

Phylogenetics: General Outline

- **Basic methods:**
 - Parsimony optimization
 - Maximum likelihood
 - Bayesian methods
- **Matrix structure:**
 - Parameters affecting character distributions
 - Compatibility:
 - General theory
 - Character correlation
 - Inverse modeling for relative rates
- **Stratigraphic data**
 - Tree-based methods for assessing sampling
 - Testing trees with stratigraphy
- **Tree-based tests**

Important Terms

- **Phylogeny (= tree):** ancestor-descendant relationships over time.
- **Cladogram:** graph depicting general relationships only (no temporal component or designated ancestors).
- **Clade:** descendants of a common ancestor.
- **Node:** inferred common ancestor between taxa (which might or might not match a sampled species); = **Hypothetical taxonomic unit (HTU)**.
- **Polytomy:** node giving rise to 3+ lineages (as opposed to **bifurcation**).
- **Outgroup:** taxon used to root tree & “polarize” states.
- **Sister-taxa** or **sister-groups:** taxa derived from a common ancestor (i.e., linked to the same node).

Important Terms (con't)

- **Synapomorphy:** shared derived states;
 - Ideally, **homologies** are synapomorphies, but homologies cannot be proven.
 - In contrast to **symplesiomorphy** (shared primitive state).
- **Autapomorphy:** character that is invariant save for one taxon.
- **Homoplasy:** “redundancy”.
 - Reversals: re-evolving a primitive condition;
 - Paralellisms: derived feature appearing 2+ times;
 - Like homologies, these cannot be proven.
- **Branch length:** either:
 - temporal duration of a branch;
 - number of changes along a branch.

Cladogram + Venn Diagram for Metazoans

How to “Write” Cladograms

- Nexus Format:
 - (Snails,(Fish,(Chimps,Us)));
 - (0,(1,(2,3)));
- If a new taxon add: (say, clams):
 - ((Snails,Clams),(Fish,(Chimps,Us)));
 - ((0,4),(1,(2,3)));
- Format used by PAUP, MacClade, etc.

How to “Write” Cladograms for Computer

- 0-4 give taxon #'s (e.g., snails, clams, fish, chimps, us);
- 5-7 are taxon #'s for nodes (i.e., molluscs, vertebrates, apes).

How to “Write” Cladograms for Computer

$m[x][\bullet]$ gives clade information for clade x ;

$m[x][0]$ gives # of taxa in clade;

$m[x][1]$ & $m[x][2]$ are taxa in clade x .

Polytomy: 3+ lineages attached to node

- Multiple possible interpretations
- Written as (A, (B,C,D)).

Multiple phylogenetic interpretations for Polytomy

- **Soft Polytomy**: reflects uncertainty.
- **Hard Polytomy A**: Ancestor and 2+ descendants sampled.
- **Hard Polytomy B**: Sudden radiation (e.g., species flocking).

Innumerable Phylogenies correspond to any one Cladogram

Both phylogenies have same cladistic topologies but different divergent times among sampled taxa.

Innumerable Phylogenies correspond to any one Cladogram

One phylogeny includes numerous sampled ancestors; other does not. Both fit the same cladistic topology.

Parsimony Optimization: Sankoff Vectors

Each cell gives the number of steps required if state 0 or state 1 ancestral condition at that node;

Lowest number at basal node gives the minimum steps.

Parsimony Optimization: Sankoff Vectors

Re-write cells to give steps need above and below the node;

∴ 2 steps now needed to have state 1 in basal node.

Parsimony Optimization: Multistate Characters

- **Ordered:** State X is X steps from state 0.
 - State 2 is 2 steps from 0, state 3 is 3 steps from state 0;
 - State 2 is 1 step from 1, state 3 is 2 steps from state 1.
- **Unordered:** All states are 1 step from each other.
- **Binary** is essentially a special case of either.

Parsimony Optimization: Sankoff Vectors & Unordered 3-State Character

Because all steps are equidistant, it is simply counting the needed changes.

Parsimony Optimization: Sankoff Vectors & Unordered 3-State Character

In this example, any of the three states can be the two most basal nodes.
Unimportant for *cladogram*, but important for *phylogeny*!

Parsimony Optimization: Sankoff Vectors & Ordered 3-State Character

More change now required for some ancestral reconstructions;

Parsimony Optimization: Sankoff Vectors & Ordered 3-State Character

More change now required for some ancestral reconstructions;

∴ 3 steps needed to make state 3 ancestral or to make state 0 the condition of the second node.

Parsimony Optimization: Sankoff Vectors & Ordered 3-State Character

After downwards pass, either 0 or 1 might be ancestral;
However, second node now needs to be state 1.

Step (= Cost) Matrices

Optimization & Inapplicable Characters

State 1 in (01) gives the presence of a complex structure (e.g. a feather) and states (01) give different conditions for that structure (e.g., feather color). “-” means not possible. Do not let the computer assume that there is a “primitive” feather color for the whole clade!

Optimization & Inapplicable Characters

Add an “inapplicable” “state” to the step matrices that is distance 0 from all other states.

From\To:	0	1	-
0	0	1	0
1	1	0	0
-	0	0	0

However, condition at node must be set to “-” if the independent character is absent.

Optimization & Inapplicable Characters

Independent character optimized as binary character: 1 in uppermost node and 0 at most basal node;

Inapplicable now impossible for uppermost node (optimally state 1) but necessary for most basal node.

Sankoff vectors for independent character now altered, too....

Optimization & Inapplicable Characters

Independent character now optimized as 0 at second most basal node;

Inapplicable now necessary for most that node.

Independent now needs to be 0 for the next two nodes.

Optimization & Inapplicable Characters

Dependent and independent now fully optimized.

NOTE: The dependent character actually makes 0 changes here; all of the change is by the independent character.

Finding the Parsimony Tree(s)

- Exhaustive: Examine all trees
 - $3 \times 5 \times 7 \times \dots (2n-3)$ rooted bifurcating trees for n taxa!
 - $3 \times 5 \times 7 \times \dots (2n-5)$ unrooted bifurcating trees for n taxa!
 - 316 billion rooted trees for 13 taxa alone.....
- Branch and Bound
 - Begin with nearest-neighbor reconstruction to get maximum estimate of parsimony length (the *bound*);
 - Start with three taxa, then add one (*branch*) and examine all topologies;
 - Repeat; however, once bound is surpassed, give up on these trees;
 - Limited by homoplasy: if there is a lot of it, then there will be too many trees shorter than the bound.

Finding the Parsimony Tree(s)

- Heuristic: trial and error search.
 - Nearest neighbor interchange: link taxa and then swap adjacent branches or whole branches;
 - Star decomposition: begin with n-taxon polytomy, and begin linking taxa.
 - Above algorithms are “greedy”: if a rearrangement does not work, then they do not revisit it.
 - Simulated annealing: accepts new tree if better, and sometimes if the new tree is worse;
 - initially more tolerant of worse trees;
 - Allows search to wander downhill and then uphill, possibly finding a higher peak.

Common Summaries of Parsimony Trees

- **Consistency Index (CI)** = m / s , where:
 - s = # of steps;
 - m = minimum possible # of steps;
 - = number of derived states unless inapplicable characters are involved;
 - often calculated without uninformative characters (i.e., invariant or autapomorphic characters).
- **Retention Index (RI)** = $(M-s)/(M-m)$, where:
 - M = maximum # of steps;
 - m & s as above.

Association between C.I. and Taxon Sampling

- Sanderson & Donoghue (1989): C.I. drops as taxon sampling increases for morphological and molecular data.

Association between C.I. and Taxon Sampling

- Association strongly pronounced when examining only fossil data.

Association between C.I. and Taxon Sampling

- Not a methodological artifact, but reflects limitations on recognizable variation.

Parsimony & Probability

Under what circumstances is the character vector [01100] more probable given tree A than given tree B?

I.e., under what circumstances is tree A more likely than tree B given [01100]?

Parsimony & Probability

- P[change] is the same on each branch;
 - Branch length unimportant;
 - No rate shifts on tree;
 - Other characters do not affect probability of change;
 - P[gain] = P[loss].
- Only a single ancestral reconstruction is considered per node.

Parsimony & Probability

Tree A requires only one change.

Parsimony & Probability

The probability of the character vector is:

$$P[\text{change}]^{\text{changes}} \times (1-P[\text{change}])^{\text{static branches}}$$

Log-likelihood of tree is:

$$\text{changes} \times \ln(P[\text{change}]) + \text{statics} \times \ln(1-P[\text{change}])$$

Parsimony & Probability

If P[change] = 0.1, then:

$$P[\text{character} \mid \text{tree}] = 0.1^1 \times 0.9^9 = \mathbf{3.87 \times 10^{-2}}$$

$$\ln L[\text{tree} \mid \text{character}] = \ln(0.1) + (9 \times \ln[0.9]) = \mathbf{-3.25}$$

Parsimony & Probability

If P[change] = 0.1, then:

$$P[\text{character} \mid \text{tree}] = 0.1^2 \times 0.9^8 = \mathbf{4.30 \times 10^{-3}}$$

$$\ln L[\text{tree} \mid \text{character}] = (2 \times \ln[0.1]) + (8 \times \ln[0.9]) = \mathbf{-5.45}$$

Parsimony & Probability

If P[change] = 0.01, then:

$$P[\text{character} \mid \text{tree}] = 0.01^1 \times 0.99^9 = \mathbf{9.14 \times 10^{-3}}$$

$$\ln L[\text{tree} \mid \text{character}] = \ln(0.01) + (9 \times \ln[0.99]) = \mathbf{-4.70}$$

Parsimony & Probability

If P[change] = 0.01, then:

$$P[\text{character} \mid \text{tree}] = 0.01^2 \times 0.99^8 = \mathbf{9.23 \times 10^{-5}}$$

$$\ln L[\text{tree} \mid \text{character}] = (2 \times \ln[0.01]) + (8 \times \ln[0.99]) = \mathbf{-9.29}$$

Parsimony & Probability

If $P[\text{change}] = 0.001$, then:

$$P[\text{character} \mid \text{tree}] = 0.001^1 \times 0.999^9 = \mathbf{9.91 \times 10^{-4}}$$

$$\ln L[\text{tree} \mid \text{character}] = \ln(10^{-3}) + (9 \times \ln[0.999]) = \mathbf{-6.91}$$

Parsimony & Probability

If $P[\text{change}] = 0.001$, then:

$$P[\text{character} \mid \text{tree}] = 0.001^2 \times 0.999^8 = \mathbf{9.92 \times 10^{-7}}$$

$$\ln L[\text{tree} \mid \text{character}] = (2 \times \ln[10^{-3}]) + (8 \times \ln[0.999]) = \mathbf{-13.82}$$

Infinity and beyond.....

$P[\text{change}]$	$\ln L[\text{tree A}]$	$\ln L[\text{tree B}]$	Difference
10^{-1}	-3.25	-5.45	2.20
10^{-2}	-4.70	-9.29	4.60
10^{-3}	-6.92	-13.82	6.91
$10^{-\infty}$	$-\infty$	$-2 \times \infty$	∞

Shorter tree is more likely while $P[\text{change}] < 0.5$

$P[\text{change}]$	$\ln L[\text{tree A}]$	$\ln L[\text{tree B}]$	Difference
0.2	-3.62	-5.00	1.62
0.4	-5.51	-5.92	0.41
0.5	-6.93	-6.93	0.00
0.6	-8.76	-8.35	-0.41

Relaxing assumptions of parsimony

- Low vs. high rates of change.
- Homogeneous vs. heterogeneous rates.
- Unit vs. variable branch lengths.
- Certain vs. uncertainty in ancestral reconstructions.
- Correlated character change.

Phylogeny Likelihood

- Calculate the exact probability of character matrix given a particular phylogeny.
 - Branch length affects expectations;
 - Relative rates affect expectations.
- $L[\tau, \pi | m] = \prod P[c_i | \tau, \pi_i]$
 - τ : phylogeny;
 - π : rate;
 - c : distribution of character i
 - Product over all n characters.
- Cladograms do not have likelihoods *unless* you sum over numerous phylogenies!

“Weights” and likelihood

Doubling a character’s weight invokes two step matrices:

From\To:	0	1	From\To:	0	1
0	0	1	0	0	2
1	1	0	1	2	0

This assumes that $P[\text{change char. B}] = P[\text{change char. A}]^2$, *not* $P[\text{change char. B}] = 2 \times P[\text{change char. A}]$.

From\To:	0	1	From\To:	0	1
0	$1-p_a$	p_a	0	$1-(p_a)^2$	$(p_a)^2$
1	p_a	$1-p_a$	1	$(p_a)^2$	$1-(p_a)^2$

Thus, weights reflect exponents of “base” rate.

“Ordered states” and likelihood

Doubling a character’s weight invokes two step matrices:

From\To:	0	1	2
0	0	1	2
1	1	0	1
2	2	1	0

Instead of implying that 1 must evolve between 0 and 2, it now implies that $P[0 \leftrightarrow 1] = P[0 \leftrightarrow 2]^2$.

From\To:	0	1	2
0	$1-(p+p^2)$	p	p^2
1	p	$1-2p$	p
2	p^2	p	$1-(p+p^2)$

Note: Each row must sum to 1.0.

“Unordered states” and likelihood

Doubling a character's weight invokes two step matrices:

From\To:	0	1	2
0	0	1	1
1	1	0	1
2	1	1	0

The probability of changing to any one state is simply one divided by the number of options (e.g., 2 if 3 states):

From\To:	0	1	2
0	1-p	p/2	p/2
1	p/2	1-p	p/2
2	p/2	p/2	1-p

Shorter tree is more likely while $P[\text{change}] < 0.5$

$P[\text{change}]$	In L[tree A]	In L[tree B]	Difference
0.5	-6.93	-6.93	0.00

Shift does not occur at $P[\text{change}] > 0.15$ because only a single way of generating one or two changes is considered.