Plan for this Afternoon

Character state-dependent diversification

* Theory: Speciation, extinction, and trait transitions
* Practical: Using BiSSE

Break

Biogeography

* Theory: Modeling range shifts and geographic mode of speciation
* Practical
    - Using DEC
    - Using ClaSSE

[Emma Goldberg. RevBayes NIMBioS. 10 Aug 2017.]

Outline

Motivating Biological Questions

Macroevolutionary processes

* Traits that affect speciation or extinction
  - Is this trait a key innovation?
  - Does species-level selection act on this trait?
* Trait evolution
  - How quickly does this trait evolve?
  - Is evolution of this trait irreversible?

Central point: Questions of trait evolution and diversification are intrinsically linked. . .

Methods History

Some methods consider these processes separately

* Lineage diversification: sister clades (Mitter et al. 1988)
* Trait evolution: models like sequence evol. (Pagel 1999; Lewis 2001)

[board: sister clades, transition model; unequal freqs]

Methods History

Some methods consider these processes separately

* Lineage diversification: sister clades (Mitter et al. 1988)
* Trait evolution: models like sequence evol. (Pagel 1999; Lewis 2001)

But asymmetry in one process can mislead inference of other

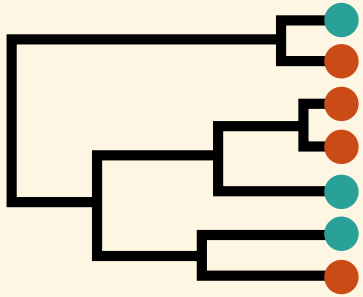
* Simulation demonstration (Maddison 2006)
* BiSSE model developed as a solution (Maddison et al. 2007)
  "Binary State Speciation and Extinction"

[board: sister clades, transition model; unequal freqs]

How BiSSE Works

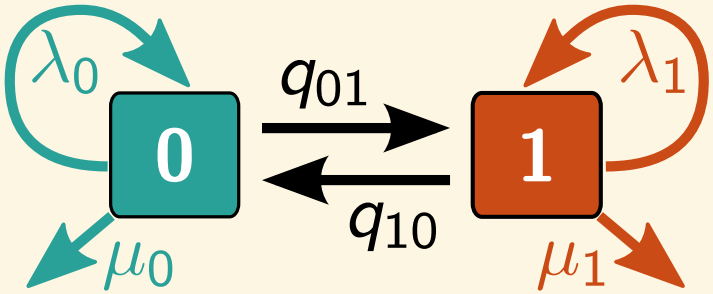
Input data

* "Known" tree
* Observed tip states



Goal

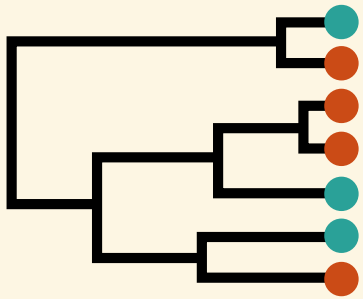
* Likelihood of data under 6-parameter model



How BiSSE Works

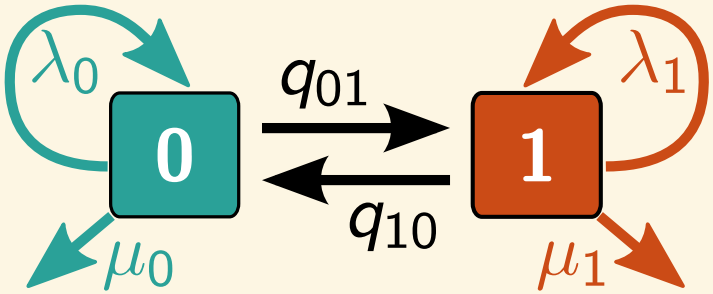
Input data

* "Known" tree
* Observed tip states



Goal

* Likelihood of data under 6-parameter model

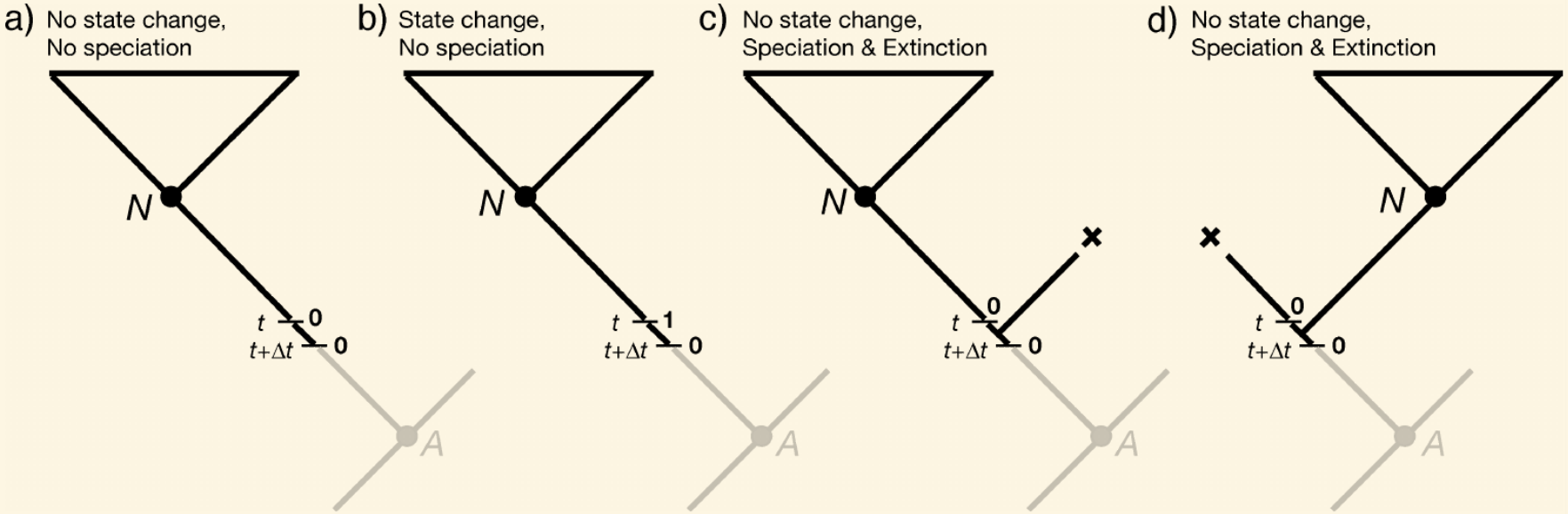


Approach

* Evolve two probability vectors down the tree
  - Clade probabilities, $D_{Ni}(t)$
  - Extinction probabilities, $E_i(t)$

How BiSSE Works

Within branches



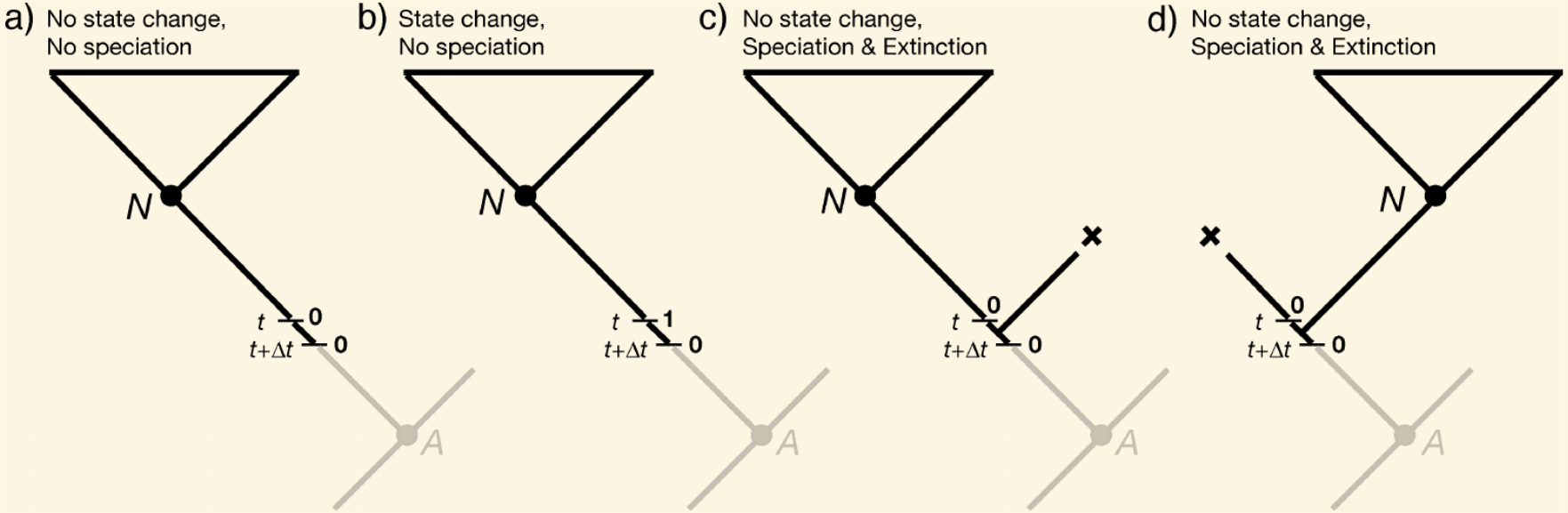
$$\frac{dD_{N0}(t)}{dt} = -\left(\lambda_0 + \mu_0 + q_{01}\right) D_{N0}(t) + q_{01} D_{N1}(t) + 2\lambda_0 D_{N0}(t) E_0(t)$$

$$\frac{dD_{N1}(t)}{dt} = -\left(\lambda_1 + \mu_1 + q_{10}\right) D_{N1}(t) + q_{10} D_{N0}(t) + 2\lambda_1 D_{N1}(t) E_1(t)$$

[board: walk down from tips]

How BiSSE Works

Within branches



$$\frac{dD_{N0}(t)}{dt} = -\left(\lambda_0 + \mu_0 + q_{01}\right) D_{N0}(t) + q_{01} D_{N1}(t) + 2\lambda_0 D_{N0}(t) E_0(t)$$

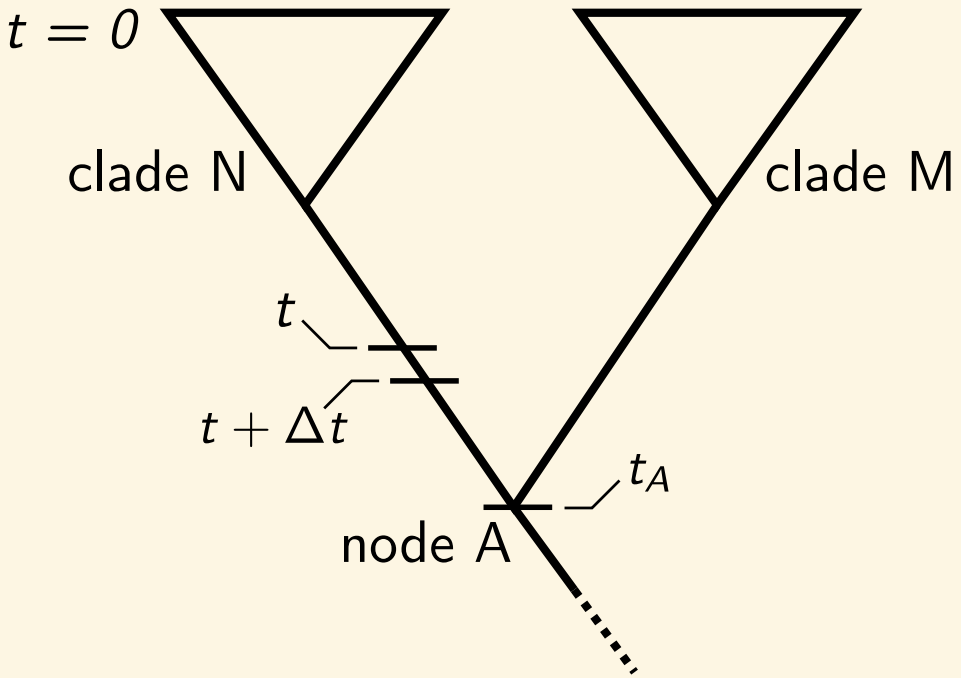
$$\frac{dD_{N1}(t)}{dt} = -\left(\lambda_1 + \mu_1 + q_{10}\right) D_{N1}(t) + q_{10} D_{N0}(t) + 2\lambda_1 D_{N1}(t) E_1(t)$$

$$\frac{dE_0(t)}{dt} = \mu_0 - \left(\lambda_0 + \mu_0 + q_{01}\right) E_0(t) + q_{01} E_1(t) + \lambda_0 E_0(t)^2$$

$$\frac{dE_1(t)}{dt} = \mu_1 - \left(\lambda_1 + \mu_1 + q_{10}\right) E_1(t) + q_{10} E_0(t) + \lambda_1 E_i(t)^2$$

How BiSSE Works

At nodes



$$D_{A0}(t_A) = D_{N0}(t_A) D_{M0}(t_A) \lambda_0$$
$$D_{A1}(t_A) = D_{N1}(t_A) D_{M1}(t_A) \lambda_1$$

[board: walk through node]

How BiSSE Works

At the root

* Have $D_{R0}(t_R)$ and $D_{R1}(t_R)$
* Combine them to get the overall likelihood of the data

The result

* Likelihood of the data given the model
  - Data = tree shape and tip states
  - Model = BiSSE with 6 parameter values

Then Bayesian inference, etc...

[board: root weightings]

Current State of Methods

BiSSE is a smart approach

* Processes suited to biological questions
* Correct algorithm and implementations

Caution 1: Correlation, not causation

* Could be other, related trait

  (Maddison et al. 2007)

Current State of Methods

BiSSE is a smart approach

* Processes suited to biological questions
* Correct algorithm and implementations

Caution 1: Correlation, not causation

* Could be other, related trait

  (Maddison et al. 2007)

Caution 2: Correlations reported when not appropriate

* Phylogenetic pseudoreplication
* Other diversification rate heterogeneity

  (Maddison & FitzJohn 2015; Rabosky & Goldberg 2015)

Current State of Methods

Caution 2: Correlations often reported when not appropriate

What do we do?

* Less-model-based approaches
  (Rabosky & Huang 2016; Rabosky & Goldberg 2017)
* Add other processes to BiSSE
  e.g., hidden state (Beaulieu & O'Meara 2016)
* Ongoing research on other approaches...

[board: hisse states]

Outline

Character State-Dependent Diversification

Biogeography

- Trait Coding
- Range Shifts
- Geographic Mode of Speciation
- Methods

Biogeography

Why is "biogeography" a topic this week?

- Geographic range trait applies to all organisms
- Modeling it raises interesting methodological considerations
  - How to code the trait
  - Anagenetic and cladogenetic trait change
  - Geology informs time-stratified rate matrices

Trait Coding

Usual traits: discrete values, maybe some uncertainty

Geographic range: often multiple values simultaneously

Approach: model states are all combinations of areas

[board: triminoes]

Trait Coding

For 3 areas...

absent: 

present: 

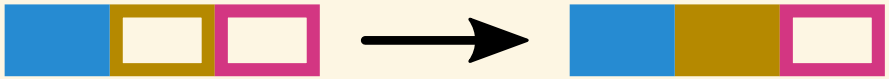
... these states are possible:

picture	*bits*	*index*
	0 0 0	0
	0 0 1	1
	0 1 0	2
	0 1 1	3
	1 0 0	4
	1 0 1	5
	1 1 0	6
	1 1 1	7

Range Shifts

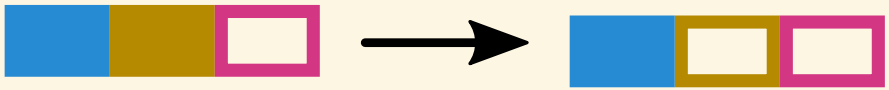
Range expansion or contraction

* State transition for one species
  Anagenetic change
* **Dispersal:** range expansion



at rate d_A

* **Extirpation** (a.k.a. local extinction): range contraction

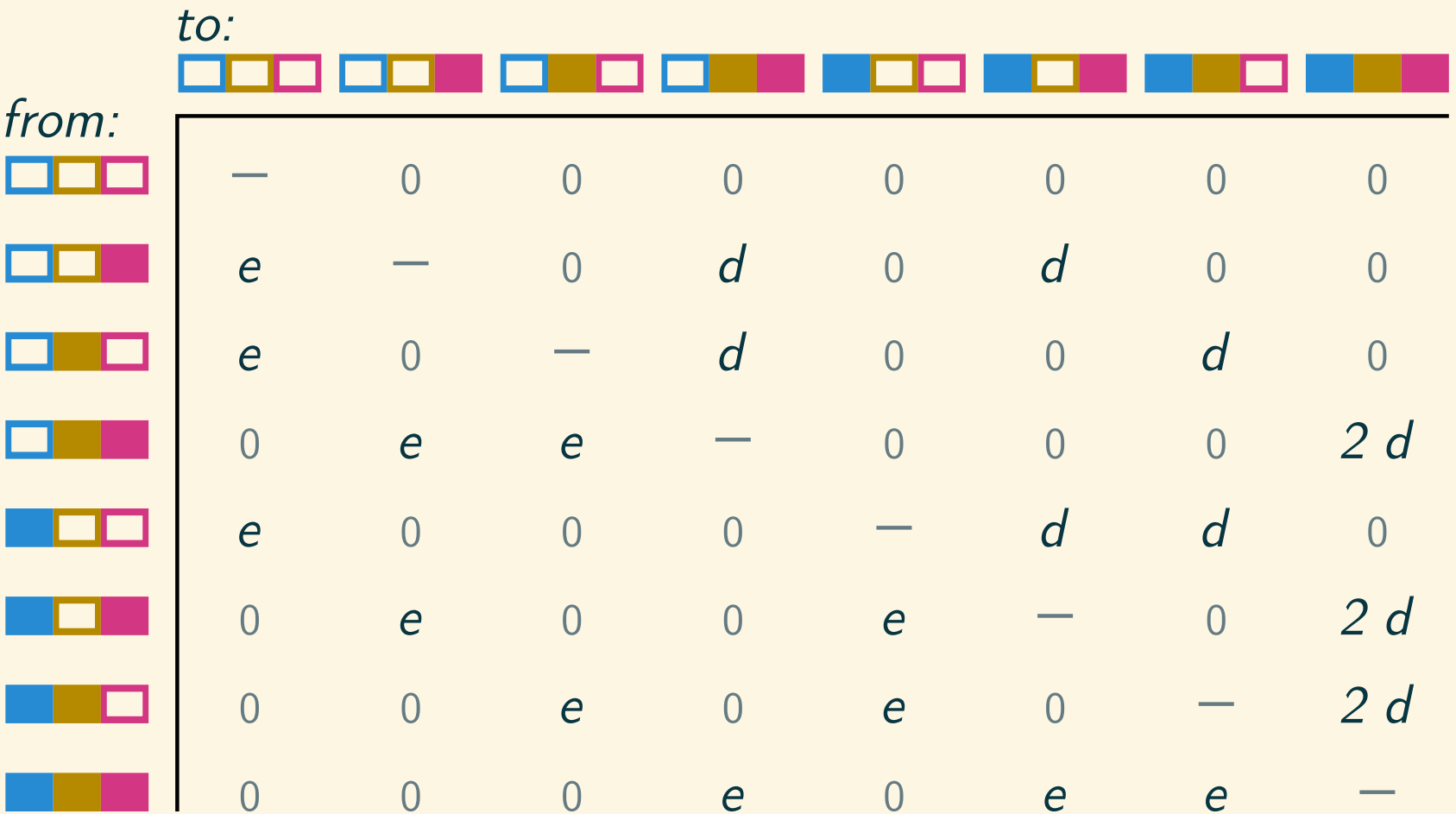


at rate e_B

[board: more examples]

Range Shifts

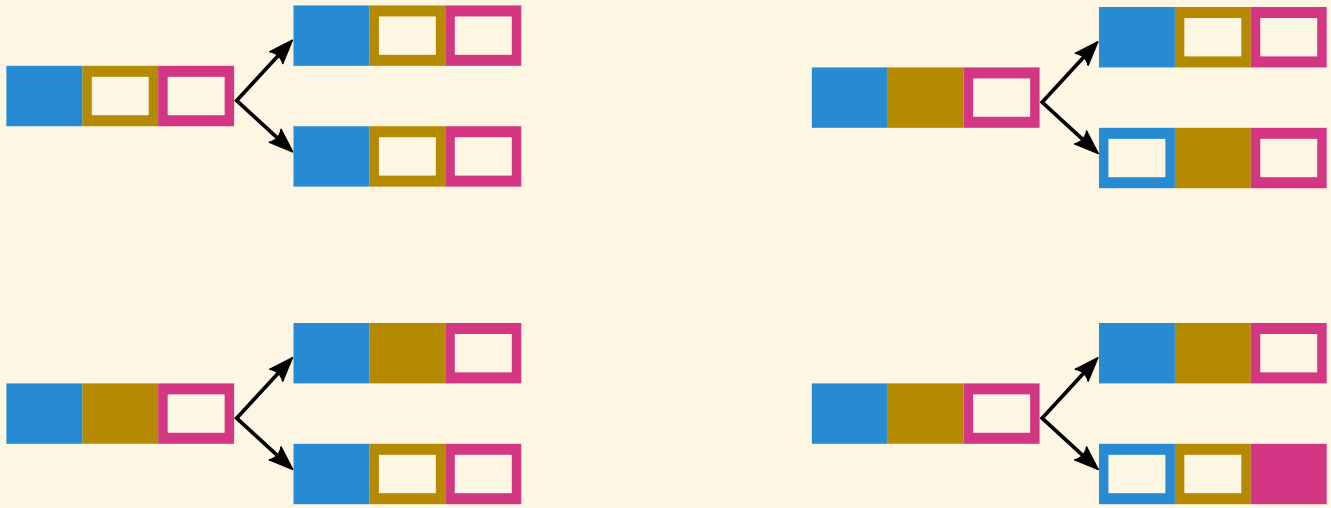
Which transitions to allow is *your* design decision. For example:



Geographic Mode of Speciation

Range changes "during" speciation

* State change as one species splits into two
  Cladogenetic change
  - Speciation within one region
  - Speciation between regions
  - Speciation with dispersal



Many possibilities. Which ones to allow is *your* design decision.

[board: more examples]

Methods

DEC

* Dispersal-Extinction-Cladogenesis
  (Ree et al. 2005; Ree & Smith 2008; Landis et al. 2013; Matzke 2014)
* Range evolution on a fixed tree

ClaSSE (GeoSSE)

* Cladogenetic (Geographic) State Speciation and Extinction
  (Goldberg et al. 2011; Goldberg & Igić 2012)
* Range evolution during a birth-death process

Methods: DEC

Anagenetic range change

* Along branches
* Rates of dispersal and local extinction

Cladogenetic range change

* At nodes
* Probabilities of the various modes
* A logical problem: nodes "hidden" by extinction

[board: hidden nodes]

Methods: ClaSSE

Anagenetic range change

* Along branches
* Rates of dispersal and local extinction

Cladogenetic range change

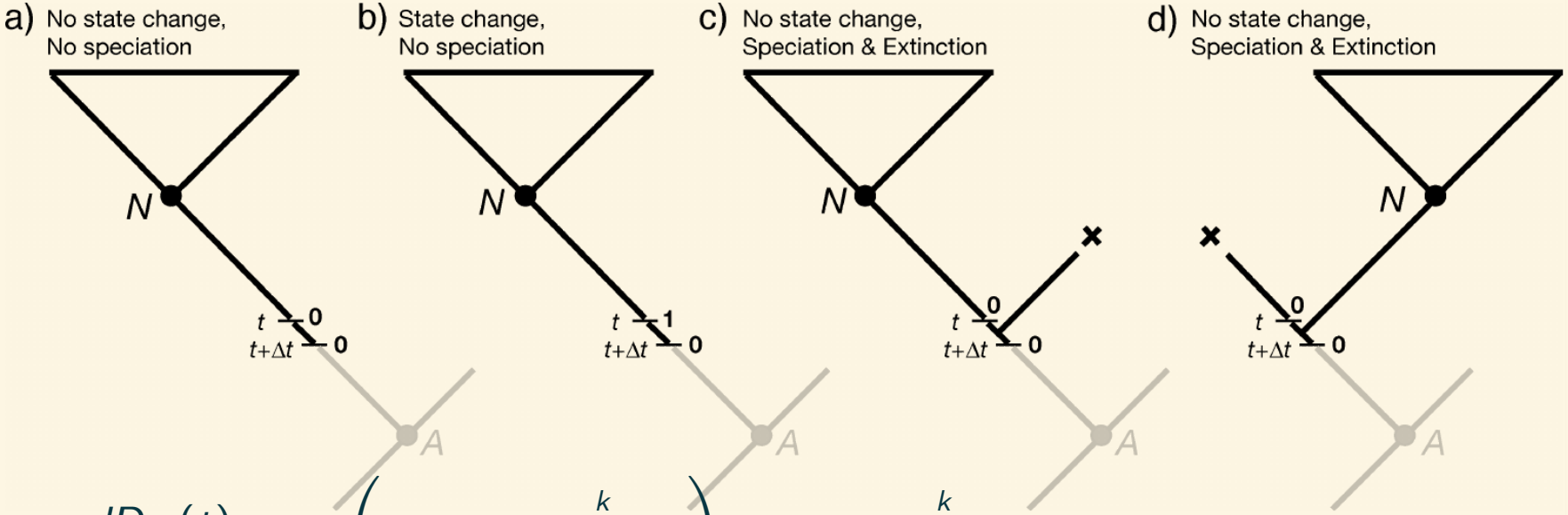
* At all speciation events
* Rates of speciation, via the various modes

Global extinction

* Branches pruned from the tree
* Allows for hidden nodes

Methods: BiSSE (MuSSE)

Within branches



$$\frac{dD_{Ni}(t)}{dt} = -\left(\lambda_i + \mu_i + \sum_{j \neq i}^{k} q_{ij}\right) D_{Ni}(t) + \sum_{j \neq i}^{k} q_{ij} D_{Nj}(t) + 2\lambda_i E_i(t) D_{Ni}(t)$$

$$\frac{dE_i(t)}{dt} = \mu_i - \left(\lambda_i + \mu_i + \sum_{j \neq i}^{k} q_{ij}\right) E_i(t) + \sum_{j \neq i}^{k} q_{ij} E_j(t) + \lambda_i E_i(t)^2$$

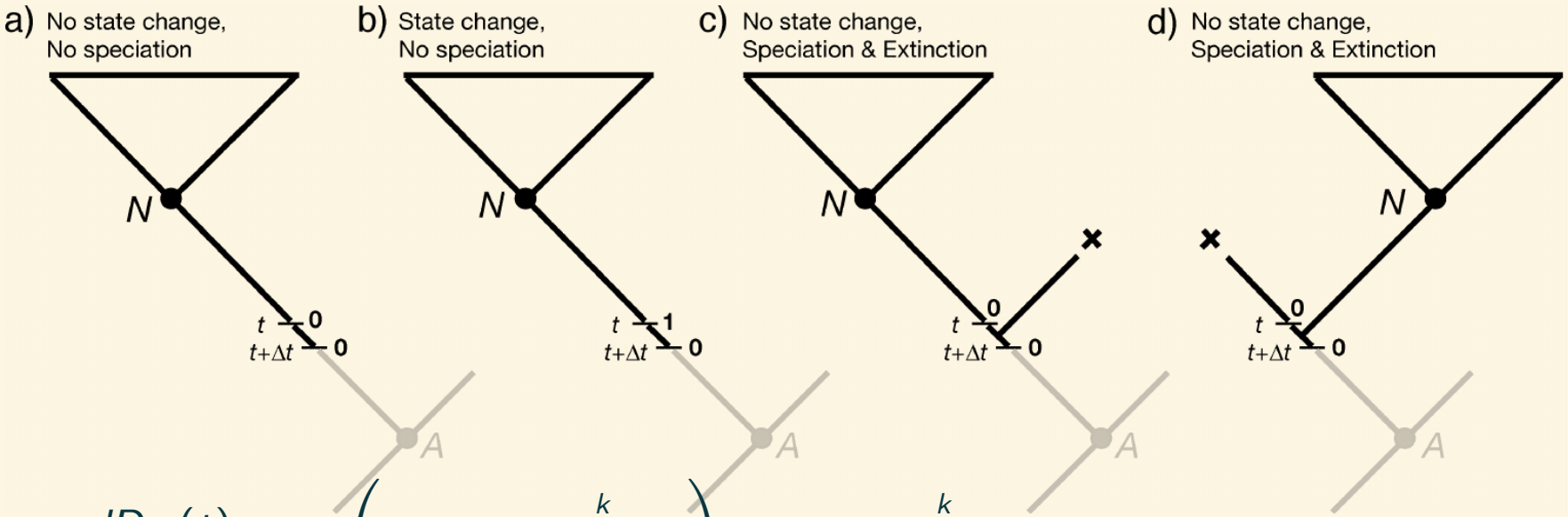
At nodes

$$D_{Ai}(t_A) = \lambda_i D_{Ni}(t_A) D_{Mi}(t_A)$$

[board: N, M reminder]

Methods: ClaSSE

Within branches



$$\frac{dD_{Ni}(t)}{dt} = -\left(\lambda_i + \mu_i + \sum_{j \neq i}^{k} q_{ij}\right) D_{Ni}(t) + \sum_{j \neq i}^{k} q_{ij} D_{Nj}(t) + 2\lambda_i E_i(t) D_{Ni}(t)$$

$$\frac{dE_i(t)}{dt} = \mu_i - \left(\lambda_i + \mu_i + \sum_{j \neq i}^{k} q_{ij}\right) E_i(t) + \sum_{j \neq i}^{k} q_{ij} E_j(t) + \lambda_i E_i(t)^2$$

At nodes

$$D_{Ai}(t_A) = \frac{1}{2} \sum_{j,k} \lambda_{ijk} \left[D_{Nj}(t_A) D_{Mk}(t_A) + D_{Nk}(t_A) D_{Mj}(t_A)\right]$$

Methods: ClaSSE

One example of parameterizing ClaSSE for biogeography:

Two areas: A B

Three states: 1 2 3

Dispersal: $q_{13} = d_B$ $q_{23} = d_A$

Local extinction: $q_{31} = e_A$ $q_{32} = e_B$

Global extinction: $\mu_1 = e_B$ $\mu_2 = e_A$

Speciation: $\lambda_{111} = s_B$ $\lambda_{222} = s_A$ $\lambda_{312} = \lambda_{321} = s_{AB}$

$\lambda_{313} = \lambda_{331} = s_B$ $\lambda_{323} = \lambda_{332} = s_A$

(all other rates are 0)