The goals and the plan

- neutral theory
- dN/dS
- mechanistic process
- phenomenological outcomes

part 1: introduction
part 2: mechanistic process
part 3: data analysis
part 4: phenomenological load

- MutSel framework
  - freq dependent selection
  - episodic selection
  - shifting balance

- types of models
- 3 analysis tasks
- analysis of deviance
- biological inferences
part 1: introduction

Evolutionary rate depends on intensity of selection.

- Selectively constrained = slower than neutral (drift alone)
- Adaptive divergence = faster than neutral (drift alone)

**conserved sites:** slower than neutral?

**fast sites:** neutral? or faster than neutral?

What is the neutral expectation?
neutral theory of molecular evolution (Kimura 1968)

The number of new mutations arising in a diploid population

\[ 2N\mu \]

The fixation probability of a new mutant by drift

\[ \frac{1}{2N} \]

The substitution (fixation) rate, \( k \)

\[ k = 2N\mu \times \frac{1}{2N} \]

The elegant simplicity of neutral theory: \( k = \mu \)

---

genetic code determines impact of a mutation

Kimura (1983)

\( d_S \): number of synonymous substitutions per synonymous site (\( K_S \))

\( d_N \): number of nonsynonymous substitutions per nonsynonymous site (\( K_A \))

\( \omega \): the ratio \( d_N/d_S \); it measures selection at the protein level

The genetic code determines how random changes to the gene brought about by the process of mutation will impact the function of the encoded protein.
an index of selection pressure

<table>
<thead>
<tr>
<th>rate ratio</th>
<th>mode</th>
<th>example</th>
</tr>
</thead>
<tbody>
<tr>
<td>$dN/dS &lt; 1$</td>
<td>purifying (negative) selection</td>
<td>histones</td>
</tr>
<tr>
<td>$dN/dS = 1$</td>
<td>Neutral Evolution</td>
<td>pseudogenes</td>
</tr>
<tr>
<td>$dN/dS &gt; 1$</td>
<td>Diversifying (positive) selection</td>
<td>MHC, Lysin</td>
</tr>
</tbody>
</table>

Why use $d_N$ and $d_S$?
(Why not use raw counts?)

example of counts:
- 300 codon gene from a pair of species
- 5 synonymous differences
- 5 nonsynonymous differences

$$5/5 = 1$$

why don't we conclude that rates are equal (i.e., neutral evolution)?
the genetic code & mutational opportunities

<table>
<thead>
<tr>
<th>Type</th>
<th>Expected number of changes (proportion)</th>
<th>All 3 Positions</th>
<th>1st positions</th>
<th>2nd positions</th>
<th>3rd positions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total mutations</td>
<td>549 (100)</td>
<td>183 (100)</td>
<td>183 (100)</td>
<td>183 (100)</td>
<td></td>
</tr>
<tr>
<td>Synonymous</td>
<td>134 (25)</td>
<td>8 (4)</td>
<td>0 (0)</td>
<td>126 (69)</td>
<td></td>
</tr>
<tr>
<td>Nonsynonymous</td>
<td>392 (71)</td>
<td>166 (91)</td>
<td>176 (96)</td>
<td>57 (27)</td>
<td></td>
</tr>
<tr>
<td>nonsense</td>
<td>23 (4)</td>
<td>9 (5)</td>
<td>7 (4)</td>
<td>7 (4)</td>
<td></td>
</tr>
</tbody>
</table>

Modified from Li and Graur (1991). Note that we assume a hypothetical model where all codons are used equally and that all types of point mutations are equally likely.

Why do we use $d_N$ and $d_S$?

same example, but using $d_N$ and $d_S$:

Synonymous sites = 25.5%

$S = 300 \times 3 \times 25.5\% = 229.5$

Nonsynonymous sites = 74.5%

$N = 300 \times 3 \times 74.5\% = 670.5$

So, $d_S = 5/229.5 = 0.0218$

$d_N = 5/670.5 = 0.0075$

$d_N/d_S (\omega) = 0.34$, purifying selection !!!
an index of selection pressure acting on the protein

**conserved sites:** $dN/dS < 1$

**fast sites:** $dN/dS > 1$

**conclusion:** $dN$ differs from $dS$ due to the effect of selection on the protein.

mutational opportunity vs. physical site

**Relative proportion of different types of mutations in hypothetical protein coding sequence.**

<table>
<thead>
<tr>
<th>Type</th>
<th>Expected number of changes (proportion)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All 3 Positions</td>
</tr>
<tr>
<td>Total mutations</td>
<td>549 (100)</td>
</tr>
<tr>
<td>Synonymous</td>
<td>134 (25)</td>
</tr>
<tr>
<td>Nonsynonymous</td>
<td>392 (71)</td>
</tr>
<tr>
<td>nonsense</td>
<td>23 (4)</td>
</tr>
</tbody>
</table>

*Note* that by framing the counting of sites in this way we are using a “mutational opportunity” definition of the sites. Thus, a synonymous or non-synonymous site is not considered a physical entity!

*Note* that we assume a hypothetical model where all codons are used equally and that all types of point mutations are equally likely.
real data have biases (Drosophila GstD1 gene)

**transitions** vs. **transversions:**

\[
\text{ts/tv} = 2.71
\]

**preferred** vs. **un-preferred** codons:

<table>
<thead>
<tr>
<th>Codon</th>
<th>Phe</th>
<th>Ser</th>
<th>Tyr</th>
<th>Cys</th>
<th>Leu</th>
<th>Pro</th>
<th>His</th>
<th>Arg</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTT</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td>27</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TTC</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td>15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TTA</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCA</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TGG</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
\omega = \frac{dN}{dS}
\]

Don’t worry: we will improve upon the counting method later in this lecture via likelihood!

correcting \(dS\) and \(dN\) for underlying mutational process of the DNA makes them sensitive to assumptions about the process of evolution!
Macroevolutionary time-scale

Mutation: $\mu_{ij}$
Drift: $N$
Selection: $s_i$

Population time-scale

$dN^h_i/dS^h_i$

Mechanistic models

Phenomenological models

Macroevolutionary time-scale
- Wright-Fisher population
- drift: \( N \)
- mutation: \( \mu \)
- selection: \( s_{ij} \)
- \( s_{ij} \) vary among sites AND amino acids
- expected \( dN/dS \)

\[
Pr = \begin{cases} 
\mu_{ij} & \text{if neutral} \\
\mu_{ij}N \times \frac{2s_{ij}}{1-e^{-2Ns_{ij}}} & \text{if selected}
\end{cases}
\]

\[s_{ij} = \Delta f_{ij}\]

Halpern and Bruno (1998)

---

**fixation probability with selection**

**population genetics at a single codon site (\( h \))**

**fitness coefficients**

\[f^h = \{ f_1, \ldots, f_{61} \} \]

**selection coefficients**

\[s_{ij}^h = f_j^h - f_i^h\]

**fixation probability (Kimura, 1962)**

\[
Pr(s_{ij}^h) = \frac{2s_{ij}^h}{1-e^{-2Ns_{ij}^h}}
\]
realism: fitness expected to differ among sites and amino acids according to protein function

the cost of realism: too complex to fit such a model to real data (but simplified versions will allow new ways of data analysis)
"OMEGA MODELS"

\[
q_{ij} = \begin{cases} 
0 & \text{if } i \text{ and } j \text{ differ by } > 1 \\
\pi_j & \text{for synonymous tv.} \\
\kappa \pi_j & \text{for synonymous ts.} \\
\omega \pi_j & \text{for non-synonymous tv.} \\
\omega \kappa \pi_j & \text{for non-synonymous ts.}
\end{cases}
\]

Goldman and Yang (1994)
Muse and Gaut (1994)

- phenomenological parameters
- ts/tv ratio: \( \kappa \)
- codon frequencies: \( \pi_j \)
- \( \omega = dN/dS \)
- parameter estimation via ML
- stationary process

the instantaneous rate matrix, \( Q \), is very big: \( 61 \times 61 \)

phenomenological codon models: just a few parameters are needed to cover the 3721 transitions between codons!

<table>
<thead>
<tr>
<th>From codon below:</th>
<th>TTT (Phe)</th>
<th>TTC (Phe)</th>
<th>TTA (Leu)</th>
<th>TGG (Leu)</th>
<th>CTT (Leu)</th>
<th>CTC (Leu)</th>
<th>---</th>
<th>GGG (Gly)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTT (Phe)</td>
<td>---</td>
<td>( \kappa \pi_{TTT} )</td>
<td>( \omega \pi_{TTT} )</td>
<td>( \omega \pi_{TTG} )</td>
<td>( \omega \pi_{TTT} )</td>
<td>0</td>
<td>---</td>
<td>0</td>
</tr>
<tr>
<td>TTC (Phe)</td>
<td>( \kappa \pi_{TTT} )</td>
<td>---</td>
<td>( \omega \pi_{TTT} )</td>
<td>( \omega \pi_{TTG} )</td>
<td>0</td>
<td>( \omega \kappa \pi_{TTC} )</td>
<td>---</td>
<td>0</td>
</tr>
<tr>
<td>TTA (Leu)</td>
<td>( \omega \pi_{TTT} )</td>
<td>( \omega \pi_{TTT} )</td>
<td>---</td>
<td>( \omega \pi_{TTG} )</td>
<td>0</td>
<td>0</td>
<td>---</td>
<td>0</td>
</tr>
<tr>
<td>TGG (Leu)</td>
<td>( \omega \pi_{TTT} )</td>
<td>( \omega \pi_{TTT} )</td>
<td>( \kappa \pi_{TTA} )</td>
<td>---</td>
<td>0</td>
<td>0</td>
<td>---</td>
<td>0</td>
</tr>
<tr>
<td>CTT (Leu)</td>
<td>( \omega \pi_{TTT} )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>---</td>
<td>( \kappa \pi_{CTC} )</td>
<td>---</td>
<td>0</td>
</tr>
<tr>
<td>CTC (Leu)</td>
<td>0</td>
<td>( \omega \pi_{TTT} )</td>
<td>0</td>
<td>0</td>
<td>( \kappa \pi_{TTT} )</td>
<td>---</td>
<td>---</td>
<td>0</td>
</tr>
<tr>
<td>GGG (Gly)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>

* This is equivalent to the codon model of Goldman and Yang (1994). Parameter \( \omega \) is the ratio \( d_N/d_S \), \( \kappa \) is the transition/transversion rate ratio, and \( \pi_j \) is the equilibrium frequency of the target codon \( j \).
intentional simplification: all amino acid substitutions have the same $\omega$!

1. ATI (Ile) $\rightarrow$ TTA (Leu): $\omega_{\text{ile-leu}}$ (conservative)

2. ATI (Ile) $\rightarrow$ AAA (Lys): $\omega_{\text{ile-lys}}$ (radical)

contradiction? selection should favour amino acids with higher fitness.

substitution probability with selection

probability of substitution between codons over time, $P(t)$

$$Q_i = \begin{cases} 
0 & \text{if } i \text{ and } j \text{ differ by } > 1 \\
\pi_j & \text{for synonymous tv.} \\
k\pi_j & \text{for synonymous ts.} \\
\epsilon\pi_j & \text{for non-synonymous tv.} \\
\epsilon\kappa\pi_j & \text{for non-synonymous ts.}
\end{cases}$$

$$P(t) = \{p_j(t)\} = e^{Qt}$$

recall that Paul Lewis introduced $Q$ matrices and how to obtain transition probabilities
likelihood of the data at a site

\[ L_h(CCC, CCT) = \sum_k \pi_k p_{CCC}(t_0) p_{CCT}(t_1) \]

The likelihood is a sum over all possible ancestral codon states that could have been observed at node \( k \).

Recall that Paul Lewis described how to compute the likelihood of the data at a site for a DNA model. The only difference here is that the states are codons rather than nucleotides.

Note: Analysis is typically done by using an unrooted tree.

likelihood of the data at all sites

The likelihood of observing the entire sequence alignment is the product of the probabilities at each site.

\[ L = L_1 \times L_2 \times L_3 \times \ldots \times L_N = \prod_{h=1}^{N} L_h \]

Paul Lewis covered this with the “AND” rule in his likelihood lecture.

The log likelihood is a sum over all sites.

\[ \ell = \ln\{L\} = \ln\{L_1\} + \ln\{L_2\} + \ln\{L_3\} + \ldots + \ln\{L_N\} = \sum_{h=1}^{N} \ln\{L_h\} \]

See Paul Lewis’s lecture slides for more about likelihoods vs. log-likelihoods.
we made some progress...

1. we are now being explicit about phenomenological and mechanistic models
2. we are more cautious about mechanistic interpretation of phenomenological parameters
3. we have learned how to connect evolutionary mechanisms to the substitution process
4. we introduced the idea that we can compute expectations from mechanistic parameters

Let’s look at some mechanism of evolution and “see” what we should expect!

part 2: mechanistic processes of codon evolution
MutSel time-scale is infinitesimal compared to substitution scale

MutSel probabilities approximate the instantaneous site-specific rate matrix, \( A \)

\( \mu_j = \text{nucleotide GTR process (before the effect of selection)} \)
two explicit ways to reconcile population genetics and macroevolution:

1. map fitness to equilibrium frequencies

2. macroevolution index of selection intensity

(1) Sella and Hinh 2005; (2) Jones et al. 2016

1. fitness coefficients map to stationary codon frequencies

\[ f^h = \{ f_{i_1}, \ldots, f_{i_61} \} \]

\[ \pi^h = \{ \pi_1, \ldots, \pi_{61} \} \]
2. from fitness coefficients to $dN/dS$

**MutSel rate matrix**

\[
dN^h / dS^h = \frac{E[\text{evolution w/ selection}]}{E[\text{evolution by drift alone}]}
\]

\[
dN^h / dS^h = \frac{\sum_{i \neq j} \pi_i^h A^h_{ij} I_N}{\sum_{i \neq j} \pi_i^h \mu_i I_N}
\]

- $dN/dS = \omega$ when matrix $A^h$ is replaced by matrix $Q$ of model M0
- $dN/dS$ is an analog of $\omega$ under MutSel

**positive selection: 3 evolutionary scenarios**

1. frequency dependent selection

2. episodic adaptation

3. shifting balance
scenario 1: frequency dependent selection

1. antagonistic evolutionary interaction

host-pathogen

sexual-conflict

molecular-interactions

frequency-dependent selection: MutSelM0

1. amino acid at a site has $f^a$; all others have $f^a + s$

2. fitness values swap when a substitution occurs

MutSelM0: (1) and (2) above imply Markov chain properties with the same rate matrix $Q$ as codon model M0
conclusion: phenomemological codon models assume frequency-dependent selection

[frequency-dependent selection: MutSelM0]

Generating process: Mutation-selection

Analytic Model: Pairs were fitted to M0, the standard model with one omega category, to get maximum likelihood estimates (MLEs).

Conclusion: Standard models assume frequency-dependent selection.

[ dos Reis (2015); Jones et al. (2016) ]

scenario 2: adaptive peak shift

exploitation of a new niche

darwinian adaptation

gene duplication

lateral gene transfer (LGT)
adaptive peak shift: evolution of novel function

optimal function in a stable environment

- **population**: at fitness peak
- **fitness peak**: stationary
- **FFTNS**: keeps population at peak

adapted peak shift: evolution of novel function

sub-optimal function in a novel environment

- **population**: lower fitness
- **fitness peak**: moving
- **FFTNS**: increase population mean fitness
  (non-stationary process)
adaptive peak shift: evolution of novel function

**episodic adaptive evolution** of a novel function

populations: returns to peak
fitness peak: stabilized
FFNFS: increases population mean fitness until at peak

---

BIOLOGY LETTERS

Research

Cite this article: dos Reis M. 2015 How to calculate the non-synonymous to synonymous rate ratio of protein-coding genes under the Fisher–Wright mutation–selection framework. BIOL. Lett. 11: 20141031. http://dx.doi.org/10.1098/rsbl.2014.1031

Received: 8 December 2014
Accepted: 10 March 2015

Molecular evolution

How to calculate the non-synonymous to synonymous rate ratio of protein-coding genes under the Fisher–Wright mutation—selection framework

Mario dos Reis
Department of Evolution, Genetics and Environment, University College London, Gower Street, London WC1E 6BT, UK

First principles of population genetics are used to obtain formulas relating the non-synonymous to synonymous substitution rate ratio to the selection coefficients acting at codon sites in protein-coding genes. Two theoretical cases are discussed and two examples from real data (mitochondrial gene and a virus polymerase) are given. The formulas give much insight into the dynamics of non-synonymous substitutions and may inform the development of methods to detect adaptive evolution.

4. The non-synonymous rate during adaptive evolution
**conclusion:** episodic models “work” because \( w > 1 \) is a consequence of a system moving towards a new fitness peak.

**conclusion:** episodic models “work” because they are sensitive to non-stationary behavior

---

**Scenario 3: non-adaptive evolution**

3. fitness coefficients are constant (fixed-peak)

Spielman and Wilke (2015)

- \( dN/dS \) must be \( \leq 1 \) when fitness coefficients are fixed.
- Positive selection is not possible on a stationary fitness peak

[Spielman and Wilke, (2015); Jones et al., (2016)]
shifting balance: movement around peak

mutation and drift can move a pop. off a fitness peak

shifting balance: the MutSel landscape (Jones at el. 2016)

fitness peak

most of the time

occasionally

never (if lethal)

dwelling time of the “SB” process
shifting balance: positive selection on a MutSel landscape

(1) amino acid at site varies over time
(2) selection acts to “repair” shifts to deleterious amino acids

**Expected proportion of mutations fixed by selection**

\[ p_i^h = \frac{\sum_{i(j)} \pi_i^h (A_{ij}^h - \mu_i) I_j}{\sum_{i(j)} \pi_i^h A_{ij}^h} \]

**Conclusion:** \( p_i > 0 \) as long as number of viable amino acids > 1 at a site

shifting balance: the MutSel landscape

\( dN^h/dS^h \) depends on the current amino acid

**Conclusion:** positive selection operates on a stationary fitness peak in the same way as when there is an adaptive peak shift
landscapes have unique structures

**conclusion:** A population can get to a sub-optimal codon (E) by drift and reside there for some time (b/c moving between T and E requires changes ≥ 2 codons).

landscape structure depends on N

**same site... 10x decrease in N** *(f*h have not changed!)*

**conclusion:** decreasing N changes:
  i. the “space” for shifting balance
  ii. mean dN/dS
  iii. equilibrium frequencies
shifting balance: the MutSel landscape

\[ \frac{dN}{dS} \] depends on the current amino acid

\[ \text{temporal average } \frac{dN}{dS} = 0.61 \]

shifting balance: a mechanistic model

\[ \omega^h < 1 = \frac{\sum_{\text{mut}_{h}} \frac{\pi_i^h}{p_i} A_{i}^{h} I_{N}}{\sum_{\text{mut}_{h}} \frac{\pi_i}{p_i} I_{N}} \]

\[ \omega^h > 1 = \frac{\sum_{\text{mut}_{h}} \frac{\pi_i}{p_i} A_{i}^{h} I_{N}}{\sum_{\text{mut}_{h}} \frac{\pi_i}{p_i} I_{N}} \]

\[ \text{Expected no. of switches per sub.} \]

\[ \delta^h = \frac{\sum_{t_{h=0}} \pi_i^h A_{i}^{h} I_{\text{switch}}}{\sum_{t_{h=0}} \mu_i^h I_{N}} \]

\[ \text{"SB" process} \]
**shifting balance: a mechanistic model**

<table>
<thead>
<tr>
<th>Shifting balance over landscape</th>
<th>high</th>
<th>moderate</th>
<th>low</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median switching rate ($\delta$)</td>
<td>0.45</td>
<td>0.25</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Expected $dN/dS$ in the “tail” of the landscape</th>
</tr>
</thead>
<tbody>
<tr>
<td>high</td>
</tr>
<tr>
<td>$\sim 1.1$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Expected $dN/dS$ near the “peak” of the landscape</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sim 0.95$</td>
</tr>
</tbody>
</table>

**Rate of evolution (i.e., “type of site”)**
- “fast”
- “informative”
- “conserved”

**gene sequences**

<table>
<thead>
<tr>
<th>Human</th>
<th>Cow</th>
<th>Rabbit</th>
<th>Rat</th>
<th>Opossum</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCTG</td>
<td>CTGC</td>
<td>TCTC</td>
<td>CCTG</td>
<td>GCCG</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>G</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>G</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>G</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>G</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>G</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>G</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>G</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>G</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>G</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>G</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>G</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>G</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>G</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>G</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>G</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>G</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>G</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>C</td>
</tr>
<tr>
<td>G</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>C</td>
</tr>
</tbody>
</table>

27
shifting balance: a mechanistic model

\[
\delta^h = \frac{\sum_{i,j} \pi_i^h I_{SW} p^h_{ij} t_{SW}}{\sum_{i,j} \pi_i^h I_{N} p^h_{ij} t_{N}}
\]

Expected no. of switches per sub.

\[
\omega^h < 1 = \frac{\sum_{i \neq j} \pi_i^h I_{N} p^h_{ij} t_{N}}{\sum_{i,j} \pi_i^h I_{N} p^h_{ij} t_{N}}
\]

\[
\omega^h > 1 = \frac{\sum_{i \neq j} \pi_i^h I_{N} p^h_{ij} t_{N}}{\sum_{i,j} \pi_i^h I_{N} p^h_{ij} t_{N}}
\]

covarion-like model of evolution

\[
Q = \begin{cases} 
\text{evolutionary regime 1:} \\
\quad \omega_1 = \text{low} \\
\quad \text{("near the peak"')} \\
\text{switching process:} \\
\quad \omega_2 \rightarrow \omega_1 \\
\end{cases}
\]

\[
\text{switching process:} \\
\quad \omega_1 \rightarrow \omega_2 \\
\text{evolutionary regime 2:} \\
\quad \omega_1 = \text{high} \\
\quad \text{("in the tail")}
\]

[ Guindon et al., (2004); Jones et al. (2016)]
covarion-like model of evolution (phenomenological)

2 selective regimes (low & high): sites CAN switch regime

![Diagram of site transitions]

\( \rho_1 \): proportion of time sites are in \( \omega_1 \)
\( \rho_2 \): proportion of time sites are in \( \omega_2 \)
switching: \( \delta \)

the covarion-like codon model can be fit to real data

shifting balance: a mechanistic model

<table>
<thead>
<tr>
<th>Landscape</th>
<th>shifting balance over landscape</th>
</tr>
</thead>
<tbody>
<tr>
<td>high</td>
<td>0.45</td>
</tr>
<tr>
<td>moderate</td>
<td>0.25</td>
</tr>
<tr>
<td>low</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

- median switching rate (\( \delta \))
- expected probability of a site being in the “tail” of the landscape (\( p_{wv} \))
- Expected \( dN/dS \) in the “tail” of the landscape
- Expected \( dN/dS \) near the “peak” of the landscape
- rate of evolution (i.e., “type of site”)

This “signal” is detectable with covarion and branch-site codon models!

<table>
<thead>
<tr>
<th>landscapes:</th>
</tr>
</thead>
<tbody>
<tr>
<td>250 ( \sigma )</td>
</tr>
<tr>
<td>(0.0001, 0.001, 0.01)</td>
</tr>
<tr>
<td>( N = 1000 )</td>
</tr>
</tbody>
</table>

recall: no adaptive evolution in this case (stationary fitness peak)!!!
summary

- standard codon models (single \( \omega \)) assume frequency dependent selection, which yields a persistent \( dN/dS > 1 \)

- episodic adaptive evolution leads to transient \( dN/dS > 1 \) (non-stationary process, with \( \omega \) upwardly biased)

- MutSel landscapes can be complex and a site can reside at a sub-optimal state for extended periods of time

- protein evolution on a static fitness landscape has temporal dynamics that include positive selection

- rate variation among sites reflects the interplay between mutation, drift, and selection (i.e., shifting balance dynamics)

REFERENCES


