The Structure, Function, and Evolution of Biological Systems

Instructor: Van Savage

Spring 2017 Quarter

4/12/2017
Decay Test—Modeling more than two mutations with interactions

\[ \omega_{epi} = 1 - sk^{1+\delta} \sim (1-s)^{k^{1+\delta}} \sim e^{-sk^{1+\delta}} \]
Pair Test—Measures of epistasis (interactions)

In symbols, no interaction means

\[ \frac{w_{xy}}{w_x w_y} \]

So if the difference does not equal zero, this means there is an interaction. Measure of deviation from additivity

\[ DA = \frac{w_{xy}}{w_x w_y} - \frac{w_x w_y}{w_{xy}} \]

This is known as additive (or sometimes multiplicative).
How do we interpret synergy and antagonism?

Mutations to steps in sequence are antagonistic (green). Mutations to steps in parallel can be synthetically lethal if it knocks out a loop, which is extreme synergy (red), or multiplicative (black).
Papers using models of epistasis: Lenski, Ofria, Collier, Adami
Definition of digital organism

Carrying capacity of 3600 individuals

Probability of point mutation was 0.0075 per instruction copied and probability of insertion and deletion is 0.05 per division

Each generation is 5-10 updates and each update is execution of on average 30 instructions per individual

Start with genome length of 20 instructions

28 different types of instructions (like amino acids)

Phenotypic rewards are multiplicative

Instructions are mathematical operations
Advantages of digital organisms

1. Allow us to choose environment and seek generalizations beyond organic life forms

2. Allow us to perform experiments, in terms of time scales and numbers, that are unattainable with real systems

3. Use evolving programs to solve computational problems
Complex organisms

Selection criteria:

1. Baseline allocation of CPU time (fitness) is proportional to genome size

   Why? Larger genomes do not necessarily imply more complex or better at “solving problems” or getting resources. See next plot.

2. Certain mathematical operations, which require novel combinations of instructions (e.g., performing an XOR operation using NAND operations), are rewarded with additional CPU time.

   This is a type of selection. Solving computational problem is solving fitness problem or getting more resources.
Simple organisms

Selection criteria starting with complex organisms:

1. Baseline allocation of CPU time is independent of genome size

2. Mathematical operations are not rewarded with additional CPU time

   Removes a type of selection. Does not seem biological. But, there is still selection for shorter replication time, so some biological analogy.
Complex vs. Simple organisms

Complex organisms average genome size = 91.3 instructions (Because of assumption 1 or 2?)

Simple organisms average genome size = 19.8 instructions
Simple organisms have more lethal mutations
Decay Test-Tests for epistasis (interactions)

Decay test—see whether successive mutations (1-10) become increasingly worse (synergy), better (antagonistic), or are multiplicative.

\[ w_{epi} = 1 - sk^{1+\delta} \sim (1 - s)^{k^{1+\delta}} \sim e^{-sk^{1+\delta}} \]

Both show an average type of antagonistic epistasis. But average can be quite misleading. Antagonism and synergy can balance out of average.
Pair tests for epistasis—average is misleading! LOTS of interactions!
Pair test—Explicitly calculate and compare double mutant fitness \( (w_{xy}) \) to the product of each single mutant fitness \( (w_xw_y) \) for each fitness for all pairs (double mutants).
M=Multiplicative (same as additive!)
(no interaction), A=Antagonism,
S=Synergism, L=Lethal, E=Epistatic = S+A

\[
DA = w - w \cdot w
\]

Synergy and antagonism balancing out.
<table>
<thead>
<tr>
<th>Response variable</th>
<th>Mean complex (± s.d.)</th>
<th>Mean simple (± s.d.)</th>
<th>Mean difference (± s.d.)</th>
<th>$P^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genome length</td>
<td>91.25 (69.07)</td>
<td>19.80 (14.18)</td>
<td>71.45 (64.76)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Decay test, $\alpha$</td>
<td>0.581 (0.207)</td>
<td>1.141 (0.591)</td>
<td>-0.560 (0.562)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Decay test, $\beta$</td>
<td>0.896 (0.081)</td>
<td>0.972 (0.192)</td>
<td>-0.077 (0.201)</td>
<td>0.0011</td>
</tr>
<tr>
<td>Pair test, proportion epistatic of total</td>
<td>0.191 (0.093)</td>
<td>0.045 (0.080)</td>
<td>0.146 (0.122)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Pair test, proportion epistatic of non-lethal</td>
<td>0.743 (0.243)</td>
<td>0.781 (0.234)</td>
<td>-0.038 (0.303)</td>
<td>0.4374</td>
</tr>
<tr>
<td>Pair test, proportion synergistic of epistatic</td>
<td>0.271 (0.093)</td>
<td>0.168 (0.159)</td>
<td>0.103 (0.175)</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

*Two-tailed Wilcoxon signed-ranks test of the differences between 87 paired complex and simple organisms.*
Recent papers using models of epistasis: Segre, DeLuna, Church, Kishony
Quantitative Epistatic Interactions

Perturbation X

Perturbation Y

Phenotype (Growth Rate)

Synergy

Antagonism

Additivity

Masking

Synthetic Lethality

See also:
Interactions between mutations in yeast metabolism using Flux Balance Analysis

Varma and Palsson, 1994
Famili et al, 2003

- 829 - metabolic reactions
- 343206 gene pairs

Rate of biomass production
(growth rate)
Flux Balance Analysis

• Computational model of metabolism

• Growth rate predictions for wild-type and deletion mutants

• Main Assumptions:
  – Steady-state
  – Mass-conservation
  – Optimality

• Developed and experimentally verified in *E. coli* and yeast by Palsson et al:

  *Nature* 2002; *PNAS* 2003; *Nat. Genetics* 2004
Measures of yeast epistasis

Sort of unimodal distribution
Measures of epistasis for RNA viruses
Look more carefully at distribution of interactions! We have non-zero and sign information? What does magnitude tell us? Strength of interactions?

Frequency
(Percent or fraction of total possible interactions with this value)

How would you interpret this?
Same type and information can be mapped to several different points

(Lethal Synergy)
DA = 0 - (0.9)(0.9) = -0.81
DA = 0 - (0.5)(0.5) = -0.25
DA = 0 - (0.1)(0.1) = -0.01

(Antagonistic Buffering)
DA = 0.9 - (0.9)(0.9) = 0.09
DA = 0.5 - (0.5)(0.5) = 0.25
DA = 0.1 - (0.1)(0.1) = 0.09

Frequency
(Percent or fraction of total possible interactions with this value)
Same point can represent very different types and information

synergistic lethal
DA = 0 - (0.5)(0.5) = -0.25
just a little synergistic
DA = 0.39 - (0.8)(0.8) = -0.25

antagonistic buffering
DA = 0.1 - (0.1)(0.1) = 0.09
just a little antagonistic
DA = 0.73 - (0.8)(0.8) = 0.09
Need to rescale or make magnitude represent information we want

\[ D \tilde{A} = \frac{DA}{\tilde{w}_{xy} - w_x w_y} = \frac{w_{xy} - w_x w_y}{\tilde{w}_{xy} - w_x w_y} \]

Want to map synergistic lethals to -1, antagonistic buffering to 1, no interaction to 0, and everything else in between.
choices of rescaling

Additive

\[ |\tilde{w}_{xy} - w_x w_y| = 0 \]

All synergistic is binned as lethal

\[ |\tilde{w}_{xy} - w_x w_y| = |0 - w_x w_y| = w_x w_y \]

All antagonistic is binned as buffering \((w_{xy} = w_x < w_y)\)

\[ |\tilde{w}_{xy} - w_x w_y| = w_x - w_x w_y \]
limiting cases

Additive

\[ D\widetilde{A} = \frac{w_{xy} - w_x w_y}{|\tilde{w}_{xy} - w_x w_y|} = 0 \]

Synergistic lethal

\[ D\widetilde{A} = \frac{-w_x w_y}{|-w_x w_y|} = -1 \]

Antagonistic buffering (\(w_{xy} = w_x < w_y\))

\[ D\widetilde{A} = \frac{w_x - w_x w_y}{|w_x - w_x w_y|} = 1 \]
Measures of epistasis

Since covariance is as fundamental as fitness, why not define relative covariance instead of relative fitness. We define it relative to tri-modally binned covariance that itself varies, so relative to a shifting baseline.

Absolute covariance

\[ DA = \text{Cov}(w_x, w_y) = w_{xy} - w_x w_y \]

Relative covariance

\[ D\tilde{A} = \frac{\text{Cov}(w_x, w_y)}{\text{BinnedCov}(w_x, w_y)} = \frac{w_{xy} - w_x w_y}{\tilde{w}_{xy} - w_x w_y} \]
Measures of epistasis

Additive

\[ |BinnedCov(w_x, w_y)| = w_{xy} - w_x w_y = 0 \]

Synergistic is binned as synthetic lethal

\[ |BinnedCov(w_x, w_y)| = |0 - w_x w_y| = w_x w_y \]

Antagonistic is binned as buffering \((w_{xy}=w_x<w_y)\)

\[ |BinnedCov(w_x, w_y)| = w_x - w_x w_y \]
Can think of this as relative fitness being relative to product of fitnesses and again a shifting baseline

Additive

\[
\left| \text{BinnedCov}(w_x, w_y) \right| = 0
\]

Synergistic is binned as synthetic lethal

\[
\tilde{\varepsilon} = w_{xy}^r - 1 \sim -1
\]

Antagonistic is binned as buffering \((w_{xy}=w_x<w_y)\)

\[
\tilde{\varepsilon} = \frac{w_{xy}^r - 1}{w_x^r - 1} \sim 1
\]
Measures of yeast epistasis

Sort of unimodal distribution goes to trimodal distribution
Measures of epistasis for RNA viruses
Higher level epistasis—interactions among functional groups rather than loci

Interactions are mostly monochromatic. No reason a priori that this should be, except it signifies functional organization.

\[
\text{Cov}(W_{\text{ModA}}, W_{\text{ModB}}) = W_{\text{ModA}, \text{ModB}} - W_{\text{ModA}} W_{\text{ModB}}
\]
Do clustered groups correspond to functionally annotated groups?

Statistical Test:  
\[ N = \text{total gene pairs, 1034} \]
\[ n = \text{interacting pairs, 278} \]
\[ K = \text{same-annotation pairs, 104} \]

\[ P_{ANN} = 1 - \sum_{i=0}^{k-1} \frac{K^i (N - K)^i}{n^i (n - i)^i} / \binom{N}{n} \]

Each term is probability of choosing \( i \) same-annotation pairs out of subset of \( n \) pairs relative to any subset of \( n \) pairs being the interacting ones. Sum is probability of having \( k-1 \) or less.
Can we do reverse and cluster monochromatically to find functional groups?

Construct network for all pairwise interactions, start with each gene in its own group. Cluster by pairs if they interact with other genes in the same way. Require monochromaticity, each group must interact with all other groups in the same way. Within a group there is no requirement for monochromaticity. Make cluster sizes as large as possible.
Can we do reverse and cluster monochromatically to find functional groups?

How clusterable are networks?
Is clustering unique?
If not, which instantiation is chosen?
Overlay of monochromatic and functionally annotated groups?
A system-level view of modular genetic interactions in yeast metabolism
Monochromatic organization exists in the yeast metabolic network, but is very unlikely in random networks.
Identifying Biological Functions in Higher Hierarchical Levels

Fermentation

Respiration

ACAL

GLUCN

GLYC

PENT

ETHxt

ATPs

RESPIR

TCA
Measures of epistasis

How does $\varepsilon$ vary with relative covariance?

$\tilde{\varepsilon} \sim \varepsilon^2$

Will accentuate differences in distribution
Future directions research using digital organisms

1. I would use mixed selection criteria because I think they are more biological:

   a. Baseline allocation of CPU time is independent of genome size or decreases with genome size

   b. Mathematical operations are rewarded with additional CPU time

2. Run Prism algorithm on existing digital data. Is it clusterable? What would make it clusterable?