

Introduction.....	2
Installation.....	2
Starting the program.....	2
Loading interfaces.....	2
Running a simulation .....	3
Simulation Input.....	3
a. The Load/Run section .....	3
i. Processes .....	4
ii. Configuration File.....	4
iii. Output directory .....	4
iv. Output files base name.....	4
b. Configuration Info section .....	4
i. Configuration File Name.....	4
ii. Model Name .....	5
c. The Population section.....	5
i. N0 (Newborns) .....	5
ii. Nb/Nc.....	6
iii. Nb/Ne.....	6
iv. Nb .....	6
vi. Ages .....	6
vii. Female,Male Survival.....	6
viii. Female,Male Fecundity .....	6
ix. Force Skip.....	6
x. Litter .....	6
xi. Reproductive cycles.....	7
xii. Monogamous .....	7
xiii. Probability of male birth.....	7
xiv. Population size.....	7
d. The Genome section.....	7
i. Mutation frequency .....	7
ii. Number of microsatellites.....	7
iii. Number of SNPs .....	7
iv. Starting Msat allele total.....	8
e. The Simulation section.....	8
i. Cull method .....	8
ii. Replicates.....	8

iii.	Nb and census adjustment .....	8
iv.	Skip breeding probability .....	8
v.	Cycles of burn-in .....	9
vi.	Start recording at cycle number .....	9

## Introduction

The AgeStructureNb GUI interface offers a user interface to allow easy access to simuPOP-based simulations<sup>1</sup> and the LDNe<sup>2</sup>-based Nb and Ne estimations as implemented in Tiago Antao's python program, AgeStrucureNe, available at <https://github.com/tiagoantao/AgeStructureNe.git>. We also offer an interface for plotting Nb and Ne estimations, and regressions based on the estimations. The original analyses based on Tiago's code, with their applications to many species, are available in several publications, including [refs to AgeStructureNe - based pubs].

Our program offers a separate interface to perform three functions: population simulation, Nb and Ne estimation, and estimate visualization. The genepop file output from a simulation can be loaded into an Nb estimation interface, and in turn, the output from an Nb estimation can be loaded into a visualization interface. The Nb estimation interface can also use any genepop file for input.

## Installation

The program can be downloaded from <https://www.github.com/popgengui/negui>. Click on the button labeled “clone or download”. You can keep the program in any directory to which it can be written. For dependency details and other installation advice, see the README.md file provided with the program files.

## Starting the program

The program is launched using a python 2.7 or python 3.6 executable, invoking the negui.py module. Please see the README.md file for details on different ways to start the program.

## Loading interfaces

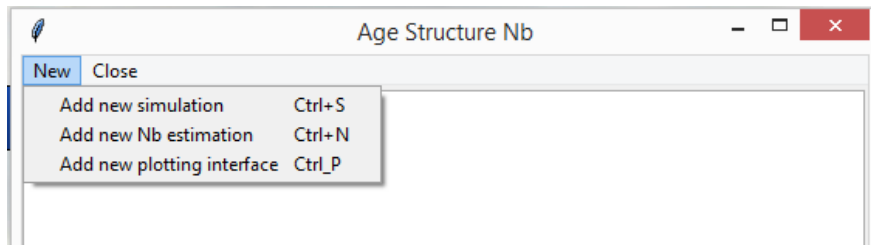
To load one or more of the three interfaces for performing simulations, Nb/Ne estimations, or plotting programs, from the main menu click on the “New” menu (fig. 1). You can load

---

<sup>1</sup> Bo Peng and Marek Kimmel, “SimuPOP: A Forward-Time Population Genetics Simulation Environment,” *Bioinformatics* 21, no. 18 (September 15, 2005): 3686–87, doi:10.1093/bioinformatics/bti584.

<sup>2</sup> Robin S. Waples and Chi Do, “Ldne: A Program for Estimating Effective Population Size from Data on Linkage Disequilibrium,” *Molecular Ecology Resources* 8, no. 4 (July 1, 2008): 753–56, doi:10.1111/j.1755-0998.2007.02061.x. Note that our program uses the beta version 2, the source code generously supplied to us by the authors.

any number of interfaces and run them simultaneously, though running too many at once can tax your computer's CPU and/or memory capacity to a standstill.



*Figure 1 Adding an interface*

## Running a simulation

- a. Load a simulation interface with the add menu (fig. 1), and set the parameters with the provided controls. Steps for preparing the interface to run a simulation follow.
- b. Load a configuration file. The initial simulation interface requires the user to load a configuration file (fig. 2). The user can load a configuration file provided with the program, found inside the configuration/simulation directory inside the main program directory. Note that you can open one of these files and change the parameters manually, if you prefer it to setting them in the interface
- c. Adjust the simulation parameters. With a configuration file loaded (fig. 3) you can change the values in the editable controls. These are detailed below in the “Simulation Input” section.
- d. Click the button labeled “run simulation,” and the simulation will start. The button’s text now changes to say “cancel simulation,” and next to it a new label notes that a simulation is in progress. While the simulation is in progress, the parameter controls are disabled.

## Simulation Input

The simulation interface has 5 subframes that divide the parameters by category.

- a. **The Load/Run section** (fig. 2) of the Simulation interface offers parameters related to input and output files.

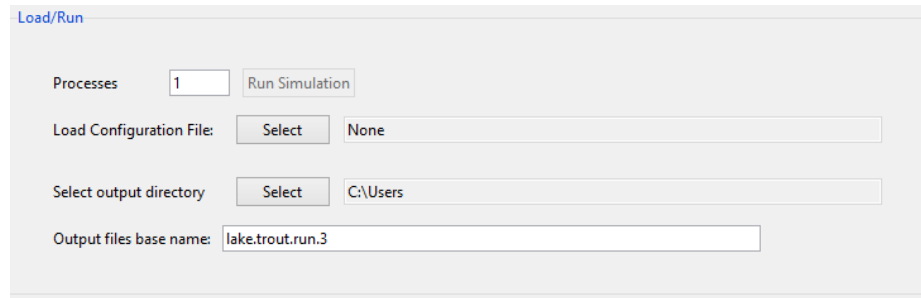


Figure 2 Simulation interface, Load/Run section.

- i. **Processes**, with valid values between 1 and the total number of virtual cores in your computer. Multiple processes are only useful if you have set the “Replicates” parameter (see the Simulation subframe details below) to a value greater than one.
  - ii. **Configuration File**. Press the Select button next to the label, *Load Configuration file* to load a configuration file into the interface. We have included configuration files for many species. These can be found in the “configuration\_files/simulation” path inside the main program folder.
  - iii. **Output directory**. Press the select button next to the label, *Select output directory*, select a folder for the output files written by the simulation
  - iv. **Output files base name**. You can type in a base name for the simulation output files. The simulation will prepend this to the \*.genepop, \*.conf, \*\_age\_totals.tsv and \*\_nb\_values.tsv output files (see the “Simulation Output” section below).
- b. **Configuration Info section** (fig. 3). This group simply shows you the input file information and has no settable parameters.

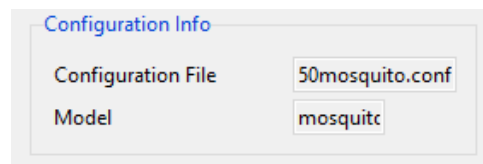


Figure 3 Simulation interface, Configuration info

- i. **Configuration File Name**. This gives the file name of the loaded configuration file.

- ii. **Model Name.** This gives the name of the model parameterized by the configuration file. In our example configuration files, the model name is usually a species' common name.

The screenshot shows the 'Population' section of a simulation interface. It contains several input fields and rows of values:

- N0 (Newborns):** 32
- Nb:** 206
- Nb/Nc:** 0.986
- Nb/Ne:** 0.267
- Nb Tolerance:** 0.05
- Ages:** 32
- Female relative fecundity:** A row of 10 input fields, all containing 0.0.
- Male relative fecundity:** A row of 10 input fields, all containing 0.0.
- Force skip:** 0
- Reproductive cycles:** 100
- Monogamous:** An unchecked checkbox.
- Litter:** None
- Probability male birth:** 0.5
- Population size:** 1505
- Female survival:** A row of 10 input fields, all containing 0.85.
- Male survival:** A row of 10 input fields, all containing 0.85.

Figure 4, Simulation interface, population section

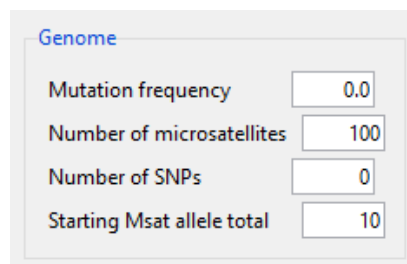
- c. **The Population section** (fig. 4) offers many parameter settings that characterize the population's size and fecundity.
- i. **N0 (Newborns).** This gives the number of newborns added at each simulated reproductive cycle. This value is not editable directly, but is calculated using several values, all of which are editable. These including Nb, Nb/Nc, survival rates, and the probability of male birth. The N0 is recalculated whenever any of these values changes, using the following procedure:
1. Assign an Nc value, as Nb divided by Nb/Nc.
  2. Assign a current\_male\_proportion equal to the probability of male birth.
  3. Assign a current\_female\_proportion equal to 1 – the probability of male birth.
  4. Assign a cumulative\_proportion=1.
  5. For each age value age\_val giving a male and female survival rate:
    - a. Update, current\_male\_proportion=current\_male\_proportion x male\_survival at age\_val.
    - b. Update, current\_female=current\_remale x female\_survival at age\_val.
    - c. Update cumulative\_proportion =cumulative\_proportion + current\_male\_proportion.

- d. Update  $\text{cumulative\_proportion} = \text{cumulative\_proportion} + \text{current\_female\_proportion}$ .
6. Set  $N_0 = N_c / \text{cumulative\_proportion}$ , rounding it to the nearest integer.
- ii. **Nb/Nc** is the effective number of breeders in one reproductive cycle divided by the cohort size.
- iii. **Nb/Ne** is the ratio of the effective number of breeders in one reproductive cycle to the effective population size per generation. This value is not used in the simulation itself, but is written to the output genepop file, and can be used in the Nb estimation interface to make a bias correction in the LDNe estimation of Nb (see the section “Nb estimation input,” below).
- iv. **Nb** is the effective number of breeders in one reproductive cycle.
- v. **Nb Tolerance** is the proportion of the *Nb* parameter by which a new population is added to the simulation after at each reproductive cycle. For example, if the *Nb* is set at 600, and the *Nb Tolerance* is set at 0.02 then populations created at each reproductive cycle (after the burn-in period, explained below), must have an *Nb* value, as calculated using the parentage analysis without parents (PwoP) procedure<sup>3</sup>.
- vi. **Ages** gives the number of age classes for the population to be simulated. Note that this is disabled, and that the length of the lists for *Fecundity* and *Survival* values (see below) are set to length *Ages* minus one for the former and *Ages* minus two for the latter. The age value and changes in these lists, therefore, need to be edited in either a life table or configuration file (see the section “Manually editing life tables and configuration files.”)
- vii. **Female, Male Survival** are lists whose  $i^{\text{th}}$  item gives the probability of survival for an individual of the  $i^{\text{th}}$  age category.
- viii. **Female, Male Fecundity** are lists whose  $i^{\text{th}}$  item gives the probability of reproducing for individuals of the  $i^{\text{th}}$  age category.
- ix. **Force Skip** gives a probability, for each non-zero value,  $f_a$ , in the female fecundity list, that during a given reproductive cycle  $r$ , the value will be replaced with zero. Such replacement means that females belonging to the age class  $a$ , given by  $f_a$ , for cycle  $r$ , are infertile. This parameter is set (assigned a non-zero value) in only a few of the configuration files we copied from the AgeStructureNe program, and we have not made it editable in our interface. In your own custom configuration files you can set it to any value 0 through 100 (the value shown in the interface will be the file’s value divided by 100).
- x. **Litter**, if not a “None” value, will be a list of integers, affecting litter sizes. Note that we do not allow interface editing of these parameters, but note that, as above for the *Force Skip* setting, you can enter this parameter value in a configuration file. This should be a list, and can have one of 2 valid configurations:
  - a. The list can have a single value  $l$ , and  $l < 0$ , then at each reproductive cycle the maximum possible number of offspring available to each reproducing female is given by  $l * -1$ .
  - b. Otherwise, the list should have (positive) integers. In this case these integers proportionally allot litter sizes, as given by their indices in the

<sup>3</sup> Robin S. Waples and Ryan K. Waples, “Inbreeding Effective Population Size and Parentage Analysis without Parents,” *Molecular Ecology Resources* 11 (March 1, 2011): 162–71, doi:10.1111/j.1755-0998.2010.02942.x.

list. In particular, at each reproductive cycle, as a female is chosen to mate:

- i. An age  $a$  is chosen randomly.
  - ii. A female  $f_a$  is chosen randomly from the females of age  $a$ .
  - iii. A list index  $i$  (i.e. one of  $1, 2, 3 \dots n$ , where  $n$  is the number of items in the *litter* list), is selected by weighted probability, proportionally according to the ratio of each list value to the sum of the list values.
  - iv. Female  $f_a$  is then the mother of the next  $i$  offspring (i.e. the female selection steps are skipped for the next  $i$  pairings, since  $f_a$  is the female of the pair.). Thus, she will parent the next  $i$  offspring, unless the  $j$ th of her offspring assignments produces the maximum total offspring for the cycle (i.e.  $NO$  is reached), and  $j < i$ .
- xi. **Reproductive cycles** shows the total number of reproductive cycles that will be simulated.
- xii. **Monogamous**, when checked, tells the simulation to enforce monogamy.
- xiii. **Probability of male birth** is used during reproductive cycles to determine the sex of new individuals. As noted above in the description of the  $NO$ , it is used to in the  $NO$  calculation, and so the latter is recalculated when this value is changed. When the *Cull method* parameter the *Simulation* section is set to *equal\_sex\_ratio*, this parameter is automatically set to 0.5, and it's entry box is disabled.
- xiv. **Population size** shows the number of individuals that will be created in the simulation's initial population. Thereafter the size will change according to the reproductive parameters, notably  $NO$ .
- d. **The Genome section** (fig 5 ) parameters determine the simulated individuals' allelic content.



Genome	
Mutation frequency	0.0
Number of microsatellites	100
Number of SNPs	0
Starting Msat allele total	10

Figure 5 Simulation interface, genome section

- i. **Mutation frequency**, if non-zero, is applied to microsatellites (not to SNPs). It will be used to set the simuPOP simulation StepwiseMutater's rate parameter.
- ii. **Number of microsatellites**, simulated as diploid. Note that in cases in which you specify both  $n$  microsatellites and  $m$  SNPs, in the output genepop file, the first  $n$  loci are the microsatellites and the last  $m$  loci are the SNPs.
- iii. **Number of SNPs**, simulated as diploid.

- iv. **Starting Msat allele total** gives the initial number of microsatellite alleles present in each microsatellite in the initial population. For each microsatellite, the initial genotype frequencies are drawn from the Dirichlet distribution. With Msat total 10, for example, each microsatellite will have 10 alleles with frequencies given by the Dirichlet distribution of order 10, with alpha's uniformly set to 1.0. The maximum allowed number of alleles is 100.
- e. **The Simulation section** (fig. 6) parameters determine several per-cycle behaviours:

Figure 6, Simulation interface, simulation section

- i. **Cull method** indicates one of two possible per-cycle methods whereby individuals are removed from the population.
- ii. **Replicates** sets the number of independent simulations run with the current parameter set. These can be run in parallel if you specify more than one process in the *Processes* parameter in the Load/Run section.
- iii. **Nb and census adjustment** offers one or more specifications that will change the target Nb and the number of individuals in the population by a fixed rate and at a range of cycles (one or more). Entries are of the form min-max:rate, specifying a change in Nb and census size applied at cycle numbers min through max. The values conform to  $\text{min} \leq 2 \leq \text{max}$ , and  $\text{rate} \geq 0.0$ . No adjustment is made with  $\text{rate} = 0.0$ . For example, to reduce the Nb and the total number of individuals by a tenth at cycle 3, you would edit the entry to read, 3-3:0.1. The adjustments are different, depending whether the rate is less than or greater than 1.0.
  1. If the rate is less than 1.0, the target Nb value, and each age class in the current census is reduced by the proportion given by rate. Note that the change in Nb will result in a change to *NO* as described above the Population section's description of *NO*.
  2. If the rate is more than 1.0, the target Nb value will be multiplied by the rate, with a resulting recalculation of *NO*. No change will be made to the current census.
- iv. **Skip breeding probability**, if its value is not "None," should be a list of percentages. It effects the number of available females of a given age at a given cycle number  $c$ . The  $i$ th percent  $p_i$  gives the probability ( $p/100$ ) that a female of age= $i$ , is not able to breed in cycle  $c$ . Like the *Litter* and *Force Skip*



parameters described in the *Population section*, this parameter is not settable in the interface, but can be included in your configuration file.

- v. **Cycles of burn-in**, give an integer  $n$  in the range  $1 \leq n \leq r$ , with  $r$  giving the total *Reproductive cycles*. This value tells simulation that the Nb tolerance test (see the *Nb tolerance* parameter description in the *Population section*) should not be performed for the first  $n$  cycles. The default value for this parameter equals the number of ages in the model.
- vi. **Start recording at cycle number**  $c$  will result in the genepop file containing only the populations of cycles  $c$  through  $r$ ,  $r = \text{total } \textit{Reproductive cycles}$ . This can greatly reduce the size of the output genepop file, when you are interested only in the last  $r-c$  cycles, but want to simulate many cycles before recording, and/or have large populations/loci to simulate.