DeFinetti 1.0

 Anton E. Weisstein, Ed Louis, Tia Johnson, and John R. Jungck
BioQUEST Curriculum Consortium

Summary
DeFinetti is an Excel workbook that simulates evolution of a single gene with three alleles. The user assigns fitness values to each of the six possible genotypes, then enters starting allele frequencies for each of up to 22 individual populations. The program tracks each population through 100 generations and plots the results in a ternary graph. Additional output includes a 3D adaptive landscape and a time plot of local populations' mean fitness, genetic variance in fitness, and expected heterozygosity.

Output: What do I see?

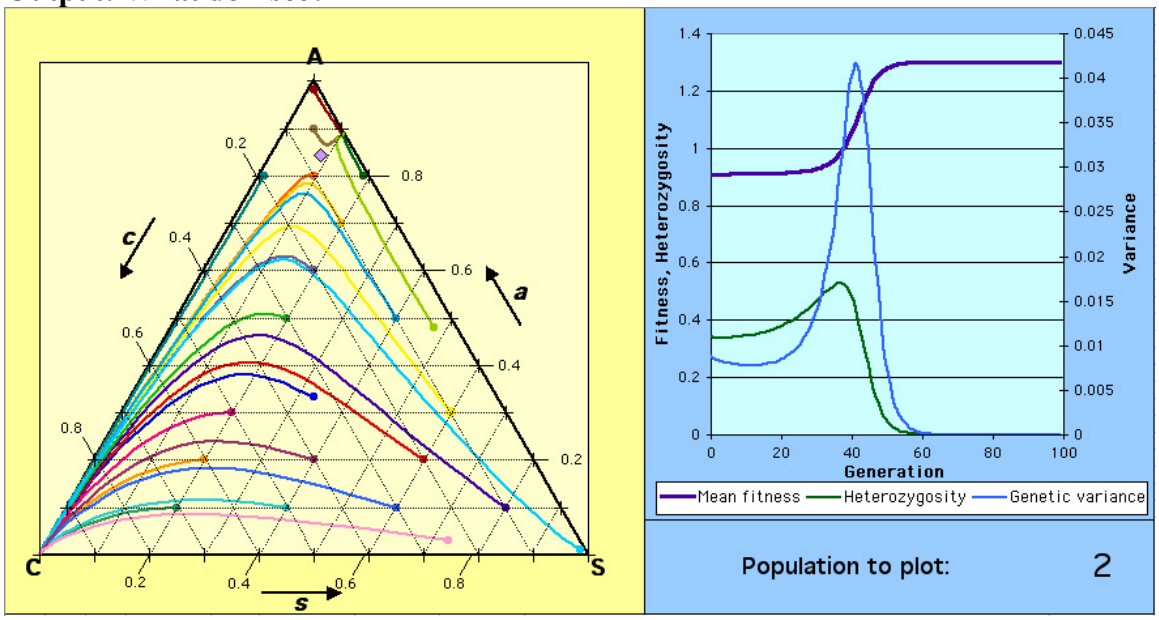

Figure A. A screen shot of *DeFinetti*'s Output worksheet.

The main output of *DeFinetti* consists of two plots on the appropriately-named "Output" worksheet. The yellow plot on the left is a ternary graph of the allele frequencies in each individual population. Each colored dot depicts the initial allele frequency for one population; in the above figure, for example, the purple dot near the lower right-hand corner corresponds to population #14. (To identify the other populations, click in the chart area of the Excel sheet, then move the cursor onto a dot or curve. A small box will pop up and display the appropriate label.) The vertices represent fixation points for each allele; for example, the "A" vertex at the top of the graph represents fixation of the *A* allele ($a = 1$, $c = 0$, $s = 0$).

To determine the allele frequencies for a given point, first trace the dotted lines up and to the left until you reach the "c" axis and note the appropriate value. For example, if we begin with the purple dot (population #14) and then follow the dotted line up and to the left, we intersect the "c" axis at a value of 0.1, so the *C* allele in this particular population has an initial frequency of 0.1. Similarly, tracing the dotted line to the right gives us an initial *A* frequency of 0.1, and tracing the dotted line down and to the left gives us an initial *S* frequency of 0.8. By definition, these frequencies should always add up to one.

Of course, just because the population begins with particular allele frequencies doesn't mean it will stay there! Natural selection acting on each population causes allele frequencies to change over time, resulting in the evolutionary trajectories you see plotted. If you alter the fitnesses of individual genotypes (see the **Controls** section below), this will change the selective forces acting on the individual populations. As a result, they may follow different evolutionary trajectories and even arrive at completely different evolutionary outcomes.

Depending on the genotype fitnesses, a large diamond may also appear in the ternary graph. This diamond represents the system's evolutionary equilibrium. Its color reflects the type of equilibrium: blue indicates a stable fitness peak, purple denotes an unstable saddle point, and red indicates an unstable fitness valley. Sometimes the diamond may lie outside the boundaries of the ternary graph: this is not an error! We leave the interpretation of this result to you as an exercise.

The blue graph on the right plots a single population's mean fitness, genetic variance in fitness, and expected heterozygosity as functions of time. Note that fitness and heterozygosity are plotted on the left-hand y-axis, while variance is plotted on the right-hand y-axis.

The "Interface" worksheet contains another useful plot, a 3D adaptive landscape. (See Figure Y.) You can access this worksheet using the appropriate tab at the bottom of the workbook. The adaptive landscape is a contour map showing the mean fitness of populations as a function of allele frequencies. Cool colors (purple and blue) indicate low fitness, while warm colors (orange and red) denote high fitness. Remember that natural selection acts to increase fitness, so populations will tend to move uphill on this adapative surface: peaks are therefore stable equilibria and valleys are unstable. You can determine allele frequencies for a given point using the same procedure as before, although we were not able to draw the dotted lines on this three-dimensional graph.

The yellow "Equilibrium calculations" table shows the allele frequencies at the system's equilibrium point. The parameters T, U, and V represent partial second derivatives of the fitness surface at the equilibrium point, while Q is the determinant of the corresponding matrix. The values of these parameters hence determine whether the equilibrium is a peak, a saddle point, or a valley within the fitness surface. (See **Model details** for specifics.)

The calculations for the population genetics of each individual population are contained on a separate worksheet. These sheets tabulate numeric values for the following variables:

Column(s)	**Variable**
A	Generation number (starting population is generation 0)
B – G	Genotype frequencies before selection
H – J	Allele frequencies before selection
K	Mean fitness
L	Genetic variance in fitness
N – S	Genotype frequencies after selection
T – V	Allele frequencies after selection
W	Expected heterozygosity
Y & Z	Euclidean x- and y- coordinates corresponding to frequencies in columns U – W.

Table 1. Variables tabulated on the individual population worksheets.

Controls: What can I do?

Allele names:	
Allele #1:	A
Allele #2:	S
Allele #3:	C

Fitnesses:			
AA:	0.90	SS:	0.20
AS:	1.00	SC:	0.71
AC:	0.90	CC:	1.30

Equilibrium calculations:			
A eq:	0.845	T:	-1.800
S eq:	0.090	U:	0.370
C eq:	0.065	V:	1.240
		Q:	-2.369

Initial allele frequencies:	A	S	C
Pop 2:	0.98	0.01	0.01
Pop 3:	0.8	0.1	0.1
Pop 4:	0.9	0.05	0.05
Pop 5:	0.8	0.19	0.01
Pop 6:	0.8	0.01	0.19
Pop 7:	0.333	0.334	0.333
Pop 8:	0.6	0.2	0.2
Pop 9:	0.2	0.6	0.2
Pop 10:	0.2	0.2	0.6
Pop 11:	0.48	0.48	0.04
Pop 12:	0.1	0.2	0.7
Pop 13:	0.1	0.4	0.5
Pop 14:	0.1	0.6	0.3
Pop 15:	0.1	0.8	0.1
Pop 16:	0.3	0.2	0.5
Pop 17:	0.7	0.2	0.1
Pop 18:	0.3	0.6	0.1
Pop 19:	0.5	0.2	0.3
Pop 20:	0.5	0.4	0.1
Pop 21:	0.01	0.98	0.01
Pop 22:	0.2	0.4	0.4
Pop 23:	0.03	0.73	0.24

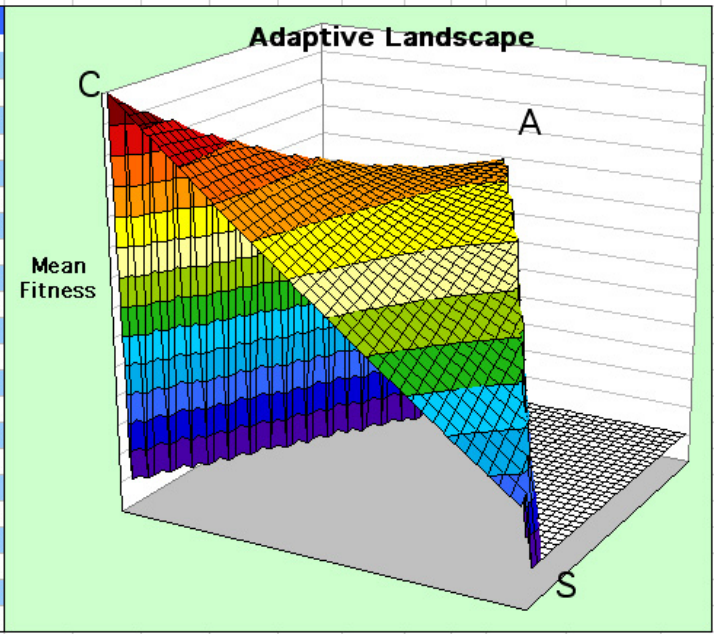

Figure B. A screen shot of *DeFinetti*'s Interface worksheet.

The controls are located on the "Interface" worksheet. Individual parameters are:

Cells	Parameter
B2 – B4	Name of each allele
E2 – E4; G2 – G4	Fitness of each genotype
B8 – D29	Initial allele frequencies in each of up to 22 individual populations

Table 2. *DeFinetti* model parameters.

You can change the values of most parameters by typing the new value directly into the appropriate cell. Note that if you change the name of an allele, *DeFinetti* automatically updates the name throughout the workbook. Remember: by definition, fitnesses can't be negative, and allele frequencies must be lie in the interval [0, 1]. If you enter biologically impossible values for these parameters, the cell will turn red to indicate that there is an error.

To help visualize the adaptive landscape, you can rotate it in three dimensions. Click carefully on one of the corners of the plot; for example, slightly below and to the right of the "Mean Fitness" label. The corners will highlight and the cursor will turn into crosshairs. Then drag one of the corners to rotate the graph. If it doesn't work, you may not have selected the corner correctly: click somewhere else on the worksheet, then try again. The allele labels (for example, "A", "S", and "C" in Figure B) remain fixed, so you will need to make sure that you know which vertex is which.

On the "Output" worksheet, you can plot the fitness, variance, and expected heterozygosity of whichever individual population you choose. Simply enter the population you want to plot into the labeled cell.

How it works: Model details

DeFinetti first calculates the initial genotype frequencies for each individual population. To do this, it assumes that all populations are in Hardy-Weinberg equilibrium: each homozygote frequency is set to the square of the corresponding allele frequency, while each heterozygote frequency is set to twice the product of the two corresponding allele frequencies.

The workbook then tracks each population through one entire generation or life cycle. First, *DeFinetti* calculates each population's mean fitness by multiplying each genotype's frequency by that genotype's fitness, then adding together the resulting products:

$$(1) \qquad \overline{W} = \sum_{i,j} f_{ij} W_{ij} .$$

Similarly, the genetic variance in fitness is calculated by taking the difference between a genotype's fitness and the population's mean fitness, squaring, and multiplying by the genotype's frequency, again summing the results over all genotypes:

$$(2) \qquad \mathrm{Var}_G(W) = \sum_{i,j} f_{ij} (W_{ij} - \overline{W})^2 .$$

The expected heterozygosity is calculated as one minus the summed squared frequency of each allele:

$$(3) \qquad H_E = 1 - \sum_i f_{ii}^2 .$$

This quantity thus measures the frequency of heterozygotes that would be present <u>if</u> the population were in Hardy-Weinberg equilibrium. Note that natural selection will often drive populations out of their initial H-W equilibrium, so the <u>expected</u> heterozygosity will not always match the <u>observed</u> heterozygosity. Because random mating re-establishes H-W equilibrium each generation, however, this deviation will usually be small.

As the next step in the life cycle, *DeFinetti* models the effects of natural selection. Each genotype's post-selection frequency g_{ij} is calculated by multiplying its pre-selection frequency f_{ij} by its fitness W_{ij}, then divided by the population's mean fitness $\overline{W}$. (This last step is necessary to re-normalize the frequencies so that they will still add up to one.)
Each allele's post-selection frequency g_i is then calculated by adding the frequency of that allele's homozygote to one-half times the frequency of each heterozygote involving that allele. Mathematically,

$$(4) \qquad g_{ij} = f_{ij} W_{ij} / \overline{W} , \quad g_a = g_{aa} + \tfrac{1}{2}(g_{ab} + g_{ac}),$$

where *a*, *b*, and *c* represent the three different alleles.

Finally, the workbook completes the life cycle by computing the pre-selection genotype frequencies f'_{ij} in the next generation. The model assumes random mating, so these frequencies are just the Hardy-Weinberg frequencies for the previous generation's post-selection allele frequencies. *DeFinetti* then repeats the above series of calculations to simulate generations 2 through 100.

The ternary plot on the "Output" worksheet is actually a standard scatterplot. *DeFinetti* converts each set of allele frequencies into x- and y-coordinates by placing an allele at each vertex of an equilateral triangle, turning each vertex's coordinates into a vector, weighting each vector according to the allele's frequencies, and adding the three vectors together (Figure C).

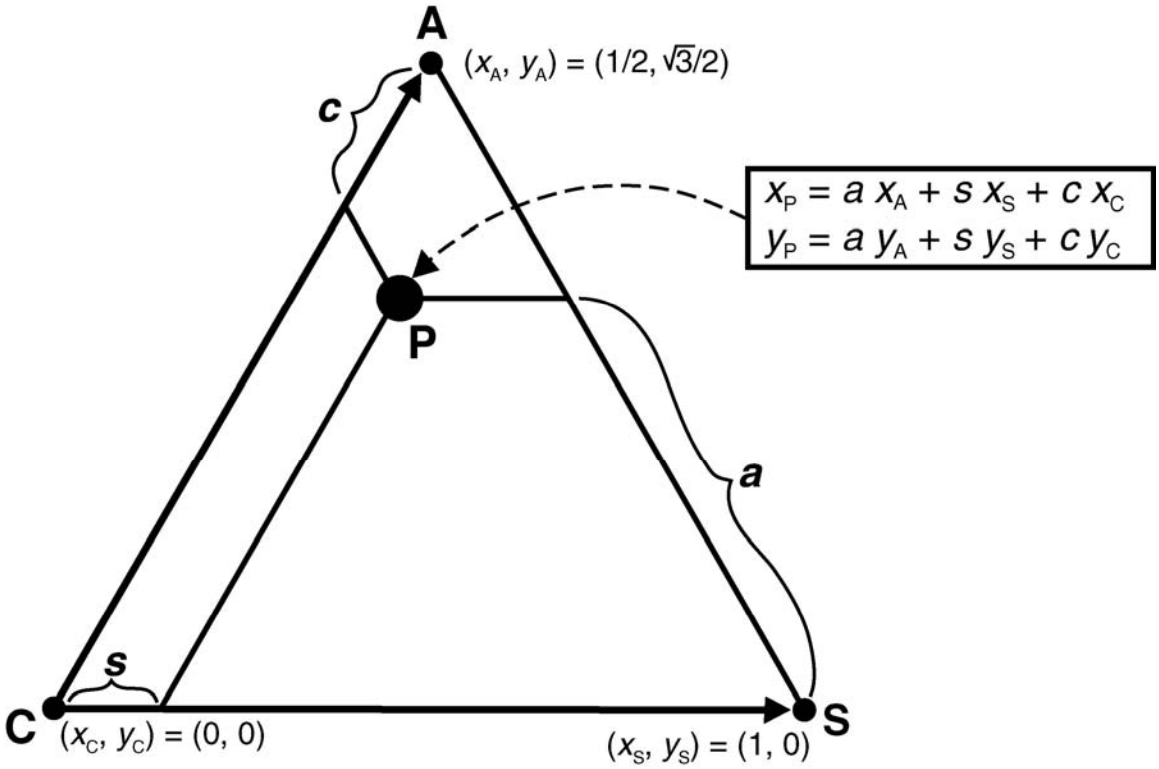

Figure C. Transformation from ternary coordinates to Cartesian coordinates in *DeFinetti*.

The equations for this transformation are

$$(5a) \qquad x = q + \tfrac{1}{2}r, \quad y = \tfrac{1}{2}r\sqrt{3};$$

for the reverse transformation, the equations are

$$(5b) \qquad p = 1 - x - y\sqrt{3}/3, \quad q = x - y\sqrt{3}/3, \quad r = 2y\sqrt{3}/3.$$

The equilibrium calculations follow the method described in Crow (1986). Given allele frequencies p_1, p_2, and p_3, the respective allele frequencies in the next generation are given by

$$(6) \qquad \begin{aligned} p_1' &= p_1(p_1W_{11} + p_2W_{12} + p_3W_{13})/\overline{W}, \\ p_2' &= p_2(p_1W_{12} + p_2W_{22} + p_3W_{23})/\overline{W}, \\ p_3' &= p_3(p_1W_{13} + p_2W_{23} + p_3W_{33})/\overline{W}, \end{aligned}$$

where $\overline{W}$ is the population's mean fitness. By definition, equilibrium is a set of frequencies that remains constant from one generation to the next. We can therefore solve for the equilibrium allele frequencies $\hat{p}_1$, $\hat{p}_2$, and $\hat{p}_3$ by setting each p_i' equal to the corresponding p_i and solving. The resulting solutions are

$$(7) \qquad \hat{p}_1 = D_1/D, \quad \hat{p}_2 = D_2/D, \quad \hat{p}_3 = D_3/D,$$

where

$$(8) \qquad \begin{aligned} D_1 &= (W_{12} - W_{22})(W_{13} - W_{33}) - (W_{12} - W_{23})(W_{13} - W_{23}), \\ D_2 &= (W_{23} - W_{33})(W_{12} - W_{11}) - (W_{23} - W_{13})(W_{12} - W_{13}), \\ D_3 &= (W_{13} - W_{11})(W_{23} - W_{11}) - (W_{13} - W_{12})(W_{23} - W_{12}), \\ D &= D_1 + D_2 + D_3\,. \end{aligned}$$

The next step is to determine whether the equilibrium is a local maximum, a local mininum, or a saddle point of the fitness surface. First, we substitute equations *(5b)* into equation *(1)* to obtain an expression for $\overline{W}$ in terms of x- and y- coordinates. We then calculate the matrix of second partial derivatives (known as the Hessian matrix), obtaining

$$(9) \qquad \begin{bmatrix} \frac{\partial^2 \overline{W}}{\partial x^2} & \frac{\partial^2 \overline{W}}{\partial x \partial y} \\ \frac{\partial^2 \overline{W}}{\partial y \partial x} & \frac{\partial^2 \overline{W}}{\partial y^2} \end{bmatrix} = \begin{bmatrix} T & U \\ U & V \end{bmatrix},$$

where

$$(10) \qquad \begin{aligned} T &= 2(W_{11} - 2W_{12} + W_{22}), \\ U &= \tfrac{2\sqrt{3}}{3}(W_{11} - W_{22} + 2W_{13} + 2W_{23}), \\ V &= \tfrac{2}{3}(W_{11} + W_{22} + 4W_{33} + 2W_{12} - 4W_{13} - 4W_{23}). \end{aligned}$$

Finally, we take the determinant of this matrix, obtaining

$$(11) \qquad \begin{aligned} Q = TV - U^2 = -\tfrac{16}{3}(&W_{12}^2 + W_{13}^2 + W_{23}^2 - W_{11}W_{22} - W_{11}W_{33} - W_{22}W_{33} \\ &+2W_{11}W_{23} + 2W_{22}W_{13} + 2W_{33}W_{13} - 2W_{12}W_{13} - 2W_{12}W_{23} - 2W_{13}W_{23}). \end{aligned}$$

We can now apply the second derivative test to determine the nature of the equilibrium, using the following table:

Criteria	Equilibrium Type	Equilibrium Stability
$Q > 0$, $T > 0$, $T + V > 0$	Local minimum	Unstable
$Q > 0$, $T < 0$, $T + V < 0$	Local maximum	Stable
$Q < 0$	Saddle point	Unstable
$Q = 0$	Higher-order tests needed	

Resources and further reading

Barman, C.R., A. Collins, E.J. Louis, and J.R. Jungck. 1985. "Sickle cell anemia: 'interesting pathology' and 'rarely told stories'". American Biology Teacher 47 (3): 183-187.

Crow, J.F. 1986. *Basic Concepts in Population, Quantitative, and Evolutionary Genetics*, pp. 97-102. W.H. Freeman & Co.: New York.

Crow, J.F., and M. Kimura. 1970. *An Introduction to Population Genetics*. Harper & Row: New York.

Hartl, D.L. 2000. *A Primer of Population Genetics*, 3rd. ed. Sinauer Associates: Sunderland, Mass.

Li, C.C. 1976. *First Course in Population Genetics.* Boxwood Press: Pacific Grove, California.

Templeton, A.R. 1982. "Adaptation and integration of evolutionary forces." Pages 15-31 in: Milkman, R., ed. *Perspectives on Evolution.* Sinauer Associates: Sunderland, Mass.

History:
Date: July 28, 2004
Title: *DeFinetti 1.0*
Name: Anton E. Weisstein, Ed Louis, Tia Johnson, and John R. Jungck
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, and EOT-PACI.
Modifications: None (original version).