Potential effects of life history on demographic genetic structure in stage-structured plant populations

Yoichi Tsuzuki[†], Takenori Takada[†], Masashi Ohara[†]

† Graduate School of Environmental Science, Hokkaido University, Sapporo City, Hokkaido Prefecture, Japan, 060-0810

running title: demographic genetic structure

key words: expected heterozygosity, genetic differentiation, genetic diversity, non-identical-by-descent, stable stage distribution, temporal dynamics

corresponding author: Yoichi Tsuzuki
N10W5, Kita-ku, Sapporo City, Hokkaido Prefecture, Japan, 060-0810
phone number: (+81)-011-706-2225
email address: yoichi.tsuzuki.95@gmail.com

Abstract

Standing genetic variation, or genetic diversity, is a source of adaptive evolution, and is crucial for long-term population persistence under environmental changes. One empirical method to predict the temporal dynamics of standing genetic variation in age- or stage-structured populations is to compare genetic diversity and composition among age/stage classes. The resultant within-population genetic structure, sometimes referred to as demographic genetic structure, has been regarded as a proxy of potential genetic changes that accompany sequential generation turnover. However, especially in stage-structured plant populations, individuals in more juvenile stages do not necessarily represent future populations, as they might die, stop growing, or retrogress over the course of life history. How demographic genetic structure is subjected to life history and whether it is a good proxy of temporal genetic dynamics had remained unclear. Here, we developed a matrix model which well describes temporal dynamics of expected heterozygosity, a common proxy of genetic diversity, for a neutral locus in stage-structured populations under equilibrium assumption. Based on the model, two indices of demographic genetic structure were formulated: relative ratio of expected heterozygosity and genetic differentiation among stage classes. We found that the two indices were largely determined by stable stage distribution and population size, and that they did not show clear correlations with the change rate of genetic diversity, indicating that inferring future genetic diversity from demographic genetic structure conventionally is misleading. Our study facilitates reliable interpretation on empirical demographic genetic data.

Standing genetic variation, or genetic diversity, is one of the factors crucial to population viability. Adaptation from standing genetic variation contributes to population persistence under environmental changes (Agashe *et al.* 2011; Barrett and Schluter 2008; Ramsayer and Hochberg 2012). Therefore, understanding the temporal dynamics of genetic diversity is of great necessity to predict ecological and evolutionary dynamics of populations.

The rate of change in genetic diversity per generation time is primarily determined by effective population size (N_e): the larger N_e is, the weaker genetic drift works, and the more likely genetic diversity is maintained (Crow and Kimura 1970). Effective population size is the size of an ideal population which fulfills several assumptions, such as constant population size, discrete generations, and no immigration and emigration. While these assumptions are advantageous in deriving simple mathematical predictions, many organisms, especially perennial plants, have complex life histories, taking multiple years from birth to death. In these species, populations consist of individuals in different age or stage classes, and generations are not discrete. It is thus necessary to construct methods to predict genetic dynamics of species with overlapping generations.

One solution is to mathematically derive effective population size for age- and stage-structured populations. There have been theoretical studies which formulated N_e using demographic rates (age- or stage-specific survival rates and fecundities) for life histories structured by age (Felsenstein 1971; Hill 1972; Hill 1979; Johnson 1977) or by stage (Orive 1993; Yonezawa *et al.* 2000). Based on these derivation, N_e has been estimated and studied in many "non-ideal" species (Waples *et al.* 2011; Waples *et al.* 2013).

Meanwhile, some empirical genetic studies took another approach in which genetic diversity and composition are comparatively estimated for each age or stage class at a single time point. The resultant within-population genetic structure (sometimes referred to as demographic genetic

structure) reflects potential genetic changes that accompany the replacement of mature individuals with juveniles (Mimura *et al*. 2017), and is used for inferring temporal genetic dynamics (Aldrich *et al*. 1998; Ally and Ritland 2006; Kitamura and Nakanishi 2021; Schmidt *et al*. 2018). For example, if juvenile stage is less diverse than mature stages, genetic diversity would expected to decrease over the course of generation turnover. Because species with age or stage structure are mostly long-lived and it is hard to observe actual temporal dynamics, demographic genetic structure is considered to be an instant and convenient proxy.

However, demographic genetic structure could be strongly subjected to life history, which might obscure any signs of temporal changes. Firstly, juveniles do not always stand for future populations, as some of them would die over the course of life cycle. Moreover, in stage-structured populations including most perennial plant species, individuals survive and reproduce according to physiological or developmental stage rather than to age (Ellis, 2012; Silvertown, 1987). In such species, individuals not only grow and proceed to more mature stages, but also remain in the same stage for successive periods (stasis) or reverse to more juvenile stages (retrogression), regardless of age. 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 prereproductive stage due to resource exhaustion (Knight 2004; Ohara *et al*. 2001; Tomimatsu and Ohara 2010). The static and bidirectional flows in life history complicate the order of individual turnover in a population, making it difficult to interpret demographic genetic structure as the potential temporal dynamics per se. Besides, considering that age distribution (fraction of population size occupied by each age class) is known to influence the fixation probability of neutral and selective mutants in age-structured populations (Li *et al*. 2016), stage distribution (fraction of population size occupied by each stage class) would also affect the fixation probability of alleles in stage-structured populations, potentially determining

genetic diversity and composition of each stage class. There is a need to evaluate these possible influences of life history and to judge whether demographic genetic structure can still serve as a good proxy for temporal changes.

In this study, we develop a matrix model which is made up of difference equations of the probability that two randomly-sampled genes are non-identical-by-descent for a neutral locus in stage-structured populations. The model well describes the temporal dynamics of expected heterozygosity, a commonly-used proxy of genetic diversity. Based on the model, relative ratio of expected heterozygosity among stage classes and inter-stage genetic differentiation are formulated, and their dependence on demographic rates, stage distribution, and population size are evaluated. We then assess if comparison of genetic diversity and composition among stage classes would be a good proxy of temporal trends under a variety of parameter values.

Model

Overview

Felsenstein (1971) derived inbreeding effective population size for an age-structured population, which was also described in Charesworth (2008). We partly follow mathematical formulation procedures in Felsenstein (1971) and Charesworth (2008) to develop difference equations that describe the dynamics of the probability of non-identical-by-descent for a closed, stage-structured population, supposing a diploid perennial plant species. We do not consider sex difference because most plants are hermaphrodite (Torices, Méndez, & Gómez, 2011). Census population size is set to N, which is divided into n life history stages $(N_1, N_2, \cdots, N_n)$. As in Felsenstein (1971), we assume an equilibrium state, where the census population size and stage distribution are constant over the course of time (i.e., the number of individuals in each stage is constant). The probability of transition (either of growth, stasis, or retrogression) from stage j

to stage i is t_{ij} per a unit of time (say, per year). In each time step, 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 flow of individuals from stage j to i in successive time intervals.

$$a_{ij} = t_{ij} + f_{ij} \tag{1}$$

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

$$\begin{pmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\ \vdots & & \vdots & & \vdots \\ a_{i1} & \cdots & a_{ij} & \cdots & a_{in} \\ \vdots & & \vdots & & \vdots \\ a_{n1} & \cdots & a_{nj} & \cdots & a_{nn} \end{pmatrix} \tag{2}$$

Stable stage distribution, which is the relative number of individuals among stages in the equilibrium state, is proportional to the leading right eigenvector of this matrix (Caswell, 2001).

We define $H_{ij,t}$ as the probability that two genes randomly sampled from stage i and j with replacement in time t are not identical-by-descent. $H_{ij,t}$ is equivalent to expected heterozygosity, which is a common proxy of genetic diversity in empirical studies. To derive difference equations of $H_{ij,t}$, we separately consider two situations, $i \neq j$ (case 1) and $i = j$ (case 2).

case 1. $i \neq j$

Two genes randomly sampled from stage i and j respectively in time t were either in the same stage (say, stage m, Figure 1, A; case 1-1), or in different stages (say, stage k and l, Figure 1, B

and C; case 1-2) one time step ago.

$$H_{ij,t} = H_{ij,t}|_{\text{case 1-1}} + H_{ij,t}|_{\text{case 1-2}} \tag{3}$$

In case 1-1, there are three exclusive possibilities about how the two genes that were in stage m in time $t-1$ split into stage i and j in time t: both genes transferred by survival (case 1-1-1), one by survival and the other by reproduction (case 1-1-2), and both by reproduction (case 1-1-3) (Figure 1, A). To efficiently handle these three cases, we define substage i_{ms} and i_{mr}, which consist of individuals transferred from stage m to i by survival and by reproduction, respectively. $H_{ij,t}|_{\text{case 1-1}}$ can be decomposed into $H_{i_{ms}j_{ms},t}$ (case 1-1-1), $H_{i_{ms}j_{mr},t}$, $H_{i_{mr}j_{ms},t}$ (case 1-1-2), and $H_{i_{mr}j_{mr},t}$ (case 1-1-3) weighted by the corresponding number of two-gene pairs.

$$H_{ij,t}|_{\text{case 1-1}} = \sum_m \left(\frac{N_{i_{ms}}N_{j_{ms}}}{N_iN_j} H_{i_{ms}j_{ms},t} + \frac{N_{i_{ms}}N_{j_{mr}}}{N_iN_j} H_{i_{ms}j_{mr},t} + \frac{N_{i_{mr}}N_{j_{ms}}}{N_iN_j} H_{i_{mr}j_{ms},t} + \frac{N_{i_{mr}}N_{j_{mr}}}{N_iN_j} H_{i_{mr}j_{mr},t} \right) \tag{4}$$

As for case 1-1-1, two genes, each sampled from stage i and j, belong to substage i_{ms} and j_{ms} with the chance of $N_{i_{ms}}/N_i \times N_{j_{ms}}/N_j$. Here, $N_{i_{ms}}$ and $N_{j_{ms}}$ each denotes the number of individuals in substage i_{ms} and j_{ms}. Thus the weighting coefficient for $H_{i_{ms}j_{ms},t}$ is $(N_{i_{ms}}N_{j_{ms}})/(N_iN_j)$, as shown in the first term of the right side of Equation 4. Case 1-1-2 (the second and the third term) and 1-1-3 (the fourth term) are similarly accounted for to formulate $H_{ij,t}|_{\text{case 1-1}}$.

Two genes, each sampled from substage i_{ms} and j_{ms}, should not be the same gene because one gene in stage m in the previous time step 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, which can be obtained

by dividing $H_{mm,t-1}$ by the chance of not sampling the same gene twice ($= 1 - 1/(2N_m)$).

$$H_{i_{ms}j_{ms},t} = \frac{H_{mm,t-1}}{1 - 1/(2N_m)} \tag{5}$$

Unlike case 1-1-1, the probabilities of non-identical-by-descent are exactly equal to $H_{mm,t-1}$ both in case 1-1-2 and in case 1-1-3, because the same gene can be sampled twice due to the copying during reproduction.

$$H_{i_{ms}j_{mr},t} = H_{i_{mr}j_{ms},t} = H_{i_{mr}j_{mr},t} = H_{mm,t-1} \tag{6}$$

The number of individuals in substages must satisfy the following equations.

$$N_{i_{ms}} = t_{im}N_m \tag{7}$$

$$N_{j_{ms}} = t_{jm}N_m \tag{8}$$

$$N_{i_{mr}} = f_{im}N_m \tag{9}$$

$$N_{j_{mr}} = f_{jm}N_m \tag{10}$$

Substituting Equation 5-10 to Equation 4, $H_{ij,t}$ can be formulated as follows:

$$H_{ij,t}|_{\text{case 1-1}} = \sum_m \frac{N_m^2}{N_i N_j} \left\{ \frac{t_{im}t_{jm}}{1 - 1/(2N_m)} + f_{im}t_{jm} + t_{im}f_{jm} + f_{im}f_{jm} \right\} H_{mm,t-1} \tag{11}$$

In case 1-2, two genes, each sampled from stage i and j in time t, were transferred from stage k and l, respectively (case 1-2-1, Figure 1, B), or from stage l and k, respectively (case 1-2-2, Figure 1, C). In either situation, the probability of non-identical-by-descent holds from $t - 1$, regardless of whether they were transferred by survival or reproduction. Therefore, $H_{ij,t}$ can be

formulated as the weighted sum of $H_{kl,t-1}$. Contrary to case 1-1, there is no need to distinguish survival and reproduction, and only the sum of transition probabilities and fecundities matters, as shown in File S1, Section 1.

$$H_{ij,t}|_{\text{case 1-2}} = \sum_{k,l} \left(\frac{a_{ik} a_{jl} N_k N_l}{N_i N_j} + \frac{a_{il} a_{jk} N_l N_k}{N_i N_j} \right) H_{kl,t-1} \tag{12}$$

Substituting Equation 11 and 12 to Equation 3, $H_{ij,t}$ is formulated as follows:

$$H_{ij,t} = \sum_{m} \frac{N_m^2}{N_i N_j} \left\{ \frac{t_{im} t_{jm}}{1 - 1/(2N_m)} + f_{im} t_{jm} + t_{im} f_{jm} + f_{im} f_{jm} \right\} H_{mm,t-1}$$

$$+ \sum_{k,l} \frac{N_k N_l}{N_i N_j} \left(a_{ik} a_{jl} + a_{il} a_{jk} \right) H_{kl,t-1} \tag{13}$$

case 2. $i = j$

In this case, we replace j to i. As in case 1, there are two exclusive cases that two genes in stage i were in the same stage m (case 2-1, Figure 1, D), or in different stages (case 2-2, Figure 1, E) one time step ago.

$$H_{ii,t} = H_{ii,t}|_{\text{case 2-1}} + H_{ii,t}|_{\text{case 2-2}} \tag{14}$$

$H_{ii,t}|_{\text{case 2-1}}$, ($H_{ii,t}$ under case 2-1) is the weighted sum of $H_{i_{ms} i_{ms},t}$, $H_{i_{ms} i_{mr},t}$, and $H_{i_{mr} i_{mr},t}$, as previously shown in Equation 4.

$$H_{ii,t}|_{\text{case 2-1}} = \sum_{m} \left(\frac{N_{i_{ms}}^2}{N_i^2} H_{i_{ms} i_{ms},t} + \frac{2 N_{i_{ms}} N_{i_{mr}}}{N_i^2} H_{i_{ms} i_{mr},t} + \frac{N_{i_{mr}}^2}{N_i^2} H_{i_{mr} i_{mr},t} \right) \tag{15}$$

The probability of sampling non-identical genes “without” replacement from substage i_{ms} in

time t is expected to be equal to that from stage m in time $t-1$.

$$\frac{H_{i_{ms}i_{ms},t}}{1-1/(2N_{i_{ms}})} = \frac{H_{mm,t-1}}{1-1/(2N_m)} \tag{16}$$

Therefore,

$$H_{i_{ms}i_{ms},t} = \frac{1-1/(2t_{im}N_m)}{1-1/(2N_m)} H_{mm,t-1} \tag{17}$$

Two genes, each sampled from substage i_{ms} and i_{mr}, are not identical-by-descent with the probability of $H_{mm,t-1}$ just as in case 1-1-2.

$$H_{i_{ms}i_{mr},t} = H_{mm,t-1} \tag{18}$$

As for $H_{i_{mr}i_{mr},t}$, sampling the same gene twice from substage i_{mr}, which occurs with a probability of $1/(2N_{i_{mr}}) = 1/(2f_{im}N_m)$, would make no contribution to $H_{i_{mr}i_{mr},t}$. In the remaining conditions where the two genes randomly sampled with replacement are different, there is a probability of $H_{mm,t-1}$ that the two genes are from parents that were not identical by descent in the previous time step. Therefore, $H_{i_{mr}i_{mr},t}$ can be formulated as follows:

$$H_{i_{mr}i_{mr},t} = \left(1 - \frac{1}{2f_{im}N_m}\right) H_{mm,t-1} \tag{19}$$

Substituting Equation 7, 9, 17, 18, and 19 to Equation 15, $H_{ii,t}|_{\text{case 2-1}}$ can be formulated as follows:

$$H_{ii,t}|_{\text{case 2-1}} = \sum_m \left\{ \left(\frac{t_{im}N_m}{N_i}\right)^2 \frac{1-1/(2t_{im}N_m)}{1-1/(2N_m)} + \frac{2t_{im}f_{im}N_m^2}{N_i^2} + \left(\frac{f_{im}N_m}{N_i}\right)^2 \left(1 - \frac{1}{2f_{im}N_m}\right) \right\} H_{mm,t-1} \tag{20}$$

In case 2-2 (Figure 1, E), we just need to change j to i in Equation 12.

$$H_{ii,t}|_{\text{case 2-2}} = \sum_{k,l} \frac{2a_{ik}a_{il}N_kN_l}{N_i^2} H_{kl,t-1} \tag{21}$$

Following Equation 14, $H_{ii,t}$ can be formulated as the sum of Equation 20 and 21.

$$H_{ii,t} = \sum_m \left\{ \left(\frac{t_{im}N_m}{N_i}\right)^2 \frac{1-1/(2t_{im}N_m)}{1-1/(2N_m)} + \frac{2t_{im}f_{im}N_m^2}{N_i^2} + \left(\frac{f_{im}N_m}{N_i}\right)^2 \left(1 - \frac{1}{2f_{im}N_m}\right) \right\} H_{mm,t-1}$$

$$+ \sum_{k,l} \frac{2a_{ik}a_{il}N_kN_l}{N_i^2} H_{kl,t-1} \tag{22}$$

Developing matrix model

Combining case 1 (Equation 13) and case 2 (Equation 22), we construct a matrix equation.

$$\boldsymbol{H_t} = \boldsymbol{M}\boldsymbol{H_{t-1}} \tag{23}$$

$\boldsymbol{H_t}$ 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. 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 Equation 13 and 22. Detailed elements of $\boldsymbol{M}$ are shown in File S1, Section 2 in the case of $n = 2$. In general, multiplying matrix $\boldsymbol{M}$ will be 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.

$$\boldsymbol{H_t} \simeq \eta \boldsymbol{H_{t-1}} \tag{24}$$

$$\boldsymbol{H_t} \simeq c\boldsymbol{w} \tag{25}$$

where η and $\boldsymbol{w}$ are the leading eigenvalue and right eigenvector of matrix $\boldsymbol{M}$, respectively, and c is a constant. Equation 24 means that $H_{ij,t}$ for any i and j changes with a constant rate η over the course of time. Moreover, Equation 25 indicates that the relative ratio of expected heterozygosity among stage classes can be approximated by comparing the elements of $\boldsymbol{w}$.

H_t, which denotes expected heterozygosity of a whole population in time t, is the sum of $H_{ij,t}$ weighted by the number of individuals in stage i and j.

$$H_t = \frac{N_1^2}{N^2}H_{11,t} + \frac{2N_1N_2}{N^2}H_{12,t} + \cdots + \frac{2N_iN_j}{N^2}H_{ij,t} + \cdots + \frac{N_n^2}{N^2}H_{nn,t} \tag{26}$$

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

Materials and Methods

Accuracy of the model

To ensure that the procedures of model formulation described above are adequate, we tested whether η, which is the theoretical change rate of the probability of non-identical-by-descent and is obtained as the dominant eigenvalue of $\boldsymbol{M}$ in Equation 23, matched with the observed change rate of expected heterozygosity computed by stochastic simulation. We considered a life history with two stages: juvenile and adult (Figure 2). In this case, Equation 23 can be written

as follows:

$$\boldsymbol{H_t} = \begin{pmatrix} H_{11,t} \\ H_{12,t} \\ H_{22,t} \end{pmatrix} = \boldsymbol{M} \begin{pmatrix} H_{11,t-1} \\ H_{12,t-1} \\ H_{22,t-1} \end{pmatrix} \tag{27}$$

where $\boldsymbol{M}$ is a 3×3 matrix. The nine elements of $\boldsymbol{M}$ depend on transition probabilities (t_{11}, t_{21}, t_{12}, and t_{22}), fecundity (f_{12}), and the number of individuals in each stage (N_1 and N_2), as shown in Equation S12-S20 in File S1, Section 2. The four transition probabilities were set to be either of 0, 0.2, 0.4, 0.6, 0.8, and 1, while satisfying several conditions: $0 \leq t_{11} + t_{21} \leq 1$ (sum of probabilities should not exceed 1), $0 \leq t_{12} + t_{22} \leq 0.8$ (individuals in stage 2 were let to die at a minimum rate of 0.2 to avoid immortality), $t_{21} \neq 0$ (the life cycle should not be broken off), and $t_{11} + t_{22} > 0$ (to avoid eternal segregation between genes in different stages, either of the stasis probabilities must be greater than 0). We adjusted fecundity (f_{12}) so that populations were in an equilibrium state. As a result, 200 sets of parameter values that covered a wide range of life histories were generated. Total population size N was set to 100 and was divided into N_1 and N_2 as proportional to stable stage distribution. For each parameter set, we obtained η: the dominant eigenvalue of matrix $\boldsymbol{M}$ in Equation 27 (theoretical change rate of the probability of non-identical-by-descent). Meanwhile, we simulated 200 years of temporal dynamics of expected heterozygosity of a biallelic neutral locus 100 times, and calculated the mean ($\hat{H}_t$) for each time step. All simulations were initiated with maximum H_t (i.e., $H_0 = 0.5$), in which two alleles share the gene pool half-and-half in both stages. R script for the simulation is available in File S1, Section 3.2. We then calculated the annual change rate of $\hat{H}_t$ by

$$r_t = \frac{\hat{H}_t}{\hat{H}_{t-1}} \tag{28}$$

where $1 \leq t \leq 200$. We omitted the first year as burn-in, and tested if the geometric mean of the remaining 199 values of r_t ($2 \leq t \leq 200$) were significantly different from η. We logarized r_t and implemented t-test, following Alf and Grossberg (1979). We expected that the null hypothesis that r_t was equal to η would not be rejected. Although not rejecting the null hypothesis cannot fully guarantee that the null hypothesis is correct, it could be one practical way to verify the validity of our theoretical model.

Demographic genetic structure

Using our model, we formulated two indices of demographic genetic structure, (i) the ratio of expected heterozygosity between stages and (ii) inter-stage genetic differentiation, for the two-stage life history in Figure 2. The ratio of expected heterozygosity was defined by $\log(H_{11,t}/H_{22,t})$. When $\log(H_{11,t}/H_{22,t})$ is positive, $H_{11,t}$ is larger than $H_{22,t}$, and when negative vice versa. Based on Equation 25, $\boldsymbol{H_t}$ in Equation 27 can be written as follows:

$$\begin{pmatrix} H_{11,t} \\ H_{12,t} \\ H_{22,t} \end{pmatrix} = c\boldsymbol{w} = c \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} \tag{29}$$

where $\boldsymbol{w}$, the transpose of (w_1, w_2, w_3), is the leading right eigenvector of matrix $\boldsymbol{M}$. Therefore, $\log(H_{11,t}/H_{22,t})$ can be formulated as follows:

$$\log\left(\frac{H_{11,t}}{H_{22,t}}\right) = \log\left(\frac{w_1}{w_3}\right) \tag{30}$$

It should be noted that $\log(H_{11,t}/H_{22,t})$ is time-invariant, although $H_{11,t}$ and $H_{22,t}$ themselves change with time.

As for genetic differentiation, we defined inter-stage F_{st} as follows:

$$F_{st} = \frac{H_t - \overline{H_{ii,t}}}{H_t} \tag{31}$$

$\overline{H_{ii,t}}$ is the arithmetic mean of $H_{11,t}$ and $H_{22,t}$. Substituting Equation 26 and 29 to Equation 31, F_{st} turns out to be time-independent, as with $\log(H_{11,t}/H_{22,t})$:

$$F_{st} = \frac{(w_1 N_1^2 + 2w_2 N_1 N_2 + w_3 N_2^2)/N^2 - (w_1 + w_3)/2}{(w_1 N_1^2 + 2w_2 N_1 N_2 + w_3 N_2^2)/N^2} \tag{32}$$

It should be noted that the constant coefficient c in Equation 29 was cancelled out when put into Equation 31.

We obtained $\log(H_{11,t}/H_{22,t})$ and F_{st} under the same parameter values as the "Accuracy of the model" section, except N being 50, 100, and 1000, using Equation 30 and 32. We then examined the dependence of the two indices on demographic rates (t_{11}, t_{21}, t_{12}, and t_{22}), total population size N, and stable stage distribution $\log(N_1/N_2)$. Moreover, to evaluate if demographic genetic structure could serve as a proxy of temporal genetic dynamics, the relationships with η and η^T were also examined, where T is generation time defined as the expected age of parents of a newborn cohort. η and η^T reflect the change rate of expected heterozygosity per year and per generation time, respectively. T was calculated by the same method for calculating generation length in Waples et al. (2013).

All analyses, including simulation, were carried out in R ver. 4.1.1 (R Core Team 2021).

Data availability statement

The authors state that all data necessary for confirming the conclusions presented in the article are presented within the article. Supplemental material available at figshare.

Results

Comparison of simulation results and theoretical expectation

The average expected heterozygosity at each time step computed by stochastic simulation depicted monotonic and exponential decrease over the course of time for all 200 sets of parameter values. The geometric mean of the annual change rate of the simulated expected heterozygosity (r_t) turned out not to be significantly different from theoretical expectation (η), which was obtained as a dominant eigenvalue of matrix $\boldsymbol{M}$ in Equation 27 in 195 out of 200 sets of parameter values with significance level 0.05. Although there were 5 out of 200 cases where theoretical expectation did not match with simulation results, it is sufficiently possible because the proportion of 5/200 was below the significance level of 0.05. We also compared the kernel density distribution of $\log(r_t)$ ($2 \leq t \leq 200$) and $\log(\eta)$, confirming that η well matched with the distributional center of r_t for all parameter settings (File S1, Section 3.1). Thus, our model seemed to be valid across a wide range of the parameter space.

Relative ratio of expected heterozygosity

$\log(H_{11,t}/H_{22,t})$, which was obtained from our model (Equation 30), became gradually large with increasing t_{11} and t_{12} and decreasing t_{21} and t_{22} (Figure 3, A). These results exactly mean that $H_{11,t}$ was relatively larger than $H_{22,t}$ under high t_{11} and t_{12}, while vice versa under high t_{21} and t_{22}. The same pattern held regardless of total population size N (File S1, Section 4). However, as N increased, $\log(H_{11,t}/H_{22,t})$ approached to zero, which means $H_{11,t}$ and $H_{22,t}$ became equal to each other (Figure 4, A).

The ratio of expected heterozygosity was clearly associated with stable stage distribution: $\log(H_{11,t}/H_{22,t})$ was positively correlated with $\log(N_1/N_2)$ (Figure 5, A). The correlation became weaker with increasing N, as $\log(H_{11,t}/H_{22,t})$ converged to zero. On the other hand, as for

temporal dynamics of expected heterozygosity, $\log(H_{11,t}/H_{22,t})$ did not have apparent correlation with either η or η^T (Figure 5, B; File S1, Section 6).

Genetic differentiation between stages

F_{st}, a proxy of inter-stage genetic differentiation, was obtained from Equation 32. Overall, F_{st} took small values ranging from 0 to 0.038. F_{st} did not change monotonously with an increase or a decrease in the four transition probabilities, showing no clear patterns compared with $\log(H_{11,t}/H_{22,t})$ (Figure 3, B; File S1, Section 5). As N increased, F_{st} became smaller and approached to zero, which is a situation where the two stages are genetically homogeneous (Figure 4, B). Besides, Fst showed convex downward when plotted against $\log(N_1/N_2)$ (Figure 5, C). The vertex of the convex curve stood at around $\log(N_1/N_2) = 0$, where the both stage consist of the same number of individuals. In other words, F_{st} was high when the number of individuals was skewed to either stages. F_{st} was negatively correlated with both η and η^T, while the correlations were along the gradient of total population size N (Figure 5, D; File S1, Section 6).

Discussion

Determinants of demographic genetic structure

Comparison of genetic diversity and composition among different life history stages within a population is often considered as a proxy of temporal genetic dynamics (Aldrich *et al.* 1998, Mimura *et al.* 2017, Schmidt *et al.* 2018). When interpreting the inter-stage genetic structure (demographic genetic structure), it is necessary to account for factors that would obscure the signal of temporal change. Although there is a study that used simulation to evaluate the impacts of genetic and demographic stochasticity (Tsuzuki *et al.* 2021), no studies have examined how

life history affects demographic genetic structure.

Our theoretical model, whose validity was confirmed by comparison with simulation, clarified the major impacts of stable stage distribution on demographic genetic structure. Relative ratio of $H_{11,t}$ and $H_{22,t}$ corresponded to the skewness of stable stage distribution: expected heterozygosity was relatively high in stage with more individuals, and low in stage with less individuals (Figure 5, A). This result indicates that relative ratio of genetic diversity among life history stages might strongly reflect stage distribution. Skewed stage distribution also affected F_{st}: F_{st} was high when the number of individuals was unbalanced between the two stages (Figure 5, C).

In general, small number of individuals intensifies stochastic genetic drift due to increased sampling bias in gene frequencies between successive generations, leading to loss of genetic diversity and promoting genetic differentiation (Kimura and Crow 1970). When stage distribution is skewed, the two stages consist of different number of individuals, which might make the degree of stochasticity to vary between the stages. Stage with smaller number of individuals are made up of genes that were sampled fewer times from the gene pool of the previous time step, either by survival or by reproduction, than stage with larger number of individuals, thus suffering random perturbation in gene frequencies to a greater extent. The alleviated stochasticity must have resulted in the loss of genetic diversity and the high degree of genetic differentiation.

As the total population size N increased, inter-stage difference in genetic diversity and composition were erased even if the stage distribution was skewed (Figure 4), indicating that the number of individuals of either stage became large enough to diminish stochasticity. Therefore, not only the skewness of stage distribution but also population size determine demographic genetic structure because of the sampling bias effect.

How we should interpret demographic genetic structure

A common practice in interpreting demographic genetic structure is that if juvenile stages are less diverse than, and are highly differentiated from, mature stages, genetic diversity would decrease with time over the course of generation turnover. Our model showed that relative ratio of expected heterozygosity between stages did not correlate with either η or η^T: as η and η^T decreased, $\log(H_{11,t}/H_{22,t})$ did not necessarily become lower (i.e., $H_{11,t}$ did not necessarily get smaller than $H_{22,t}$). Similarly, F_{st} did not change monotonously with η and η^T. Although F_{st} had a negative correlation with both η and η^T, the correlation seemed to result from the confounding effect of total population size N. These results indicate that inferring temporal trends in genetic diversity solely from demographic genetic structure, which has been often done by previous studies, is potentially misleading. This study, to our knowledge, for the first time draws caution on the conventional interpretation on demographic genetic structure.

For a more reliable interpretation, it is necessary to account for both stage distribution and total population size, which turned out to be the two major determinants of demographic genetic structure in this study. One way is to compare raw demographic genetic structure, which is obtained by any genetic markers, with the theoretical expectation calculated based on the equations we derived. The deviations of observed structure from expectation reflect factors unexplored in our analyses, such as fluctuating population size, non-random mating, selection, and immigration. Thus, our model can work as a null model for a variety of purposes. To make the most use of our formulas, it is necessary to monitor individuals from year to year and to estimate transition probabilities and fecundity 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 the effects stage distribution. Combining demographic survey with demographic genetic data would yield

reliable insights into the genetic dynamics of stage-structured populations.

Funding Information

This research was financially supported by Grants-in-Aid for Scientific Research from the JSPS KAKENHI (Grant no. 19H03294, 20K06821, and 21J10814).

Conflicts of Interest

The authors are competing no conflicts of interest.

References

Agashe, D., Falk, J. J., and Bolnick, D. I. (2011). Effects of founding genetic variation on adaptation to a novel resource. Evolution 65(9): 2481-2491.

Aldrichl, P. R., Hamrick, J. L., Chavarriaga, P., and Kochert, G. (1998). Microsatellite analysis of demographic genetic structure in fragmented populations of the tropical tree *Symphonia globulifera*. Mol. Ecol. 7: 933-944.

Alf, E. F., and Grossberg, J. M. (1979). The geometric mean: confidence limits and significance tests. Percept. Psychophys. 26(5): 419-421.

Ally, D., and Ritland, K. (2007). A case study: looking at the effects of fragmentation on genetic structure in different life history stages of old-growth mountain hemlock (*Tsuga mertensiana*). J. Hered. 98(1): 73-78.

Barrett, R. D. H., and Schluter, D. (2008). Adaptation from standing genetic variation. Trends Ecol. Evol. 23(1): 38-44.

Caswell, H. (2001). *Matrix population models*, 2nd edn. Cambridge University Press, New York.

Charlesworth, B. (2008). *Evolution in age-structured populations*, 2nd edn. Cambridge University Press, New York.

Crow, J. F., and, Kimura, M. (1970). *An introduction in population genetic theory*. Harper and Row, New York.

Ellis, M. M. (2012). Matrix population models from 20 studies of perennial plant populations. Ecology 93(4): 951-951.

Felsenstein, J. (1971). Inbreeding and variance effective numbers in populations with overlapping generations. Genetics 68: 581-597.

Hill, W. G. (1972). Effective size of populations with overlapping generations. Theor. Popu. Biol. 3: 278-289.

Hill, W. G. (1979). A note on effective population size with overlapping generations. Genetics 92(1): 317-322.

Johnson, D. L. (1977). Inbreeding in populations with overlapping generations. Genetics 87(3): 581-591.

Kitamura, K., and A. Nakanishi, 2021 Recovery process of genetic diversity through seed and pollen immigration at the northernmost leading-edge population of *Fagus crenata*. Plant Species Biol. 36(3): 489-502.

Knight, T. M. 2004 The effects of herbivory and pollen limitation on a declining population of *Trillium grandiflorum*. Ecol. Appl. 14(3): 915-928.

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

Mimura, M., T. Yahara, D. P. Faith, E. Vázquez-Domínguez, R. I. Colautti *et al.*, 2017 Understanding and monitoring the consequences of human impacts on intraspecific variation. Evol. Appli. 10: 121-139.

Orive, M. E. 1993 Effective population size in organisms with complex life-histories. Theor. Popu. Biol. 44: 316-340.

Ohara, M., T. Takada, and S. Kawano, 2001 Demography and reproductive strategies of a polycarpic perennial, *Trillium apetalon* (Trilliaceae). Plant Species Biol 16(3): 209-217.

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., O. Kaltz, and M. E. Hochberg, 2012 Evolutionary rescue in populations of *Pseudomonas flurescens* across an antibiotic gradient. Evol. Appli. 6: 608-616.

Schmidt, D. J., S. Fallon, D. T. Roberts, T. Espinoza, A. McDougall *et al.*, 2018 Monitoring age-related trends in genomic diversity of Australian lungfish. Mol. Ecol. 27: 3231-3241.

Silvertown, J. 1987 *Introduction to plant population ecology*, 2nd edn. Longman Scientific & Technical, New York.

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

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

Tsuzuki, Y., M. P. Sato, A. Matsuo, Y. Suyama, and M. Ohara, 2021 Genetic consequences of habitat fragmentation in a perennial plant *Trillium camschatcense* are subjected to its slow-paced life history. Popu. Ecol.

Waples, R. S., C. Do, and J. Chopelet, 2011 Calculating Ne and Ne/N in age-structured populations: a hybrid Felsenstein-Hill approach. Ecology 92(7): 1513-1522.

Waples, R. S., G. Luikart, J. R. Faulkner, and D. A. Tallmon, 2013 Simple life-history traits explain key effective population size ratios across diverse taxa. Proc. R. Soc. B 280:

20131339.

Yonezawa, K., E. Kinoshita, Y. Watano, and H. Zentoh, 2000 Formulation and estimation of the effective population size of stage-structured populations in *Fritillaria camtschatcensis*, a perennial herb with a complex life history. Evolution 54(6): 2007-2013.

Figure captions

Figure 1. Five possible transition patterns of genes between two successive time steps. (A) case 1-1, (B) (C) case 1-2, (D) case 2-1, (E) case 2-2. Single- and double-line arrows represent transition by survival and reproduction, respectively. In (A), sampling of two genes in time t is visualized with gray arrows. There are four possible pairs of substages, which are classified into three cases (1-1-1, 1-1-2, and 1-1-3) in terms of the combination of survival and reproduction

Figure 2. The two-stage life history used in analyses. We assumed that reproduction only took place in the direction from the adult stage to the juvenile stage (i.e., $f_{12} > 0$, $f_{11} = f_{21} = f_{22} = 0$). Population dynamics can be modeled as the matrix equation. Stable stage distribution is proportional to the leading right eigenvector of the 2×2 matrix in the right side of the equation

Figure 3. Parameter dependence of (A) $\log(H_{11,t}/H_{22,t})$ and (B) F_{st} on transition probabilities. Rows and columns of large square correspond to t_{22} and t_{12}, respectively. The positions of small squares within a large one represent t_{21} and t_{11}. The color filled in each small square shows the value of $\log(H_{11,t}/H_{22,t})$ or F_{st}

Figure 4. The dependence of (A) $\log(H_{11,t}/H_{22,t})$ and (B) F_{st} on the total population size N. The dotted line in (A) represents $\log(H_{11,t}/H_{22,t}) = 0$ (i.e., $H_{11,t} = H_{22,t}$), while that in (B) shows $F_{st} = 0$, in which the juvenile and the adult stages are genetically homogeneous

Figure 5. The relationships between the indices of demographic genetic structure (i.e., $\log(H_{11,t}/H_{22,t})$ and F_{st}) and either of stable stage distribution ($\log(N_1/N_2)$, A and C) and η (B and D). Color of the dots stands for N: $N = 50$ (blue), $N = 100$ (black), and $N = 1000$ (red)

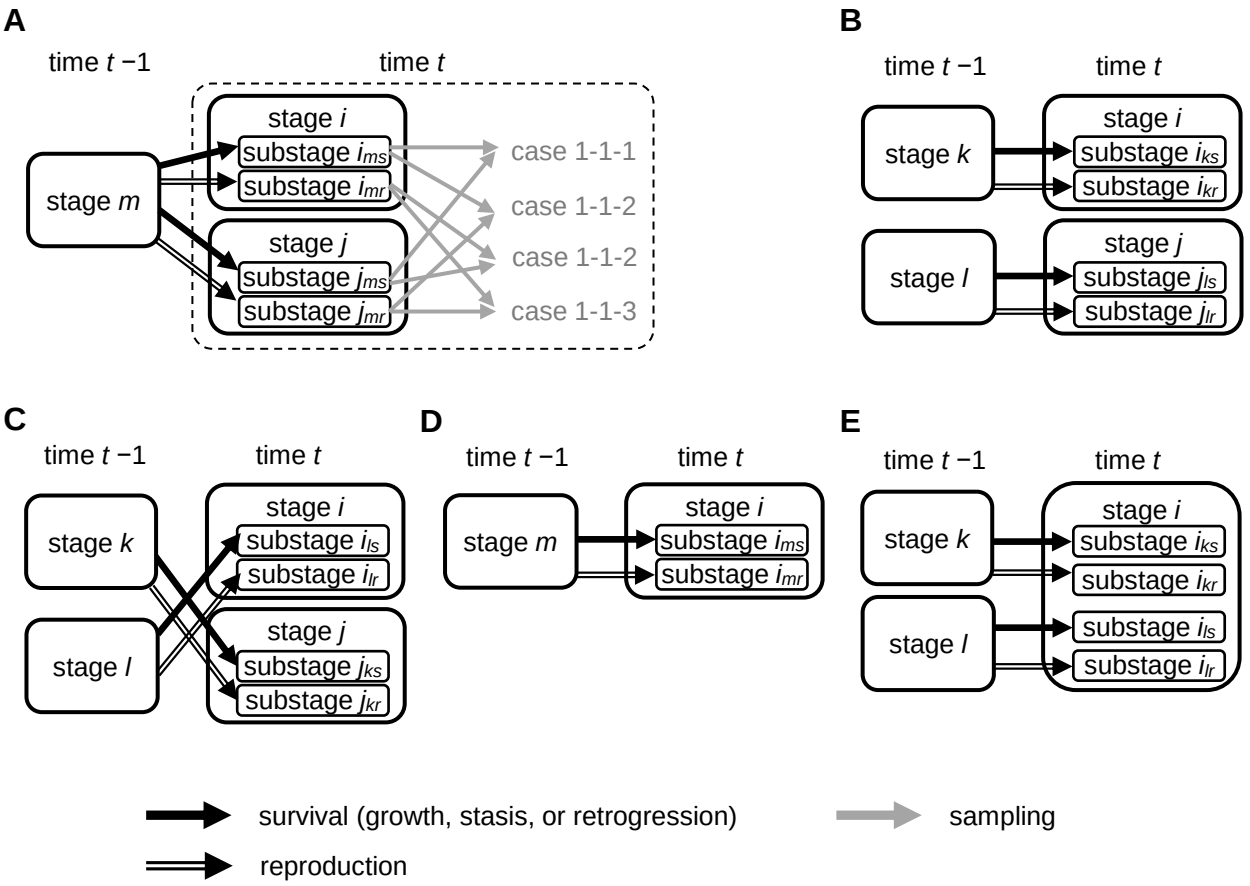

Figure 1. Five possible transition patterns of genes between two successive time steps. (A) case 1-1, (B) (C) case 1-2, (D) case 2-1, (E) case 2-2. Single- and double-line arrows represent transition by survival and reproduction, respectively. In (A), sampling of two genes in time t is visualized with gray arrows. There are four possible pairs of substages, which are classified into three cases (1-1-1, 1-1-2, and 1-1-3) in terms of the combination of survival and reproduction

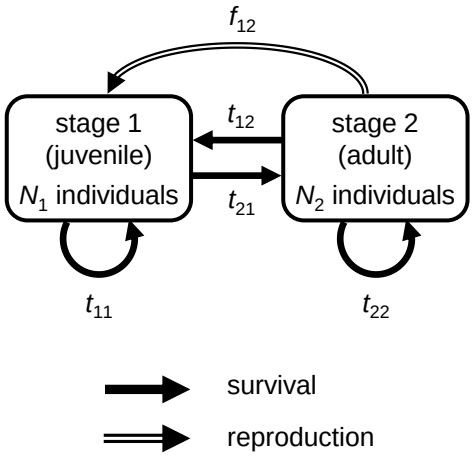

Figure 2. The two-stage life history used in analyses. We assumed that reproduction only took place in the direction from the adult stage to the juvenile stage (i.e., $f_{12} > 0$, $f_{11} = f_{21} = f_{22} = 0$). Population dynamics can be modeled as the matrix equation. Stable stage distribution is proportional to the leading right eigenvector of the 2×2 matrix in the right side of the equation

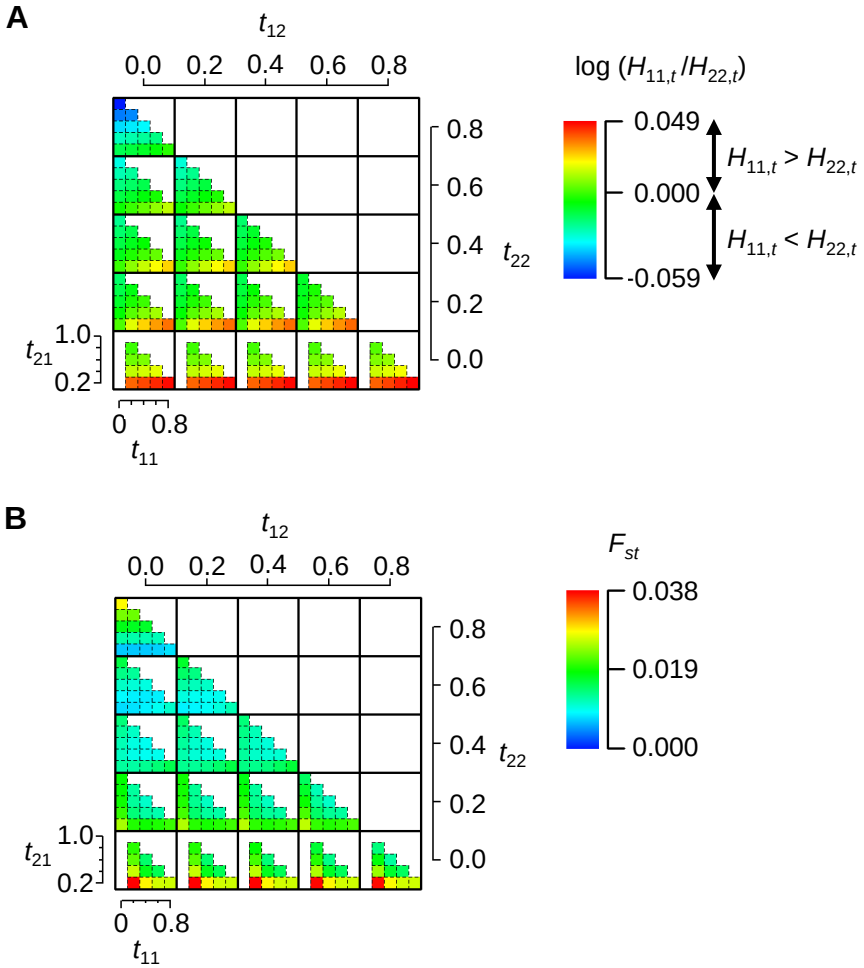

Figure 3. Parameter dependence of (A) $\log(H_{11,t}/H_{22,t})$ and (B) F_{st} on transition probabilities. Rows and columns of large square correspond to t_{22} and t_{12}, respectively. The positions of small squares within a large one represent t_{21} and t_{11}. The color filled in each small square shows the value of $\log(H_{11,t}/H_{22,t})$ or F_{st}

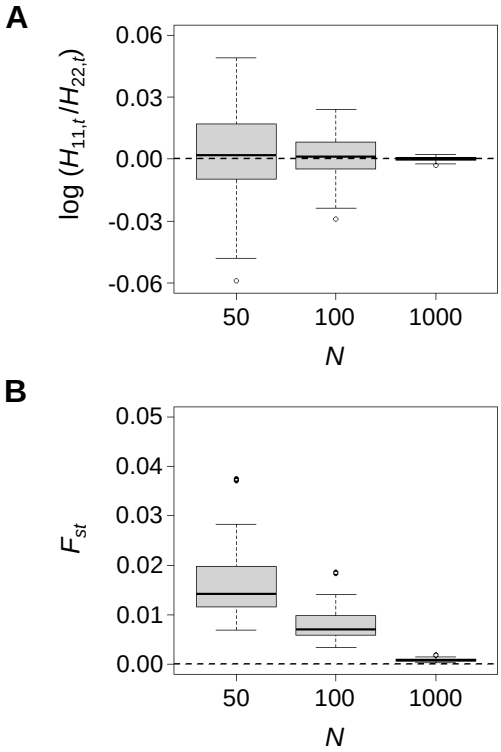

Figure 4. The dependence of (A) $\log(H_{11,t}/H_{22,t})$ and (B) F_{st} on the total population size N. The dotted line in (A) represents $\log(H_{11,t}/H_{22,t}) = 0$ (i.e., $H_{11,t} = H_{22,t}$), while that in (B) shows $F_{st} = 0$, in which the juvenile and the adult stages are genetically homogeneous

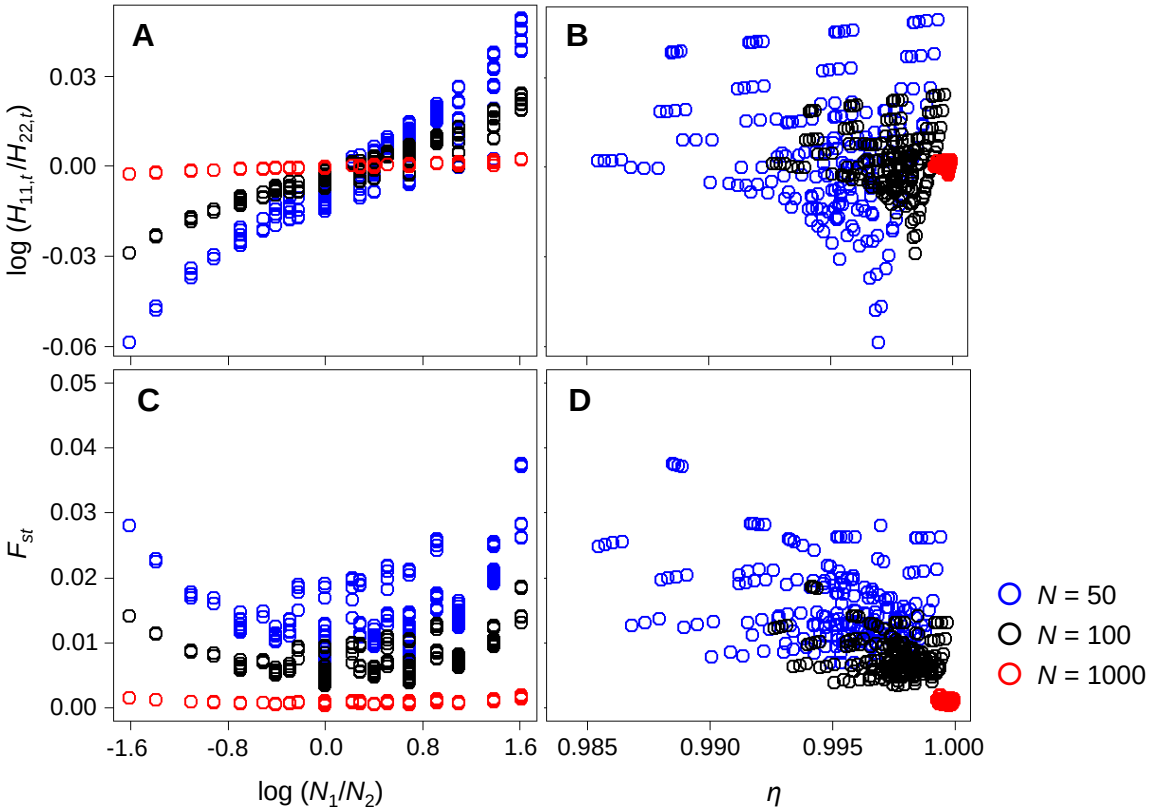

Figure 5. The relationships between the indices of demographic genetic structure (i.e., $\log(H_{11,t}/H_{22,t})$ and F_{st}) and either of stable stage distribution ($\log(N_1/N_2)$, A and C) and η (B and D). Color of the dots stands for N: $N = 50$ (blue), $N = 100$ (black), and $N = 1000$ (red)