatcense*. Assessing these recent abiotic and biotic environmental changes and their effects on demographic rates would be

necessary to understand how the life history of fragmented *T. camschatcense* is currently determined.

Although the duration of the field census survey was relatively shorter than those of the preceding studies (4 years in this study, while 12 years in Ohara et al. (2006) and 7 years in Tomimatsu and Ohara (2010)), the results of the present study satisfies the basic life history characteristics of *T. camschatcense*. In *T. camschatcense*, most individuals are likely to die at young and even if they survive, they spend more than 10 years in pre-reproductive juvenile stages in general (Ohara et al. 2006). The overall low probabilities of growth observed in this study (g_0, 0.02-0.21; g_1, 0.04-0.68; g_2, 0.02-0.31; g_3, 0.05-0.40, Figures 3.5 and 3.6) correspond to the high mortality and the slow progressive growth of *T. camschatcense*. Besides, flowering individuals of *T. camschatcense* are generally long-lived and survive for many years, reaching 40-50 years old (Ohara et al. 2006). The high survival rate of FL individuals (the sum of s_4 and r_4, 0.84-0.98, 3.5) is in line with the long life expectancy of flowering individuals.

It should be noted that previous studies on the demography of *Trillium camschatcense* did not estimate germination rate, although estimated in other *Trillium* species (Knight et al. 2009). In this study, I succeeded in estimating g_0 by adopting the integrated population model (IPM). The strength of IPM lies in incorporating and estimating unobserved variables in the model (Schaub and Kéry 2021). It was pointed out that the underground seeds should generally be taken into consideration for successful assessment of plant population dynamics (Arroyo-Cosultchi et al. 2022). While previous studies highlighted reduced seed production as a limiting factor of new recruitment in fragmented *T. camschatcense* populations (Tomimatsu and Ohara 2002), the present study showed that the germination rate was consistently low (ranging from 0.02 to 0.21), indicating that not only seed production but also germination

are limiting newborn seedlings.

3.4.2 Mismatch between theoretical and observed demographic genetic structure and effective population size

Demographic genetic structure was different between the observed data and the theoretical expectation. In all the seven core populations, the centered logarithms of theoretically obtained H_{ii} were almost equivalent to one another and were distributed close to zero. In chapter 2, I showed that inter-stage difference in H_{ii} was erased under large population size ($N = 1,000$) in the two and in the three stage model. Considering that the estimated population size N were far above one thousand ($10^{5.62} \leq N \leq 10^{8.37}$, Figures 3.5 and 3.6), the large N should have been responsible for the insignificant inter-stage difference.

Contrary to the theoretical expectation, observed demographic genetic structure showed the signs of small population size (N). The observed H_{ii} were uneven and were significantly different between some pairs of stages (Figure 3.10). Considering that inter-stage difference in H_{ii} is amplified under small N due to strong genetic drift, as shown in the theoretical model analysis in Chapter 2, the results suggests that population size was small enough for demographic genetic structure to be uneven in the seven populations.

The discrepancy between the theoretical expectation and the observation was also seen in effective population size (N_e). The estimates based on demographic data N_e[DE] were larger than those based on linkage disequilibrium of flowering individuals N_e[LD]. Again, the large N used in obtaining the theoretical expectation should be responsible for the difference between the two estimates, because large N leads to large N_e, as shown in the theoretical model analysis (Figure 2.9). This result is in line with the previous studies which showed that demographic N_e was relatively larger than the genetic counterpart (Rowe and Beebee

2004, Schmeller and Merilä 2007).

One possible explanation for the discrepancy between the theoretical expectation and the observed data was that not all individuals in the same forest effectively serve as one "population" in terms of genes. As for the theoretical prediction, I used the total number of individuals in the study site as population size, which was estimated by the extrapolation of the fine-scale population density (m^{-2}). In other words, the theoretical expectation implicitly assumed that individuals survive and reproduce randomly across the whole population. However, because plants are sessile and the dispersal of pollen and seed is often spatially limited, not all flowering individuals randomly mate with one another and only nearby individuals effectively serve as population, resulting in spatial aggregation of similar genotypes (fine-scale genetic structure) (Hardy et al. 2006, Vekemans and Hardy 2004). A previous study showed the presence of fine-scale genetic structure in *T. camschatcense* populations in the Tokachi region (Yamagishi et al. 2007). Therefore, population might not be a one unit but could be spatially sub-structured and the spatial range that effectively serves as an ideal population might be far limited than the whole population, resulting in the symptoms of small population size with regard to demographic genetic structure and effective population size. Although the possibility of confounding phylogenetic constraint should be carefully considered, the results that the population-level genetic diversity was not correlated with population size but with population density per 1 m^2 (Figure 3.9) also suggests that the amount of nearby individuals, not the whole individuals, is relevant to genetic diversity.

3.4.3 Life history stage as a genetic reservoir

Contrary to the theoretical expectation, observed demographic genetic structure was not flat, indicating that some life history stages serve as reservoirs of genetic diversity compared

to the other stages. Especially, three-leaves (3L) stage showed relatively higher genetic diversity than the others (Figure 3.10), indicating genetic variation was likely to be stored in 3L stage. Previous studies on demographic genetic structure of the fragmented *Trillium camschatcense* populations in the Tokachi region showed that not only 3L but also FL had comparably high genetic diversity due to the stasis of individuals (Tsuzuki et al. 2022a), which is not in line with the present results. The inconsistency might be explained by the presence of a fine-scale spatial genetic structure within a population, as with the inconsistency with the theoretical prediction discussed in the previous section. While the samples were randomly collected from the whole population in this study, Tsuzuki et al. (2022a) collected the samples from a fine-scale 5 m $\times$ 5 m quadrat within a population. Because the fine-scale genetic structure would decay towards mature stages due to random thinning of individuals (Berens et al. 2014, Chung et al. 2003), genetic composition could be more spatially divergent in more juvenile stages. Therefore, 3L might be more spatially heterogeneous than FL, resulting in comparable genetic diversity at the fine scale (5 m $\times$ 5 m) but higher diversity at the population level.

To sum up, the fact that 3L stage possessed high genetic diversity in spite of the theoretical prediction suggests that the life history trajectory interacts with spatial heterogeneity under small "genetically effective" population size and maintains genetic diversity at 3L stage at the population level in the fragmented *T. camschatcense* populations. Incorporating the genetic effectiveness and the within-population spatial genetic heterogeneity into the theoretical model would be necessary to predict the life history stage that serves as the genetic reservoir more realistically.

3.4.4 Conclusion and future direction

In accordance with the previous studies that evaluated genetic diversity of fragmented *Trillium camschatcense* populations (Tomimatsu and Ohara 2003b, Tsuzuki et al. 2022a), empirical genetic analyses showed that population size was not significantly related to genetic diversity (Figure 3.9) while genetic diversity was accumulated at particular stages over the course of life history (Figure 3.10). The discrepancy with the theoretical model might be attributed to the small number of individuals that effectively serve as population in terms of genes. The dispersal limitation of pollens and seeds, which are candidate factors that downsize the ”effective population,” are known to differ in their degrees among species depending on the types of vectors and the demographic trajectories from birth to death (Epperson 2000, Vekemans and Hardy 2004), which would interfere with the generality of the present results. How many individuals are actually form the ”effective” population and the candidate factors that control the effective size should be further studied in the future.

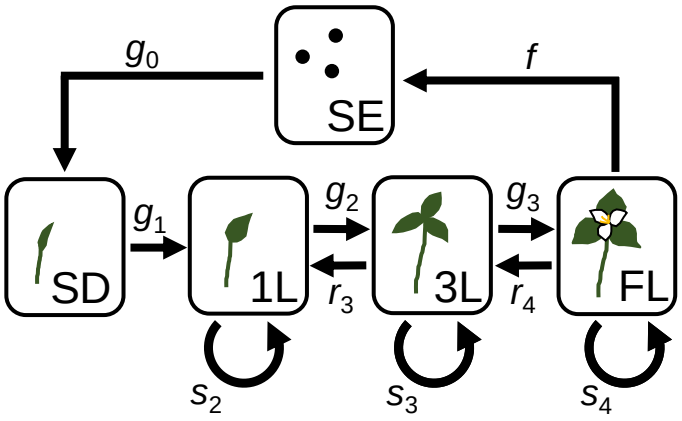


Figure 3.1: Life history of *Trillium camschatcense*. Arrows represent the possible annual trajectories of individuals. Demographic rates are denoted by the alphabets with numerical subscripts. g_0, g_1, g_2, and g_3 represent growth, s_2, s_3, and s_4 stasis, r_3 and r_4 retrogression, and f reproduction. SE, seed; SD, seedling; 1L, one-leaf; 3L, three-leaves; FL, flowering

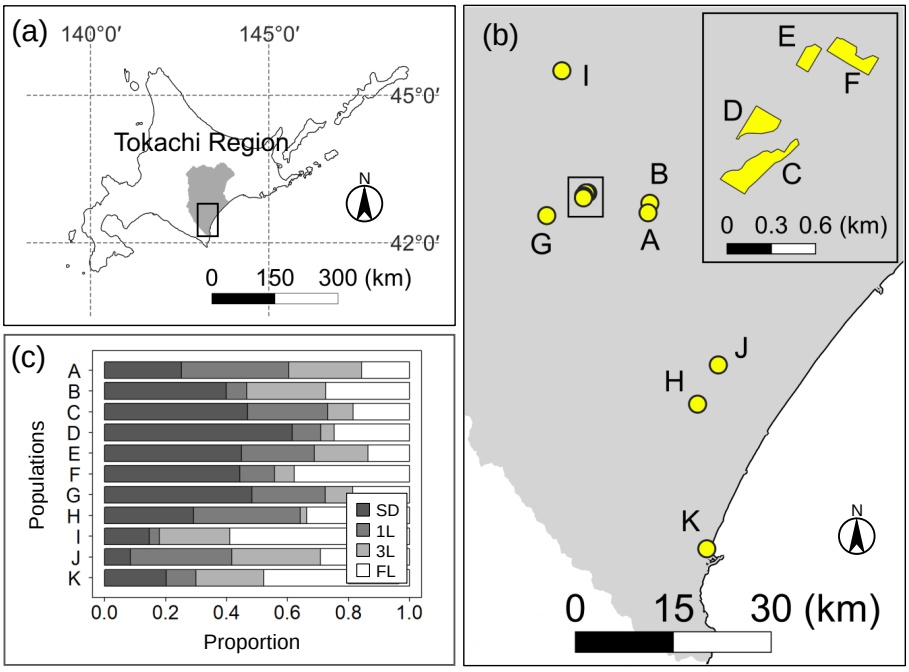


Figure 3.2: (a) Geographical location of the Tokachi region, which is shown in gray. (b) The enlarged view of the rectangle in (a), in which the 11 study sites are distributed. population C-F are shown in the top-right window, which is the enlargement of the rectangle that is located between population A and G. (c) Stage structure of the 11 populations in 2018

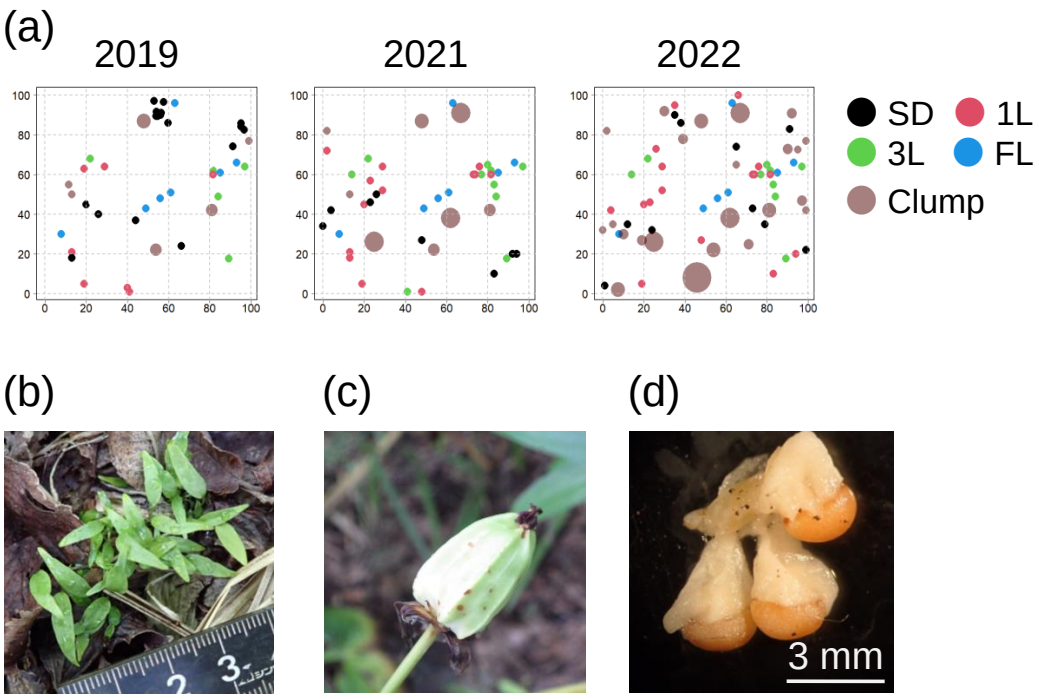


Figure 3.3: (a) The spatial distribution of individuals recorded during the field census in one quadrat established in population B. (b) Clump of seedlings. (c) Fruit. (d) Mature seeds with elaiosomes, which are to be dispersed by ants

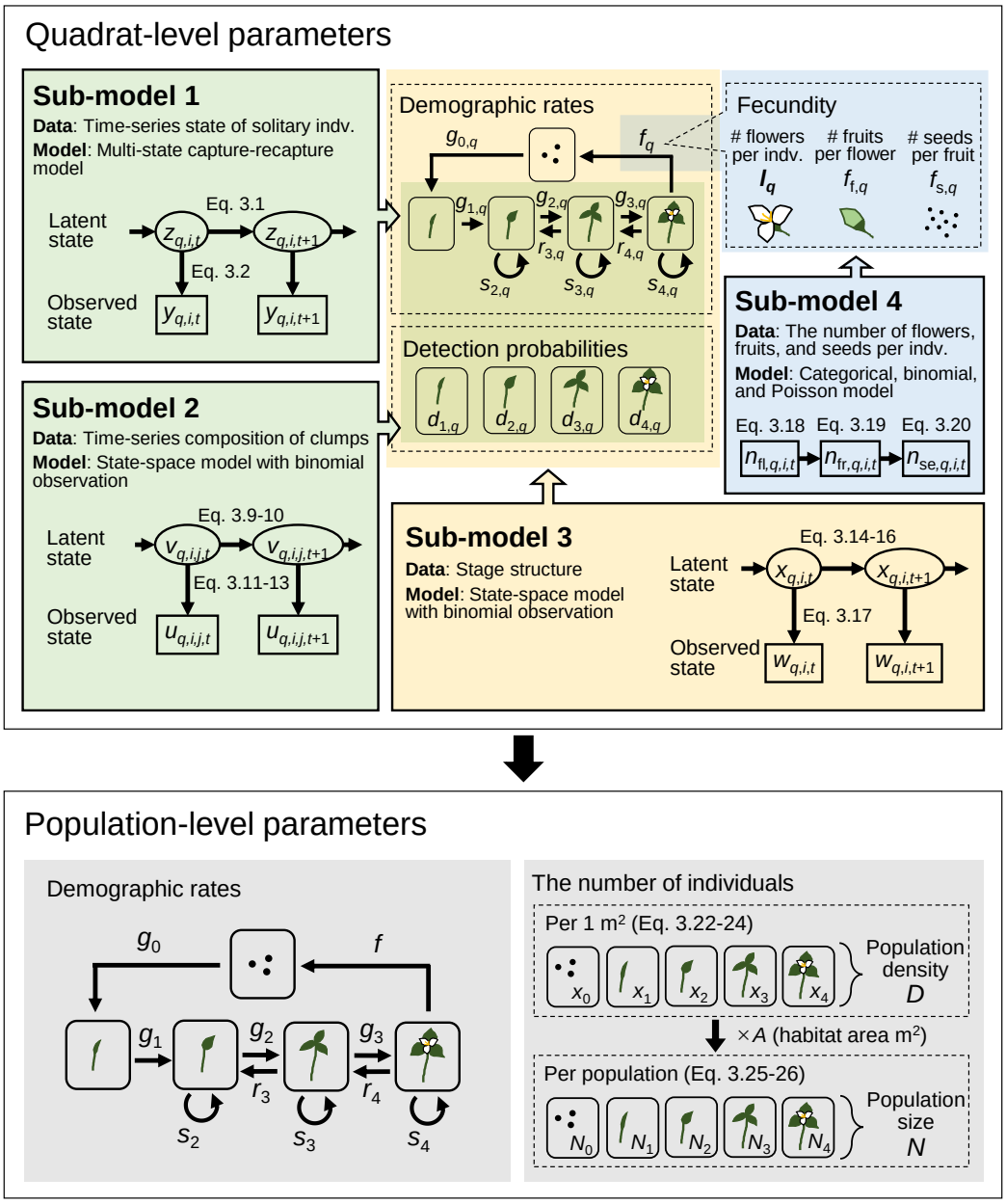


Figure 3.4: Graphical overview of the integrated population model (IPM). At the quadrat level, each sub-model was used to estimate parameters shadowed with the corresponding color. In each sub-model, latent and observed variables are shown in circle and rectangle, respectively

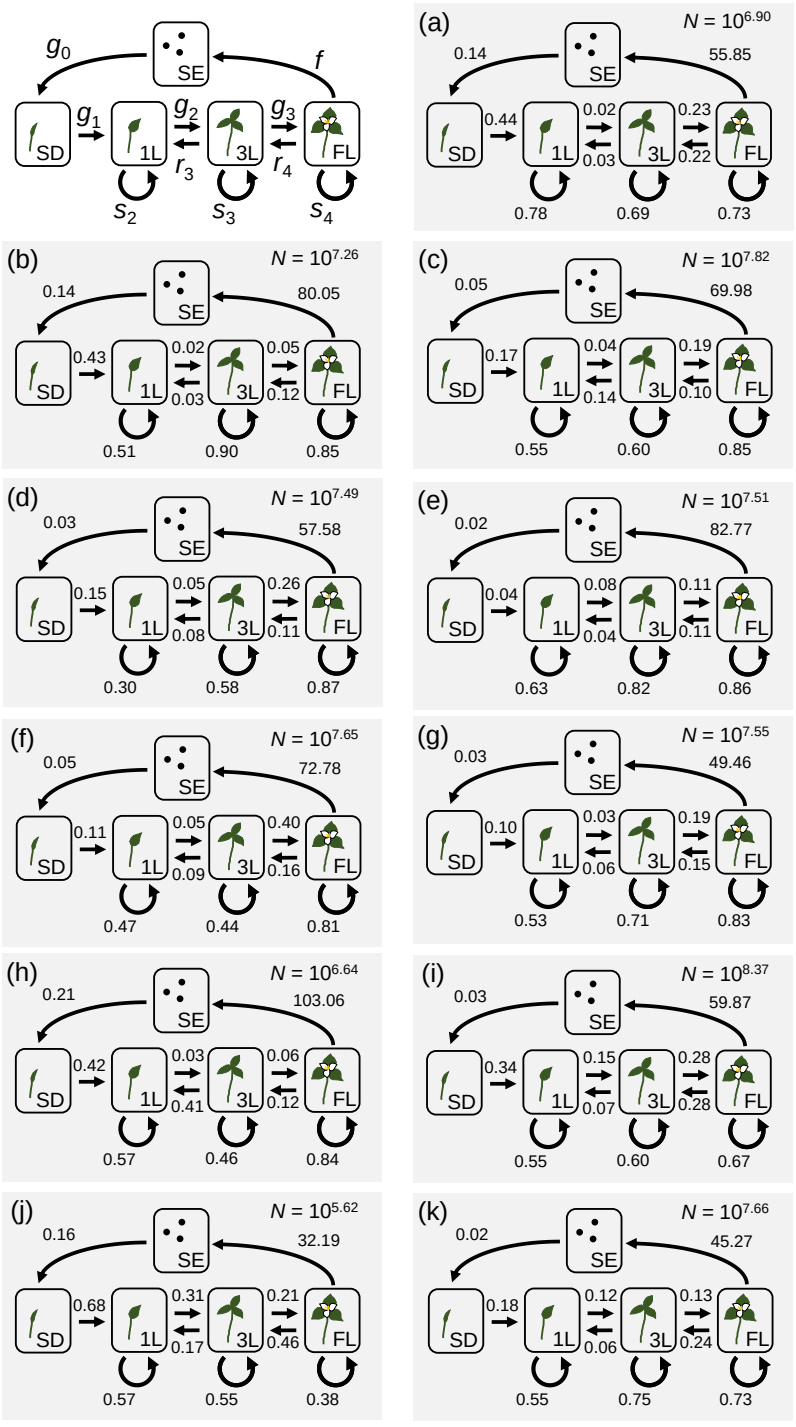


Figure 3.5: The mean of posterior distributions of demographic rates and population size. The top-left alphabet in each panel (a-k) corresponds to population A-K, respectively

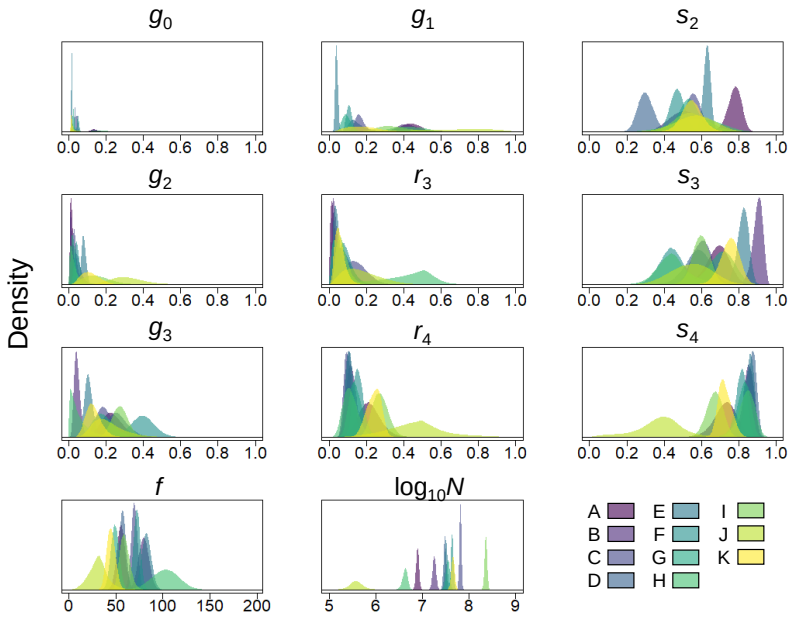


Figure 3.6: The posterior distributions of demographic rates and population size. The x axis represents the parameter values while the y axis represents the probabilistic density. Color denotes population

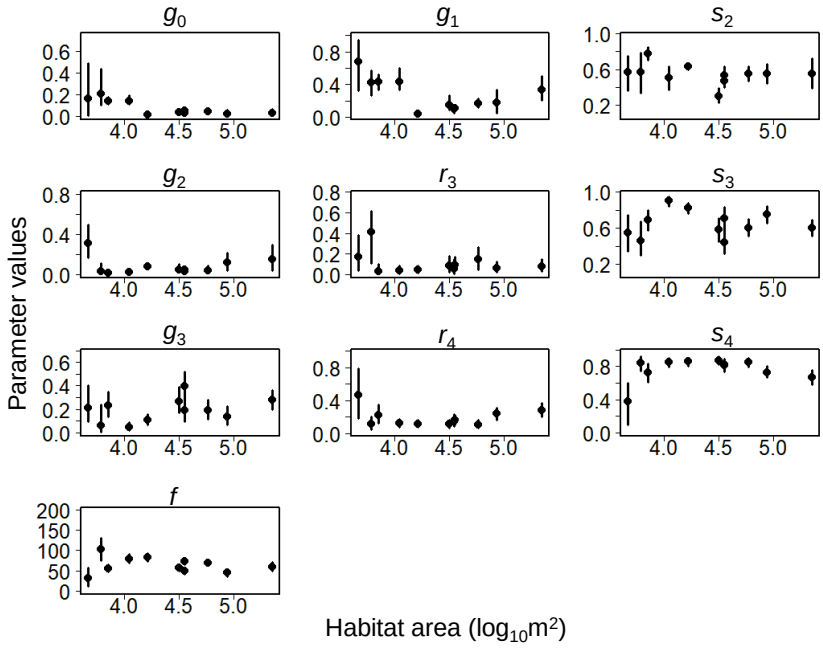


Figure 3.7: The relationship between habitat area and demographic rates. Demographic rates are shown with mean and 95% Bayesian confidence interval

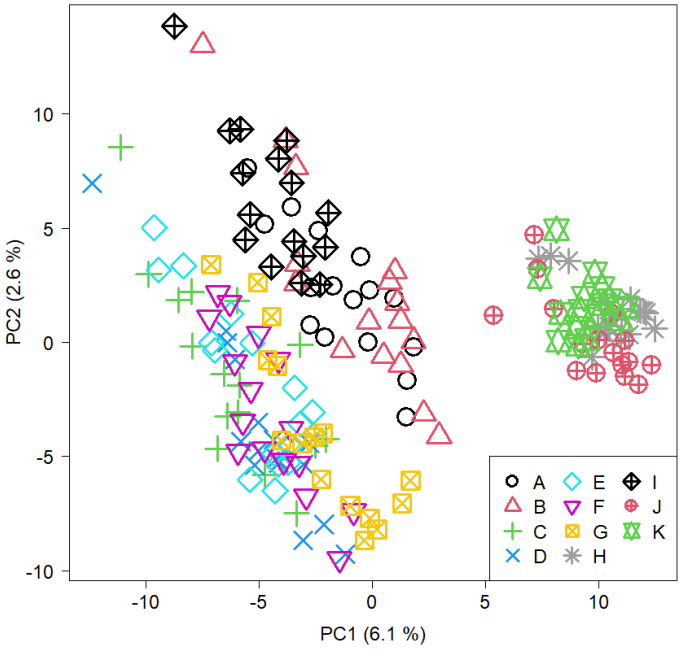


Figure 3.8: The results of principal component analysis of 176 samples (= 16 flowering individuals $\times$ 11 populations)

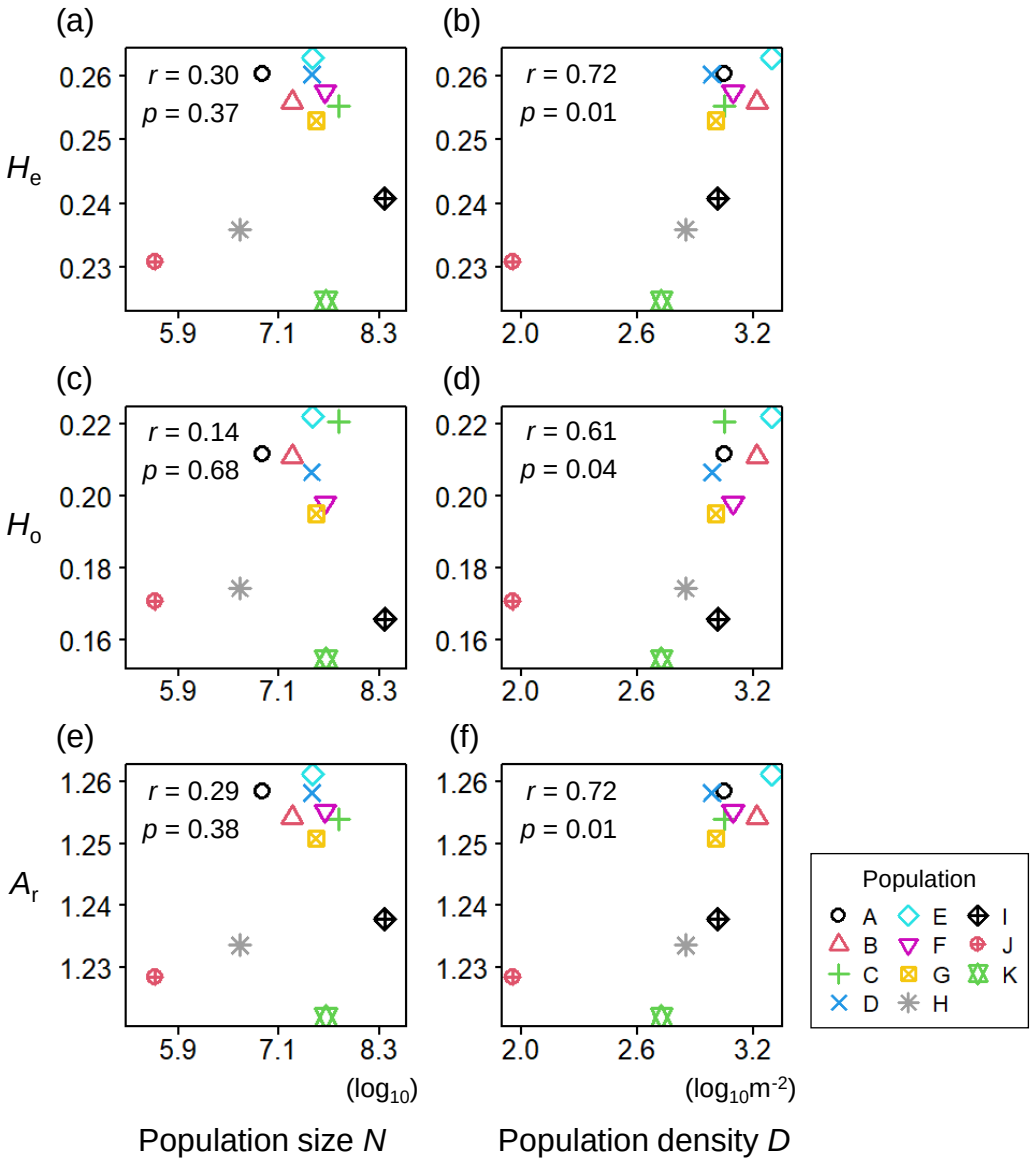


Figure 3.9: The relationships of genetic diversity (expected heterozygosity H_e, observed heterozygosity H_o, allelic richness A_r) with population size N (a, c, e) and with population density D (b, d, f). Genetic diversity was estimated from the 176 FL samples of the 11 populations. Population size N and density D were the mean of the posterior distributions estimated by the integrated population model. Pearson correlation coefficient and its P value are shown at the top-left corner of each panel

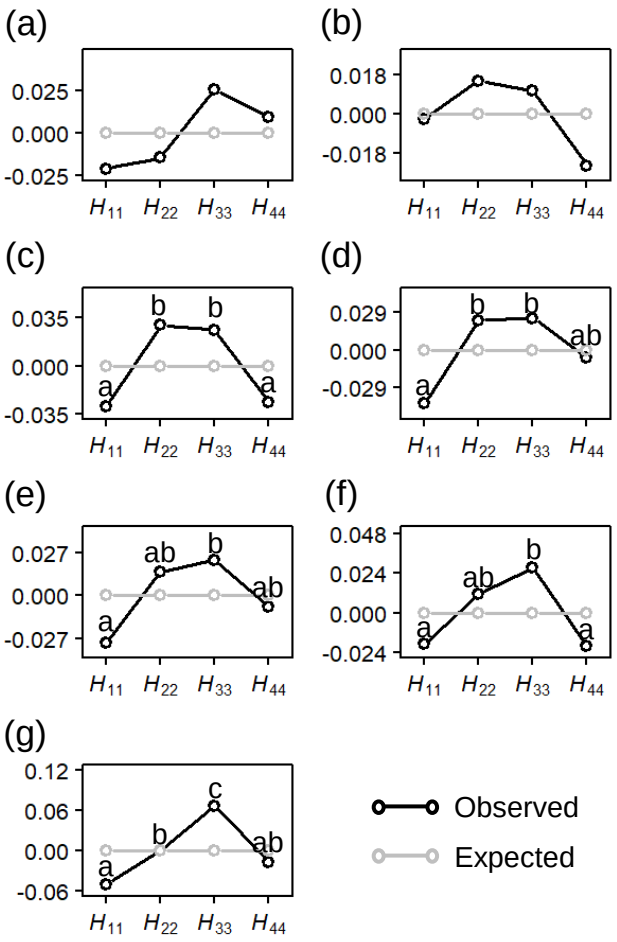


Figure 3.10: Demographic genetic structure of the seven core populations (Population A-G). The theoretical expectation and the observed data are shown in gray and black, respectively. H_{ii} ($i = 1$: SD; $i = 2$: 1L; $i = 3$: 3L; $i = 4$: FL) was log transformed and centered. The alphabet in the parentage corresponds to the population. As for the observed data, significant difference was detected by permutation test for pairs marked with different alphabets (significance level = 0.05)

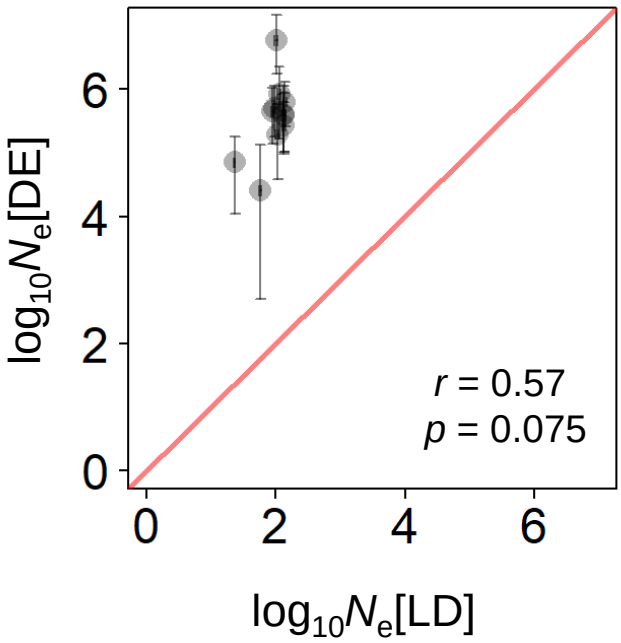


Figure 3.11: The comparison of effective population size N_e between estimation based on the linkage disequilibrium of the flowering stages (N_e[LD], x-axis) and the theoretical expectation obtained by using the demographic rates and population size estimated from IPM (N_e[DE] y-axis). The red line represents N_e[LD] $= N_e$[DE]. The Pearson correlation coefficient and its P value, which are shown at the bottom-right corner, indicate that the two estimated values were not significantly correlated

Table 3.1: Geographic location and the basic information on census survey for each study site. The number of solitary individuals, clumps, and FL individuals are the sum over quadrats

Population	Longitude (°E)	Latitude (°N)	Area ($\log_{10}$m^2)	No. of quadrat	No. of solitary indv.	No. of clumps	No. of FL indv.
A	143.215	42.774	3.85	2	222	41	73
B	143.218	42.787	4.05	3	359	76	136
C	143.095	42.795	4.77	5	515	89	212
D	143.096	42.797	4.50	4	359	30	183
E	143.100	42.801	4.22	3	268	76	175
F	143.103	42.801	4.55	5	347	81	207
G	143.027	42.770	4.55	2	129	27	107
H	143.307	42.513	3.79	2	44	24	33
I	143.055	42.968	5.35	5	238	25	230
J	143.345	42.567	3.67	2	32	2	13
K	143.324	42.315	4.94	5	135	25	134
Average			4.67	3.5	240.7	45.1	136.6

Table 3.2: The number of SNPs obtained through Stacks and PCAdapt. The first row corresponds the result of the *de novo* assembly of 176 flowering individuals from the all populations (16 indv. $\times$ 11 populations). The second to eighth rows correspond to the assembly of 64 individuals (16 indv. $\times$ 4 stages) in the seven core populations (population A-G). The last row is the average of the seven core populations.

Sample set	Population	Output of Stacks	Removed by PCAdapt (potentially under selection)	Remaining (neutral)
1	A-K (only FL stage)	4,823	717	4,176
2	A	6,834	502	6,332
3	B	6,590	402	6,188
4	C	6,905	570	6,335
5	D	6,490	525	5,965
6	E	6,986	460	6,526
7	F	5,554	419	5,135
8	G	5,826	485	5,341
Average of 2-8		6,455	480.4	5974.6

Chapter 4

Genetically effective life history that represents the flow of genes along the life cycle

Abstract

How individual survive and reproduce along the life cycle can be parameterized by demographic rates, which are survival probabilities and fecundities of each developmental stage. Demographic rates determine not only the life history trajectory from birth to death but also the temporal changes in population size and genetic composition of a population. Therefore, estimating demographic rates is important to understand ecological and evolutionary dynamics of a population. Previous studies usually carried out field demographic census (capture-mark-recapture survey) to estimate demographic rates, which requires long time and large number of replicates to obtain reliable results. In this study, I propose a new method that makes the most use of genetic information to estimate demographic rates of seven populations of the perennial herb *Trillium camschatcense* in the Tokachi region, Hokkaido, northern Japan. Using double digest restriction site associated DNA sequencing (ddRAD-seq), I quantified genetic diversity of each life history stage and pairwise genetic differentiation among stages, from which I inversely estimated stage-specific demographic rates and population size based on the theoretical model developed in Chapter 2. Besides, I carried out field demographic census to estimate the same parameters by the conventional methods. As a result, the estimated values were different between the two methods, which could potentially attributed to the difference in the spatio-temporal scale each method can handle. While the conventional census well describes contemporary fine-scale demography, the results of the new method reflects life history of the recent past that formed the present genetic data at the whole population level. Although previous methods based on field census have been poor at describing the observed population genetic data, the new method might reflect genetically effective life history and is suitable to explain and predict genetic dynamics of a population.

4.1 Introduction

Demography, or life history, is the lifetime trajectory of individual from birth to death. In many species, life history can be divided into multiple stage classes. In perennial plants, for

example, newly germinated individuals belong to the seedling stage at first and subsequently to the juvenile stages for multiple years. After vegetative growth, they become reproductive and join the flowering stage. How individuals survive and reproduce along the life history can be described by stage-specific demographic rates, that is, probabilities of growth, stasis, and retrogression and fecundity of each stage class (Caswell 2001, Lefkovitch 1965).

Demographic rates are different among species and populations, which raises variation in life history traits, such as survivorship curve type, age at sexual maturity, and longevity (Capdevila et al. 2020, Caswell 2001, Salguero-Gómez et al. 2016). In addition to these life history traits, population growth rate is subjected to demographic rates: changes in survivorship and fecundity of each stage class will impact the population dynamics of the whole population (Caswell 2001). Moreover, genetic dynamics also depends on demographic rates. Effective population size, which determines the strength of genetic drift per generation time, is a function of demographic rates (Felsenstein 1971, Hill 1972, Waples et al. 2011). The probability of fixation of adaptive and neutral alleles are also subjected to demographic rates (Li et al. 2016). Therefore, stage-specific demographic rates are the basic parameter to understand ecological and evolutionary dynamics of stage-structured populations (Metcalf and Pavard 2007, Silvertown 1987).

Many previous studies estimated demographic rates by long-term field census survey (Kéry and Schaub 2011, Schaub and Kéry 2021), in which individuals are marked one by one and their fate is tracked by capture-mark-recapture with the fixed time interval. Based on the relative occurrence of progressive, static, and retrogressive growth between the two consecutive observation time steps, the probabilities of growth, stasis, and retrogression are estimated. Fecundity is estimated by the fraction of newborns per reproductive individuals. Because survival and reproduction are stochastic processes, it is necessary to obtain enough

replicates over space and time to improve the accuracy of parameter estimation (Doak et al. 2005).

Here, I propose a new method that makes the most use of stage-wise genetic data to estimate demographic rates while easing the demand for survey duration. Assuming that life history stage is a sub-population, growth, retrogression, and reproduction can be regarded as gene flow among sub-populations, which erases the genetic differentiation among stages while increases genetic diversity per stage. Stasis would contribute to genetic diversity of each stage class, by maintaining existing variation in the same stage over time. In other words, genetic diversity and genetic differentiation at the stage-level at a one time point should have information on the movement of genes in the life cycle, and can be potentially used for estimating the probabilities of growth, stasis, and retrogression as well as fecundity.

A recent theoretical model (Tsuzuki et al. 2022b), which was developed in Chapter 2 of my dissertation, revealed mathematical relationships between demographic rates and the stage-level genetic diversity and differentiation. In the model, genetic diversity and differentiation is quantified by the metric $H_{ij,t}$, which is the probability that two genes randomly sampled from stage i and j with replacement at time t are non-identical-by-descent. $H_{ii,t}$ represents expected heterozygosity of stage i, while $H_{ij,t}$ ($i \neq j$) reflects the genetic differentiation between stage i and j. The array of $H_{ij,t}$ for all i and j is asymptotically proportional to the leading eigenvector of matrix $\boldsymbol{M}$, which is the coefficient matrix of the difference equations of $H_{ij,t}$ and is parameterized by the stage-specific demographic rates and population size (Equation 2.10). Using the model, we can explore the parameter space to yield almost the same ratio of $H_{ij,t}$ as the empirically observed data.

In this study, based on the theoretical model developed in Chapter 2, I inversely estimated demographic rates and population size from observed $H_{ij,t}$ in seven populations of a perennial

herb *Trillium camschatcense* using approximate Bayesian computation. At the same time, I carried out a field census survey for four years to estimate the same parameters by the conventional field census method. Then, I compared the results between the two methods to discuss the applicability of the newly developed genetic-based estimation approach.

4.2 Materials and Methods

4.2.1 Study species and study sites

Trillium camschatcense is an understory polycarpic perennial herb. The life history of *T. camschatcense* consists of five stages: seed (SE), seedling (SD), one-leaf (1L), three-leaves (3L), and flowering (FL) (Ohara and Kawano 2005). Individuals annually move among the five stages through nine possible transition paths including growth, stasis, and retrogression, as described in Chapter 3 (Figure 3.1). There is also a reproduction path from FL to SE. The study was conducted in population A-G in the Tokachi region (Figure 2.2b).

4.2.2 Stage-wise genetic diversity and differentiation

As explained in Chapter 3, I randomly collected leaves of 16 individuals of seedling (SD), one-leaf (1L), three-leaves (3L), and flowering (FL) from each population in May 2021, which were genotyped by double digest restriction-site associated DNA sequencing technique (ddRAD-seq). Details on the procedures of SNP call are explained in chapter 3. Using the resultant SNPs, I calculated $H_{ij,t}$ for all pairs of i and j ($1 \leq i \leq 4$, $i \leq j \leq 4$) for each population separately. Here, I treated SD, 1L, 3L, and FL as stage 1, 2, 3, and 4, respectively. As in Chapter 3, I omit t from $H_{ij,t}$ hereafter for tidiness. Denoting p_i as the allele frequency

in stage i, H_{ij} is calculated for each SNP as follows:

$$H_{ij} = p_i(1 - p_j) + (1 - p_i)p_j \tag{4.1}$$

I calculated the mean of H_{ij} over all SNPs, which was used as the observed H_{ij} in the following analyses.

4.2.3 Inverse estimation of life history parameters

I adopted approximate Bayesian computation (ABC) based on sequential Monte Carlo (SMC) to estimate demographic rates and populations size from observed H_{ij}. ABC is a computational approach to estimate parameters: instead of evaluating the exact likelihoods, ABC computes a predefined summary statistics for numerous sets of parameter values which are randomly drawn from a given prior distribution. The parameter values that yield close summary statistics compared to the observed one are accepted to form the posterior distribution. Therefore, ABC can be applicable to situations where likelihoods are too complicated to calculate (Beaumont 2010, Bertorelle et al. 2010, Csilléry et al. 2010). I defined the summary statistics $\boldsymbol{S}$ as the array of H_{ij} whose sum is scaled to be unity.

$$\boldsymbol{S} = \boldsymbol{h} / \sum_{i=1}^{4} \sum_{j \geq i}^{4} H_{ij}, \tag{4.2}$$

$$\boldsymbol{h} = {}^{t}(H_{11}\ H_{22}\ H_{33}\ H_{44}\ H_{12}\ H_{13}\ H_{14}\ H_{23}\ H_{24}\ H_{34}), \tag{4.3}$$

where t indicates transpose. Here I denote the observed and the computed summary statistics as $\tilde{\boldsymbol{S}}$ and $\hat{\boldsymbol{S}}$, and their elements as $\{\tilde{H_{ij}}\}$ and $\{\hat{H_{ij}}\}$, respectively. $\hat{\boldsymbol{S}}$ can be computed from the leading right eigenvector $\boldsymbol{w}$ of the matrix $\boldsymbol{M}$ in Equation (2.10). Because H_{ij}

is proportional to the elements of $\boldsymbol{w}$, as shown in Equation (2.13), $\hat{H_{ij}}$ is formulated as follows:

$$\hat{H_{ij}} = w_{ij} / \sum_{i=1}^{4} \sum_{j \geq i}^{4} w_{ij}. \tag{4.4}$$

The detailed elements of the matrix $\boldsymbol{M}$ is shown in Appendix 2.

The comparison between observed and simulated summary statistics is based on the distance metrics which can be flexibly defined. I evaluated the distance between $\tilde{\boldsymbol{S}}$ and $\hat{\boldsymbol{S}}$, which is denoted by D, by the sum of squared errors between the corresponding elements.

$$D = \sum_{i=1}^{4} \sum_{j \geq i}^{4} (\hat{H_{ij}} - \tilde{H_{ij}})^2 \tag{4.5}$$

ABC-SMC searches the parameter values that yield small D using a large set of parameter values called "population" (Figure 4.1a) (Fountain et al. 2018, Scranton et al. 2014, Speich et al. 2021). The member of the "population" is called "particle," to which a set of parameter values is assigned. The parameter values of particles are updated towards smaller D step by step (Figure 4.1). After updating the particles for sufficient time steps, the parameter values of the final population approximates the posterior distributions. The advantage of SMC method lies in tuning parameter values of a large number of particles in parallel, which allows us to avoid being stuck in local optima.

In implementing ABC-SMC, I first determined prior distributions of all parameters (Table 4.1). The number of particles (N_p) were set to 10,000, and I randomly drew parameter values N_p times from the prior distributions to arrange the initial particles. For each particle, the summary statistics $\hat{\boldsymbol{S}}$ and the distance to the observed summary statistics D was calculated. Then, I assigned weight of $1/N_p$ to each particle. To update the population, particles were sampled with replacement N_p times based on their assigned weights. Parameter values

of the resampled particles were perturbed by a multivariate Gaussian kernel with means equal to the original parameter values and twice the variance-covariance of the particles before resampling. Random perturbation was repeated until D became smaller than the 30 percentile of the population before resampling. By perturbing the 10,000 resampled particles in parallel, I updated the particle assemblages to fit more to the observed data. Weights were also updated with the same method as Beaumont et al. (2009). The updated particles were used as a new population. These procedures were repeated 15 times, and the parameter values of the final particles were used as the posterior distribution. The percentile used for the threshold of accepting perturbed particles was set to become smaller by a factor of 0.95 per step (see Figure 4.1b for the pseudocode).

4.2.4 Estimation based on the field census

I used the estimation results of the integrated population model (IPM) in Chapter 3, which were 2,700,000 (= 900,000 iterations $\times$ 3 chains) sets of parameter values obtained by Markov Chain Monte Carlo (MCMC).

4.2.5 Effective population size

For each of the seven populations, parameter values of the 10,000 final particles of ABC-SMC were used to calculate demographic effective population size (N_e[DE]) based on Equation (2.18). N_e[DE] was also calculated using the parameter values estimated by the IPM of the field census data. The resultant two posterior distributions of N_e[DE] were compared with N_e[LD], which was the N_e estimated from linkage disequilibrium of the flowering individuals in Chapter 3.

4.3 Results

4.3.1 Demographic parameters estimated from observed H_{ij}

H_{ij} ranged approximately from 0.155 to 0.185 (Figure 4.2). ABC-SMC yielded posterior distributions of demographic rates and population size from the observed H_{ij} (Figure 4.3). Except germination rate (g_0), posterior distributions converged to a unimodal shape. The peak of the distributions differed among populations. As for g_0, posterior distributions depicted a flat and broad shape.

4.3.2 Comparison with the estimation based on field census

The parameter values estimated from the genetic data did not match with those estimated from the field census data (Figure 4.4). Fecundity and the probabilities of stasis (s_2, s_3, and s_4) tended to be estimated larger in demographic census, while retrogression and growth rates were larger in genetic estimates. Population size (N) was smaller in the genetic estimates by approximately six orders of 10.

4.3.3 Effective population size

When using demographic rates estimated from the field census data, N_e[DE] was larger than N_e[LD], as shown in chapter 3. On the other hand, When using parameter values estimated from the observed H_{ij}, N_e[DE] ranged 10 to 100, which were smaller than N_e[LD] estimated from linkage disequilibrium. Taking the logarithm of 10, it turned out that parameter values estimated from the observed genetic H_{ij} yielded close order of N_e[DE] to N_e[LD] compared to those estimated from the field census (Figure 4.5).

4.4 Discussion

4.4.1 Difference in the spatio-temporal scale between the field census and the genetic data

I found that the estimated demographic rates and population size differed between the newly-developed method based on genetic data (H_{ij}) and the conventional method based on the field census data. A similar discrepancy is reported when estimating dispersal rate (Cayuela et al. 2018, Lowe and Allendorf 2010, Moore et al. 2017, Yu et al. 2010). Dispersal rate is often estimated either by interpopulation genetic differentiation or by individual capture-mark-recapture. Estimated values generally differ between the two methods, which is often attributed to the mismatch in the spatial and temporal scale between the two types of data (Cayuela et al. 2018). Field census survey that monitors individual over time is often carried out with limited number of samples, and is more vulnerable to missing rare dispersal events. On the other hand, genetic data is the result of all dispersal events including irregular and stochastic movements, and would potentially reflect the net dispersal rate. Besides, while field census captures contemporary movements per se, genetic data reflects dispersal of recent past that constructed the present genetic structure.

The same explanation on the mismatch in the spatio-temporal scale might apply to the present study. As for the spatial scale, the field census was carried out in a limited area of the population (several 1 m^2 quadrats), while genetic samples were collected from the entire population. The mismatch of the spatial scale becomes more serious when considering the within-population spatial genetic structure. When there is a spatial genetic structure within a population, as discussed in Chapter 3 and as shown in Yamagishi et al. (2007), the estimation based on the field census data could only apply to a small portion of the gene pool with

limited spatial area, and the extrapolation to the whole population would be inappropriate to describe the overall genetic dynamics of the population. The same possibility is also discussed in Smith et al. (2020), which revealed that demographic census did not explain the population-level genetic diversity in a perennial plant *Plantago lanceolata*. In terms of the temporal scale, the estimated value of the newly-developed genetic method would reflect growth, stasis, retrogression, and reproduction of the recent past that contributed to the present stage-wise genetic diversity and inter-stage genetic differentiation, while the demographic census reflects the contemporary life history.

4.4.2 Genetically effective life history and population size

Demographic rates estimated from the observed H_{ij} did not satisfy the typical life history characteristics of *Trillium camschatcense* in some populations. *T. camschatcense* is characterized by slow progressive growth and high stasis during juvenile stages (Ohara and Kawano 2005). Although the demography estimated from field census is consistent with the basic life history characteristics, as shown in Chapter 3, the growth rates of 1L and 3L stages (g_1 and g_2) were high in some populations and the stasis rates (s_2 and s_3) were generally low when estimated based on the genetic data. It might be that demographic rates varied over space and time, resulting in the stage-wise genetic composition which cannot be explained by the typical life history of *T. camschatcense*. There are several possible factors that alter demographic rates over space and time. In the fragmented forests in the Tokachi region, strong wind has been causing tree falling and subsequent recovery of the aboveground tree biomass (Tomimatsu et al. 2015), indicating that the openness of the forest crown is changing over time. Besides, human-induced disturbances in relation to the adjacent agricultural fields, such as drainage construction and partial logging, are continuously taking place, influencing

the available sunlight and water. These spatio-temporal environmental changes might have altered demographic rates over space and time, resulting in the necessity to account for quite different life history cycle when describing long-term population-level genetic dynamics.

As for the estimated values of N (the number of individuals in population), the genetic method yielded smaller estimates. While the field census data estimated the raw number of individuals that exist in a whole study site, the new method could have estimated the number of individuals that effectively fulfill the premises of the theoretical model, such as random mating and equal probabilities of growth, stasis, and retrogression among genotypes.

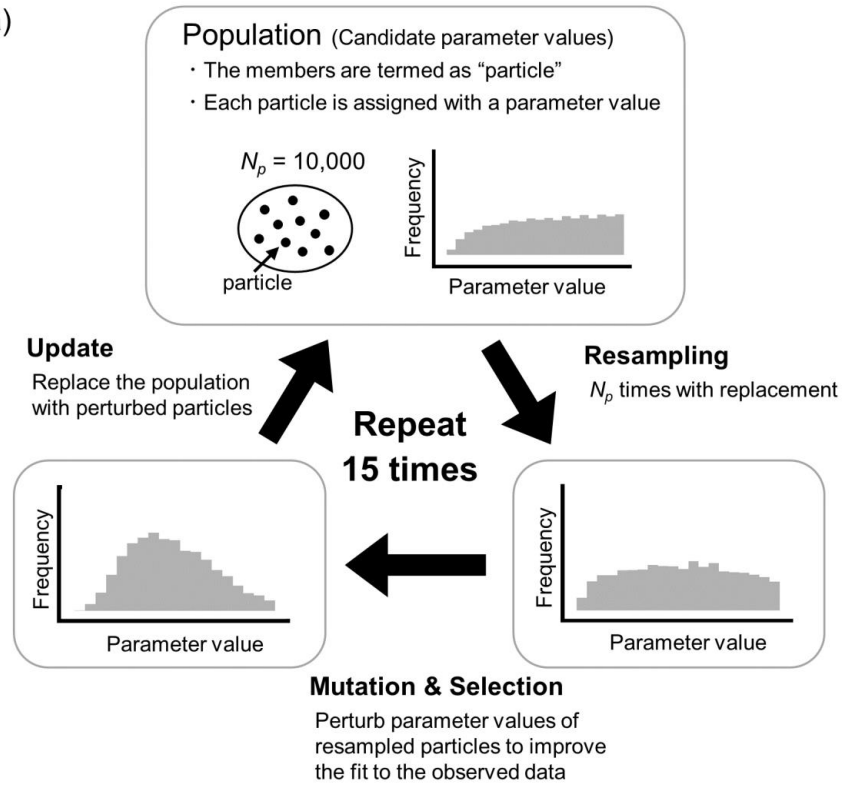
Overall, the new method can estimate life history cycle and population size that explain the observed stage-wise genetic diversity and differentiation at the population level, rather than the fine-scale contemporary population dynamics. The effectiveness in explaining genetic data might be supported by the consistency between N_e[DE] and N_e[LD] compared to the field census data (Figure 4.5).

4.4.3 Problems to be solved

There are some methodological problems to be solved to guarantee the validity of the new method. Firstly, parameter identifiability should be examined. The posterior distribution of g_0, which is the germination rate of seed, did not converge to a unimodal shape, indicating that the estimation was not successful. This is probably because seed (i.e., stage 0) was not sampled and genotyped and H_{ij} that involves with seed (i.e., H_{00}, H_{01}, H_{02}, H_{03}, and H_{04}) were omitted during parameter estimation. Secondly, the robustness to annual variation in genetic sampling remains unclear, which should be examined by estimating the demographic rates and population size with the same protocol in different years. Lastly, the spatio-temporal gaps with field census should be further confirmed. One solution is to

carry out spatially explicit simulation with fine-scale genetic structure to check if the field census and the new method yielded a similar mismatch in silico. Another solution is to fit the scale of census data to that of genetic data. Species for which long-term and large-scale demographic census data is available is a promising target to try the new genetic method. These methodological examination will deepen the understanding on the efficacy of the new method to reconstruct life history from stage-structured genetic data and to describe genetic dynamics of a population.

(a)



(b) **ABC-SMC Algorithm Pseudo-code**

Create the initial population (t = 0)

1. Draw N_p particles (parameter values) from the prior distributions
2. For each particle,
   - 2.1 Calculate $\hat{\boldsymbol{S}}$ using the theoretical model
   - 2.2 Calculate the distance D (= residual sum of squares between $\hat{\boldsymbol{S}}$ and $\tilde{\boldsymbol{S}}$)
   - 2.3 Set weight $1/N_p$

Update the population (t = 1, 2, …,15)

3. For t ($1 \leq t \leq 15$),
   - 3.1 Repeat N_p times
     - 3.1.1 Resample one particle with the probability of the corresponding weight
     - 3.1.2 Perturb the parameter values of the resampled particle using a multivariate Gaussian kernel
     - 3.1.3 Calculate $\hat{\boldsymbol{S}}$ using the perturbed parameter values
     - 3.1.4 Calculate the distance D
     - 3.1.5 If the distance is above the $30 \times 0.95^{t-1}$ percentile of $\{D\}$ of the previous population, go back to 3.1.2. Otherwise, accept the perturbed particles.
     - 3.1.6 Calculate the weight based on Beaumont *et al.* 2009
   - 3.2 Replace the previous population with the accepted particles
   - 3.3 Divide the weights by their sum (normalization)
4. Particles of the final population (t = 15) are used as the posterior distributions

Figure 4.1: (a) A graphical overview of ABC-SMC (approximate Bayesian computation based on sequential Monte Carlo) procedures. (b) Pseudo-code of the ABC-SMC algorithm

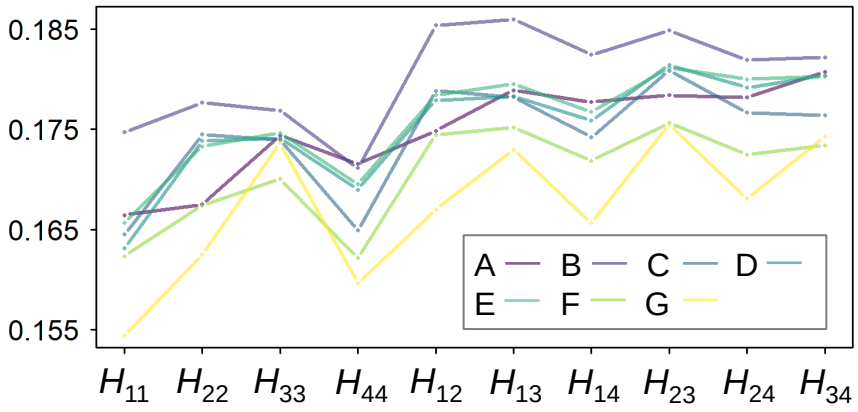


Figure 4.2: Observed H_{ij} in the 7 populations of *Trillium camschatcense* (population A-G). Color denotes population

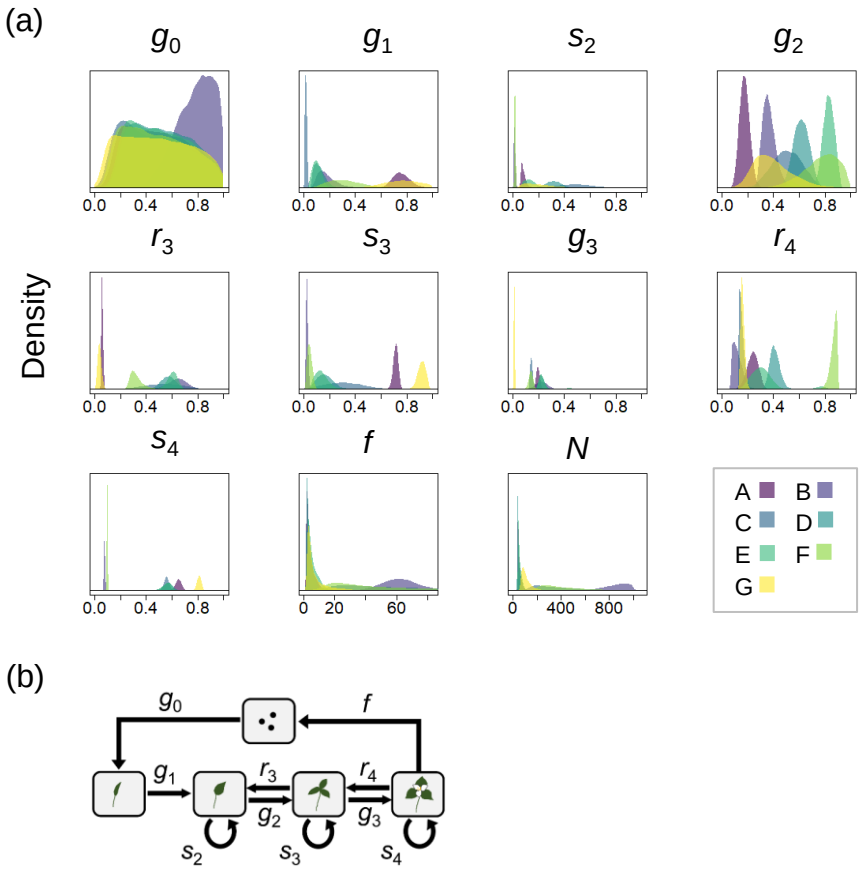


Figure 4.3: (a) The posterior distributions of demographic rates and population size estimated from genetic data (H_{ij}). Color represents population A-G. (b) Life history diagram of *Trillium camschatcense* showing which parameter corresponds to which life history process

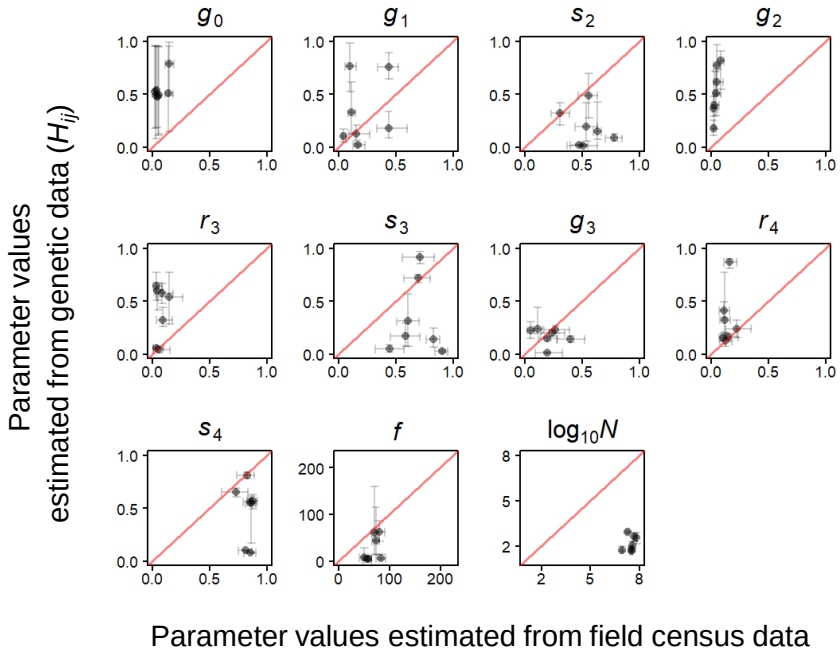


Figure 4.4: The comparison between parameter values estimated from field census data and those estimated from genetic data (H_{ij}). Dots and bars represent the means and the 95% Bayesian credible intervals of the posterior distribution. The red diagonal line in each panel represents $x = y$

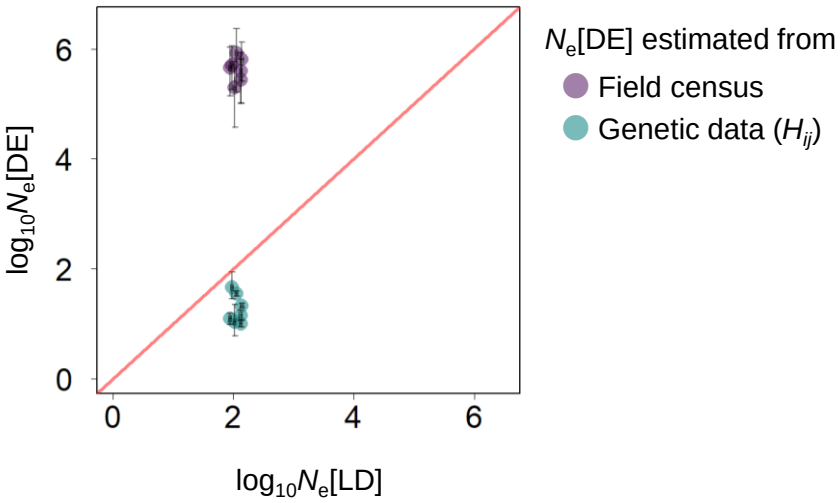


Figure 4.5: Effective population size estimated from linkage disequilibrium (N_e[LD]) was compared with effective population size calculated using demographic rates and population size N_e[DE]. There are two variants in N_e[DE]: those estimated from field census data using integrated population model (shown in purple), and those estimated from genetic data using approximate Bayesian computation based on sequential Monte Carlo (shown in green). The estimated values of N_e[DE] are shown with 95% Bayesian credible interval, While N_e[LD] shown with 95% confidential interval

Table 4.1: Prior distributions of the demographic rates and population size used in ABC-SMC analysis. Uniform distributions were used for all parameters. Because the sum of s_2 and g_2, which is the survival rate of stage 2 (1L), should not exceed 1, the upper limit of g_2 was set to $1 - s_2$. Similarly, the sum of r_3, s_3, and g_3 (survival rate of 3L), and that of r_4 and s_4 (survival rate of FL) were set to range from 0 to 1

Parameter	Prior distribution
g_0	U(0, 1)
g_1	U(0, 1)
s_2	U(0, 1)
g_2	U(0, $1 - s_2$)
r_3	U(0, 1)
s_3	U(0, $1 - r_3$)
g_3	U(0, $1 - r_3 - s_3$)
r_4	U(0, 1)
s_4	U(0, $1 - r_4$)
N	U(0, 1000)

Chapter 5

Temporal skewness of pollination success in the spring ephemeral *Trillum camschatcense*

Abstract

Phenological overlap with pollinators is crucial for reproductive success in insect-pollinated plants. In this study, I examined whether pollinator visitation successfully occurred during an entire flowering season in two populations of the insect-pollinated spring ephemeral *Trillium camschatcense* in the Tokachi region of Hokkaido, northern Japan. I bagged flowers and excluded pollinator visitation during either the first or the last half of the entire flowering season to compare pollination success between the two periods. The two populations have experienced differing levels of climate warming in the last 60 years, which impacted pollinator visitation. In the population experiencing temperature rise more rapidly, fertilization rate and seed set decreased sharply when bagged during the first half period, indicating that pollinator visitation is skewed to the early part of the flowering season. The temporal skewness of pollination success would be an early warning signal of the impacts of climate warming on the reproductive success of *T. camschatcense*.

5.1 Introduction

Climate warming is altering the flowering phenology of many plant species (Root et al. 2003, Walther et al. 2002), particularly in early blooming spring ephemerals that sprout and flower directly after snowmelt (Fitter and Fitter 2002). Because the phenological response to rising temperature differs between taxa and species (Parmesan 2007, Post 2019), pollinator activities does not necessarily shift in synchrony with flowering, potentially resulting in a temporal mismatch and reduced pollinator visitation (Kudo and Cooper 2019). It is crucial to evaluate the degree of phenological overlap with pollinators for insect-pollinated spring ephemeral plants as temperatures continue to rise.

Trillium camschatcense is an insect-pollinated, long-lived spring ephemeral that sprouts in mid to late April and flowers in May in temperate deciduous forests. Although most populations of *T. camschatcense* in Japan are self-compatible, those in the Tokachi region of Hokkaido, northern Japan, are self-incompatible and owe seed production completely to insect pollinators (Ohara et al. 1996). Therefore, a phenological mismatch with pollinators could be a serious threat to the reproductive success of *T. camschatcense*.

In this study, I examined whether pollination successfully occurred during a complete flowering season of *Trillium camschatcense* at two locations in the Tokachi region, where long-term temperature data was available. I hypothesized that pollination was more temporally limited due to a phenological mismatch at the location experiencing more rapid warming.

5.2 Materials and Methods

5.2.1 Study species

Trillium camschatcense is an understory herb whose primary pollinators are Coleoptera and Diptera (Tomimatsu and Ohara 2003a). *T. camschatcense* does not produce inflorescences made up of multiple flowers. Flowering individuals usually have one flowering stem (sometimes two stems), with one flower and three leaves per stem. *T. camschatcense* does not compensate for reproductive failure by producing additional flowers later in the flowering season: the number of flowers per individual is pre-determined before flowering onset. Most flowering stems synchronously sprout in April, start anthesis in early May, and finish flowing around late May to early June. After the flowering season, fruits ripen in July (Ohara and Kawano 2005, Ohara et al. 2006).

5.2.2 Study Sites

The present study was conducted in two populations located in: Obihiro (inland, 42.802 °N, 143.103 °E) and Hiroo (seacoast, 42.316 °N, 143.324 °E) (Figure 5.1 (a)). The two populations were approximately 57 kilometers apart. Obihiro is a remnant forest surrounded by agricultural lands. Hiroo is a part of a windbreak forest that stretches along the coastline. The two populations were relatively large for the region in terms of population density and habitat area (Figures 5.2 and 5.3).

5.2.3 Temperature

I obtained the daily temperature of Obihiro and Hiroo from 1958 to 2019 from the website of the Japan Meteorological Agency (JMA, https://www.jma.go.jp/jma/index.html). I extracted the data for April and May, the growing and flowering months of *Trillium camschatcense*. For each site, I carried out a linear regression against the mean daily temperature per year, using year as an explanatory variable.

5.2.4 Flowering phenology and precipitation

In early April 2021, I randomly established two 3 m $\times$ 3 m quadrats in each population. Before flowering onset, I labeled all flowering stems within the quadrats in late April (Figure 5.4a). Flowering stems that sprouted later were labeled right after I noticed their emergence. I monitored the labeled flowers at an interval of 2 to 11 days. I defined the first and the last day of the anthesis of each flower as the date on which we first confirmed that the flower became open/withered. Along with phenological surveys, daily precipitation during the flowering season in the study year (2021), as well as that in the past years (from 1958 to 2019), was obtained from the website of the JMA.

5.2.5 Bagging experiment

I divided the entire flowering period of *Trillium camschatcense* into two parts, the first half and last half, and implemented three treatments: (1) the control, in which flowers were left intact throughout the flowering period; (2) the first half, in which flowers were left intact during the first half, but then bagged with cellophane bags during the last half, (3) and the last half, in which flowers were bagged prior to anthesis and then opened during the last half (Figure 5.4b). The cellophane bags excluded pollinators from flowers. Therefore, comparing seed production among the three treatments could reveal the relative frequency of pollinator visitation during an unbagged period.

Before flowering onset, I randomly selected 30 individuals for the first half and the last half treatments around each 3 m $\times$ 3 m quadrat, as well as 30 individuals within the quadrat for control. Approximately 40 flowers were selected for each treatment in each quadrat. Flowers for the last half treatment were bagged at that time. Because *Trillium camschatcense* in the Tokachi region usually flowers in May and early June, I opened flowers for the last half treatment but bagged those for the first half treatment within one week from the midpoint of May (May 16): on May 17 and 22 in Obihiro and Hiroo, respectively (Figure 5.4c). I measured the length and width of a leaf of each flowering stem to account for the confounding effects of plant size on reproduction.

In early July, I collected fruits and counted the number of unfertilized ovules, fertilized ovules that did not develop to seeds due to abortion, and seeds (Figure 5.4d). I recorded the presence/absence of stem injury caused by trampling and herbivory on seeds by caterpillars, both of which took place before collecting fruits and might have overridden the effects of bagging treatments by decreasing the number of seeds. I defined "fertilization rate" and

"seed set" as follows:

$$\text{fertilization rate} = \frac{\text{sum of fertilized ovules and seeds}}{\text{total sum of ovules and seeds}} \tag{5.1}$$

$$\text{seed set} = \frac{\text{sum of seeds}}{\text{total sum of ovules and seeds}} \tag{5.2}$$

I constructed a generalized linear mixed model (GLMM) with binomial error and logit link for fertilization rate and seed set, using the function "glmer" in the R package "lme4" (Bates et al. 2015). Models were separately constructed for Obihiro and Hiroo. For each response variable, I incorporated the treatment, size (= leaf length $\times$ leaf width), presence/absence of herbivory and stem injury as fixed effects, and individual and quadrat as random effects. I carried out a multiple comparison test for treatment by a Tukey method with Benjamini-Hochberg correction, using the R package "multcomp" (Hothorn et al. 2008).

5.3 Results

5.3.1 Temperature

The mean daily temperature of April and May in Obihiro significantly increased from 1958 to 2019 at a rate of 0.021 °C per year (P = 0.005, Figure 5.1b). Although the temperature in Hiroo was also rising, the rate of 0.012 °C per year was not statistically significant (P = 0.106), indicating that Obihiro was experiencing more rapid warming compared to Hiroo.

5.3.2 Flowering phenology and precipitation

In the four quadrats used for phenological survey (two in each population), all flowers continued anthesis across the first and the last half period, depicting unimodal trajectories (Figure 5.5a). The timing of the onset, the peak, and the end of flowering was earlier overall in

Obihiro than in Hiroo. Daily precipitation of the study year (2021), as well as the average daily precipitation over the past 60 years (1958-2019), mostly remained low throughout the flowering period in both sites (Figure 5.5b).

5.3.3 Bagging experiment

The fertilization rate of the last half treatment was lower than that of control and first half treatment in Obihiro (Figure 5.6a). On the other hand, there were no significant differences among the three treatments in Hiroo (Figure 5.6b). The seed set followed the same pattern as the fertilization rate: the last half treatment was significantly lower than the others in Obihiro (Figure 5.6c), while no significant differences were detected in Hiroo (Figure 5.6d).

5.4 Discussion

In Obihiro, the last half treatment decreased both fertilization rate and seed set, while the first half treatment was comparable to the control (Figure 5.6a, c), suggesting that pollinators mostly visited flowers during the first half period, and that pollinator visitation was infrequent during the last half. On the other hand, in Hiroo, neither bagging treatment caused reproductive failure compared to the control (Figure 5.6b, d), indicating that pollinators actively visited flowers during the entire flowering period.

Pollinator scarcity during the last half period in Obihiro may be a result of climate warming. Temperature rise was more pronounced in Obihiro than in Hiroo (Figure 5.1b). Because phenological advance in response to warming is generally slower in plants than in other taxa (Parmesan 2007, Post 2019), *T. camschatcense* may not have caught up with the accelerating phenology of pollinators during the 60 years of rapid temperature rise. Another possibility is pollinator-mediated plant-plant competition (Rathcke 1983), in which

the flowering onset of other plant species that had formerly flowered after *T. camschatcense* advanced and overlapped with the last half period. I think it is less likely because the study sites were thoroughly dominated by *T. camschatcense* during its flowering period, and no other species appeared to compete evenly (Figure 5.3). Considering that precipitation remained low throughout the flowering season (Figure 5.5b), it was also unlikely that precipitation specifically inhibited pollinator visitation during the last half period in Obihiro.

There are some remaining issues to be solved in future studies. Firstly, the seed set was lower than the fertilization rate (Figure 5.6c, d), indicating not all fertilized ovules acquired enough resources to become seeds (i.e., resource limitation). A pollen supplementation experiment is necessary to examine if the lowered reproductive success during the last half period in Obihiro could be entirely attributed to reduced pollination or could be partly caused by resource limitation as well. Secondly, the pollinating fauna of Obihiro is not completely the same as that of Hiroo at a species level (Tomimatsu and Ohara 2003a). Considering the species-specific phenological shifts under climate warming (Parmesan 2007, Post 2019), not only do the trends of temperature rise but the pollinators' response might be essentially different between the two populations. Direct observation of pollinating insects will help reassess the temporal skewness of visitation with special attention to pollinator species composition.

This study reported the possibility of a phenological mismatch in *Trillium camschatcense* with their pollinators. While habitat fragmentation has been considered the primary threat to reproductive success in *T. camschatcense* in the Tokachi region (Tomimatsu and Ohara 2002), phenological mismatches with pollinators have not been explored. The impacts of the temporal skewness of pollination success on reproductive success should be further examined in future studies.

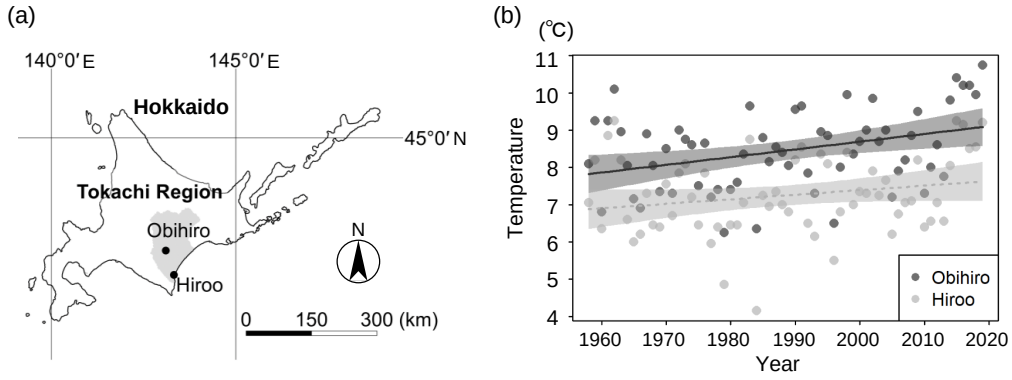


Figure 5.1: (a) The location of the Tokachi region (gray) and the two studied populations: Obihiro and Hiroo in the Tokachi region of Hokkaido. (b) Mean daily temperature in April and May from 1958 to 2019 at Obihiro city (dark) and Hiroo town (light). Regression lines and their 95 % confidence intervals are shown. Solid and dotted lines indicate significant and non-significant effects of year, respectively

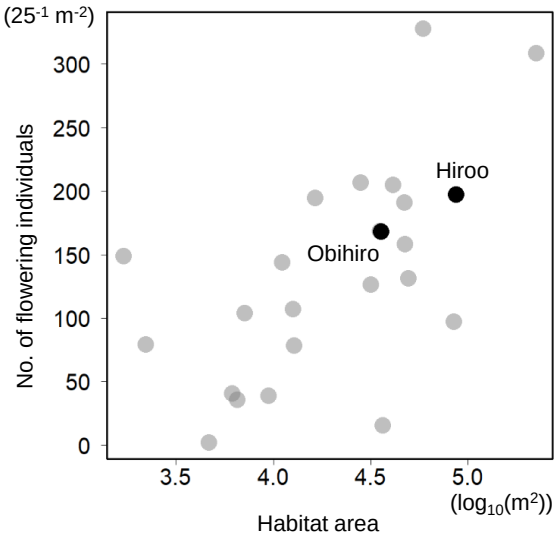


Figure 5.2: In May 2018, we selected 24 *Trillium camschatcense* populations in the Tokachi region. We established one to three 5 m × 5 m quadrats in each population and recorded the number of flowering individuals within the quadrats. We found a significant positive correlation between the logarithm of habitat area and flowering density (correlation coefficient = 0.602, P = 0.002). The two study sites, Obihiro and Hiroo, were relatively large compared to the other populations

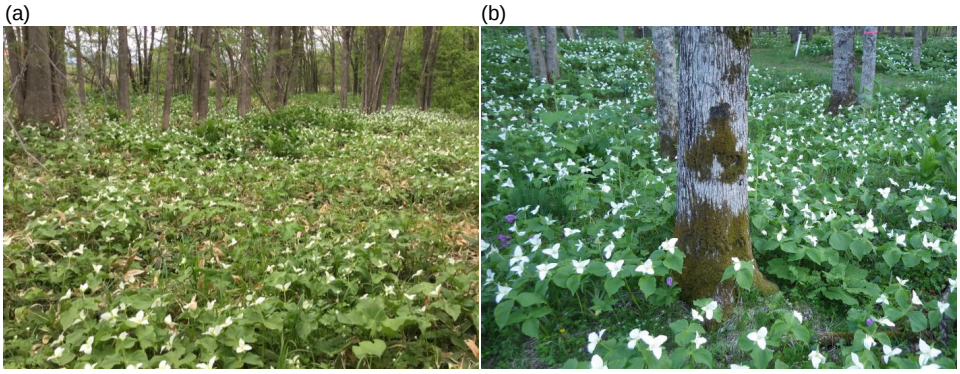


Figure 5.3: The study sites during the flowering season. a: Obihiro; b: Hiroo. One flower of *Trillium camschatcense* consists of three white petals, forming a white triangular shape. In both sites, flowering individuals of *T. camschatcense* distributed thoroughly on the forest floor

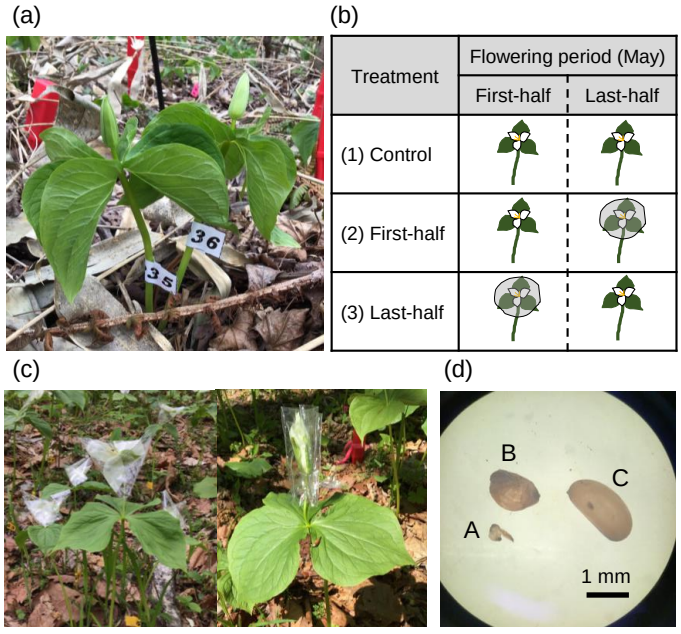


Figure 5.4: (a) Flower buds were labeled by attaching number tags to stems. (b) A schematic view of the three treatments of the bagging experiments. Translucent circles overlaid on flowers stand for cellophane bags, which keep pollinators away from flowers. (c) Individuals bagged with cellophane bags. Left and right images show first- and last-half treatment, respectively. (d) Three types of ovule/seed in fruits. A: an unfertilized ovule; B: a fertilized ovule that did not developed to seed; C: a mature seed

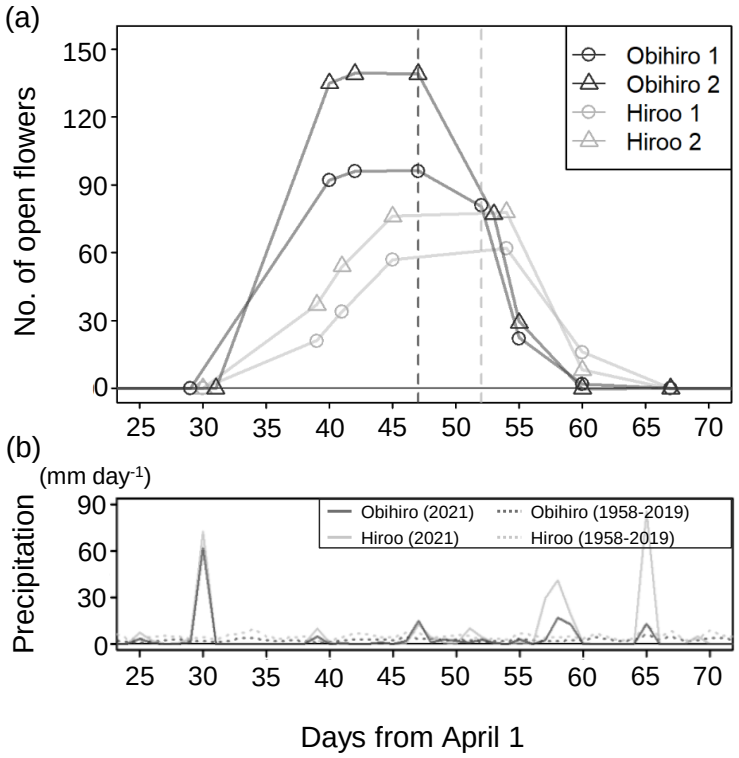


Figure 5.5: (a) Flowering phenology of the four quadrats: two in Obihiro (dark) and in Hiroo (light), respectively. The intervals between observations (dots) are connected by straight lines. Two quadrats in the same population are distinguished by the shape of dots. Dotted vertical lines show the turning point days on which bagging was ceased in the last-half treatment but was applied to the first-half instead. We labeled all flowering stems on the first observation day. As for the two quadrats in Obihiro, 71 and 106 individuals with 98 and 141 flowers were found in each, while 54 and 70 individuals with 62 and 79 flowers were found in Hiroo. In each quadrat, only one or two flowering stems sprouted later than the first observation (in Obihiro 1, one stem on May 10; in Obihiro 2, two on May 10; in Hiroo 1, one on May 11; in Hiroo 2, one on May 15). We confirmed the start of anthesis for these delayed flowers on the same day as we first noticed their sprouting. (b) Daily precipitation during flowering period. Two solid lines represent precipitation in the study year (2021) in Obihiro (dark) and Hiroo (light). Two dotted lines represent the average from 1958 to 2019 in Obihiro (dark) and Hiroo (light). Precipitation remained low during flowering season in the study year, which was the same as in the past years

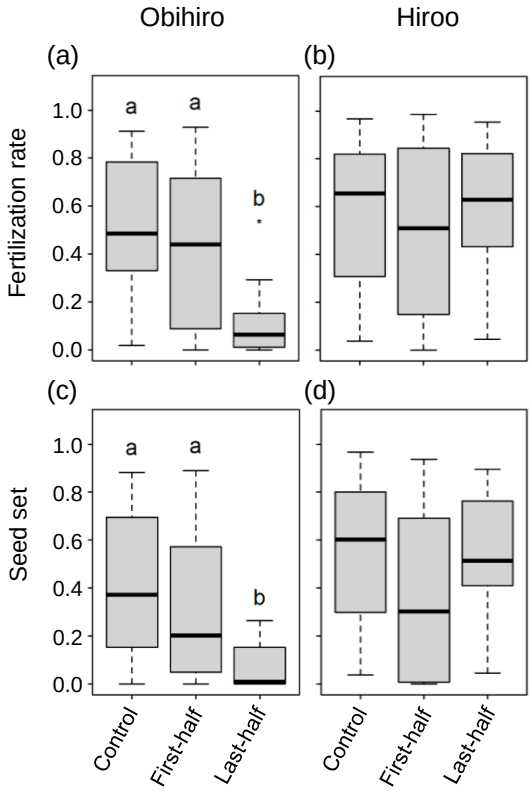


Figure 5.6: Fertilization rate (a and b) and seed set (c and d) of the three treatments in Obihiro and Hiroo: control (unbagged), first half (bagged during the last half of the flowering period), and last half (bagged during the first half). Significant difference was detected for pairs marked with different letters in each panel (significance level = 0.05). No letters are assigned when there were no significantly different pairs

Chapter 6

General discussion

6.1 The genetic effectiveness of life history in perennial plant populations

In Chapter 2, I derived the theoretical model that describes demographic genetic structure in stage-structured plant populations. Analyzing the model, I showed that life history stages with high static and retrogressive inflows accumulated genetic diversity compared to the other stages, and that slow growth and prolonged lifespan contributed to maintaining genetic diversity over time. The effects of life history were concealed by large population size: genetic diversity was maintained regardless of life history under large population size. Although previous studies revealed that particular life history characteristics, such as long generation time, contributed the maintenance of genetic diversity (Aparicio et al. 2012, Austerlitz and Garnier-Géré 2003, Hamrick and Godt 1996, Loveless and Hamrick 1984), the conditions on which the effects of life history was crucial to genetic diversity remained unclear. My theoretical analysis revealed that stasis and retrogression were the key demographic process for delaying generation turnover and accumulating genetic diversity under a small population size (Figure 2.8). Besides, the results that the contribution of life history to maintaining genetic diversity becomes apparent under small population size highlights the importance of considering life history processes for evaluating genetic diversity in anthropogenically impacted small populations.

In Chapter 3, I applied the theoretical model to the small fragmented populations of

the understory perennial herb *Trillium camschatcense* in the Tokachi region, Hokkaido. I estimated demographic rates and population size by the four years of field census survey, and the estimated values were subsequently substituted to the model to obtain the theoretical prediction of demographic genetic structure. The theory predicted no apparent difference in genetic diversity among stage classes, indicating that the estimated population size N was still large enough to obscure the effects of life history despite being fragmented and downsized. However, empirical genetic analysis using ddRAD-seq technique revealed that the theoretical expectation did not match the real demographic genetic structure. Genetic diversity varied significantly among life history stages, with 3L stage possessed higher genetic diversity than the others (Figure 3.10).

The discrepancy between observation and theoretical prediction is not special to the present study, and has been continually on the agenda in the long history of population genetics. Lewontin et al. (1974) pointed out that the observed genetic variation was much more narrower among species than expected from the large variation in census population size (Lewontin's paradox). The explanation of this paradox has been sought for years (Bobay and Ochman 2018, Buffalo 2021, Ellegren and Galtier 2016, Filatov 2019, Romiguier et al. 2014), and one flagship solution is to consider "effective" population size (N_e), which is the size of the ideal discrete-generation population that can explain the observed level of genetic diversity in parallel to the real population size. The term "effective" implies that the real number of individuals, which is estimated by the census, is merely an apparent population size and is ineffective in explaining genetic dynamics per generation. Effective population size is usually far smaller than census population size (Frankham 1995, Leffler et al. 2012), because wild populations violate the assumptions of the theoretical model such as constant population size, random mating, and the even sex ratio (Peart et al. 2020).

In analogous to the concept of effective population size, I propose the concept of "genetically effective life history" to account for the observed demographic genetic structure. In Chapter 4, I inversely estimated demographic rates and population size N from the observed genetic data, which turned out to be different from those estimated from the field census data (Figure 4.4). What deviates from the effective population in this study from the conventional effective population size (N_e) is that this study evaluated the effectiveness in the form of stage-structured populations with overlapping generation using the model developed in Chapter 2, not in the form of the discrete-generation population which was parameterized only by population size. Therefore, the effective population in this study enables us to estimate genetically effective stage-specific demographic rates and compare them with those estimated from the census. The discrepancy between the two might indicate that not only population size but also the overall lifetime trajectory from birth to death should be evaluated through the lens of genes, not through the lens of real individuals, when understanding genetic diversity of stage-structured populations.

Considering the small size of the genetically effective population compared to that of the census-based estimated values (Figure 4.4), the largeness of population size would only slightly contributed to the maintenance of genetic diversity while life history might play an important role in the fragmented *Trillium camschatcense* populations. Therefore, merely maintaining the number of individuals may be insufficient in the conservation of genetic diversity, and we also have to pay attention to whether the whole life history process from birth to death is going as usual as undisturbed conditions.

6.2 Trade-off between maintaining genetic diversity and rapid evolution

In Chapter 5, I showed that there is a temporal skewness of pollination success in *Trillium camschatcense* in the Tokachi region. In Obihiro, in which relatively rapid climate warming is taking place, successful pollination that contributed to seed set was concentrated during the first-half period of the entire flowering season (Figure 5.6), indicating that phenological overlap with its pollinators is temporally limited to the first-half period. The cause of the temporal skewness might be the ongoing climate warming, which is known to provoke phenological mismatch between mutually interacting plants and pollinators. Flowering phenology of *Trillium* species is considered to be a trait under selection (Barfield et al. 2011, Knight 2004, Knight et al. 2008), and the ongoing climate change might work as a selection pressure to facilitate phenological advance in the fragmented *T. camschatcense* populations. Genetic diversity could play an important role in immediate and successful adaptation to climate change.

It should be noted, however, that high genetic diversity does not necessarily result in successful rapid adaptation. Maintaining genetic diversity means that the potential of adaptive evolution increases on one hand, but it also means that not only adaptive genes but also non-adaptive ones are more likely to be maintained on the other hand (Orive et al. 2017, Yamamichi et al. 2019). In other words, life history strategies that can maintain high genetic diversity can be poor at rapid adaptation. Both high genetic diversity and rapid adaptation are required for successful evolutionary rescue, and the lack of either of the two requirements threatens population viability. Previous studies showed that life history with dormant stages, such as seed bank, leads to the accumulation of both adaptive and non-adaptive individuals, and slows down the speed of adaptation, resulting in the collapse of the population in some

conditions (Yamamichi et al. 2019).

Therefore, the trade-off between maintaining genetic diversity and promoting rapid evolution should be considered when evaluating viable life history strategies under environmental changes. Viability along the fast-slow continuum of life history can change in response to the mode of environmental changes, including the speed (abrupt or gradual), the stochasticity, the frequency, and the intensity: the more stable population size is, the less necessary it is to hurry up the population recovery, and thus long-lived life histories (i.e., the strategies with high genetic diversity and slow adaptation) might be favored. Adaptive evolution from standing genetic variation and the subsequent population persistence in stage-structured populations, especially those with high stasis and retrogressive characteristics, is awaited to be further examined under various scenarios of environmental changes in future studies.

6.3 Concluding remarks

In this dissertation, I showed the complex relationships between life history and genetic diversity in stage-structured perennial plant populations, which was not appreciated until recently. Charlesworth (1994), which is the pioneering textbook on the evolution in age-structured populations, argues as follows:

> "*So far, such models* [stage-structured models] *have had little application to population genetics and life-history evolution. This is partly due to the difficulty of providing concise descriptions of evolutionary dynamics with these models, and in part to the fact that they fail to include any direct effect of age on survival and fecundity. They will not be considered further in this book.*" (p. 11)

Stage-structured plant life histories are indeed complex and sometimes difficult to handle. Contrary to most animals whose life history is age-dependent, individuals do not simply

grow and reproduce with increasing age but also stay in the same stage for successive years and retrogress to more juvenile stages. In Chapter 2, However, I showed that stasis and retrogression, which are the emergent properties of stage-structured life histories, are the key drivers of maintaining genetic diversity. Besides, it is acknowledged that the survival and reproduction of plants usually do not show age-dependent changes over the course of life (Jones et al. 2014) and are more subjected to developmental stage (Silvertown 1987). Considering these things, there is no reason to look away from the stage-structured plant life histories. Together with the recent theoretical achievements (Barfield et al. 2011, Cotto et al. 2020), this study is a promising step towards the understanding of the genetic and evolutionary dynamics of plant populations.

In evolutionary ecology, the life cycle as a whole is considered to be a trait of evolution (Cole 1954), and life history strategy has been studied for its evolutionary background (Stearns 1992). The fact that life history affects genetic diversity (i.e., the potential capacity of future adaptation) indicates that life history is one of the determinants of evolvability (Riederer et al. 2022) and serves not only as a target but also a driver of adaptive evolution. The bilateral aspect of life history might be a key factor in driving the chains of evolution in biological populations. The genetic and evolutionary consequences of life history, which are shown in my dissertation, would contribute to elucidating the possible roles life history play in evolutionary ecology.

Acknowledgements

I would like to thank my supervisor Professor Masashi Ohara for his guidance and encouragement throughout my studies. He allowed me to explore what I want to do, which gave me opportunities to try new things such as theoretical modeling and population genetics analyses. With his patience and support, I think I have grown as a researcher. I am also grateful to Professor Shigeyuki Koshikawa, Associate Professor Shunsuke Utsumi, and Professor Emeritus Takenori Takada for giving me invaluable advice during my studies. I thank Dr. Mitsuhiko P. Sato, who was an excellent collaborator and helped me a lot with regard to NGS (next generation sequencing) analyses. I appreciate the cooperation from Professor Koji Suzuki, Professor Emeritus Hitoshi Suzuki, Associate Professor Takashi Hayakawa, and Research Associate Kiyohito Yoshida during genetic experiments. I thank the people in Hiroo town, Obihiro city, and Memuro town for kindly allowing my field work in the Tokachi region. Especially, Hiroo town supported the lodging facility, which helped me a lot to concentrate on my field work. Lastly, I would like to thank the members of the Ohara laboratory and the course in ecological genetics for daily interesting discussions, which always motivated my passion for research. This research was financially supported by the Japanese Society for the Promotion of Science, Fujiwara Natural History Foundation, and the Japanese Science Society.

References

Agashe, D., Falk, J.J., Bolnick, D.I., 2011. Effects of founding genetic variation on adaptation to a novel resource. *Evolution*, 65, 2481–2491.

Agrawal, A.A., Hastings, A.P., Maron, J.L., 2021. Evolution and seed dormancy shape plant genotypic structure through a successional cycle. *Proceedings of the National Academy of Sciences*, 118, e2026212118.

Aguilar, R., Cristóbal-Pérez, E.J., Balvino-Olvera, F.J., de Jesús Aguilar-Aguilar, M., Aguirre-Acosta, N., Ashworth, L., Lobo, J.A., Martén-Rodríguez, S., Fuchs, E.J., Sanchez-Montoya, G., et al., 2019. Habitat fragmentation reduces plant progeny quality: a global synthesis. *Ecology Letters*, 22, 1163–1173.

Aguilar, R., Quesada, M., Ashworth, L., Herrerias-Diego, Y., Lobo, J., 2008. Genetic consequences of habitat fragmentation in plant populations: susceptible signals in plant traits and methodological approaches. *Molecular Ecology*, 17, 5177–5188.

Aldrich, P.R., Hamrick, J.L., Chavarriaga, P., Kochert, G., 1998. Microsatellite analysis of demographic genetic structure in fragmented populations of the tropical tree *Symphonia globulifera*. *Molecular Ecology*, 7, 933–944.

Ally, D., Ritland, K., 2006. A case study: Looking at the effects of fragmentation on genetic structure in different life history stages of old-growth mountain hemlock (*Tsuga mertensiana*). *Journal of Heredity*, 98, 73–78.

Anderson, J.T., Inouye, D.W., McKinney, A.M., Colautti, R.I., Mitchell-Olds, T., 2012. Phenotypic plasticity and adaptive evolution contribute to advancing flowering phenology in response to climate change. *Proceedings of the Royal Society B: Biological Sciences*, 279, 3843–3852.

Aparicio, A., Hampe, A., Fernández-Carrillo, L., Albaladejo, R.G., 2012. Fragmentation and comparative genetic structure of four mediterranean woody species: complex interactions between life history traits and the landscape context. *Diversity and Distributions* 18, 226–235.

Arroyo-Cosultchi, G., Golubov, J., Mandujano, M.C., Salguero-Gómez, R., Martínez, A.J., 2022. What are the demographic consequences of a seed bank stage for columnar cacti? *Population Ecology*, 64, 35–46.

Austerlitz, F., Garnier-Géré, P.H., 2003. Modelling the impact of colonisation on genetic diversity and differentiation of forest trees: interaction of life cycle, pollen flow and seed long-distance dispersal. *Heredity*, 90, 282–290.

Bacles, C.F., Jump, A.S., 2011. Taking a tree's perspective on forest fragmentation genetics. *Trends in Plant Science*, 16, 13–18.

Bakker, J., Van Rijswijk, M., Weissing, F., Bijlsma, R., 2010. Consequences of fragmentation for the ability to adapt to novel environments in experimental *Drosophila* metapopulations. *Conservation Genetics*, 11, 435–448.

Barfield, M., Holt, R.D., Gomulkiewicz, R., 2011. Evolution in stage-structured populations. *The American Naturalist*, 177, 397–409.

Barrett, R.D., Schluter, D., 2008. Adaptation from standing genetic variation. *Trends in Ecology and Evolution*, 23, 38–44.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48.

Beaumont, M.A., 2010. Approximate Bayesian computation in evolution and ecology. *Annual Review of Ecology, Evolution, and Systematics*, , 379–406.

Beaumont, M.A., Cornuet, J.M., Marin, J.M., Robert, C.P., 2009. Adaptive approximate bayesian computation. *Biometrika*, 96, 983–990.

Benítez-Malvido, J., Lázaro, A., Ferraz, I.D., 2018. Effect of distance to edge and edge interaction on seedling regeneration and biotic damage in tropical rainforest fragments: A long-term experiment. *Journal of Ecology*, 106, 2204–2217.

Berens, D., Braun, C., González-Martínez, S.C., Griebeler, E., Nathan, R., Böhning-Gaese, K., 2014. Fine-scale spatial genetic dynamics over the life cycle of the tropical tree *Prunus africana*. *Heredity* 113, 401–407.

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.

Bezemer, N., Krauss, S.L., Roberts, D.G., Hopper, S.D., 2019. Conservation of old individual trees and small populations is integral to maintain species' genetic diversity of a historically fragmented woody perennial. *Molecular Ecology*, 28, 3339–3357.

Bobay, L.M., Ochman, H., 2018. Factors driving effective population size and pan-genome evolution in bacteria. *BMC Evolutionary Biology*, 18, 1–12.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics*, 30, 2114–2120.

Buffalo, V., 2021. Quantifying the relationship between genetic diversity and population size suggests natural selection cannot explain lewontin's paradox. *Elife*, 10, e67509.

Cannon, C.H., Piovesan, G., Munné-Bosch, S., 2022. Old and ancient trees are life history lottery winners and vital evolutionary resources for long-term adaptive capacity. *Nature Plants*, 8, 136–145.

Capdevila, P., Beger, M., Blomberg, S.P., Hereu, B., Linares, C., Salguero-Gómez, R., 2020. Longevity, body dimension and reproductive mode drive differences in aquatic versus terrestrial life-history strategies. *Functional Ecology*, 34, 1613–1625.

Carey, J.R., Roach, D.A., 2020. Biodemography: An Introduction to Concepts and Methods. Princeton University Press.

Carlson, S.M., Cunningham, C.J., Westley, P.A., 2014. Evolutionary rescue in a changing world. *Trends in Ecology and Evolution*, 29, 521–530.

Caswell, H., 2001. Matrix Population Models. 2 ed., Sinauer Associates, Inc.

Cayuela, H., Rougemont, Q., Prunier, J.G., Moore, J.S., Clobert, J., Besnard, A., Bernatchez, L., 2018. Demographic and genetic approaches to study dispersal in wild animal populations: A methodological review. *Molecular Ecology* 27, 3976–4010.

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., Lee, J.J., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, 4, s13742–015.

Charlesworth, B., 1994. Evolution in Age-Structured Populations. 2 ed., Cambridge University Press.

Chung, M.Y., Epperson, B.K., Gi Chung, M., 2003. Genetic structure of age classes in *Camellia japonica* (theaceae). *Evolution* 57, 62–73.

Cole, L.C., 1954. The population consequences of life history phenomena. *The Quarterly Review of Biology*, 29, 103–137.

Cotto, O., Schmid, M., Guillaume, F., 2020. Nemo-age: Spatially explicit simulations of eco-evolutionary dynamics in stage-structured populations under changing environments. *Methods in Ecology and Evolution*, 11, 1227–1236.

Crow, J., Kimura, M., 1970. An Introduction to Population Genetics Theory. Harper and Row.

Csilléry, K., Blum, M.G., Gaggiotti, O.E., François, O., 2010. Approximate Bayesian computation (ABC) in practice. *Trends in Ecology and Evolution*, 25, 410–418.

De Kort, H., Prunier, J.G., Ducatez, S., Honnay, O., Baguette, M., Stevens, V.M., Blanchet, S., 2021. Life history, climate and biogeography interactively affect worldwide genetic diversity of plant and animal populations. *Nature Communications*, 12, 1–11.

Do, C., Waples, R.S., Peel, D., Macbeth, G., Tillett, B.J., Ovenden, J.R., 2014. NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (N_e) from genetic data. *Molecular Ecology Resources*, 14, 209–214.

Doak, D.F., Gross, K., Morris, W.F., 2005. Understanding and predicting the effects of sparse data on demographic analyses. *Ecology*, 86, 1154–1163.

Ehrlén, J., Lehtilä, K., 2002. How perennial are perennial plants? *Oikos*, 98, 308–322.

Ellegren, H., Galtier, N., 2016. Determinants of genetic diversity. *Nature Reviews Genetics*, 17, 422–433.

Ellis, E.C., Klein Goldewijk, K., Siebert, S., Lightman, D., Ramankutty, N., 2010. Anthropogenic transformation of the biomes, 1700 to 2000. *Global Ecology and Biogeography*, 19, 589–606.

Ellner, S., 1996. Environmental fluctuations and the maintenance of genetic diversity in age or stage-structured populations. *Bulletin of Mathematical Biology*, 58, 103–127.

Ellner, S., Hairston Jr, N.G., 1994. Role of overlapping generations in maintaining genetic variation in a fluctuating environment. *The American Naturalist*, 143, 403–417.

Epperson, B.K., 2000. Spatial genetic structure and non-equilibrium demographics within plant populations. *Plant Species Biology* 15, 269–279.

Fahrig, L., 2003. Effects of habitat fragmentation on biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, , 487–515.

Felsenstein, J., 1971. Inbreeding and variance effective numbers in populations with overlapping generations. *Genetics*, 68, 581–597.

Filatov, D.A., 2019. Extreme lewontin's paradox in ubiquitous marine phytoplankton species. *Molecular Biology and Evolution*, 36, 4–14.

Fisher, R.A., 1930. The genetical theory of natural selection. The Clarendon Press.

Fitter, A., Fitter, R., 2002. Rapid changes in flowering time in British plants. *Science*, 296, 1689–1691.

Fountain, E.D., Kang, J.K., Tempel, D.J., Palsbøll, P.J., Pauli, J.N., Zachariah Peery, M., 2018. Genomics meets applied ecology: Characterizing habitat quality for sloths in a tropical agroecosystem. *Molecular Ecology*, 27, 41–53.

Frankham, R., 1995. Effective population size/adult population size ratios in wildlife: a review. *Genetics Research*, 66, 95–107.

Fukano, Y., Guo, W., Uchida, K., Tachiki, Y., 2020. Contemporary adaptive divergence of plant competitive traits in urban and rural populations and its implication for weed management. *Journal of Ecology*, 108, 2521–2530.

González, A.V., Gómez-Silva, V., Ramírez, M.J., Fontúrbel, F.E., 2020. Meta-analysis of the differential effects of habitat fragmentation and degradation on plant genetic diversity. *Conservation Biology*, 34, 711–720.

Hairston, N.G., Ellner, S.P., Geber, M.A., Yoshida, T., Fox, J.A., 2005. Rapid evolution and the convergence of ecological and evolutionary time. *Ecology Letters*, 8, 1114–1127.

Hamrick, J.L., Godt, M.W., 1996. Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 351, 1291–1298.

Hardy, O.J., Maggia, L., Bandou, E., Breyne, P., Caron, H., Chevallier, M.H., Doligez, A., Dutech, C., Kremer, A., Latouche-Hallé, C., et al., 2006. Fine-scale genetic structure and gene dispersal inferences in 10 Neotropical tree species. *Molecular Ecology*, 15, 559–571.

Hill, W.G., 1972. Effective size of populations with overlapping generations. *Theoretical Population Biology*, 3, 278–289.

Hill, W.G., 1979. A note on effective population size with overlapping generations. *Genetics*, 92, 317–322.

Hothorn, T., Bretz, F., Westfall, P., 2008. Simultaneous inference in general parametric models. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 50, 346–363.

Hughes, P.W., 2017. Between semelparity and iteroparity: Empirical evidence for a continuum of modes of parity. *Ecology and Evolution*, 7, 8232–8261.

Johnson, D.L., 1977. Inbreeding in populations with overlapping generations. *Genetics*, 87, 581–591.

Jones, O.R., Scheuerlein, A., Salguero-Gómez, R., Camarda, C.G., Schaible, R., Casper, B.B., Dahlgren, J.P., Ehrlén, J., García, M.B., Menges, E.S., Quintana-Ascencio, P.F., Caswell, H., Baudisch, A., Vaupel, J.W., 2014. Diversity of aging across the tree of life. *Nature*, 505, 169–173.

Kellner, K., 2021. jagsUI: A wrapper around 'rjags' to streamline 'JAGS' analyses. URL: https://CRAN.R-project.org/package=jagsUI. R package version 1.5.2.

Kéry, M., Schaub, M., 2011. Bayesian population analysis using WinBUGS: a hierarchical perspective. Academic Press.

Kettle, C.J., Hollingsworth, P.M., Jaffre, T., Moran, B., Ennos, R.A., 2007. Identifying the early genetic consequences of habitat degradation in a highly threatened tropical conifer, *Araucaria nemorosa* Laubenfels. *Molecular Ecology*, 16, 3581–3591.

Knight, T.M., 2004. The effects of herbivory and pollen limitation on a declining population of *Trillium grandiflorum*. *Ecological Applications*, 14, 915–928.

Knight, T.M., Barfield, M., Holt, R.D., 2008. Evolutionary dynamics as a component of stage-structured matrix models: an example using *Trillium grandiflorum*. *The American Naturalist*, 172, 375–392.

Knight, T.M., Caswell, H., Kalisz, S., 2009. Population growth rate of a common understory herb decreases non-linearly across a gradient of deer herbivory. *Forest Ecology and Management*, 257, 1095–1103.

Kramer, A.T., Ison, J.L., Ashley, M.V., Howe, H.F., 2008. The paradox of forest fragmentation genetics. *Conservation Biology*, 22, 878–885.

Kudo, G., Cooper, E.J., 2019. When spring ephemerals fail to meet pollinators: mechanism of phenological mismatch and its impact on plant reproduction. *Proceedings of the Royal Society B*, 286, 20190573.

Leffler, E.M., Bullaughey, K., Matute, D.R., Meyer, W.K., Segurel, L., Venkat, A., Andolfatto, P., Przeworski, M., 2012. Revisiting an old riddle: what determines genetic diversity levels within species? *Plos Biology*, 10, e1001388.

Lefkovitch, L., 1965. The study of population growth in organisms grouped by stages. *Biometrics*, , 1–18.

Leimu, R., Mutikainen, P., Koricheva, J., Fischer, M., 2006. How general are positive relationships between plant population size, fitness and genetic variation? *Journal of Ecology*, 94, 942–952.

Lewontin, R.C., et al., 1974. The genetic basis of evolutionary change. volume 560. Columbia University Press New York.

Li, X.Y., Kurokawa, S., Giaimo, S., Traulsen, A., 2016. How life history can sway the fixation probability of mutants. *Genetics*, 203, 1297–1313.

Linhart, Y.B., Mitton, J.B., Sturgeon, K.B., Davis, M.L., 1981. Genetic variation in space and time in a population of poderosa pine. *Heredity*, 46, 407–426.

Loveless, M.D., Hamrick, J.L., 1984. Ecological determinants of genetic structure in plant populations. *Annual Review of Ecology and Systematics*, , 65–95.

Lowe, A., Boshier, D., Ward, M., Bacles, C., Navarro, C., 2005. Genetic resource impacts of habitat loss and degradation; reconciling empirical evidence and predicted theory for neotropical trees. *Heredity*, 95, 255–273.

Lowe, A., Cavers, S., Boshier, D., Breed, M., Hollingsworth, P.M., 2015. The resilience of forest fragmentation genetics—no longer a paradox—we were just looking in the wrong place. *Heredity*, 115, 97–99.

Lowe, W.H., Allendorf, F.W., 2010. What can genetics tell us about population connectivity? *Molecular Ecology*, 19, 3038–3051.

Luu, K., Bazin, E., Blum, M.G., 2017. pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, 17, 67–77.

Martins, E., Lamont, R.W., Martinelli, G., Lira-Medeiros, C., Quinet, A., Shapcott, A., 2015. Genetic diversity and population genetic structure in three threatened *Ocotea* species (Lauraceae) from Brazil's Atlantic Rainforest and implications for their conservation. *Conservation Genetics*, 16, 1–14.

Metcalf, C.J.E., Pavard, S., 2007. Why evolutionary biologists should be demographers. *Trends in Ecology and Evolution* 22, 205–212.

Mimura, M., Yahara, T., Faith, D.P., Vázquez-Domínguez, E., Colautti, R.I., Araki, H., Javadi, F., Núñez-Farfán, J., Mori, A.S., Zhou, S., Hollingsworth, P.M., Neaves, L.E., Fukano, Y., Smith, G.F., Sato, Y.I., Tachida, H., Hendry, A.P., 2017. Understanding and monitoring the consequences of human impacts on intraspecific variation. *Evolutionary Applications*, 10, 121–139.

Moore, J.S., Harris, L.N., Le Luyer, J., Sutherland, B.J., Rougemont, Q., Tallman, R.F., Fisk, A.T., Bernatchez, L., 2017. Genomics and telemetry suggest a role for migration harshness in determining overwintering habitat choice, but not gene flow, in anadromous Arctic Char. *Molecular Ecology*, 26, 6784–6800.

Moore, J.W., Yeakel, J.D., Peard, D., Lough, J., Beere, M., 2014. Life-history diversity and its importance to population stability and persistence of a migratory fish: steelhead in two large North American watersheds. *Journal of Animal Ecology*, 83, 1035–1046.

Murray, M., Thompson, W., 1980. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Research*, 8, 4321–4326.

Murren, C.J., 2003. Spatial and demographic population genetic structure in *Catasetum viridiflavum* across a human-disturbed habitat. *Journal of Evolutionary Biology*, 16, 333–342.

Nunney, L., 2002. The effective size of annual plant populations: the interaction of a seed bank with fluctuating population size in maintaining genetic variation. *The American Naturalist*, 160, 195–204.

Ohara, M., Kawano, S., 2005. Life-history monographs of japanese plants. 2: *Trillium camschatcense* Ker-Gawl.(Trilliaceae). *Plant Species Biology*, 20, 75–82.

Ohara, M., Takada, T., Kawano, S., 2001. Demography and reproductive strategies of a polycarpic perennial, *Trillium apetalon* (Trilliaceae). *Plant Species Biology*, 16, 209–217.

Ohara, M., Takeda, H., Ohno, Y., Shimamoto, Y., 1996. Variations in the breeding system and the population genetic structure of *Trillium kamtschaticum* (Liliaceae). *Heredity*, 76, 476–484.

Ohara, M., Tomimatsu, H., Takada, T., Kawano, S., 2006. Importance of life history studies for conservation of fragmented populations: a case study of the understory herb, *Trillium camschatcense*. *Plant Species Biology*, 21, 1–12.

Orive, M., 1993. Effective population size in organisms with complex life-histories. *Theoretical Population Biology*, 44, 316–340.

Orive, M.E., Barfield, M., Fernandez, C., Holt, R.D., 2017. Effects of clonal reproduction on evolutionary lag and evolutionary rescue. *The American Naturalist*, 190, 469–490.

Orr, H.A., Unckless, R.L., 2014. The population genetics of evolutionary rescue. *PLoS Genetics*, 10, e1004551.

Orsini, L., Schwenk, K., De Meester, L., Colbourne, J.K., Pfrender, M.E., Weider, L.J., 2013. The evolutionary time machine: using dormant propagules to forecast how populations can adapt to changing environments. *Trends in Ecology and Evolution*, 28, 274–282.

Parmesan, C., 2007. Influences of species, latitudes and methodologies on estimates of phenological response to global warming. *Global Change Biology*, 13, 1860–1872.

Peart, C.R., Tusso, S., Pophaly, S.D., Botero-Castro, F., Wu, C.C., Aurioles-Gamboa, D., Baird, A.B., Bickham, J.W., Forcada, J., Galimberti, F., et al., 2020. Determinants of genetic variation across eco-evolutionary scales in pinnipeds. *Nature Ecology and Evolution*, 4, 1095–1104.

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., Hoekstra, H.E., 2012. Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PloS One*, 7, e37135.

Peterson, G.W., Dong, Y., Horbach, C., Fu, Y.B., 2014. Genotyping-by-sequencing for plant genetic diversity analysis: a lab guide for SNP genotyping. *Diversity*, 6, 665–680.

Post, E., 2019. Time in Ecology. Princeton University Press.

R Core Team, 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Ramsayer, J., Kaltz, O., Hochberg, M.E., 2013. Evolutionary rescue in populations of *Pseudomonas fluorescens* across an antibiotic gradient. *Evolutionary Applications*, 6, 608–616.

Rathcke, B., 1983. Chapter 12 - competition and facilitation among plants for pollination, in: Real, L. (Ed.), Pollination Biology. Academic Press, pp. 305–329.

Reed, D.H., Frankham, R., 2003. Correlation between fitness and genetic diversity. *Conservation Biology*, 17, 230–237.

Riederer, J.M., Tiso, S., van Eldijk, T.J., Weissing, F.J., 2022. Capturing the facets of evolvability in a mechanistic framework. *Trends in Ecology and Evolution*, 37, 430–439.

Rochette, N.C., Rivera-Colón, A.G., Catchen, J.M., 2019. Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology*, 28, 4737–4754.

Romiguier, J., Gayral, P., Ballenghien, M., Bernard, A., Cahais, V., Chenuil, A., Chiari, Y., Dernat, R., Duret, L., Faivre, N., et al., 2014. Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature*, 515, 261–263.

Root, T.L., Price, J.T., Hall, K.R., Schneider, S.H., Rosenzweig, C., Pounds, J.A., 2003. Fingerprints of global warming on wild animals and plants. *Nature*, 421, 57–60.

Rowe, G., Beebee, T., 2004. Reconciling genetic and demographic estimators of effective population size in the anuran amphibian *Bufo calamita*. *Conservation Genetics*, 5, 287–298.

Salguero-Gómez, R., Jones, O.R., Jongejans, E., Blomberg, S.P., Hodgson, D.J., Mbeau-Ache, C., Zuidema, P.A., De Kroon, H., Buckley, Y.M., 2016. Fast-slow continuum and reproductive strategies structure plant life-history variation worldwide. *Proceedings of the National Academy of Sciences*, 113, 230–235.

Schaub, M., Kéry, M., 2021. Integrated population models: theory and ecological applications with R and JAGS. Academic Press.

Schindler, D.E., Hilborn, R., Chasco, B., Boatright, C.P., Quinn, T.P., Rogers, L.A., Webster, M.S., 2010. Population diversity and the portfolio effect in an exploited species. *Nature*, 465, 609–612.

Schmeller, D.S., Merilä, J., 2007. Demographic and genetic estimates of effective population and breeding size in the amphibian *Rana temporaria*. *Conservation Biology*, 21, 142–151.

Schmidt, D.J., Fallon, S., Roberts, D.T., Espinoza, T., McDougall, A., Brooks, S.G., Kind, P.K., Bond, N.R., Kennard, M.J., Hughes, J.M., 2018. Monitoring age-related trends in genomic diversity of Australian lungfish. *Molecular Ecology*, 27, 3231–3241.

Schoener, T.W., 2011. The newest synthesis: understanding the interplay of evolutionary and ecological dynamics. *Science*, 331, 426–429.

Scranton, K., Knape, J., de Valpine, P., 2014. An approximate bayesian computation approach to parameter estimation in a stochastic stage-structured population model. *Ecology*, 95, 1418–1428.

Shea, K., Kelly, D., 1998. Estimating biocontrol agent impact with matrix models: *Carduus nutans* in New Zealand. *Ecological Applications*, 8, 824–832.

Silvertown, J., 1987. Introduction to Plant Population Ecology. 2 ed., Longman Scientific and Technical.

Silvertown, J., Franco, M., Menges, E., 1996. Interpretation of elasticity matrices as an aid to the management of plant populations for conservation. *Conservation Biology* 10, 591–597.

Silvertown, J., Franco, M., Pisanty, I., Mendoza, A., 1993. Comparative plant demography–relative importance of life-cycle components to the finite rate of increase in woody and herbaceous perennials. *Journal of Ecology* , 465–476.

Smith, A.L., Hodkinson, T.R., Villellas, J., Catford, J.A., Csergő, A.M., Blomberg, S.P., Crone, E.E., Ehrlén, J., Garcia, M.B., Laine, A.L., et al., 2020. Global gene flow releases invasive plants from environmental constraints on genetic diversity. *Proceedings of the National Academy of Sciences*, 117, 4218–4227.

Speich, M., Dormann, C.F., Hartig, F., 2021. Sequential Monte-Carlo algorithms for bayesian model calibration—A review and method comparison . *Ecological Modelling*, 455, 109608.

Stearns, S.C., 1992. The Evolution of Life Histories. Oxford University Press.

Tomimatsu, H., Ohara, M., 2002. Effects of forest fragmentation on seed production of the understory herb *Trillium camschatcense*. *Conservation Biology*, 16, 1277–1285.

Tomimatsu, H., Ohara, M., 2003a. Floral visitors of *Trillium camschatcense* (Trilliaceae) in fragmented forests. *Plant Species Biology*, 18, 123–127.

Tomimatsu, H., Ohara, M., 2003b. Genetic diversity and local population structure of fragmented populations of *Trillium camschatcense* (Trilliaceae). *Biological Conservation*, 109, 249–258.

Tomimatsu, H., Ohara, M., 2004. Edge effects on recruitment of *Trillium camschatcense* in small forest fragments. *Biological Conservation*, 117, 509–519.

Tomimatsu, H., Ohara, M., 2010. Demographic response of plant populations to habitat fragmentation and temporal environmental variability. *Oecologia*, 162, 903–911.

Tomimatsu, H., Yamagishi, H., Suzuki, S.N., Sato, C., Konno, Y., 2015. Long-term dynamics of small fragmented forests inferred from patterns along a gradient of fragment sizes. *Ecological Research*, 30, 1057–1064.

Tomimatsu, H., Yamagishi, H., Tanaka, I., Sato, M., Kondo, R., Konno, Y., 2011. Consequences of forest fragmentation in an understory plant community: extensive range expansion of native dwarf bamboo. *Plant Species Biology*, 26, 3–12.

Torices, R., Méndez, M., Gómez, J.M., 2011. Where do monomorphic sexual systems fit in the evolution of dioecy? Insights from the largest family of angiosperms. *New Phytologist*, 190, 234–248.

Tsuzuki, Y., Sato, M.P., Matsuo, A., Suyama, Y., Ohara, M., 2022a. Genetic consequences of habitat fragmentation in a perennial plant *Trillium camschatcense* are subjected to its slow-paced life history. *Population Ecology*, 64, 5–18.

Tsuzuki, Y., Takada, T., Ohara, M., 2022b. Modeling temporal dynamics of genetic diversity in stage-structured plant populations with reference to demographic genetic structure. *Theoretical Population Biology*, 148, 76–85.

Vekemans, X., Hardy, O.J., 2004. New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology* 13, 921–935.

Vranckx, G., Jacquemyn, H., Mergeay, J., Cox, K., Kint, V., Muys, B., Honnay, O., 2014. Transmission of genetic variation from the adult generation to naturally established seedling cohorts in small forest stands of pedunculate oak (*Quercus robur* L.). *Forest Ecology and Management*, 312, 19–27.

Waddle, E., Piedrahita, L.R., Hall, E.S., Kendziorski, G., Morris, W.F., DeMarche, M.L., Doak, D.F., 2019. Asynchrony in individual and subpopulation fecundity stabilizes reproductive output of an alpine plant population. *Ecology*, 100, e02639.

Walther, G.R., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T.J., Fromentin, J.M., Hoegh-Guldberg, O., Bairlein, F., 2002. Ecological responses to recent climate change. *Nature* 416, 389–395.

Waples, R.S., Do, C., Chopelet, J., 2011. Calculating N_e and N_e/N in age-structured populations: a hybrid Felsenstein-Hill approach. *Ecology*, 92, 1513–1522.

Waples, R.S., Luikart, G., Faulkner, J.R., Tallmon, D.A., 2013. Simple life-history traits explain key effective population size ratios across diverse taxa. *Proceedings of the Royal Society B: Biological Sciences*, 280, 20131339.

Williams, S.L., 2001. Reduced genetic diversity in eelgrass transplantations affects both population growth and individual fitness. *Ecological Applications*, 11, 1472–1488.

Wright, S., 1931. Evolution in Mendelian populations. *Genetics* 16, 97.

Yamagishi, H., Tomimatsu, H., Ohara, M., 2007. Fine-scale spatial genetic structure within continuous and fragmented populations of *Trillium camschatcense*. *Journal of Heredity* 98, 367–372.

Yamamichi, M., Hairston, N.G., Rees, M., Ellner, S.P., 2019. Rapid evolution with generation overlap: the double-edged effect of dormancy. *Theoretical Ecology* 12, 179–195.

Yonezawa, K., Kinoshita, E., Watano, Y., Zentoh, H., 2000. Formulation and estimation of the effective size of stage-structured populations in *Fritillaria camtschatcensis*, a perennial herb with a complex life history. *Evolution*, 54, 2007–2013.

Young, A., Boyle, T., Brown, T., 1996. The population genetic consequences of habitat fragmentation for plants. *Trends in Ecology and Evolution*, 11, 413–418.

Yu, H., Nason, J.D., Ge, X., Zeng, J., 2010. Slatkin's Paradox: when direct observation and realized gene flow disagree. A case study in *Ficus*. *Molecular Ecology*, 19, 4441–4453.

Chapter A1

Appendix of chapter 2

A1.1 Model development

A1.1.1 Formulation of $H_{ij,t}$

As explained in the main text, $H_{ij,t}$ is split into six subsets:

$$\begin{aligned} H_{ij,t} =& H_{ij,t}|_{1\cap A\cap\alpha} + H_{ij,t}|_{1\cap A\cap\beta} + H_{ij,t}|_{1\cap A\cap\gamma} \\ &+ H_{ij,t}|_{1\cap B\cap\alpha} + H_{ij,t}|_{1\cap B\cap\beta} + H_{ij,t}|_{1\cap B\cap\gamma}, \end{aligned} \tag{A1.1}$$

where $H_{ij,t}|_{X\cap Y\cap Z}$ stands for $H_{ij,t}$ under the concurrence of case X, Y, and Z ($X = 1, 2$; $Y = A, B$; $Z = \alpha, \beta, \gamma$).

We define sub-stage i_{ms} and i_{mr}, which consist of individuals transferred from stage m to i by survival and by reproduction, respectively. Each $H_{ij,t}|_{1\cap A\cap Z}$ can be formulated using $H_{i_{ms}j_{ms},t}$ (case $1\cap A\cap\alpha$), $H_{i_{ms}j_{mr},t}$, $H_{i_{mr}j_{ms},t}$ (case $1\cap A\cap\beta$), and $H_{i_{mr}j_{mr},t}$ (case $1\cap A\cap\gamma$) weighted by the corresponding number of two-gene pairs.

$$H_{ij,t}|_{1\cap A\cap\alpha} = \sum_{m=1}^{n} \frac{N_{i_{ms}} N_{j_{ms}}}{N_i N_j} H_{i_{ms}j_{ms},t} \tag{A1.2}$$

$$H_{ij,t}|_{1\cap A\cap\beta} = \sum_{m=1}^{n} \left(\frac{N_{i_{ms}} N_{j_{mr}}}{N_i N_j} H_{i_{ms}j_{mr},t} + \frac{N_{i_{mr}} N_{j_{ms}}}{N_i N_j} H_{i_{mr}j_{ms},t} \right) \tag{A1.3}$$

$$H_{ij,t}|_{1\cap A\cap\gamma} = \sum_{m=1}^{n} \frac{N_{i_{mr}} N_{j_{mr}}}{N_i N_j} H_{i_{mr}j_{mr},t}. \tag{A1.4}$$

Here, $N_{i_{ms}}$ and $N_{j_{ms}}$ denote the number of individuals in sub-stage i_{ms} and j_{ms}, respectively. As for case $1\cap A\cap\alpha$, two genes, each sampled from stage i and j, belong to sub-stage i_{ms} and j_{ms} with the chance of $\frac{2N_{i_{ms}}}{2N_i} \times \frac{2N_{j_{ms}}}{2N_j}$. The number of genes is twice as many as that of individuals because we assume diploid species. Thus, $H_{i_{ms}j_{ms},t}$ is weighted by $\frac{N_{i_{ms}} N_{j_{ms}}}{N_i N_j}$, as shown in equation A1.2 . Case $1\cap A\cap\beta$ (equation A1.3) and $1\cap A\cap\gamma$ (equation A1.4) are similarly formulated.

In the concurrence of case 1, A, and α, two genes, each sampled from sub-stage i_{ms} and j_{ms}, cannot be the same gene because one gene in stage m in year $t-1$ could not move to both stage i and j simultaneously by survival. Therefore, $H_{i_{ms}j_{ms},t}$ is equal to the probability that two genes randomly sampled from stage m "without" replacement in time $t-1$ are not identical-by-descent.

Here, we define $H'_{ij,t}$ as the probability that two genes sampled from stage i and j

"without" replacement in time t are non-identical-by-descent. Because $H_{i_{ms}j_{ms},t}$ is equal to $H'_{mm,t-1}$ in case $1 \cap A \cap \alpha$, we formulate $H'_{mm,t-1}$. When sampling two genes from stage m with replacement in year $t-1$, the same gene can be sampled twice with the probability of $\frac{1}{2N_m} \times \frac{1}{2N_m} \times 2N_m = \frac{1}{2N_m}$, which makes no contribution to $H_{mm,t-1}$. Therefore, $H_{mm,t-1}$ can be formulated as follows.

$$H_{mm,t-1} = \frac{1}{2N_m} \times 0 + \left(1 - \frac{1}{2N_m}\right) \times H'_{mm,t-1} \tag{A1.5}$$

As a result, $H'_{mm,t-1}$ is obtained.

$$H_{i_{ms}j_{ms},t} = H'_{mm,t-1} = \frac{1}{1 - 1/(2N_m)} H_{mm,t-1} \tag{A1.6}$$

Unlike transfer by survival, transfer by reproduction allows the same gene to move multiple pathways simultaneously, because genes are replicated. In the case of $1 \cap A \cap \beta$ and $1 \cap A \cap \gamma$, sampling in year t does not preclude the chance of sampling the same gene twice, because at least one of the two genes are transferred by reproduction. Therefore,

$$H_{i_{ms}j_{mr},t} = H_{i_{mr}j_{ms},t} = H_{i_{mr}j_{mr},t} = H_{mm,t-1} \tag{A1.7}$$

The number of genes in each sub-stage is given by

$$N_{i_{ms}} = t_{im} N_m \tag{A1.8}$$

$$N_{j_{ms}} = t_{jm} N_m \tag{A1.9}$$

$$N_{i_{mr}} = f_{im} N_m \tag{A1.10}$$

$$N_{j_{mr}} = f_{jm} N_m. \tag{A1.11}$$

Substituting equations A1.6 -A1.11 to equations A1.2 -A1.4 ,

$$H_{ij,t}|_{1 \cap A \cap \alpha} = \sum_{m=1}^{n} \left\{ \frac{t_{im} t_{jm} N_m^2}{N_i N_j} \times \frac{1}{1 - 1/(2N_m)} H_{mm,t-1} \right\} \tag{A1.12}$$

$$H_{ij,t}|_{1 \cap A \cap \beta} = \sum_{m=1}^{n} \left\{ \frac{(t_{im} f_{jm} + f_{im} t_{jm}) N_m^2}{N_i N_j} \times H_{mm,t-1} \right\} \tag{A1.13}$$

$$H_{ij,t}|_{1 \cap A \cap \gamma} = \sum_{m=1}^{n} \left(\frac{f_{im} f_{jm} N_m^2}{N_i N_j} \times H_{mm,t-1} \right). \tag{A1.14}$$

As for $H_{ij,t}|_{1 \cap B \cap Z}$ $(Z = \alpha, \beta, \gamma)$, two genes, each sampled from stage i and j in time t, were in stage k and l, or in stage l and k, in year $t-1$ respectively. In either situation, the probability of non-identical-by-descent remains the same as that in year $t-1$, which is $H_{kl,t-1}$, regardless of whether they were transferred only by survival (case $1 \cap B \cap \alpha$), both

by survival and by reproduction (case $1 \cap B \cap \beta$), or only by reproduction (case $1 \cap B \cap \gamma$).

$$
\begin{aligned}
H_{ij,t}|_{1\cap B\cap\alpha} &= \sum_{k=1}^{n}\sum_{l>k}^{n}\left(\frac{N_{i_{ks}}N_{j_{ls}}}{N_iN_j}H_{i_{ks}j_{ls},t} + \frac{N_{i_{ls}}N_{j_{ks}}}{N_iN_j}H_{i_{ls}j_{ks},t}\right)\\
&= \sum_{k=1}^{n}\sum_{l>k}^{n}\left(\frac{t_{ik}t_{jl}N_kN_j}{N_iN_j}H_{kl,t-1} + \frac{t_{il}t_{jk}N_kN_j}{N_iN_j}H_{kl,t-1}\right)\\
&= \sum_{k=1}^{n}\sum_{l>k}^{n}\left\{\frac{(t_{ik}t_{jl}+t_{il}t_{jk})N_kN_j}{N_iN_j}H_{kl,t-1}\right\} \qquad \text{(A1.15)}\\
H_{ij,t}|_{1\cap B\cap\beta} &= \sum_{k=1}^{n}\sum_{l>k}^{n}\left(\frac{N_{i_{ks}}N_{j_{lr}}}{N_iN_j}H_{i_{ks}j_{lr},t} + \frac{N_{i_{kr}}N_{j_{ls}}}{N_iN_j}H_{i_{kr}j_{ls},t}\right.\\
&\qquad \left. + \frac{N_{i_{ls}}N_{j_{kr}}}{N_iN_j}H_{i_{ls}j_{kr},t} + \frac{N_{i_{lr}}N_{j_{ks}}}{N_iN_j}H_{i_{lr}j_{ks},t}\right)\\
&= \sum_{k=1}^{n}\sum_{l>k}^{n}\left(\frac{t_{ik}f_{jl}N_kN_l}{N_iN_j}H_{kl,t-1} + \frac{f_{ik}t_{jl}N_kN_l}{N_iN_j}H_{kl,t-1}\right.\\
&\qquad \left. + \frac{t_{il}f_{jk}N_kN_l}{N_iN_j}H_{kl,t-1} + \frac{f_{il}t_{jk}N_kN_l}{N_iN_j}H_{kl,t-1}\right)\\
&= \sum_{k=1}^{n}\sum_{l>k}^{n}\left\{\frac{(t_{ik}f_{jl}+f_{ik}t_{jl}+t_{il}f_{jk}+f_{il}t_{jk})N_kN_l}{N_iN_j}H_{kl,t-1}\right\} \qquad \text{(A1.16)}\\
H_{ij,t}|_{1\cap B\cap\gamma} &= \sum_{k=1}^{n}\sum_{l>k}^{n}\left(\frac{N_{i_{kr}}N_{j_{lr}}}{N_iN_j}H_{i_{kr}j_{lr},t} + \frac{N_{i_{lr}}N_{j_{kr}}}{N_iN_j}H_{i_{lr}j_{kr},t}\right)\\
&= \sum_{k=1}^{n}\sum_{l>k}^{n}\left(\frac{f_{ik}f_{jl}N_kN_j}{N_iN_j}H_{kl,t-1} + \frac{f_{il}f_{jk}N_kN_j}{N_iN_j}H_{kl,t-1}\right)\\
&= \sum_{k=1}^{n}\sum_{l>k}^{n}\left\{\frac{(f_{ik}f_{jl}+f_{il}f_{jk})N_kN_j}{N_iN_j}H_{kl,t-1}\right\} \qquad \text{(A1.17)}
\end{aligned}
$$

Substituting equations A1.12 -A1.17 to equation A1.1 , we can formulate $H_{ij,t}$ as follows.

$$
\begin{aligned}
H_{ij,t} &= \sum_{m=1}^{n}\left\{\frac{t_{im}t_{jm}N_m^2}{N_iN_j}\times\frac{1}{1-1/(2N_m)} + \frac{(t_{im}f_{jm}+f_{im}t_{jm})N_m^2}{N_iN_j}\right.\\
&\qquad \left. \frac{f_{im}f_{jm}N_m^2}{N_iN_j}\right\}H_{mm,t-1}\\
&\quad + \sum_{k=1}^{n}\sum_{l>k}^{n}\left\{\frac{(t_{ik}t_{jl}+t_{il}t_{jk})N_kN_l}{N_iN_j} + \frac{(t_{ik}f_{jl}+f_{ik}t_{jl}+t_{il}f_{jk}+f_{il}t_{jk})N_kN_l}{N_iN_j}\right.\\
&\qquad \left. + \frac{(f_{ik}f_{jl}+f_{il}f_{jk})N_kN_l}{N_iN_j}\right\}H_{kl,t-1}\\
&= \sum_{m=1}^{n}\frac{N_m^2}{N_iN_j}\left\{\frac{t_{im}t_{jm}}{1-1/(2N_m)} + f_{im}t_{jm} + t_{im}f_{jm} + f_{im}f_{jm}\right\}H_{mm,t-1}
\end{aligned}
$$

$$
\begin{aligned}
&+\sum_{k=1}^{n}\sum_{l>k}^{n}\frac{N_kN_l}{N_iN_j}\left((t_{ik}+f_{ik})(t_{jl}+f_{jl})+(t_{il}+f_{il})(t_{jk}+f_{jk})\right)H_{kl,t-1}.\\
&=\sum_{m=1}^{n}\frac{N_m^2}{N_iN_j}\left\{\frac{t_{im}t_{jm}}{1-1/(2N_m)}+f_{im}t_{jm}+t_{im}f_{jm}+f_{im}f_{jm}\right\}H_{mm,t-1}\\
&+\sum_{k=1}^{n}\sum_{l>k}^{n}\frac{N_kN_l}{N_iN_j}(a_{ik}a_{jl}+a_{il}a_{jk})H_{kl,t-1}. \qquad \text{(A1.18)}
\end{aligned}
$$

A1.1.2 Formulation of $H_{ii,t}$

$H_{ii,t}$ is split into mutually exclusive six subsets:

$$
\begin{aligned}
H_{ii,t}=&H_{ii,t}|_{2\cap A\cap\alpha}+H_{ii,t}|_{2\cap A\cap\beta}+H_{ii,t}|_{2\cap A\cap\gamma}\\
&+H_{ii,t}|_{2\cap B\cap\alpha}+H_{ii,t}|_{2\cap B\cap\beta}+H_{ii,t}|_{2\cap B\cap\gamma}, \qquad \text{(A1.19)}
\end{aligned}
$$

Considering which sub-stages two genes are sampled from, we can formulate the six $H_{ii,t}$ on the right side of equation A1.19 .

$$
\begin{aligned}
H_{ii,t}|_{2\cap A\cap\alpha}&=\sum_{m=1}^{n}\left\{\left(\frac{N_{i_{ms}}}{N_i}\right)^2H_{i_{ms}i_{ms},t}\right\}\\
&=\sum_{m=1}^{n}\left\{\left(\frac{t_{im}N_m}{N_i}\right)^2H_{i_{ms}i_{ms},t}\right\} \qquad \text{(A1.20)}\\
H_{ii,t}|_{2\cap A\cap\beta}&=\sum_{m=1}^{n}\left(\frac{N_{i_{ms}}N_{i_{mr}}}{N_i^2}H_{i_{ms}i_{mr},t}+\frac{N_{i_{mr}}N_{i_{ms}}}{N_i^2}H_{i_{mr}i_{ms},t}\right)\\
&=\sum_{m=1}^{n}\left(\frac{2N_{i_{ms}}N_{i_{mr}}}{N_i^2}H_{i_{ms}i_{mr},t}\right)\\
&=\sum_{m=1}^{n}\left(\frac{2t_{im}f_{im}N_m^2}{N_i^2}H_{i_{ms}i_{mr},t}\right) \qquad \text{(A1.21)}\\
H_{ii,t}|_{2\cap A\cap\gamma}&=\sum_{m=1}^{n}\left\{\left(\frac{N_{i_{mr}}}{N_i}\right)^2H_{i_{mr}i_{mr},t}\right\}\\
&=\sum_{m=1}^{n}\left\{\left(\frac{f_{im}N_m}{N_i}\right)^2H_{i_{mr}i_{mr},t}\right\} \qquad \text{(A1.22)}\\
H_{ii,t}|_{2\cap B\cap\alpha}&=\sum_{k=1}^{n}\sum_{l>k}^{n}\left(\frac{N_{i_{ks}}N_{i_{ls}}}{N_i^2}H_{i_{ks}j_{ls},t}+\frac{N_{i_{ls}}N_{i_{ks}}}{N_i^2}H_{i_{ls}j_{ks},t}\right)\\
&=\sum_{k=1}^{n}\sum_{l>k}^{n}\left(\frac{2N_{i_{ls}}N_{i_{ks}}}{N_i^2}H_{i_{ls}j_{ks},t}\right)\\
&=\sum_{k=1}^{n}\sum_{l>k}^{n}\left(\frac{2t_{ik}t_{il}N_kN_l}{N_i^2}H_{i_{ls}j_{ks},t}\right) \qquad \text{(A1.23)}
\end{aligned}
$$

$$
\begin{aligned}
H_{ij,t}|_{2\cap B\cap\beta} &= \sum_{k=1}^{n}\sum_{l>k}^{n}\left(\frac{N_{i_{ls}}N_{i_{kr}}}{N_i^2}H_{i_{kr}i_{ls},t}+\frac{N_{i_{kr}}N_{i_{ls}}}{N_i^2}H_{i_{kr}i_{ls},t}\right.\\
&\qquad\left.+\frac{N_{i_{ks}}N_{i_{lr}}}{N_i^2}H_{i_{ks}i_{lr},t}+\frac{N_{i_{lr}}N_{i_{ks}}}{N_i^2}H_{i_{ks}i_{lr},t}\right)\\
&= \sum_{k=1}^{n}\sum_{l>k}^{n}\left(\frac{2N_{i_{kr}}N_{i_{ls}}}{N_i^2}H_{i_{kr}i_{ls},t}+\frac{2N_{i_{ks}}N_{i_{lr}}}{N_i^2}H_{i_{kr}i_{ls},t}\right)\\
&= \sum_{k=1}^{n}\sum_{l>k}^{n}\left(\frac{2f_{ik}t_{il}N_kN_l}{N_i^2}H_{i_{kr}i_{ls},t}+\frac{2t_{ik}f_{il}N_kN_l}{N_i^2}H_{i_{ks}i_{lr},t}\right) \qquad \text{(A1.24)}\\
H_{ii,t}|_{2\cap B\cap\gamma} &= \sum_{k=1}^{n}\sum_{l>k}^{n}\left(\frac{N_{i_{kr}}N_{i_{lr}}}{N_i^2}H_{i_{kr}j_{lr},t}+\frac{N_{i_{lr}}N_{i_{kr}}}{N_i^2}H_{i_{lr}j_{kr},t}\right)\\
&= \sum_{k=1}^{n}\sum_{l>k}^{n}\left(\frac{2N_{i_{lr}}N_{i_{kr}}}{N_i^2}H_{i_{lr}j_{kr},t}\right)\\
&= \sum_{k=1}^{n}\sum_{l>k}^{n}\left(\frac{2f_{ik}f_{il}N_kN_l}{N_i^2}H_{i_{lr}j_{kr},t}\right) \qquad \text{(A1.25)}
\end{aligned}
$$

As in the case of $H_{i_{ms}i_{ms},t}$, two genes are sampled from sub-stage i_{ms} with replacement. Because all genes in sub-stage i_{ms} were transferred by survival from stage m, sub-stage i_{ms} consist of genes that were randomly sampled $2t_{im}N_m$ times 'without' replacement from stage m. Therefore, the probability of sampling two genes that are non-identical-by-descent without replacement should remain the same between stage m in year $t-1$ and sub-stage i_{ms} in year t (i.e., $H'_{mm,t-1} = H'_{i_{ms}i_{ms},t}$). As with equation A1.5 , $H_{i_{ms}i_{ms},t}$ is formulated as follows.

$$
H_{i_{ms}i_{ms},t} = \frac{1}{2t_{im}N_m}\times 0 + \left(1-\frac{1}{2t_{im}N_m}\right)\times H'_{i_{ms}i_{ms},t}, \qquad \text{(A1.26)}
$$

From equations A1.5 and A1.26 ,

$$
\begin{aligned}
H'_{i_{ms}i_{ms},t} &= H'_{mm,t-1}\\
H_{i_{ms}i_{ms},t} &= \frac{1-1/(2t_{im}N_m)}{1-1/(2N_m)}H_{mm,t-1}. \qquad \text{(A1.27)}
\end{aligned}
$$

It should be noted that $H_{i_{ms}i_{ms},t}$ should not be equal to $H_m m, t-1$. Two genes are always sampled from a common subset of stage m (i.e, sub-stage i_{ms}), which means that two genes are not sampled from separate and independent surrogates of stage m of the previous year. Therefore, sampling two genes from sub-stage i_{ms} with replacement is not equivalent to that from stage m with replacement. Unlike $H_{i_{ms}i_{ms},t}$, $H_{i_{ms}i_{mr},t}$ should be equal to $H_{mm,t-1}$, because sub-stage $i_m s$ and $i_m r$, from which two genes are sampled, are independently formed from stage m.

$$
H_{i_{ms}i_{mr},t} = H_{mm,t-1} \qquad \text{(A1.28)}
$$

In the case of $H_{i_{mr}i_{mr},t}$, the sources of two genes sampled are the same (i.e., sub-stage i_{mr}) and thus are not independent surrogates of stage m of the previous year, as with the case of $H_{i_{ms}i_{ms},t}$. Sampling the same gene twice occurs with the probability of $\frac{1}{2f_{im}N_m}$, which makes no contribution to $H_{i_{mr}i_{mr},t}$. In the remaining conditions where the two genes are sampled without replacement, the two genes are not identical-by-descent with a chance of $H_{mm,t-1}$ because sub-stage i_{mr} were formed by reproduction, or sampling with replacement. Therefore, $H_{mm,t-1}$ is discounted by the fraction of $\frac{1}{2f_{im}N_m}$.

$$H_{i_{mr}i_{mr}} = \left(1 - \frac{1}{2f_{im}N_m}\right) H_{mm,t-1} \tag{A1.29}$$

In the case of $H_{i_{ks}i_{ls},t}$, $H_{i_{ks}i_{lr},t}$, $H_{i_{kr}i_{ls},t}$, and $H_{i_{kr}i_{lr},t}$, two genes are sampled from independent subset or copy of stage k and l of the previous year. Therefore,

$$H_{i_{ks}i_{ls},t} = H_{i_{ks}i_{lr},t} = H_{i_{kr}i_{ls},t} = H_{i_{kr}i_{lr},t} = H_{kl,t-1} \tag{A1.30}$$

Substituting equations A1.27 -A1.30 to equations A1.20 -A1.25 ,

$$\begin{aligned}
H_{ii,t}|_{2\cap A\cap\alpha} &= \sum_{m=1}^{n}\left\{\left(\frac{t_{im}N_m}{N_i}\right)^2 \times \frac{1-1/(2t_{im}N_m)}{1-1/(2N_m)} H_{mm,t-1}\right\} \\
H_{ii,t}|_{2\cap A\cap\beta} &= \sum_{m=1}^{n}\left(\frac{2t_{im}f_{im}N_m^2}{N_i^2} \times H_{mm,t-1}\right) \\
H_{ii,t}|_{2\cap A\cap\gamma} &= \sum_{m=1}^{n}\left\{\left(\frac{f_{im}N_m}{N_i}\right)^2 \times \left(1 - \frac{1}{2f_{im}N_m}\right) H_{mm,t-1}\right\} \\
H_{ii,t}|_{2\cap B\cap\alpha} &= \sum_{k=1}^{n}\sum_{l>k}^{n}\left(\frac{2t_{ik}t_{il}N_kN_l}{N_i^2} \times H_{kl,t-1}\right) \\
H_{ii,t}|_{2\cap B\cap\beta} &= \sum_{k=1}^{n}\sum_{l>k}^{n}\left\{\frac{2(t_{ik}f_{il}+f_{ik}t_{il})N_kN_l}{N_i^2} \times H_{kl,t-1}\right\} \\
H_{ii,t}|_{2\cap B\cap\gamma} &= \sum_{k=1}^{n}\sum_{l>k}^{n}\left(\frac{2f_{ik}f_{il}N_kN_l}{N_i^2} \times H_{kl,t-1}\right).
\end{aligned} \tag{A1.31}$$

Substituting equations A1.31 to A1.19 , we can formulate $H_{ii,t}$.

$$\begin{aligned}
H_{ii,t} = \sum_{m=1}^{n}\Bigg\{&\left(\frac{t_{im}N_m}{N_i}\right)^2 \frac{1-1/(2t_{im}N_m)}{1-1/(2N_m)} + \frac{2t_{im}f_{im}N_m^2}{N_i^2} \\
&+ \left(\frac{f_{im}N_m}{N_i}\right)^2 \left(1 - \frac{1}{2f_{im}N_m}\right)\Bigg\} H_{mm,t-1} \\
&+ \sum_{k=1}^{n}\sum_{l>k}^{n}\frac{2(t_{ik}t_{il}+t_{ik}f_{il}+f_{ik}t_{il}+f_{ik}f_{il})N_kN_l}{N_i^2} H_{kl,t-1}
\end{aligned}$$

$$= \sum_{m=1}^{n} \left\{ \left(\frac{t_{im} N_m}{N_i} \right)^2 \frac{1 - 1/(2t_{im}N_m)}{1 - 1/(2N_m)} + \frac{2t_{im} f_{im} N_m^2}{N_i^2} + \left(\frac{f_{im} N_m}{N_i} \right)^2 \left(1 - \frac{1}{2f_{im}N_m} \right) \right\} H_{mm,t-1} + \sum_{k=1}^{n} \sum_{l>k}^{n} \frac{2a_{ik}a_{il}N_k N_l}{N_i^2} H_{kl,t-1}. \quad \text{(A1.32)}$$

A1.1.3 Definition of generation time T

We use generation time T to formulate effective population size N_e. Here, we explain the definition of generation time.

Firstly, we decompose the population projection matrix into two: $\boldsymbol{U}$ matrix, which is made up of t_{ij} and describes the survival process, and $\boldsymbol{F}$ matrix, which is made up of stage-specific fecundity f_{ij}. In the case of the two-stage model,

$$\begin{pmatrix} t_{11} & t_{12} + f_{12} \\ t_{21} & t_{22} \end{pmatrix} = \begin{pmatrix} t_{11} & t_{12} \\ t_{21} & t_{22} \end{pmatrix} + \begin{pmatrix} 0 & f_{12} \\ 0 & 0 \end{pmatrix} = \boldsymbol{U} + \boldsymbol{F}. \quad \text{(A1.33)}$$

In the case of the three-stage model,

$$\begin{pmatrix} t_{11} & 0 & f_{13} \\ t_{21} & t_{22} & t_{23} \\ 0 & t_{32} & t_{33} \end{pmatrix} = \begin{pmatrix} t_{11} & 0 & 0 \\ t_{21} & t_{22} & t_{23} \\ 0 & t_{32} & t_{33} \end{pmatrix} + \begin{pmatrix} 0 & 0 & f_{13} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \boldsymbol{U} + \boldsymbol{F}. \quad \text{(A1.34)}$$

By multiplying $\boldsymbol{U}$ matrix x times, we can obtain transition probabilities per x years. In the case of the two-stage model,

$$\boldsymbol{U}^x = \begin{pmatrix} t_{11} & t_{12} \\ t_{21} & t_{22} \end{pmatrix}^x = \begin{pmatrix} \tilde{u}_{11} & \tilde{u}_{12} \\ \tilde{u}_{21} & \tilde{u}_{22} \end{pmatrix}. \quad \text{(A1.35)}$$

Here, $\tilde{u}_{11}$ and $\tilde{u}_{21}$ are the probabilities that an individual in stage 1 remain in stage 1, or move to stage 2, after x years, respectively. Now, we can formulate age-specific survival rate l_x, which denotes the probability of a newborn individual to survive until age x, and age-specific fecundity m_x, which is a expected number of newborns that an individual of age x can make.

$$l_x = \tilde{u}_{11} + \tilde{u}_{21} \quad \text{(A1.36)}$$

$$m_x = 0 \times \frac{\tilde{u}_{11}}{\tilde{u}_{11} + \tilde{u}_{21}} + f_{12} \times \frac{\tilde{u}_{21}}{\tilde{u}_{11} + \tilde{u}_{21}}. \quad \text{(A1.37)}$$

In the case of the three-stage,

$$\boldsymbol{U}^x = \begin{pmatrix} t_{11} & 0 & 0 \\ t_{21} & t_{22} & t_{23} \\ 0 & t_{32} & t_{33} \end{pmatrix}^x = \begin{pmatrix} \tilde{u}_{11} & \tilde{u}_{12} & \tilde{u}_{13} \\ \tilde{u}_{21} & \tilde{u}_{22} & \tilde{u}_{23} \\ \tilde{u}_{31} & \tilde{u}_{32} & \tilde{u}_{33} \end{pmatrix} \tag{A1.38}$$

$$l_x = \tilde{u}_{11} + \tilde{u}_{21} + \tilde{u}_{31} \tag{A1.39}$$

$$m_x = f_{13} \times \frac{\tilde{u}_{31}}{\tilde{u}_{11} + \tilde{u}_{21} + \tilde{u}_{31}}. \tag{A1.40}$$

Then, we formulate generation time (T) as the expected age of a parent of a cohort.

$$T = \frac{\sum_{x=1}^{x_{max}} x l_x m_x}{\sum_{x=1}^{x_{max}} l_x m_x}, \tag{A1.41}$$

where x_{max} is the maximum age defined as the age at which either of the two criteria (quoted from Waples et al. (2013)) is satisfied.

1. oldest age for which l_x was ≥ 1 % of the value at age at maturity (L_α)
2. oldest age for which the product $l_x v_x$ was ≥ 1 % of the maximum $l_x v_x$ for any age, where v_x is the reproductive value of an individual of age x

Equation A1.41 is exactly the mean age of net fecundity in the cohort (Carey and Roach 2020).

A1.2 How to determine parameter values

We randomly produced 500 parameter sets for each of the two- and the three-stage model for simulation and model analysis. Here, we explain how we determined the values of each parameter (demographic rates and the number of individuals in each stage).

Step 1 We draw four random numbers from the uniform distribution $U(0, 1)$ for the two-stage model. In the case of the three-stage model, six random numbers are drawn from the same uniform distribution.

Step 2 We rearrange the random numbers in an increasing order.

Step 3 We multiply the random numbers by 100, and round them off to be integers. Moreover, we add 0 and 100 to the sequences.

Step 4 We take the difference between the neighboring numbers: we subtract each number from its next smaller one. As a result, five and seven numbers are generated for the two-stage and the three-stage models, respectively.

Step 5 Each number is assigned to one of the demographic processes (i.e., growth, stasis, retrogression, and reproduction) of each stage. As for the two-stage model, the first

to fifth numbers are assigned to (1) stasis at juvenile, (2) growth from juvenile to adult, (3) retrogression from adult to juvenile, (4) stasis at adult, and (5) reproduction, respectively. In the case of the three-stage model, seven numbers are sequentially assigned to (1) stasis at seed, (2) growth from seed to juvenile, (3) stasis at juvenile, (4) growth from juvenile to adult, (5) retrogression from adult to juvenile, (6) stasis at adult, and (7) reproduction.

Step 6 We calculate the number of individuals of each stage as the sum of flows coming into each stage.

Step 7 We calculate demographic rates by dividing the number of individuals of corresponding flows by that of the stages from which the flows come out.

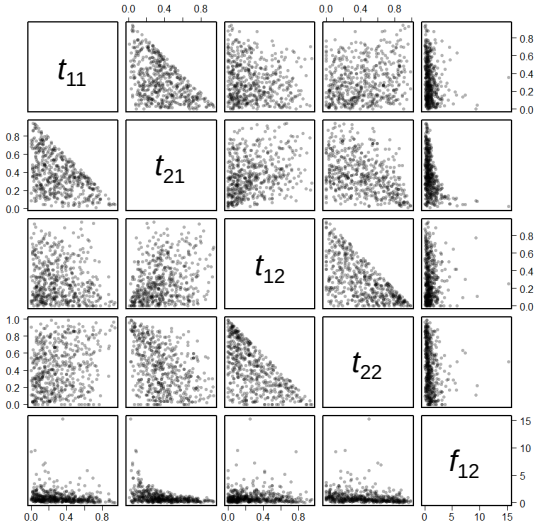
Step 8 We assess if the parameter values calculated in step 7 completely satisfy the following three criteria. If they do, the values are added to the parameter sets. If not, the values are discarded and we restart the procedures from Step 1.

1. Growth probability and fecundity should be greater than 0, otherwise the life cycle would be broken off.
2. Survival probability of each stage (i.e., sum of growth, stasis, and retrogression probabilities of each stage) is within the range of $[0, 1]$ to be a probability.
3. At least one stasis probability is not zero, otherwise genes in different stages do not mix with one another and are segregated eternally.

Step 9 We repeated Step 1 to 8 until the number of parameter sets reached 500.

The resultant parameter sets range the parameter space widely both for the two-stage and for the three-stage models, indicating that our model is examined for a variety of demographic strategies (Figure A1).

(a) Two-stage model



(b) Three-stage model

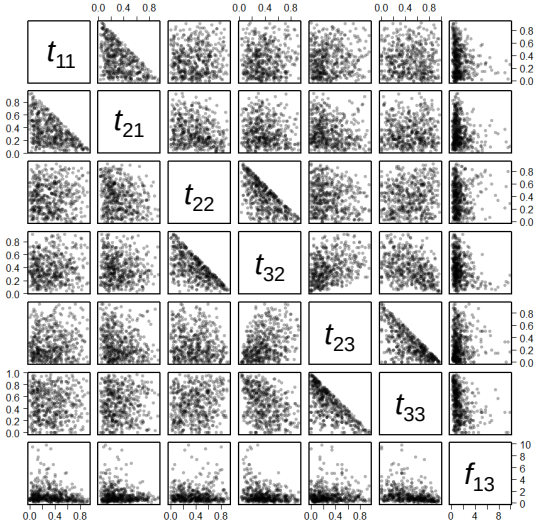


Figure A1: The 500 parameter sets used in simulation and model analysis. One dot corresponds to one parameter set. (a) Two-stage model, (b) three-stage model. There are some parameter pairs where the dots occupy only the lower-left triangle (e.g., t_{11} and t_{21} in (a); t_{22} and t_{32} in (b)). This is because of the second criterion in step 8, that is, their sum should not exceed 1

A1.3 Additional results

A1.3.1 Validation of our model

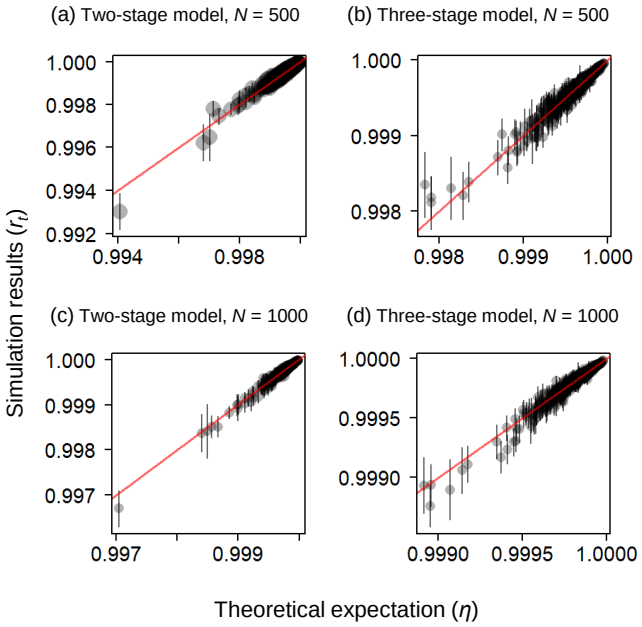


Figure A2: Comparison between the theoretical expectation of the annual change rate of the probability of non-identical-by-descent (η) and the simulation results of that of expected heterozygosity (r_t) when N =500 and 1000 for both the two- and the three-stage model. (a) Two-stage model, $N = 500$, (b) three-stage model, $N = 500$, (c) two-stage model, $N = 1000$, (d) three-stage model, $N = 1000$. Vertical bars represent standard error of r_t. The red lines indicate $r_t = \eta$

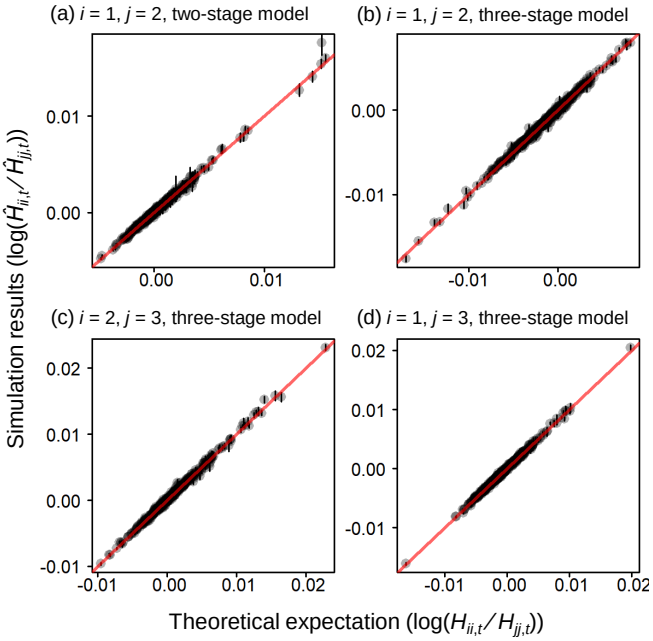


Figure A3: Comparison of demographic genetic structure between theoretical expectation ($\log(H_{ii,t}/H_{jj,t})$) and simulation results ($\log(\hat{H}_{ii,t}/\hat{H}_{jj,t})$) when N =500 for both the two- and the three-stage model. (a) $i = 1$, $j = 2$, two-stage model, (b)$i = 1$, $j = 2$, three-stage model, (c) $i = 2$, $j = 3$, three-stage model, (d) $i = 1$, $j = 3$, three-stage model. Vertical bars represent standard error of r_t. The red lines indicate $\log(H_{ii,t}/H_{jj,t})$=$\log(\hat{H}_{ii,t}/\hat{H}_{jj,t})$

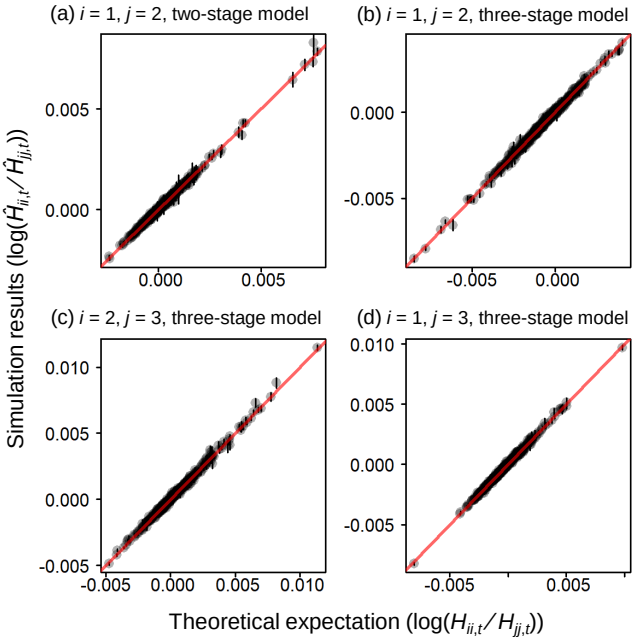


Figure A4: Comparison of demographic genetic structure between theoretical expectation ($\log(H_{ii,t}/H_{jj,t})$) and simulation results ($\log(\hat{H}_{ii,t}/\hat{H}_{jj,t})$) when N =1000 for both the two- and the three-stage model. (a) $i = 1$, $j = 2$, two-stage model, (b)$i = 1$, $j = 2$, three-stage model, (c) $i = 2$, $j = 3$, three-stage model, (d) $i = 1$, $j = 3$, three-stage model. Vertical bars represent standard error of r_t. The red lines indicate $\log(H_{ii,t}/H_{jj,t})$=$\log(\hat{H}_{ii,t}/\hat{H}_{jj,t})$

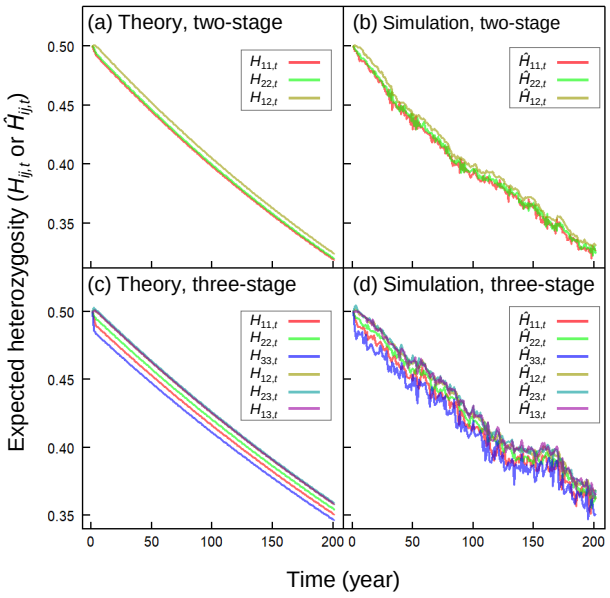


Figure A5: Graphical comparison of temporal dynamics of expected heterozygosity between theoretical expectation (i.e., dynamics of $H_{ij,t}$) and simulation results (i.e., dynamics of $\hat{H}_{ij,t}$) under a particular parameter set in the two- and the three-stage model. Colors stand for combinations of i and j. (a) Theoretical expectations and (b) simulation results of the two-stage model, (c) theoretical expectations and (d) simulation results of the three-stage model. Parameter set of the two-stage model is $t_{11} = 0.115, t_{21} = 0.750, t_{12} = 0.333, t_{22} = 0.188, f_{12} = 0.625, N_1 = 52, N_2 = 48$, while that of the three-stage model is $t_{11} = 0.476, t_{21} = 0.405, t_{22} = 0.568, t_{32} = 0.273, t_{23} = 0.143, t_{33} = 0.143, f_{13} = 1.57, N_1 = 42, N_2 = 44, N_3 = 14$. Each parameter set was randomly picked from the 500 parameter sets under $N = 100$ as an example case

A1.3.2 Comparison of demographic genetic structure with N_e and η

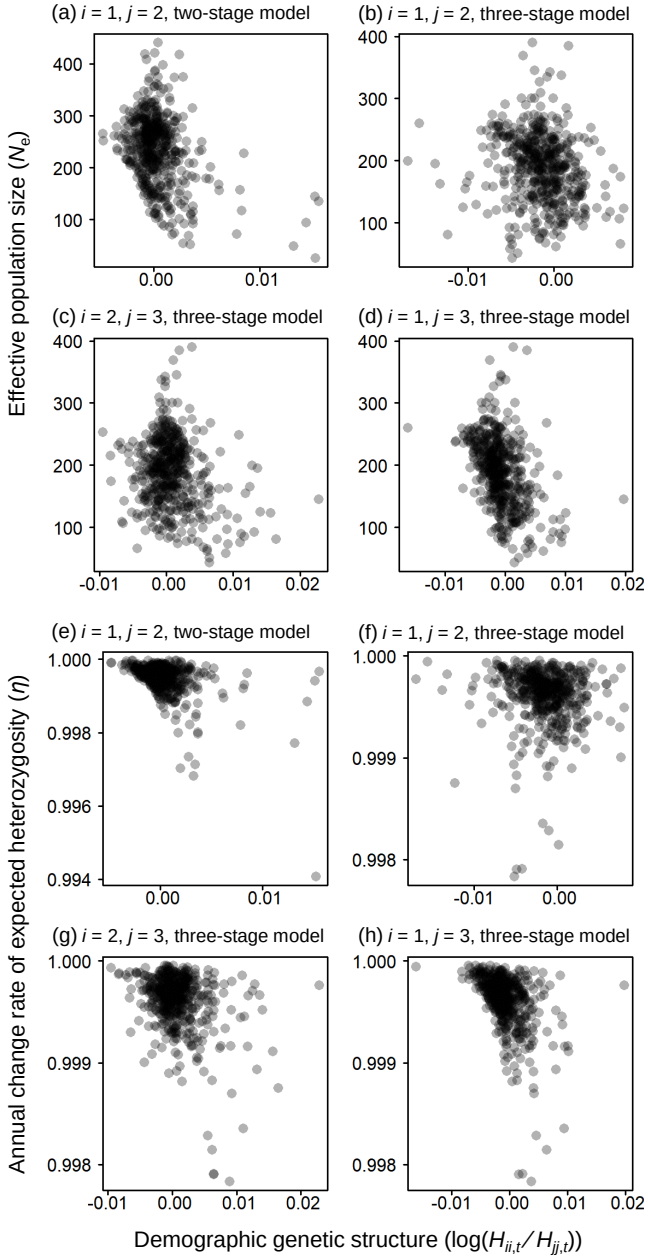


Figure A6: Comparison of demographic genetic structure ($\log(H_{ii,t}/H_{jj,t})$) with effective population size N_e (a-d) and the annual change rate of expected heterozygosity η (e-h) when $N = 500$. (a, e) $i = 1$ and $j = 2$ of the two-stage model, (b, f) $i = 1$ and $j = 2$, (c, g) $i = 2$ and $j = 3$, (d, h) $i = 1$ and $j = 3$ of the three-stage model

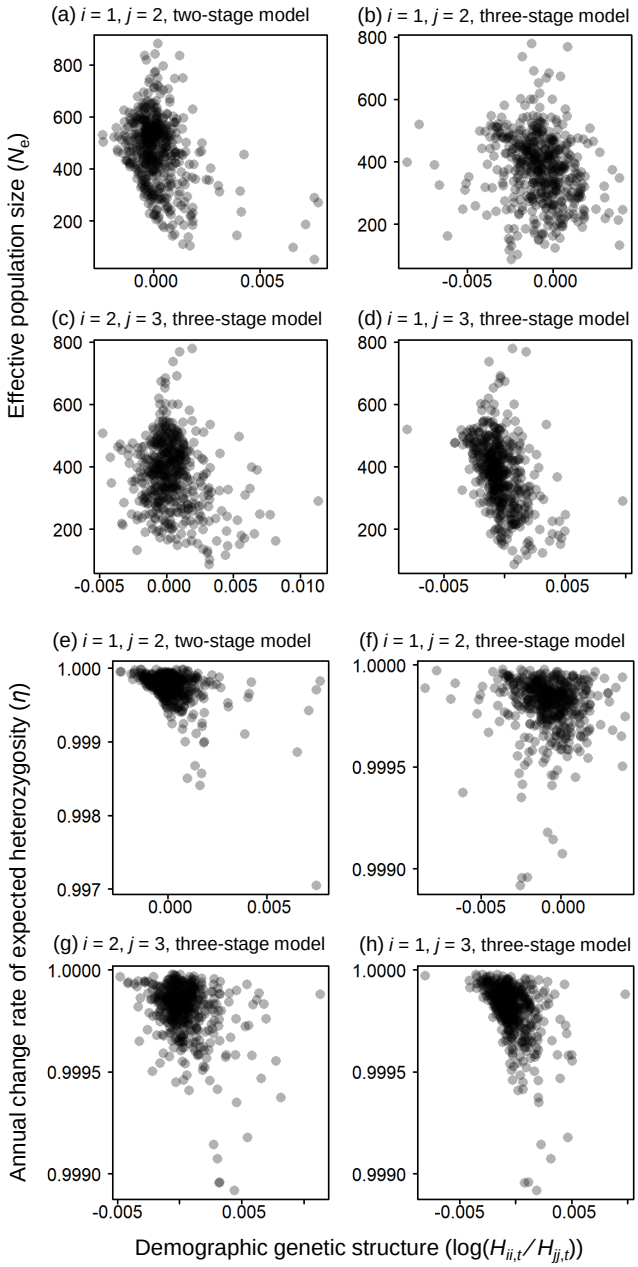


Figure A7: Comparison of demographic genetic structure ($\log(H_{ii,t}/H_{jj,t})$) with effective population size N_e (a-d) and the annual change rate of expected heterozygosity η (e-h) when $N = 1000$. (a, e) $i = 1$ and $j = 2$ of the two-stage model, (b, f) $i = 1$ and $j = 2$, (c, g) $i = 2$ and $j = 3$, (d, h) $i = 1$ and $j = 3$ of the three-stage model

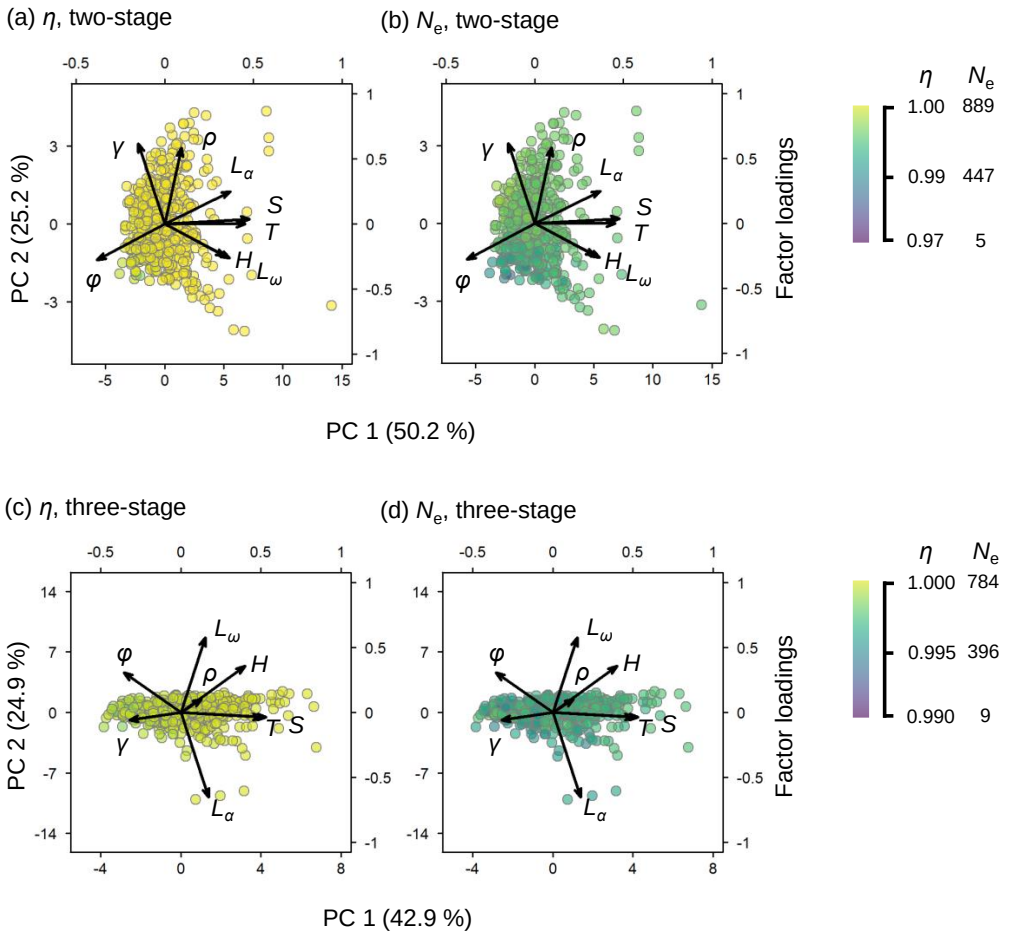


Figure A8: Principal components of the eight life history traits of the 500 parameter sets in the two- and the three stage model. Color denotes either annual change rate of expected heterozygosity (η) or effective population size (N_e) when $N = 500$. (a) η of the two-stage model, (b) N_e of the two-stage model, (c) η of the three-stage model, (d) N_e of the three-stage model

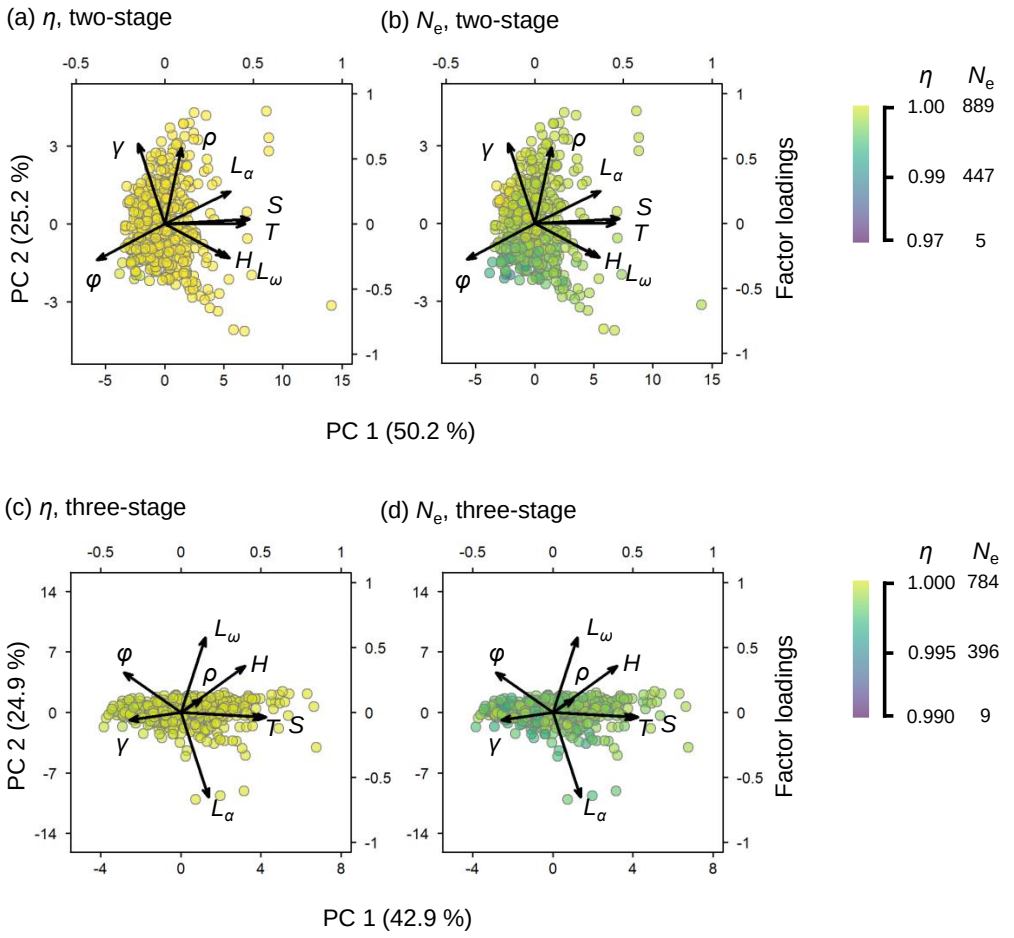


Figure A9: Principal components of the eight life history traits of the 500 parameter sets in the two- and the three stage model. Color denotes either annual change rate of expected heterozygosity (η) or effective population size (N_e) when $N = 1000$. (a) η of the two-stage model, (b) N_e of the two-stage model, (c) η of the three-stage model, (d) N_e of the three-stage model

Chapter A2

Applying the model to *Trillium camschatcense*

A2.1 matrix M

Life history of *Trillium camschatcense* consists of 5 stages, which are seed (stage 0), seedling (stage 1), one-leaf (stage 2), three-leaves (stage 3), and flowering (stage 4). Here, I define the order of $H_{ij,t}$ in $\boldsymbol{h_t}$ as follows.

$$\boldsymbol{h_t} = \begin{pmatrix} H_{00,t} \\ H_{11,t} \\ H_{22,t} \\ H_{33,t} \\ H_{44,t} \\ H_{01,t} \\ H_{02,t} \\ H_{03,t} \\ H_{04,t} \\ H_{12,t} \\ H_{13,t} \\ H_{14,t} \\ H_{23,t} \\ H_{24,t} \\ H_{34,t} \end{pmatrix}. \tag{A2.1}$$

Quoting Equation (2.10), the difference equation of $\boldsymbol{h_t}$ can be rewritten as follows.

$$\boldsymbol{h_{t+1}} = \boldsymbol{M}\boldsymbol{h_t}, \tag{A2.2}$$

where $\boldsymbol{M}$ is the 15×15 square matrix. The detailed elements of $\boldsymbol{M}$ are shown in Appendix A2.2. For sufficiently large t, $\boldsymbol{h_t}$ converges to the scalar multiplication of $\boldsymbol{w}$, which is the leading right eigenvector of $\boldsymbol{M}$. Therefore, I can obtain the relative ratio among $H_{ij,t}$ by calculating $\boldsymbol{w}$.

For a given parameter set, I calculated $\boldsymbol{w}$ by solving the eigenvalue problem of the matrix $\boldsymbol{M}$. Here, I denote w_{ij} as the element of $\boldsymbol{w}$ corresponding to $H_{ij,t}$. Because seed (stage 0) was not sampled and genotyped, I omitted w_{00}, w_{01}, w_{02}, w_{03}, and w_{04} from $\boldsymbol{w}$ and divided the remaining elements by their sum for normalization. Similarly, the observed H_{ij} ($1 \leq i \leq 4$, $i \leq j \leq 4$) were normalized. Then, the squared errors between the

corresponding elements of the two normalized vectors were calculated and summed, which was used as the distance metric between the observed and expected H_{ij}.

A2.2 Elements of the matrix *M*

$m_{i,j}$, which is the element of $\boldsymbol{M}$ in the i-th row and the j-th column, is zero except the following 53 elements.

$$m_{1,5} = \left(\frac{fN_4}{N_0}\right)^2 \left(1 - \frac{1}{2fN_4}\right), \tag{A2.3}$$

$$m_{2,1} = \left(\frac{g_0N_0}{N_1}\right)^2 \frac{1 - 1/(2g_0N_0)}{1 - 1/(2N_0)}, \tag{A2.4}$$

$$m_{3,2} = \left(\frac{g_1N_1}{N_2}\right)^2 \frac{1 - 1/(2g_1N_1)}{1 - 1/(2N_1)}, \tag{A2.5}$$

$$m_{3,3} = s_2^2 \frac{1 - 1/(2s_2N_2)}{1 - 1/(2N_2)}, \tag{A2.6}$$

$$m_{3,4} = \left(\frac{r_3N_3}{N_2}\right)^2 \frac{1 - 1/(2r_3N_3)}{1 - 1/(2N_3)}, \tag{A2.7}$$

$$m_{3,10} = \frac{2g_1s_2N_1}{N_2}, \tag{A2.8}$$

$$m_{3,11} = \frac{2g_1r_3N_1N_3}{N_2^2}, \tag{A2.9}$$

$$m_{3,13} = \frac{2s_2r_3N_3}{N_2}, \tag{A2.10}$$

$$m_{4,3} = \left(\frac{g_2N_2}{N_3}\right)^2 \frac{1 - 1/(2g_2N_2)}{1 - 1/(2N_2)}, \tag{A2.11}$$

$$m_{4,4} = s_3^2 \frac{1 - 1/(2s_3N_3)}{1 - 1/(2N_3)}, \tag{A2.12}$$

$$m_{4,5} = \left(\frac{r_4N_4}{N_3}\right)^2 \frac{1 - 1/(2r_4N_4)}{1 - 1/(2N_4)}, \tag{A2.13}$$

$$m_{4,13} = \frac{2g_2s_3N_2}{N_3}, \tag{A2.14}$$

$$m_{4,14} = \frac{2g_2r_4N_2N_4}{N_3^2}, \tag{A2.15}$$

$$m_{4,15} = \frac{2s_3r_4N_4}{N_3}, \tag{A2.16}$$

$$m_{5,4} = \left(\frac{g_3N_3}{N_4}\right)^2 \frac{1 - 1/(2g_3N_3)}{1 - 1/(2N_3)}, \tag{A2.17}$$

$$m_{5,5} = s_4^2 \frac{1 - 1/(2s_4N_4)}{1 - 1/(2N_4)}, \tag{A2.18}$$

$$m_{5,15} = \frac{2g_3s_4N_3}{N_4}, \tag{A2.19}$$

$$m_{6,9} = 1, \tag{A2.20}$$

$$m_{7,12} = \frac{g_1 N_1}{N_2}, \tag{A2.21}$$

$$m_{7,14} = s_2, \tag{A2.22}$$

$$m_{7,15} = \frac{r_3 N_3}{N_2}, \tag{A2.23}$$

$$m_{8,5} = \frac{r_4 N_4}{N_3}, \tag{A2.24}$$

$$m_{8,14} = \frac{g_2 N_2}{N_3}, \tag{A2.25}$$

$$m_{8,15} = s_3, \tag{A2.26}$$

$$m_{9,5} = s_4, \tag{A2.27}$$

$$m_{9,15} = \frac{g_3 N_3}{N_4}, \tag{A2.28}$$

$$m_{10,6} = \frac{g_1 N_1}{N_2}, \tag{A2.29}$$

$$m_{10,7} = s_2, \tag{A2.30}$$

$$m_{10,8} = \frac{r_3 N_3}{N_2}, \tag{A2.31}$$

$$m_{11,7} = \frac{g_2 N_2}{N_3}, \tag{A2.32}$$

$$m_{11,8} = s_3, \tag{A2.33}$$

$$m_{11,9} = \frac{r_4 N_4}{N_3}, \tag{A2.34}$$

$$m_{12,8} = 1 - s_4, \tag{A2.35}$$

$$m_{12,9} = s_4, \tag{A2.36}$$

$$m_{13,3} = \frac{N_2}{N_3} \frac{s_2 g_2}{1 - 1/(2N_2)}, \tag{A2.37}$$

$$m_{13,4} = \frac{N_3}{N_2} \frac{r_3 s_3}{1 - 1/(2N_3)}, \tag{A2.38}$$

$$m_{13,10} = \frac{g_1 g_2 N_1}{N_3}, \tag{A2.39}$$

$$m_{13,11} = \frac{g_1 s_3 N_1}{N_2}, \tag{A2.40}$$

$$m_{13,12} = \frac{g_1 r_4 N_1 N_4}{N_2 N_3}, \tag{A2.41}$$

$$m_{13,13} = s_2 s_3 + g_2 r_3, \tag{A2.42}$$

$$m_{13,14} = \frac{s_2 r_4 N_4}{N_3}, \tag{A2.43}$$

$$m_{13,15} = \frac{r_3 r_4 N_4}{N_2}, \tag{A2.44}$$

$$m_{14,4} = \frac{N_3^2}{N_2 N_4} \frac{r_3 g_3}{1 - 1/(2N_3)}, \tag{A2.45}$$

$$m_{14,11} = \frac{g_1(1 - s_4)N_1}{N_2}, \tag{A2.46}$$

$$m_{14,12} = \frac{g_1 s_4 N_1}{N_2}, \tag{A2.47}$$

$$m_{14,13} = s_2(1 - s_4), \tag{A2.48}$$

$$m_{14,14} = s_2 s_4, \tag{A2.49}$$

$$m_{14,15} = \frac{r_3 s_4 N_3}{N_2}, \tag{A2.50}$$

$$m_{15,4} = \frac{s_3(1 - s_4)}{1 - 1/(2N_3)}, \tag{A2.51}$$

$$m_{15,5} = \frac{N_4}{N_3} \frac{r_4 s_4}{1 - 1/(2N_4)}, \tag{A2.52}$$

$$m_{15,13} = \frac{g_2 g_3 N_2}{N_4}, \tag{A2.53}$$

$$m_{15,14} = \frac{g_2 s_4 N_2}{N_3}, \tag{A2.54}$$

$$m_{15,15} = s_3 s_4 + r_4 g_3. \tag{A2.55}$$