SSB 2015 Model-Based Biogeography
The Dispersal-Extinction-Cladogenesis Model

Jeet Sukumaran

Ecology and Evolutionary Biology
University of Michigan
Ann Arbor, MI

jeetsukumaran@gmail.com

May 20, 2015

The DEC Model

The DEC Model Family

- Probabilistically models the processes of (D)ispersal, (E)xtinction, and (C)ladogenesis in a likelihood framework.
- Introduced by Ree et al. (2005) and Ree and Smith (2008); extended by Matzke (2014).
- Data:
    - An ultrametric phylogeny
    - Distribution of tips (areas occupied by each tip lineage).
    - List of possible ranges.
    - Area connectivity, representing relative weighting of dispersal between areas.
- Produces:
    - Likelihood of the model given the data
    - Estimates of the global dispersal and extinction rates
    - Likelihoods of the ranges of the internal nodes.

Conceptual Components

Lineage	A taxon, represented on a phylogeny by a node (or an edge, depending on whether think of your tree as a graph or not).
Area	A fundamental geographic unit of analyses. A region is typically divided into a set of mutually-exclusive and distinct areas.
Range	A distinct combination or set of areas. The distribution of a lineage is given by its range.
Dispersal	Acquistion of a new area in the range of a lineage.
Extinction	Loss of an area from the range of a lineage.
Cladogenesis	Splitting of a lineage into two daughter lineages.

Anagenetic vs. Cladogenetic Range Evolution

- Geographical *range evolution* is the primary aspect modeled under the DEC approach.
- Range evolution: changes in areas included in the range of lineages over time.
- The DEC model distinguishes between *anagenetic* and *cladogenetic* range evolution.
- Anagenetic range evolution occurs *between* nodes (i.e., along an edge, in between speciation events).
- Cladogenetic range evolution occurs *at* a node (technically, an infinitesimal amount of time right *after* the splitting event).

Anagentic Range Evolution: Dispersal and Extinction

- Assume that each distinct range (set of areas) is a state in a Continuous-Time Markov model, for a total of 2^n states, where n is the number of areas.

- The instantaneous rate of change of a range of a lineage is then given by the matrix, $\mathbf{Q}$, shown opposite, where:

  - D_{ij} is the instantaneous rate of dispersal of the lineage from area i to area j.
  - E_i is the instantaneous rate of local extinction of the lineage from area i.
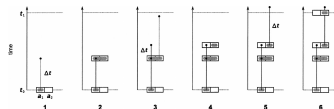
$$
Q = \begin{bmatrix}
 & \emptyset & 1 & 2 & 3 & 12 & 13 & 23 & 123 \\
\hline
\emptyset & - & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & E_1 & - & 0 & 0 & D_{12} & D_{13} & 0 & 0 \\
2 & E_2 & 0 & - & 0 & D_{21} & 0 & D_{23} & 0 \\
3 & E_3 & 0 & 0 & - & 0 & D_{31} & D_{32} & 0 \\
12 & 0 & E_2 & E_1 & 0 & - & 0 & 0 & D_{13} + D_{23} \\
13 & 0 & E_3 & 0 & E_1 & 0 & - & 0 & D_{12} + D_{32} \\
23 & 0 & 0 & E_3 & E_2 & 0 & 0 & - & D_{21} + D_{31} \\
123 & 0 & 0 & 0 & 0 & E_3 & E_2 & E_1 & -
\end{bmatrix}.
$$

Anagentic Range Evolution: Dispersal and Extinction

$$Q = \begin{bmatrix}
 & \varnothing & 1 & 2 & 3 & 12 & 13 & 23 & 123 \\
\hline
\varnothing & - & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & E_1 & - & 0 & 0 & D_{12} & D_{13} & 0 & 0 \\
2 & E_2 & 0 & - & 0 & D_{21} & 0 & D_{23} & 0 \\
3 & E_3 & 0 & 0 & - & 0 & D_{31} & D_{32} & 0 \\
12 & 0 & E_2 & E_1 & 0 & - & 0 & 0 & D_{13} + D_{23} \\
13 & 0 & E_3 & 0 & E_1 & 0 & - & 0 & D_{12} + D_{32} \\
23 & 0 & 0 & E_3 & E_2 & 0 & 0 & - & D_{21} + D_{31} \\
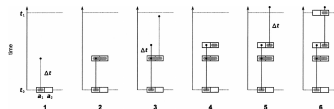123 & 0 & 0 & 0 & 0 & E_3 & E_2 & E_1 & -
\end{bmatrix}.$$

- D_{ij} is given by the product of the *global dispersal rate*, d, and the *connectivity* of area i to j.
- The global dispersal rate, d, is estimated from the data.
- The connectivity of area i to j is the relative weight of dispersal from area i to j, and is specified by the model.
    - If 1.0 for all i, j: all areas are equally connected.
    - If 0.0: no dispersal possible between area i and j.
    - Connectivity can be scheduled to change at fixed times to represent or test the geological history of a region.
- E_i is given by e, the *extinction rate*, which is estimated from the data.

Anagentic Range Evolution: Dispersal and Extinction

- If the duration of a lineage is t, then the transition probability matrix, $\mathbf{P}$, is given by exponentiating the instantaneous rate of change matrix:

$$\mathbf{P}(t) = e^{-\mathbf{Q}t}.$$

- With this, the probability of starting at a particular range state and ending at another particular range state under a given rate of dispersal and extinction can be calculated for any edge in the phylogeny.

- The product of this calculation across all edges in the phylogeny gives us (anagentic range evolution) part of the probability of any particular estimated history.

$$Q = \begin{bmatrix}
 & \emptyset & 1 & 2 & 3 & 12 & 13 & 23 & 123 \\
\hline
\emptyset & - & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & E_1 & - & 0 & 0 & D_{12} & D_{13} & 0 & 0 \\
2 & E_2 & 0 & - & 0 & D_{21} & 0 & D_{23} & 0 \\
3 & E_3 & 0 & 0 & - & 0 & D_{31} & D_{32} & 0 \\
12 & 0 & E_2 & E_1 & 0 & - & 0 & 0 & D_{13} + D_{23} \\
13 & 0 & E_3 & 0 & E_1 & 0 & - & 0 & D_{12} + D_{32} \\
23 & 0 & 0 & E_3 & E_2 & 0 & 0 & - & D_{21} + D_{31} \\
123 & 0 & 0 & 0 & 0 & E_3 & E_2 & E_1 & -
\end{bmatrix}.$$

Cladogenetic Range Evolution: Sympatry and Vicariance

- At a speciation event, the range of the parent lineage is inherited with modification by the daughter lineages.
- (Original) DEC modeled inheritance under one of three modes:
    - single-area sympatry (if ancestor range consists of only one area)
    - narrow-vicariance or peripheral allopatry (one daugther inherits a single area; other daughter inherits remainder of range)
    - subset sympatry (one daughter inherits a single area; other daughter inherits entire range)

Cladogenetic Range Evolution: Sympatry and Vicariance

- DEC integrates over conditional likelihoods of range inheritance scenarios at internal nodes (under a flat prior over the scenarios) rather than the range states itself.

- Intuitively: for calculation purposes, we can imagine an infinite-length "pseudo-edge" spannning the speciation event, where the rate of transition from the range state just before cladogenesis to any other range state is equal.

Optimization and Estimation

- Pruning algorithm to calculate likelihood of entire phylogenetic history.
- All parameters and range states optimized to find maximum likelihood solution.
  - Original "*lagrange*" optimizes ancestral range states locally.
  - "*BioGeoBEARS*" optimizes ancestral range states globally.

Extending the Classic DEC Model by Adding Other Anagenetic and Cladogenetic Range Evolution Processes