

Deme 1.0

No doc on Deme 2.0

©2004 Anton E. Weisstein
BioQUEST Curriculum Consortium

No discussion of mutation, not realistic in this 50-gen

Summary

Deme (pronounced "deem") is an Excel workbook that simulates the population genetics of a single gene with two alleles. The user sets initial allele frequencies and enters parameters for selection, genetic drift, mutation, migration, and changes in population size. The program then tracks the population through 50 generations and plots the results. The user may choose whether to plot genotype frequencies, allele frequencies, mean fitness, and/or the population size.

Introduction

Population genetics is the study of changes in allele and genotype frequencies within populations, and of the processes that cause such changes. Applications of population genetics include:

- Predicting a population's evolution. For example, a mutation of the CCR5 gene in humans seems to confer substantial resistance to HIV infection. We can use population genetics to explore how this gene might evolve in response to the AIDS epidemic and to determine whether this process will bring the epidemic under natural control.
- Testing specific evolutionary hypotheses. For example, many human genetic disorders such as Tay-Sachs disease, cystic fibrosis, and sickle-cell anemia are thought to be adaptive responses to past epidemics in the affected populations' histories. Based on measured levels of disease resistance conferred by these alleles, and on the known time and severity of specific epidemics, we may be able to eliminate certain hypotheses of this sort while supporting others.
- Informing debates about public policy. For example, during first half of the 20th Century, proponents of eugenics advocated compulsory sterilization of people with undesirable genetic traits. Using population genetics, we can explore the effectiveness of such a policy, determine what proportion of the population would be affected, and predict any potential secondary consequences.

In population genetics, *evolutionary forces* are defined as processes that cause allele frequencies to change. Four such forces are generally recognized:

1. **Selection** — differential survival or reproduction of individuals with different genotypes;
2. **Genetic drift** — random changes in allele frequency due to chance events;

3. **Mutation** — genetic change of one allele into another; and

4. **Migration** — movement of individuals (and the alleles they carry) between local populations.

A fifth force, **nonrandom mating**, is not considered a true evolutionary force because it does not alter allele frequencies. However, it does alter genotype frequencies, and can have important effects in combination with other forces such as selection and migration.

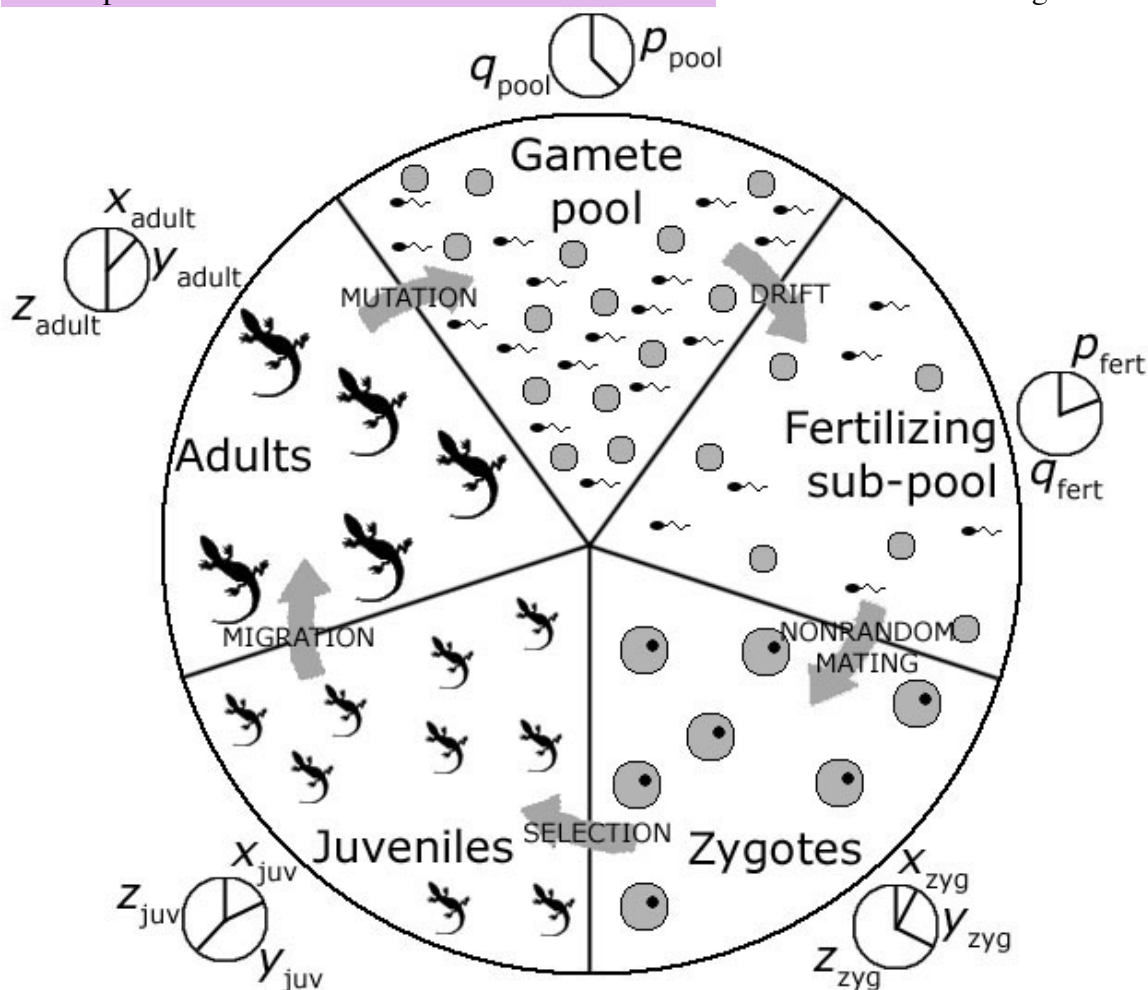


Figure 1. A population's life cycle as modeled by *Deme*. A generation is defined as one complete turn of the cycle. We measure allele frequencies p and q during the two haploid stages (gamete pool and fertilizing sub-pool) and genotype frequencies x , y , and z during the three diploid stages (zygote, juvenile, and adult).

x , y , & z ?

Output: What do I see?



Figure 2. A screen shot of *Deme*'s main worksheet.

The main output of *Deme* consists of one or more plots on the "Main" worksheet. The user chooses from the following plot options: frequency of any genotype, frequency of either allele, mean fitness of the population, and population size (N). Each of these variables is plotted against the number of generations elapsed. All independent variables except for population size are plotted on the left-hand y-axis (linear scale), population size is plotted on the right-hand y-axis (logarithmic scale), and time is plotted on the x-axis. The legend below the graph explains the interpretation of each individual plot.

The colorful tables on the right display current parameter values. The "Plot Variables" box shows which variables are currently plotted. More details on individual parameters are given in the **Controls** section below.

You can also access the "Calculations" worksheet using the tab of the same name at the bottom of the workbook. This sheet tabulates numeric values for the following variables:

Column(s)	Variable
A	Generation number (starting population is generation 0)
B	Population size N
C – E	<u>Non-normalized</u> genotype frequencies of juveniles
F – H	<u>Normalized</u> genotype frequencies of juveniles
I – K	Genotype frequencies of adults
L	Mean fitness of the population
M & N	Allele frequencies in the gamete pool
O & P	Allele frequencies in the subset of gametes drawn from the pool to form the next generation of zygotes

Table 1. Variables tabulated on the "Calculations" worksheet.

Columns in boldface (A, B, I – L, M & N) correspond to variables that can be plotted on the main plot. *Deme* considers the allele frequencies input by the user to correspond to the gamete pool frequencies in generation 0. As a result, selection, migration, and mutation do not occur in generation 0, so the corresponding entries are italicized. Note, however, that genetic drift does occur, so columns O and P are not italicized.

Controls: What can I do?

Most of the controls are located on the "Main" worksheet. Individual parameters are:

Parameter	Description
p_0	Initial frequency of the <i>A</i> allele
q_0	Initial frequency of the <i>B</i> allele
W_{AA}, W_{AB}, W_{BB}	Relative fitness of each genotype
N	Total population size
$\mu_{A \rightarrow B}, \mu_{B \rightarrow A}$	Mutation rates from <i>A</i> to <i>B</i> and vice versa
M_{AA}, M_{AB}, M_{BB}	Number of immigrants of each genotype arriving each generation

Table 2. *Deme* model parameters.

You can change the values of most parameters by typing the new value directly into the appropriate cell. Exception: by definition, $p_0 + q_0 = 1$. *Deme* calculates q_0 as $1 - p_0$, so you cannot edit the value of q_0 directly. If you want to set q_0 , simply enter the complementary value of p_0 .

You can turn genetic drift on and off using the radio control button under the "Drift" header. You can also control the population size in each generation; for example, to simulate a population bottleneck or boom/bust cycles. To do this, go to the "Calculations" sheet and enter the new values directly into column B. To set the population size back to its initial baseline, enter the text " $=N$ " in each of the cells in this column that you previously modified.

On the "Main" sheet, use the check boxes in the "Plot Variables" box to determine which variables are plotted.

When you make any changes to the worksheet, it will automatically perform a new simulation. You can also use Excel's "Calculate Now" command to run a new simulation without changing parameter values. You can access this command in the Calculation tab of the Preferences panel, or by using the keyboard shortcut (usually *Control* + "=" or *Command* + "="). If you do not want *Deme* to run a new simulation every time you make a change, use the Preferences panel to set Calculation to Manual.

Both sheets are protected so that users will not unintentionally overwrite key portions of the workbook. If you want to make changes beyond those outlined above, such as hiding or un-hiding specific columns or changing the model, you must first use the "Unprotect Sheet" command. In Excel 2001, this command is on the Tools menu. You can re-protect the sheet when you are done, and (if you wish) even add a password to prevent

others from later un-protecting it again. This is useful for assigning exercises where you want the parameter values to be unknown; see the **Ideas** section below.

How it works: Model details

Deme begins by calculating q_0 , the initial frequency of the B allele among gametes that will unite to form zygotes in generation 1, as $q_0 = 1 - p_0$. The population is then tracked through the following life cycle stages:

1) Zygotes. *Deme* assumes random mating among the gametes that form the fertilizing sub-pool. Zygotes of genotype AA , AB , and BB thus have respective frequencies p^2 , $2pq$, and q^2 , where p and q are the respective frequencies of the A and B alleles in the fertilizing sub-pool. Gametes from any individual adult are assumed to be compatible with those from any other (or even the same) adult, corresponding to a population of self-compatible hermaphrodites with non-overlapping generations.

2) Juveniles. *Deme* assumes that all natural selection occurs between the zygote and juvenile stages. Each genotype's relative fitness is thus equal to the proportion of zygotes of that genotype that survive to the juvenile stage. We can therefore obtain the raw "frequency" of juveniles with a given genotype by just multiplying the frequency of zygotes with that genotype by the genotype's fitness. However, these raw "frequencies" add up to the average fitness of the population, which in most cases will not be one. To turn them into true frequencies, we must re-normalize them by dividing each raw "frequency" by the population's average fitness. Note that this is a model of *soft selection*: frequencies are altered, but not the population's overall size.

3) Adults. *Deme* assumes that all migration occurs between the juvenile and adult stages. Moreover, the population is assumed to have a strict carrying capacity equal to the population's size at that time point. In *Deme*, therefore, migration affects genotype frequencies but not overall population size. The frequency of adults of a given genotype is calculated as the weighted average of individuals of that genotype already in that population and individuals of that genotype immigrating into the population.

It is possible to model emigration in *Deme* by just entering a negative value for the number of immigrants of a specific genotype. However, this approach can cause problems if the number of adults of that genotype ever falls below the number emigrating in every generation. Probably a better strategy is to have a specific percentage of each genotype emigrate. The easiest way to do this is by modifying genotypes' relative fitnesses appropriately. For example, if 90% of AA zygotes survive to become juveniles, but 20% of them then emigrate, we might set W_{AA} to $0.90 * (1 - 0.20) = 0.72$.

4) Gamete pool. In the absence of mutation, AA adults produce only A gametes, BB adults produce only B gametes, and AB adults produce both in a 1:1 ratio. Under mutation, a proportion $\mu_{A \rightarrow B}$ of A gametes mutate into B , while a proportion $\mu_{B \rightarrow A}$ of B gametes change into A . These two mutation rates can be equal or unequal.

5) Fertilizing gamete sub-pool. Only a small fraction of the overall gamete pool will either fertilize or be fertilized to form zygotes in the next generation. Due to sampling effects, allele frequencies within this sub-pool may not be the same as those in the overall gamete pool: this is the process of genetic drift. *Deme* models genetic drift by assuming that the change in allele frequency follows a normal distribution with mean zero and variance equal to that of a binomial distribution ($\sigma^2 = pq/2N$, where p and q are the allele frequencies in the overall gamete pool and N is the population size in the next generation).

The equations corresponding to steps 1 – 5 are given below. In these equations, x , y , and z denote the respective frequencies of the AA , AB , and BB genotypes, while p and q denote the respective frequencies of the A and B alleles. The subscripts zyg , juv , $adult$, $pool$, and $fert$ respectively denote the life cycle stages of zygote, juvenile, adult, overall gamete pool, and fertilizing gamete sub-pool. The number in parentheses, e.g. "(1)", represents the generation in which the value is being measured.

$$1) \quad x_{zyg}(1) = p_{fert}^2(0), \quad y_{zyg}(1) = 2p_{fert}(0) \cdot q_{fert}(0), \quad z_{zyg}(1) = q_{fert}^2(0).$$

$$2) \quad x_{juv}(1) = \frac{x_{zyg}(1) \cdot W_{AA}}{\bar{W}(1)}, \quad y_{juv}(1) = \frac{y_{zyg}(1) \cdot W_{AB}}{\bar{W}(1)}, \quad z_{juv}(1) = \frac{z_{zyg}(1) \cdot W_{BB}}{\bar{W}(1)},$$

where

$$\bar{W}(1) = x_{zyg}(1) \cdot W_{AA} + y_{zyg}(1) \cdot W_{AB} + z_{zyg}(1) \cdot W_{BB}.$$

$$3) \quad x_{adult}(1) = \frac{x_{juv}(1) + M_{AA}}{N(1) + M}, \quad y_{adult}(1) = \frac{y_{juv}(1) + M_{AB}}{N(1) + M}, \quad z_{adult}(1) = \frac{z_{juv}(1) + M_{BB}}{N(1) + M},$$

where

$$M = M_{AA} + M_{AB} + M_{BB}.$$

$$4) \quad \begin{aligned} p_{pool}(1) &= [x_{juv}(1) + \frac{1}{2}y_{juv}(1)](1 - \mu_{A \rightarrow B}) + [\frac{1}{2}y_{juv}(1) + z_{juv}(1)](\mu_{B \rightarrow A}), \\ q_{pool}(1) &= [\frac{1}{2}y_{juv}(1) + z_{juv}(1)](1 - \mu_{B \rightarrow A}) + [x_{juv}(1) + \frac{1}{2}y_{juv}(1)](\mu_{A \rightarrow B}). \end{aligned}$$

$$5) \quad \begin{aligned} p_{fert}(1) &= p_{pool}(1) + \Delta p(1), \quad \text{where} \quad \Delta p(1) \sim N\left(0, \frac{p_{pool}(1)q_{pool}(1)}{2N(2)}\right); \\ q_{fert}(1) &= 1 - p_{fert}(1). \end{aligned}$$

Ideas for classroom use

Deme can be used to demonstrate and explore individual evolutionary forces, in either a lecture format or as a computer lab exercise. For example, students can model selection against recessive alleles by setting $W_{AA} = W_{AB} > W_{BB}$. They are often surprised to discover that such selection, no matter how strong, can never entirely eliminate the harmful allele. It is often useful to model selection against dominant alleles ($W_{AA} = W_{AB}$

$< W_{BB}$) as well, because students often confuse the concepts of dominance and selective advantage (Soderberg and Price 2003). Cases of overdominance ($W_{AB} > W_{AA} > W_{BB}$) illustrate the action of balancing selection and introduce the concept of (stable) equilibria, while underdominance cases ($W_{AA} > W_{BB} > W_{AB}$) allow exploration of unstable equilibria. A particularly powerful exercise is to ask students to predict the behavior of an underdominant system, then to have them model such a system for a range of initial allele frequencies and discuss the results. The observation that the fittest genotype can be selectively eliminated when the corresponding allele is rare provides an excellent opportunity to challenge the notion that natural selection means the survival of the fittest.

Deme can also be used to model specific scenarios. For example, Freeman and Herron (2004) present a discussion of evolution at the CCR5 locus in humans due to the AIDS epidemic. They conclude that HIV prevalence, and hence selective pressure, is too low in European populations to cause substantial evolution; while the mutant $\Delta 32$ allele is too rare in higher-prevalence African populations to allow rapid evolution. However, they note that this analysis assumes that the mutant allele is fully recessive with respect to fitness. Recent studies (Kokkotou et al. 1998, Marmor et al. 2001) suggest that heterozygotes may confer partial resistance to HIV infection. Students can explore the enormous evolutionary impact of these findings (if corroborated), potentially motivating a discussion of model sensitivity to specific parameters.

Model	W_{AA}	W_{AB}	W_{BB}	p_0
European population ($\Delta 32$ allele recessive)	0.995	0.995	1.000	0.2
African population ($\Delta 32$ allele recessive)	0.75	0.75	1.00	0.001
African population ($\Delta 32$ allele co-dominant)	0.750	0.925	1.000	0.001

Table 3. Examples of parameter values for exploring CCR5 evolution in humans.

A third possible use for *Deme* is to pose evolutionary problems for students to explore. This can be done by printing out plots and/or table columns (see Figure 3), or by hiding the columns that display the model's parameter values and password-protecting the sheet so that students cannot access them directly. Students may then be asked to determine the evolutionary and population parameters from the information given. This approach works best in the context of a specific scenario, such as the genetics of an isolated human population that suffers a severe epidemic or natural disaster.

Figure 3. Partial screen shot of *Deme*, showing only the frequency of allele A over 50 generations. What can you determine about the specific genotype frequencies, population size, and other parameters underlying these data?

Resources and further reading

Alstad D.N. 2003. Populus 5.3. <<http://www.cbs.umn.edu/software/populus.html/>>

Freeman S. and Herron J. C. 2004. Evolutionary Analysis, 2nd edition. Pearson Prentice Hall: Upper Saddle River, New Jersey.

Kaplan R. 1997. PopBio. <<http://academic.reed.edu/biology/software.html>>

Kokkotou E., Philippon V., Gueye-Ndiaye A., et al. 1998. Role of the CCR5 delta 32 allele in resistance to infection in west Africa. J. Hum. Virol. 1: 469-474.

Louisiana State University. 2000. Software for Population Genetic Analyses. <<http://www.biology.lsu.edu/general/software.html>>

Marmor M., Sheppard H.W., Donnell D., et al. 2001. Homozygous and heterozygous CCR5-Δ32 genotypes are associated with resistance to HIV infection. J. Acquir. Immune Defic. Syndr. 27: 472-481.

Price F. and Vaughan V. 1998. EVOLVE. In *The BioQUEST Library, Vol. VI* (Jungck J.R., Calley J.N., Donovan S., Everse S.J., Soderberg P., Stanley E., Stewart J., eds). Academic Press: San Diego, CA.

Soderberg P. and Price F. 2003. An examination of problem-based teaching and learning in population genetics and evolution using *EVOLVE*, a computer simulation. Int. J. Sci. Educ. 25: 35-55.

Terms and Conditions:

You may use, reproduce, and distribute this module, consisting of both the software and this associated documentation, freely for all nonprofit educational purposes. You may also make any modifications to the module and distribute the modified version. If you do, you must:

- Give the modified version a title distinct from that of the existing document, and from all previous versions listed in the "History" section.
- In the line immediately below the title, replace the existing text (if any) with the text "© YEAR NAME", where YEAR is the year of the modification and NAME is your name. If you would prefer not to copyright your version, then simply leave that line blank.
- Immediately below the new copyright line (even if you left it blank), add or retain the lines:

Original version: *Deme 1.0* © 2004 Anton E. Weisstein
See end of document for full modification history

- Retain this "Terms and Conditions" section unchanged.
- Add to the "History" section an item that includes at least the date, title, author(s), and a description of the modifications, while retaining all previous entries in that section.

These terms and conditions form a kind of "copyleft," a type of license designed for free materials and software. Note that because this section is to be retained, all modified versions and derivative materials must also be made freely available in the same way. This text is based on the GNU Free Documentation License v1.2, available from the Free Software Foundation at <http://www.gnu.org/copyleft/>.

History:

Date: May 3, 2004

Title: *Deme 1.0*

Name: Anton E. Weisstein

Institution: BioQUEST Curriculum Consortium, Beloit College

Acknowledgements: Support for this work was provided, in part, by the National Science Foundation Division of Undergraduate Education, the Howard Hughes Medical Institute, EOT-PACI, and the Shodor Foundation.

Modifications: None (original version).