Evolutionary Relationships: A Mathematical Concept?

A look into the mathematics of distance-based phylogenetics.

Nazareth College
Sara Reynolds
Dr. Cheri Boyd
SUNY Geneseo Undergraduate Research in Biomathematics Program - April 22, 2006

Background Information

Phylogenetic Tree - a graphical representation of the evolutionary relationship among three or more genes or organisms

- Terminal Nodes (A, B, C, D, E)
- Internal Nodes (1, 2, 3, 4)
- Branches
Rooted and Unrooted Trees

Rooted

Unrooted

Finding Possible (Bifurcating) Trees

3 Possible Rooted Trees

1 Possible Unrooted Tree
Formulas for Calculating Possible Trees

Number of Rooted Trees: 
\[ N_R = \frac{(2n - 3)!}{2^{n-2}(n-2)!} \]

Number of Unrooted Trees: 
\[ N_U = \frac{(2n - 5)!}{2^{n-3}(n-3)!} \]

<table>
<thead>
<tr>
<th>(N_D)</th>
<th>(N_R)</th>
<th>(N_U)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>15</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>105</td>
<td>15</td>
</tr>
<tr>
<td>10</td>
<td>34,459,425</td>
<td>2,027,025</td>
</tr>
<tr>
<td>15</td>
<td>213,458,046,767,875</td>
<td>7,905,853,580,625</td>
</tr>
<tr>
<td>20</td>
<td>8,200,794,532,637,891,559,375</td>
<td>221,643,095,476,699,771,875</td>
</tr>
</tbody>
</table>

Distance Matrix Methods

Distance - a measure of the overall pairwise differences between two data sets such as DNA or protein sequences

UPGMA - Unweighted-Pair-Group Method with Arithmetic Mean

- Oldest distance matrix method (1960’s)
- Simplest of all methods for tree reconstruction
- Largely Statistically Based
**Example**

A: GTGCTGCACGG CTCAATTTG AGATCTTGGAA
B: GTGCTGCACGG CTCAATTTG AGATCTTGGAA
C: GTGCTGCACGG CTCAATTTG AGATCTTGGAA
D: GTGCTGCACGG CTCAATTTG AGATCTTGGAA
E: GTGCTGCACGG CTCAATTTG AGATCTTGGAA

<table>
<thead>
<tr>
<th>Species</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>9</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>C</td>
<td>8</td>
<td>11</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>D</td>
<td>12</td>
<td>15</td>
<td>10</td>
<td>--</td>
</tr>
<tr>
<td>E</td>
<td>15</td>
<td>18</td>
<td>13</td>
<td>5</td>
</tr>
</tbody>
</table>

**Example (continued)**

(10+13)/2 = 11.5
(15+18)/2 = 16.5
(10+13)/2 = 11.5
Example (continued)

<table>
<thead>
<tr>
<th>Species</th>
<th>B</th>
<th>AC</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC</td>
<td>(9+11)/2 = 10</td>
<td>--</td>
</tr>
<tr>
<td>DE</td>
<td>16.5</td>
<td>(13.5+11.5)/2 = 12.5</td>
</tr>
</tbody>
</table>

Limitations of UPGMA

1) Assumes that the evolutionary rates are the same for all lineages

2) Doesn’t take into account nucleic acid replacements that will yield the same amino acid, and other replacements with similar affects

3) Doesn’t take into account the differences in frequency of amino acid replacements
Alternative Methods

- Transformed Distance Method
  ~ Uses an outgroup to produce an additive more accurate distance tree

- Neighbor’s Relation Method
  ~ Emphasizes pairing species in such a way that a tree is created with the smallest possible branch lengths overall

- Neighbor-Joining Methods
  ~ Similar to the Neighbor's Relation Method with a few alternatives

- Maximum Likelihood Approaches
  ~ Statistically based method, which takes into consideration the probabilities for every individual nucleotide substitution in a set of sequence alignments

- Multiple Sequence Alignments
  ~ Simultaneously generates an alignment and a phylogeny

Special © Thanks © to:

- SUNY Geneseo
- Dr. Cheri Boyd (Nazareth College Mathematics Department)
- Dr. Matthew Temple (Nazareth College Biology Department)

Bibliography


...and the AlignX Computer Program