

Tree Shape

- Different phylogenies can have very different shapes.
- Shape typically in terms of *symmetry*: sister taxa with comparable diversity stem from a symmetric node;
 - Number of symmetric nodes;
 - Quantification of how symmetrical nodes are.

Clade Shape Metrics: Fould-Robinson Balance (B_{FR})

- $B_{FR} = (\sum B_n) / \text{nodes}$
 - B_n : balanced node
 - Node is balanced if both sister taxa are sister clades;
 - If a polytomy, the node is fractionally balanced:
 - $B_n = ([\sum \text{clades}] - 1) / \text{daughters}$.
-
- Given a perfectly symmetrical tree, $B_{FR} = 1.0$.
-
- Given a perfectly *pectinate* topology, $B_{FR} = 0.00$.

Clade Shape Metrics: Colless' Imbalance (I_C)

- $I_C = (2 \times \sum |T_R - T_L|) / (\text{nodes} \times [\text{nodes} - 1])$
 - T_R : number of sampled taxa (OTUs) within “right” descendant of node;
 - T_L : number of OTUs within “left” descendant of node;
 - \therefore numerator sums differences in diversity between sister taxa.
- If every sister taxon is a clade with the same diversity, then $I_C = 0.0$.
- Metric not initially designed for polytomies, and must be modified;
- $I_C = (2 \times \text{average}[|T_i - T_j|]) / (\text{nodes} \times [\text{nodes} - 1])$
 - T_i : one of the descendants of the node;
 - If there are three descendant clades with 2 taxa each, then the average is 0 and the node is considered symmetrical.
- Given a perfectly symmetrical topology, $I_C = 0.00$.
- Given a perfectly pectinate topology, $I_C = 1.00$.
- Given a perfectly symmetrical topology, $I_C = 0.00$.

Clade Shape Metrics: Intermediate shapes

- Each node yielding a species is asymmetrical.
- Nodes need not be completely balanced for B_{FR} to treat the tree as symmetrical.
- I_C is sensitive to this imbalance.
- Each node yielding a species is asymmetrical.
- Asymmetries higher in the tree deflate I_C less than do asymmetries lower in the tree.

Clade Shape Metrics: Relationship between metrics

- Relationship much tighter given bifurcating cladogenesis than given budding cladogenesis.
- Based on 500 simulations sampling 32 taxa over time with $\mu = 0.5$, $\lambda = 0.45$ & $R = 0.5$.

Clade Shape Metrics: $P[B_{FR} \mid \lambda=0.45, \mu=0.5, R=0.5, 32 \text{ taxa}]$

- Expected B_{FR} somewhat higher given bifurcating cladogenesis than given budding cladogenesis.
- Indicates that bifurcation increases symmetry.

Clade Shape Metrics: $P[I_C \mid \lambda=0.45, \mu=0.5, R=0.5, 32 \text{ taxa}]$

- Expected I_C somewhat lower given bifurcating cladogenesis than given budding cladogenesis.
- This again indicates that bifurcation increases symmetry.

Importance of Tree Shape for Macroevolutionary Studies

- Hypotheses about shifts in diversification rates predict asymmetrical phylogenies;
- Tests using only extant taxa limited to testing net rates of cladogenesis;
 - Sampled diversity already conditioned upon having survived;
 - At $\mu=0$, $P[\text{clade extant}] = 1$; however, if $\mu>0$, then $P[\text{clade extant}] < 1.0$
 - This makes $\mu=0$ more likely than $\mu\geq 0$.
 - If fossil taxa included, then their extinctions must be taken into account.

Identifying Rate Shifts

- Standard rates of diversification apply;
- Tests using only extant taxa limited to testing net rates of cladogenesis;
 - Sampled diversity already conditioned upon having survived;
 - At $\mu=0$, $P[\text{clade extant}] = 1$; however, if $\mu>0$, then $P[\text{clade extant}] < 1.0$
 - This makes $\mu=0$ more likely than $\mu\geq 0$.
 - If fossil taxa included, then their extinctions must be taken into account.

Identifying Rate Shifts

- Standard diversification equations apply for predicting the probability of X taxa over Y million years give origination (λ) and extinction (μ).
- Phylogenetic context used to identify particular points from which to test rates.
- Problem: identifying exact point of rates very difficult because a diverse clade will make its larger clade overly diverse.

Identifying Rate Shifts

- Blue clade is much more diverse than the yellow clade...

Identifying Rate Shifts

- ... which makes the green clade significantly more diverse than the orange clade.

Using Origination & Extinction Rates can put prior probabilities of phylogenetic topologies

- No need for flat priors given that certain topologies are very improbable.
- One can modify prior probabilities to allow for differential diversification.