# EVOLVE Thoughts

## Goals

As a biology instructor, I need a cross-platform program (app?) my students can use to

1. Learn about the major forces of evolution;
2. Learn to design experiments;
3. Learn to interpret graphs and analyze data;
4. Solidify their knowledge of Mendelian genetics
5. Export data for more sophisticated analyses (and import scenario?)

It should be open–source in a cross-platform language

## Basic Genetic & Evolutionary Concepts

A *Population* consists of *Organisms*

Each population has 1 *Gene* (or more) that affects 1 or more traits (e.g., pigments, proteins, etc.)

Genes may have 2 or 3 Alleles (more in real world) symbolized by A, B, C

Alleles have

Mutation Rates (A–> B, B–>A, A–>C, C–>A, B–>C, C–>B)

Patterns of inheritance*, see below*

Ea. Organism has 2 copies of each gene

Thus, organisms may have one of 3 (or 6) *Genotypes*.

Symbolized AA, AB, BB; AC, BC, CC

Organisms may be *Homozygotes* (2 copies of same allele) or *Heterozygotes* (2 different alleles)

Organisms have *Phenotypes*: observable traits (e.g., pigments, proteins, etc.) produced by genotype.

Phenotypes of interest in EVOLVE include:

*Survival rate* (avg. % young surviving to adulthood)

*Reproductive rate* (avg. # young / adult)

*Absolute Fitness* (SurvR8 \* ReprR8) Values > 1 result in increasing population size

*Relative Fitness* (AbsFit/MaxAbsFit) Values 0–1

*Migration rates* (% individuals moving to another population)

*Mating preferences* (QUERY: model or not? Adds significant complexity. )

*Patterns of inheritance* are determined by interactions between alleles and include

*Dominant–Recessive*: *heterozygote* has same phenotype as dominant homozygote, recessive phenotype is homozygous recessive. E.g., if AA = 5, AB = 5, BB = 10, A is dominant, B is recessive

*Incomplete Dominance*: heterozygote is in between phenotypes of homozygotes. E.g., If AA = 5, AB = 8, BB = 10, alleles are incompletely dominant.

*Codominance*: heterozygote displays phenotypes of both homozygotes. E.g., AA = Protein1, BB = Protein 2, AB = Proteins 1 & 2.

*Heterosis / Heterozygote superiority*: heterozygote phenotype is more extreme than either homozygote. E.g., AA = 5, AB = 13, BB =8

*Underdominant*: heterozygote phenotype is less extreme than either homozygote. E.g., AA = 5, AB = 3, BB = 10

A Species can be conceived of as a Metapopulation, a set of more–or–less isolated subpopulations that are capable of reproducing within themselves and which are isolated from other populations.

#### Hardy–Weinberg Concept

A population can be modeled as a *gene pool* made up of all of the alleles in a population.

Gene Pools have

*Allele Frequencies* (p = freq. A, q = freq. B, r = freq. C)

*Genotype frequencies*

A population is conceived of as a *gene pool* with allele frequencies p *&* q *(or* r*)*, by definition, p + q = 1 or p + q + r = 1

*Mating* consists of randomly sampling pairs of alleles without replacement from the gene pool

**Assuming** non–overlapping generations and no:

*Mutation*: alleles don’t change (violation is Mutational Disequilibrium)

*Selection*: no differences in allele/genotyp survival, reproduction (violation is Natural Selection)

*Nonrandom mating*: no differences in probability of genotypes mating w/ ea. other (violation is Sexual Selection or Preferential Mating).

*Gene flow*: population is closed, no differential emigration/immigration of alleles/genotypes

*Genetic Drift*: sampling errors

Then **allele frequencies will not change and genotype frequencies will be**

p2 + 2pq + q2(2–alleles) or

p2 + 2pq + 2pr + 2qr + q2 + r2 (3–alleles)

This is the null case of no evolution. Violation of any of the 5 assumptions leads to changes in allele/genotype frequencies = *evolution*

#### Modeled Organisms

*Diploid* (have 2 copies of each gene), with *non–overlapping generations*; think of insects where adults mate, give birth, lay eggs, set seeds, etc., then die before the young mature.

*Hermaphroditic*, facultatively self–fertilizing (if cannot find a mate, will fertilize themselves). QUERY: make non–selfing? Make an instructor option?

*Life cycle:* adults mate, produce zygotes (eggs/seed), die. A % of zygotes emigrate; migration is random between populations; Immigrants are added to resident zygotes. Zygotes hatch/germinate to produce juveniles which undergo selection; remainder are adults of next generation.

## Model Parameters

**TITLE** [text for this run/experiment**]**

Require entry? Customize?

Add more structured text fields? E.g., [Comment], [Question], [Method], [Results], [Conclusion]

**▼ CONFIGURE SIMULATION**

**Seed [**integer#**]**

**Number of alleles: ◉ Two ○ Three**

**Name alleles:**

**1 [A**/Letter**]** & **[**Long text**]**

**2 [B**/Letter**]** & **[**Long text**]**

**3 [C**/Letter**]** & **[**Long text**]**

**Name genotypes** {customize from above}

**Evolutionary forces**

◉ **Population size** (AKA Genetic drift)

◉ **Natural**

◉ **Mutation**

◉ **Migration** (AKA Gene Flow)

◉ **Non–Random Mating** (AKA Sexual Selection)

**[#] Number of populations to run**

**[#] Number of generations to run**

**◎ Change evolutionary scenario**

**Seed:** an initial value for generating uniform & Gaussian random numbers.

**Number of Alleles** determines how many alleles to model, configuration of inputs & displays

**Names of alleles & genotypes**

Default to A, B, C & AA, AB, BB, AC, BC, CC; allow user to rename, e.g.,   
S = Normal, s = Sickle, & R = Red, y = Yellow  
SS = Normal, Ss = Sickle trait, ss = Anemia RR = Red, Ry = Red, yy = Yellow

**Evolutionary Forces**

Population size must always be ON; turn other evolutionary forces ON | OFF.

Setting individual forces OFF hides input parameters and simplifies the user interface.

**Change Evolutionary Scenario**

This will require some thought. One brainstorm: Drop down list or dialog box of parameters: have user click checkboxes of the ones to change, when done, display list with current values & blank for new ones. Will need + – icons to add/delete.

**▼ INITIAL POPULATION**

**◉ Genotypes** (establish equilibrium population)

**◎ Frequencies** (establish non–equilibrium population)

**▼ INITIAL POPULATION**

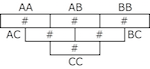
◉ **Frequencies** (Calculate equilibrium population)

**[#] Initial population Size** (integer 10–10,000)

**[#] Frequency of A** (if 2 alleles, decimal 0–1.0)

**[#] Frequency of B** (if 3 alleles) (if 3 alleles must input this, if 2 alleles calculate automatically)

User would enter values (decimal, 0.0–1.0) for A &/or B, program calculates genotype numbers of B/C per H-Weinberg and display:

If 2 alleles, can collapse all of these input boxes to top row

**▼ INITIAL POPULATION**

◉ **Genotypes** (establish non–equilibrium population)

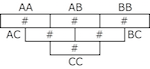
**[#] Initial population Size** (integer 10–10,000)

**[#] Frequency of A** (if 2 alleles)

**[#] Frequency of B** (if 3 alleles) (if 3 alleles must input this, & C)

**[#] Frequency of C** (if 3 alleles)

User would enter values (decimal, 0.0–1.0) for A &/or B, program calculates genotype numbers of B/C per H-Weinberg and display:



———————————————————————————————————————

**▼ EVOLUTIONARY SITUATION**

▼ POPULATION SIZE (Genetic Drift)

[#] Carrying Capacity (1–10,000)

[#] Post–Crash Size (1–10,000)

Required parameters, Sets ranges for a population’s size (integer, 1–10,000); Post–crash must be < Carrying capacity

If the population grows in size and the adult population exceeds the carrying capacity, randomly reduce the number of each genotype to the post-crash size. This could be a random or a stochastic process.

If the population declines, then allow it to decline to extinction.

**▼ EVOLUTIONARY SITUATION**

▼ SELECTION:

**◎** Off

◉ Input Survival & Reproductive Rates

**◎** Input Absolute Fitness

▼ SELECTION:

◉ Off

If off in configuration menu, this would be the only item under selection and others would be hidden

▼ SELECTION:

**◎** Off

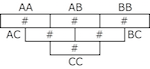
◉ Input Survival & Reproductive Rates

**◎** Input Absolute Fitness

Show Survival, Reproductive rate as inputs, and calculate Absolute & Relative fitness displays as shown below.

Survival Rate:

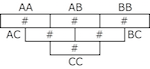
Avg.% young surviving to adulthood per genotype (decimal, 0–100%)



Reproduction Rate:

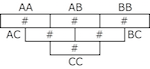
Avg. # young per genotype (decimal, 0–10, QUERY: higher if fast enough?)

Input boxes for ReprR8



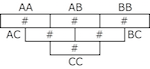
Absolute Fitness:

Calculated for each genotype = SurvR8 \* ReprR8



Relative Fitness:

Calculated = Absolute fitness (genotype) / Max Absolute



▼ SELECTION:

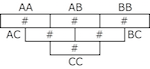
○ Off

○ Survival & Reproductive Rates

◉ Input Absolute Fitness

If ON, don't show Survival or Reproductive Rates; show Absolute fitness as input and relative fitness as calculated values

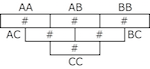
Absolute Fitness:



Input values (0.0–10.0)

This will require careful thinking about reproduction algorithm. As first approximation, could multiply genotype Mates array by absolute fitness values to generate Young array

Relative Fitness:



Calculated = Absolute fitness (genotype) / Max Absolute

**▼ EVOLUTIONARY SITUATION**

**▼ MUTATION**

○ **On**

**▼ MUTATION**

**◎** On

set all mutation rates to 0

**▼ MUTATION**

**◉ On**

**From–>To From–>To**

**[#] A–>B [#] B–>A**

**[#] A–>C [#] B–>C**

**[#] C–>A [#] C–>B**

If On, display rates: forward & back mutation rates (decimal, range? 10^-6 – 10^?? Major problem w/ range, how to guide students? Use sci notation?)

Top line only for 2–allele, all for 3–allele)

**▼ EVOLUTIONARY SITUATION**

**▼ Migration** (Gene Flow)

**◉ Off**

**◎ All Genotypes**

**◎ By Individual Genotype**

**▼ Migration** (Gene Flow)

**◎ Off**

Hide below, set mutation rates to 0.0

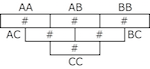
**▼ Migration** (Gene Flow)

◉ **All Genotypes**

[#] % migrating between populations (integer, 0 – 95%)

All genotypes have same migration rate.

◉ **By Individual Genotypes**

[#] % migrating between populations (integer, 0 – 95%)  
 

**▼ EVOLUTIONARY SITUATION**

▼ Non–Random Mating (AKA Sexual Selection)

**◎** Off

◉ Mating Preferences

Mating Preferences:

Not sure how to model. Random would be no preference, 0.0; positive & negative values indicate preference & aversion. But what limits to put on values?

AA w/ AA, AB, BB also (3 alleles) AC, BC, CC

AB w/ AA, AB, BB also (3 alleles) AC, BC, CC

BB w/ AA, AB, BB also (3 alleles) AC, BC, CC

CC w/ AA, AB, BB also (3 alleles) AC, BC, CC

CA w/ AA, AB, BB also (3 alleles) AC, BC, CC

CB w/ AA, AB, BB also (3 alleles) AC, BC, CC

### List of Input Parameters

**TITLE** [text for this run/experiment**]**

**▼ CONFIGURE SIMULATION**

**Seed [**integer#**]**

**Number of alleles: ◉ Two ○ Three**

**Evolutionary forces**

◉ **Population Size** (AKA Genetic drift)

◉ **Natural Selection**

◉ **Mutation**

◉ **Migration** (AKA Gene Flow)

**[#] Number of Populations to run**

**[#] Number of generations to run**

**◎ Change evolutionary scenario**

**▼ INITIAL POPULATION**

**◉ Frequencies** (establish non–equilibrium population)

◉ **Frequencies** (Calculate equilibrium population)

**[#] Initial population Size** (integer 10–10,000)

**[#] Frequency of A** (if 2 alleles, decimal 0–1.0)

**[#] Frequency of B** (if 3 alleles) (if 3 alleles must input this, if 2 alleles calculate automatically)\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

◉ **Genotypes** (establish non–equilibrium population)

**[#] Initial population Size** (integer 10–10,000)

**[#] Number of A** (if 2 alleles)

**[#] Number of B** (if 3 alleles) (if 3 alleles must input this, & C)

**[#] Number of C** (if 3 alleles)

**▼ EVOLUTIONARY SITUATION**

▼ POPULATION SIZE (Genetic Drift)

[#] Carrying Capacity (1–10,000)

[#] Post–Crash Size (1–10,000)

▼ SELECTION:

**◎** Off

◉ Input Survival & Reproductive Rates

**◎** Input Relative Fitnesses

◉ Off

**◎** Off

◉ Input Survival & Reproductive Rates............

**◎** Input Absolute Fitness

○ Off

○ Survival & Reproductive Rates

◉ Input Absolute Fitness.....................

**▼ MUTATION**

**◎ On**

**◎** On

**◉ On**

**From–>To From–>To**

**[#] A–>B [#] B–>A**

**....................**

**▼ Migration** (Gene Flow)

**◉ Off  
 ◎ All Genotypes** **◎ By Individual Genotype**

◉ **All Genotypes**  
 [#] % migrating between populations (integer, 0 – 100%)

◉ **By Individual Genotype**  
 % Emigrating (integer, 0 – 90%)...............

▼ Non–Random Mating (AKA Sexual Selection)

**◎** Off

◉ Mating Preferences

**AA w/ AA, AB, BB also (3 alleles) AC, BC, CC**

**AB w/ AA, AB, BB also (3 alleles) AC, BC, CC**

.................

## Menus

File

New

Open

Save As…

Export…

Page Setup…

Print…

Edit

Cut ⌘X

Copy ⌘C

Paste ⌘V

Run

Continue, Current Parameters ⌘^C

Continue, Different Parameters ⌘D

New Run, Current Parameters ⌘N

New Run, Change Parameters ⌘^N

Will need to give thought to UI for modifying parameters, file structure, format, how to export, etc.

## Output

### **SCREEN**

Will need careful thought about user interface to select what to display, how to present displays, & how to let users customize displays (e.g., line colors, widths, types) & arrange windows.

Basic UI: Allele frequencies, Genotype freq.s, Population Size, Mean fitness, Time

Advanced UI: Any data vs any other data (e.g., data on mate combinations (#, freq.s), young allele & genotype freq.s, fitness of genotypes…)

2D: Time vs any other data(separate tracks if multiple pop.s)

Allele Freq.s, Genotype Freq.s, Population size, Mean fitness of 2 & 3 alleles

Ternary plot of 3 alleles

If multiple populations

Line plots of individual pop.s

Box–&–Whisker plots of mean, SD, max, min

Frequency density, color Gaussian ridge

**3D (rotating) of Time vs**

Bars of Allele Freq.s, Genotype Freq.s, Pop. Size

Ternary (tetrahedral) plots for 3 alleles

If multiple populations

Line plots of raw pop. freq.s

Histograms of frequencies &/or Normal distribution of mean, SD

**Density/statistical plots** from multi–pop simulations: box & whisker, color Gaussian ridge?

Mean, SD, pop fitness

Adaptive landscape if have explored parameter space, see *DeFinetti.xlsx*

### Printouts, Export

**Copy/paste** and/or **export** text, graphic & data files for reports and for analysis in other programs.  
Need Basic & Advanced (any variable’s data)

See *EVOLVE Print/Export Ideas.xlsx*

Print/Export Data

Input parameters

Results Data

### Additional Possibilities

* Allow Instructor customization
  + Could configure & rename EVOLVE file to provide incremental learning   
    e.g., “EVOLVE Sel for Dominant” or “EVOLVE Migration + Drift”
  + # alleles: 2 or 3
  + Evol forces: Instructor can turn off any or all except Pop. Size
  + Set default values or ranges for any or all variables
    - MaxPopSize, PostCrashSize
    - Selection: turn On | Off (If Off, will need to set fixed SurvR8, ReprR8)
    - Automate Inheritance patterns? (Dom/Rec, Incomplete Dom, Under– & Over–dominance (heterosis), Codominance? Difficult to design for combination of selection, mutation, migration
* Scenarios/ScriptingE.g., Catastrophic bottlenecks in pop size
  + Perhaps a scripting language or template text file that would allow execution of multiple exp’ts?
  + Automate exploration of a parameter space like *DeFinetti.xlsx*?
  + Cycle of oscillating selection parameters?
* Notebook of series of exp'ts? A place to store a series of exp’ts, allow paste-up of graphs for comparision?

|  |  |
| --- | --- |
| **EXPERIMENTS** Show hidden Show Summaries   Print Visible Exp’ts |  |
| No. Hide Title |  |
| 1. Selection for Dominant (w/ error) |  |
| 2. Selection for Dominant |  |
| 3. Selection for Recessive |  |
|  |  |
| **GRAPHS** |  |
|  |  |
|  |  |
|  |  |

* Should this be a separate input/output file or built into main program?
* Just have unstructured space for Copy/Paste of text, images?

## Sample Files

* *EVOLVE Old UI Ideas.docx* — 20–year old rough design for a v4 UI
* *EVOLVE Print/Export Ideas.xlsx* — Possible designs for printouts & data exports
* *EVOLVE v1 FORTRAN Source.txt* — Original Mac FORTRAN code
* *EVOLVE v1 Manual Complete.txt* ­— Original Mac FORTRAN documentation. Will provide
* *EVOLVE v4 Thoughts.docx* — This file
* EVOLVE v4 UI Demo.html— Folder w/ draft of UI; open file *index.html*
* *My EVOLVE Workbook* — Folder w/ An Excel implementation of the basic design of EVOLVE. It deals only w/ Selection & is deterministic. The following 5 files illustrate how it’s used & provide explanations of some of the concepts
  + *Lab06 - Sel vs Lethal Rec.pdf*
  + *Lab07 Sel Dom-Rec.pdf*
  + *Lab07 Sel expts KEY.xlsx*
  + *Lab08 Sel of Other Inheritance Patterns.pdf*
  + *Lab10 drift, flow, mut 2015-2.pages*
  + *Lab10 drift, flow, mut.pdf*
* *Other Models* — Folder w/ Excel based models that use selection coefficients instead of survival & reproductive rates.
  + *DeFinetti 3–Allele Model* — Folder w/ documentation & a 3-allele model with ternary & fitness landscape, but deterministic model of only selection.
  + *DEME Model* — Folder w/ documentation & a 2-allele model that models drift, selection, mutation & migration.

*App PopG 3.3* — Folder w/ a­­­­nother 2-allele model that models drift, selection, mutation & migration w/ a better UI