What is epistasis?

Epistasis can produce wholes greater than the sum of their parts.

...a phenomenon where an expression of an organism is the result of a combination of interacting genes.
Example of Epistasis

Epistasis drives variation of dog coat color
Does nature have a preference for epistasis?

In other words, on the whole, will natural selection promote genetic configurations having a higher degree of genetic interaction over configurations that are less interdependent?
Variables and Methods

**Variables**

\[ N = \text{number of loci within a species genome} \]

\[ K = \text{epistatic value of the organism } \{ K \mid 0 < K \leq N - 1 \} \]

\[ v_i = \text{locus fitness value } \{ v_i \mid -1, 0, 1 \} \]

\[ \psi_{i,j} = \text{epistatic fitness coefficient} \]

**Fitness Calculations**

**Case 1: No Epistasis (K=0):**

Total Organism Fitness = \( \sum_{i=0}^{n} v_i \)

**Case 2: Epistasis (0 < K \leq N-1):**

Total Organism Fitness = \( \sum_{i=0}^{n} v_i + \sum_{i=0}^{n} \sum_{j=0}^{k+1} \psi_{i,j} v_{i,j} \)

**Key Methods**

- Each organism is haploid and has 8 genes.
- Each offspring is given a 6.67% chance for stepwise mutation of its K value.
- Replication may only occur when an organism has a total fitness value > 80.
- Organisms with total fitness values between 0 and 79 have a chance of being eliminated at each pass.
- Organisms having a total fitness value < 0 are instantly eliminated.
Indexed Lookup Tables

- Each $K$ has its own lookup table consisting of $2^{k+1}$ rows representing every possible combination of Boolean values.

- A single uniform probability distribution is used to generate random fitness coefficients ($\psi_n$) for each row in every table.

- The range of fitness values is identical for every table, regardless of the number of rows.

$K = 1$

$\begin{array}{ccc}
0 & 0 & \psi_1 \\
0 & 1 & \psi_2 \\
1 & 0 & \psi_3 \\
1 & 1 & \psi_4 \\
\end{array}$

$K = 2$

$\begin{array}{ccc}
0 & 0 & 0 & \psi_1 \\
0 & 1 & 0 & \psi_2 \\
0 & 0 & 1 & \psi_3 \\
0 & 1 & 1 & \psi_4 \\
1 & 0 & 0 & \psi_5 \\
1 & 1 & 0 & \psi_6 \\
1 & 0 & 1 & \psi_7 \\
1 & 1 & 1 & \psi_8 \\
\end{array}$
Software Framework

Simulation Engine (SE)
- C PROGRAM
  - *.csv output

Analysis Engine (AE)
- MYSQL
- R / SPREADSHEET

Forward Time Simulation
- PASS 1
- PASS 2
- PASS 3
- PASS 4
- PASS 5
- PASS 6
- PASS 7
Samples containing ~20 identically configured simulations for three distinct durations (time spans) were compared.

### SAMPLE COMPARISON

<table>
<thead>
<tr>
<th></th>
<th>SAMPLE 1</th>
<th>SAMPLE 2</th>
<th>SAMPLE 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DURATION</strong></td>
<td>20 passes</td>
<td>25 passes</td>
<td>30 passes</td>
</tr>
<tr>
<td><strong># SIMULATIONS</strong></td>
<td>19 simulations*</td>
<td>20 simulations</td>
<td>19 simulations*</td>
</tr>
<tr>
<td><strong>AVG FITNESS, START</strong></td>
<td>18.262</td>
<td>-19.416</td>
<td>-20.676</td>
</tr>
<tr>
<td><strong>AVG FITNESS, END</strong></td>
<td>364.674</td>
<td>354.490</td>
<td>351.545</td>
</tr>
<tr>
<td><strong>AVG FITNESS, CHANGE</strong></td>
<td>346.412</td>
<td>373.907</td>
<td>372.220</td>
</tr>
<tr>
<td><strong>AVG INITIAL COUNT</strong></td>
<td>10 organisms</td>
<td>10 organisms</td>
<td>10 organisms</td>
</tr>
<tr>
<td><strong>AVG FINAL COUNT</strong></td>
<td>330 organisms</td>
<td>332 organisms</td>
<td>503 organisms</td>
</tr>
</tbody>
</table>
### Results: Zooming In on Each Sample

<table>
<thead>
<tr>
<th># OF PASSES</th>
<th>FIRST PASS</th>
<th>LAST PASS</th>
<th>LAST - FIRST</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>3.421</td>
<td>3.965</td>
<td>0.544</td>
</tr>
<tr>
<td>25</td>
<td>3.565</td>
<td>4.314</td>
<td>0.749</td>
</tr>
<tr>
<td>30</td>
<td>3.563</td>
<td>4.380</td>
<td>0.817</td>
</tr>
</tbody>
</table>

#### AVERAGE K, BEGINNING OF SIMULATION VS END OF SIMULATION

<table>
<thead>
<tr>
<th>NUMBER OF PASSES</th>
<th>AVERAGE K</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>3.421</td>
</tr>
<tr>
<td>25</td>
<td>3.565</td>
</tr>
<tr>
<td>30</td>
<td>3.563</td>
</tr>
</tbody>
</table>

Over the duration of each simulation:

- 20 pass sample had a 15.9% increase in K
- 25 pass sample had a 21.0% increase in K
- 30 pass sample had a 22.9% increase in K
Summary

The results suggest that:

• The mechanics of natural selection provide an arithmetic incentive for epistasis.

• Epistatic networks may leverage mutations for outsized fitness gains (aka “radical differentiation”).

• Outsized fitness gains may fragment a population and drive speciation.
Acknowledgements

Tracy Davidson, InTech Collegiate High School science teacher

Dr. Zach Gompert, Biology, Utah State University

Dr. Christoph Adami, Microbiology, Molecular Genetics, Physics and Astronomy, Michigan State University

Dr. Stephen Stearns, Ecology and Evolutionary Biology, Yale University
Cagnoni, S., Midolli, M., Villani, M. (2014), Evolution, Complexity and Artificial Life


DeDeo, S., Krakauer, D.C., Flack, J.C. (2010), Inductive Game Theory and the Dynamics of Animal Conflict


Gilbert, K.J., Whitlock, M.C. (2016), The Genetics of Adaptation to Discrete Heterogeneous Environments: Frequent Mutation or Large-Effect Alleles Can Allow Range Expansion

LaBar, T., Adami, C. (2016), Different Evolutionary Paths to Complexity for Small and Large Populations of Digital Organisms

Bibliography


Payne, R.J.H., Krakauer, D.C. (1997), Sexual Selection, Space, and Speciation

Schmiegelt, B., Krug, J. (2013), Evolutionary Accessibility of Modular Fitness Landscapes


Vincent, T.L. & Brown, J.S. (2005), Evolutionary Game Theory, Natural Selection and Darwinian Dynamics

Watkins, K.S., Rose, K.A. (2017), Simulating Individual-Based Movement in Dynamic Environments

Watts, D.J., (2003), Six Degrees: The Science of a Connected Age