Orthologs versus Paralogs

- **Orthologs**: derived from a speciation event
- **Paralogs**: derived from a gene duplication event
  - Ancient: before speciation
  - Recent: after speciation

**OrthoMCL Flow Chart**

- For ESTs: cluster and find ORFs that are >= 75 a.a.
- BLASTP cut-off P-value = 1e-5
- **Orthologs**: reciprocal best between-species hits
- **Paralogs** (recent only): within-species hits that are better than between-species hits
OrthoMCL Sample Output

Similarity matrix

<table>
<thead>
<tr>
<th></th>
<th>A1</th>
<th>A2</th>
<th>B1</th>
<th>B2</th>
<th>B3</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>-</td>
<td>25</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>A2</td>
<td>-</td>
<td>-</td>
<td>19</td>
<td>16</td>
<td>-</td>
</tr>
<tr>
<td>B1</td>
<td>25</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>B2</td>
<td>-</td>
<td>19</td>
<td>-</td>
<td>40</td>
<td>-</td>
</tr>
<tr>
<td>B3</td>
<td>-</td>
<td>16</td>
<td>40</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Score = \(-\log_{10}(P\text{-value})\), cut-off = 5

Clustering

Group 1: A1 B1
Group 2: A2 B2 B3

Relationships of Taxa in OrthoMCL-DB

Bacteria

Archaea

Eukaryota

Plants & Algae

Apicomplexa

Unikonts

Fungi