Adaptive protein evolution: Detecting changes in selection

Belinda Chang

Department of Ecology & Evolutionary Biology
Department of Cell & Systems Biology
University of Toronto

@opsinlab

Phylogenetic approaches to the study of protein structure and function

• Ancestral protein reconstruction

• Computational analyses of selection (dN/dS)

-> Combining computational with experimental approaches allows us to test hypotheses of selection in protein evolution
Codon models of evolution - Ratio of nonsynonymous/synonymous substitution rates

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

A measure of evolutionary constraint which can provide insight into the form and strength of selection on proteins.

- purifying selection (AA changes are purged) \( \omega < 1 \)
- neutral evolution \( \omega = 1 \)
- positive selection (AA changes are favoured) \( \omega > 1 \)

Adaptive protein evolution

ADAPTIVE EVOLUTION:
Combining computational and experimental approaches to test hypotheses of selection in protein evolution

- Testing hypotheses of selection in a phylogenetic framework
- Identification of positively selected sites/regions
- Experimental studies of shifts in protein function
ADAPTIVE EVOLUTION:
Combining computational and experimental approaches to
test hypotheses of selection in protein evolution

- Testing hypotheses of selection in a phylogenetic framework
- Identification of positively selected sites/regions
- Experimental studies of shifts in protein function

1. Formulate hypothesis of selection acting on evolution of particular group of sequences
2. Collect target group of sequences
3. Test hypothesis of selection using codon models in a phylogenetic framework
4. Identify amino acid sites/regions that have been targets of selection
5. Formulate hypothesis of selection acting on particular aspect of protein function
6. Select proteins/mutants for experimental studies
7. Test hypothesis of selection using experiments investigating shifts in protein function
Testing hypotheses of selection in protein evolution: Combining computational and experimental approaches

1. Formulate hypothesis of selection acting on evolution of particular group of sequences
2. Collect target group of sequences
3. Test hypothesis of selection using codon models in a phylogenetic framework
4. Identify amino acid sites/regions that have been targets of selection
5. Formulate hypothesis of selection acting on particular aspect of protein function
6. Select proteins/mutants for experimental studies
7. Test hypothesis of selection using experiments investigating shifts in protein function

Testing hypotheses of adaptive evolution in protein function: Combining computational and experimental approaches

Computational analyses
- Statistical test for positive, or divergent selection in a phylogenetic context
- Identifies amino acid residues that may have been the targets of positive selection
- Does NOT prove adaptive evolution -> Need function

Experimental studies
- Empirical test for shifts in protein function as a consequence of substitutions at sites targeted by selection
- Does NOT prove anything about selection & adaptive evolution, need evolutionary context
- Ideally combined with ancestral reconstruction and mutagenesis studies
Testing hypotheses of adaptive evolution in protein function: Combining computational and experimental approaches

Computational analyses
- Statistical test for positive, or divergent selection in a phylogenetic context
- Identifies amino acid residues that may have been the targets of positive selection
- Does NOT prove adaptive evolution -> Need function

Experimental studies
- Empirical test for shifts in protein function as a consequence of substitutions at sites targeted by selection
- Ideally combined with ancestral reconstruction and mutagenesis studies

-> Taken together, computational and experimental approaches can be used to test for ADAPTIVE EVOLUTION in protein function as a consequence of natural selection!

Formulating specific and testable hypotheses of adaptive evolution

Pervasive vs. episodic selection
Sites targeted by selection
Effect on protein function

-> Answering the question, not only when/where but WHY did selection occur?

-> The role of a priori information in formulating hypotheses about function
Examples of positive selection on proteins: Experimental models

Host-viral arms races
- Primate intracellular immune response TRIM5a, protein kinase R (Sawyer et al. 2005, Elde et al. 2009)
- Potato virus Y (Moury & Simon 2011)
- Neutrophil-mediated host response to infection (Loughran et al. 2012)

Sperm-egg recognition proteins
- Abalone lysin, egg coat (Aagaard et al. 2013)

Sensory proteins
- Visual pigments in whales, fishes, birds, bats (Dungan & Chang 2017, Hauser et al. 2017)

Transcription factors in development
- Stem cell pluripotency (Baker et al. 2016)

Primate protein kinase R locked in arms race with viral protein mimic K3L

---

Elde et al. Nature 2009
Positive selection has shaped PKR in primate evolution

Elde et al. Nature 2009

Positively selected sites on distinct surfaces of the PKR kinase domain are crucial to K3L resistance

Elde et al. Nature 2009
Beyond random sites and branch-sites models: Clade models

The form and strength of selection may vary across both the phylogeny and the protein.

Codon models can incorporate variation across sites and lineages, and can be used to detect positive and divergent selection.

The codon model of divergent selection: Clade model C (CmC)

CmC assumes sites evolve in three ways:

1. **purifying selection** \(0 < \omega_0 < 1\) constrains evolution at some sites, operating consistently across the tree.

2. some sites evolve free of constraint (neutral: \(\omega_1 = 1\)) across the entire tree.

3. the remaining sites evolve **divergently** \(\omega_2 \neq \omega_3\) between particular sections of the tree.


Codon model tests of divergent selection

CmC assumes sites evolve in three ways:

1. **always purifying selection** \(0 < \omega_0 < 1\)

2. **always neutral** \(\omega_1 = 1\)

3. **divergent** \(\omega_2 \neq \omega_3\)

3 (alt). **no divergence** \(\omega_2 > 0\)

An improved null model for testing for divergent selection, M2a_rel model (Weadick & Chang 2012).

M2a_rel retains the third site class, but simply estimates a single \(\omega\) ratio that applies to all branches of the phylogeny.
Vision in aquatic environments

- Marine environments become increasingly blue with depth
- Deep-sea fishes have blue-shifted Rhodopsin
- Rivers are red-shifted and dimmer than marine

Freshwater fishes have red-shifted rhodopsin pigments

Molecular evolutionary models + experiments support functional divergence of rhodopsin in Neotropical cichlids

Frances Hauser
Hernan Lopez-Fernandez, ROM
Clade models used to test different hypotheses of Neotropical cichlid rhodopsin evolutionary divergence

<table>
<thead>
<tr>
<th>Model (foreground clade)</th>
<th>np</th>
<th>lnL</th>
<th>k</th>
<th>AIC</th>
<th>ΔAIC</th>
<th>Parameters</th>
<th>Null</th>
<th>LRT</th>
<th>df</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>M2a_rel</td>
<td>212</td>
<td>-5909.85</td>
<td>3.12</td>
<td>12239.70</td>
<td>17.46</td>
<td>p: 0.868 0.091 0.041</td>
<td>M2a</td>
<td>250.160</td>
<td>2</td>
<td>0.0000</td>
</tr>
<tr>
<td>C+H+G</td>
<td>211</td>
<td>-5904.67</td>
<td>3.06</td>
<td>12235.34</td>
<td>13.10</td>
<td>site proportion 0.866 0.097 0.034</td>
<td>M2a_rel</td>
<td>10.360</td>
<td>3</td>
<td>0.0157</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>background</td>
<td>0.000 1 3.979</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cichlasomatini</td>
<td>0.000 1 6.707</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Heroniusi</td>
<td>0.000 1 8.710</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Geophagini</td>
<td>0.000 1 1.100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central America (clade)</td>
<td>211</td>
<td>-5900.12</td>
<td>2.41</td>
<td>12222.24</td>
<td>7.34</td>
<td>proportion</td>
<td>0.866 0.097 0.034</td>
<td>M2a_rel</td>
<td>10.464</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>background</td>
<td>0.000 1 4.476</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>foreground</td>
<td>0.000 1 11.600</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>site</td>
<td>0 1 2</td>
<td>M2a_rel</td>
<td>19.464</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>proportion</td>
<td>0.866 0.091 0.040</td>
<td>M2a_rel</td>
<td>0.020</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>background</td>
<td>0.000 1 5.630</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>foreground</td>
<td>0.000 1 5.710</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>site</td>
<td>0 1 2</td>
<td>M2a_rel</td>
<td>7.280</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>proportion</td>
<td>0.866 0.095 0.037</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>background</td>
<td>0.000 1 4.630</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>foreground</td>
<td>0.000 1 8.020</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>site</td>
<td>0 1 2</td>
<td>M2a_rel</td>
<td>4.300</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>proportion</td>
<td>0.866 0.094 0.038</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>background</td>
<td>0.000 1 6.520</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>foreground</td>
<td>0.000 1 4.200</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>branch</td>
<td>0.000 1 4.200</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>M2a_rel</td>
<td>3.440</td>
<td>1</td>
<td>0.0119</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Central America</td>
<td>0.000 1 4.500</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>site</td>
<td>0 1 2</td>
<td>M2a_rel</td>
<td>1.670</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>proportion</td>
<td>0.866 0.092 0.040</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>background</td>
<td>0.000 1 5.410</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>foreground</td>
<td>0.000 1 10.250</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Adaptive protein evolution
Accelerated rhodopsin divergence during invasion of Central America

Rhodopsin site 83 identified as under positive selection and undergoes independent transitions in Central American cichlids
Spectroscopic assays of rhodopsin function

In vitro expression & purification

Light-activated retinal release

hv

Relative Fluorescence

Time (min)

0 30 60

Wavelength (nm)

Dark-light difference spectrum

Absorbance

400 500 450

Bickelmann et al. 2012
Morrow et al. 2015

Light-activated retinal release

Adaptive protein evolution

The N83D mutation produces a significant shift in rhodopsin function

N83D in Central American fishes mediates faster kinetics

Hauser et al. 2017 MBE

Adaptive protein evolution
Accelerated rhodopsin divergence during invasion of Central America

Positive selection in RH1
3.4% of sites

| Background (South America) | ω = 4.5 |
| Foreground (Central America) | ω = 14.0 |

Hauser et al. 2017 MBE

Combination of computational and functional studies suggest adaptation to increased light levels in clear streams of Central America

Rhodopsin evolution in marine-derived Amazonian anchovies

Hypothesis: Adaptive evolution in rhodopsin gene of freshwater anchovies

- Closely related marine and freshwater clades
- Ancestrally marine

Van Nynatten et al. 2015 Biol Lett
Hypothesis: Adaptive evolution in rhodopsin gene of freshwater anchovies

- Increased $dN/dS$ (CmC) in rhodopsin of the freshwater clade
- No difference in non-visual control genes

Amino acid substitutions in anchovy rhodopsin

- Longer branch lengths in freshwater clade when scaled by number of amino acid substitutions in rhodopsin
- More remodelled chromophore binding pocket in freshwater clade

$Larger$ $circles$ $at$ nodes $represent$ $more$ $amino$ $acid$ $differences$ $in$ retinal binding pocket residues with common ancestor (arrowhead)
Vision at high altitudes: Andean catfish rhodopsin evolution

Gianni Castiglione
Nathan Lujan
ROM, UTSC

Castiglione et al. 2017 PNAS

Positive divergent selection in rhodopsin of high altitude catfishes

Castiglione et al. 2017 PNAS
Positive divergent selection in rhodopsin of high altitude catfishes

Castiglione et al. 2017 PNAS

Adaptive protein evolution

17-07-21
High-altitude variants increase kinetic rates

Light-activated stability
- Castiglione et al. 2017 PNAS

Cold Adaptation: Activity-Stability trade-off
- Dark-state thermal decay

Parallel shifts in evolutionary rates at high altitudes
- Evolutionary Convergence in Andean and Himalayan High Altitude Specialist Rhodopsins
- Castiglione et al. In prep
Himalayan and Andean sites target the same structural motifs

Parallel shifts in evolutionary constraint are targeting *nonspectral* function

Adaptive protein evolution

Andean and Himalayan variants modulate the nonspectral kinetic rates of ligand-bound rhodopsin

Dark-state thermal decay

Evolutionary convergence predicts non-parallel functional convergence

Light-activated stability

Adaptive protein evolution
The future of studies in adaptive molecular evolution?

Combining computational and experimental approaches to study adaptive evolution within a hypothesis testing framework enables us to draw conclusions about adaptive molecular evolution that would not otherwise be possible with either method alone.