Simulation as Experiment by Using a Metapopulation Model for Exploring a New Way of Landscape Genetics

Abstract

A metapopulation model, concerning an annual species inhabiting in a limited number of patches, is discussed in this paper. A simulation is conducted with R (R Core Team 2017). Two questions are focused on: how landscape properties influence genetic differentiation of the metapopulation and how conservation planning alters the local adaptation and metapopulation viability The results show that at a neutral locus, higher carrying capacity (K) of the patches and higher dispersal rate (d) both lead to lower genetic differentiation; at a non-neutral locus, increasing maximum fitness (W_{max}) leads to better local adaptation and metapopulation viability, while higher dispersal rate (d) does the opposite. According to the results, suggestion for conservation planning could be given that W_{max} and d should be controlled to a critical point in order to maintain the species.

Key words：simulation, genetic differentiation, metapopulation viability, landscape genetics

Introduction

The combination of landscape ecology and population genetics forms a new subject, Landscape Genetics (Manel et al. 2003). It's a study of genetic variation and focuses on how landscape properties are reflected in gene-flow patterns (Holderegger 2008). Landscape genetics, as a tool, is also applied in the field of conservation planning (van Strien *et al.* 2014). This paper combines the methods of landscape ecology and population genetics, in order to get close to the frontier of the discipline and explore a new way in conservation planning.

To understand landscape genetics, a basic knowledge of population genetics is required. Population genetics started with the separately published papers of Godfrey Hardy (Hardy 1908)and Wilhelm Weinberg (Weinberg 1908). The core theory is later named as Hardy-Weinberg Equilibrium Principle, which states that genotype frequencies will remain the same in a population, if there are no selection, no mutation, no migration, no genetic drift or gene flow and no mate choice. It is however only an ideal assumption. Heterozygosity is inevitably reduced by genetic drift in actual populations.

In order to study and simulate a real population, the concept of effective population size, N_e, is raised. An ideal population with this size would lose heterozygosity just as the real one (Freeman & Herron 2007). Sex ratios, fluctuations in population size, variance in reproductive success, overlapping generations and spatial population structure will all influence N_e.

Another important concept about a population is the genetic differentiation. In a metapopulation, which was first raised by Levins (Levins 1970) as a population composed of several subpopulations that frequently extinguish locally but recolonize and regenerate, fixation index is used to assess the genetic differentiation. In 1973, Masatoshi Nei (Nei 1973) introduced a method of measuring genetic diversity within and between populations. The method is later developed into the standard measurement of fixation index. The fixation index in the scale of between total population and subpopulation is defined as F_{st} that can be calculated as:

$$F_{st} = 1 - \overline{H_s}/H_t$$

$\overline{H_s}$ is the mean expected heterozygosity within subpopulations.

H_t is the expected heterozygosity at metapopulation level.

F_{st} is used to assess how spatial structure influence genetic structure. It can have a value between 0 and 1. When F_{st} is below 0.05, the genetic variation is low, which means the subpopulations are sharing many of the alleles. On the other hand, a F_{st} lager than 0.25 refers to a very high genetic variation. A value of around 0.15 is considered to be moderate. (Hartl & Grant 1997)

To conduct simulations and calculate the indices mentioned above, computer simulation technology is introduced in recent decades. The new technology simplifies the procedure as well as deepens and widens the view of researches. Computer simulation can reveal the historical characteristics in the evolutionary process and predict the influence of changes in landscape properties on genetic differentiation. This is becoming a new trend in the research of metapopution (Balloux 2001). In the evolution of metapopulations, the natural factors such as mutation, migration, dispersal and drift, and human-related factors such as acclimation, habitat fragmentation and selection, can all be simulated by computer (Peck 2004).

In this paper, a metapopulation model is set up and simulated based on the computer simulation technology. The following two questions are focused on and discussed:

1. What's the influence of landscape properties on genetic differentiation at a neutral locus?
2. What's the influence of landscape management on local adaptation and metapopulation viability?

University of Hohenheim
Community & Evolutionary Ecology   Haiyin Ye 718002

Methodology

A metapopulation consists of two or more subpopulations or local populations that are spatially separated but functionally related. The subpopulations form a system of patch populations, which can be described by Levins model (Levins 1970):

$$\frac{dp}{dt} = cp(1-p) - ep$$

P is the patches occupancy ratio. c is colonization coefficient. e is extinction coefficient.

This model has mathematical and conceptual similarities with the balance model of island biogeography (MacArthur & Wilson 1967). Levins' definition of metapopulation emphasizes the local population turnover, which means that individuals disappear in some patches then recolonize, and repeat this process. Therefore, metapopulation has two prerequisites: one is the frequent local extinction of subpopulations, the other is the dispersal and recolonization possibility from one patch to another.

The simulation model of this study is based on levins model. The model is designed as follows:

1. A certain annual species with sexual reproduction is assumed. The individuals are diploid.
2. The individuals (adults) have the ability of selfing so all of them are hermaphroditic.
3. The habitats of the species are assumed to be a limited number ($N_{patches}$) of patches. Dispersal of individuals is possible from one patch to another.
4. The same carrying capacity K is assumed for all patches.
5. There's a "ceiling density-dependence", which means that the population growth is density-dependent.

According to the basic design, the simulation modules in R platform (R Core Team 2017) are further set up to follow the flow chart as shown in Figure 1. The package hierfstat (Goudet & Jombart 2015) is used in the R script.
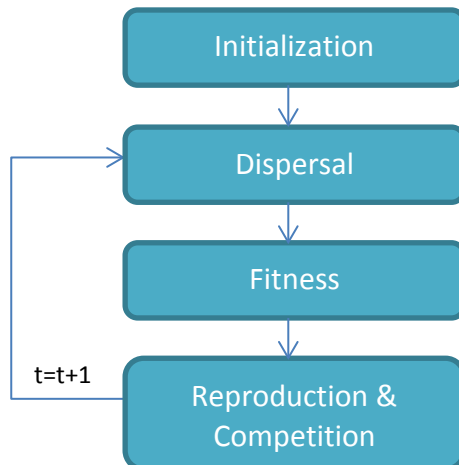

Figure 1. Simulation Flow Chart

The detailed designs within the modules are described as follows:

3

1. Initialization

    The number of adults in each patch is initialized as N_{init}. The genotype of each adult is assembled randomly to a number from 0.01 to 1 (100 possible alleles).

2. Dispersal

    The dispersal follows the Island Model (Wright 1931). An assumed proportion (d) of adults disperses and lands randomly on one of the patches. It is possible that a dispersing adult lands on the same patch that it just leaves.

3. Fitness

    Two different conditions are considered in this step.

    One is the neutral case, in which the same fitness is expected for all adults.

    $$w = w_{max}$$

    The other is the non-neutral case, in which the expected fitness of an adult (w) depends on its breeding value (B) and local environment (E).(Leimu *et al.* 2006)

    $$w = w_{max} e^{-(E-B)^2}$$

    The breeding value of an adult is determined by the two alleles it carries.

    $$B = a_1 + a_2$$

4. Reproduction

    Each adult has a number ($N_{offspring}$) of offspring, which is drawn from Poisson distribution with mean w. The adults that have offspring contribute $N_{offspring}$ of gametes to a "gamete pool". A procedure called random mating is conducted to determine the genotype of the offspring, which means to form an offspring by drawing two random gametes from the gamete pool.

    Competition

    The offspring, however, cannot all live to the next round due to "ceiling density-dependence". If the total number of offspring exceeds the carrying capacity K, only K offspring can recruit.

5. Loop: The survived offspring become new adults and start the next round.

The neutral case in module in module Fitness is first run to research the first question.

1. What's the influence of landscape properties on genetic differentiation at a neutral locus?

The landscape properties are considered to be:

(1) Geographic properties that influence the carrying capacity of each patch.

For example, plain may carry more individuals than other landforms.

In this case, simulation is delivered with a constant d=0.2 and varying K = {10,20,50,100,200,500}. F_{st} and $N_{metapop}$ by different K are recorded.

(2) Transportation conditions that influence the dispersal rate.

For example, the patches may completely isolated or far away from each other, or on the other hand very close to each other.

In this case, simulation is delivered with a constant K=50 and varying d={0.00, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45}. F_{st} and $N_{metapop}$ by different d are recorded.

Then, the non-neutral case in Fitness is run for the second question.

2. What's the influence of landscape management on local adaptation and metapopulation viability?

A non-neutral case simulation will be conducted for this question. Landscape management may change:

(1) Maximum fitness of individuals

In this case, simulation is delivered with a constant K=50, a constant d=0.2 and varying w_{max}={0.90, 0.95, 1.00, 1.05, 1.10, 1.15, 1.20, 1.25, 1.30, 1.35, 1.4}. w_{mean} and $N_{metapop}$ by different w_{max} are recorded.

(2) Dispersal rate

In this case, simulation is delivered with a constant K=50, a constant w_{max}=1.2 and varying d={0.00, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45}. w_{mean} and $N_{metapop}$ by different d are recorded.

The simulations are conducted more than once. The results are slightly different each time due to the random process but general trends can be concluded. In the next section, only one typical result of the simulation for each sub-question is reported.

Results

The results of the two groups of simulation are reported as follows:

1. What's the influence of landscape properties on genetic differentiation at a neutral locus?
(1) Vary carrying capacity, K (for d=0.2)

As a whole, this metapopulation has a low fixation rate (always below 0.05).With the growth of carrying capacity of each patch, the F_{st} index declines, which means that the genetic differentiation decreases (shown in Figure 2 and table 1). This trend becomes more obvious if the data is plotted in a log scale (Figure 3).

Table 1. Records of K, N and F_{st}

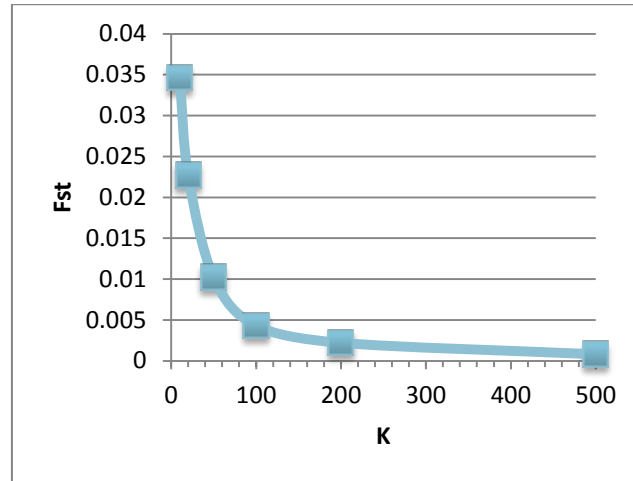| K	10	20	50	100	200	500
N	151	379	984	1991	3993	10000
Fst	0.0346	0.0227	0.0102	0.0043	0.0022	0.0008

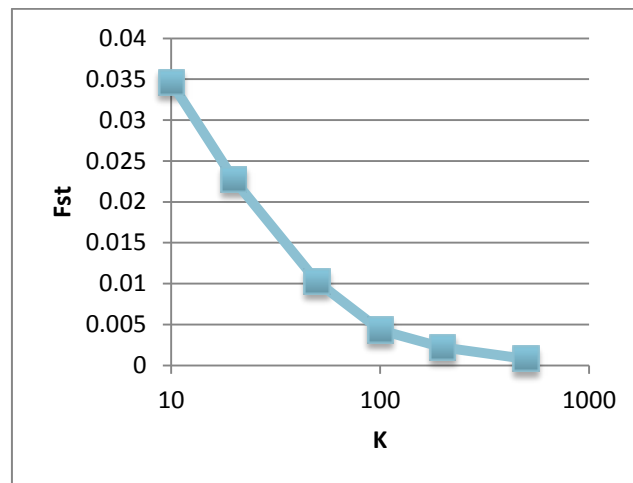*Figure 2. F_{st} with changing K*

Figure 3. F_{st} with changing K (in log scale)

(2) Vary dispersal rate, d (for K=50)

When dispersal rate is 0, the fixation index is higher than moderate (>0.15). Once there's possibility of dispersal, the fixation index becomes lower than 0.05, which could be defined as low. As the dispersal rate grows, the F_{st} index also has a declining trend, though not strictly, which means the genetic differentiation also has a trend of decrease (as shown in Figure 4 and table 2).

Table 2. Records of d, N and F_{st}

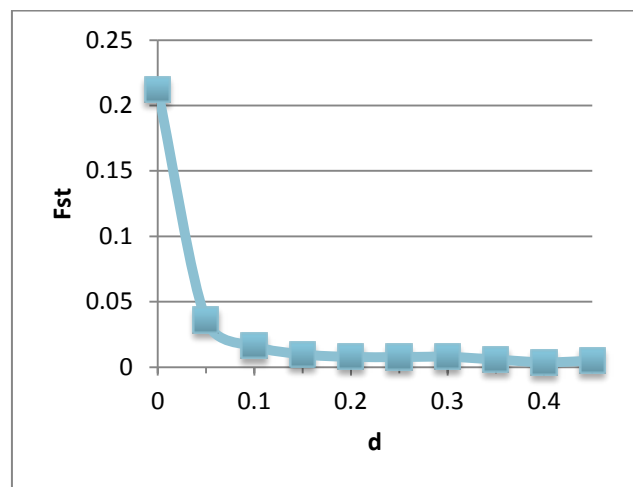| d	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45
N	985	979	995	984	989	990	986	985	984	989
Fst	0.2120	0.0360	0.0162	0.0098	0.0078	0.0077	0.0079	0.0058	0.0037	0.0049

Figure 4. F_{st} with changing d

If we also plot the number of individuals left at the end of simulation, it fluctuates within a small range (shown in Figure 5). This will be compared with the non-neutral case in question 2.
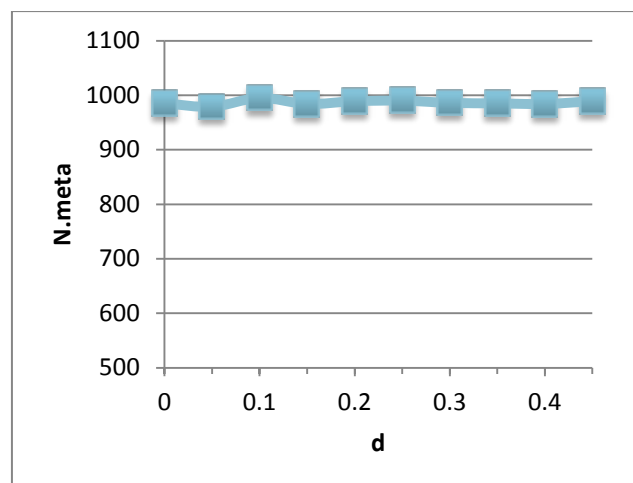

Figure 5. N_{meta} with changing d (neutral case)

7

To sum up, at a neutral locus, an increase in the carrying capacity of the patches leads to a decrease in fixation index; an increase in dispersal rate of individuals also leads to a decrease in fixation rate. The conclusion might be drawn that both the growth of carrying capacity and the growth of dispersal rate have a negative effect on genetic differentiation.

2. What's the influence of landscape management on local adaptation and metapopulation viability?
(1) vary maximum fitness, w_{max} (for d=0.2, K=50)

When the maximum fitness is too low (below 1.05), the metapopulation goes extinct at a high possibility. As it reaches 1.1, the number of surviving individuals starts to increase as w_{max} grows (shown in Figure 6 and table 3).At 1.1, the remaining population is a very small number of 25, which would probably drop to zero if the simulation runs for several more rounds. After w_{max} reaches 1.2, the final population number exceeds the initial number (700). The metapopulation is growing in each round and will possibly keep growing.

At the same time, the mean fitness also grows in the same direction (shown in Figure 7 and table 3).

Table 3. Records of W_{max}, N and W_{mean}

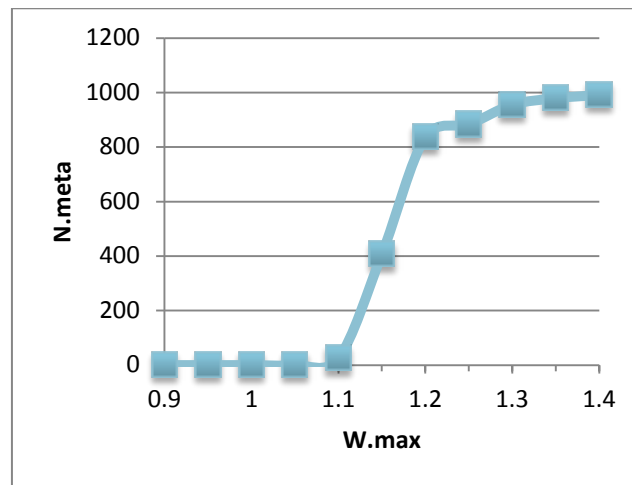| W.max	0.9	0.95	1	1.05	1.1	1.15	1.2	1.25	1.3	1.35	1.4
N	0	0	0	0	25	406	836	883	955	979	991
w.mean	NA	NA	NA	NA	0.930	1.058	1.070	1.116	1.141	1.198	1.242

Figure 6. N_{meta} with changing w_{max}
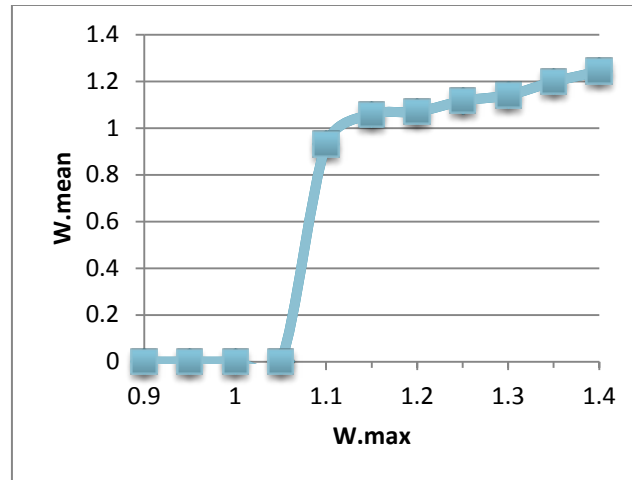
Figure 7. w_{mean} with changing w_{max}

(2) vary dispersal rate, d (for w_{max}=1.2, K=50)

The number of surviving individuals remains high with the dispersal rate in the range of 0 to 0.15. It drops sharply when dispersal rate continues to increase. When the dispersal rate is lager than 0.3, the metapopulation is very likely to go extinct (shown in Figure 8 and table 4). The mean fitness also drops gradually as dispersal rate grows(shown in Figure 9 and table 4). It falls to zero when the population fails to survive.

Table 4. Records of d, N and W_{mean}

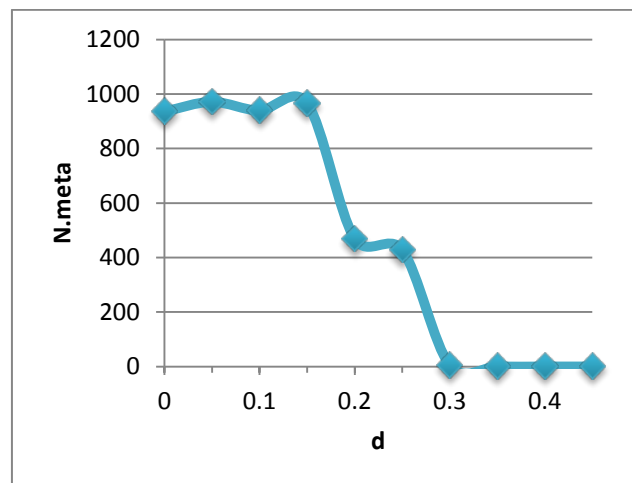| d	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45
N	936	971	939	966	469	428	5	0	0	0
w.mean	1.170	1.156	1.122	1.083	1.060	1.052	0.968	NA	NA	NA

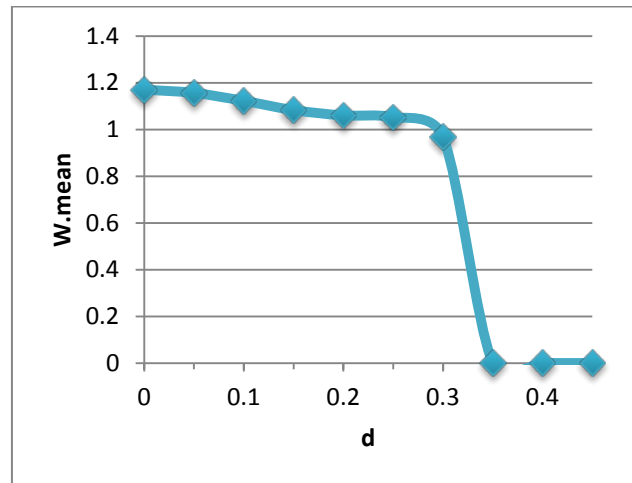Figure 8. N_{meta} with changing d (non-neutral case)

Figure 9. w_{mean} with changing d

   In conclusion, higher maximum fitness leads to better local adaptation and metapopulation viability, while higher dispersal rate does the opposite. Compared to the neutral case (Figure 5), the negative effect of dispersal rate is quite obvious in the non-neutral case.

Discussion

With the results stated above, both the questions could be answered.

1. What's the influence of landscape properties on genetic differentiation at a neutral locus?

At a neutral locus, as the landscape properties change, both the growth of carrying capacity (K) and the growth of dispersal rate (d) have a negative effect on genetic differentiation.

A larger carrying capacity reduces the local competition and allows more individuals carrying different alleles to survive in one patch. With random mating process, the offspring of these individuals will be more "mixed" in genotype, so that the subpopulations will tend to share more alleles and become less differentiated in genotype.

A lager dispersal rate allows the individuals to travel around patches more easily. Different alleles will be brought to all patches in a smaller time period. The subpopulations become more similar in genotype sooner so the fixation rate drops more quickly.

2. What's the influence of landscape management on local adaptation and metapopulation viability?

Higher maximum fitness (w_{max}) leads to better local adaptation and metapopulation viability, while higher dispersal rate (d) does the opposite.

With higher maximum fitness, all individuals facing the same environment have a higher fitness of their own ($w = w_{max}e^{-(E-B)^2}$). They can thus adapt to local environment better and reproduce offspring that also adapt to the environment. The metapopulation grows better in this way.

Higher dispersal rate, however, increase the number of individuals that land in a "wrong" environment. In this case, less offspring are reproduced and those survived may also disperse and land in another unsuitable environment. Compared to the neutral case, the negative effect of dispersal rate is quite obvious in the non-neutral case.

It could also be concluded that there is a critical point for a certain factor. When the factor goes below/over the point, the metapopulation can have a stable growth while in the other way the population will probably extinguish in the end. This provides evidence for some advices in conservation planning, such as to keep a certain factor over/ below a certain point. In this study, for example, maximum fitness (w_{max}) above 1.2 and dispersal rate (d) below 0.15 could be advised to maintain the species.

This result gives an idea of planning in a view of evolution of the metapopulation. However, most research and applications in conservation planning nowadays are focused on the biodiversity patterns (Cowling & Pressey 2001). The evolutionary processes should also be noticed. With the computer simulation technology, planning for an evolutionary future could become a new trend in conservation planning.

For future study in this field, computer simulation would play an important role. It has huge potential in predicting evolutionary process, recurring history evolution and assessing the impact of landscape properties on genetic diversity (Aberer & Stamatakis 2013). A combination with geographic information platform to establish a dynamic model (Costanza & Voinov 2004) could also be a feasible way to support the research more vividly.

Reference

Aberer, A.J. & Stamatakis, A. (2013). Rapid forward-in-time simulation at the chromosome and genome level. *BMC Bioinformatics*, 14, 216.

Balloux, F. (2001). EASYPOP (Version 1.7): A Computer Program for Population Genetics Simulations. *J. Hered.*, 92, 301–302.

Costanza, R. & Voinov, A. (2004). *Landscape Simulation Modeling: A spatially explict, dynamic approach*. Springer New York.

Cowling, R.M. & Pressey, R.L. (2001). Rapid plant diversification: Planning for an evolutionary future. *Proc. Natl. Acad. Sci.*, 98, 5452–5457.

Freeman, S. & Herron, J.C. (2007). *Evolutionary analysis*. *Am. Biol. Teach.* 4th edn. Pearson Education Inc., New York.

Goudet, J. & Jombart, T. (2015). hierfstat: Estimation and Tests of Hierarchical F-Statistics.

Hardy, G. (1908). Mendelian Proportions in a Mixed Population. *Science (80-.).*, 28, 49–50.

Hartl, D.L. & Grant, A. (1997). *Principles of Population Genetics. 4th Edition*. *Sinauer Assoc.* 4th edn. Sinauer Associates.

Holderegger, R. (2008). Landscape Genetics, 58, 199.

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

Levins, R. (1970). Extinctions. In: *Some mathematical problems in biology* (ed. Gerstenhaber, M.). American Mathematical Society, Rhode Island, USA, pp. 77–107.

MacArthur, R.H. & Wilson, E.O. (1967). *The Theory of Island Biogeography*. Princetor:Princeton University Press.

Manel, S., Schwartz, M.K., Luikart, G. & Taberlet, P. (2003). Landscape genetics: Combining landscape ecology and population genetics. *Trends Ecol. Evol.*, 18, 189–197.

Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proc. Nat. Acad. Sci*, 70, 3321–3323.

Peck, S.L. (2004). Simulation as experiment: A philosophical reassessment for biological modeling. *Trends Ecol. Evol.*, 19, 530–534.

R Core Team. (2017). R: A Language and Environment for Statistical Computing.

van Strien, M.J., Keller, D., Holderegger, R., Ghazoul, J., Kienast, F. & Bolliger, J. (2014). Landscape genetics as a tool for conservation planning: predicting the effects of landscape change on gene flow. *Ecol. Appl.*, 24, 327–339.

Weinberg, W. (1908). Über Vererbungsgesetze beim Menschen. *Mol. Genet. Genomics*, 1, 377–392.

Wright, S. (1931). Evolution in Mendelian Populations. *Genetics*, 16, 97–159.