Integrative Biology 200A "PRINCIPLES OF PHYLOGENETICS"

University of California, Berkeley Kipling Will- In revision 2014

Host - Associate
Organism - genes
Host organism - parasite organism
Geographic area - organism

I. Historical explanation for disjunct patterns can be either dispersal or fragmentation of a continuous range (vicariant).

A. Vicariance biogeography considers dispersal to be an ad hoc explanation that could be used to account for any pattern. Dispersal is not falsifiable.

B. Implied process assumptions in pattern methods, which are more explicit in event-based methods

1. allopatric speciation is relatively more common
2. dispersal is relatively less common than stasis
3. current distributions generally reflect ancient events

C. The basics

1. You need monophyletic groups that occur in three or more areas
    a. Requires that you assume the phylogeny is correct
2. Replace the OTU names with area names to make taxon-area cladograms
    a. Requires that you assume ranges are adequately known
    b. Requires assuming the areas are meaningful/correct
3. Summarize the set of taxon-area cladograms to make a general area cladogram that shows the congruent vicariant pattern
    a. Assumes that the majority of taxa respond the same most of the time
4. "Missing", widespread taxa and redundant areas are problems.
    a. Areas may be absent in groups because
        i. no member ever occur in the area
        ii. sampling error
        iii. extinction
    b. Taxa may occur in more than one area because
        i. non-response to a vicariant event
        ii. dispersal (secondary sympatry)
    c. Redundant distributions
        i. dispersal in one or more taxa
        ii. sympatric speciation

II. *A priori methods*- Derive general biogeographic patterns from shared cladistic patterns and distributional data. No secondary or "total evidence" analysis.

A. Component Analysis. TAC, FAC, RAC, GAC: Use taxon area cladograms (TAC) to derive reduced area cladograms (sometimes called fundamental area cladograms (FAC)) from phylogenetic hypotheses and distributional data under Assumptions 0, 1, 2 (Zandee & Roos 1987; Platnick & Nelson 1978). General Area cladograms (GAC) are the intersection of reduced area cladograms (RAC) from two or more sets of trees.

Assumption 0: assumes that widespread taxa are the result of non-response to only the most recent vicariant event, so that areas that are occupied by widespread taxa are interpreted as being most closely related. Clearly, the existence of more than one widespread taxon, each occupying overlapping but not identical sets of areas are conflicting evidence for area relationships. The conflict is resolved for these redundant areas for the most apical cladistic position.

Assumption 1: assumes that widespread taxa are the result of non-response to the most recent OR earlier vicariant event. Redundant area representations are all considered valid. The assumption will often produce more trees (the convex set) than the more restrictive assumption 0, but will always include Assumption 0 trees as a subset.

Assumption 2: allows for the possibility that dispersal, as well as non-response to a vicariant event, may be the explanation for widespread taxa. (fix one move the other to all possible positions).

B. Reconciled Trees (Page 1990, 1994): Mapping one tree onto another by the inclusion of unobserved events.

1. User tree (host or area relationships) mapped to observed or associate tree (taxa or parasites)
2. Items of Error (IOE): Mis-fit of the trees. Number of nodes in reconciled tree – number of nodes in associate tree (in Componant 2.0 "leaves added" = ½ IOE)
3. Various area cladograms are reconciled to find an optimal fit.
4. Heuristic searches may focus on trees that minimizes leaves added, however this may overestimate the number of events.

III. *A Posteriori* methods- taxon distributional data is compiled into a single matrix and subsequently analyzed.

A. Brooks parsimony analysis -BPA (Brooks 1985,1990; Wiley 1987)

See the debate:

Siddall, M.E., Perkins, S.L., 2003. Brooks Parsimony Analysis: a valiant failure. Cladistics, 19:554-564.

Brooks, Daniel R., Dowling, Ashley P. G., van Yeller, Marco G. P. & Hoberg, Eric P. 2004. Ending a decade of deception: a valiant failure, a not-so-valiant failure, and a success story. Cladistics 20 (1):32-46.

1. Convert taxon cladogram an area x taxon/node matrix (MRP)
2. Analyze to find the parsimony solution (primary BPA)
3. Widespread, redundant and missing taxa may cause the solution to include impossible ancestor-descendent relationships. If so, re-code offending areas (duplicate) and re-run analysis or force characters (taxa and their ancestor-descendent relationships) to be polarized and irreversible (secondary BPA).
4. Combine many such matrices and read the resulting relationships from the parsimony tree(s).

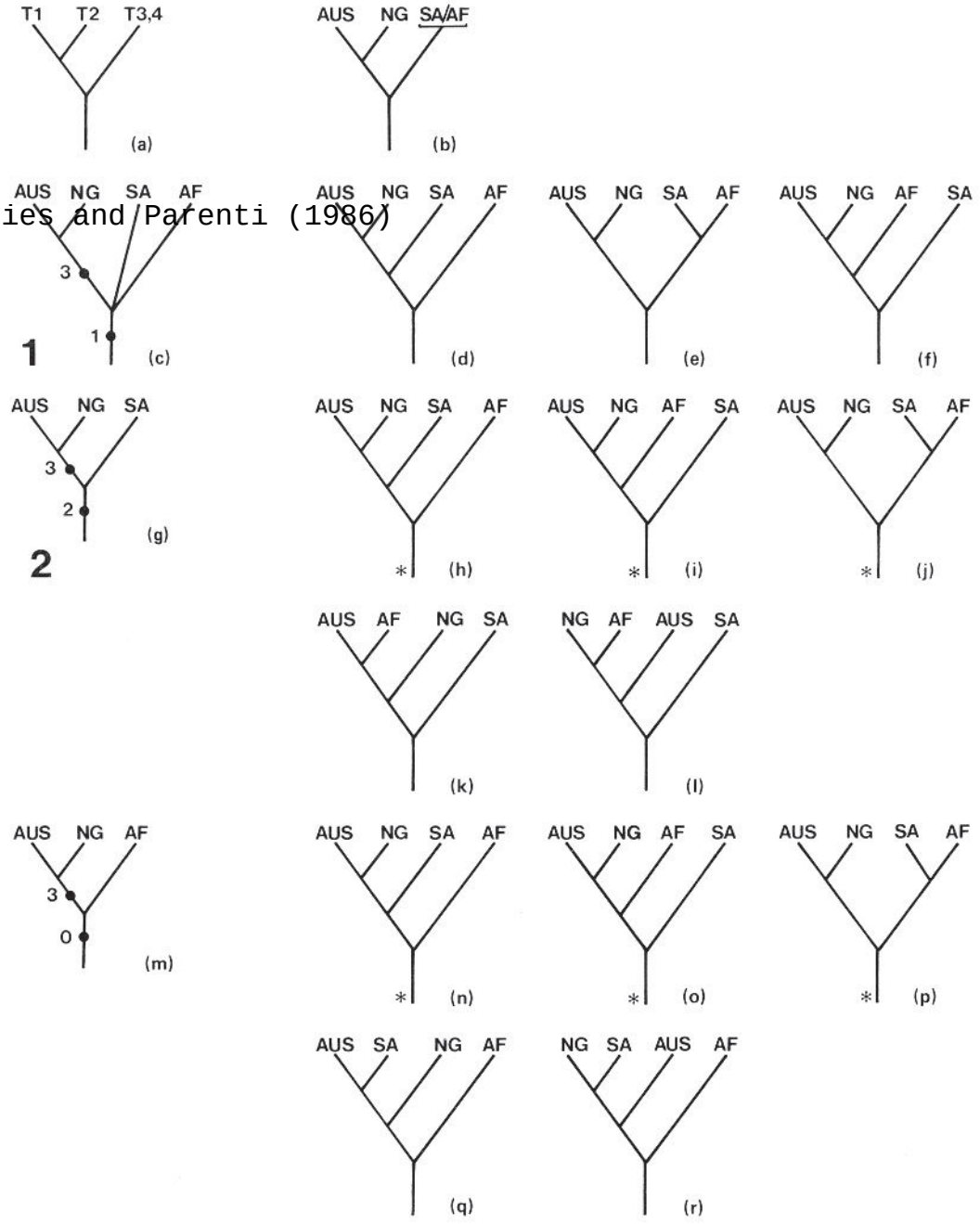
B. Comparative Biogeography (sensu Parenti & Ebach 2009)

1. Focus is on homology of natural areas not on speciation or other diversity generation processes.
2. Defined as “the comparative study of biotic areas and their relationships through time”
3. Uses three area statements for “paralogy-free subtree analysis”

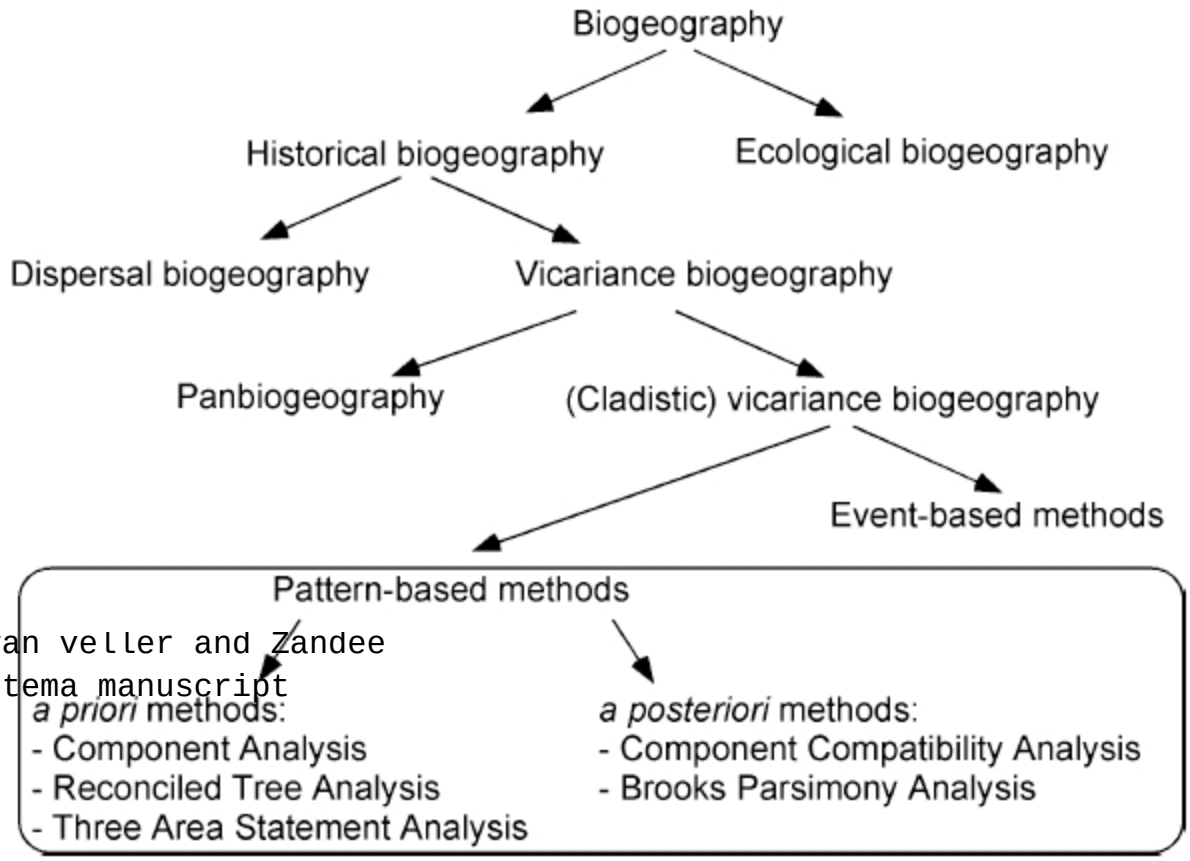
IV. Event based optimality methods for estimation of biogeographic history

A. Event-based methods is rapidly expanding due to advances in ML and Bayesian methods, parametric statistics, DNA data and fossil data.

1. Calibrated phylogenies
2. DIVA, DEC, HVM…



from Humpries and Parenti (1986)



from van veller and Zandee
Biosystema manuscript

Fig.1. Hierarchical outline of the scientific discipline of biogeography.

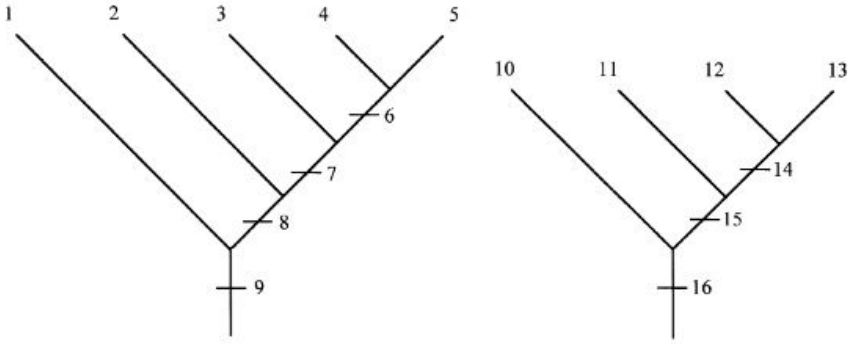


Figure 6 Phylogenetic trees for species 1–5 and 10–13, with internal branches numbered for matrix representation.

Table 6 Matrix listing areas A–E, the species that inhabit them (1–5 and 10–13), and the binary codes representing those species and their phylogenetic relationships

Areas	Taxa	Binary code*
A	1, 10	1000000011000001
B	2, 11	0100000110100011
C	3, 12	0010001110010111
D	4, 13	0001011110001111
E	5	0000111112222222

*? = taxa missing from an area.

lists the binary codes for the members of clades 1–5 and 10–13 in each area. Areas lacking members of one clade are coded as missing data ('?') as suggested by Wiley (1988a, b), so as not to bias the analysis with *a priori* assertions of either extinction of dispersal. The area cladogram produced by

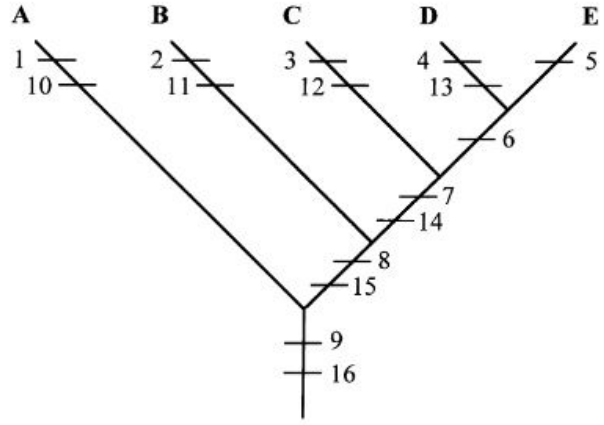
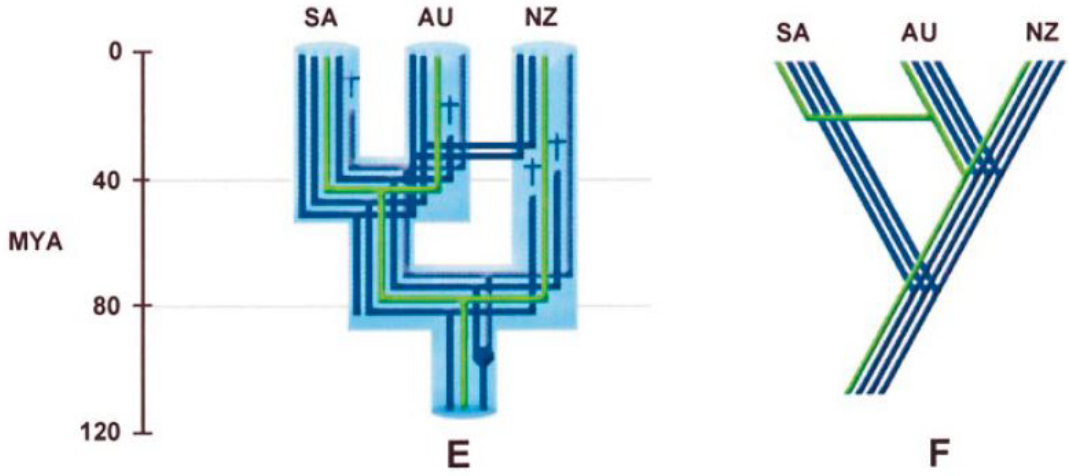


Figure 7 Area cladogram for areas A–E based on phylogenetic relationships of species 1–5 and 10–13.



Annu. Rev. Ecol. Evol. Syst. 2011.42:441-464. Downloaded from www.annualreviews.org
by University of California - Berkeley on 04/18/12. For personal use only.

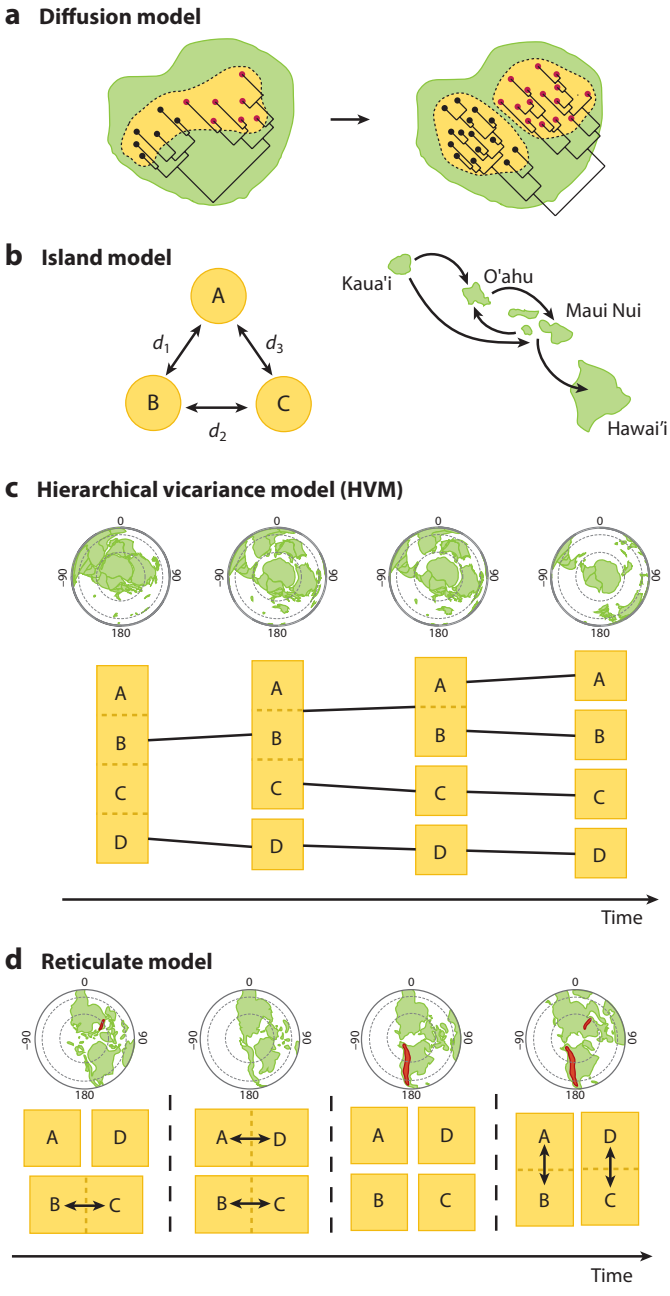


Figure 1

Model types considered in phylogenetic biogeography. Diffusion models are used primarily for phylogeographic analysis but may also be of interest in classical historical biogeography (*a*). Models that use discrete areas (A–D) fall into the categories of static island models (*b*), hierarchical vicariance models (HVMs) (*c*), and reticulate models (*d*). Static island models may allow variation over time in dispersal rates, d_1 to d_3, but areas are constant. Both HVMs and reticulate models assume different area configurations over time. The empirical examples shown here are the Hawaiian archipelago (island model), Gondwana breakup (HVM), and Northern Hemisphere biogeography (reticulate model).