

# GESP: A computer program for modelling genetic effective population size, inbreeding and divergence in substructured populations

Fredrik Olsson<sup>1</sup> | Linda Laikre<sup>2</sup> | Ola Hössjer<sup>1</sup> | Nils Ryman<sup>2</sup>

<sup>1</sup>Department of Mathematics, Division of Mathematical Statistics, Stockholm University, Stockholm, Sweden

<sup>2</sup>Department of Zoology, Division of Population Genetics, Stockholm University, Stockholm, Sweden

## Correspondence

Fredrik Olsson, Department of Mathematics, Division of Mathematical Statistics, Stockholm University, Stockholm, Sweden.  
Email: GESP.info@gmail.com

## Present address

Fredrik Olsson, Department for Development of Processes and Methods, Statistics Sweden, Stockholm, Sweden

## Funding information

Swedish Research Council Formas; BONUS; EU; Swedish Environmental Protection Agency; Swedish Research Council, Grant/Award Number: 621-2008-4946, 621-2013-4633; Gustafsson Foundation for Research in Natural Sciences and Medicine

## Abstract

The genetically effective population size ( $N_e$ ) is of key importance for quantifying rates of inbreeding and genetic drift and is often used in conservation management to set targets for genetic viability. The concept was developed for single, isolated populations and the mathematical means for analysing the expected  $N_e$  in complex, subdivided populations have previously not been available. We recently developed such analytical theory and central parts of that work have now been incorporated into a freely available software tool presented here. GESP (Genetic Effective population size, inbreeding and divergence in Substructured Populations) is R-based and designed to model short- and long-term patterns of genetic differentiation and effective population size of subdivided populations. The algorithms performed by GESP allow exact computation of global and local inbreeding and eigenvalue effective population size, predictions of genetic divergence among populations ( $G_{ST}$ ) as well as departures from random mating ( $F_{IS}$ ,  $F_{IT}$ ) while varying (i) subpopulation census and effective size, separately or including trend of the global population size, (ii) rate and direction of migration between all pairs of subpopulations, (iii) degree of relatedness and divergence among subpopulations, (iv) ploidy (haploid or diploid) and (v) degree of selfing. Here, we describe GESP and exemplify its use in conservation genetics modelling.

## KEYWORDS

eigenvalue effective size, inbreeding coefficient, inbreeding effective size, kinship coefficient, metapopulation effective size, migration, software, subpopulation differentiation

## 1 | INTRODUCTION

Assessing and monitoring the expected rate of loss of genetic variation and the degree of population differentiation is of key importance in molecular ecology and conservation genetics. It is therefore important to quantify inbreeding within individuals and kinship coefficients between them as a function of time. One fundamental parameter in this respect is the genetically effective

population size  $N_e$  (Wright, 1931, 1938), defined as the size of an ideal population exhibiting the same rate of increased inbreeding as the nonideal population under study.  $N_e$  is a key tool in conservation genetics, but guidelines are based on models referring to a single, isolated population of constant size (Allendorf & Ryman, 2002; Franklin, 1980; Soulé, 1986; Traill, Brook, Frankham, & Bradshaw, 2010). This is primarily because much less is known about the behaviour of  $N_e$  in substructured populations (i.e. metapopulation-

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2017 The Authors. *Molecular Ecology Resources* Published by John Wiley & Sons Ltd.

$N_e$ ) as compared to isolated homogeneous ones (Wang & Caballero, 1999; Waples, 2010). This is in spite of a long history of theoretical developments for  $N_e$  of substructured populations that includes effects of symmetrical migration between subpopulations of the island model (Wright, 1951) and stepping stone migration models (Kimura, 1953; Weiss & Kimura, 1965), impact of strong migration (Nagylaki, 1980) and relationships between different types of  $N_e$  (Whitlock & Barton, 1997).

The coefficient of gene differentiation  $G_{ST}$  of Nei (1973) is another essential parameter that quantifies the proportion of genetic variation due to genetic differences between subpopulations. It extends the fixation index  $F_{ST}$  (Wright, 1943, 1951) to multiallelic and multilocus situations. In some applications, it is also important to quantify the degree of nonrandom mating in terms of departures from Hardy–Weinberg proportions, either within subpopulations ( $F_{IS}$ ) or within the total population ( $F_{IT}$ ).

We have recently developed a unified mathematical framework for haploid and diploid structured populations that can be used to compute expected inbreeding and kinship coefficients, effective population size, genetic divergence and departures from random mating in populations that consist of various numbers of more or less interconnected subpopulations whose size may vary over space and time (Hössjer, Laikre, & Ryman, 2016; Hössjer, Olsson, Laikre, & Ryman, 2014, 2015). This newly developed theory allows computation and modelling of parameters of complex metapopulations that has previously not been possible. It is applicable to selectively neutral inheritance at Y-chromosomes, mitochondrial DNA (haploid populations) and autosomes (diploid populations).

Here, we present a computer program *GESP* (Genetic Effective population size, inbreeding and divergence in Substructured Populations) that performs several of the analytical computations outlined in Hössjer et al. (2014, 2015). *GESP* can be used to model exact local and global rates of inbreeding, haploid and diploid inbreeding as well as eigenvalue effective size, and population divergence in a substructured population. *GESP* focuses on geographic subdivision and ignores other types of structure such as overlapping generations.

Several software exist for addressing various issues relating to genetically effective population size, and they can be classified into the following three categories: computer programs that (i) simulate drift or other processes and produce data that can be used for estimating  $N_e$ , (ii) estimate  $N_e$  from empirical data and (iii) compute predictions of  $N_e$  from demographic parameters by an exact algorithm. Examples of category 1 software include simulation programs *POPSIM* (Hampe, Wienker, Schreiber, & Nürnberg, 1998), *VORTEX* (Lacey, 2000) and *EASYPop* (Balloux, 2001). *TEMPOFS* (Jorde & Ryman, 2007), *LDNE* (Waples & Do, 2008), *ONESAMP* (Tallmon, Koyuk, Luikart, & Beaumont, 2008), *GONE* (Coombs, Letcher, & Nislow, 2012) and *NEESTIMATOR* (Do et al., 2014) represent the second category of programs that estimate  $N_e$  from empirical genotype or allele frequency data. Category 3 includes *GESP* and other programs, such as *AGENE* (Waples, Do, & Chopelet, 2011), that iteratively compute a forward prediction of  $N_e$ . However, in contrast to *GESP*, *AGENE* focuses on a single, isolated population with age structure. *GESP* complements *AGENE* and other

available software for  $N_e$ -modelling by performing exact calculations for spatially substructured populations using theory that has not previously been available. No other program currently performs such computations.

*GESP* can be used to model, for example metapopulation- $N_e$  and such modelling can use simulated, empirical or hypothetical estimates of local effective size, inbreeding levels and rates of migration as input. This can aid in avoiding incorrect conservation management recommendations for  $N_e$  of substructured populations as have been reported (Holley et al., 2014). For instance, we have recently applied *GESP* to the case study of the Fennoscandian wolf metapopulation to address questions of how large local effective sizes and what rates of gene flow that are needed to reach conservation genetic goals (Laikre, Olsson, Jansson, Hössjer, & Ryman, 2016).

Below we describe *GESP*, its parameters, and exemplify how the program can be used to aid researchers of molecular ecology and conservation genetics to explore the impact that various future demographic scenarios may have on effective size, inbreeding and subpopulation differentiation.

## 2 | WHAT *GESP* Does

In this section, we briefly introduce notation and describe the mathematics behind *GESP*. The metapopulation is assumed to have  $s$  subpopulations. The theory allows  $s$  to change over time (Hössjer et al., 2014, 2015), but in the current implementation of *GESP*, it is constant. On the other hand, the sizes of all subpopulations or the migration rates between them can differ and vary over time. We consider a selectively neutral and polymorphic locus and study how the genetic composition of the population at this locus is expected to evolve over discrete time steps  $t = 0, 1, \dots, t_{\max}$ , typically generations. Here,  $t = 0$  represents the present,  $t > 0$  the future and  $t_{\max} + 1$  is the number of time points.

### 2.1 | Identity-by-descent parameters and their recursions

Let  $\mathcal{A}$  be the set of all types of gene pairs, that is pairs of alleles. The most crucial building block of *GESP* is a number of probabilities

$$f_{ta} = 1 - h_{ta} \quad (1)$$

that a gene pair of type  $a \in \mathcal{A}$  is identical by descent IBD (or identical by state, IBS) if drawn randomly from the population at time  $t$ . For a haploid population, where each individual has a single gene copy, there are  $d = s^2$  different types  $a = ij$  that specify the ordered pair  $i$  and  $j$  of subpopulations to which the two genes belong. We then refer to  $f_{tij}$  as an average inbreeding coefficient between subpopulations  $i$  and  $j$  at time  $t$ . For a diploid population, each individual carries two homologous genes. Whenever the two genes are drawn from the same subpopulation  $i$ , we must distinguish whether they belong to the same ( $a = i$ ) or to different ( $a = ii$ ) individuals. This gives a total of  $d = s^2 + s$  diploid gene pair types. The corresponding

quantities in Equation (1) are referred to as average inbreeding coefficients ( $a = i$ ), average kinship coefficients within subpopulations ( $a = ii$ ) or average kinship coefficients between subpopulations ( $a = ij, i \neq j$ ).

An important aspect of GESP is to use matrix analytic methods to describe time progression of the inbreeding and kinship coefficients, with genetic drift, migration and mutation as the three forces of genetic change. In the haploid as well as the diploid case, this is achieved by gathering all non-IBD probabilities at time  $t$  into a column vector  $\mathbf{h}_t = (h_{ta}, a \in \mathcal{A})'$  of length  $d$ . If the two genes are drawn without replacement, the vector of non-IBD probabilities obeys a linear recursion

$$\mathbf{h}_t = (1 - \mu)^2 \mathbf{D}_t \mathbf{h}_{t-1} + [\mathbf{1} - (1 - \mu)^2] \mathbf{1} \quad (2)$$

between time points  $t - 1$  and  $t$ , where  $\mathbf{D}_t$  is a square matrix of order  $d$ ,  $\mathbf{1}$  is a column vector of  $d$  ones and  $\mu$  is the probability that a gamete mutates under an infinite alleles model (Kimura, 1971). A recursion similar to Equation (2) holds if the two genes are drawn with replacement.

For haploid and diploid models, the elements of  $\mathbf{D}_t$  are functions of the local census and effective sizes  $N_{ti}$  and  $N_{eti}$  of all subpopulations  $i$  at time  $t$ , as well as the migration rates  $m_{ji}$  from one subpopulation  $j$  to another subpopulation  $i$  between time points  $t - 1$  and  $t$ . As migration is specified forward in time, from one time point to the next, we refer to  $m_{ji}$  as a forward migration rate. Some additional parameters are needed for diploid models, because reproduction may occur either by selfing or crossing. This requires specifying the rate of selfing as well as whether mating occurs before or after migration.

## 2.2 | Subpopulation weights

When computing, for example the effective population size for the metapopulation as a whole, the contribution from the different subpopulations must be weighted, and this can be done in several ways regarding the weights of separate subpopulations. The predefined weighting schemes in GESP include that the subpopulation weights are either uniform (i.e. the same for all subpopulations), proportional to size, proportional to reproductive value (i.e., populations that contribute more to the system as a whole because of migration rates and patterns are given more weight; cf. Fisher, 1958; Felsenstein, 1971) or allocated to particular subpopulations so that other subpopulations are ignored by receiving zero weights. Which of these weights to use depends on the goals of the investigator. Size proportional weights treat all individuals of the metapopulation equally, reproductive weights corresponds to the long-term behaviour of the system, and local weights focus on one particular subpopulation. Further, it is possible to define user-specified subpopulation weights as any non-negative numbers  $w_1, \dots, w_s$  that sum to one ( $\sum_{i=1}^s w_i = 1$ ). A weighting scheme is global if at least two  $w_i$  are positive, whereas it is local if one subpopulation  $i$  receives full weight ( $w_i = 1$ ). It is convenient to interpret all  $w_i$  as probabilities of sampling genes from the various subpopulations, because this

naturally defines weights  $W_a$  for all  $a$  in terms of probabilities of sampling gene pairs of type  $a$ . This can be done in different ways, and we distinguish between a number of different sampling schemes for gene pairs. In the diploid case, the three most important schemes are  $T$ ,  $S$  and  $I$ . They differ as to whether the two genes are chosen independently from the total ( $T$ ) population (weights  $W_a = W_{Ta}$ ), from the same randomly chosen subpopulation ( $S$ ; weights  $W_a = W_{Sa}$ ) or from the same randomly chosen individual ( $I$ ; weights  $W_a = W_{Ia}$ ). Given that subpopulation weights have been specified, the probability is

$$\begin{aligned} h_{Tt} &= \sum_{a \in \mathcal{A}} W_{Ta} h_{ta} = 1 - f_{Tt} \\ h_{St} &= \sum_{a \in \mathcal{A}} W_{Sa} h_{ta} = 1 - f_{St} \\ h_{It} &= \sum_{a \in \mathcal{A}} W_{Ia} h_{ta} = 1 - f_{It} \end{aligned} \quad (3)$$

that two genes are not IBD, if sampled at time  $t$  by any of the three schemes  $T$ ,  $S$  or  $I$ . If the gene pair is sampled without replacement, the formulas for  $h_{Tt}$  and  $h_{St}$  are the same, whereas the definition of  $h_{It}$  is multiplied by a factor of two.

In the diploid case,  $f_{It}$  is a weighted average of the inbreeding coefficients  $f_{ti}$  of all subpopulations  $i$  at time  $t$  with positive weights, whereas  $f_{St}$  is a weighted average of the inbreeding coefficients  $f_{ti}$  and kinship coefficients  $f_{ii}$  within subpopulations, for all subpopulations  $i$  with positive weights.

## 2.3 | Notions of effective population size in GESP

Some definitions of effective size incorporate mutations (Ewens, 1989; Maruyama & Kimura, 1980). This may be of interest for long-time scenarios, and the theory in Hössjer et al. (2014, 2015) includes the effect of mutations. However, in the present implementation of GESP, we follow the most common approach and assume there are no mutations ( $\mu = 0$ ), or equivalently, pay no attention to mutations in the definition of non-IBD probabilities in Equation (3). These probabilities are used to compute a number of different effective sizes over different time horizons  $[t, t + \tau]$ . In the diploid case, the inbreeding effective size

$$N_{ei}([t, t + \tau]) = \frac{1}{2 \left[ 1 - (h_{t,t+\tau}/h_{tt})^{1/\tau} \right]} \quad (4)$$

quantifies the average rate at which the non-IBD probabilities of a gene pair decreases between  $t$  and  $t + \tau$ , when sampled without replacement according to scheme  $I$  at both time points. Global and local inbreeding effective sizes differ as to whether global and local subpopulation weights  $w_i$  are used. In order to compute Equation (4), it is necessary to define a scenario for how the demography of the population evolves during  $[t, t + \tau]$ , and  $h_{It}$  requires knowledge of kinship and inbreeding coefficients at time  $t$ . The latter can be chosen arbitrarily from simulated or real data. In particular, it is not required that the population is in equilibrium at time  $t$ .

GESP provides inbreeding effective size for the global metapopulation and for separate subpopulations over time intervals, either from the start  $t = 0$  to another specified point  $t$ , or from one time point  $t$  to the next  $t + 1$ . The latter rate of inbreeding from one generation to

the next corresponds to an *instantaneous effective size*. We have previously suggested the term *realized effective size* ( $N_{eR}$ ) for an instantaneous effective size that is determined both by genetic drift and migration (Laikre et al., 2016). If subpopulation  $i$  receives full weight ( $w_i = 1$ ), then  $N_{eRti} = N_{el}([t, t + 1])$  is the realized effective size of  $i$  at time  $t$ . This is a local inbreeding effective size that equals  $N_{eti}$  if  $i$  is isolated. But in general, the two quantities differ, as the local effective size  $N_{eti}$  is an input parameter of the model that is only affected by genetic drift within  $i$  between time points  $t$  and  $t + 1$ , whereas  $N_{eRti}$  is an output parameter that is also influenced by immigration into  $i$  from the other subpopulations between  $t$  and  $t + 1$ . More generally,  $N_{el}([t, t + \tau])$  quantifies the average combined impact of genetic drift and migration for time intervals of any length  $\tau$ , for those subpopulations that are part of the weighting scheme. For management applications, we argue that  $N_{el}$  is a more relevant concept than those notions of  $N_e$  that only include genetic drift, as the effects of migration and drift are hard to separate (and estimate), in particular when the subpopulation structure is cryptic and partly unknown.

The eigenvalue effective population size  $N_{eE}$  gives the long-term equilibrium rate at which inbreeding increases ( $\tau \rightarrow \infty$ ). This requires some additional assumptions, such as time-invariant migration rates and subpopulation sizes, and that no group of subpopulations is isolated, to make  $\mathbf{D}_t = \mathbf{D}$  in Equation (2) time invariant with a unique largest eigenvalue  $\lambda_{\max}(\mathbf{D})$ . The inbreeding effective population size then has the long-term limit

$$\lim_{\tau \rightarrow \infty} N_{el}([t, t + \tau]) = N_{eE} = \frac{1}{2[1 - \lambda_{\max}(\mathbf{D})]}$$

For many types of migration schemes, neither the instantaneous nor the long term  $N_e$  gives the full picture. Rather, the whole curve  $\tau \rightarrow N_{el}([t, t + \tau])$  is needed to capture the rate at which inbreeding increases in a subdivided population.

## 2.4 | Fixation indices

In order to quantify subpopulation differentiation and departures from random mating, we use

$$\begin{aligned} g_{STt} &= (h_{Tt} - h_{St})/h_{Tt} \\ f_{ISt} &= (h_{St} - h_{It})/h_{St} \\ f_{ITt} &= (h_{Tt} - h_{It})/h_{Tt} \end{aligned} \quad (5)$$

to predict the fixation indices  $G_{STt}$ ,  $F_{ISt}$  and  $F_{ITt}$  at time  $t$ . The quantities on the right-hand sides of Equation (5) are all defined in Equation (3) under the assumption that pairs of genes are drawn with replacement. The predicted coefficient of gene differentiation is only applicable for global subpopulation weights, and it satisfies  $0 \leq g_{STt} \leq 1$ , with the lower and upper bounds attained when subpopulations are genetically identical or fully diverged. The other two fixation indices satisfy  $-1 \leq f_{ISt}, f_{ITt} \leq 1$ . A necessary condition for attaining the lower and upper bounds is that all or no individuals have heterozygous genotypes. Random mating and selfing give  $f_{ISt}$  a value close to 0, with a small negative bias caused by a Levene effect (Crow & Kimura, 1970).

**TABLE 1** Population genetic parameters used by GESP. They all apply to a diploid model. Some quantities are slightly different for haploid models, see the reference manual (Olsson, 2017) for details

Symbol	Definition
$s$	Number of subpopulations
$t$	Discrete time point (typically a generation number)
$t_{\max}$	Number of time points after $t = 0$
$N_{ti}$	Local census size of subpopulation $i$ at time $t$
$N_{eti}$	Local effective size of subpopulation $i$ at time $t$ under isolation
$m_{ij}$	Forward migration rate from subpopulation $j$ to subpopulation $i$ between time points $t - 1$ and $t$
$\mu$	Mutation probability per gamete
$\mathcal{A}$	Set of all types of gene pairs
$d$	Number of possible gene pairs
$f_{ti}$	Inbreeding coefficient of individuals of subpopulation $i$ at time $t$
$f_{tij}$	Kinship or coancestry coefficient of two individuals from subpopulations $i$ and $j$ at time $t$
$f_{It}$	Average inbreeding coefficient within individuals at time $t$
$f_{St}$	Average inbreeding/coancestry coefficient within subpopulations at time $t$
$f_{Tt}$	Average inbreeding/coancestry coefficient in the total population at time $t$
$h_{ti}$	$= 1 - f_{ti}$
$h_{tij}$	$= 1 - f_{tij}$
$h_{It}$	$= 1 - f_{It}$
$h_{St}$	$= 1 - f_{St}$
$h_{Tt}$	$= 1 - f_{Tt}$
$\tau$	Length of time interval of genetic drift
$N_{el}([t, t + \tau])$	Inbreeding effective size over time interval $[t, t + \tau]$
$N_{el}([t, t + 1])$	Instantaneous inbreeding effective size over one single generation at time $t$
$N_{eRti}$	Realized effective size of subpopulation $i$ at time $t$ . It is a special case of instantaneous effective size when subpopulation $i$ receives full weight. It includes the effect of genetic drift within $i$ and migration into $i$
$N_{eE}$	Eigenvalue effective size
$G_{STt}$	Coefficient of gene differentiation at time $t$
$g_{STt}$	Prediction of $G_{STt}$
$F_{ISt}$	Fixation index of individuals within subpopulations, time $t$
$f_{ISt}$	Prediction of $F_{ISt}$
$F_{ITt}$	Fixation index of individuals within the total population, time $t$
$f_{ITt}$	Prediction of $F_{ITt}$

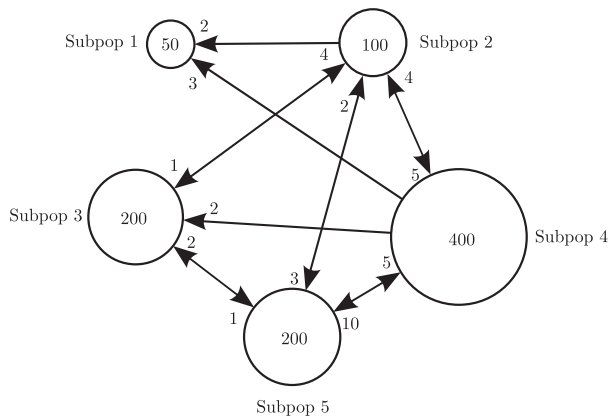
## 3 | PARAMETERS IN GESP

In GESP, all input parameters are specified using the graphical user interface. Table 1 contains a summary of some of the most important quantities used by the program. The output of GESP is shown in the

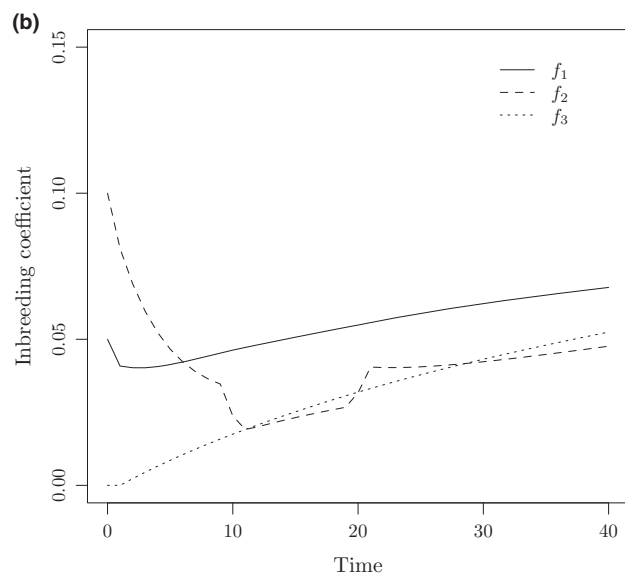
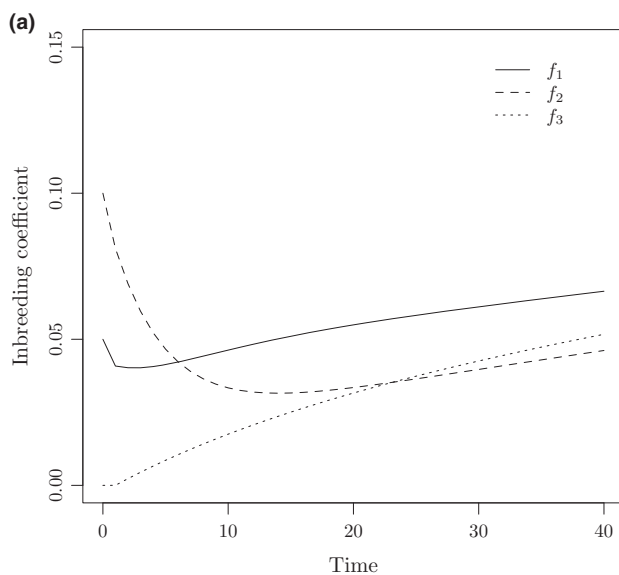
interface. Figures can be saved in various formats, and all results can be exported to a csv-file. All input and output parameters are described in detail in the manual (Olsson, 2017).

#### 4 | GESP in Conservation Genetic Modelling

One of the main purposes of GESP is to analyse how inbreeding dynamics and effective population sizes are affected by various migration scenarios, including populations with varying subpopulation sizes and local bottlenecks. Even though the number of subpopulations is kept fixed, it is still possible to put some local census sizes to zero and thereby incorporate subpopulation extinction and recolonization. Here, we describe an example population in which



**FIGURE 1** Schematic overview of a population divided into five subpopulations. All local census and local effective population sizes are the same, given as numbers inside the circles. The integer at each arrow refers to the number of migrants per generation between this pair of subpopulations



**FIGURE 2** Inbreeding coefficients for subpopulations 1–3 of Figure 1. In the left subplot (a) the subpopulation sizes are constant, whereas in the right subplot (b), the size of subpopulation 2 has been reduced from 100 to 30 between generations 10 and 20 in order to model a local bottleneck see Figure 1

one of the subpopulations exhibits a local bottleneck, although not a complete extinction. The example is further described in the manual (Olsson, 2017), where the model is specified with a step-by-step instruction. See also Laikre et al. (2016) for a case study of the Fennoscandian wolves where the theory that has been incorporated in GESP is used for practical conservation genetic modelling, including suggestions of general conservation genetic targets for metapopulations (the publication is available for download at the GESP website).

Consider a diploid population with no selfing divided into five subpopulations with migration scheme and subpopulation sizes described in Figure 1. Let the initial inbreeding and kinship coefficients be 0.05 for subpopulations 1 and 4, 0.1 for subpopulation 2 and 0 for subpopulations 3 and 5. Starting levels of kinship between subpopulation pairs is zero for all pairs. The time dynamics of the inbreeding coefficients for the first 40 generations are shown in Figure 2a), for subpopulations 1, 2 and 3. Now, we change the size of subpopulation 2 to 30 between generations 10 and 20. This is done by keeping all migration rates fixed, but the number of nonmigrants in subpopulation 2 is reduced from 90 to 20. Figure 2b) displays the effect of this local bottleneck on the inbreeding coefficients for subpopulations 1, 2 and 3.

To summarize, with GESP, it is possible to model a substructured population with a general migration scheme and compute analytical values of, for example, local and global rates of inbreeding, effective population sizes and population divergence.

#### 5 | POTENTIAL FOR FUTURE EXTENSION

At this point, only parts of the Hössjer et al. (2014, 2015, 2016) mathematical framework for modelling various genetic aspects of substructured populations have been implemented in GESP. Thus,

there is potential for further extensions of GESP, that is, to combine geographic structure with overlapping generations, as outlined in Hössjer et al. (2015). Further, we believe it is possible to extend the theory to X-chromosomes, by generalizing results in Nagylaki (1995) for isolated populations to those with geographic subdivision.

## 6 | DOWNLOAD AND USAGE

The program, together with its manual, can be downloaded from the website [www.zoologi.su.se/research/GESP](http://www.zoologi.su.se/research/GESP). The manual (Olsson, 2017) covers information about the installation process, a detailed overview of the interface and a number of examples.

## ACKNOWLEDGEMENTS

The authors thank Sara Kurland for valuable comments on the manuscript and manual and Ingvar Ståhl for extensive testing of the GESP installation process. Linda Laikre and Nils Ryman were funded by the Swedish Research Council Formas, the BONUS BAMBI project supported by BONUS (Art 185), funded jointly by the EU and the Swedish Research Council Formas, the Swedish Environmental Protection Agency, and the Swedish Research Council. Ola Hössjer was supported by the Swedish Research Council, contracts nr. 621-2008-4946 and 621-2013-4633, and the Gustafsson Foundation for Research in Natural Sciences and Medicine. Fredrik Olsson was supported by grants to LL. Comments from a subject editor and six anonymous reviewers are acknowledged as well.

## AUTHOR CONTRIBUTIONS

All the authors wrote the manuscript jointly and participated in designing and testing of the program. F.O. developed the program.

## REFERENCES

- Allendorf, F. W., & Ryman, N. (2002). The role of genetics in population viability analysis. In S. R. Bessinger, & D. R. McCullough (Eds.), *Population viability analysis*. Chicago, IL: The University of Chicago Press.
- Balloux, F. (2001). EASYPOP (version 1.7): A computer program for population genetics simulation. *Journal of Heredity*, 92, 301–302.
- Coombs, J. A., Letcher, B. H., & Nislow, K. H. (2012). GONe: Software for estimating effective population size in species with generational overlap. *Molecular Ecology Resources*, 12, 160–163.
- Crow, J. F., & Kimura, M. (1970). *An introduction to population genetics theory*. Caldwell, NJ, USA: The Blackburn Press.
- Do, C., Waples, R. S., Peel, G., MacBeth, G. M., Tillett, B. J., & Ovenden, 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.
- Ewens, W. J. (1989). The effective population size in the presence of catastrophes. In M. Feldman (Ed.), *Mathematical evolutionary theory* (p. 9). Princeton, NJ: Princeton University.
- Felsenstein, J. (1971). Inbreeding and variance effective numbers in populations with overlapping generations. *Genetics*, 68, 581–597.
- Fisher, R. A. (1958). *The genetical theory of natural selection*, 2nd ed.. New York, NY: Dover.
- Franklin, I. R. (1980). Evolutionary change in small populations. In M. E. Soulé & B.A. Wilcox (Eds.), *Conservation biology: An evolutionary-ecological perspective* (pp. 135–150), Sunderland, MA: Sinauer Associates.
- Hampe, J., Wienker, T., Schreiber, S., & Nürnberg, P. (1998). POPSIM: A general population simulation program. *Bioinformatics*, 14, 458–464.
- Holley, C. E., Nachols, R. A., Whitehead, M. R., Acamack, A. T., Bunn, M. R., & Sherwin, W. B. (2014). Testing single-sample estimators of effective population size in genetically structured populations. *Conservation Genetics*, 15, 23–35.
- Hössjer, O., Laikre, L., & Ryman, N. (2016). Effective sizes and time to migration-drift equilibrium in geographically subdivided populations. *Theoretical Population Biology*, 112, 139–156.
- Hössjer, O., Olsson, F., Laikre, L., & Ryman, N. (2014). A new general analytical approach for modeling patterns of genetic differentiation and effective size of subdivided populations over time. *Mathematical Biosciences*, 258, 113–133. doi:10.1016/j.mbs.2014.10.001
- Hössjer, O., Olsson, F., Laikre, L., & Ryman, N. (2015). Metapopulation inbreeding dynamics, effective size and subpopulation differentiation – A general analytical approach for diploid organisms. *Theoretical Population Biology*, 102, 40–59. doi:10.1016/j.tpb.2015.03.006
- Jorde, P., & Ryman, N. (2007). Unbiased estimator for genetic drift and effective population size. *Genetics*, 177, 927–935.
- Kimura, M. (1953). 'Stepping stone' model of population. *Annual Report of the National Institute of Genetics Japan*, 3, 62–63.
- Kimura, M. (1971). Theoretical foundations of population genetics at the molecular level. *Theoretical Population Biology*, 2, 174–208.
- Lacey, R. C. (2000). Structure of the VORTEX simulation model for population viability analysis. *Ecological Bulletins*, 48, 191–203.
- Laikre, L., Olsson, F., Jansson, E., Hössjer, O., & Ryman, N. (2016). Metapopulation effective size and conservation genetic goals for the Fennoscandian wolf (*Canis lupus*) population. *Heredity*, 117, 279–289.
- Maruyama, T., & Kimura, M. (1980). Genetic variability and effective size when local extinction and recolonization of subpopulations are frequent. *Proceedings of the National Academy of Sciences of the United States of America*, 77, 6710–6714.
- Nagylaki, T. (1980). The strong migration limit in geographically structured populations. *Journal of Mathematical Biology*, 9, 101–114.
- Nagylaki, T. (1995). The inbreeding effective population number in dioecious populations. *Genetics*, 139, 473–485.
- Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America*, 70, 3321–3323.
- Olsson, F. (2017). GESP: A User's Manual. Retrieved from [www.zoologi.su.se/research/GESP](http://www.zoologi.su.se/research/GESP)
- Soulé, M. E. (Ed.) (1986). *Conservation biology. The science of scarcity and diversity*. Sunderland, MA: Sinauer Associates.
- Tallmon, D. A., Koyuk, A., Luikart, G., & Beaumont, M. A. (2008). ONE-SAMP: A program to estimate population size using approximate Bayesian computation. *Molecular Ecology Resources*, 8, 299–301.
- Trails, L. W., Brook, B. W., Frankham, R. R., & Bradshaw, C. J. A. (2010). Pragmatic population viability targets in a rapidly changing world. *Biological Conservation*, 143, 28–34.
- Wang, J. L., & Caballero, A. (1999). Developments in predicting the effective size of subdivided populations. *Heredity*, 82, 212–226.
- Waples, R. (2010). Spatial-temporal stratifications in natural populations and how they affect understanding and estimation of effective population size. *Molecular Ecology Resources*, 10, 785–796.
- Waples, R. S., & Do, C. (2008). LDNE: A program for estimating effective population size from data on linkage disequilibrium. *Molecular Ecology Resources*, 8, 753–756.
- Waples, R. S., Do, C., & Choquet, J. (2011). Calculating  $N_e$  and  $N_e/N$  in age-structured populations: A hybrid Felsenstein-Hill approach. *Ecology*, 92(7), 1513–1522.

- Weiss, G. H., & Kimura, M. (1965). A mathematical analysis of the stepping stone model of genetic correlation. *Journal of Applied Probability*, 2, 129–149.
- Whitlock, M. C., & Barton, N. H. (1997). The effective size of a subdivided population. *Genetics*, 146, 427–441.
- Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, 16, 97–159.
- Wright, S. (1938). Size of population and breeding structure in relation to evolution. *Science*, 87, 430–431.
- Wright, S. (1943). Isolation by distance. *Genetics*, 28, 114–138.
- Wright, S. (1951). The general structure of populations. *Annals of Eugenics*, 15, 323–354.

**How to cite this article:** Olsson F, Laikre L, Hössjer O, Ryman N. GESP: A computer program for modelling genetic effective population size, inbreeding and divergence in substructured populations. *Mol Ecol Resour.* 2017;17:1378–1384.

<https://doi.org/10.1111/1755-0998.12673>