SSB 2015 Model-Based Biogeography

Background and Concepts

Jeet Sukumaran

Ecology and Evolutionary Biology
University of Michigan
Ann Arbor, MI

jeetsukumaran@gmail.com

May 20, 2015

The Field of Biogeography

Historical vs. Ecological Biogeography

- Historical biogeography: "reconstruct the origin, dispersal, and extinction of taxa and biotas"; the relationship between geographical history and evolutionary structure
- Ecological biogeography: "present distributions and geographic variation in diversity in terms of interactions between organisms and their physical and biotic environments"
- Different time scales and different questions (though with a large overlap)
- Different community and methods/approaches (much smaller overlap)
- Different (academic) histories

(Historical) Biogeography vs. Phylogeography

- Both: study of relationships and distribution of lineages under geographical processes; structuring of genetic information by the spatial history of the information carriers.
- Phylogeography
  - Population-level processes dominate
  - Coalescent theory provides a robust, powerful, flexible and highly-informative statistical framework to explore a broad range of questions and issues by explicitly modeling the relationship between a phylogeny and the demographic history of a system.
- Historical biogeography
  - Species-level and higher processes dominate
  - No single framework or theory that statistically relates a phylogeny to the geographic or diversification history of a system

Continuous-Area vs. Discrete-Area Approaches

- Continuous-area models treat the geographical landscape as continuous field.
  - Movement is through a diffusion process.
  - Suitable at time- and spatial-scales where local movement dominate (typically, phylogeography).
- Discrete-area models regionalize and partition space.
  - Require some degree of understanding and assumptions about the system on how to partition space.
  - Usually straightforward (e.g., continents, islands, island/systems, faunal zones, etc.)
  - Allow for selective focus on type of movement to study without noise of non-significant movement (e.g., ignore saturated local movement)
  - Typically preferred for studies at the biogeographic scale.
- Hybrid approaches: discretize region into a very large number of very small areas (e.g. BayArea)

A Brief, Idiosyncratic, and Opinionated Synopsis of the History of Historical Biogeographic Analysis

A Brief, Idiosyncratic, and Opinionated Synopsis of the History of Historical Biogeographic Analysis

- Discussions of the history of biogeography usually start with Wallace (or, in more pedantic versions, Aristotle).
- But here I am going to be discussing the history of historical biogeographical *analysis* (and associated theory).
- So, we will give our just a quick nod to Wallace and his predecessors, and move on ...

Wallace in the Malay Archipelago. A painting by Evstafieff in the collection of Down House. Copyright English Heritage Photo Library.

A Brief, Idiosyncratic, and Opinionated Synopsis of the History of Historical Biogeographic Analysis

1. Center-of-Origin and dispersal: Important questions defined, with answers provided by epic evolutionary sagas
2. Panbiogeography: Incorporation of formal methods of analyses, unfortunately based on a *very* kooky system
3. Cladisitic biogeography and variants: Incorporation of phylogenetics, accompanied by crippling philosophical fanaticism
4. Event-based biogeography: modeling of biogeographical processes in a parsimony framework
5. Statistical process-based biogeography: modeling of biogeographical processes in a probabilistic framework

"Centers-of-Origin" and Other "Just-So" Biogeographical Stories

- Dominated historical biogeography till the 1970's/80's
- Concerned with:
  - identifying the "birthplace" or "cradle" of a taxon
  - tracking the "success" (expansion/dispersal) in relation to evolutionary adaptations or features.
- Study based on formulation and application of "rules", e.g.:
  - boreal superiority (Matthew, 1915): "new, succcessful forms arose in response to the challenges of temperate climates; these forms eventually supplanted their progenitors and other lineages, forcing them inyo peripheral habitats, then to lower latitudes, and eventually into the Southern Hemisphere"; most derived forms found near center of origin
  - progression rule (Henning, 1996): "stepping stone" pattern of sequential dispersal/speciation events laed to more derived forms being found at farther distances away from the center of origin

“Centers-of-Origin” and Other “Just-So” Biogeographical Stories: Issues

- The “rules” are a set of “plot points” of a (great!) evolutionary story
- Rules serve to select evidence that fit the plot
- Identification of “origin” and dispersal/expansion trajectories closely tied with evolutionary “success sagas”
- Failure to consider alternate explanations for patterns (the “Sherlock Holmes Explanation” syndrome).
- Difficult to test, compare, or evaluate competing perspectives.
- Set the stage for the field of enquiry that we pursue to this day: fundamental questions and research programmes.

Croziat's Panbiogeography

- Observation: common patterns of disjunction across multiple groups
- Explanation: Fragmentation of a biota that originally inhabited interconnected regions
- Analysis: Overlapping *tracks* (lines connecting distributions) of taxa
- *Many* issues, some of which are highly problematic, e.g.:
  - Categorical rejection of phylogentics (i.e., rejection of *data*!)
  - Categorical rejection of continental drift
  - Categorical rejection dispersal (even in the case of oceanic islands!)
  - Forced to propose numerous fantastical "land-bridges" spanning the entire planet
- Yet, while the concepts underlying the model were deeply-flawed to the point of delusional, Croziat *did* advance the field by emphasizing the need for formalized methods of analysis (as opposed to story-telling).

Phylogenetics + Bigeography + Vicariance

- Brundin: phylogenetics + biogeography
- Nelson and Platnick: "vicariant biogeography"
- Enshrined vicariance and only vicariance as the only explanatory principle worth studying.
- Rejection of non-vicariant processes based on *philosophical* grounds, rather than scientific or statistical.
- More extreme views reject any discussion of processes, viewing the analysis as a strict pattern-fitting exercise.
- Disagreements over arcane methodological details led to splintering into many related variants: "cladistic biogeography", "phylogenetic biogeography" (Brooks), "parsimony analysis of endemicity" etc.

Phylogenetics + Bigeography + Fundamental(ist) Rejection of Dispersal

- Dispersal and extinction are important processes for many systems studied at most time scales.
- Many times this is so self-evident in the data that, like the epicycles for the earth-centric pre-Kepplerian astronomers, convoluted devices and arcane contrivinces were needed to reconcile reality to deliberately blinded models.
    - Treated as noise to be handled by *a priori* manipulation
    - Explained away through *a posteriori* interpretation (a.k.a., speculative/imaginative story-telling)
- Yet some important concepts were developed and refined that we still find useful today: apart from the incorporation of phylogenetics, area cladograms, treatment of areas as phylogenetic characters, etc.

Event-based Methods: Parsimony Analysis of Bigeographic Processes (DIVA)

- Parsimony character optimization based on minimizing costs of vicariance, dispersal, and extinction events reconstructed on phylogeny to explain pattern of areas on tips.
- Allow for treatment of dispersal and extinction, even if associated with costs (no reason dispersal/extinction are biologically less plausible then vicariance; this is baggage from cladistic biogeography)
- Problematical:
  - Costs of various events need to be assigned *a priori*, and somewhat arbitrarily
  - Reconstructed history is an artifact of optimization and cannot be interpreted as meaningful or interpretable in any way
  - Does not consider time element (same pattern can arise through different histories in different time-scales)
  - Cannot easily compare different geological hypotheses
  - Cannot easily account for phylogenetic uncertainty
- Despite these problems, DIVA was a very important advance in the field as it rescued biogeography from the morass into which it

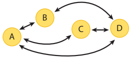
Statistical Modeling of Biogeographic Processes‘

- Today we use methodological paradigms that explicitly model the processes of interest using the biological patterns as *data* instead of *monuments* to support pre-conceived narratives.
- *c.f.* A similar revolution in archaeological thought over a century ago: documents as monuments vs. monuments as documents
- Likelihood-based methods:
  - DEC family
  - GeoSSE
  - BayArea
- Simulation-based methods:
  - Approximate Bayesian Computation
  - Machine Learning

Statistical Modeling of Biogeographical Processes

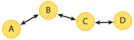
Modeling Geographic States as “Nucleotides”

a



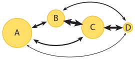
	[A]	[B]	[C]	[D]
[A]	—	μ	μ	μ
[B]	μ	—	μ	μ
[C]	μ	μ	—	μ
[D]	μ	μ	μ	—

b



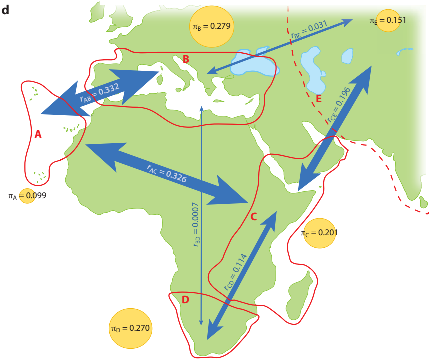
	[A]	[B]	[C]	[D]
[A]	—	μ	0	0
[B]	μ	—	μ	0
[C]	0	μ	—	μ
[D]	0	0	μ	—

c



	[A]	[B]	[C]	[D]
[A]	—	$\pi_B r_{AB}$	$\pi_C r_{AC}$	$\pi_D r_{AD}$
[B]	$\pi_A r_{AB}$	—	$\pi_C r_{BC}$	$\pi_D r_{BD}$
[C]	$\pi_A r_{AC}$	$\pi_B r_{BC}$	—	$\pi_D r_{CD}$
[D]	$\pi_A r_{AD}$	$\pi_B r_{BD}$	$\pi_C r_{CD}$	—

d



The Dispersal-Extinction-Cladogenesis (DEC) Model

- Seminal work by Ree (2005) and Ree and Smith (2008) described a probabilistic framework that models the evolution of *ranges* of lineages under various biogeographical processes.
- A range is a collection of areas constituting a lineage distribution.
- Ranges can expand or contract over the history of the lineage (anagenetic range evolution: dispersal, extinction) or through differential inheritance by daughter lineages at speciation (cladogenetic range evolution: sympatric speciation, vicariance).
- Connectivity between areas can be controlled or changed over time, allowing for incorporation or testing of geological history.
- Allows for inference of ancestral or origin areas for different clades, testing of hypotheses of different geographical connectivities, timing of vicariance events that have structured biota, etc.

The Geographic State Speciation and Extinction (GeoSSE) Model

- Goldberg and Ree (2011) described a probablistic model by which related geographic distribution to rates of diversification in a probabilistic framework.
- Based on BiSSE (which related character traits to rates of diversification).
- By using geographic areas occupied by a species as proxies for habitat preferences, the GeoSSE approach can be used to study the relationships between ecology and diversity.

BayArea: Stochastic Mapping of Biogeographical Histories

- Landis et al. (2013) presented a Bayesian approach that allows for handling of a very large number of areas.
- Biogeographic history is simulated on the phylogeny, with rejection sampling to only accept histories that are consistent with the data (data augmentation).
- As the probabilities for each history can be computed just based on instantaneous rate matrix, avoiding the need for matrix exponentiation, likelihood calculations are relatively fast.
- Allows for dispersal rates to vary with geographic distance.
- MCMC sampling to produce approximation of posterior probability of biogeographic histories.
- For any phylogenetic node, can summarize marginal posterior probability of area occupancy or range state.

Other Approaches

- Simulation-based approaches (ABC, machine learning) offer many advantages over full likelihood-based (both ML and Bayesian) methods.
- Not so for speed (speed advantage misleading: does not take into account summary statistic and theoretical development needed on a case-by-case basis), but flexibility.
- Flexibility is fantastic: if you can simulate it, you can analyze it!
- Downside: lots of work needed to characterize performance, robustness, sensitivity, and other behavior; often needs to be done on a case-by-case basis, since this can vary with data, analytical objectives, summary statistics, or details of simulation.