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 adaptive 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:

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

Table 1. Variables tabulated on the individual population worksheets.
Controls: What can I do?

<table>
<thead>
<tr>
<th>Allele names:</th>
<th>Fitnesses:</th>
<th>Equilibrium calculations:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allele #1:</td>
<td>AA: 0.99</td>
<td>A eq: 0.845, T: -1.900</td>
</tr>
<tr>
<td>Allele #2:</td>
<td>AS: 1.00</td>
<td>S eq: 0.090, U: 0.370</td>
</tr>
<tr>
<td>Allele #3:</td>
<td>AC: 0.99</td>
<td>C eq: 0.065, V: 1.240</td>
</tr>
</tbody>
</table>

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

- **Cells:**
  - 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

Figure B. A screen shot of DeFinetti’s Interface worksheet.

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 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:

\[
W = \sum_{i,j} f_{ij} W_{ij}.
\]

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:

\[
\text{Var}_G(W) = \sum_{i,j} f_{ij} (W_{ij} - \bar{W})^2.
\]

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

\[
H_E = 1 - \sum_i f_{ii}^2.
\]

This quantity thus measures the frequency of heterozygotes that would be present if the population were in Hardy-Weinberg equilibrium. Note that natural selection will often drive populations out of their initial H-W equilibrium, so the expected heterozygosity will not always match the observed 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 \(\bar{W}\). (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,

\[
g_{ij} = f_{ij} \frac{W_{ij}}{\bar{W}}, \quad g_a = g_{aa} + \frac{1}{2} (g_{ab} + g_{ac}),
\]

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).

The equations for this transformation are

\[
(5a) \quad x = q + \frac{1}{2} r, \quad y = \frac{1}{2} r \sqrt{3};
\]

for the reverse transformation, the equations are

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

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

\[
(6) \quad p'_1 = p_1 (p_1 W_{11} + p_2 W_{12} + p_3 W_{13}) / \bar{W}, \quad p'_2 = p_2 (p_1 W_{12} + p_2 W_{22} + p_3 W_{23}) / \bar{W}, \quad p'_3 = p_3 (p_1 W_{13} + p_2 W_{23} + p_3 W_{33}) / \bar{W},
\]

where \(\bar{W}\) 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

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

where

\[
D_1 = (W_{12} - W_{22})(W_{13} - W_{33}) - (W_{12} - W_{23})(W_{13} - W_{23}),
\]

\[
D_2 = (W_{23} - W_{33})(W_{12} - W_{11}) - (W_{23} - W_{13})(W_{12} - W_{13}),
\]

\[
D_3 = (W_{13} - W_{11})(W_{23} - W_{11}) - (W_{13} - W_{12})(W_{23} - W_{12}),
\]

\[
D = D_1 + D_2 + D_3.
\]
The next step is to determine whether the equilibrium is a local maximum, a local minimum, or a saddle point of the fitness surface. First, we substitute equations (5b) into equation (1) to obtain an expression for $W$ in terms of $x$- and $y$-coordinates. We then calculate the matrix of second partial derivatives (known as the Hessian matrix), obtaining

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

where

$$T = 2(W_{11} - 2W_{12} + W_{22}),$$

$$U = \frac{2}{3}(W_{11} - W_{22} + 2W_{13} + 2W_{23}),$$

$$V = \frac{2}{3}(W_{11} + W_{22} + 4W_{33} + 2W_{12} - 4W_{13} - 4W_{23}).$$

Finally, we take the determinant of this matrix, obtaining

$$Q = TV - U^2 = -\frac{16}{3}(W_{12}^2 + W_{13}^2 + W_{23}^2 - W_{11}W_{22} - W_{11}W_{33} - W_{22}W_{33} + 2W_{11}W_{23} + 2W_{22}W_{13} + 2W_{33}W_{13} - 2W_{12}W_{13} - 2W_{12}W_{23} - 2W_{13}W_{23}).$$

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

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Equilibrium Type</th>
<th>Equilibrium Stability</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Q &gt; 0$, $T &gt; 0$, $T + V &gt; 0$</td>
<td>Local minimum</td>
<td>Unstable</td>
</tr>
<tr>
<td>$Q &gt; 0$, $T &lt; 0$, $T + V &lt; 0$</td>
<td>Local maximum</td>
<td>Stable</td>
</tr>
<tr>
<td>$Q &lt; 0$</td>
<td>Saddle point</td>
<td>Unstable</td>
</tr>
<tr>
<td>$Q = 0$</td>
<td>Higher-order tests needed</td>
<td></td>
</tr>
</tbody>
</table>

**Resources and further reading**


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: DeFinetti 1.0 © 2004
    Anton E. Weisstein, Ed Louis, Tia Johnson, and John R. Jungck
    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: 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).