Molecular evolution 1

Please sit in row H or forward
Random biological fact of the day: the Venus fly trap can count

“The Venus Flytrap Dionaea muscipula Counts Prey-Induced Action Potentials to Induce Sodium Uptake”
Topics for today

• Enzymes, inhibitors, and AZT
• The nucleotide substitution rate, with and without selection
• A little reflection: mechanisms of sequence evolution
Early efforts at therapy: targeting reverse transcriptase
The activity of the reverse transcriptase enzyme

Reverse transcriptase is an enzyme, like carbonic anhydrase...

https://pdb101.rcsb.org/motm/49
Azidothymidine (AZT): how it works

• Reverse transcriptase prone to incorporating AZT
• Native polymerases less so
• Once AZT incorporated, no further extension possible
The failure of AZT

Topics for today

• Enzymes, inhibitors, and AZT
• The nucleotide substitution rate, with and without selection
• A little reflection: mechanisms of sequence evolution
Sequences and populations differ over time
The substitution rate: measuring how sequences evolve over time

A simulation with ongoing mutation

Generation
The fate of a newly arising mutation

Several generations
The fate of a newly arising mutation:

- **Fixed**
- **Polymorphic**
- **Lost**

Several generations
In the long run all new mutations are either fixed or lost.
Probability of fixation of a neutral allele

Definition: A neutral allele is an allele that confers the same fitness as other alleles present.
In the long run all new mutations are either fixed or lost.

Probability of fixation of new allele in a population of size $N$: $1/N$ (assumes haploid organisms)
Definition: a substitution is a new point mutation that becomes fixed
Counting substitutions in a simulation

Is this a substitution?

How about this?
Counting substitutions in a simulation

What do we need to know to determine if a substitution has occurred at a position?
Mutation probability and the expected number of substitutions at a single neutral site

Each of the three nucleotides in this example has probability $\mu$ of mutating during a single replication.
Population size is 4, so we have 4 replications per generation here...
$\mu$ is also the expected number of mutations per site per replication
Expected number of substitutions at a single neutral site

\[
\text{Substitution rate} = \frac{\text{Expected number of substitutions per site per generation}}{\text{Number of mutations per site per generation}} \times \text{Expected proportion of mutations that will fix}
\]
Expected number of substitutions at a single neutral site

\[
\text{Substitution rate} = \text{Expected number of substitutions per site per generation} = \text{Number of mutations per site per generation} \times \text{Expected proportion of mutations that will fix}
\]

\[
= \mu \frac{\text{mutations}}{\text{site} \times \text{replication}} \times N \frac{\text{replications}}{\text{generation}} \times \frac{1}{N} \frac{\text{substitutions}}{\text{mutation}}
\]

\[
= \mu \frac{\text{substitutions}}{\text{site} \times \text{generation}}
\]
Considering selection on a single codon

Hypothetical selection regime acting at this amino acid site:

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Fitness</th>
</tr>
</thead>
<tbody>
<tr>
<td>All amino acids except proline</td>
<td>0.9</td>
</tr>
<tr>
<td>Proline</td>
<td>1.0</td>
</tr>
</tbody>
</table>
Amino acid Fitness

All amino acids except proline 0.9
Proline 1.0
Selection and the substitution rate, some definitions

**Purifying selection.** Occurs when new mutations are not advantageous. Selection acts to remove new mutations.

**Positive selection.** Occurs when new mutations are advantageous. Selection drives these advantageous mutations to fixation.
Consider coding sequence with a CCC codon (coding for proline). The starting population consists of only CCC. We then subject it to a simulation with ongoing mutations. For the three selection regimes below:

• Rank the three cases in terms of expected substitution rate
• Tell what kind of selection (if any) is going on in each case?

Be sure to explain your answers.

<table>
<thead>
<tr>
<th>Case A</th>
<th>Case B</th>
<th>Case C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amino acid</td>
<td>Fitness</td>
<td>Amino acid</td>
</tr>
<tr>
<td>All others</td>
<td>0.9</td>
<td>All others</td>
</tr>
<tr>
<td>Proline</td>
<td>1.0</td>
<td>Proline</td>
</tr>
</tbody>
</table>
Consider coding sequence with a CCC codon (coding for proline). The starting population consists of only CCC. We then subject it to a simulation with ongoing mutations. For the three selection regimes below:

- Rank the three cases in terms of expected substitution rate
- Tell what kind of selection (if any) is going on in each case?

Be sure to explain your answers.

**Worksheet**

(Rip it off from the back of your packet)

Consider coding sequence with a CCC codon (coding for proline). The starting population consists of only CCC. We then subject it to a simulation with ongoing mutations. For the three selection regimes below:

- Rank the three cases in terms of expected substitution rate
- Tell what kind of selection (if any) is going on in each case?

Be sure to explain your answers.

### Case A

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Fitness</th>
</tr>
</thead>
<tbody>
<tr>
<td>All others</td>
<td>0.9</td>
</tr>
<tr>
<td>Proline</td>
<td>1.0</td>
</tr>
</tbody>
</table>

### Case B

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Fitness</th>
</tr>
</thead>
<tbody>
<tr>
<td>All others</td>
<td>1.0</td>
</tr>
<tr>
<td>Proline</td>
<td>0.9</td>
</tr>
</tbody>
</table>

### Case C

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Fitness</th>
</tr>
</thead>
<tbody>
<tr>
<td>All others</td>
<td>1.0</td>
</tr>
<tr>
<td>Proline</td>
<td>1.0</td>
</tr>
</tbody>
</table>

**Likelihood of substitution:**

B > C > A

A: purifying selection
B: positive selection
C: no selection
A note on the homework: comparing substitution rates in our simulations

>>> startAllele='CCC'
>>> popSize=50
>>> numGens=400
>>> mutProb=0.0003
>>> numReps=120
>>> selectOn=False
>>> subL=subSim(startAllele,popSize,numGens,mutProb,numReps,selectOn,{})
>>> printSummary(subL,".5f")
0 0.00025 (0.00010–0.00040)
1 0.00019 (0.00007–0.00031)
2 0.00021 (0.00008–0.00033)

• Compare substitution rates by checking if the 95% confidence intervals overlap
• When in doubt, run it again!
Topics for today

• Enzymes, inhibitors, and AZT
• The nucleotide substitution rate, with and without selection
• A little reflection: mechanisms of sequence evolution
A little reflection: mechanisms of sequence evolution

HIV evolution in response to immune system and drugs

Four forces

Example of process: population genetic simulator written in hw2

Changed sequences (e.g. substitutions have occurred)
From populations to trees: tree from a sample of 3 HIV viruses taken from a single patient
The history underlying the tree

Viruses evolving inside patient according to the four forces.
Ancestral virus with descendants in two lineages.
From populations to trees: tree from a sample of 3 HIV viruses taken from 3 patients
The history underlying the tree

Transmission event from one person to another

HIV samples taken
A little reflection: mechanisms of sequence evolution

Themes
- Evolution as organizing principle
- Sequence information has many uses
- Molecular shape → molecular function
- Importance of stochastic processes
“Nothing in biology makes sense except in the light of evolution.”
Hand in your worksheet please!

(and be sure you put your name on it)