Lucía Martín Fernández

Simulating the evolution of a single neutral mutation

Computational Tools in Evolutionary Biology – Msc. in Computational Biology

Estimate empirically how the time to fixation (*tfix*) and fixation probability (*pfix*) of a single neutral allele changes with population and bottleneck size.

Introduction

A single neutral allele does not provide any evolutionary advantage. This means, that its evolution does not dependent on selection and depends only on random genetic drift which can be measured with bottleneck size. The purpose of this simulation is to evaluate whether and when this neutral allele gets fixed. In population genetics a fixed allele is referred to that allele that is present in homozygosis in all individuals in a population. The goal of this model its to evaluate the relationship of the time and probability of fixation for this allele taking into account different population and bottleneck sizes.

Methodology

With the objective previously mention, a simulation to assess the behavior of a single neutral mutated allele was programmed using R language. The code to run this simulation, attached to this document, is based on scripts provided by Dr. Alejandro Couce for the Computational Tools in Evolutionary Biology course from the Msc. in Computational Biology. However, for this experiment and in order to simplify the simulation, the mutation rate was set to 0 and only one mutation was introduced in the population in generation 0. In other words, the simulation starts with a population in which only one individual has the neutral mutated allele and the evolution of this mutation only depends on the randomness of neutral selection. In this context, only one locus is being studied for 10.000 days. Moreover, the simulation is run 100.000 times and the estimated values are calculated as the mean of the resulting values of all of these replicates.

As previously stated, the experiment aims to calculate the fixation probability and fixation time of a neutral mutation with different bottleneck and population sizes. The allele is considered to be fixed when all the population has the mutation and is considered lost when no one in the population has maintained this allele. As a result, the fixation time is the number of days the allele takes to get fixed. In this report, the fixation probability is calculated as the number of times the mutation is fixed divided by the total number of replicates in each situation. For this study, these parameters were calculated for populations of 100, 300, 1000 and 3000 invidious combined with 1, 0.5 and 0.25 bottleneck sizes.

Results

As a result, the fixation probability and time are obtained for each bottleneck (0.25, 0.5, 1) and population size (100, 300, 1000 and 3000) combination. In Table 1 final results of the simulation after 10.000 days with 100.000 replicates are summarized.

Table 1. **Fixation probability and fixation time values for each bottleneck and population size. A.** Fixation probability (P_{fix}) depending on bottleneck and population size. P_{fix} is calculated as the total number of times the mutated allele is fixes divided by the total number of replicates. **B.** Fixation time (T_{fix}) depending on bottleneck and population size. T_{fix} is represented in days and is calculated as the mean fixation time for all 100.000 replicates.

A **FIXATION PROBABILITY**

Population size \ Bottleneck size	0.25	0.5	1
100	0.01036	0.00963	0.00970
300	0.00314	0.00337	0.00313
1000	0.00097	0.00081	0.00075
3000	0.00032	0.00036	0.00024

B **FIXATION TIME**

Population size \ Bottleneck size	0.25	0.5	1
100	47.5695	97.75182	191.6021
300	153.8439	311.39169	586.7029
1000	480.9381	1010.64198	1834.9067
3000	1523.4062	3217.50000	5630.2083

Discussion

Once the results were obtained, the behavior of the studied parameters was assessed graphically (Figure 1). Fixation time against bottleneck size (Figure 1A), fixation time against population size (Figure 1B), fixation probability against bottleneck size (Figure 1C) and fixation probability against population size (Figure 1D) were represented. In order to evaluate how different parameters are related to fixation probability and time different regression models were fitted to the data. For each case, points were adjusted to a regression model.

Results show that fixation time is related following to both bottleneck and population size following a linear regression model with $R^2 \sim 1$ (Figure 1A,B). As population size increases, so does the fixation time. Same thing happens for bottleneck size, as it gets bigger, the time to fix a mutated neutral allele also increases. Moreover, it is observed that when the bottleneck is 1, the time taken by an individual to get fixed is approximately two times its population size (2N). When observing what happens for the rest of the bottleneck sizes, it is deduced that fixation time follows the mathematical rule:

$$Tfix = 2 * population\ size * bottleneck\ size.$$

On the other hand, the fixation probability only seems to be related to the population size. As it is represented in Figure 1C, the fixation probability does not vary when bottleneck size changes. Thus, no regression model has been fitted to these results. Hence, why no regression equation or R^2 values are represented. Moreover, fixation probability and population sizes relationship is proven to adjust to a log-log regression model Figure 1D.

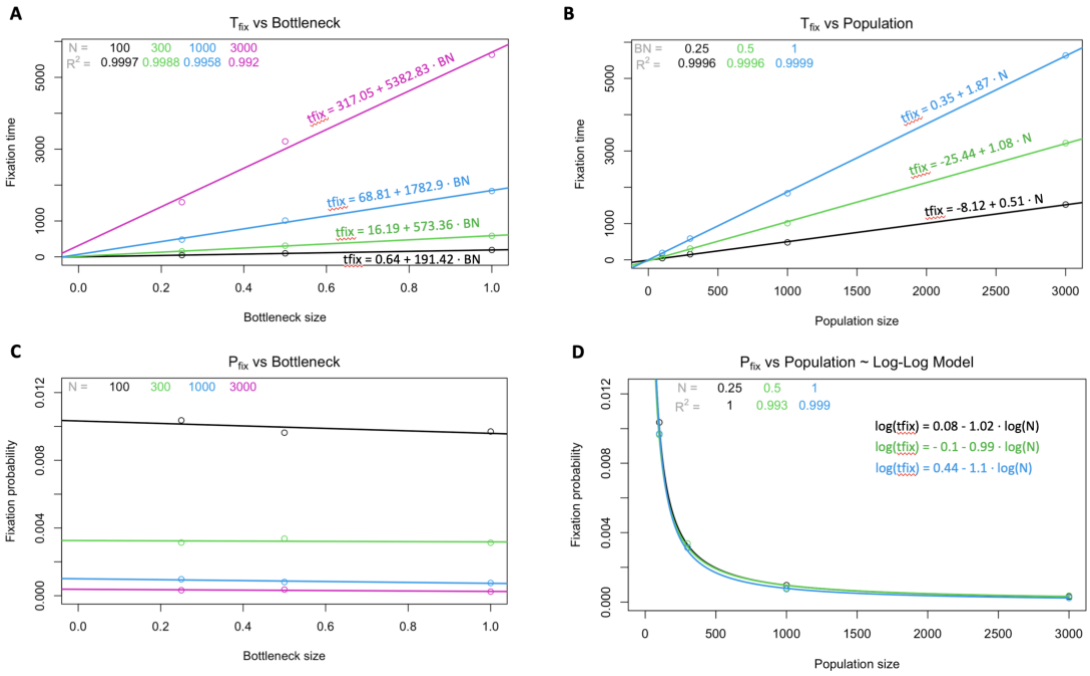

Figure 1. **Graphical representation of fixation time and fixation probability against bottleneck and population size.** Samples are represented in different colors depending on the parameter not studied in each plot. Color code legend includes the corresponding regression equations and R^2 values. **A.** Fixation times against bottleneck size is adjusted to a linear regression model for each population size. **B.** Fixation times against population size is adjusted to a linear regression model for each bottleneck size. **C.** Fixation probability against bottleneck size is not adjusted to any regression model, no equation nor R^2 values are depicted in this case. **D.** Fixation probability against population size is adjusted to a log-log regression model for each bottleneck size.

Conclusion

Given the results presented above, it can be concluded that bigger fixation times are obtained with higher population and bottleneck sizes following a linear model. Moreover, a higher fixation probability are related to smaller population sizes following a log-log regression model. However, bottleneck size is not related with fixation probability.

Limitations and possible extension

This model represents a fictional situation where specifics parameters are studied and other factors may affect the evolution of the studied allele. Moreover, the model considers an immortal population which does not exist in real life. In addition, this simulation is simplified and only works for asexual individuals. In addition, a statistical analysis including standard deviations for each calculated parameter could be applied to the graphical representations.