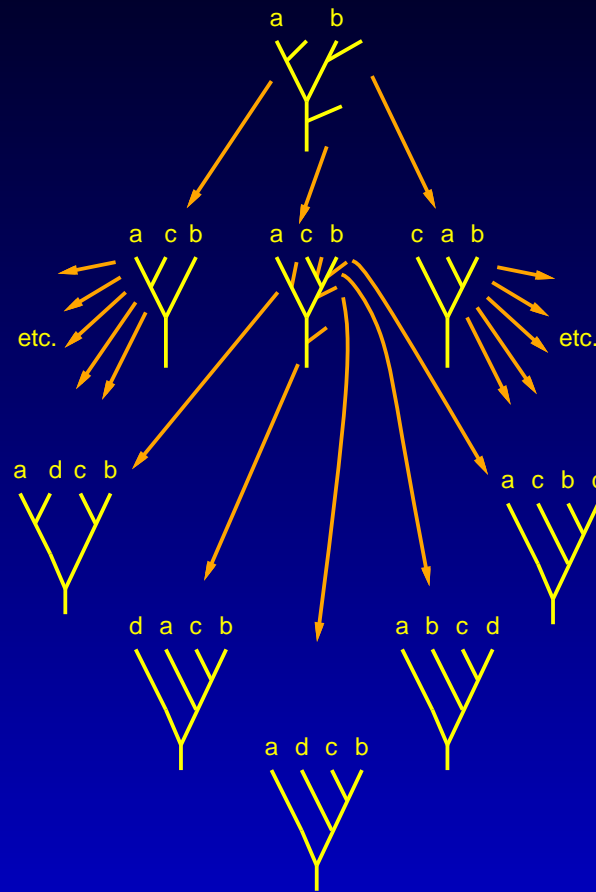# Lecture 30. Phylogeny methods, part 2 (Searching tree space)

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:

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

$$= (2n - 3)! / \left((n - 2)! \, 2^{n-2}\right)$$

which is:

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

Rooting an unrooted tree
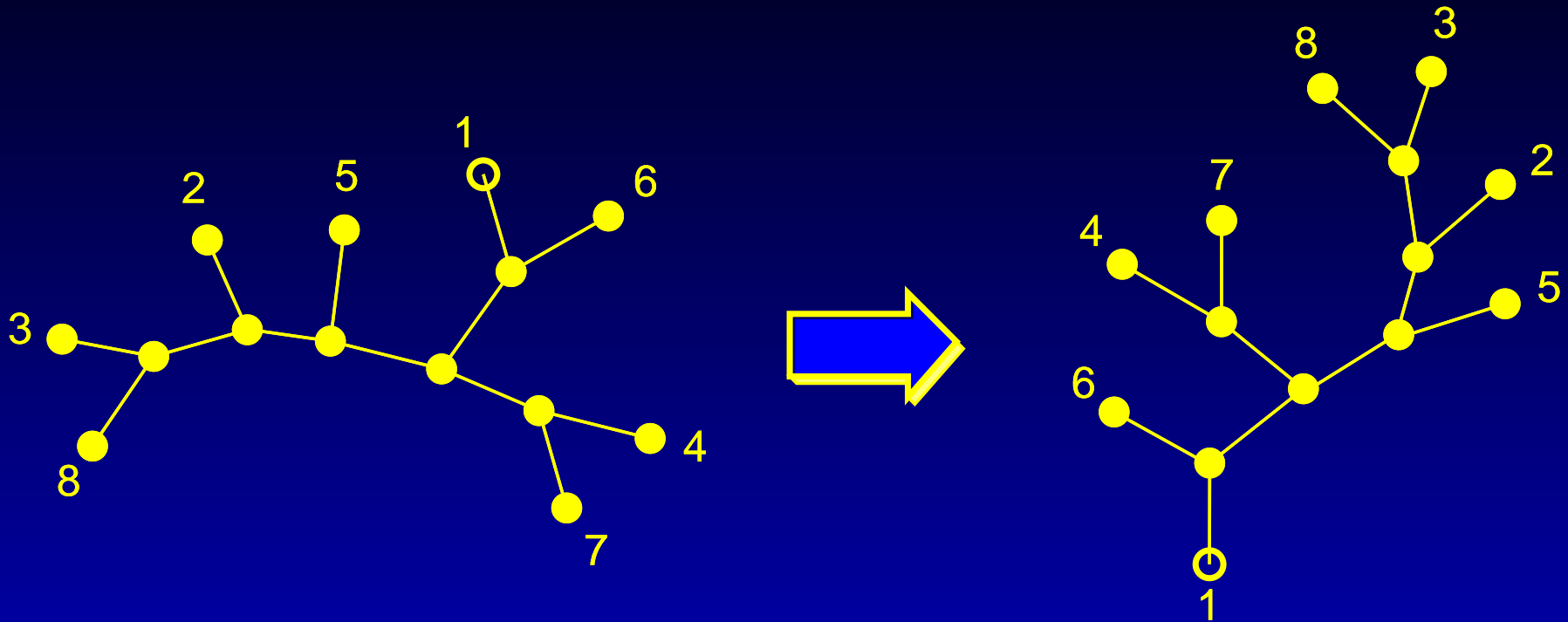
A global maximum is not easy to find

end up here

but global maximum is here

if start here
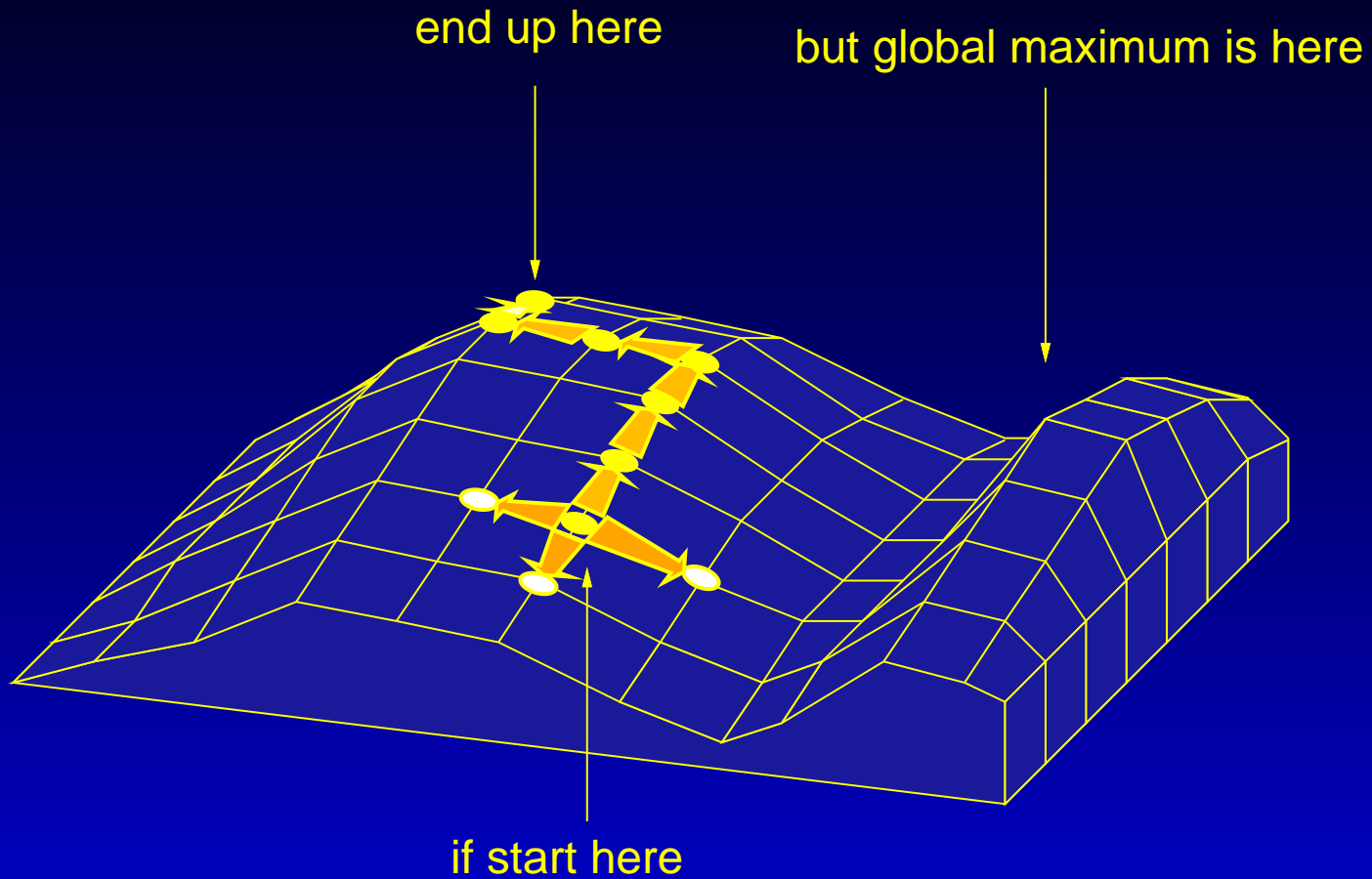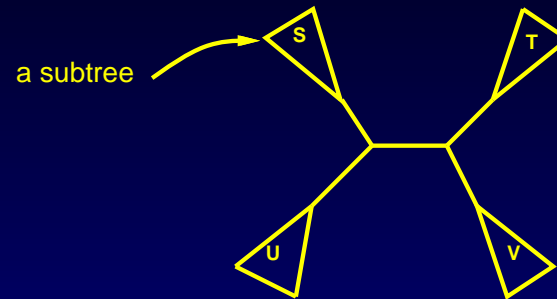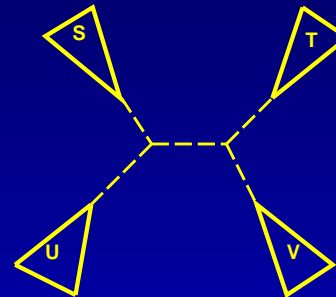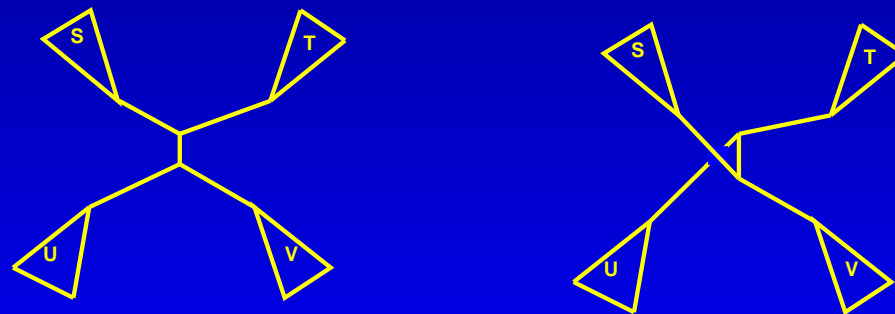
a subtree

is rearranged by dissolving the connections to an interior branch:
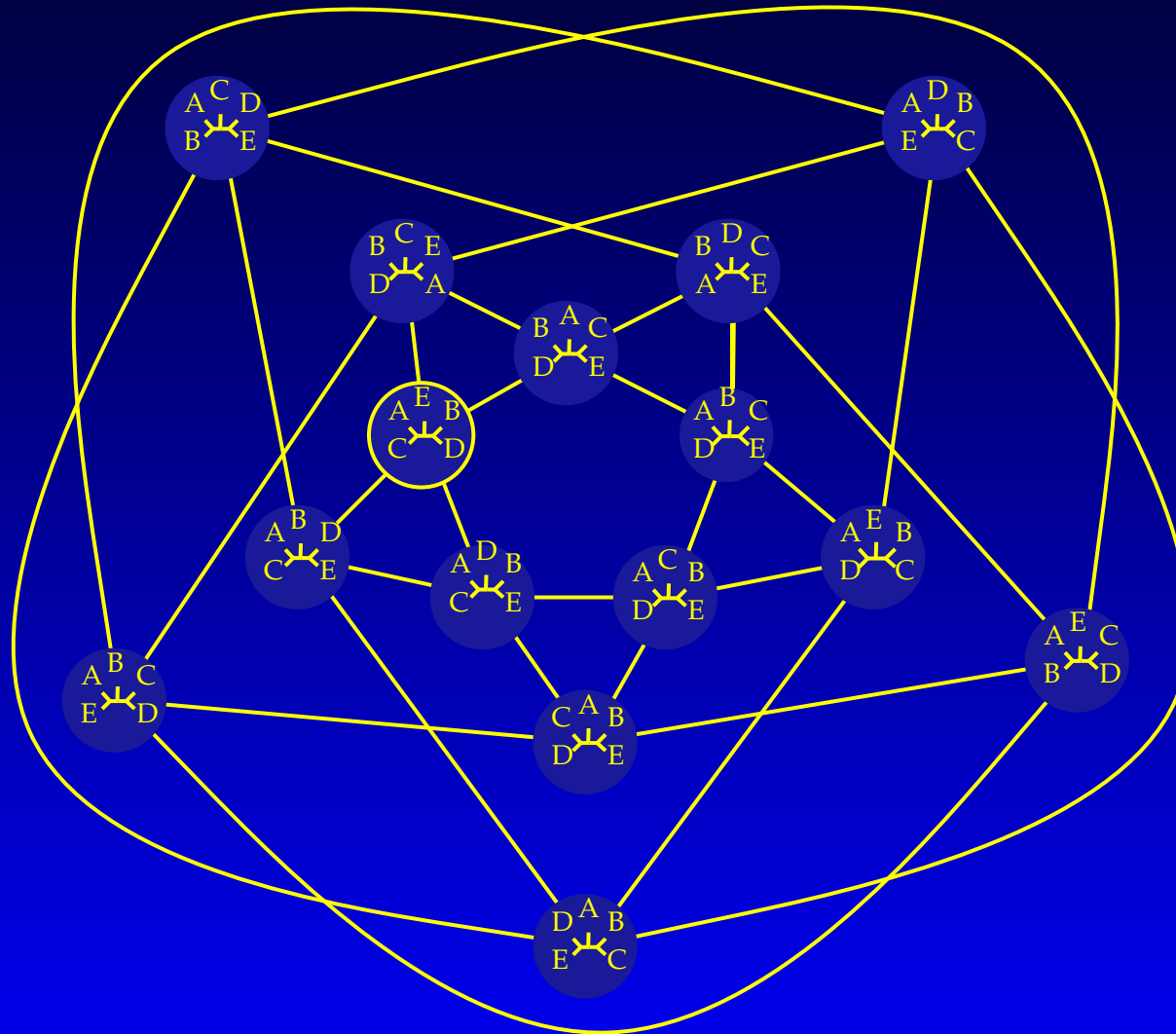
Is@

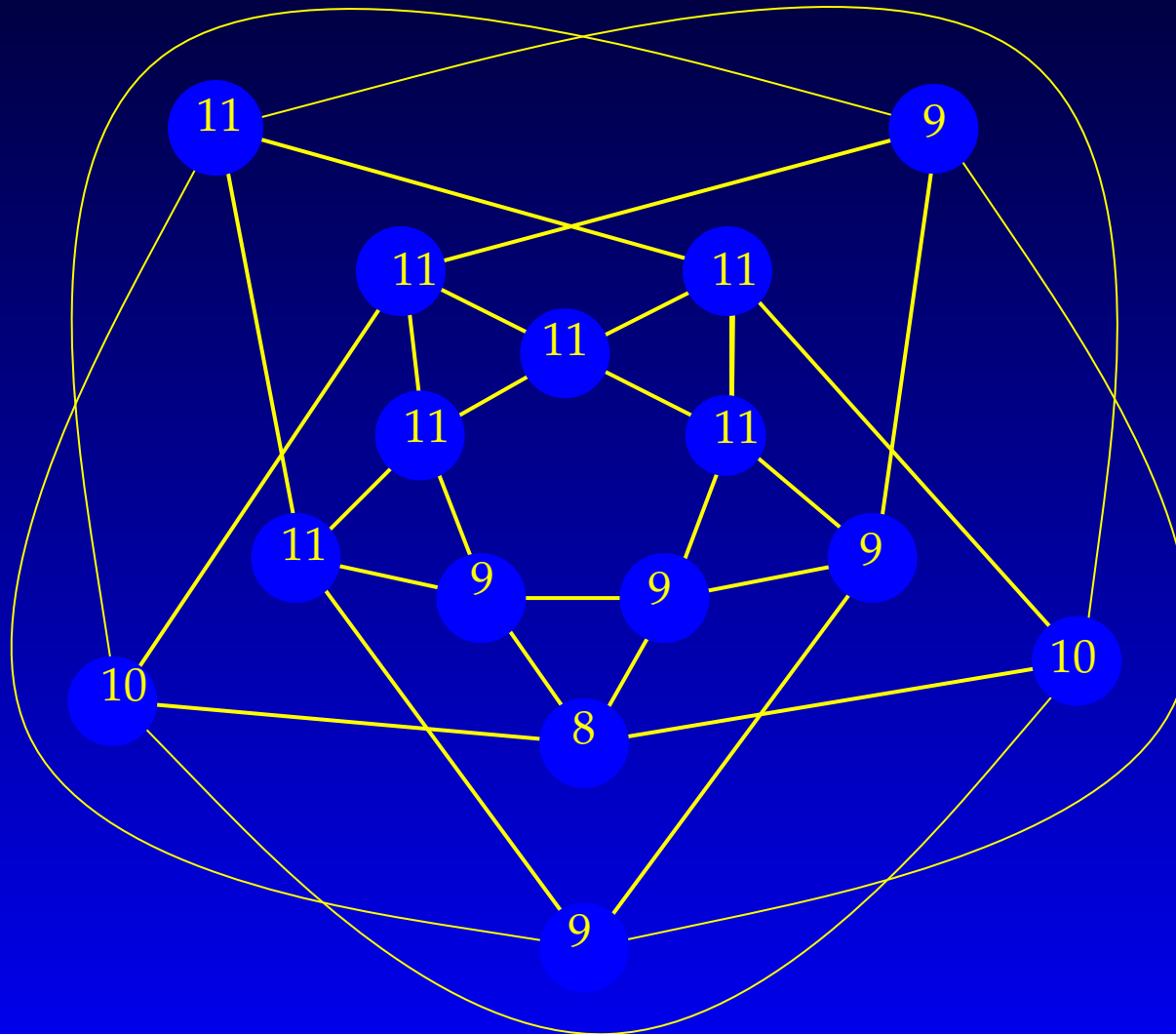and reforming them in one of the two possible alternative ways:

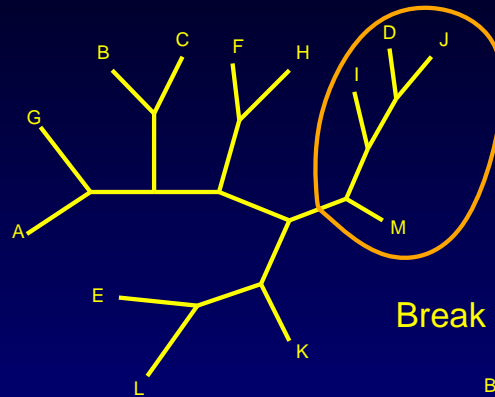Nearest-neighbor interchange (NNI) rearrangement of a tree (the triangles are subtrees)

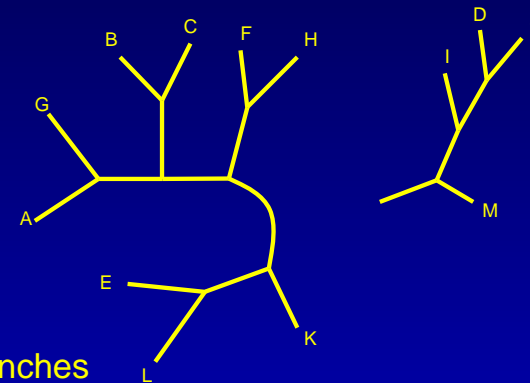# all 15 trees, connected by NNIs
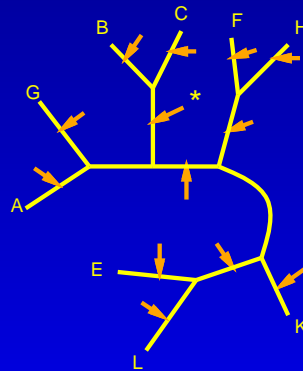
with parsimony scores

Subtree pruning and regrafting (SPR) rearrangement

Break each branch, remove a subtree
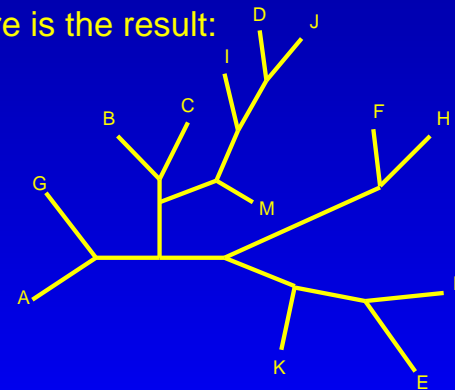
Add it in, attaching to one (*) of the other branches

Here is the result:

Tree bisection and reconnection (TBR) rearrangement
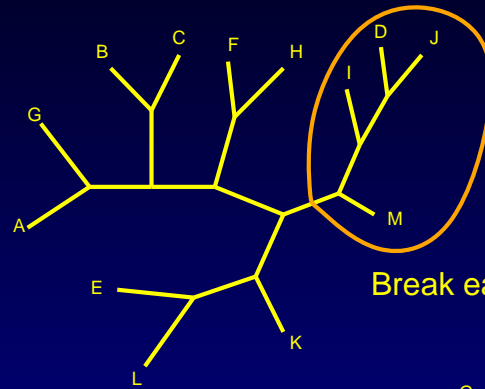
Break each branch, separate the subtrees
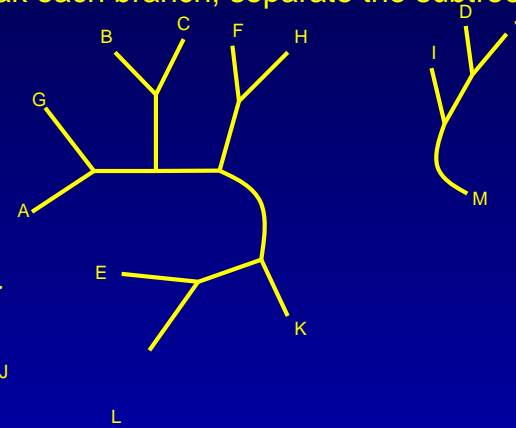
Connect a branch of one to a branch of another
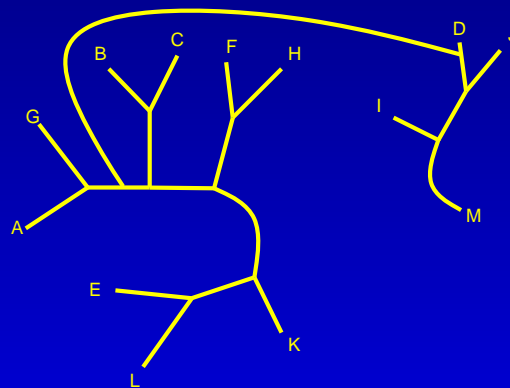
Here is that result:
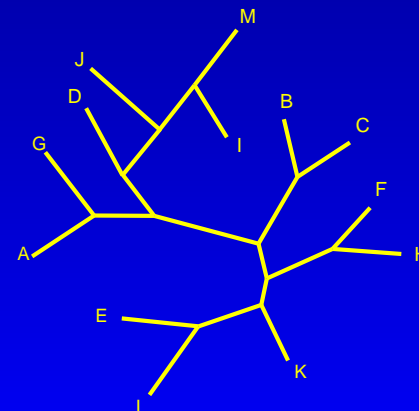
Greedy search by sequential addition
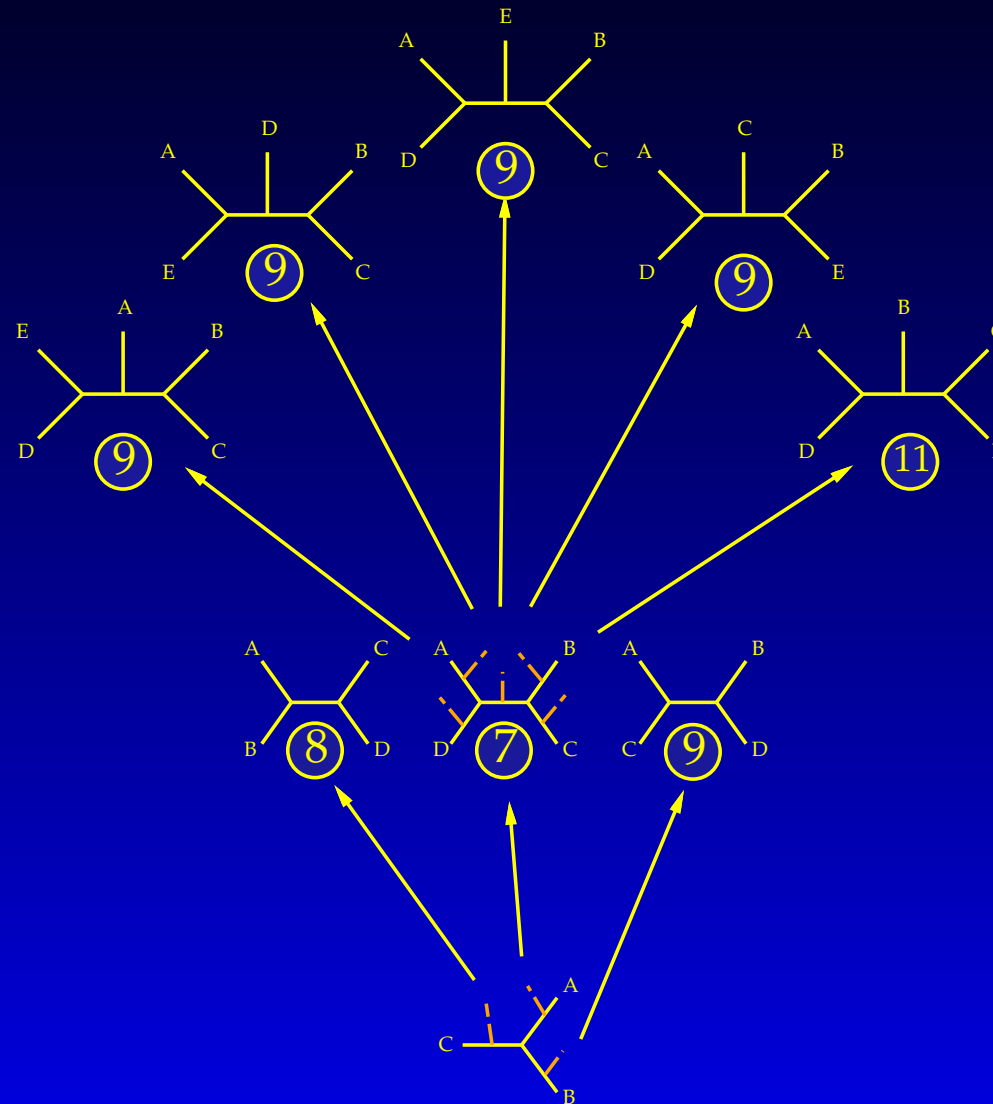
Greedy search by addition of species in a fixed order (A, B, C, D, E) in the best place each time.

Goloboff's time-saving trick
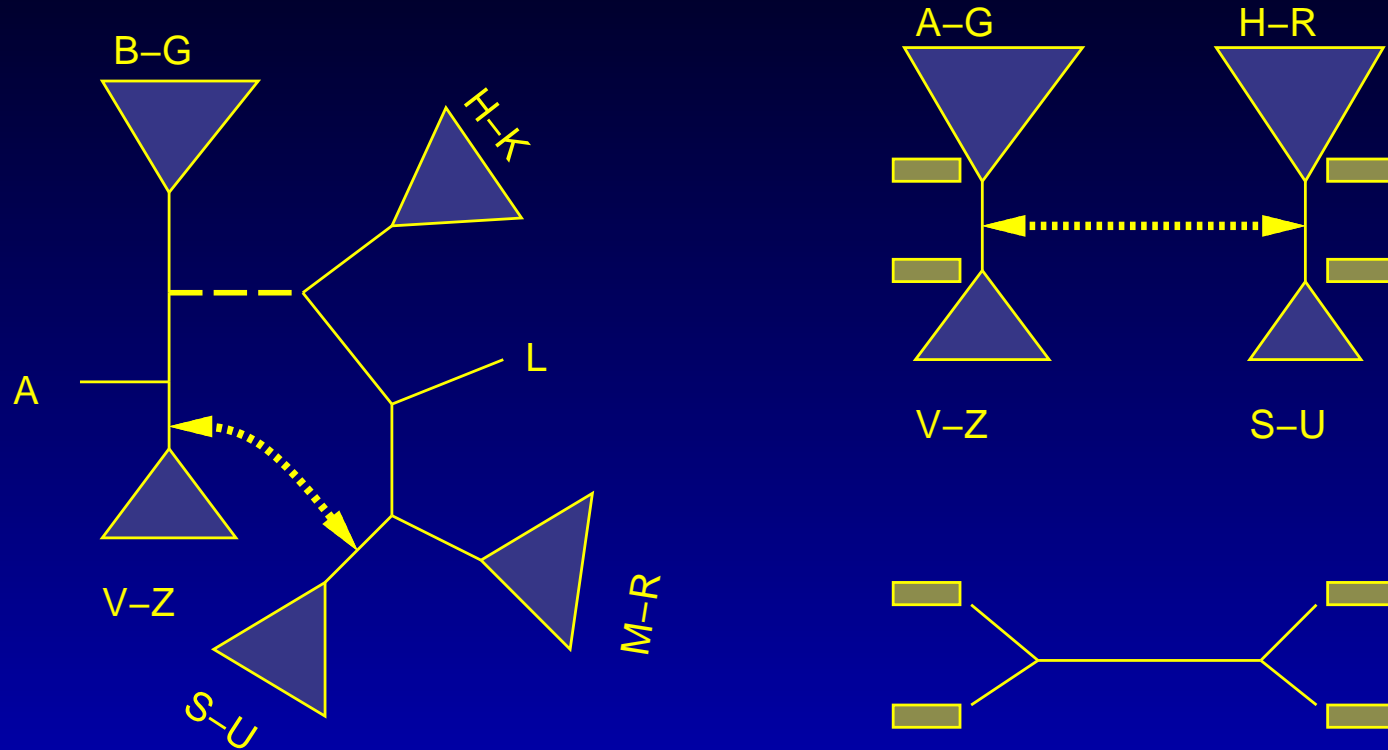
Goloboff's economy in computing scores of rearranged trees

Star decomposition
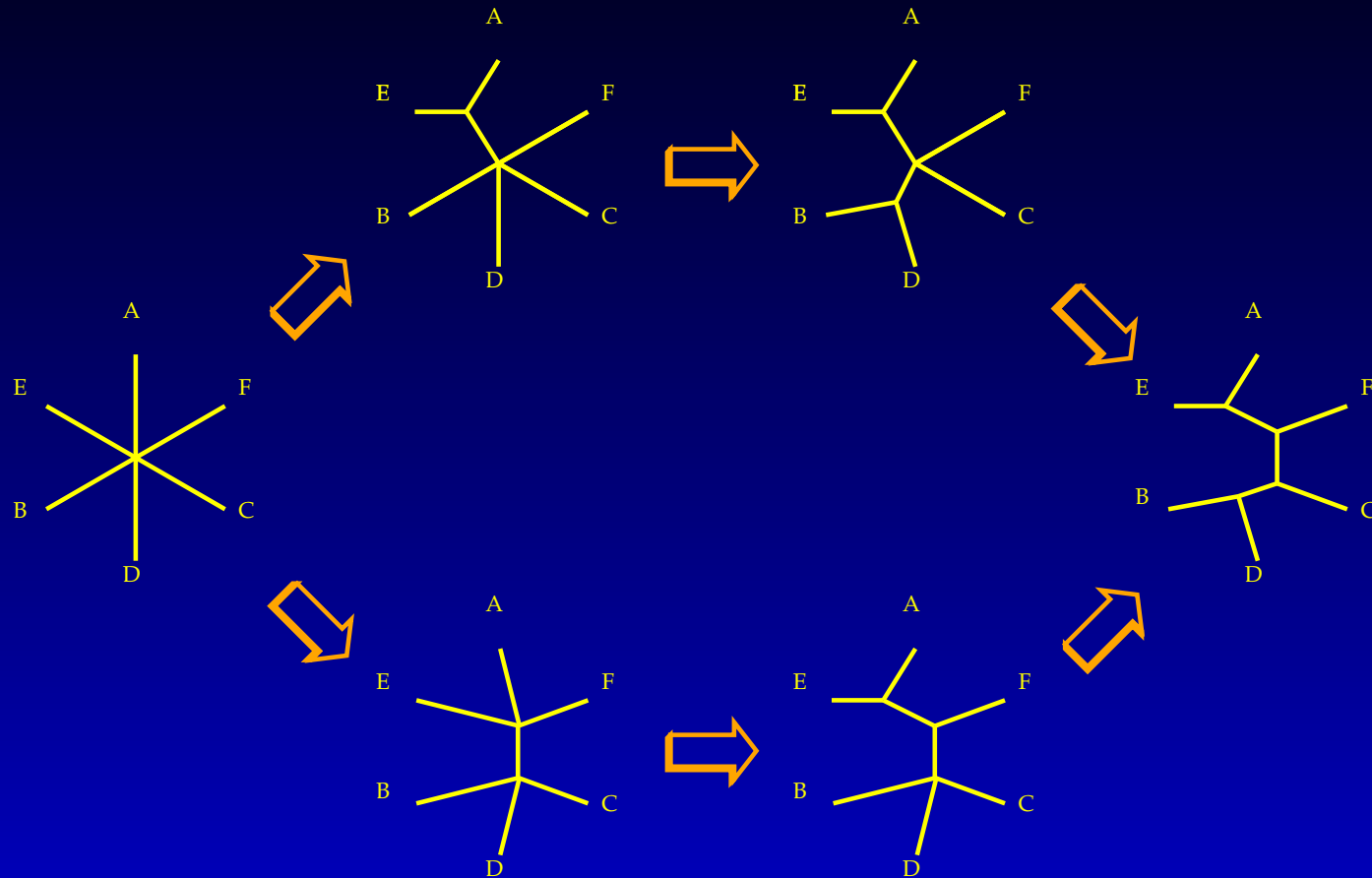
"Star decomposition" search for best tree can happen in multiple ways

Disk-covering

"Disk covering" – assembly of a tree from overlapping estimated subtrees

Shortest Hamiltonian path problem
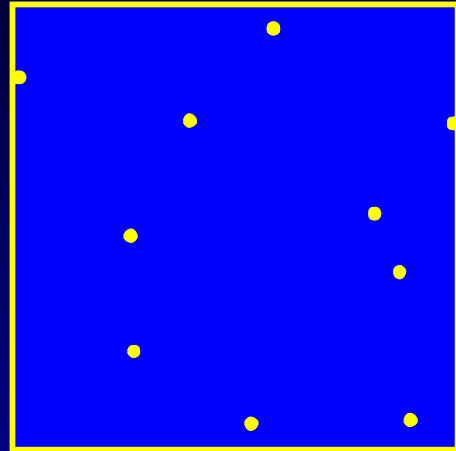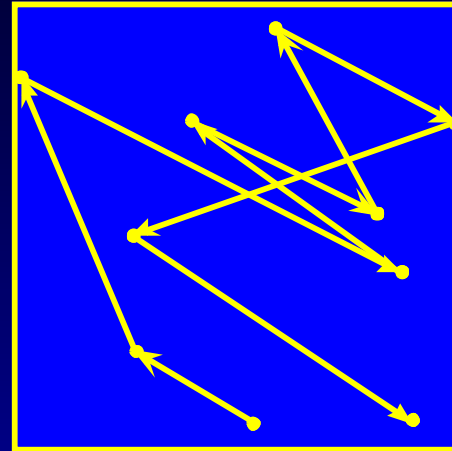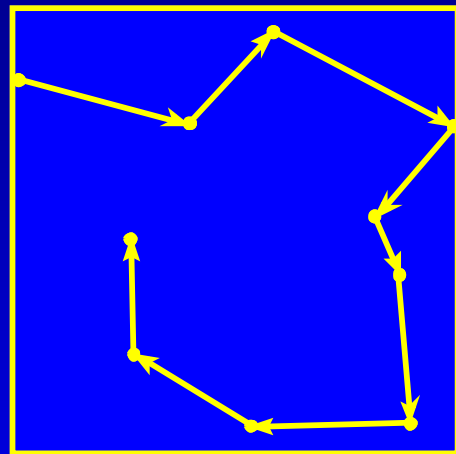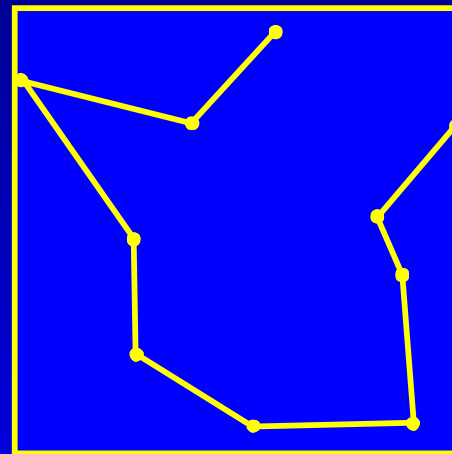
Search tree for this problem

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

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

| add 10 | add 9 | add 10 | add 8 | add 9 | add 8 |

| add 9 | add 10 | add 8 | add 10 | add 8 | add 9 |

| add 8 | | add 9 | add 10 |

... ...

etc.   etc.

| add 3 | add 4 | add 5 | ●●●   ... ...

etc.   etc.   etc.

| add 2 | add 3 | add 4 | add 5 | ●●●

etc.   etc.

| add 1 | add 2 | add 3 | ●●●

start

Search tree of trees

same, with parsimony scores in place of trees

Some references

Camin, J. H. and R. R. Sokal. 1965. A method for deducing branching sequences in phylogeny. *Evolution* **19:** 311-326. **[Early parsimony paper includes rearrangement of trees]**

Cavalli-Sforza, L. L. and A. W. F. Edwards. 1967. Phylogenetic analysis: models and estimation procedures. *American Journal of Human Genetics* **19:** 233-257. also *Evolution* **21:** 550-570. **[Includes counting and tree shapes]**

Farris, J. S. 1970. Methods for computing Wagner trees. *Systematic Zoology* **19:** 83-92. **[Early parsimony algorithms paper is one of first to mention sequential addition strategy]**

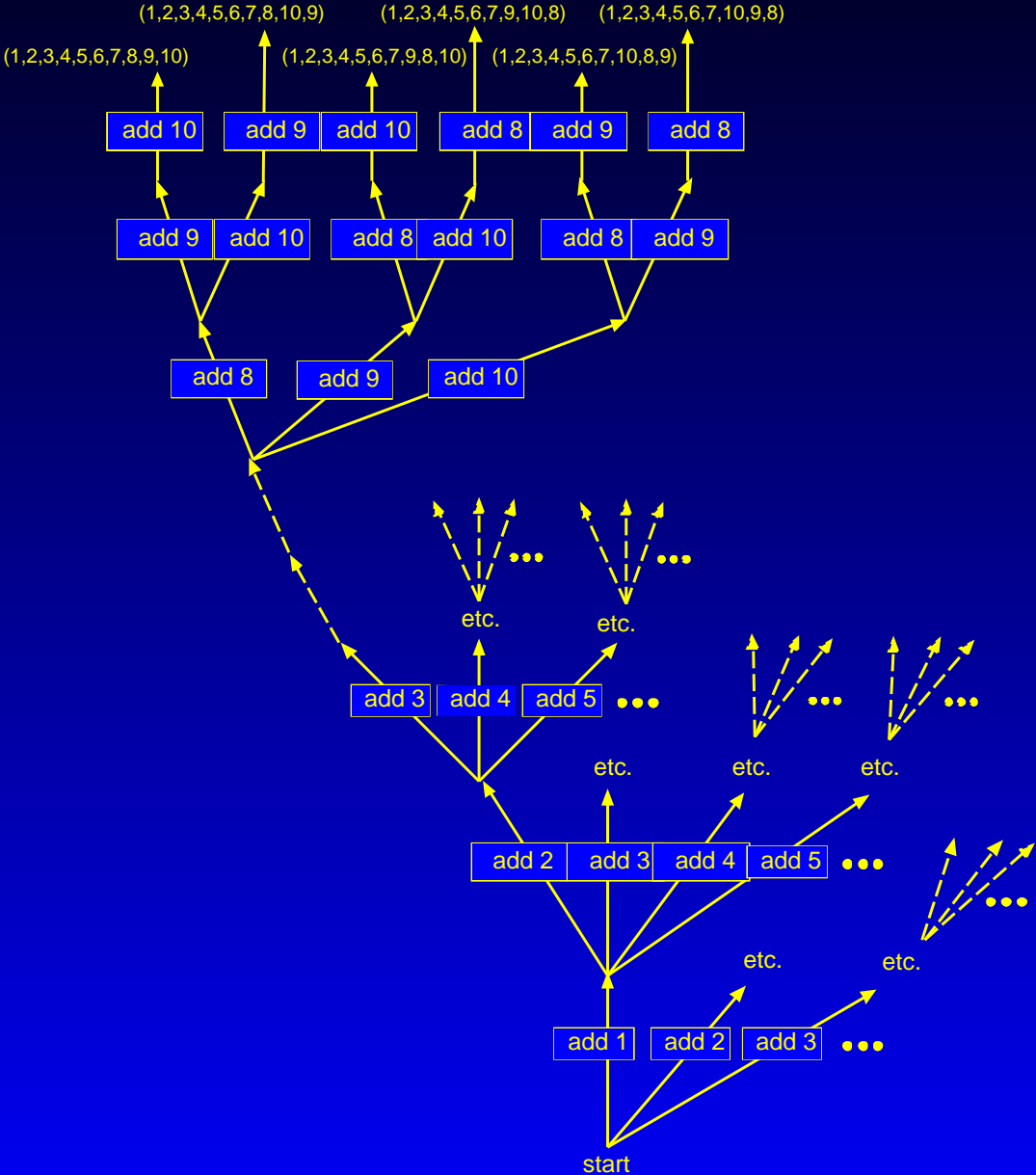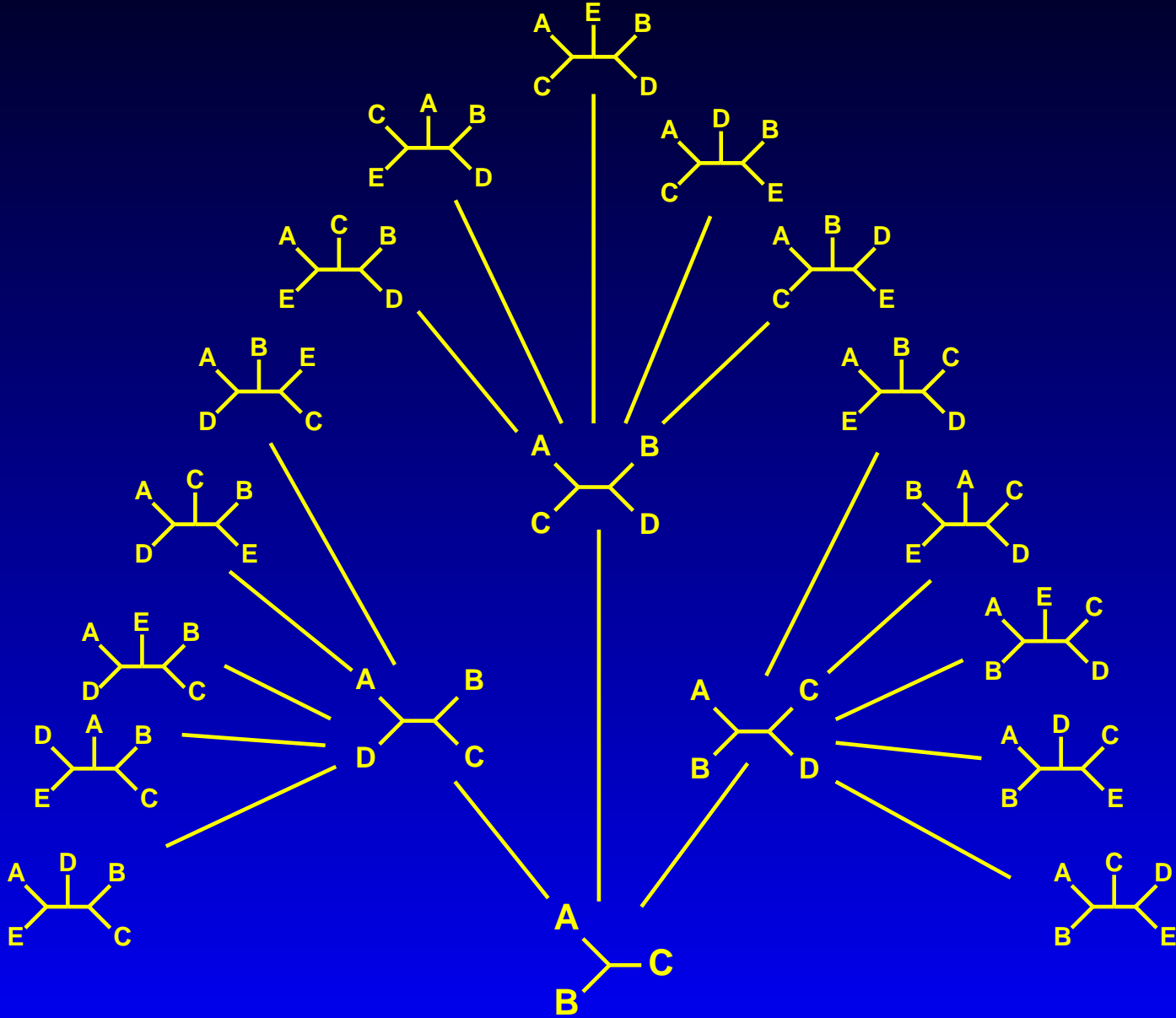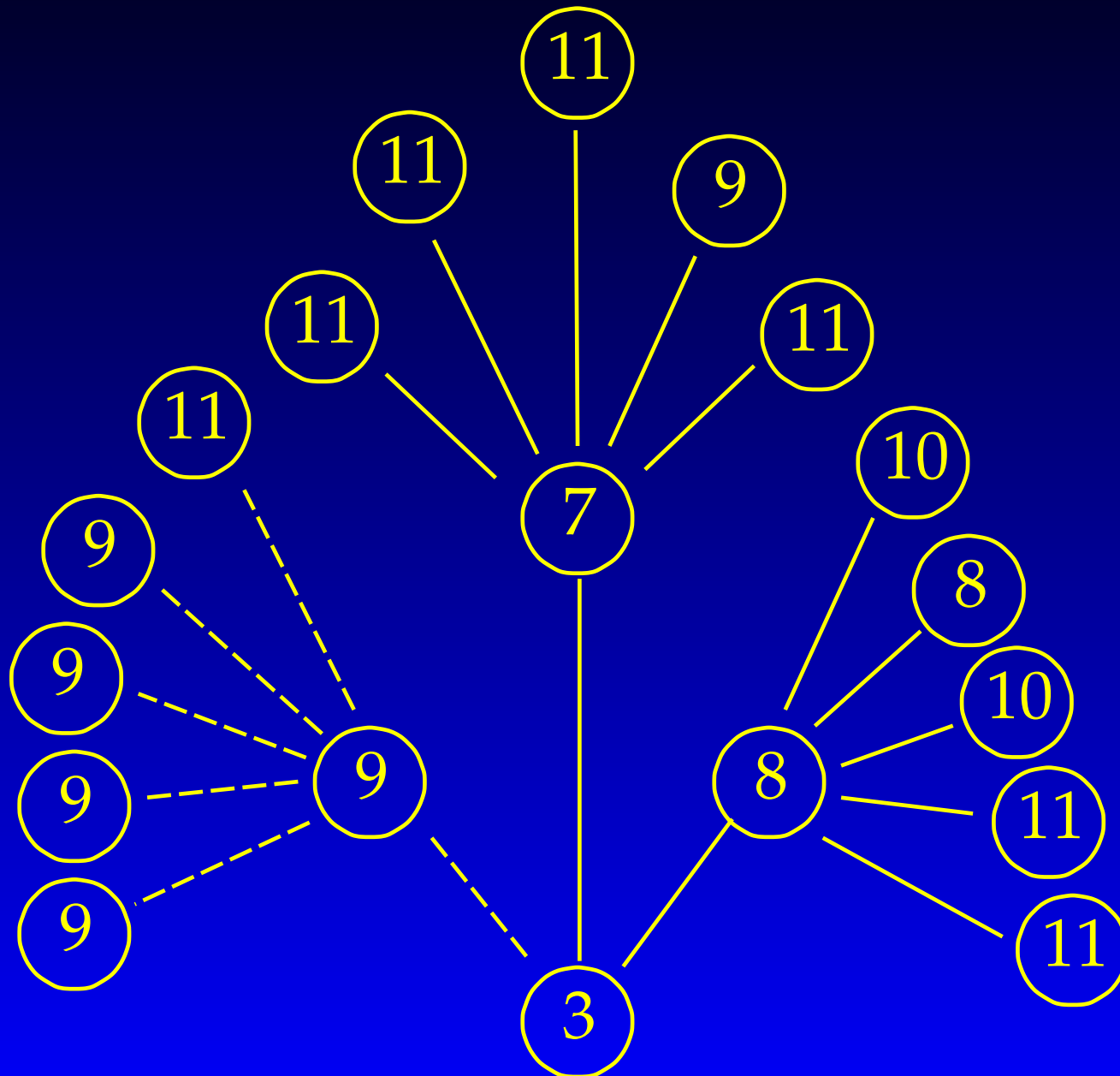Felsenstein, J. 1978. The number of evolutionary trees. *Systematic Zoology* **27:** 27-33. (Correction, vol. 30, p. 122, 1981) **[Review of counting tip-labelled trees, recursion for counting multifurcating case]**

Foulds, L. R. and R. L. Graham. 1982. The Steiner problem in phyloge ny is NP-complete. *Advances in Applied Mathematics* **3:** 43-49. **[Parsimony is NP-hard]**

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

Hendy, M. D. and D. Penny. 1982. Branch and bound algorithms to determine minimal evolutionary trees. *Mathematical Biosciences* **60:** 133-142 **[Introduced branch-and-bound for phylogenies]**

continued

Huson, D., S. Nettles, L. Parida, T. Warnow, and S. Yooseph. 1998. The disk-covering method for tree reconstruction. pp. 62-75 in *Proceedings of "Algorithms and Experiments" (ALEX98), Trento, Italy, Feb. 9-11, 1998*, ed. R. Battiti and A. A. Bertossi. **["Disk-covering method" for long stringy trees]**

Maddison, D. R. 1991. The discovery and importance of multiple islands of most-parsimonious trees. *Systematic Zoology* **40:** 315-328. **[Discusses heuristic search strategy involving ties, multiple starts]**

Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4:** 406-425. **[First mention of star-decomposition search for best trees, sort of]**

Strimmer, K., and A. von Haeseler. 1996. Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. *Molecular Biology and Evolution* **13:** 964-969. **[Assembles trees out of quartets]**

Swofford, D. L. and G. J. Olsen. 1990. Phylogeny reconstruction. Chapter 11, Pp. 411-501 in *Molecular Systematics,* ed. D. M. Hillis and C. Moritz. Sinauer Associates, Sunderland, Massachusetts. **[Review that discusses strategies, names SPR and TBR rearrangement methods]**

Waterman, M. S. and T. F. Smith. 1978. On the similarity of dendrograms. *Journal of Theoretical Biology* **73:** 789-800. **[Defines NNIs. Uses them to get a distance between trees.]**

How it was done

This projection produced

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

Result: nice slides using freeware.