Tracking the formation of a species assemblage over time: phylogenetic reconstruction of patterns of colonisation and speciation

Xia Hua, Rob Lanfear, Marcel Cardillo, Lindell Bromham
Division of Evolution, Ecology and Genetics
Australian National University

Presented by Xia Hua, 11-08-13

Are species added into an assemblage at an even rate over time?

Are species added into an assemblage at an even rate over time?

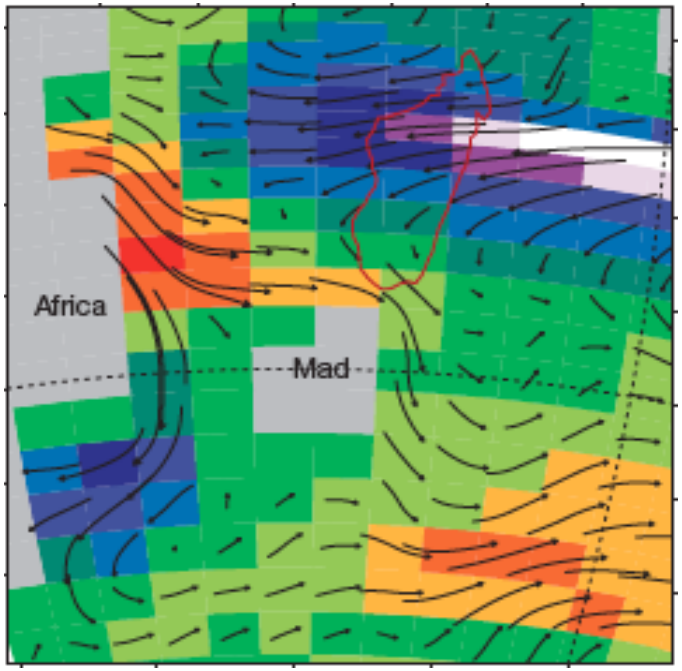
A known set of extant taxa
that co-exist in a given area

Are species added into an assemblage at an even rate over time?

A known set of extant taxa that co-exist in a given area

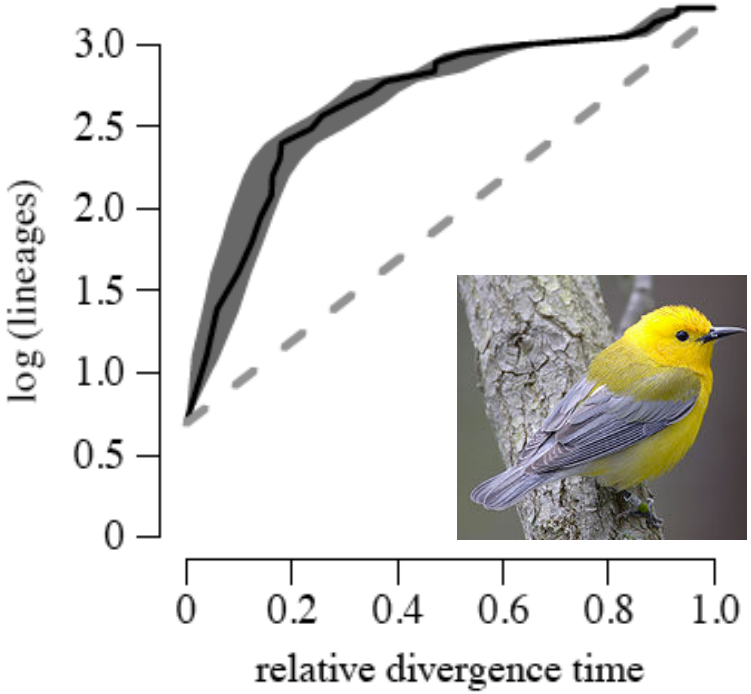
Colonisation **In-situ speciation**

Are species added into an assemblage at an even rate over time?



Uneven colonization rate due to geographical events

Ali & Huber 2010 Nature



Slow-down in speciation rate due to saturation

Rabosky & Lovette 2008 Proc.R.Soc.B

Approaches to reconstruct patterns of speciation and colonisation

Approaches to reconstruct patterns of speciation and colonisation

Use paleontological data

Need continuous fossil record and adequate taxonomic and temporal resolution

Approaches to reconstruct patterns of speciation and colonisation

Use paleontological data

Need continuous fossil record and adequate taxonomic and temporal resolution

Compare assemblages of known ages

E.g., derive assemblage ages from the formation dates of different islands

Approaches to reconstruct patterns of speciation and colonisation

Use paleontological data

Need continuous fossil record and adequate taxonomic and temporal resolution

Compare assemblages of known ages

E.g., derive assemblage ages from the formation dates of different islands

Use molecular phylogenies

Sources of error in using molecular phylogenies to infer assemblage history

Sources of error in using molecular phylogenies to infer assemblage history

Uncertainty in topology and divergence times

A “best” tree is typically used in macroecology study

Sources of error in using molecular phylogenies to infer assemblage history

Uncertainty in topology and divergence times

A “best” tree is typically used in macroecology study

Uncertainty in reconstruct ancestral geographic states, particularly taxon sampling bias

More recent events are more likely to be detected

Sources of error in using molecular phylogenies to infer assemblage history

Uncertainty in topology and divergence times

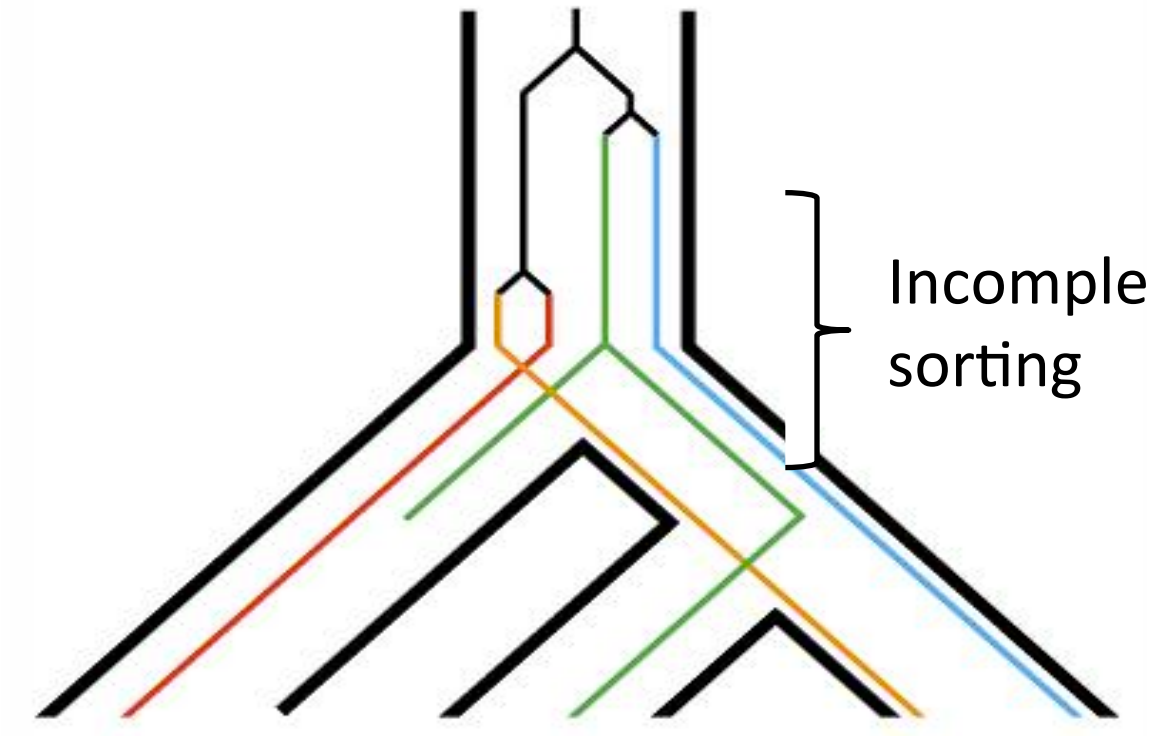
A “best” tree is typically used in macroecology study

Uncertainty in reconstruct ancestral geographic states, particularly taxon sampling bias

More recent events are more likely to be detected

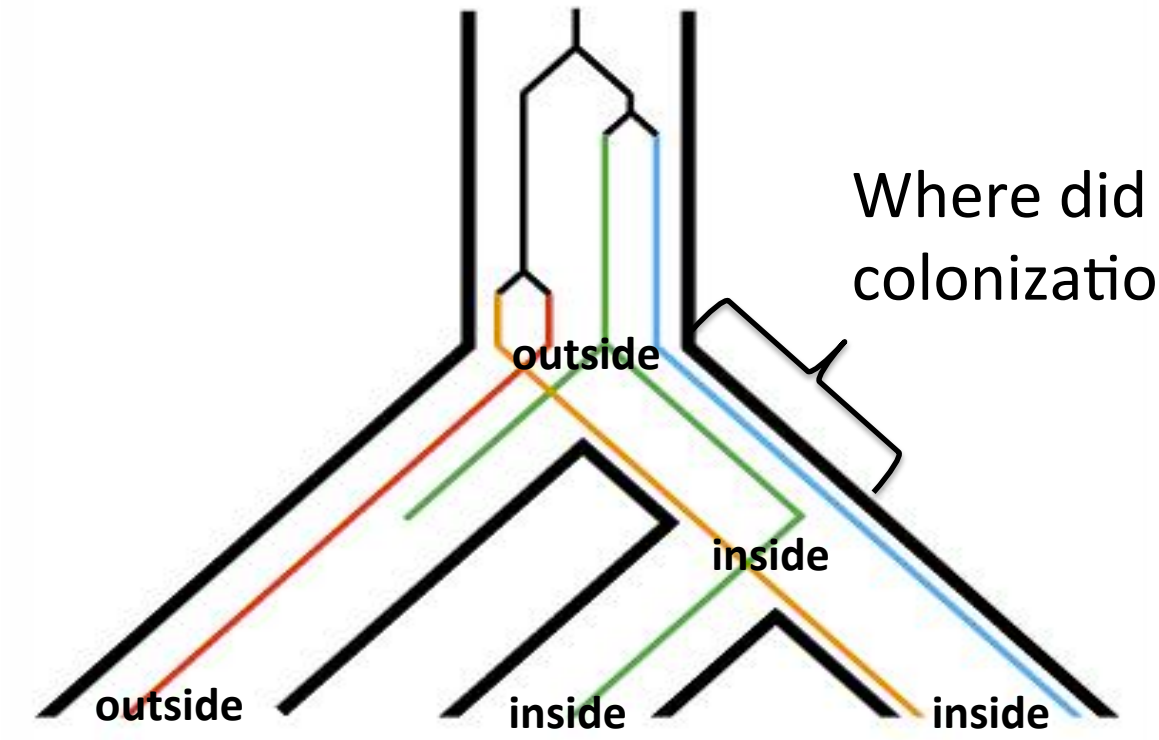
Uncertainty in localize speciation and colonization events on a branch

Sources of error in using molecular phylogenies to infer assemblage history



Uncertainty in localize speciation and colonization events on a node

Sources of error in using molecular phylogenies to infer assemblage history



Uncertainty in localize speciation and colonization events on a node

Framework of testing assemblage formation using phylogenies

How likely it is to observe the extant species assemblage if it was formed in the way as the null hypothesis predicts ?

$$L(\text{assemblage} \mid \mathrm{H}_0) =$$

$$\sum_j \sum_i P(\text{assemblage} \mid \text{history } i, \mathrm{H}_0)$$

$$\times P(\text{history } i \mid \text{phylogeny } j, \mathrm{H}_0)$$

$$\times P(\text{phylogeny } j \mid \mathrm{H}_0),$$

Framework of testing assemblage formation using phylogenies

How likely it is to observe the extant species assemblage if it was formed in the way as the null hypothesis predicts ?

$$L(\text{assemblage} \mid \text{H}_0) =$$

$$\sum_j \sum_i P(\text{assemblage} \mid \text{history } i, \text{H}_0)$$

$$\times P(\text{history } i \mid \text{phylogeny } j, \text{H}_0)$$

$$\times P(\text{phylogeny } j \mid \text{H}_0),$$

Topology and time uncertainty

Framework of testing assemblage formation using phylogenies

How likely it is to observe the extant species assemblage if it was formed in the way as the null hypothesis predicts ?

$$L(\text{assemblage} \mid \mathrm{H}_0) =$$

$$\sum_j \sum_i P(\text{assemblage} \mid \text{history } i, \mathrm{H}_0)$$

$$\times P(\text{history } i \mid \text{phylogeny } j, \mathrm{H}_0)$$ **Reconstruction uncertainty**

$$\times P(\text{phylogeny } j \mid \mathrm{H}_0),$$ **Topology and time uncertainty**

Framework of testing assemblage formation using phylogenies

How likely it is to observe the extant species assemblage if it was formed in the way as the null hypothesis predicts ?

$$L(\text{assemblage} \mid \text{H}_0) =$$

$$\sum_j \sum_i P(\text{assemblage} \mid \text{history } i, \text{H}_0)$$

Uncertainty in localizing colonization and speciation

$$\times P(\text{history } i \mid \text{phylogeny } j, \text{H}_0)$$

Reconstruction uncertainty

$$\times P(\text{phylogeny } j \mid \text{H}_0),$$

Topology and time uncertainty

Framework of testing even species addition rate using phylogenies

$$L(\text{assemblage} \mid \text{speciation } \lambda;\ \text{extinction } \mu;\ \text{colonization } q_{01};\ \text{emigration } q_{10}) =$$

$$\sum_j \sum_i P(\text{assemblage} \mid \text{history } i;\ \lambda, \mu, q_{01}, q_{10})$$

$$\times P(\text{history } i \mid \text{phylogeny } j;\ q_{01}, q_{10})$$

$$\times P(\text{phylogeny } j \mid \lambda, \mu),$$

Framework of testing even species addition rate using phylogenies

$$L(\text{assemblage} \mid \text{speciation } \lambda;\ \text{extinction } \mu;\ \text{colonization } q_{01};\ \text{emigration } q_{10}) =$$

$$\sum_{j}\sum_{i} P(\text{assemblage} \mid \text{history } i;\ \lambda, \mu, q_{01}, q_{10})$$

$$\times P(\text{history } i \mid \text{phylogeny } j;\ q_{01}, q_{10})$$

$$\times P(\text{phylogeny } j \mid \lambda, \mu),$$

Joint estimation of topology and divergence time

Framework of testing even species addition rate using phylogenies

$$L(\text{assemblage} \mid \text{speciation } \lambda;\ \text{extinction } \mu;\ \text{colonization } q_{01};\ \text{emigration } q_{10}) =$$

$$\sum_j \sum_i P(\text{assemblage} \mid \text{history } i;\ \lambda, \mu, q_{01}, q_{10})$$

$$\times P(\text{history } i \mid \text{phylogeny } j;\ q_{01}, q_{10})$$

Continuous-time Markov model

$$\times P(\text{phylogeny } j \mid \lambda, \mu),$$

Joint estimation of topology and divergence time

Framework of testing even species addition rate using phylogenies

$$L(\text{assemblage} \mid \text{speciation } \lambda;\ \text{extinction } \mu;\ \text{colonization } q_{01};\ \text{emigration } q_{10}) =$$

$$\sum_j \sum_i P(\text{assemblage} \mid \text{history } i;\ \lambda, \mu, q_{01}, q_{10})$$ **Birth-death model**

$$\times P(\text{history } i \mid \text{phylogeny } j;\ q_{01}, q_{10})$$ **Continuous-time Markov model**

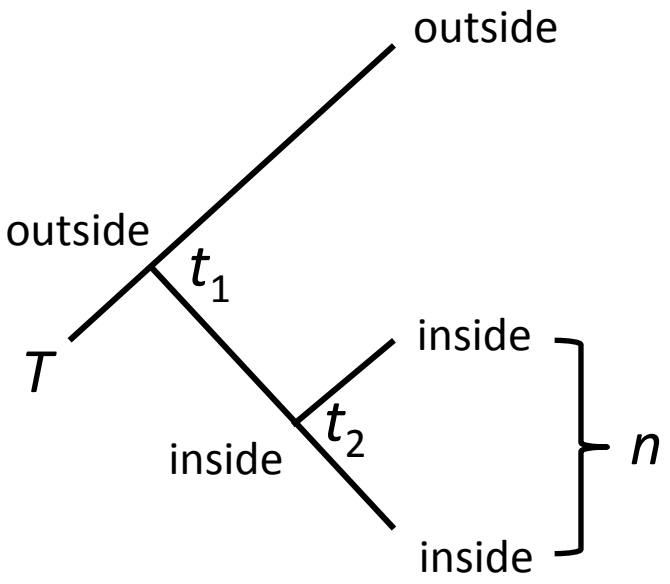
$$\times P(\text{phylogeny } j \mid \lambda, \mu),$$ **Joint estimation of topology and divergence time**

Framework of testing even species addition rate using phylogenies

$$L(\text{assemblage} \mid \text{speciation } \lambda;\ \text{extinction } \mu;\ \text{colonization } q_{01};\ \text{emigration } q_{10}) =$$

$$\sum_j \sum_i P(\text{assemblage} \mid \text{history } i;\ \lambda, \mu, q_{01}, q_{10})$$

Birth-death model

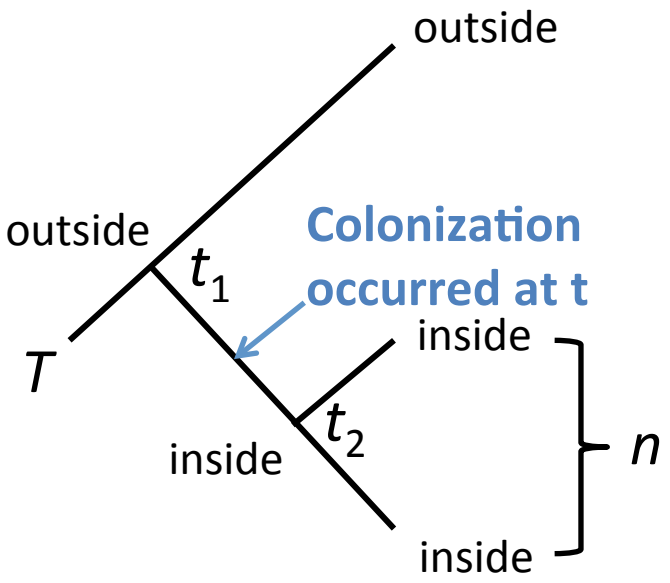


Framework of testing even species addition rate using phylogenies

$$L(\text{assemblage} \mid \text{speciation } \lambda;\ \text{extinction } \mu;\ \text{colonization } q_{01};\ \text{emigration } q_{10}) =$$

$$\sum_j \sum_i P(\text{assemblage} \mid \text{history } i;\ \lambda, \mu, q_{01}, q_{10})$$

Birth-death model

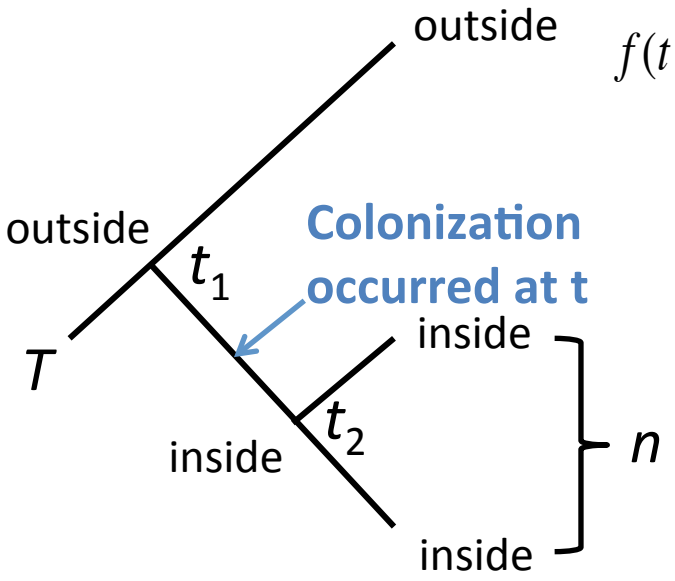


Framework of testing even species addition rate using phylogenies

$$L(\text{assemblage} \mid \text{speciation } \lambda;\ \text{extinction } \mu;\ \text{colonization } q_{01};\ \text{emigration } q_{10}) =$$

$$\sum_j \sum_i P(\text{assemblage} \mid \text{history } i;\ \lambda, \mu, q_{01}, q_{10})$$

Birth-death model



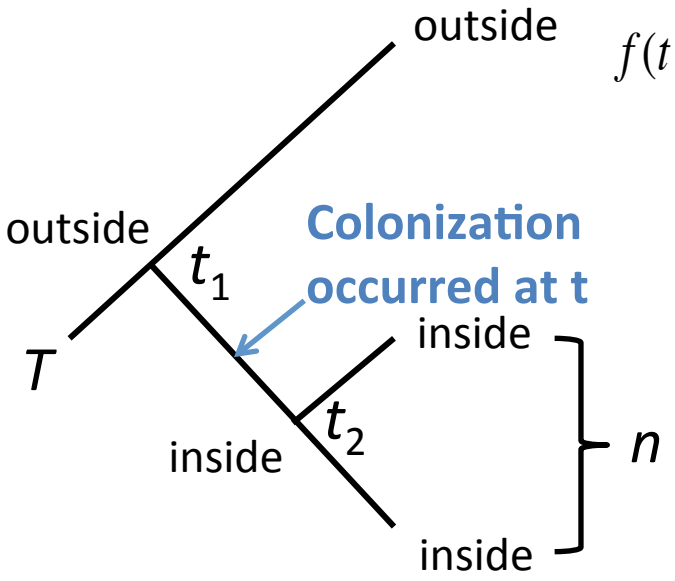
$$f(t \mid n) = \lambda^{n-1}(\lambda - \mu - q_{10})^2 \frac{(1 - e^{-(\lambda-\mu-q_{10})t})^{n-1} e^{-(\lambda-\mu-q_{10})t}}{(\lambda - \mu e^{-(\lambda-\mu-q_{10})t})^{n+1}} \frac{f(t)}{\int_0^T p(t \mid n) f(t)\, dt}$$

Framework of testing even species addition rate using phylogenies

$$L(\text{assemblage} \mid \text{speciation } \lambda;\ \text{extinction } \mu;\ \text{colonization } q_{01};\ \text{emigration } q_{10}) =$$

$$\sum_j \sum_i P(\text{assemblage} \mid \text{history } i;\ \lambda, \mu, q_{01}, q_{10})$$

Birth-death model



$$f(t \mid n) = \lambda^{n-1}(\lambda - \mu - q_{10})^2 \frac{(1 - e^{-(\lambda-\mu-q_{10})t})^{n-1} e^{-(\lambda-\mu-q_{10})t}}{(\lambda - \mu e^{-(\lambda-\mu-q_{10})t})^{n+1}} \frac{f(t)}{\int_0^T p(t \mid n) f(t)\, dt}$$

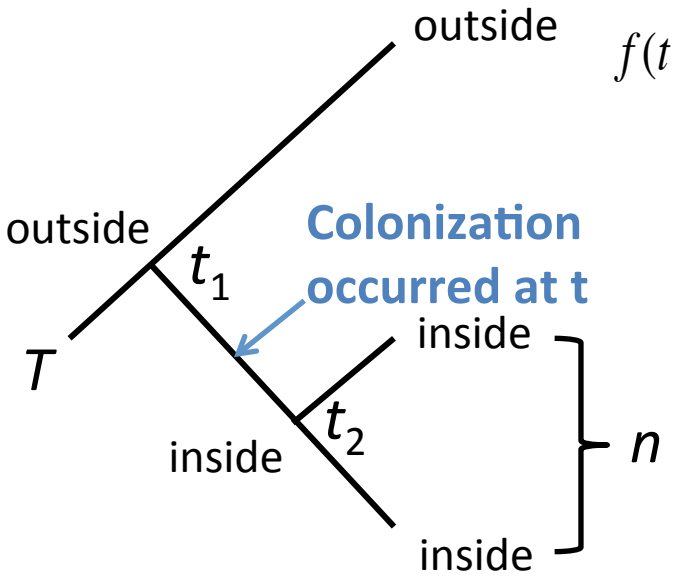
$$f(t) = [(q_{10} + q_{01} e^{(q_{10}+q_{01})(t-T)}) q_{01}] N_0(t)$$

Framework of testing even species addition rate using phylogenies

$$L(\text{assemblage} \mid \text{speciation } \lambda;\ \text{extinction } \mu;\ \text{colonization } q_{01};\ \text{emigration } q_{10}) =$$

$$\sum_j \sum_i P(\text{assemblage} \mid \text{history } i;\ \lambda, \mu, q_{01}, q_{10})$$

Birth-death model



$$f(t \mid n) = \lambda^{n-1}(\lambda - \mu - q_{10})^2 \frac{(1 - e^{-(\lambda-\mu-q_{10})t})^{n-1} e^{-(\lambda-\mu-q_{10})t}}{(\lambda - \mu e^{-(\lambda-\mu-q_{10})t})^{n+1}} \frac{f(t)}{\int_0^T p(t \mid n) f(t)\, dt}$$

$$f(t) = [(q_{10} + q_{01} e^{(q_{10}+q_{01})(t-T)}) q_{01}] N_0(t)$$

$N_0(t)$ is constant in macroecology study

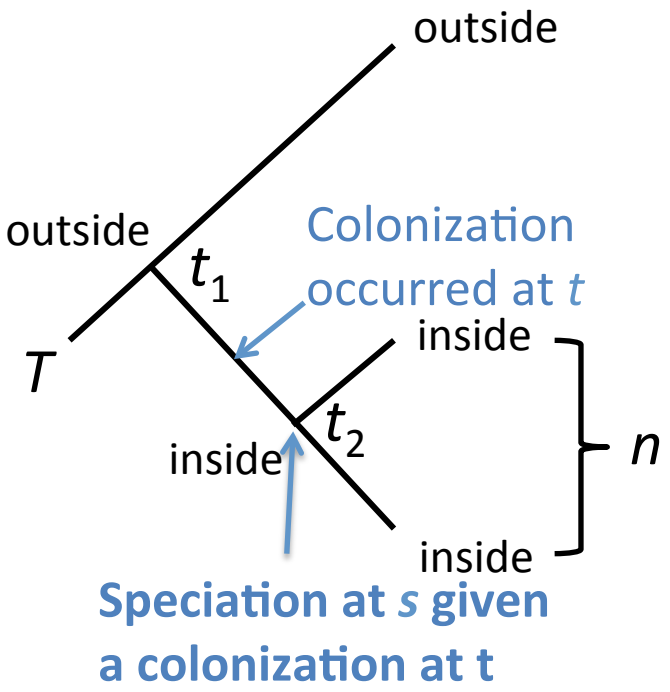
$N_0(t)$ follows deterministic two state birth-death model in macroevolution study

Framework of testing even species addition rate using phylogenies

$$L(\text{assemblage} \mid \text{speciation } \lambda;\ \text{extinction } \mu;\ \text{colonization } q_{01};\ \text{emigration } q_{10}) =$$

$$\sum_j \sum_i P(\text{assemblage} \mid \text{history } i;\ \lambda, \mu, q_{01}, q_{10})$$

Birth-death model

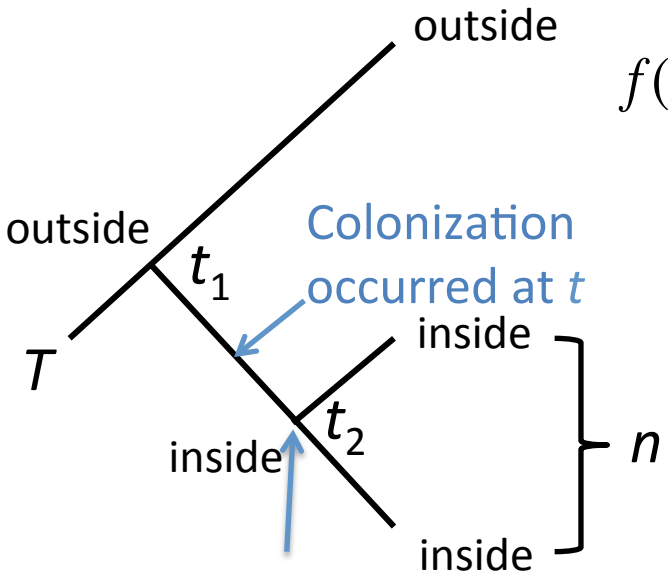


Framework of testing even species addition rate using phylogenies

$$L(\text{assemblage} \mid \text{speciation } \lambda;\ \text{extinction } \mu;\ \text{colonization } q_{01};\ \text{emigration } q_{10}) =$$

$$\sum_j \sum_i P(\text{assemblage} \mid \text{history } i;\ \lambda, \mu, q_{01}, q_{10})$$

Birth-death model



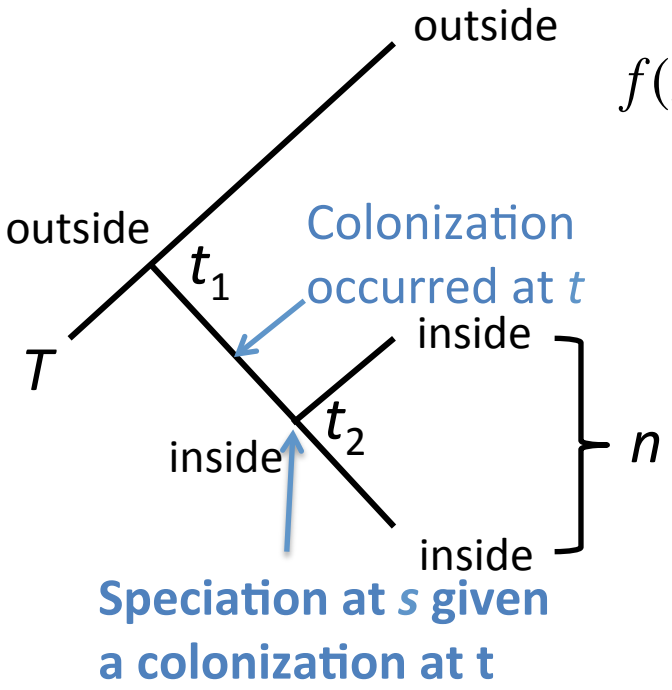
$$f(s \mid t) = \frac{(\lambda - \mu - q_{10})^2 e^{-(\lambda-\mu-q_{10})s}}{[\lambda - (\mu + q_{10})e^{-(\lambda-\mu-q_{10})s}]^2} \frac{\lambda - (\mu + q_{10})e^{-(\lambda-\mu-q_{10})t}}{1 - e^{-(\lambda-\mu-q_{10})t}}$$

Framework of testing even species addition rate using phylogenies

$$L(\text{assemblage} \mid \text{speciation } \lambda;\ \text{extinction } \mu;\ \text{colonization } q_{01};\ \text{emigration } q_{10}) =$$

$$\sum_j \sum_i P(\text{assemblage} \mid \text{history } i;\ \lambda, \mu, q_{01}, q_{10})$$

Birth-death model



$$f(s \mid t) = \frac{(\lambda - \mu - q_{10})^2 e^{-(\lambda-\mu-q_{10})s}}{[\lambda - (\mu + q_{10})e^{-(\lambda-\mu-q_{10})s}]^2} \frac{\lambda - (\mu + q_{10})e^{-(\lambda-\mu-q_{10})t}}{1 - e^{-(\lambda-\mu-q_{10})t}}$$

Or, for genus-level phylogeny

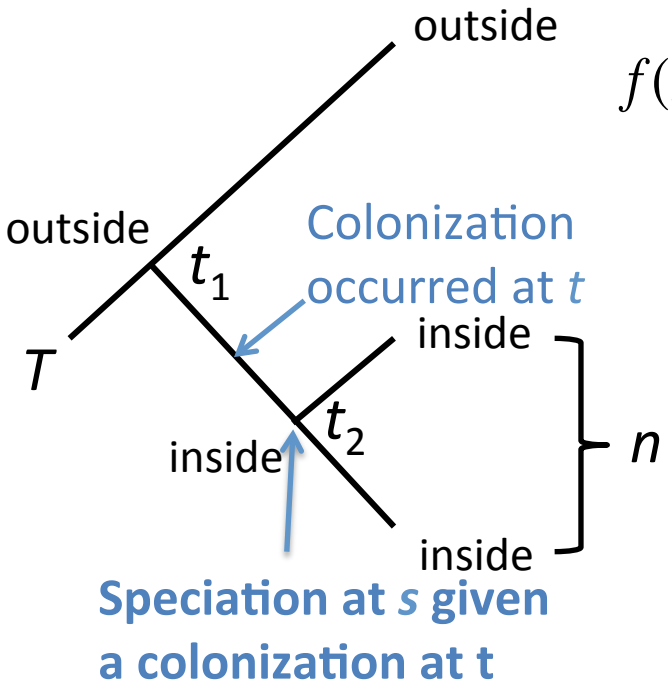
$$f(s \mid t) = [\frac{1 - e^{-(\lambda-\mu-q_{10})s}}{\lambda - (\mu + q_{10})e^{-(\lambda-\mu-q_{10})s}} \frac{\lambda - (\mu + q_{10})e^{-(\lambda-\mu-q_{10})t}}{1 - e^{-(\lambda-\mu-q_{10})t}}]^{n-1}$$

Framework of testing even species addition rate using phylogenies

$$L(\text{assemblage} \mid \text{speciation } \lambda;\ \text{extinction } \mu;\ \text{colonization } q_{01};\ \text{emigration } q_{10}) =$$

$$\sum_j \sum_i P(\text{assemblage} \mid \text{history } i;\ \lambda, \mu, q_{01}, q_{10})$$

Birth-death model



$$f(s \mid t) = \frac{(\lambda - \mu - q_{10})^2 e^{-(\lambda-\mu-q_{10})s}}{[\lambda - (\mu + q_{10})e^{-(\lambda-\mu-q_{10})s}]^2} \frac{\lambda - (\mu + q_{10})e^{-(\lambda-\mu-q_{10})t}}{1 - e^{-(\lambda-\mu-q_{10})t}}$$

Or, for genus-level phylogeny

$$f(s \mid t) = [\frac{1 - e^{-(\lambda-\mu-q_{10})s}}{\lambda - (\mu + q_{10})e^{-(\lambda-\mu-q_{10})s}} \frac{\lambda - (\mu + q_{10})e^{-(\lambda-\mu-q_{10})t}}{1 - e^{-(\lambda-\mu-q_{10})t}}]^{n-1}$$

Prior $f(s)$ follows coalescent theory if using species tree approach

Taxon sampling bias is inevitable:
Compare simulated and observed phylogenies

Taxon sampling bias is inevitable:
Compare simulated and observed phylogenies

Step 1: MCMC searches for likely sets of parameter values

Taxon sampling bias is inevitable:
Compare simulated and observed phylogenies

Step 1: MCMC searches for likely sets of parameter values

Step 2: Simulate phylogenies using those parameter sets

Constrain the extant number of lineages and total divergence time

Taxon sampling bias is inevitable:
Compare simulated and observed phylogenies

Step 1: MCMC searches for likely sets of parameter values

Step 2: Simulate phylogenies using those parameter sets

Constrain the extant number of lineages and total divergence time

Step 3: ML (observed) << ML(simulated) ?

small type I error, large type II error

Taxon sampling bias is inevitable:
Compare simulated and observed phylogenies

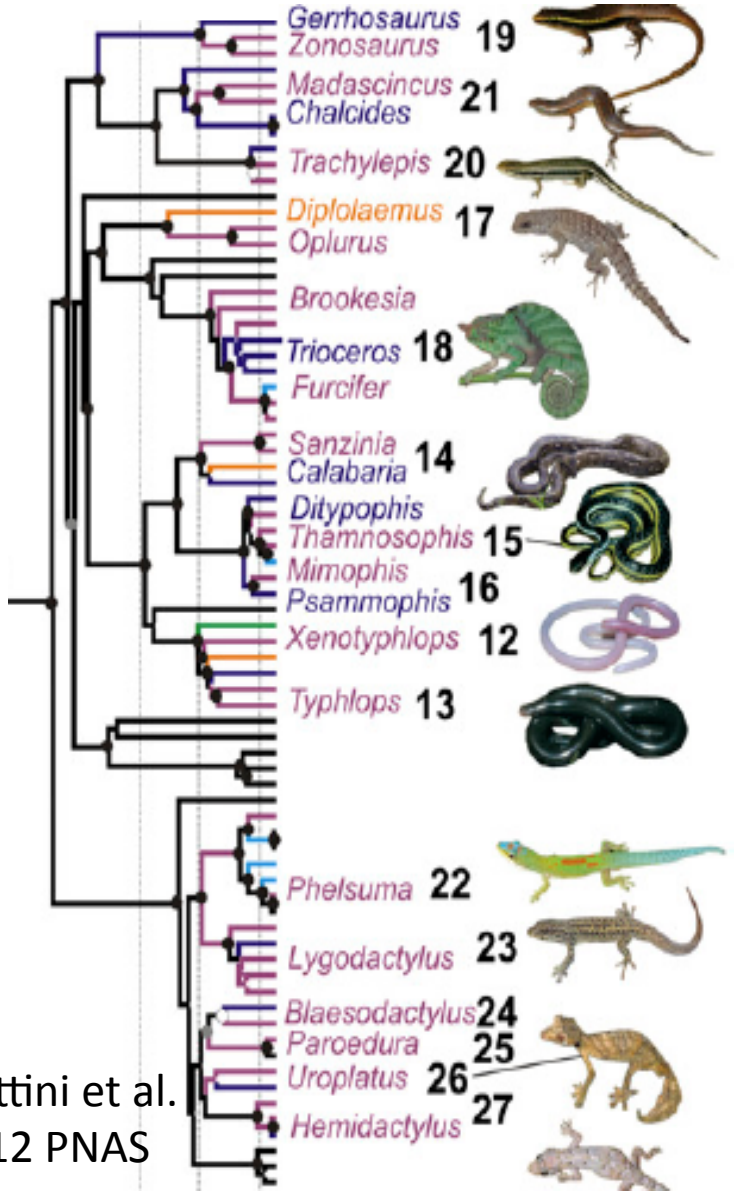
Step 1: MCMC searches for likely sets of parameter values

Step 2: Simulate phylogenies using those parameter sets
Constrain the extant number of lineages and total divergence time

Step 3: ML (observed) << ML(simulated) ?
small type I error, large type II error

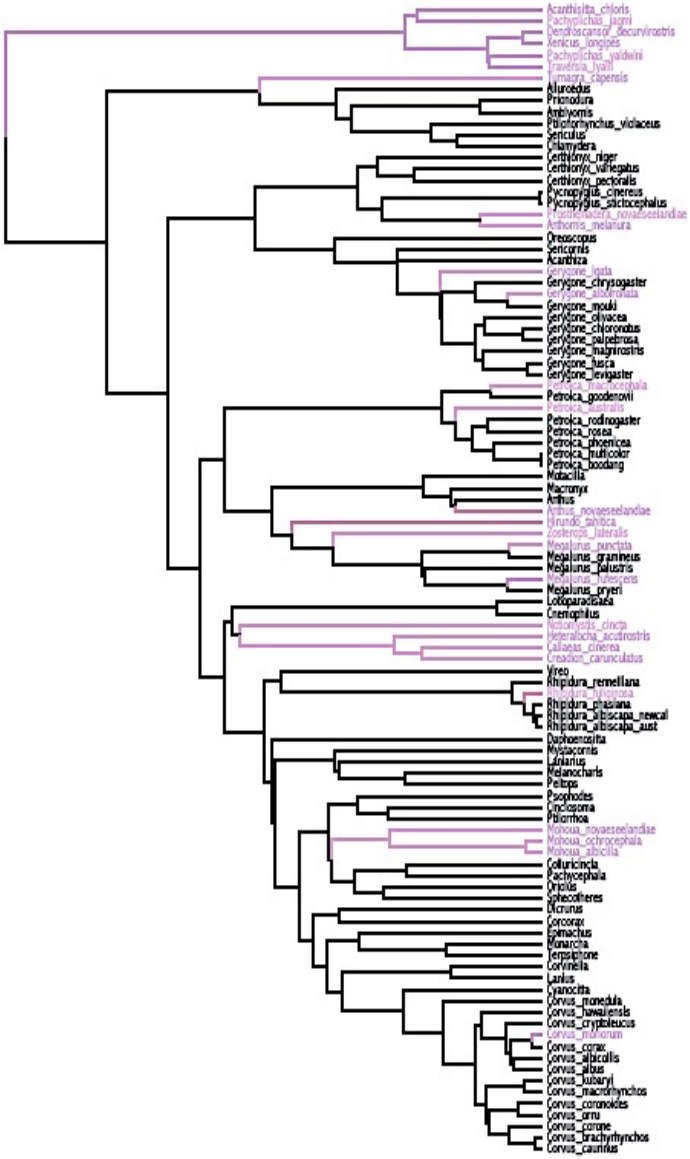
Step 4: Construct CI of simulated and observed histories:
Colonization frequency over time
Number of extant lineages per colonization

Madagascar squamates

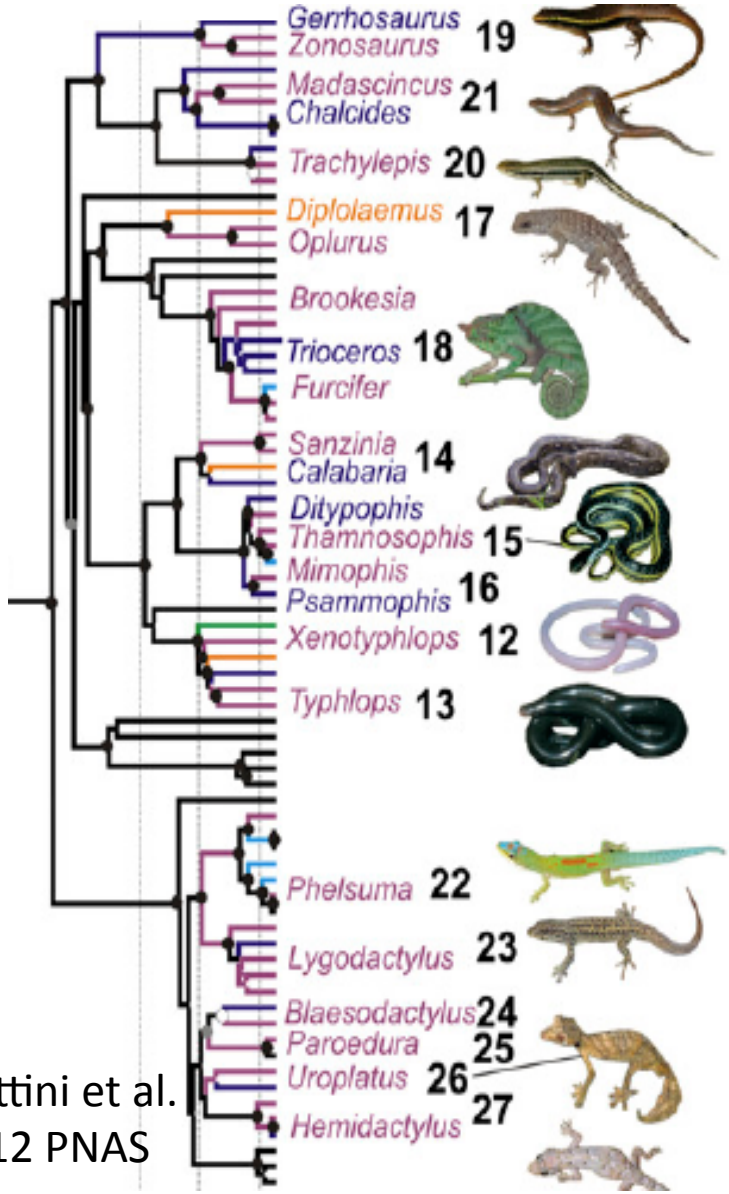


Crottini et al.
2012 PNAS

New Zealand passerines



Madagascar squamates



Crottini et al. 2012 PNAS

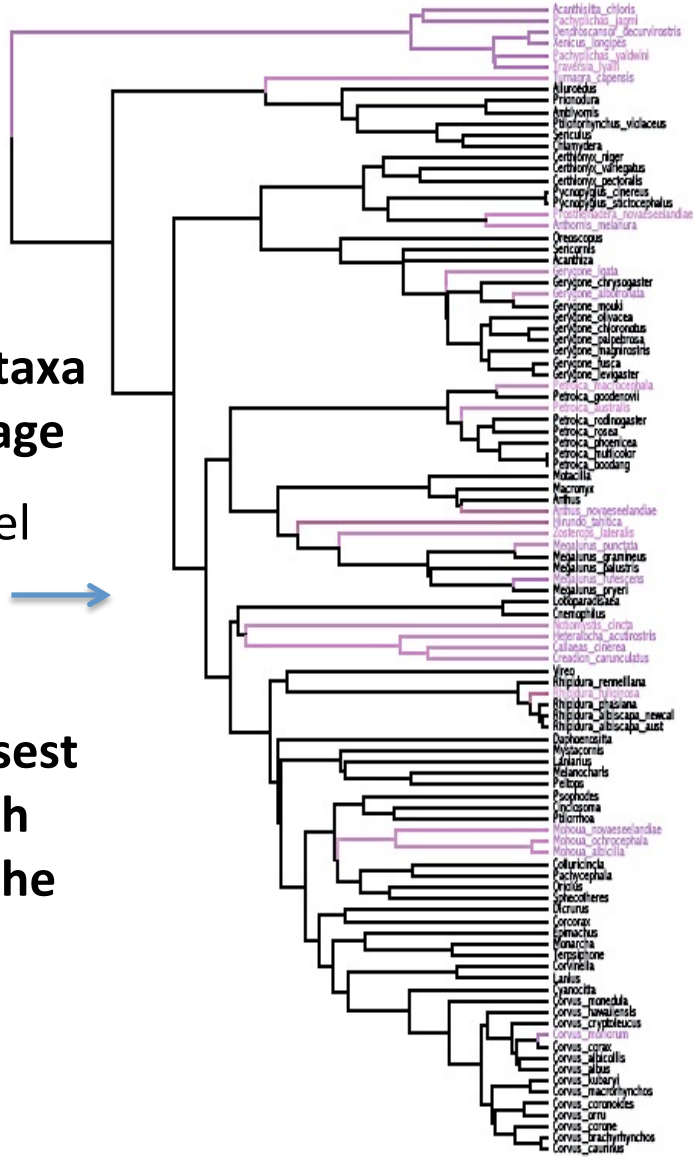
Include all the taxa in the assemblage

← Genus-level

Species-level →

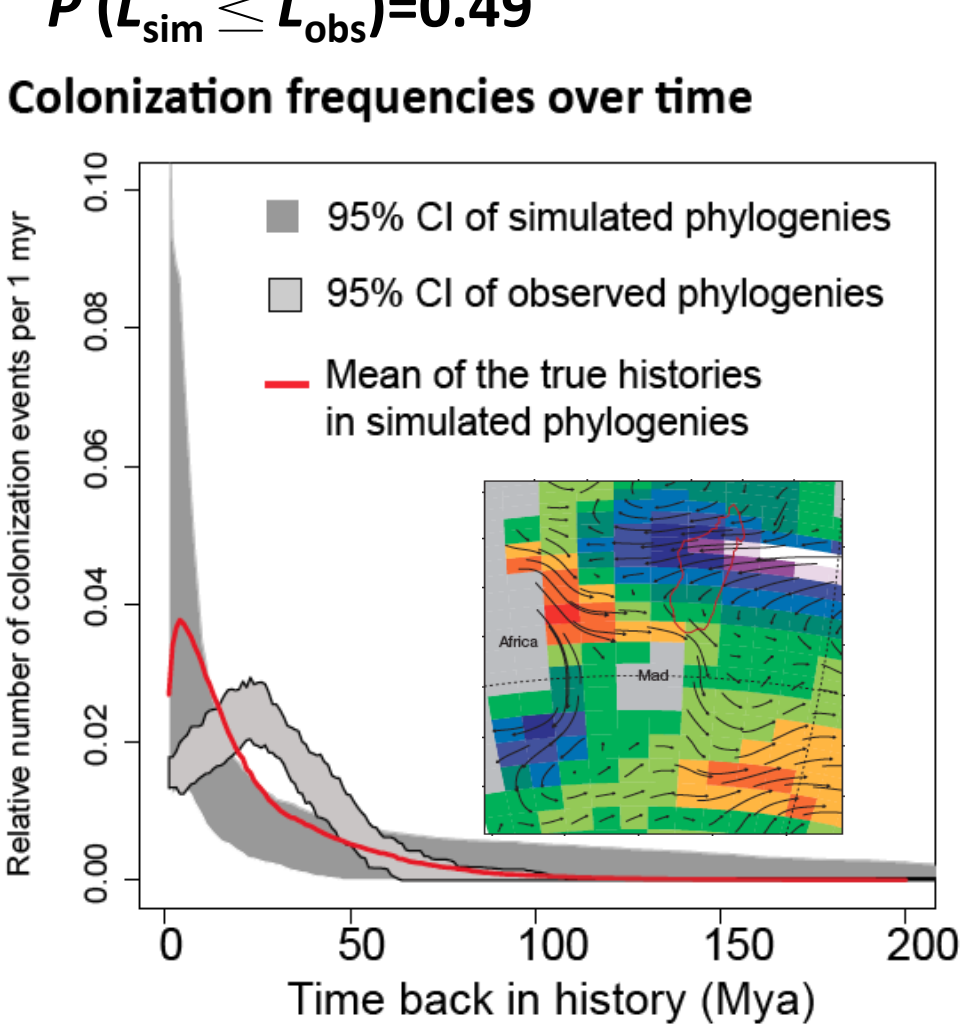
Include the closest relatives of each taxon outside the assemblage

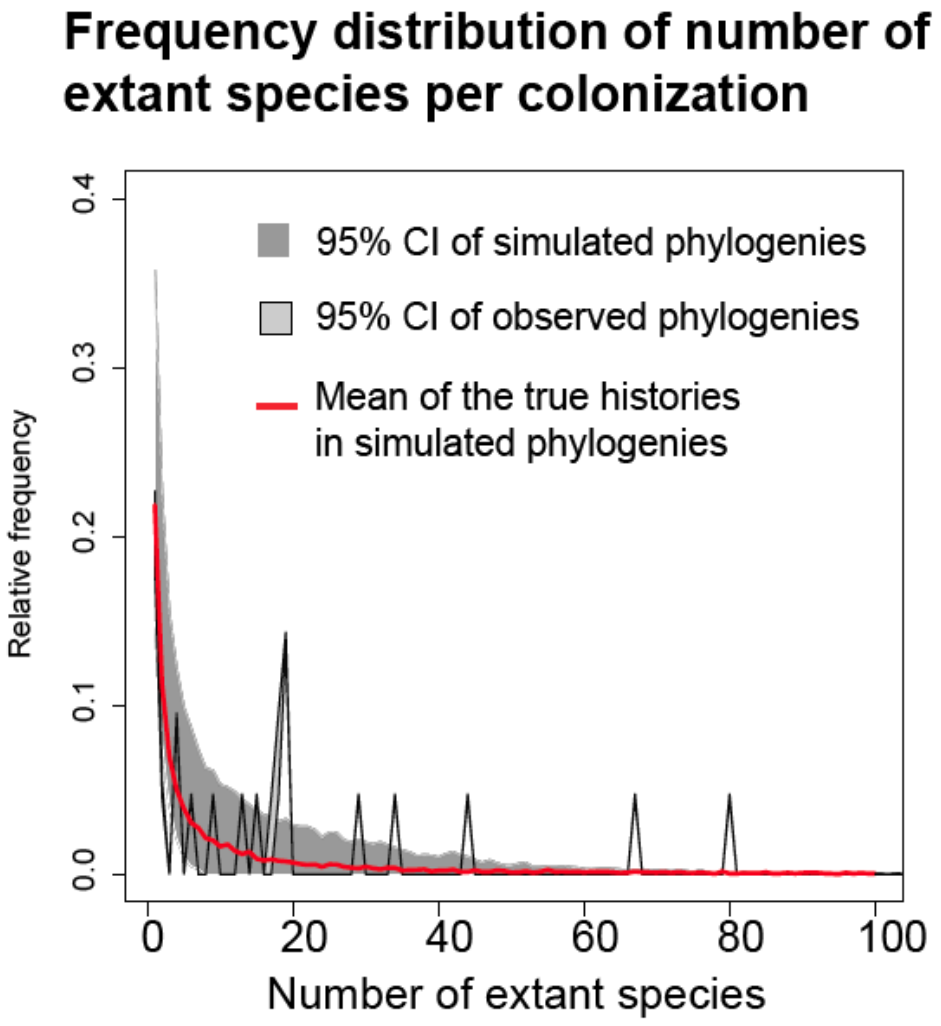
New Zealand passerines



More colorizations during Cenozoic Era in Madagascar squamates

$P\ (L_{sim} \leq L_{obs})$=0.49

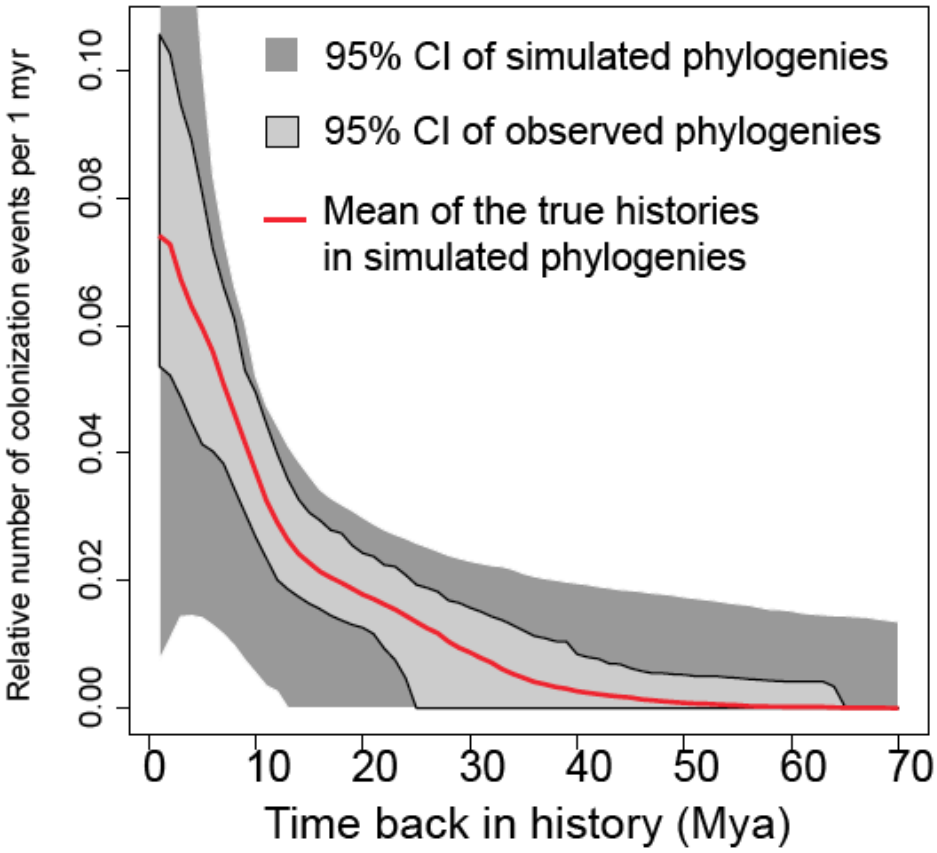




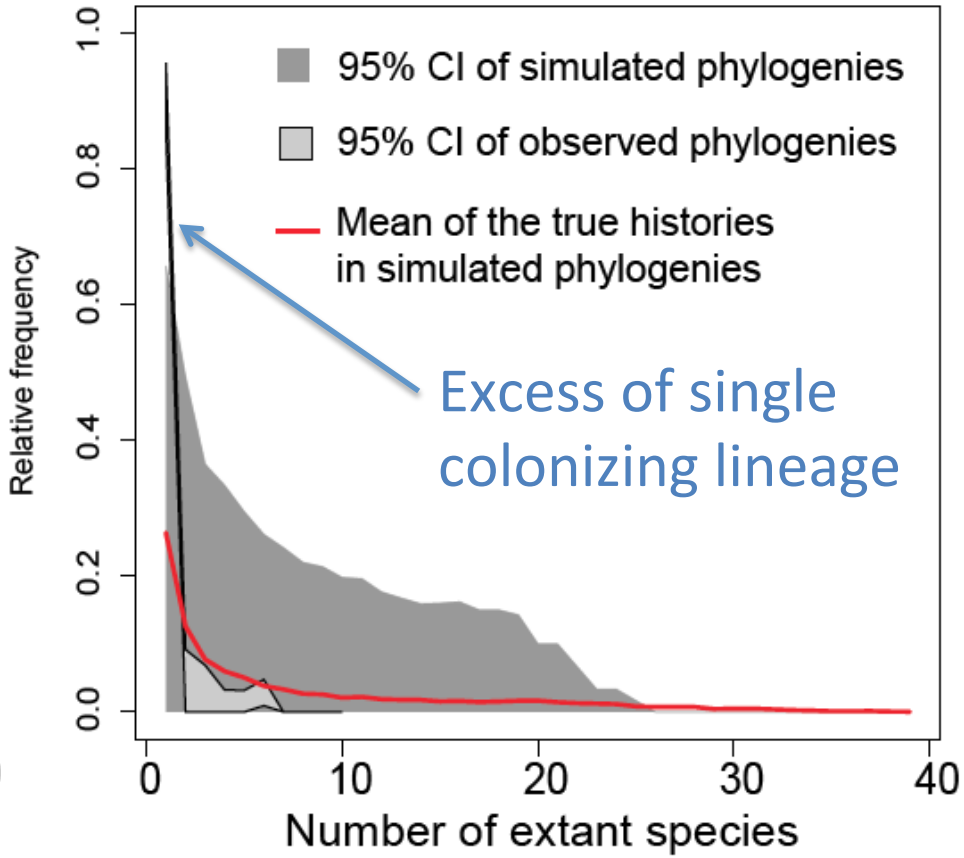
Constant colonization but inconstant speciation in New Zealand passerines

$P(L_{sim} \leq L_{obs})=0.04$

Colonization frequencies over time



Frequency distribution of number of extant species per colonization



Type I error estimated by simulations

Apply the method to 100 simulated phylogenies, each with 100 slice sampling to search for likely parameter sets

Type I error estimated by simulations

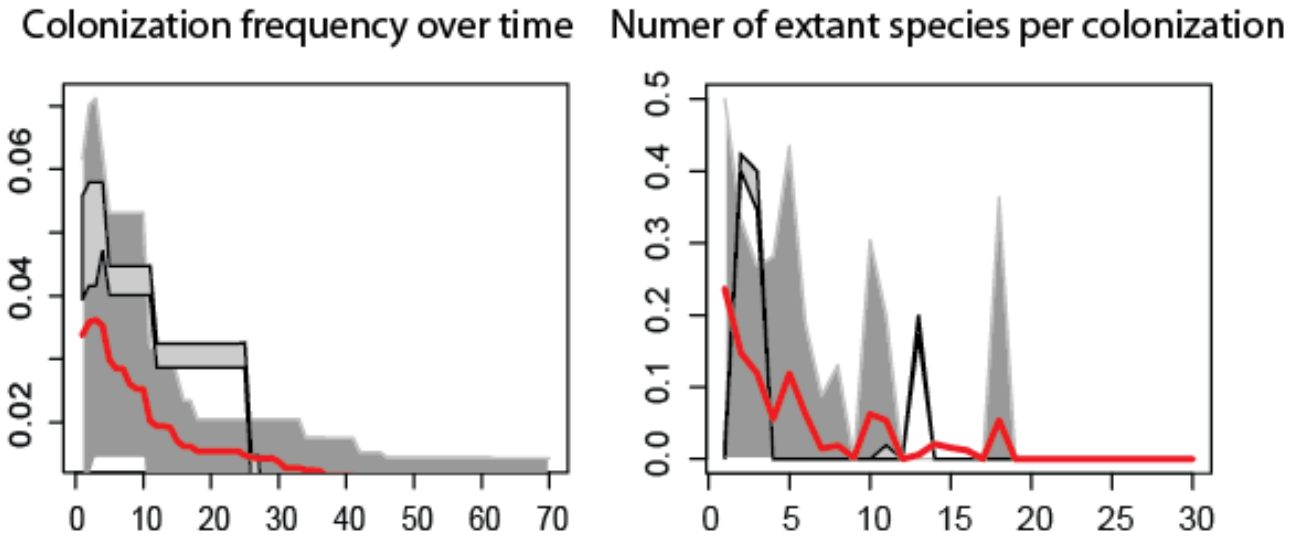
Apply the method to 100 simulated phylogenies, each with 100 slice sampling to search for likely parameter sets

Likelihoods: 0
Comparisons of reconstructed histories: 0.1

Type I error estimated by simulations

Apply the method to 100 simulated phylogenies, each with 100 slice sampling to search for likely parameter sets

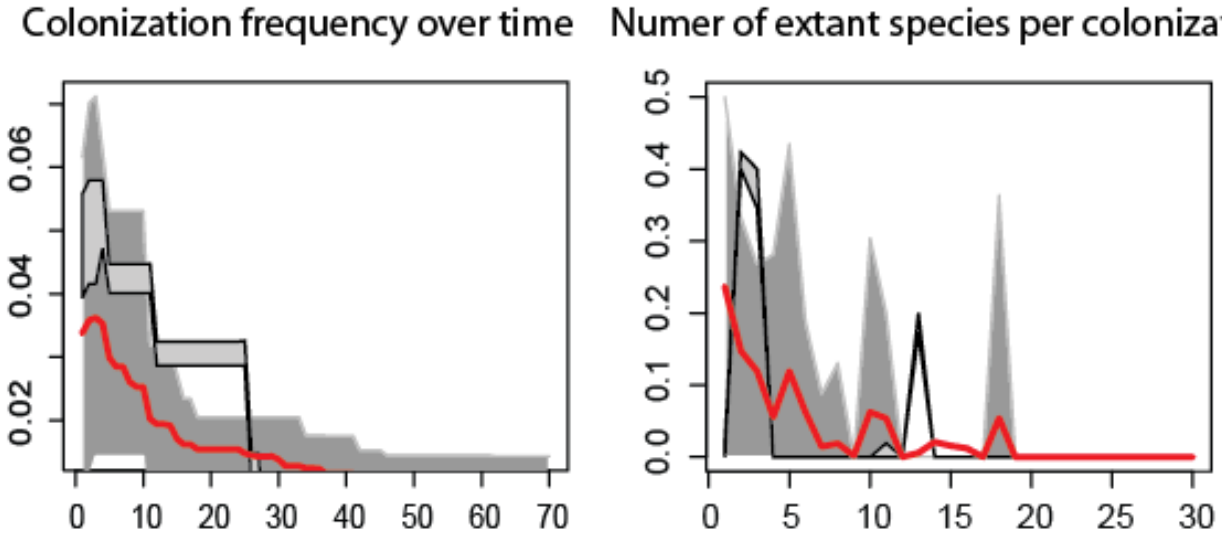
Likelihoods: 0
Comparisons of reconstructed histories: 0.1



Type I error estimated by simulations

Apply the method to 100 simulated phylogenies, each with 100 slice sampling to search for likely parameter sets

Likelihoods: 0
Comparisons of reconstructed histories: 0.1



Sample size?

MCMC sampling ?

Type II error?

Further explore taxon sampling bias

The method does not reconstruct unbiased assemblage histories. Thus, it should not be used to test effects of a specific historical or geological events.

Further explore taxon sampling bias

The method does not reconstruct unbiased assemblage histories. Thus, it should not be used to test effects of a specific historical or geological events.

A hidden Markov model may reconstruct a less biased assemblage history. Require numerical integrations and time-consuming.

Further explore taxon sampling bias

The method does not reconstruct unbiased assemblage histories. Thus, it should not be used to test effects of a specific historical or geological events.

A hidden Markov model may reconstruct a less biased assemblage history. Require numerical integrations and time-consuming.

Compare performance between hidden Markov methods and our analytical approach in reconstructing assemblage histories

Acknowledgements

To Celine Poux, Miguel Vences, and Ted Townsend for providing their data and analysis files

Funded by L. Bromham: ARC discovery projects and future fellowship

Thanks!