

CHOOSE FOCUS ANALYSE (CFA) Exercise

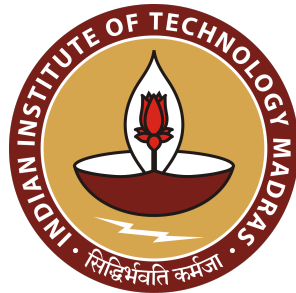
BT5051 Transport Phenomena in Biological Systems

---

**Modeling the diffusion of genes in a population**

---

N Sowmya Manojna | BE17B007  
Department of Biotechnology,  
Indian Institute of Technology, Madras



# Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Concepts Used</b>	<b>3</b>
2.1	Fick's first Law:	3
2.2	Fick's second Law	3
2.3	Continuity Equation	3
<b>3</b>	<b>Methods and Approach</b>	<b>5</b>
3.1	Analytical approach	5
3.2	Mathematical modeling	6
<b>4</b>	<b>Results</b>	<b>7</b>
4.1	Positive Selection of a gene	7
4.1.1	Weak positive selection	7
4.1.2	Moderate positive selection	7
4.1.3	Strong positive selection	8
4.2	No Selection	9
4.3	Negative Selection of a gene	9
4.3.1	Weak negative selection	9
4.3.2	Moderate negative selection	10
4.3.3	Strong negative selection	10
4.4	Cumulative Result	11
<b>5</b>	<b>Inferences</b>	<b>12</b>
<b>6</b>	<b>Conclusion</b>	<b>12</b>
<b>7</b>	<b>Further Avenues</b>	<b>12</b>
<b>8</b>	<b>Acknowledgment</b>	<b>12</b>
<b>9</b>	<b>References</b>	<b>13</b>
	<b>Appendices</b>	<b>14</b>
<b>A</b>	<b>Python code for individual selection possibilities</b>	<b>14</b>
<b>B</b>	<b>Python code for overall selection possibilities</b>	<b>15</b>

# 1 Introduction

*Change is the only constant.*  
- Heraclitus

Evolution can be defined as a change in the heritable component of an organism, mostly mediated by selection. Most living organisms are evolving based on the selection pressure imposed on them. The effects of evolution acting at the individual level, results in discernible changes at the population level.

Almost all coding regions in our DNA are under selection pressure. Positive selection pressure on a gene for very long durations could result in fixation of the gene in the population, while a negative selection pressure on a gene, would result in a decrease or complete removal of the gene in the population.

Variations of the gene that confer an organism with greater fitness are positively selected and the ones which confer the organism with lesser fitness are negatively selected.

The main objective of my CFA project is to model the spread of a gene in a population based on the selection pressures applied, using diffusion as a tool. The factors which are considered in this study are Generation rate and Consumption rate (birth and death rates) of the population. The diffusion of a gene, under different selection possibilities is modeled and studied across 100 generations.

## 2 Concepts Used

Diffusion (*diffundere* - pour out or spread out) can be defined as the net movement of a given substance from a region of higher concentration to a region of lower concentration.<sup>1</sup> It also described as the process by which information, ideas, changes in behavior, etc. spread through a group of people.<sup>2</sup>

Diffusion of an entity within a single phase is mainly driven by concentration gradients of the entity and is hence the primary driving force of diffusion.<sup>1</sup> The fundamental laws that succinctly explain diffusion are the Fick's Laws. The Continuity Equation helps us to study diffusion at a fundamental level.

### 2.1 Fick's first Law:

This law describes the relationship between the concentration gradient and molar flux of a component in a dilute solution of non-reacting solutes.

It states that the mass flux of a component is directly proportional to the negative concentration gradient of the component.

$$J_A^* = -D_A \frac{dc_A}{dt} \quad (1)$$

where,

$J_A^*$  is the molar flux of A in the solution,

$D_A$  is the mass diffusivity of A,

$c_A$  is the concentration of A in the solution

### 2.2 Fick's second Law

This law states that the rate of accumulation of a component in the system is directly proportional to the rate of curvature of the concentration gradient of the component in a dilute solution of non-reacting solutes.<sup>3</sup>

$$\frac{\partial c_A}{\partial t} = D_A \frac{\partial^2 c_A}{\partial x^2} + R_A \quad (2)$$

where,

$D_A$  is the mass diffusivity of A,

$c_A$  is the concentration of A in the solution

$R_A$  is the net rate of production of A

### 2.3 Continuity Equation

This equation is based on the principle of conservation of a particular component in the system considered (when applicable). In this project I have applied the continuity equation to conserve the number of individuals carrying our gene of interest.

Consider a closed system with input and output streams. Variations in the system composition could arise from the following:

- Input into the system ( $\dot{I}$ )
- Output out of the system ( $\dot{O}$ )
- Generation in the system ( $\dot{G}$ )
- Consumption in the system ( $\dot{C}$ )
- Accumulation in the system ( $\dot{A}$ )

$$\frac{\partial A}{\partial t} = \dot{I} - \dot{O} + \dot{G} - \dot{C} \quad (3)$$

## 3 Methods and Approach

### 3.1 Analytical approach

Taking the population under consideration to be uniformly spread on the surface of the earth, the conservation equation is applied on the number of individuals carrying our gene of interest in the population. The diffusion of the gene is modeled as follows from *equation 3*.

$$\frac{\partial A}{\partial t} = \dot{I} - \dot{O} + \dot{G} - \dot{C}$$

Assuming there is no migration and genetic drift in the population and hence, Input and Output rates of the gene, into and out of the system vanishes.

$$\frac{\partial A}{\partial t} = \overset{0}{\dot{I}} - \overset{0}{\dot{O}} + \dot{G} - \dot{C} \quad (4)$$

Taking

- $N_a$ :
  - Concentration of people having our gene of interest (number of people per unit area in spherical coordinates).
- *Generation rate  $\dot{G}$* :
  - Number of live births per thousand of population, having our gene of interest in each generation.
  - First order dependent on the number of people per unit area having our gene of interest.
- *Consumption rate  $\dot{C}$* :
  - Number of deaths per thousand in population, having our gene of interest in each generation.
  - First order dependent on the number of people per unit area having our gene of interest.

The first order rate constants of

- Generation is taken as  $K_g$
- Consumption is taken as  $K_c$

The *equation 4* becomes,

$$R^2 \sin \theta \, d\theta \, d\phi \, \frac{\partial N_a}{\partial t} = (K_g(N_a) - K_c(N_a)) R^2 \sin \theta \, d\theta \, d\phi \quad (5)$$

Dividing by  $R^2 \sin \theta \, d\theta \, d\phi$  on both sides, we get

$$\frac{\partial N_a}{\partial t} = K_g(N_a) - K_c(N_a) \quad (6)$$

Solving the equation analytically, we get

$$N_a = N_{a0} e^{(K_g - K_c)t} \quad (7)$$

### 3.2 Mathematical modeling

In order to visualize the results for different initial conditions, rates of generation and consumption (i.e.)  $N_{a0}$ ,  $K_g$ ,  $K_c$ , mathematical modeling of the system has been done. (Python code written as a part of this project is attached in Appendix A, B). The program simulates the diffusion of the gene across 100 generations.

Initially the total population is assumed to be 100 individuals per unit area. Proportion of people having the advantageous gene is calculated as follows:

$$\text{Proportion Value} = \frac{\text{No. of individuals carrying the gene of interest}}{\text{Total population}} \quad (8)$$

where the total population is defined as follows:

$$\text{Total population} = \text{No. of individuals (carrying the gene of interest + not carrying the gene of interest)} \quad (9)$$

The generation rates ( $K_g$ ) and consumption rates ( $K_c$ ) of the population not carrying our gene of interest are set to be equal to the CBR (*Crude Birth Rate*) and CDR (*Crude Death Rate*) of India respectively.<sup>4</sup>

The Crude Birth Rate can be defined as the number of live births occurring among the population of a given geographical area during a given year, per 1000 mid-year total population of the given geographical area during the same year.<sup>5</sup>

The crude death rate can be defined as the number of deaths occurring among the population of a given geographical area during a given year, per 1,000 mid-year total population of the given geographical area during the same year.<sup>6</sup> Hence,

- $K_g = CBR = 0.0216$
- $K_c = CDR = 0.007$

The proportion of individuals in the population and the number of individuals carrying our gene of interest has been plotted across the number of generations.

## 4 Results

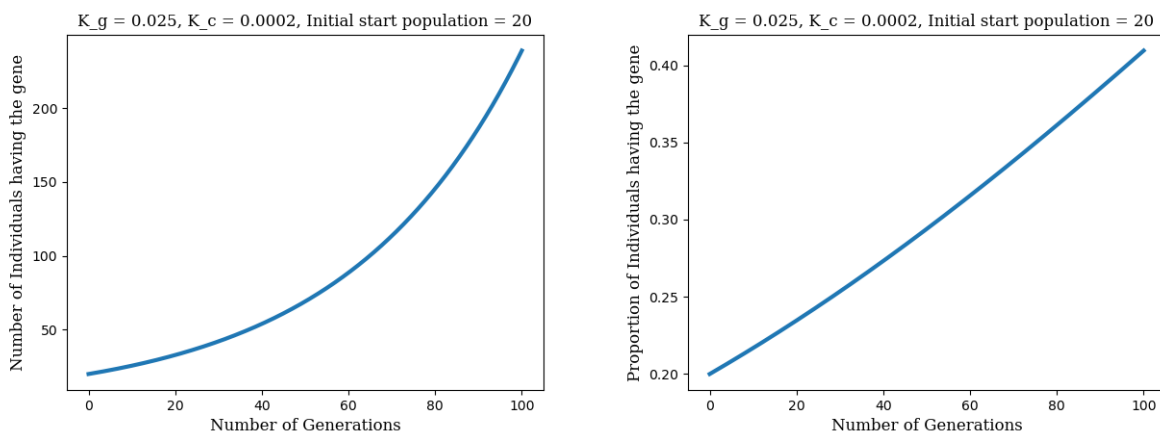
### 4.1 Positive Selection of a gene

Positive selection can be simulated by using a larger birth rate and a smaller death rate for the population having our desired gene of interest than when compared to the overall population.

#### 4.1.1 Weak positive selection

The values chosen for the simulation are:

- $K_g = 0.025$   
1 offspring is produced for every 40 individuals in the population carrying the desired gene (in each generation)
- $K_c = 0.0002$   
2 individuals die for every 10,000 individuals carrying the desired gene (in each generation)
- $N_{a0} = 20$   
Number of individuals in the population carrying the desired gene initially.



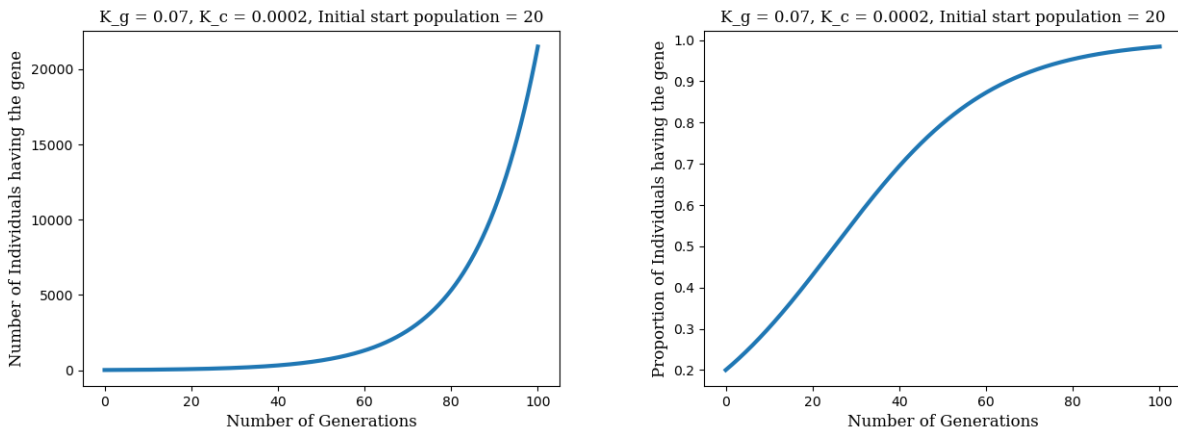
**Figure 1:** WNS Increase in number of selected individuals and proportional value is very low

#### 4.1.2 Moderate positive selection

The values chosen for the simulation are:

- $K_g = 0.07$   
7 offspring is produced for every 100 individuals in the population carrying the desired gene (in each generation)
- $K_c = 0.0002$   
2 individuals die for every 10,000 individuals carrying the desired gene (in each generation)
- $N_{a0} = 20$   
Number of individuals in the population carrying the desired gene initially.



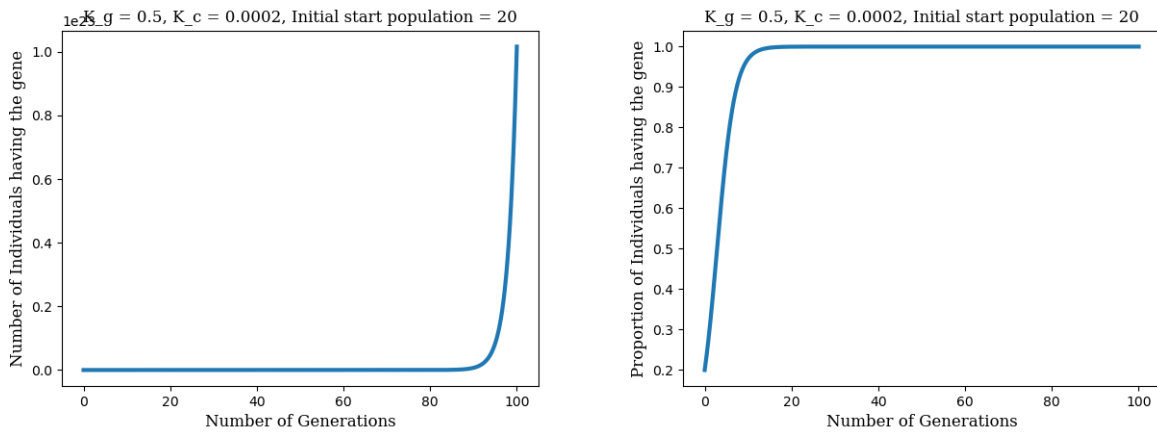


**Figure 2:** MPS Increase in number of selected individuals and proportional value is gradual

### 4.1.3 Strong positive selection

The values chosen for the simulation are:

- $K_g = 0.5$   
1 offspring is produced for every 2 individuals in the population carrying the desired gene (in each generation)
- $K_c = 0.0002$   
2 individuals die for every 10,000 individuals carrying the desired gene (in each generation)
- $N_{a0} = 20$   
Number of individuals in the population carrying the desired gene initially.



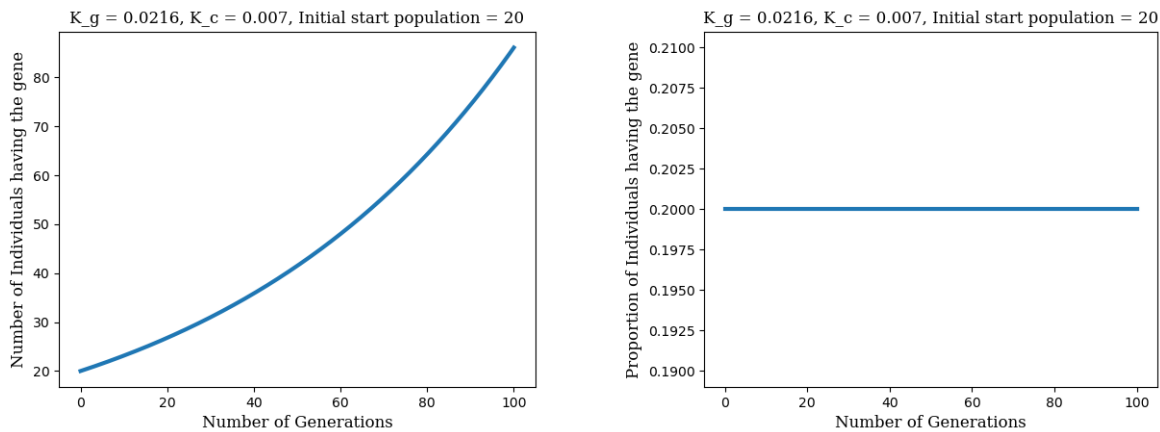
**Figure 3:** SPS Increase in number of selected individuals and proportional value is exponential

## 4.2 No Selection

Absence of selection pressure can be stimulated by making the generation rates and consumption rates of the desired gene of interest to be equal to that of the overall population.

The values chosen for the simulation are:

- $K_g = 0.0216$   
216 offspring is produced for every 10,000 individuals in the population carrying our gene of interest (in each generation)
- $K_c = 0.007$   
7 individuals die for every 1000 individuals carrying our gene of interest (in each generation)
- $N_{a0} = 20$   
Number of individuals in the population carrying our gene of interest initially.



**Figure 4:** NOS Increase in number of selected individuals is gradual and proportional value is constant

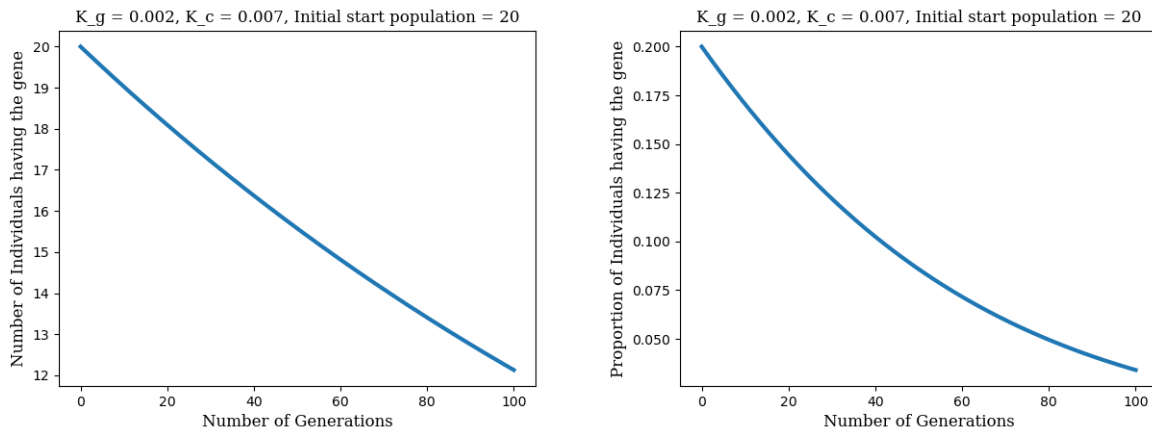
## 4.3 Negative Selection of a gene

Negative selection pressure can be stimulated by making the birth rates lesser and the death rates larger for our gene of interest when compared to that of the overall population.

### 4.3.1 Weak negative selection

The values chosen for the simulation are:

- $K_g = 0.002$   
1 offspring is produced for every 100 individuals in the population carrying our gene of interest (in each generation)
- $K_c = 0.007$   
7 individuals die for every 1000 individuals carrying our gene of interest (in each generation)
- $N_{a0} = 20$   
Number of individuals in the population carrying our gene of interest initially.

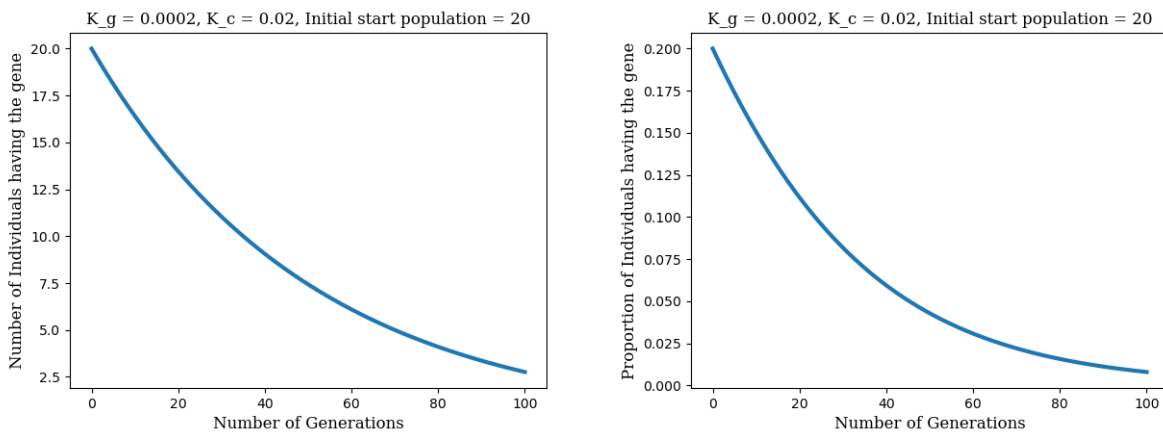


**Figure 5:** WNS Decrease in number of selected individuals and proportional value decreases slow

### 4.3.2 Moderate negative selection

The values chosen for the simulation are:

- $K_g = 0.0002$   
2 offspring is produced for every 10,000 individuals in the population carrying our gene of interest (in each generation)
- $K_c = 0.02$   
2 individuals die for every 100 individuals carrying our gene of interest (in each generation)
- $N_{a0} = 20$   
Number of individuals in the population carrying our gene of interest initially.



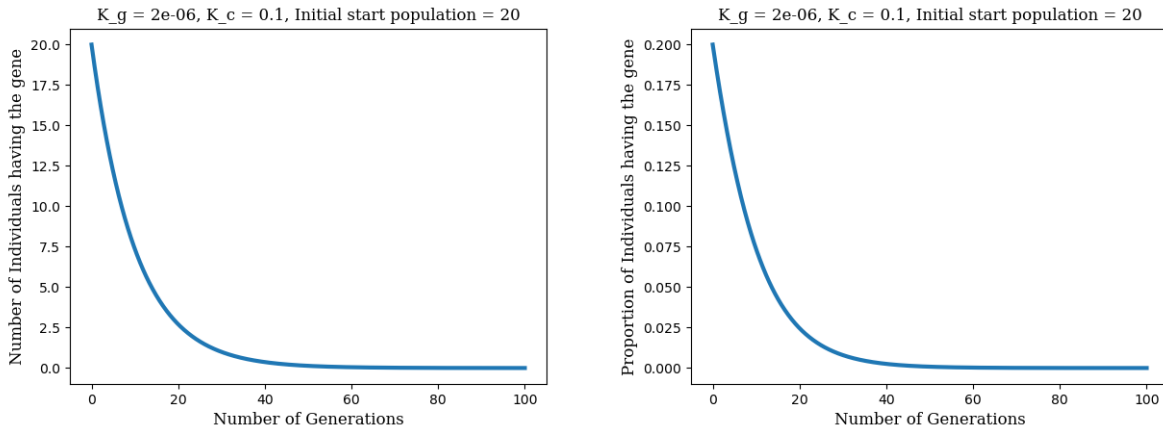
**Figure 6:** MNS Decrease in number of selected individuals and proportional value is gradual

### 4.3.3 Strong negative selection

The values chosen for the simulation are:

- $K_g = 2(10^{-6})$   
1 offspring is produced for every  $10^{-6}$  individuals in the population carrying our gene of interest (in each generation)

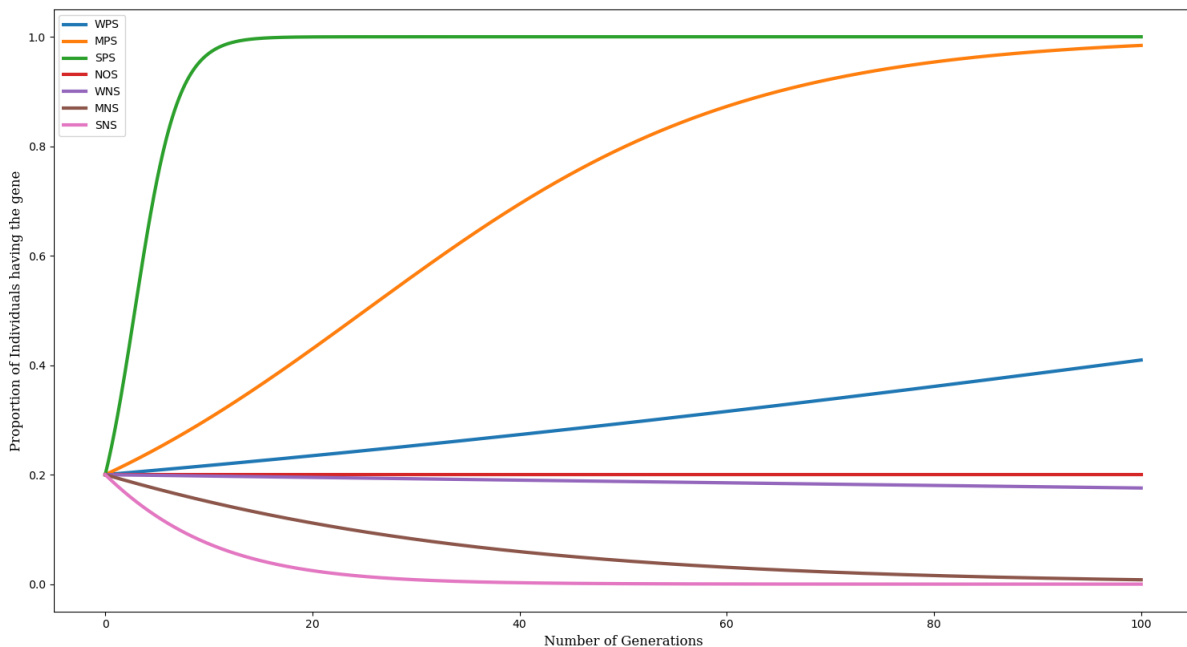
- $K_c = 0.1$   
1 individuals die for every 10 individuals carrying our gene of interest (in each generation)
- $N_{a0} = 20$   
Number of individuals in the population carrying our gene of interest initially.



**Figure 7:** SNS Decrease in number of selected individuals and proportional value is exponential

#### 4.4 Cumulative Result

The consolidated results involving the proportion of individuals carrying our gene of interest is as follows:



**Figure 8:** Proportional value of selected individuals across all selection possibilities. There is a general trend of increase (or decrease) of proportion values under positive selection (or negative selection). It remains constant when no selection is applied.

where, WPS is Weak Positive Selection, MPS is Moderate Positive Selection, SPS is Strong Positive Selection, NOS is NO Selection, WNS is Weak Negative Selection, MNS is Moderate Negative Selection and SNS is Strong Negative Selection.

## 5 Inferences

The proportion of the population carrying the gene of interest was found to

- Increase and saturate close to 1 when the gene is positively selected,
- Remain constant when there is no selection
- Decrease when the gene is negatively selected.

The number of individuals carrying the gene of interest was found to increase or decrease exponentially when the selection is strong and gradually when the selection is moderate and slowly when the selection is weak. In case of very weak selection it is seen that the number of individuals increases over generations, but the over all proportion of individuals carrying our gene of interest decreases.

The increases in proportion were found to follow a sigmoidal pattern. Similar results were observed for single allele selection studies statistically when the allele was placed under positive, neutral and negative selection.<sup>789</sup>

## 6 Conclusion

Thus, applying the diffusion, conservation equations on genes enabled us to study and predict populational level dynamics across 100 generations. **Using just two parameters - Generation rate and Consumption rate (birth rate and death rate) of a population, we have been able to model diffusion of a gene in populational level nearly accurately.** The result obtained from this study is ratified by previous studies on a similar line.<sup>789</sup>

## 7 Further Avenues

In this analysis, I have made the assumption that there are no migration, genetic drift occurrences and mutations in the population. In real life scenarios, these assumptions wouldn't hold true.

I have also made the assumption that the birth and death rates (and hence, Generation and Consumption rates) are first order dependent on the population carrying the gene of interest, while, it could in practise be skewed by factors such as epidemics, number of neonatal birth complications, malnutrition and other environmental parameters.

This model could be made better by addressing the factors listed above.

## 8 Acknowledgment

I would like to thank **Professor G.K. Suraishkumar** for giving me the opportunity to work on interesting and open-ended problems as a part of his Transport Phenomena in Biological Systems course. All the images in this report are original.

## 9 References

- <sup>1</sup> G.K. Suraishkumar. *Continuum Analysis of Biological Systems*. Springer-Verlag Berlin Heidelberg, 2014.
- <sup>2</sup> Cambridge. *Cambridge business English dictionary*. Cambridge University Press, 2011.
- <sup>3</sup> Steven L. Jacques and Scott A. Prah. Fick's Second Law. <https://omlc.org/classroom/ece532/class5/ficks2.html>, 1998.
- <sup>4</sup> Government of India. Open government data (ogd) platform india. <https://data.gov.in/>, 2011.
- <sup>5</sup> Glossary of Statistical terms and OECD. Crude Birth Rate. <https://stats.oecd.org/glossary/detail.asp?ID=490>, 2013.
- <sup>6</sup> Glossary of Statistical terms and OECD. Crude Death Rate. <https://stats.oecd.org/glossary/detail.asp?ID=491>, 2013.
- <sup>7</sup> Jon C. Herron and Scott Freeman. *Evolutionary Analysis, 5th Edition*. Pearson, 2014.
- <sup>8</sup> J. J. Martinson and N. H. Chapman. Global distribution of the CCR5 gene 32-base-pair deletion. page 100–103, 1997.
- <sup>9</sup> P. S. Dawson. Linkage and the elimination of deleterious mutant genes from experimental populations. page 147–169, 1970.

# Appendices

## A Python code for individual selection possibilities

```
1 import math
2 import numpy as np
3 import matplotlib.pyplot as plt
4
5
6 def calc_value_popu(n0, kg, kc, t):
7     n = n0*math.exp((kg-kc)*t)
8
9     return n
10
11 n0 = 20
12 n = n0
13
14 m0 = 100 - n0
15 m = m0
16 Kg = 0.0216
17 Kc = 0.007
18
19 font = {'family' : 'serif',
20         'weight' : 'normal',
21         'size' : 12}
22
23 K_g = [0.025, 0.07, 0.5, 0.0216, 0.02, 0.0002, 2*10**(-6)]
24 K_c = [0.0002, 0.0002, 0.0002, 0.007, 0.007, 0.02, 0.1]
25
26
27 for (kg,kc) in zip(K_g, K_c):
28     dt = 0.1
29     noof_generations = 100
30     t = [i*dt for i in range(int(noof_generations/dt)+1)]
31
32     n_vals = [n0]
33     m_vals = [m0]
34     for i in t[1:]:
35         n = calc_value_popu(n0, kg, kc, i)
36         n_vals.append(n)
37
38         m = calc_value_popu(m0, Kg, Kc, i)
39         m_vals.append(m)
40
41     prop = []
42     for i in range(len(n_vals)):
43         proportion = n_vals[i]/(n_vals[i]+m_vals[i])
44         prop.append(proportion)
45
46     plt.figure(1)
47     plt.plot(t,prop, linewidth = 3)
48     plt.xlabel('Number of Generations', fontdict = font)
49     plt.ylabel('Proportion of Individuals having the gene', fontdict = font...
50     )
51 plt.legend(["WPS", "MPS", "SPS", "NOS", "WNS", "MNS", "SNS"])
52 plt.show()
```

## B Python code for overall selection possibilities

```
1 import math
2 import numpy as np
3 import matplotlib.pyplot as plt
4
5
6 def calc_value_popu(n0, kg, kc, t):
7     n = n0*math.exp((kg-kc)*t)
8
9     return n
10
11 Kg = 0.0216
12 Kc = 0.007
13
14 K_g = [0.025, 0.07, 0.5, 0.0216, 0.002, 0.0002, 2*10**(-6)]
15 K_c = [0.0002, 0.0002, 0.0002, 0.007, 0.007, 0.02, 0.1]
16
17 font = {'family' : 'serif',
18         'weight' : 'normal',
19         'size': 12}
20
21 y = 'y'
22 while y == 'y':
23     n0 = float(input('Initial number of the population having the gene: '))
24     n = n0
25     m0 = 100 - n0
26     m = m0
27
28     for (kg,kc) in zip(K_g, K_c):
29         dt = 0.1
30         noof_generations = 100
31         t = [i*dt for i in range(int(noof_generations/dt)+1)]
32
33         n_vals = [n0]
34         m_vals = [m0]
35         for i in t[1:]:
36             n = calc_value_popu(n0, kg, kc, i)
37             n_vals.append(n)
38
39             m = calc_value_popu(m0, Kg, Kc, i)
40             m_vals.append(m)
41
42         prop = []
43         for i in range(len(n_vals)):
44             proportion = n_vals[i]/(n_vals[i]+m_vals[i])
45             prop.append(proportion)
46
47
48         string = 'K_g = ' + str(kg) + ', K_c = ' + str(kc) + ', Initial ...
49                 start population = ' + str(int(n0))
50         plt.figure()
51         plt.plot(t,n_vals, linewidth = 3)
52         plt.title(string, fontdict = font)
53         plt.xlabel('Number of Generations', fontdict = font)
54         plt.ylabel('Number of Individuals having the gene', fontdict = font...
```



```
55
56     plt.figure()
57     plt.plot(t,prop, linewidth = 3)
58     plt.title(string, fontdict = font)
59     plt.xlabel('Number of Generations', fontdict = font)
60     plt.ylabel('Proportion of Individuals having the gene', fontdict = ...
61               font)
62     plt.show()
63     y = input('Continue (y/n)? ').lower()
```