Lecture 2. Phylogeny methods, part 2 (Searching tree space)

Joe Felsenstein

Department of Genome Sciences and Department of Biology
All possible trees

Forming all 4-species trees by adding the next species in all possible places
The number of rooted bifurcating trees:

\[ 1 \times 3 \times 5 \times 7 \times \ldots \times (2n - 3) \]

\[ = (2n - 3)! / ((n - 2)! \ 2^{n-2}) \]
which is:

<table>
<thead>
<tr>
<th>species</th>
<th>number of trees</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>15</td>
</tr>
<tr>
<td>5</td>
<td>105</td>
</tr>
<tr>
<td>6</td>
<td>945</td>
</tr>
<tr>
<td>7</td>
<td>10,395</td>
</tr>
<tr>
<td>8</td>
<td>135,135</td>
</tr>
<tr>
<td>9</td>
<td>2,027,025</td>
</tr>
<tr>
<td>10</td>
<td>34,459,425</td>
</tr>
<tr>
<td>11</td>
<td>654,729,075</td>
</tr>
<tr>
<td>12</td>
<td>13,749,310,575</td>
</tr>
<tr>
<td>13</td>
<td>316,234,143,225</td>
</tr>
<tr>
<td>14</td>
<td>7,905,853,580,625</td>
</tr>
<tr>
<td>15</td>
<td>213,458,046,676,875</td>
</tr>
<tr>
<td>16</td>
<td>6,190,283,353,629,375</td>
</tr>
<tr>
<td>17</td>
<td>191,898,783,962,510,625</td>
</tr>
<tr>
<td>18</td>
<td>6,332,659,870,762,850,625</td>
</tr>
<tr>
<td>19</td>
<td>221,643,095,476,699,771,875</td>
</tr>
<tr>
<td>20</td>
<td>8,200,794,532,637,891,559,375</td>
</tr>
<tr>
<td>30</td>
<td>$4.9518 \times 10^{38}$</td>
</tr>
<tr>
<td>40</td>
<td>$1.00985 \times 10^{57}$</td>
</tr>
<tr>
<td>50</td>
<td>$2.75292 \times 10^{76}$</td>
</tr>
</tbody>
</table>
Mapping an unrooted tree into a rooted tree

... one with one fewer species.
A global maximum is not easy to find

end up here

but global maximum is here

if start here
Nearest-neighbor interchanges (NNIs)

is rearranged by dissolving the connections to an interior branch:

and reforming them in one of the two possible alternative ways:

(The triangles are subtrees)
all 15 trees, connected by NNIs
with parsimony scores
Subtree pruning and regrafting (SPR) rearrangement

Break a branch, remove a subtree

Add it in, attaching it to one (*) of the other branches

Here is the result:

Lecture 2. Phylogeny methods, part 2 (Searching tree space) – p.10/23
Tree bisection and reconnection (TBR) rearrangement

Break a branch, separate the subtrees

Connect a branch of one to a branch of the other

Here is the result:
Greedy search by sequential addition

Greedy search by addition of species in a fixed order (A, B, C, D, E) in the best place each time.
Goloboff’s economy in computing scores of rearranged trees
Once the “views” have been computed, they can be taken to represent subtrees, without going inside those subtrees
“Star decomposition" search for best tree can happen in multiple ways
Disk-covering

“Disk covering” – assembly of a tree from overlapping estimated subtrees
Shortest Hamiltonian path problem
Search tree for this problem

(1,2,3,4,5,6,7,8,9,10)  
(1,2,3,4,5,6,7,9,8,10)  
(1,2,3,4,5,6,7,10,9,8)  
(1,2,3,4,5,6,7,8,9,10)  
(1,2,3,4,5,6,7,10,8,9)

Start

add 1  add 2  add 3  add 4  add 5  etc.
add 4  add 5
add 3  add 5
add 2  add 4
add 1  add 2  add 3

Lecture 2. Phylogeny methods, part 2 (Searching tree space) – p.17/23
Search tree of trees
same, with parsimony scores in place of trees
Some references


Graham, R. L. and L. R. Foulds. 1982. Unlikelihood that minimal phylogenies for a realistic biological study can be constructed in reasonable computational time. *Mathematical Biosciences* **60**: 133-142. [... and more]


continued

How it was done

This projection produced

- using the \texttt{prosper} style in LaTeX,
- using \texttt{Latex} to make a \texttt{.dvi} file,
- using \texttt{dvips} to turn this into a Postscript file,
- using \texttt{ps2pdf} to make it into a PDF file, and
- displaying the slides in Adobe Acrobat Reader.

Result: nice slides using freeware.