Using GMYC for species delineation

- GMYC (which stands for Generalized Mixed Yule Coalescent) is a method that can be used to delineate species using sequence data.
- A likelihood method that determines the point of transition from species-level (speciation and extinction) to population-level (coalescence) evolutionary processes.

- It uses an ultrametric tree and attempts to detect the transition in the tree where the branching pattern switches from being attributed to speciation (one lineage per species) to when it can be attributed to the intra-species coalescent process (multiple lineages per species).

- the model used to express the expected branching pattern on the tree. We will either use:
- a Yule model (also known as pure birth) in which all branching in the tree can be explained by a constant speciation rate.
- a Coalescent model with constant population size. This is typically the prior the most adapted to model the relationships among individuals from the same species.

- The basic GMYC model assumes that all species have the same coalescent branching rates and scaling parameters.
- Therefore, for this method to work, the tree provided needs to be
- (1) fully resolved (without any mutlifurcations);
- (2) ultrametric (all the tips have the same age).

EXAMPLE

- Here,
- we use mitochondrial DNA variation to delimit species in a poorly known beetle radiation in the genus Rivacindela from Australia. Among 468 individuals sampled from 65 sites and multiple morphologically distinguishable types, sequence variation in three mtDNA genes (cytochrome oxidase subunit 1, cytochrome b, 16S ribosomal RNA) was strongly partitioned between 46 or 47 putative species identified with quantitative methods of species recognition based on fixed unique ("diagnostic") characters. The boundaries between groups were also recognizable from a striking increase in branching rate in clock-constrained calibrated trees.

1.MATERIAL AND METHODS

Collecting Information and DNA Procedures

2.Species Delimitation

3.Analyses of mtDNA Branching Times

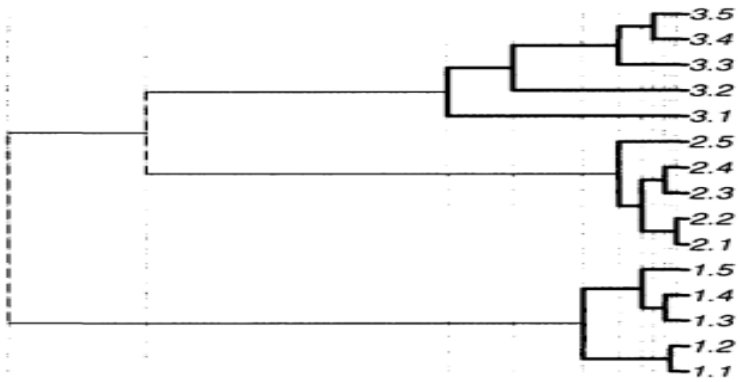
using only individuals with complete sequence information

Statistical parsimony analysis partitions the data into independent networks of haplotypes connected by changes that are non-homoplastic with a 95% probability

Intragroup and pairwise intergroup divergences were calculated with Mega version 2.1

Fst and P values at 0.05 significance were calculated from pairwise differences in Arlequin 2.0

Absolute ages were estimated setting the split of R. aurifodina (sample 204) and R. salicursoria (sample 208) to 3.2 Mya



Waiting intervals	x1	x2	x3	x4	x5	etc.
Diversification (n)	2	3	2	2	1	etc.
Species 1 (n)	0	0	2	3	3	etc.
Species 2 (n)	0	0	0	0	0	etc.
Species 3 (n)	0	0	0	0	2	etc.

Branches are categorized as either between species (thin lines) or within species branching (bold lines)

The overall aim of the procedure is to classify the observed branching time intervals defined by the nodes in a clock-constrained phylogram to either being the result of inter-specific ("diversification") or intraspecific ("coalescent") processes of lineage branching

- A key step is fitting the location of the switches from speciation to coalescent nodes,the most recent common ancestral node defining each species.

Assume that there is a threshold time, T, before which all nodes reflect diversification events and after which all nodes reflect coalescent events.

1994, 2001; Nee et al., 1994). Under a neutral coalescent, the likelihoods of the waiting times within a single population with effective population size Ne and n_i lineages present during waiting time i are given by:

$$L_{(x_i)} = \lambda n_i (n_i - 1) e^{-\lambda n_i (n_i - 1) x_i} \quad (1)$$

where the birth rate

$$\lambda = \frac{1}{2N_e} \quad (2)$$

The simplest standard approach for considering branching between species is as a Yule model (Yule, 1924); i.e., a stochastic birth-only model. The likelihoods of the waiting times in a species phylogeny (one tip per species) of a clade with constant average speciation rate, λ, and no extinction are given by:

$$L_{(x_i)} = \lambda n_i e^{-\lambda n_i x_i} \quad (3)$$

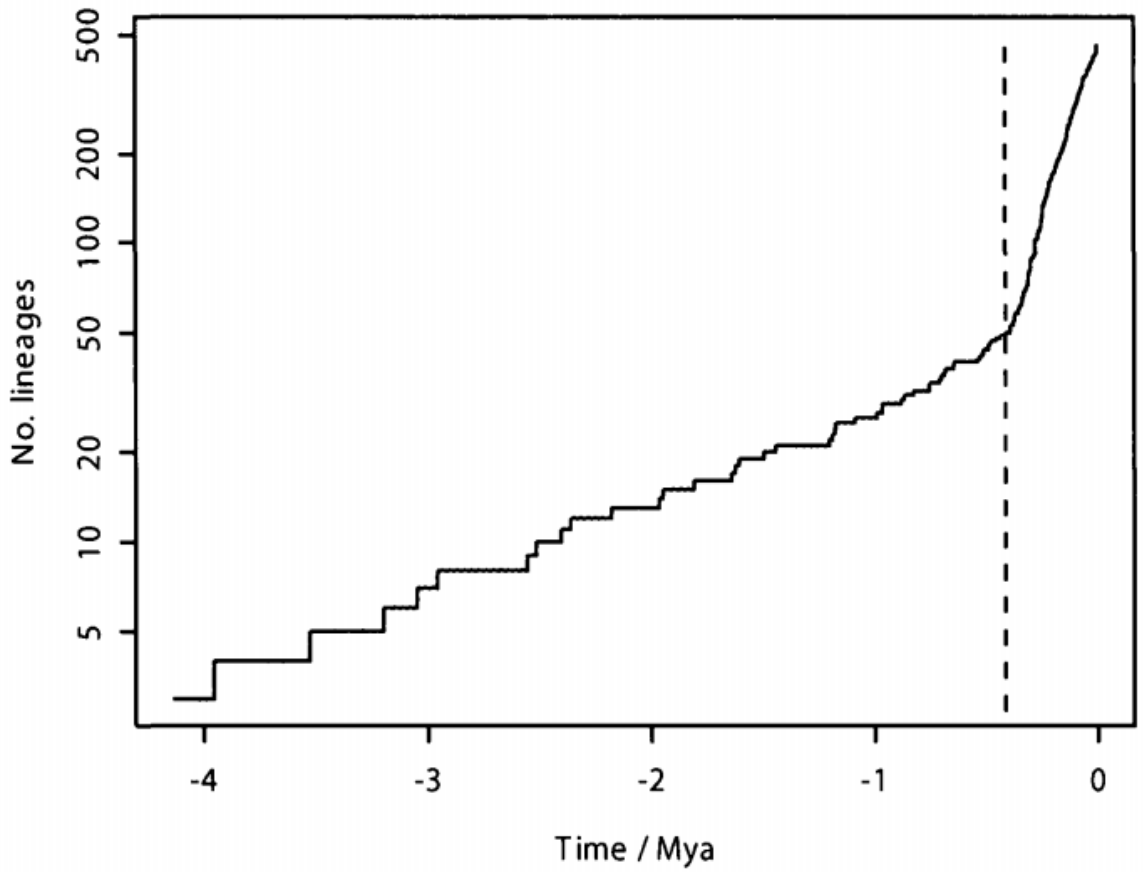
We combine the above equations describing population and speciation processes to consider a clade that has diversified into k species, each of which can be treated as a single population with effective size, N_j, $j = 1$ to k. Assuming a constant speciation rate without extinction, and neutral coalescence within each species, the likelihoods of waiting times in the entire tree under this mixed Yule coalescent (MYC) model are given by:

$$L_{(x_i)} = be^{-bx_i} \tag{4}$$

where

$$b = \lambda_{k+1} n_{i,k+1} + \sum_{j=1,k} (\lambda_j n_{i,j}(n_{i,j} - 1)) \tag{5}$$

where $k + 1$ is the index assigned to the diversification process and $n_{i,j}$ is the number of lineages in waiting interval i belonging to process j. λ_{k+1} is the speciation rate and λ_j are the branching rates for each coalescent process as defined in Equation 2. We do not consider effective population sizes explicitly hereafter, rather the coalescent branching rate parameter for each species. The term b is the probability that an event of any type happens at the end of the waiting interval and e^{-bx_i} is the probability that no event happened during the waiting interval (see appendix 1 of Nee, 2001).



The sharp increase in branching rate, corresponding to the transition from interspecies to intraspecies branching events.

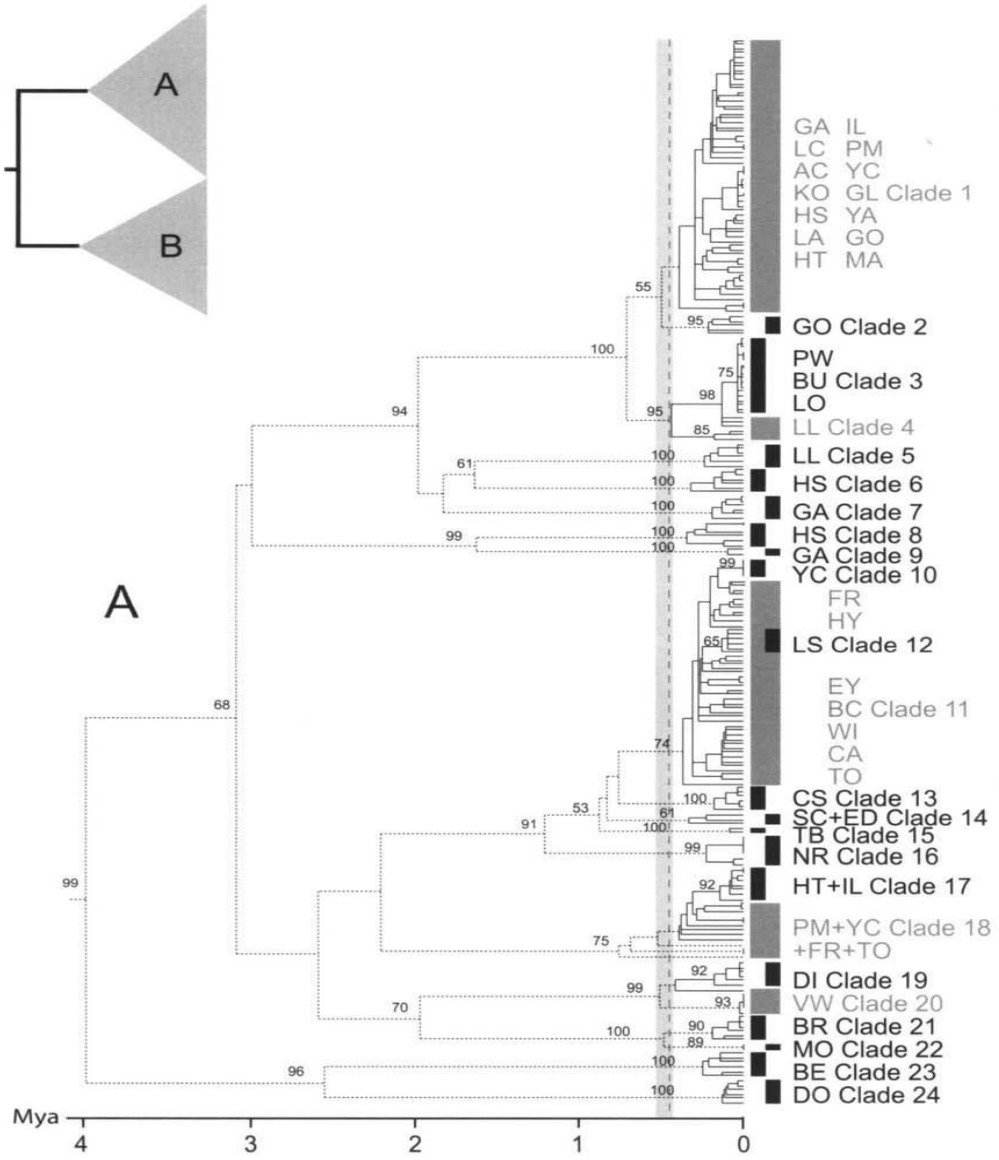


FIGURE 3. ML tree depicting relationships of *Rivacindela* mtDNA haplotypes with branch lengths fitted assuming a molecular clock. The two main sister groups are shown separately (panel A, Eastern group; panel B, Western group). Localities for each of 47 putative species ("Clades") obtained under the WP method are indicated by a two-letter code. Different shading is used for better visibility of the extent of groups. Gray bars and site names colours indicate widely distributed or paraphyletic species. The line style indicates whether branches were estimated as between-species branching (stippled) or within-species branching (solid) in the likelihood procedure. The dotted vertical line shows the maximum likelihood transition point of the switch in branching rates. The grey shading indicates the confidence limits for the transition point falling within 2 log-likelihood units of the ML solution (Material and Methods). Numbers above nodes represent bootstrap support values based on 100 pseudoreplicates and ratchet parsimony searches, shown only for the deep level clades. Bootstrap values under 50% are not shown. *(Continued)*

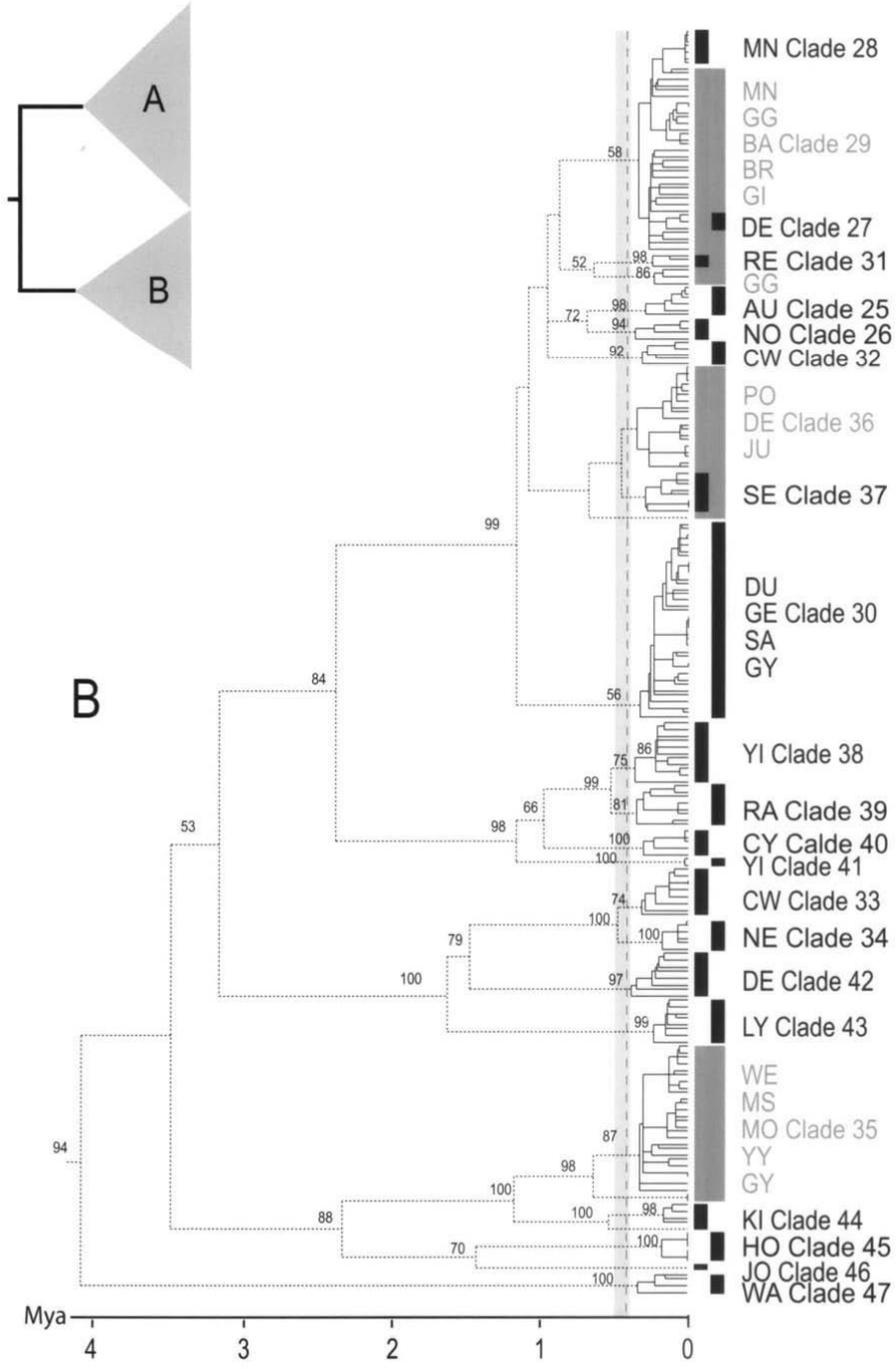


FIGURE 3. (Continued)

over/under estimate

Whether these methods actually produce differing estimates of numbers of groups will depend on the relative separation of species versus within-species coalescence times.

If many related species diverged more recently than the oldest coalescence times of the alleles they contain, the methods will underestimate species numbers, delimiting species complexes rather than individual species.

And if the number of the samples of each coalescence is too small,we will conduct a misdiagnosis of different spieces .

PROS

- haplotype
- Single locus
- Independent evolution
- All of the above makes it simple and convenient
- suitable for the delimitation of the large number of undescribed spieces