

# Exercise 1: Sergio López Padilla

## 1. Introduction

This study aims to analyze the consequences of gene drift on the probability and fixation time of the mutation produced in a single allele. This simulation is performed for population sizes 100, 300, 1000 and 3000; and for bottlenecks 0.25, 0.5 and 1. The value of bottleneck refers to the fraction of individuals that live and have the capacity to reproduce, which does not imply that they do so, since this reproduction of species must be kept constant but is carried out in such a way that an individual could reproduce more than once.

## 2. Script

The script related to this exercise can be found on the following GitHub link:

<https://github.com/shevelp/Comp-Evo>

This script has been developed using RStudio and it is divided in different parts:

- Parameters definition
- Modeling
- Data transformation
- Relationship Analysis
- Plotting

## 3. Results

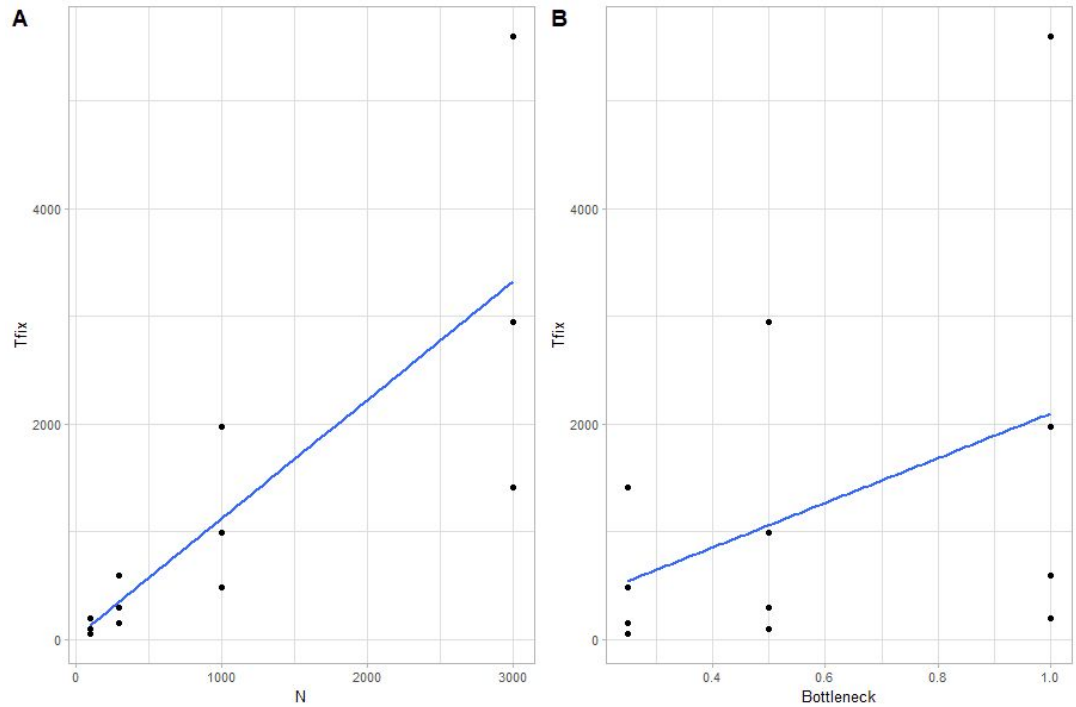
### a. Estimated values

The estimated values related to this exercise are represented in the following table (Table 1). It contains the values related to the time of fixation (Tfix) and fixation probability (Pfix) related to different values in the key parameters: population size (N) and Bottleneck.

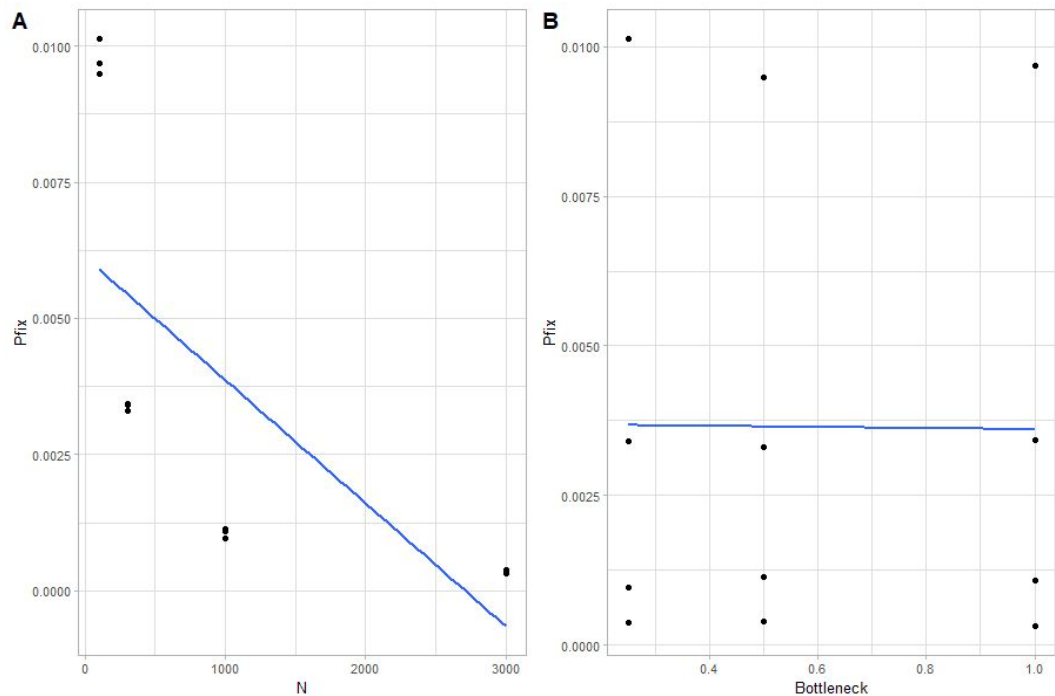
N	Bottleneck	Tfix	Pfix
100	0,25	48,29	0,013
100	0,5	96,4	0,0094
100	1	194,76	0,0096
300	0,25	149,39	0,00341
300	0,5	299,06	0,00331
300	1	593,38	0,0034
1000	0,25	479,72	0,00095
1000	0,5	993,663	0,00113
1000	1	1971,56	0,001
3000	0,25	1410,07	0,00038
3000	0,5	2944,89	0,00039
3000	1	5597,93	0,00031

**Table 1.** Estimated values.

In order to interpret these results more easily, various line charts have been developed, which allow to evaluate the relationship between the key parameters (N and Bottleneck) and Tfix (Figure 1) and Pfix (Figure 2). These analyses show how for Tfix there seems to be a positive correlation for both variables. However for Pfix the population size seems to establish a negative correlation and the bottleneck establishes a linear correlation.



**Figure 1.** Tfix values related to different population size and bottleneck values.



**Figure 2.** Pfix values related to different population size and bottleneck values.

## b. Relationship analysis

To corroborate the type of relationship existing between the variables, the R correlation test (`cor.test`) has been carried out. The results of these correlation tests (Figure 3) are in line with the conclusions obtained from the graphical representation of the results. However, the relationship that seems to be practically neutral between Pfix and bottleneck shows a correlation coefficient of 0.8, which means a remarkable positive correlation.

```
Pearson's product-moment correlation
data: fix$Tfix and fix$Bottleneck
t = 1.4383, df = 10, p-value = 0.1809
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.2097029  0.7982544
sample estimates:
      cor 
0.4140263

> cor.test(x = fix$Pfix, y = fix$Bottleneck, method = "pearson")

Pearson's product-moment correlation
data: fix$Pfix and fix$Bottleneck
t = -0.024481, df = 10, p-value = 0.981
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.5790704  0.5686867
sample estimates:
      cor 
-0.00774145

> cor.test(x = fix$Tfix, y = fix$N, method = "pearson")

Pearson's product-moment correlation
data: fix$Tfix and fix$N
t = 4.3263, df = 10, p-value = 0.001498
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.4349474 0.9438964
sample estimates:
      cor 
0.8073265

> cor.test(x = fix$Tfix, y = fix$N, method = "pearson")

Pearson's product-moment correlation
data: fix$Tfix and fix$N
t = 4.3263, df = 10, p-value = 0.001498
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.4349474 0.9438964
sample estimates:
      cor 
0.8073265
```

**Figure 3.** Results obtained from correlation-test between different variables.

## 4. Considerations

This study does not allow the results to be interpolated to any scenario of genetic drift as a number of considerations must be made. Firstly, the relationship between probability and time of fixation has not been evaluated. It is possible that the estimated parameters may affect each other. On the other hand, from a more theoretical point of view, this study discards the probability of a new mutation, and of random extinction, besides considering only asexual reproduction.

Therefore, as a final conclusion, new simulations should be developed that consider the points mentioned above in order to identify whether the variation of key parameters really affects the rates of gene drift fixation.