

Introduction to Bioinformatics

Phylogenetic trees

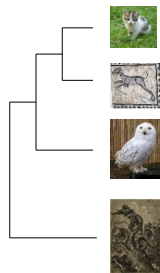
Inferring the Past: Phylogenetic Trees

- *The biological problem*
- Parsimony and distance methods
- Models for mutations and estimation of distances
- Maximum likelihood methods

344

Phylogeny

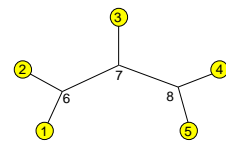
- We want to study ancestor-descendant relationships, or *phylogeny*, among groups of organisms
- Groups are called *taxa* (singular: *taxon*)
- Organisms are usually called *operational taxonomic units* or *OTUs* in the context of phylogeny



345

Phylogenetic trees

- Leaves (external nodes) ~ species, observed (OTUs)
- Internal nodes ~ ancestral species/divergence events, not observed
- Unrooted tree does not specify ancestor-descendant relationships beyond the observation "leaves are not ancestors"



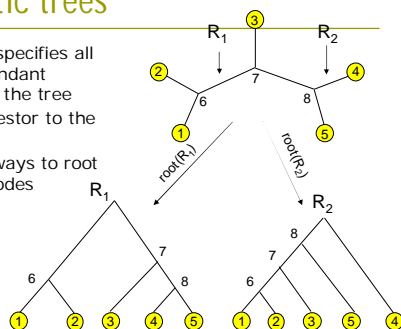
Unrooted tree with 5 leaves and 3 internal nodes.

Is node 7 ancestor of node 6?

346

Phylogenetic trees

- Rooting a tree specifies all ancestor-descendant relationships in the tree
- Root is the ancestor to the other species
- There are $n-1$ ways to root a tree with n nodes



347

Questions

- Can we enumerate all possible phylogenetic trees for n species (or sequences?)
- How to score a phylogenetic tree with respect to data?
- How to find the best phylogenetic tree given data?

348

Finding the best phylogenetic tree: naive method

- How can we find the phylogenetic tree that best represents the data?
- Naive method: enumerate all possible trees
- How many different trees are there of n species?
- Denote this number by b_n

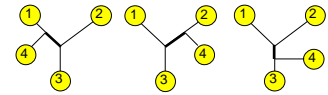
349

Enumerating unordered trees

- Start with the only unordered tree with 3 leaves ($b_3 = 1$)



- Consider all ways to add a leaf node to this tree



- Fourth node can be added to 3 different branches (edges), creating 1 new internal branch
- Total number of branches is n external and $n - 3$ internal branches
- Unrooted tree with n leaves has $2n - 3$ branches

350

Enumerating unordered trees

- Thus, we get the number of unordered trees

$$b_n = (2(n-1) - 3)b_{n-1} = (2n-5)b_{n-1}$$

$$= (2n-5) * (2n-7) * \dots * 3 * 1$$

$$= (2n-5)! / ((n-3)!2^{n-3}), n > 2$$
- Number of rooted trees b'_n is

$$b'_n = (2n-3)b_n = (2n-3)! / ((n-2)!2^{n-2}), n > 2$$
 that is, the number of unordered trees times the number of branches in the trees

351

Number of possible rooted and unrooted trees

n	B_n	b'_n
3	1	3
4	3	15
5	15	105
6	105	945
7	954	10395
8	10395	135135
9	135135	2027025
10	2027025	34459425
20	2.22E+020	8.20E+021
30	8.69E+036	4.95E+038

352

Too many trees?

- We can't construct and evaluate every phylogenetic tree even for a smallish number of species
- Better alternative is to
 - Devise a way to evaluate an individual tree against the data
 - Guide the search using the evaluation criteria to reduce the search space

353