

# Biogeography & Phylogeography

THE AMOEBA PEOPLE  
present

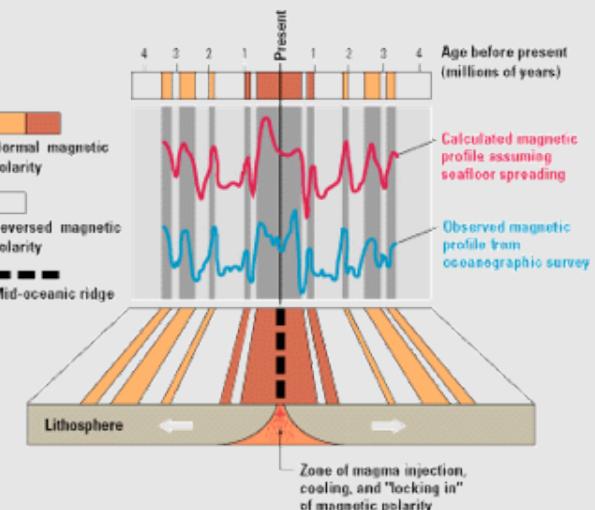
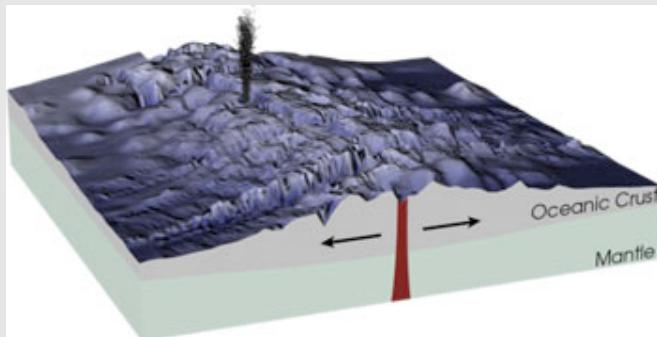
[http://www.youtube.com/  
watch?v=T1-cES1Ekto](http://www.youtube.com/watch?v=T1-cES1Ekto)

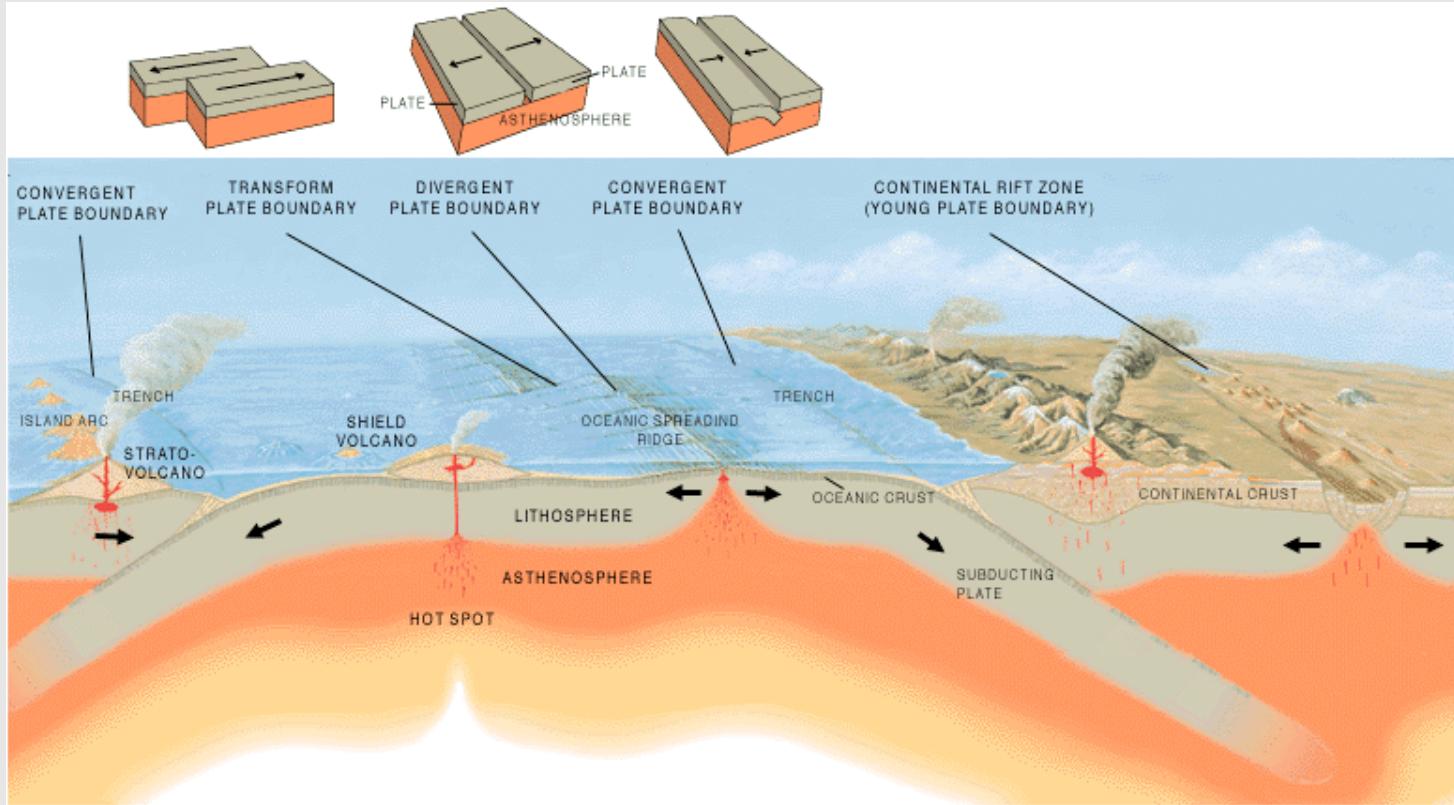
Brian O'Meara  
EEB464 Fall 2019

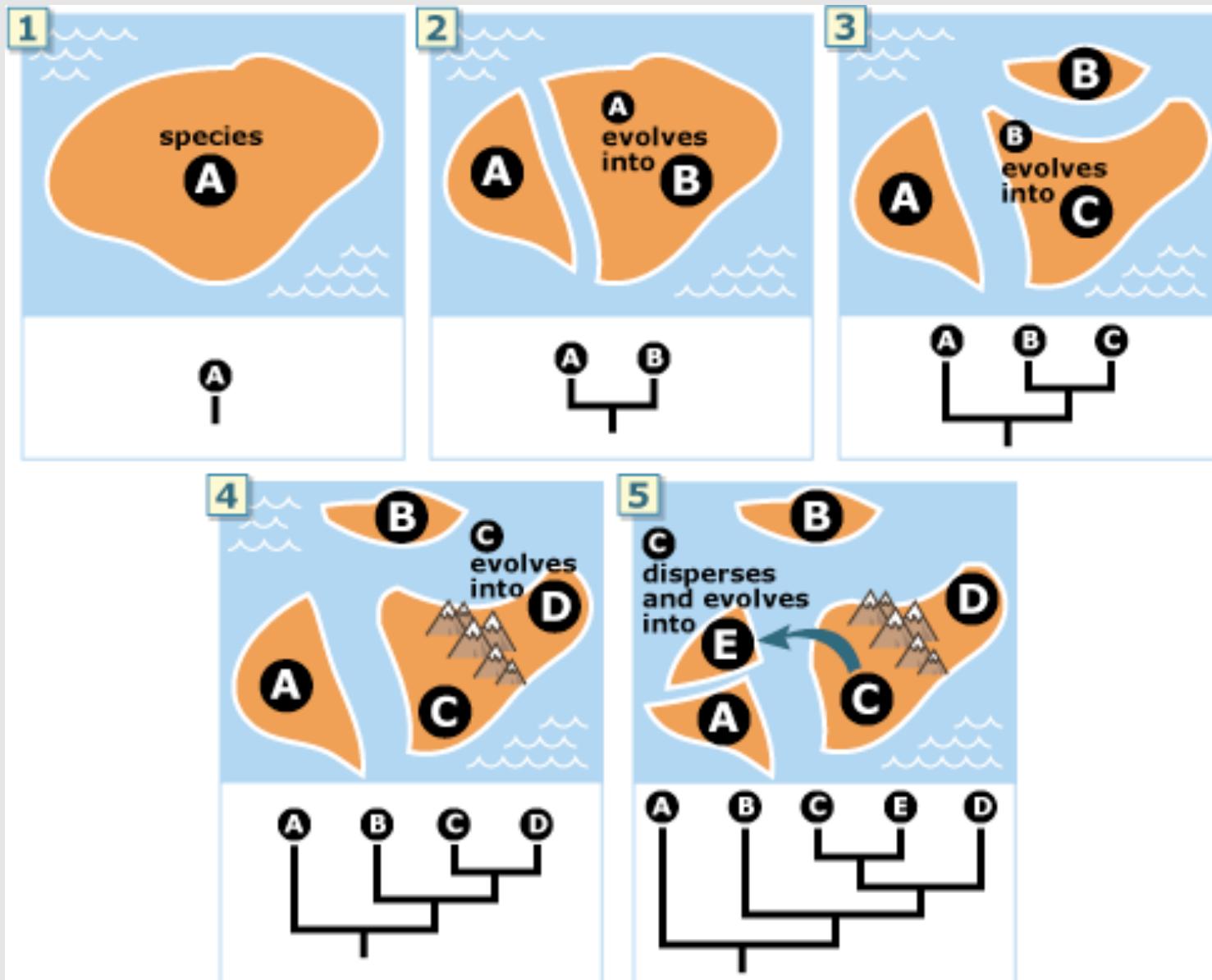
Learning objectives: Understand

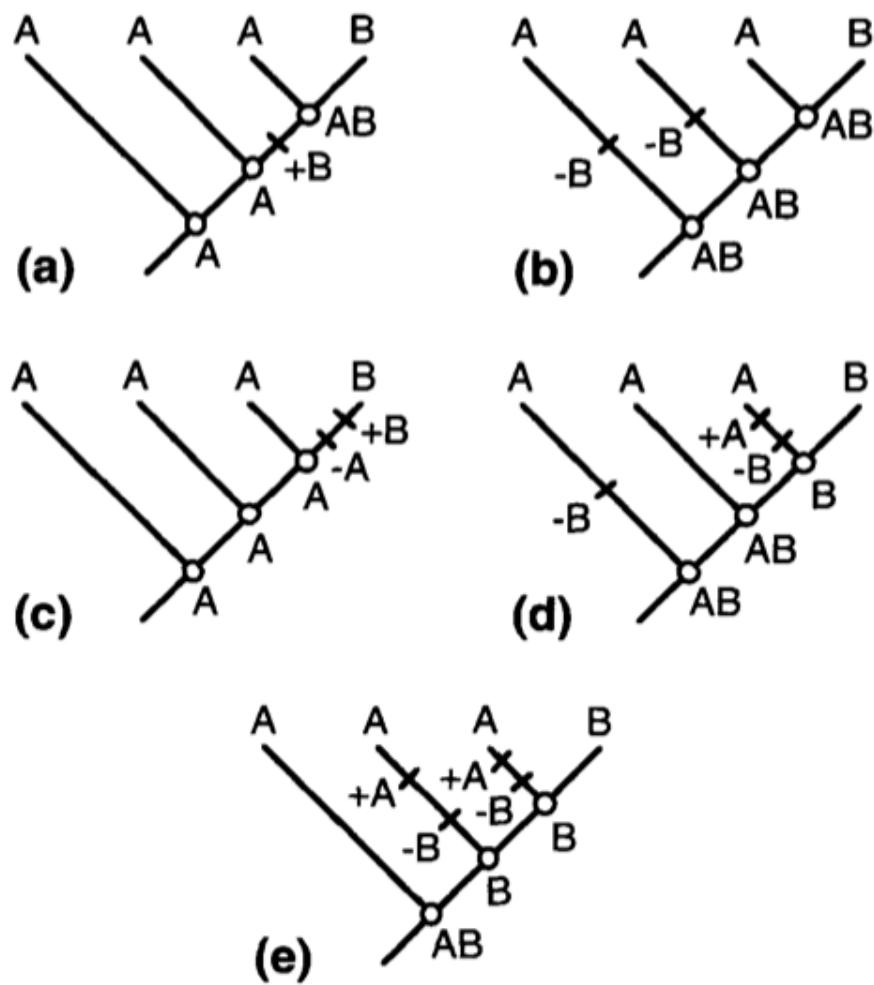
- Continental drift
- Dispersal and vicariance
- Major biogeographic events
- Phylogeography
- Perils of methods
- Uses of phylogeography

- 1912: Alfred Wegener proposes continental drift, Pangaea
- “Reaction to Wegener's theory was almost uniformly hostile, and often exceptionally harsh and scathing.... Part of the problem was that Wegener had no convincing mechanism for how the continents might move.... Another problem was that flaws in Wegener's original data caused him to make some incorrect and outlandish predictions.... Wegener's theory found more scattered support after his death, but the majority of geologists continued to believe in static continents and land bridges.” <http://www.ucmp.berkeley.edu/history/wegener.html>
- Patterns of preserved geomagnetism found in the 1950s and 1960s provided evidence that this was right



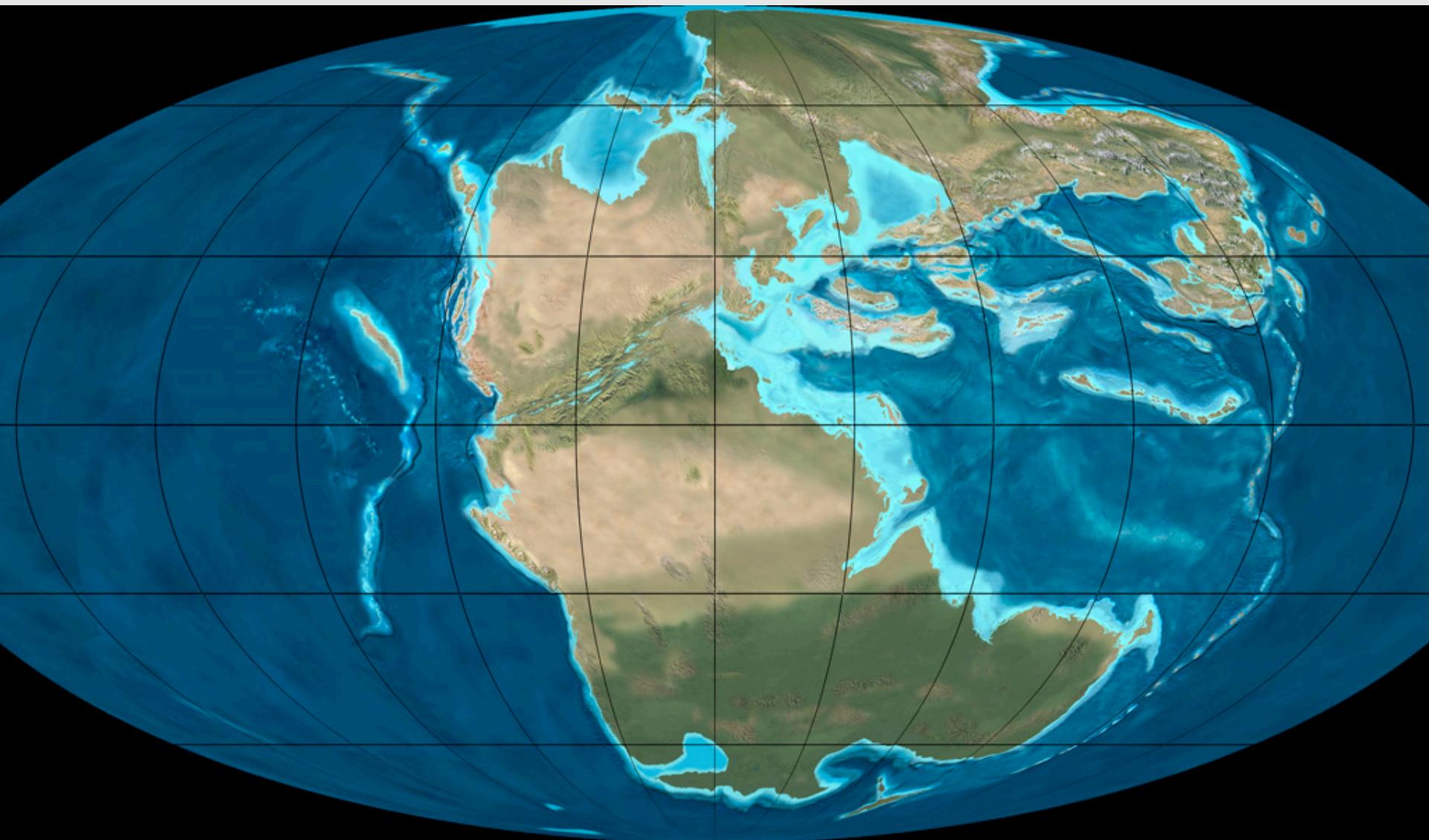






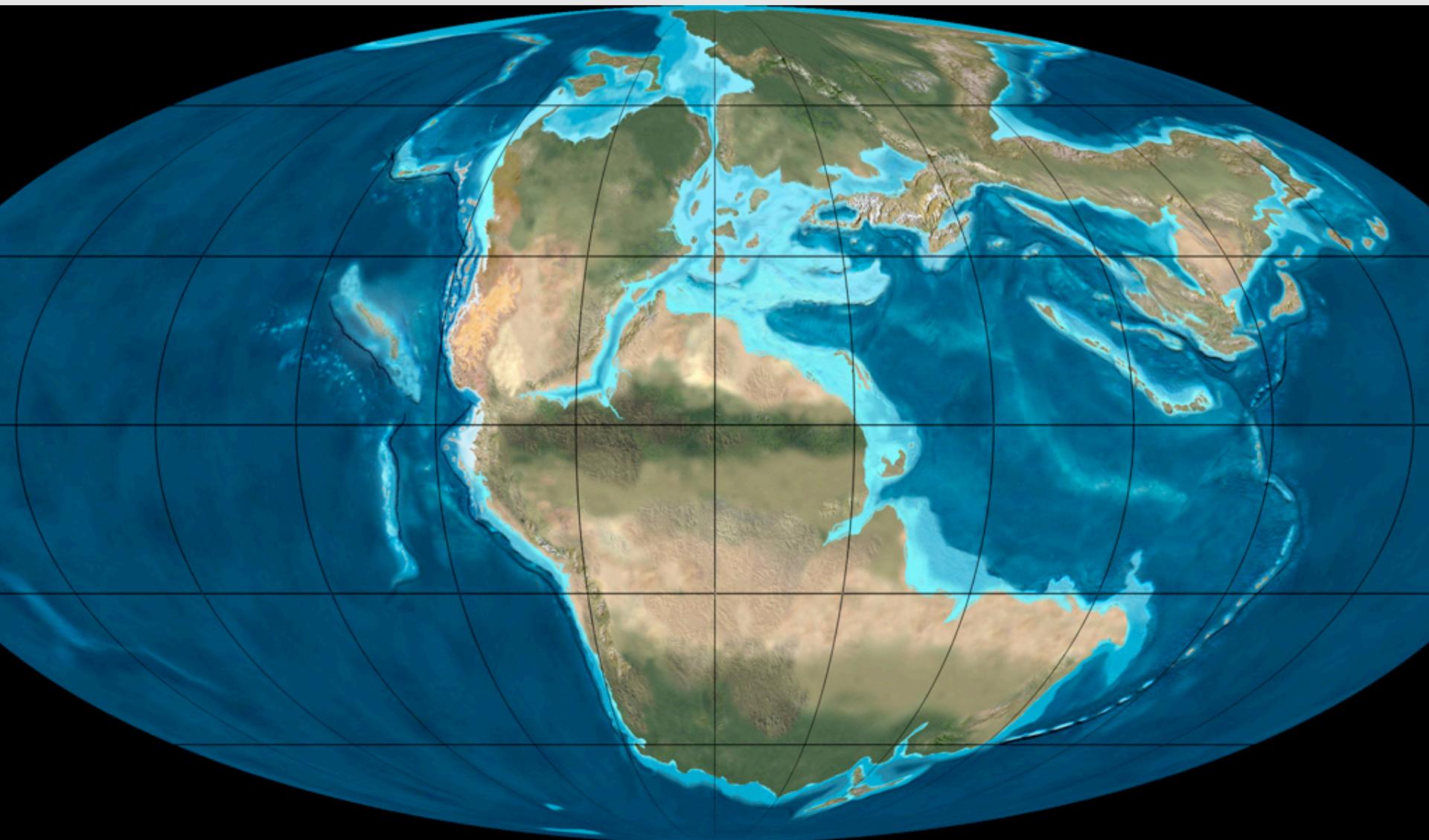
**FIGURE 3.** A hypothetical area cladogram. (a) Dispersal–vicariance analysis postulates dispersal into area B followed by vicariance. (b) Reconciliation with the only possible area cladogram suggests that two basal lineages went extinct in area B. This explanation is less parsimonious because, in addition to the extinction events, it also requires some dispersal to explain how the two earliest widespread ancestors both gave rise to two widespread descendants (duplications in the terminology of Page, 1995). (c) Brooks parsimony analysis indicates that the dispersal to area B occurred on the terminal branch. In a dispersal–vicariance framework, this reconstruction is suboptimal because it requires an extinction event in addition to the dispersal event. (d, e). Two of the three optimal solutions using the maximum cospeciation method proposed by Page (1995). The third solution is that given in cladogram b. All require more events than the dispersal–vicariance reconstruction because of the hierarchical constraint on area relationships.

Late Triassic (220Ma)



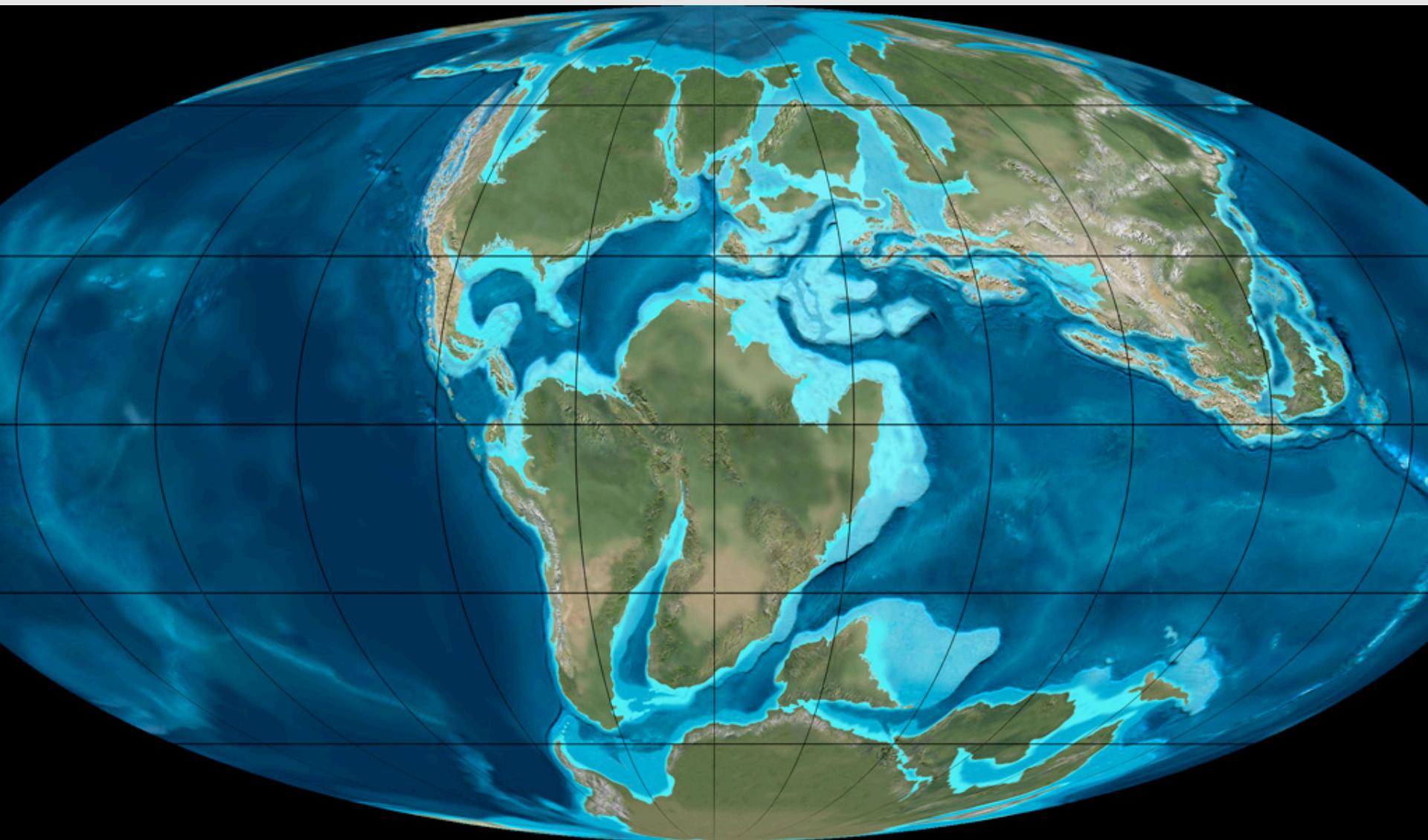
Dr. Ron Blakey

Early Jurassic (200 MYA)



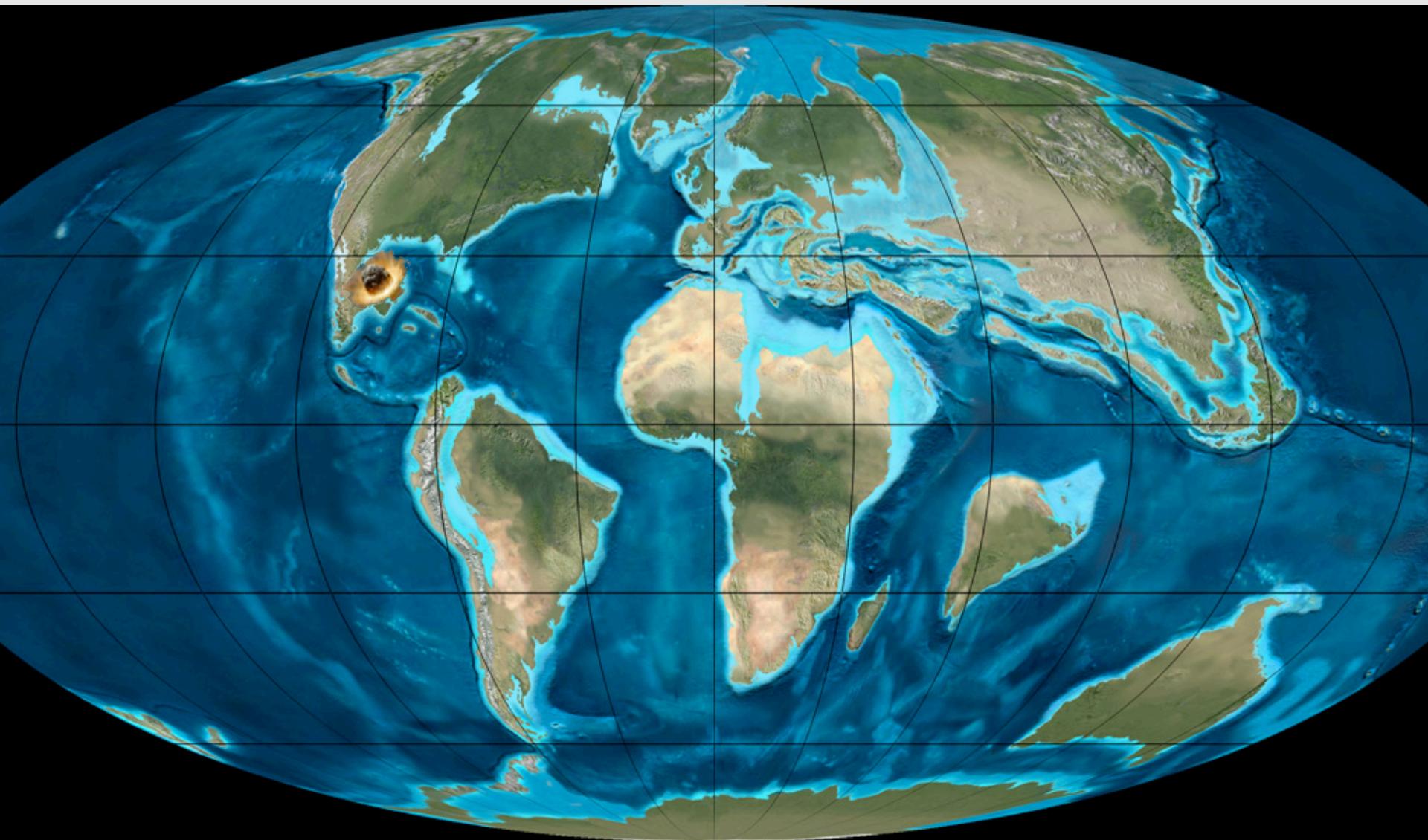
Dr. Ron Blakey

Early Cretaceous (120 MYA)



Dr. Ron Blakey

# KT boundary (65 MYA)

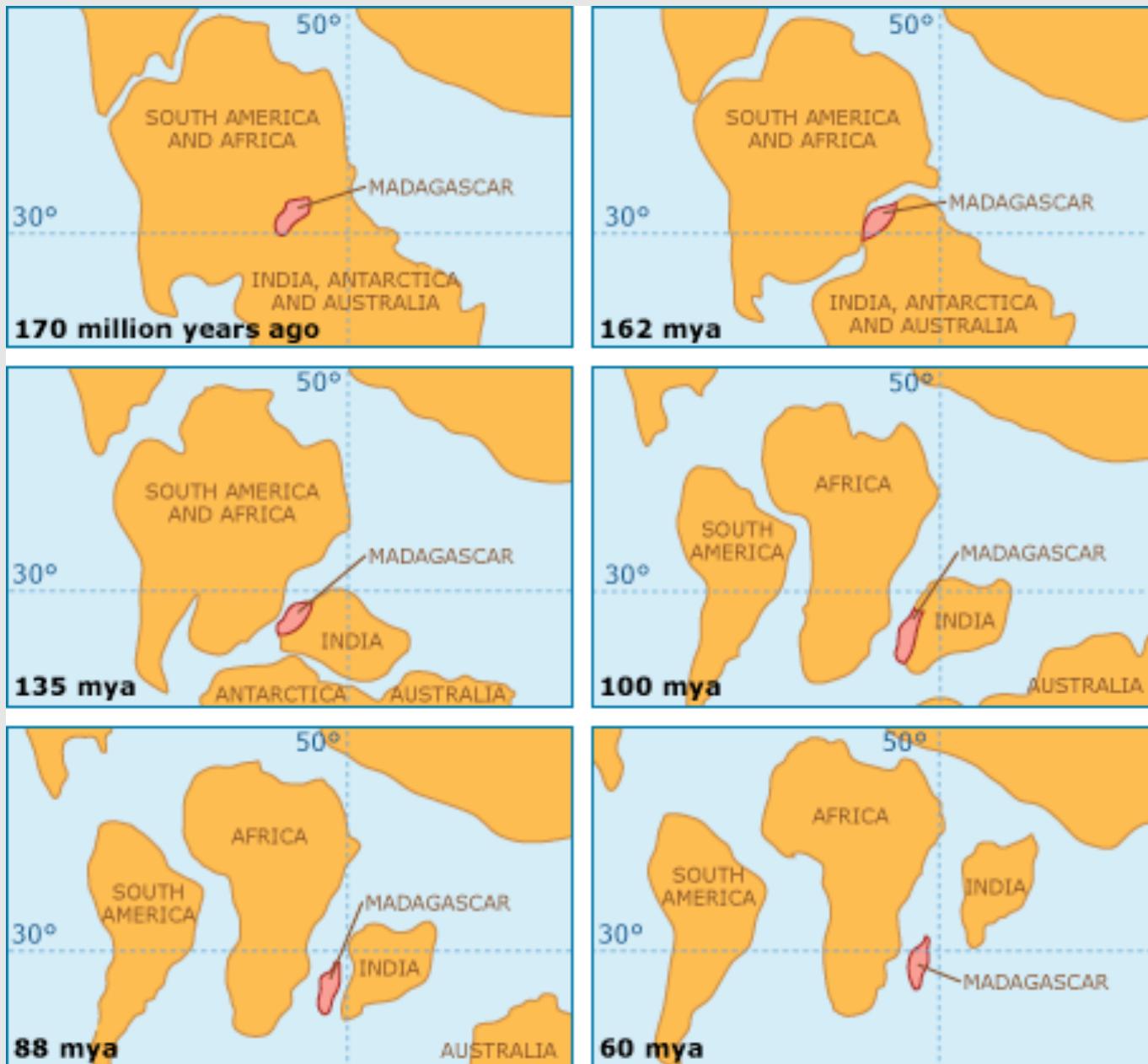


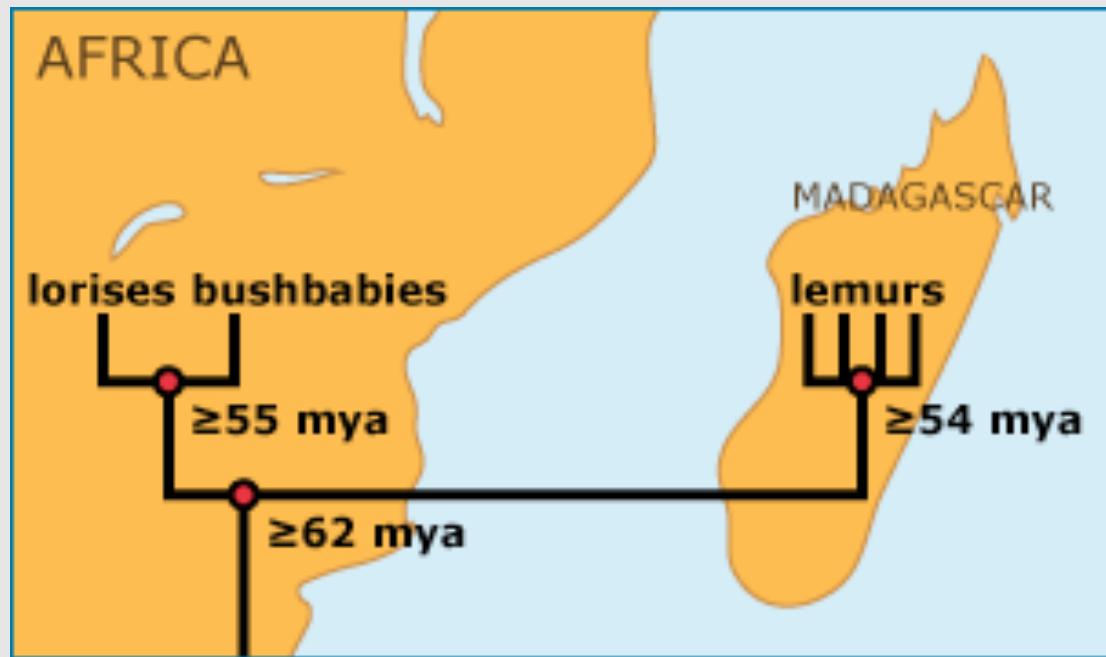
Dr. Ron Blakey

Asteroid image from David Hardy



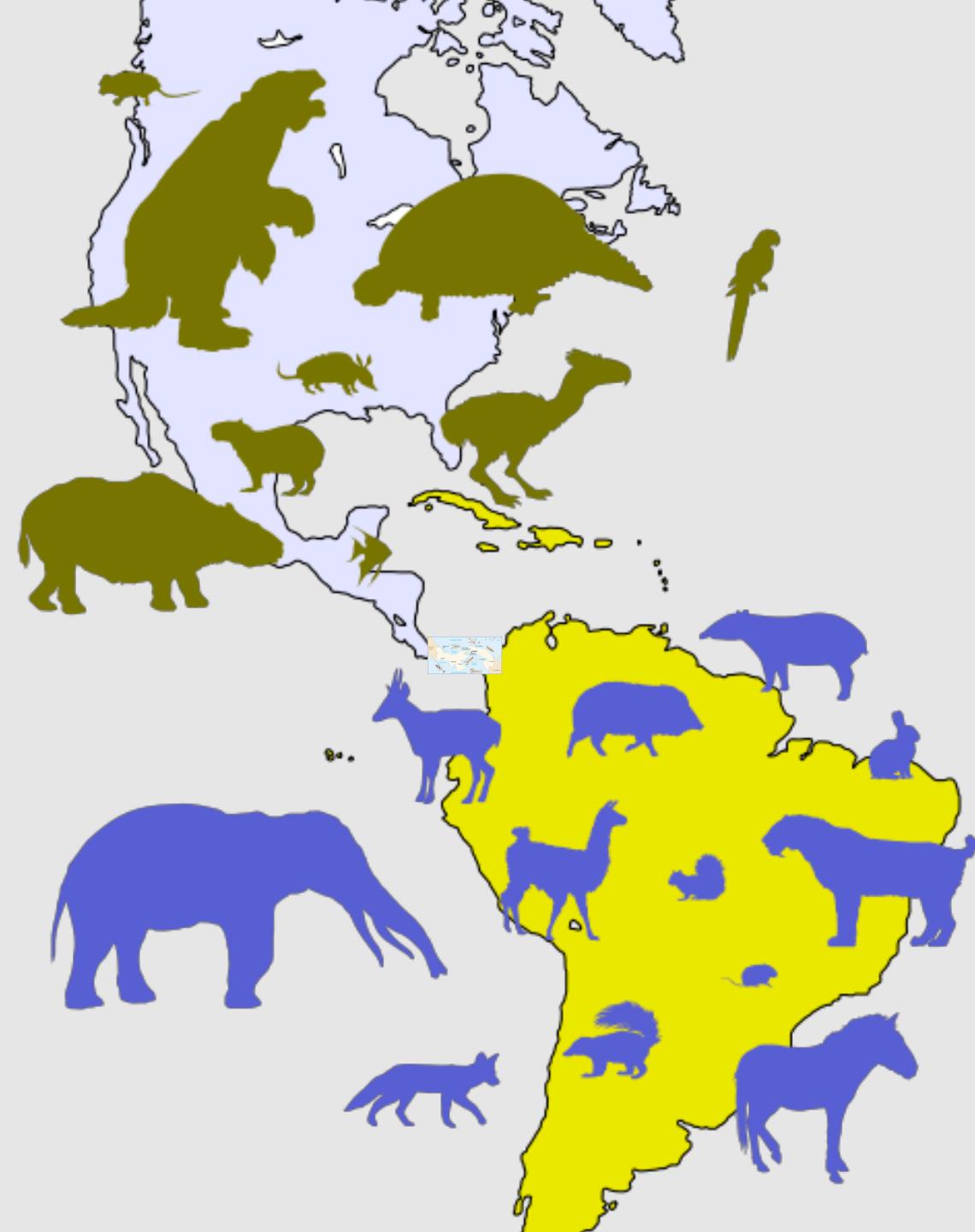
Chris Jin

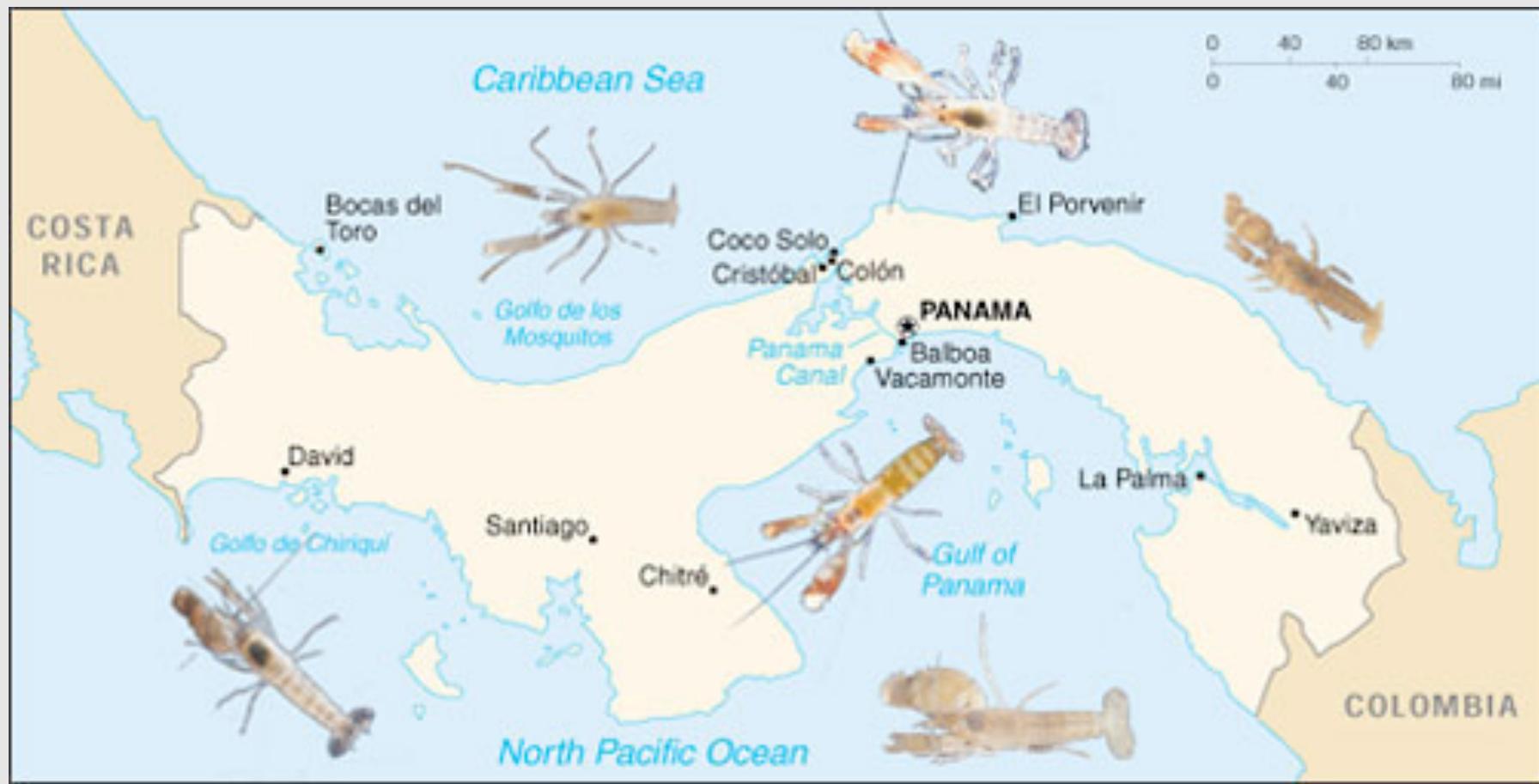




[http://evolution.berkeley.edu/evolibrary/news/091001\\_madagascar](http://evolution.berkeley.edu/evolibrary/news/091001_madagascar)

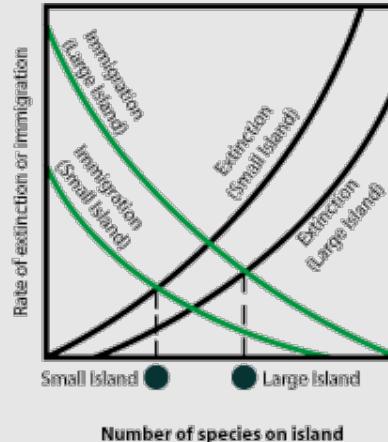
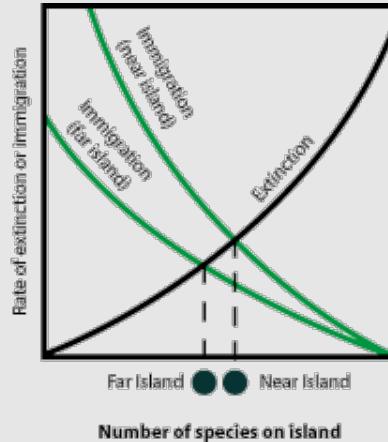
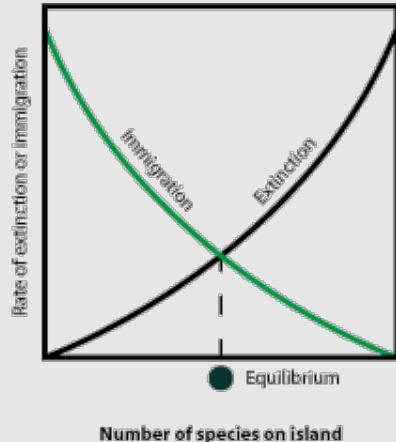
# Great American Biotic Interchange (~2.7 MYA)



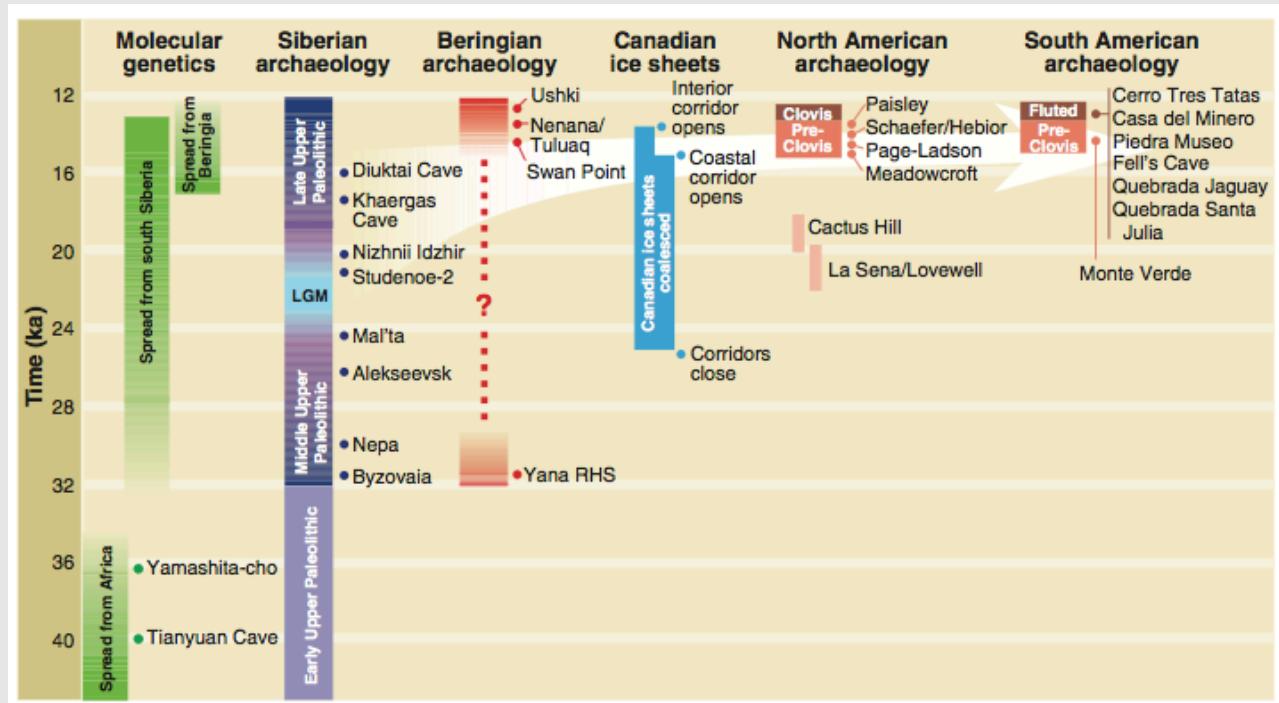
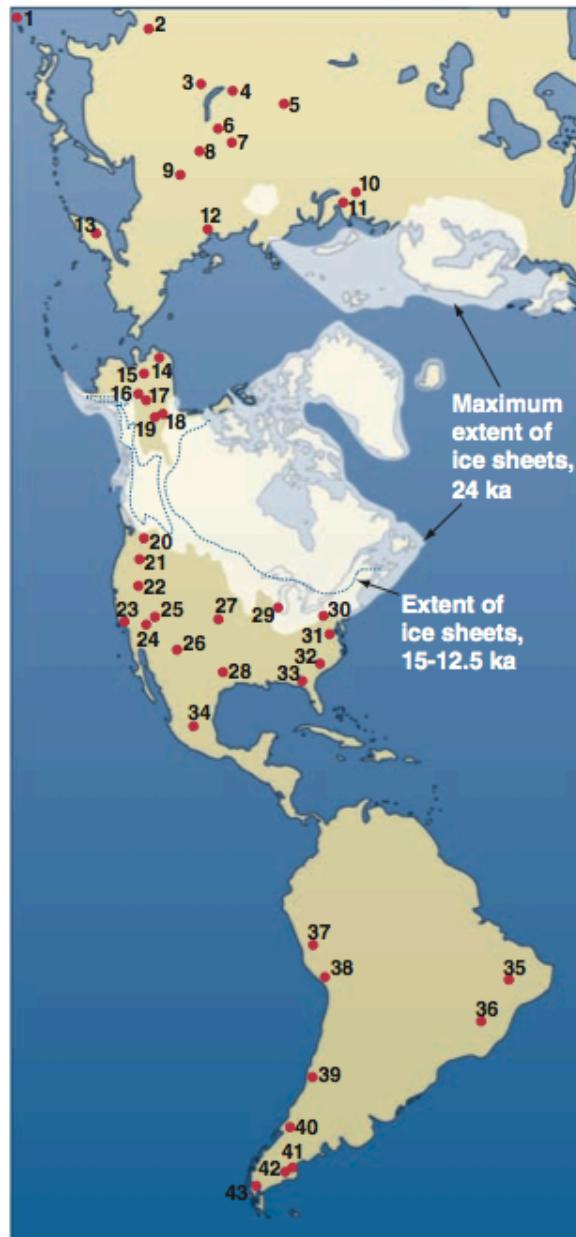


The Smithsonian Institution/Carl Hansen and Nancy Knowlton

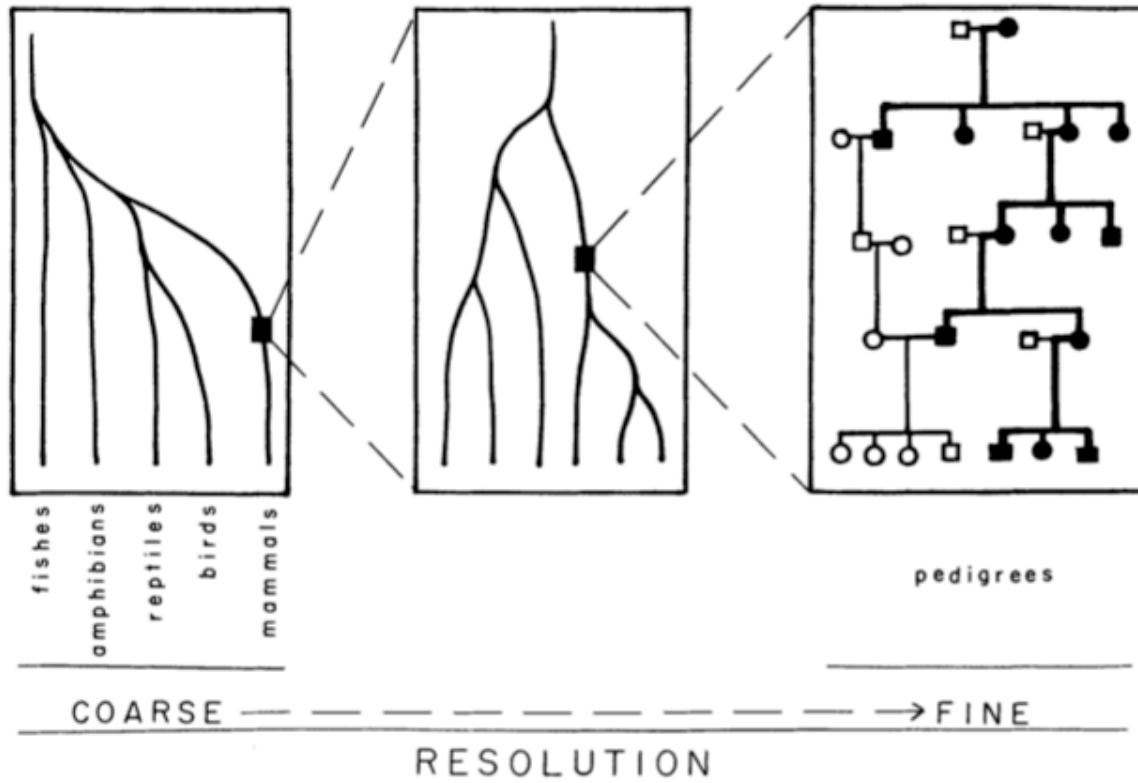
# Island Biogeography: MacArthur & Wilson



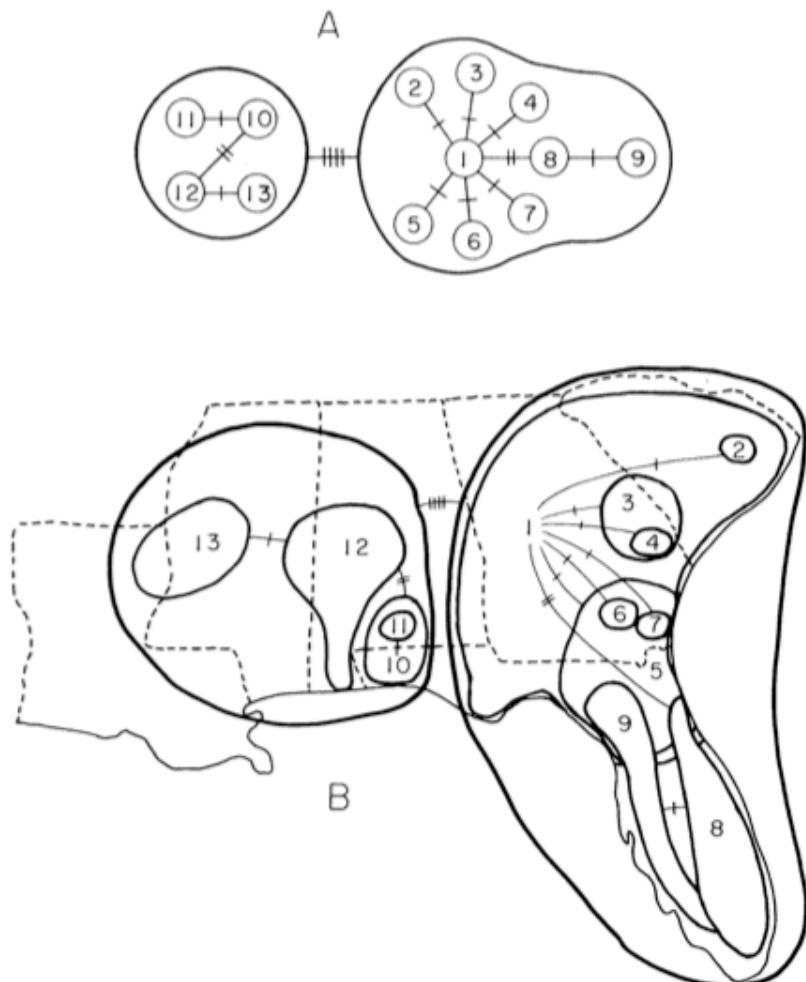




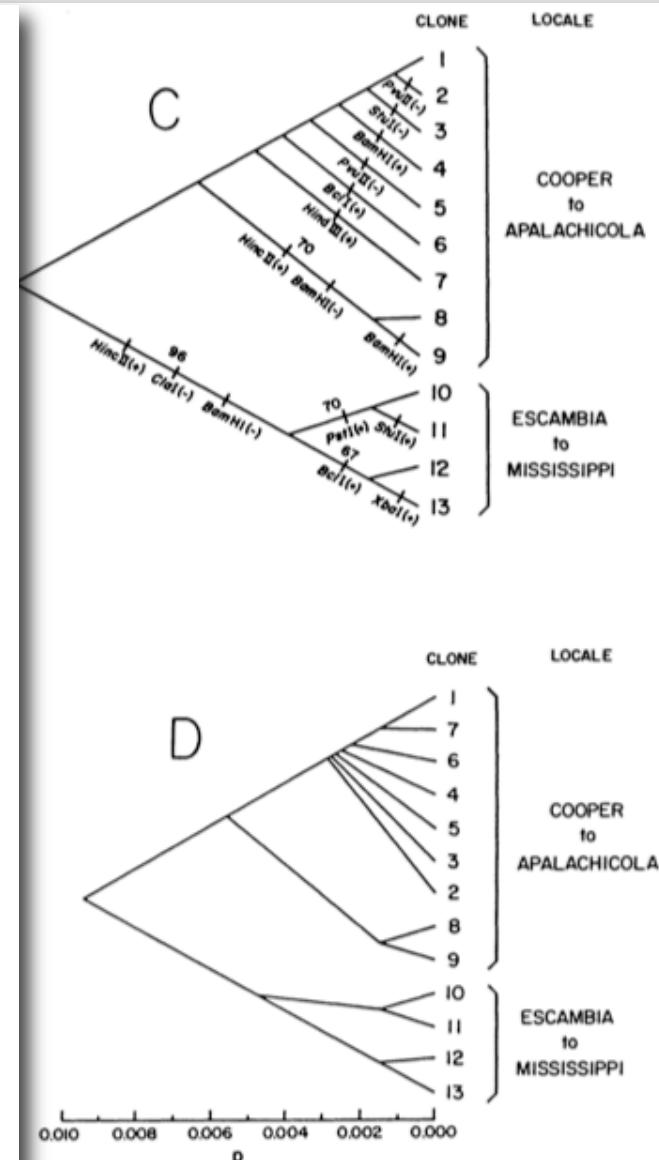
**Fig. 3.** Combined, the molecular genetic and archaeological records from Siberia, Beringia, and North and South America suggest humans dispersed from southern Siberia shortly after the last glacial maximum (LGM), arriving in the Americas as the Canadian ice sheets receded and the Pacific coastal corridor opened, 15 ka.



*Figure 1* At closer levels of examination, macroevolutionary trees (such as the one on the left summarizing relationships among some of the vertebrate classes) must in principle have a substructure consisting of smaller and smaller branches, ultimately resolvable as family pedigrees through which genes have been transmitted. Some branches in the pedigree on the right have been darkened to indicate the transmission path of mtDNA from the earliest pictured female.



**Figure 3** Phylogenetic networks and phenograms summarizing evolutionary relationships among 13 mtDNA genotypes observed in a sample of 75 bowfin fish, *Amia calva* (19). (A) Hand-drawn parsimony network. Slashes crossing branches indicate restriction site changes along a path; heavier lines encompass 2 major arrays of mtDNA genotypes distinguishable by at least 4 restriction site changes. (B) The parsimony network in A superimposed over the geographic sources of collections. (C) Wagner parsimony network computer generated from a presence-absence site matrix. Inferred restriction site changes are indicated, and numbers in the network represent levels of statistical support (by bootstrapping) for various clades. (D) UPGMA phenogram, where  $p$  is estimated nucleotide sequence divergence.



Avise et al.  
(1987)

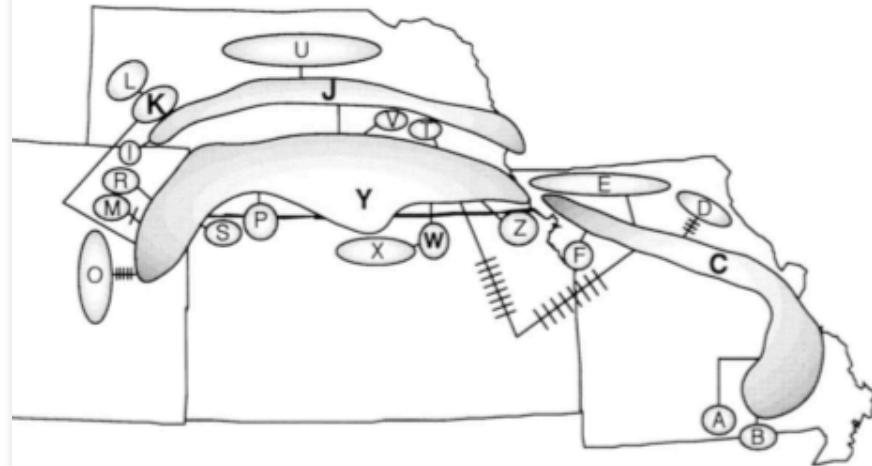
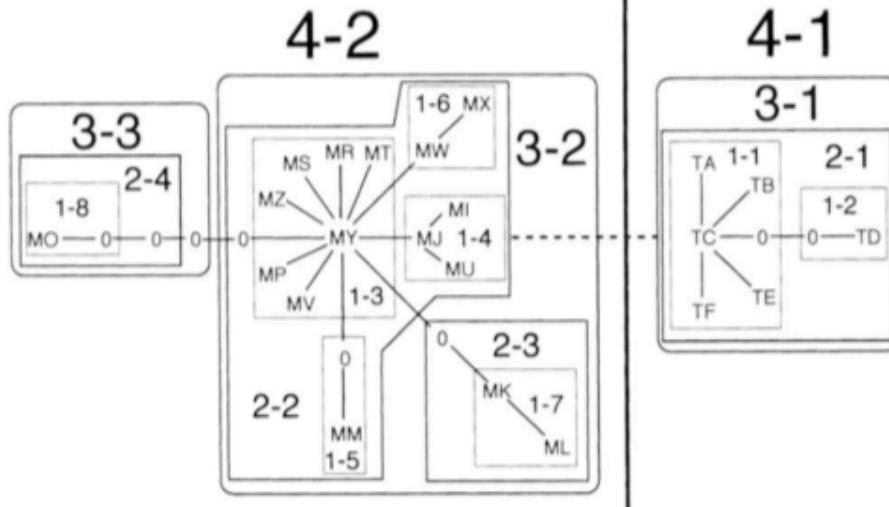


FIGURE 3.—The estimated 95% plausible set of cladograms and associated nested design for the mtDNA haplotypes found in *A. tigrinum*. Haplotypes are indicated by letter designations, as found in ROUTMAN (1993). Zeros indicate haplotype states that are necessary intermediates between observed haplotypes but that were not present in the sample. Each solid line represents a single mutational change that interconnects two haplotype states that has a probability greater than 95%. The thick dashed line indicates a multiple step mutational connection for which the exact interconnections are uncertain and for which parsimony is not supported at the 95% level. Narrow-lined boxes enclose one-step clades, which are designed by “1-*x*” where *x* is a number assigned to identify the clade; wide-lined boxes and polygons enclose two-step clades (“2-*x*”); rounded boxes enclose three-step clades (“3-*x*”); and a thick solid line separates the two four-step clades, 4-1 and 4-2. Technically, these last two clades are greater than four-step clades because of the large mutational distance (a minimum of 14 steps) that separates them, but these clades are the only nondegenerate categories at the level of four-step or above.

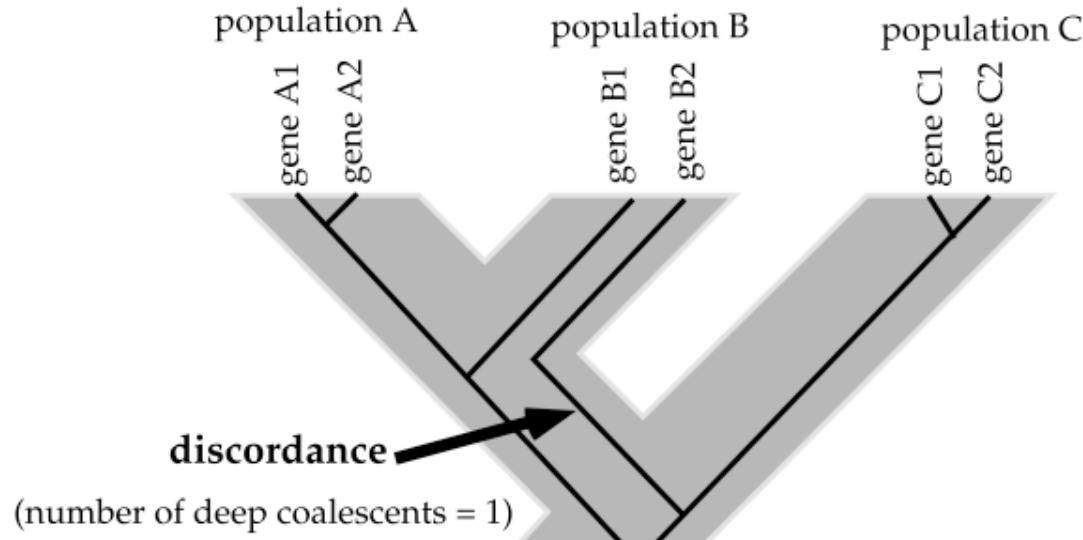
Templeton et  
al. (1995)

- interior clades significantly small or is the  $I-T D_c$  distances significantly large?
- NO: go to step 11.
  - YES: go to step 3.
  - Tip/interior status cannot be determined—Inconclusive Outcome.
3. Are any  $D_n$  and/or  $I-T D_n$  values significantly reversed from the  $D_c$  values, and/or do one or more tip clades show significantly large  $D_n$ s or interior clades significantly small  $D_n$ s or  $I-T$  significantly small  $D_n$  with the corresponding  $D_c$  values being nonsignificant?
- NO: go to step 4.
  - YES: go to step 5.
4. Do the clades (or 2 or more subsets of them) with restricted geographical distributions have ranges that are completely or mostly nonoverlapping with the other clades in the nested group (particularly interiors), and does the pattern of restricted ranges represent a break or reversal from lower level trends within the nested series (applicable to higher-level clades only)?
- NO: restricted gene flow with isolation by distance (restricted dispersal by distance in nonsexual species). This inference is strengthened if the clades with restricted distributions are found in diverse locations, if the union of their ranges roughly corresponds to the range of one or more clades (usually interiors) within the same nested group (applicable only to nesting clades with many clade members or to the highest-level clades regardless of number), and if the  $D_c$  values increase and become more geographically widespread with increasing clade level within a nested series (applicable to lower level clades only).
  - YES: go to step 9.
5. Do the clades (or 2 or more subsets of them) with restricted geographical distributions have ranges that are completely or mostly nonoverlapping with the other clades in the nested group (particularly interiors), and does the pattern of restricted ranges represent a break or reversal from lower-level trends within the nested series (applicable to higher-level clades only)?
- NO: go to step 6.
  - YES: go to step 15.
6. Do clades (or haplotypes within them) with significant reversals or significant  $D_n$  values without significant  $D_c$  values define geographically concordant subsets, or are they geographically concordant with other haplotypes/clades showing similar distance patterns?
- NO: go to step 7.
  - YES: go to step 13.
- Too few clades (<2) to determine concordance—Insufficient genetic resolution to discriminate between range expansion/colinization and restricted dispersal/gene flow: proceed to step 7 to determine if the geographical sampling is sufficient to discriminate between short vs. long distance movement.
- Are the clades with significantly large  $D_n$ s (or tip clades in general when  $D_n$  for  $I-T$  is significantly small) separated from the other clades by intermediate geographical areas that were sampled?
- NO: go to step 8.
  - YES: restricted gene flow/dispersal but with some long-distance dispersal.
- Is the species absent in the nonsampled areas?
- NO: sampling design inadequate to discriminate between isolation by distance (short distance movements) vs. long distance dispersal.
  - YES: restricted gene flow/dispersal but with some long-distance dispersal over intermediate areas not occupied by the species.
- Are the different geographically concordant clade ranges separated by areas that have not been sampled?
- NO: past fragmentation. (If inferred at a high clade level, additional confirmation occurs if the clades displaying restricted but at least partially nonoverlapping geographical distributions are mutationally connected to one another by a larger than average number of steps.)
  - YES: go to step 10.
- Is the species absent in the nonsampled areas?
- NO: geographical sampling scheme inadequate to discriminate between fragmentation and isolation by distance.
  - YES: allopatric fragmentation. (If inferred at a high clade level, additional confirmation occurs if the clades displaying restricted but at least partially nonoverlapping geographical distributions are mutationally connected to one another by a larger than average number of steps.)
- Are the  $D_n$  values for tip or some (but not all) interior clades significantly small, or the  $D_n$  for one or more interior clades significantly large, or is the  $I-T D_n$  value significantly large?
- NO: inconclusive outcome.
  - YES: go to step 4.
- Are the clades found in the different geographical locations separated by a branch length with a larger than average number of mutational steps?
- NO: geographical sampling scheme inadequate to discriminate between fragmentation, range expansion, and isolation by distance.
  - YES: geographical sampling scheme inadequate to discriminate between fragmentation and isolation by distance.

## APPENDIX

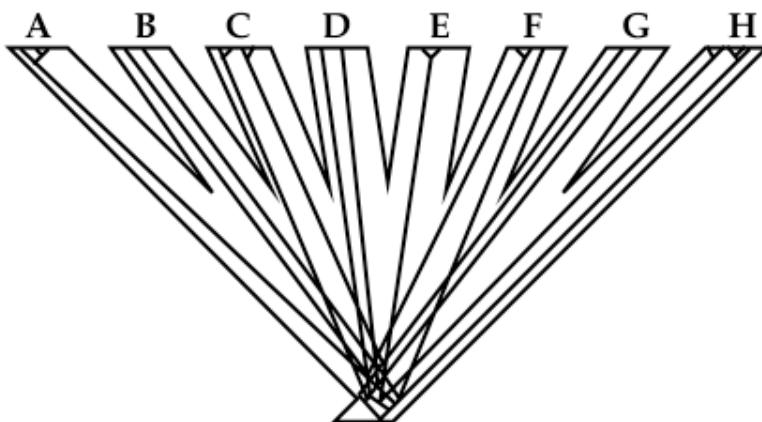
Start with haplotypes nested within a 1-step clade:

1. Are there any significant values for  $D_c$ ,  $D_n$ , or  $I-T$  within the clade?
  - NO: the null hypothesis of no geographical association of haplotypes cannot be rejected (either panmixia in sexual populations, extensive dispersal in nonsexual populations, small sample size, or inadequate geographical sampling). Move on to another clade at the same or higher level.
  - YES: go to step 2.
2. Are the  $D_c$  values for tip or some (but not all)

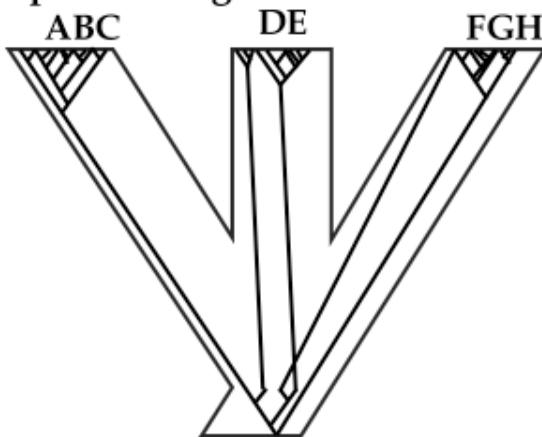


**Fig. 2** Schematic representations of a gene tree contained within a population tree; an arrow identifies the discordance between the gene and population trees. This discord can easily be quantified by counting the number of deep coalescents (Maddison 1997), or the failure of gene lineages to coalesce within their respective population lineages (assuming that there is no gene flow among the populations).

**(a) fragmentation model**



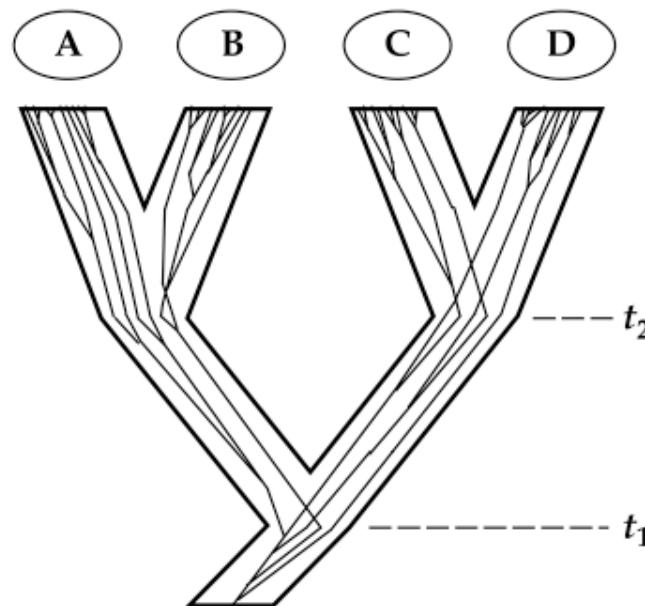
**(b) allopatric refuge model**



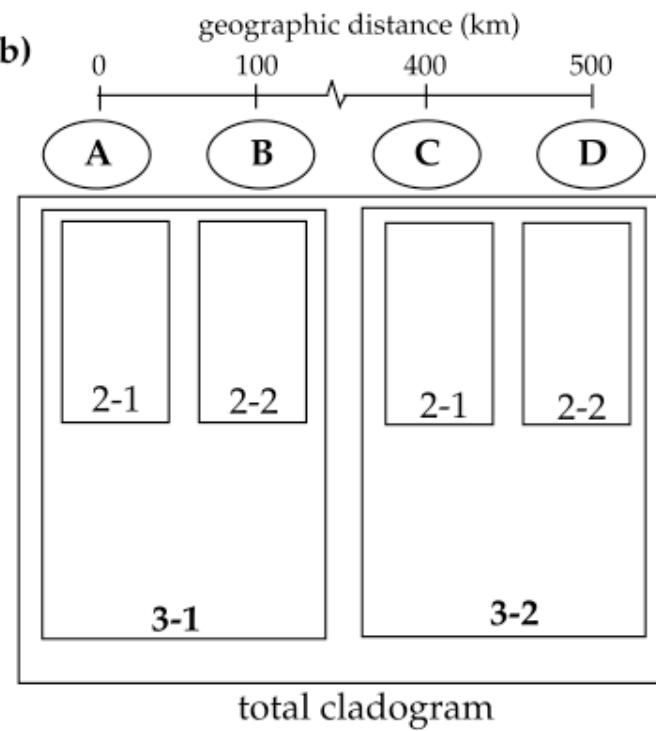
**Fig. 3** The two phylogeographic hypotheses Knowles (2001b) tested with coalescent simulations to determine if contemporary sky-island populations of montane grasshoppers were derived from (a) a single ancestral population as represented by a fragmentation model, or (b) multiple ancestral populations, using an allopatric refuge model (groupings of populations were identified from previous phylogeographic analyses (see Figs 3 and 4, Knowles 2001b)).

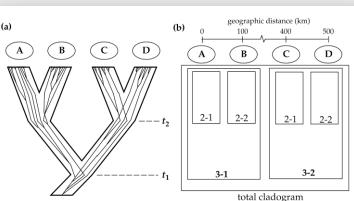
Knowles and Maddison  
(2002)

(a)



(b)





**Table 2a** Results from permutation tests of the null hypothesis that the clades are randomly distributed geographically relative to other clades within the same nesting clade. NCA results are summarized across the 10 data sets and reported separately for each of the nesting clades

NCA results	3-step clades* nested in entire cladogram	2-step clades† nested within clades 3-1 & 3-2	1-step clades‡ nested within clades 2-1 & 2-2
significant geographical association detected	90%	65%	70%
clade-groupings accurate	30%	35%	na

**Table 2b** Results from inference key summarized across the 10 data sets and reported separately for each of the nesting clades. The first value represents the percentage of the 10 data sets in which a particular inference was made, whereas the one in parentheses was calculated using only the subset of the data for which significant population structure was identified by the permutation tests

results from inference key	3-step clades* nested in entire cladogram	2-step clades† nested within clades 3-1 & 3-2	1-step clades‡ nested within clades 2-1 & 2-2
allopatric divergence inferred	0% (0%)	0.5% (7.7%)	7.7% (14.3%)
recurrent force inferred (e.g. gene flow)	20% (22%)	25% (38.5%)	15.4% (28.6%)
historical process other than allopatric divergence inferred	80% (78%)	74.5% (53.8%)	76.9% (57.1%)

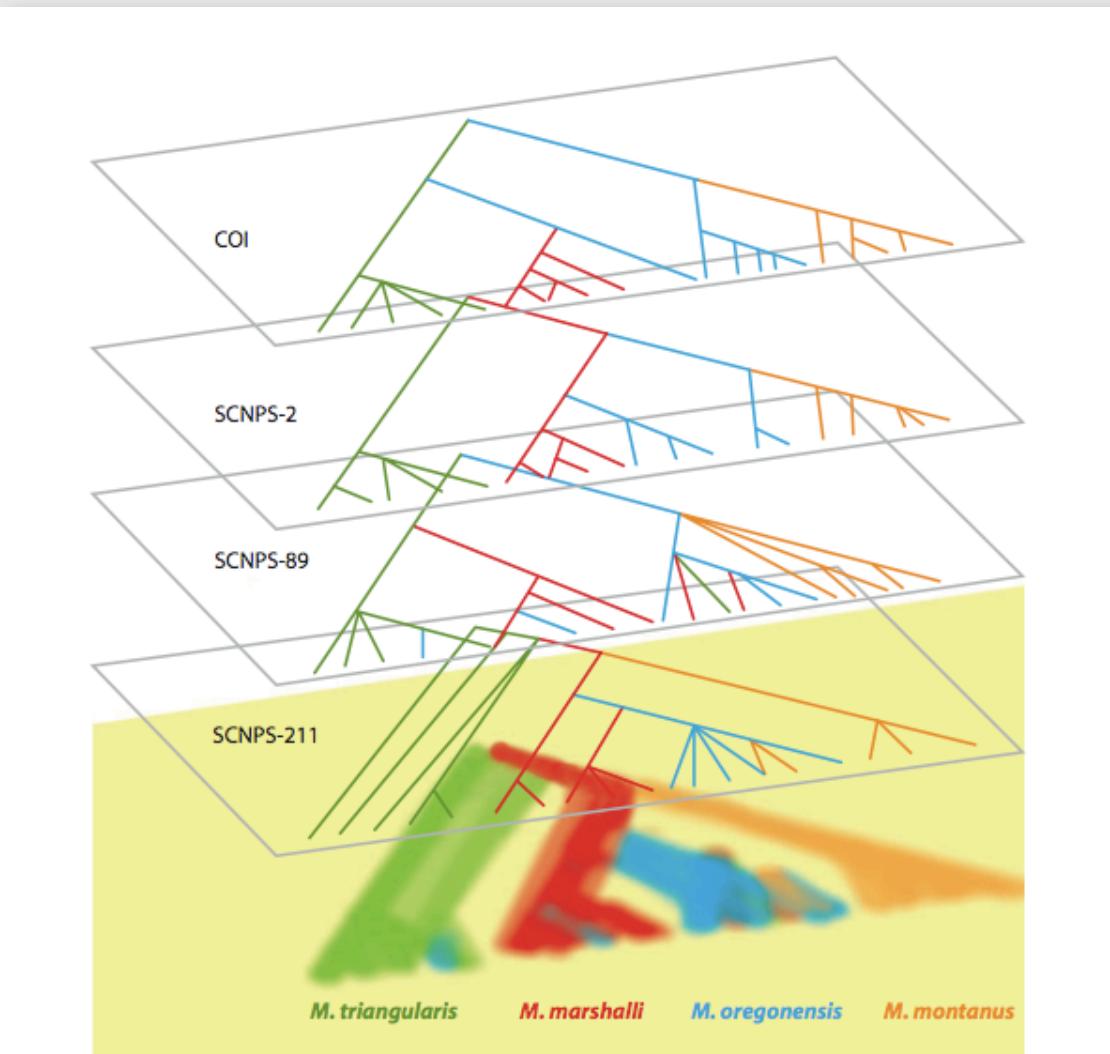
\*corresponds to the two major lineages defined by the initial vicariance at  $t_1$

†corresponds to the two subgroups within each of the major lineages defined by the subsequent vicariance at  $t_2$

‡there is no geographical structure in the model used to simulate the data that corresponds to these clades.

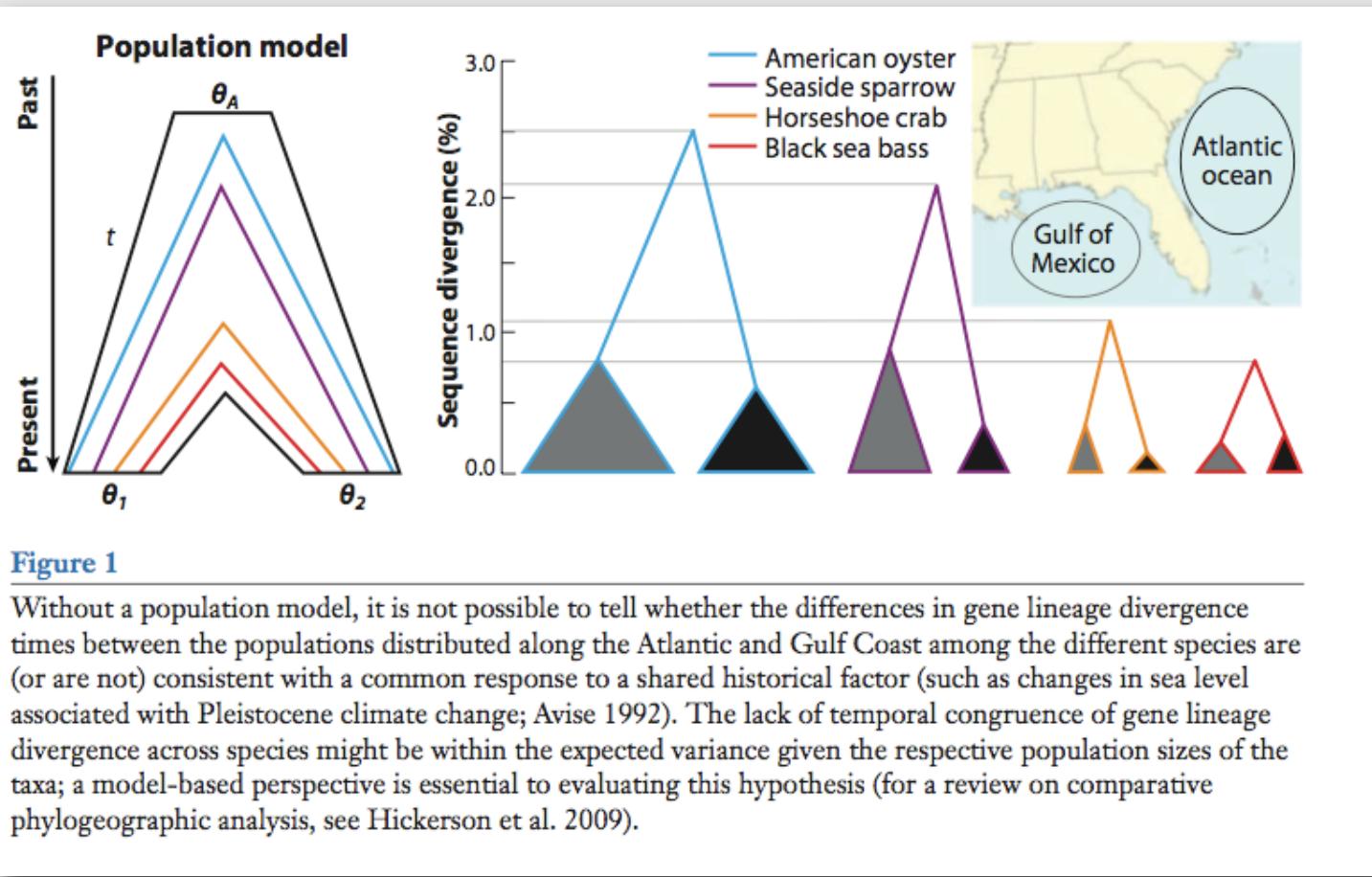
## Abstract

While studies of phylogeography and speciation in the past have largely focused on the documentation or detection of significant patterns of population genetic structure, the emerging field of statistical phylogeography aims to infer the history and processes underlying that structure, and to provide objective, rather than *ad hoc* explanations. Methods for parameter estimation are now commonly used to make inferences about demographic past. Although these approaches are well developed statistically, they typically pay little attention to geographical history. In contrast, methods that seek to reconstruct phylogeographic history are able to consider many alternative geographical scenarios, but are primarily nonstatistical, making inferences about particular biological processes without explicit reference to stochastically derived expectations. We advocate the merging of these two traditions so that statistical phylogeographic methods can provide an accurate representation of the past, consider a diverse array of processes, and yet yield a statistical estimate of that history. We discuss various conceptual issues associated with statistical phylogeographic inferences, considering especially the stochasticity of population genetic processes and assessing the confidence of phylogeographic conclusions. To this end, we present some empirical examples that utilize a statistical phylogeographic approach, and then by contrasting results from a coalescent-based approach to those from Templeton's nested cladistic analysis (NCA), we illustrate the importance of assessing error. Because NCA does not assess error in its inferences about historical processes or contemporary gene flow, we performed a small-scale study using simulated data to examine how our conclusions might be affected by such unconsidered errors. NCA did not identify the processes used to simulate the data, confusing among deterministic processes and the stochastic sorting of gene lineages. There is as yet insufficient justification of NCA's ability to accurately infer or distinguish among alternative processes. We close with a discussion of some unresolved problems of current statistical phylogeographic methods to propose areas in need of future development.



**Figure 5**

Rather than focusing on the idiosyncrasies of the tree topologies of individual loci, where coalescent and mutational variance can give rise to gene tree discord, as shown here for a sample of independent loci sequenced in montane grasshoppers (from Carstens & Knowles 2007), new coalescent-based approaches provide direct estimates of the underlying history of divergence—the actual species tree. Extracting the common historical signal of species divergence in the montane grasshoppers from the gene trees of the independent loci, despite widespread incomplete lineage sorting, highlights how this advance provides resolution of the traditionally unresolvable, estimating the pattern of species splitting during recent evolutionary radiations.



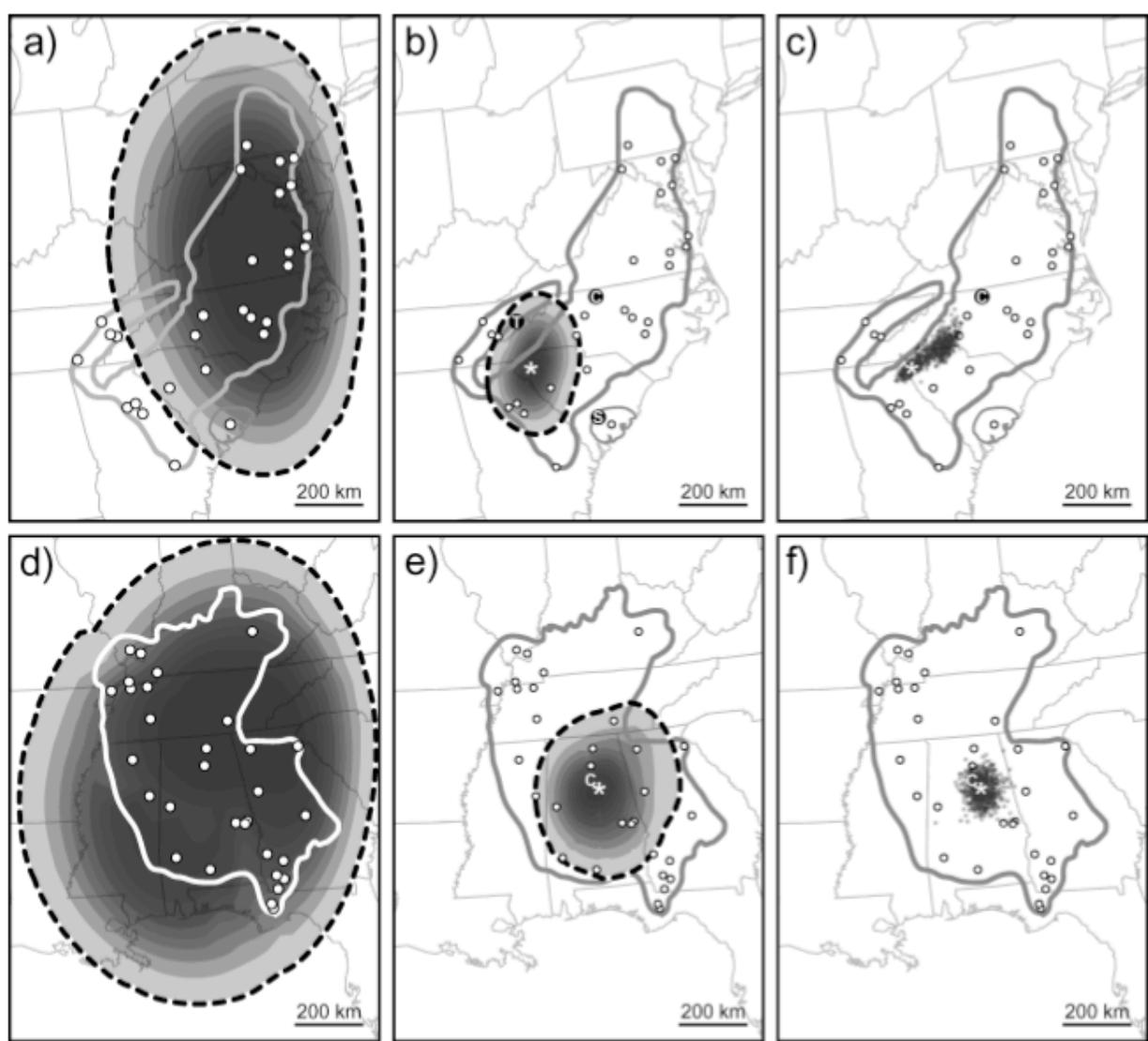


FIGURE 2. Estimates of the geographic locations for the ancestors of coastal (a–c) and inland (d–f) *Pseudacris feriarum*. Small open circles indicate the sampling locations. In a and d, we present an expected likelihood surface based on the sampling localities and an assumption of no phylogeographic association. To obtain this null expectation, we randomized the assignment of geographic location to the tips of the genealogy. The likelihood surface is represented by a series of filled contours, where the darkness of a contour corresponds to the likelihood obtained when the ancestor was constrained to be at a location inside the contour (with all other parameters optimized). The dashed oval delineates the 95% confidence envelope (all locations outside this envelope are significantly worse than the maximum likelihood location in a likelihood-ratio test). Other contours delineate the 85% confidence envelope, 75% confidence envelope, etc. In b and e, we present the likelihood surfaces for the observed phylogeographic association (assignment of geographic locations *not* randomized). Stars indicate the maximum likelihood estimate for the location of the basal ancestor of each clade. The points labeled C, S, and T indicate the estimated center of the range, the South Carolina refugium, and the Tennessee Valley refugium, respectively. In c and f, we present the effect of phylogenetic uncertainty on the estimates of the ancestral geographic location. Each black point indicates the maximum likelihood ancestral location estimated using 1 of 1000 trees sampled from the posterior probability distribution (Lemmon et al., 2007b).

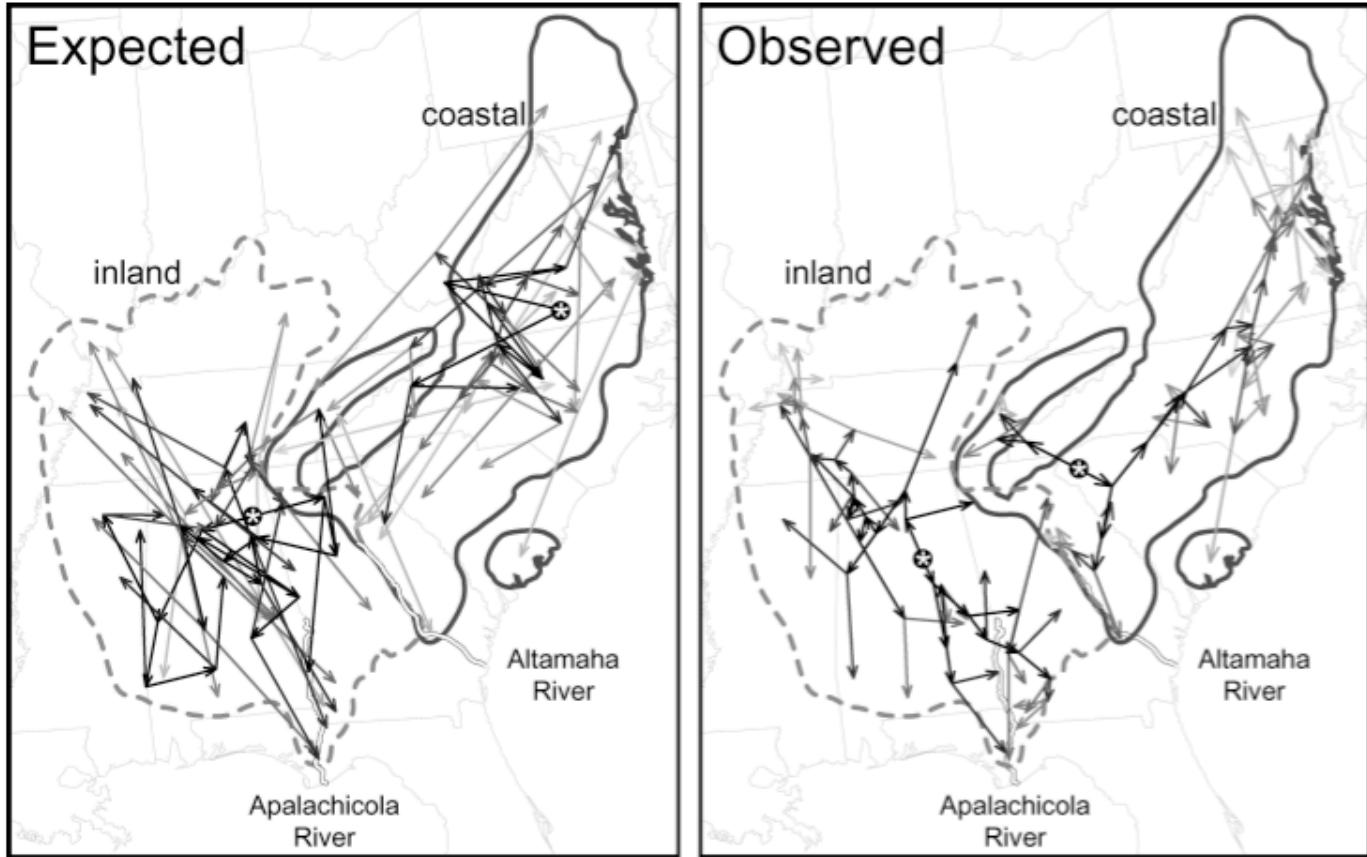
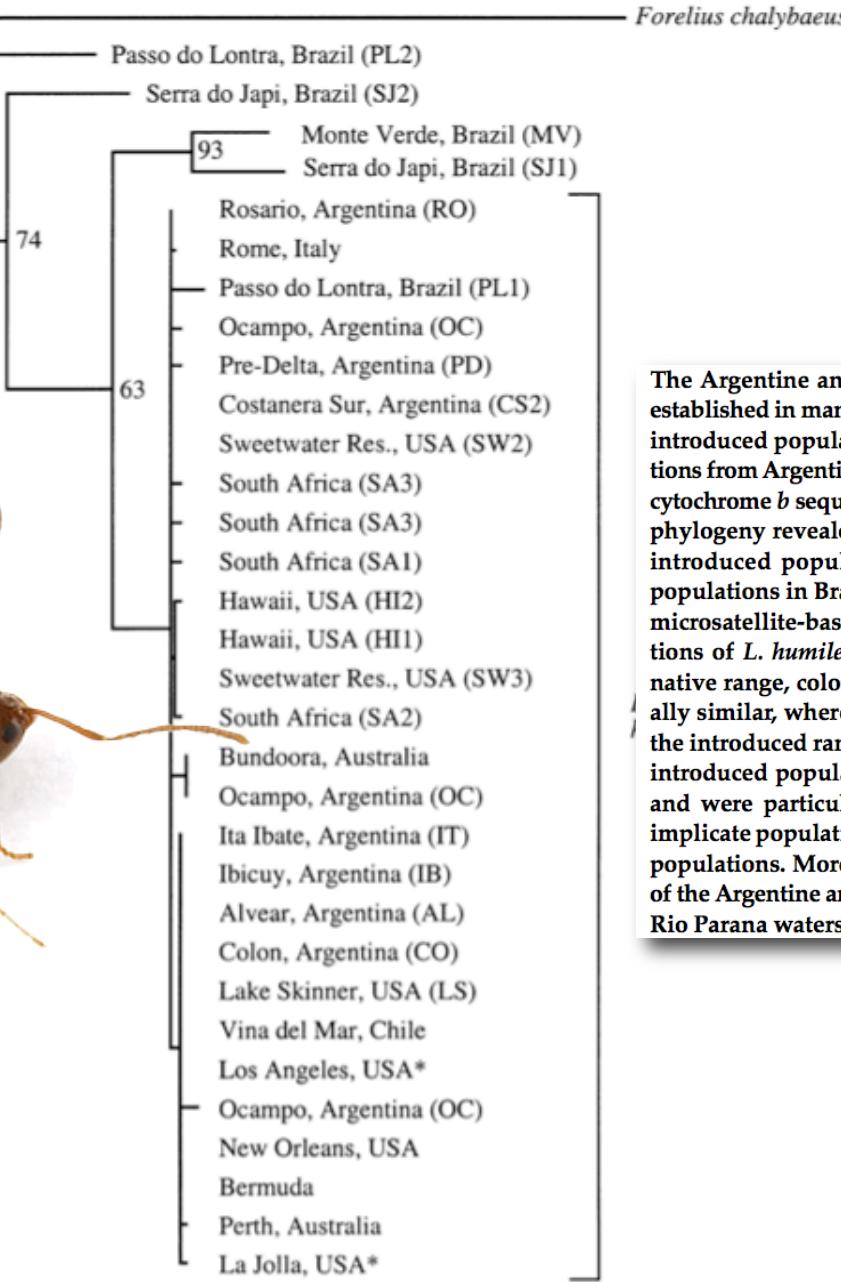


FIGURE 3. Estimated migration routes for ancestors of inland and coastal *Pseudacris feriarum*. Each arrow points from the estimated location of one ancestor (represented by an internal node on the genealogy) to the estimated location of one of its descendants. Arrows corresponding to basal branches are darker than derived branches. Stars denote the location of the most basal ancestor for each clade. The panel labeled "Expected" presents an expected pattern of migration under the assumption of no phylogeographic association. The expected ancestral migration patterns were estimated after randomizing the assignment of geographic location to the tips of the genealogy. Note that we present the results of one random replicate (e.g., location of star varies across random replicates). The panel labeled "Observed" presents the maximum-likelihood migration routes given the observed phylogeographic association (assignment of location was not randomized). Note that coastal *P. feriarum* shows a pattern of northward migration.



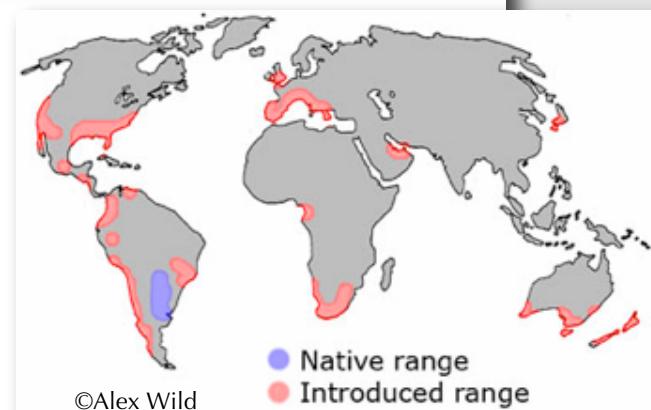
Photo ©Alex Wild,  
Myrmecos.net



Tsutsui et al. (2001)

**Fig. 2** Maximum likelihood phylogeny of native and introduced Argentine ant populations using 405 bp of the mitochondrial cytochrome *b* gene. Numbers indicate the reliability of each node as determined by the percentage of time each corresponding cluster was formed during 10 000 quartet puzzling steps (Strimmer & von Haeseler 1996). Sites that belong to the large California supercolony are designated with an asterisk

The Argentine ant (*Linepithema humile*) is a damaging invasive species that has become established in many Mediterranean-type ecosystems worldwide. To identify likely sources of introduced populations we examined the relationships among native *Linepithema* populations from Argentina and Brazil and introduced populations of *L. humile* using mitochondrial cytochrome *b* sequence data and nuclear microsatellite allele frequencies. The mitochondrial phylogeny revealed that the populations in Brazil were only distantly related to both the introduced populations and the native populations in Argentina, and confirmed that populations in Brazil, previously identified as *L. humile*, are likely a different species. The microsatellite-based analysis provided resolution among native and introduced populations of *L. humile* that could not be resolved using the mitochondrial sequences. In the native range, colonies that were geographically close to one another tended to be genetically similar, whereas more distant colonies were genetically different. Most samples from the introduced range were genetically similar, although some exceptions were noted. Most introduced populations were similar to native populations from the southern Rio Parana and were particularly similar to a population from Rosario, Argentina. These findings implicate populations from the southern Rio Parana as the most likely source of introduced populations. Moreover, these data suggest that current efforts to identify natural enemies of the Argentine ant for biological control should focus on native populations in the southern Rio Parana watershed.



©Alex Wild