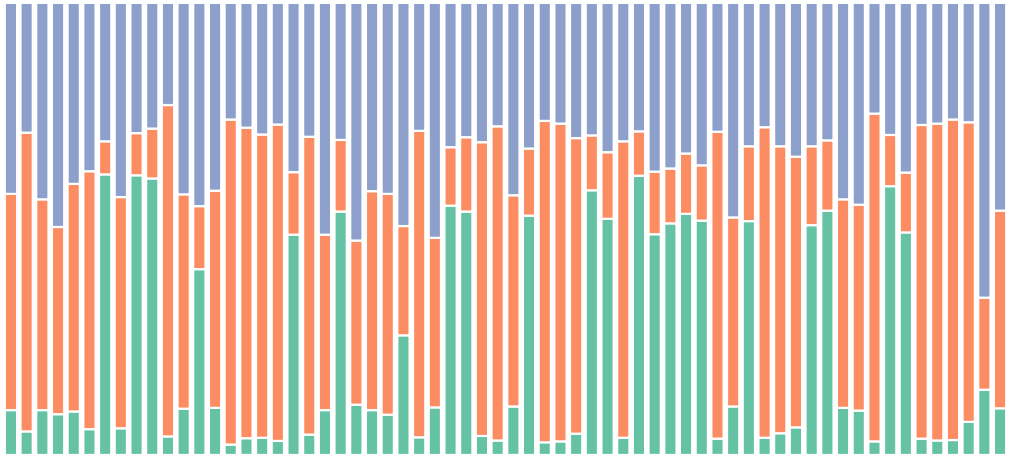
Identifying Models of Trait-Mediated Community Assembly

using random forests and approximate Bayesian computation

Megan Ruffley, Katie Peterson,
Bob Week, David C. Tank,
Luke J. Harmon



Society of Systematic Biologists

University of Idaho

SSE Society for the Study of Evolution

iDiv German Centre for Integrative Biodiversity Research

Santa Fe Institute

NSF

The Ohio State University

Botanical Society of America

FORBIO

the City College of New York

UiO : Life Science University of Oslo

ibest Institute for Bioinformatics and Evolutionary Studies

ID University of Idaho Stillinger Herbarium

Interested in identifying neutral and non-neutral processes of assembly

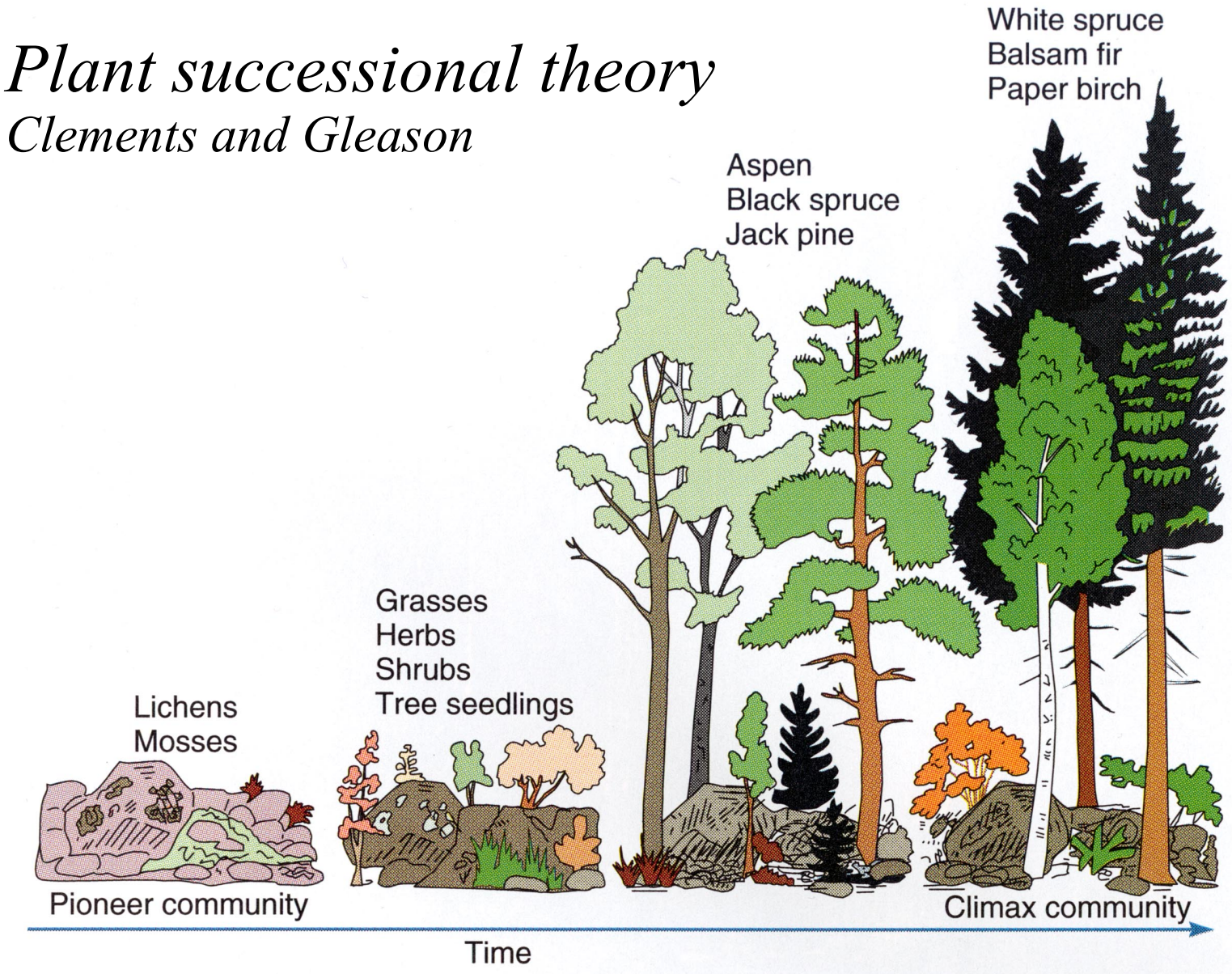
Environmental Filtering

Competitive Exclusion

Plant successional theory

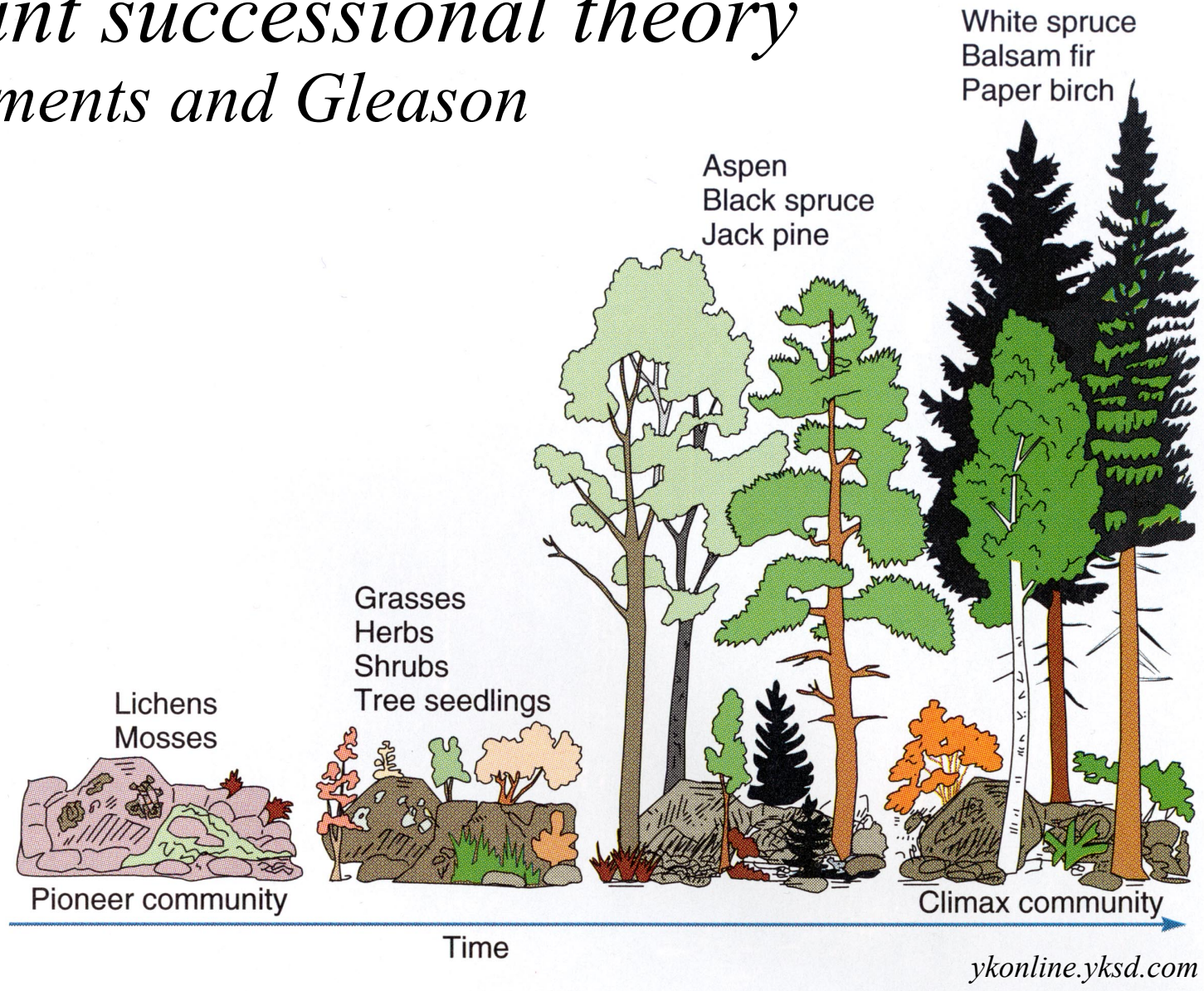
Clements and Gleason

Are communities a result of deterministic (Clements 1938) or stochastic (Gleason 1927) processes



Plant successional theory

Clements and Gleason



Frederic Clements (1916)

"Succession is deterministic"

-stable 'climax' community that is in perfect balance with the climate

"Communities are super-organisms"

-Communities are dependent on the species interactions among members

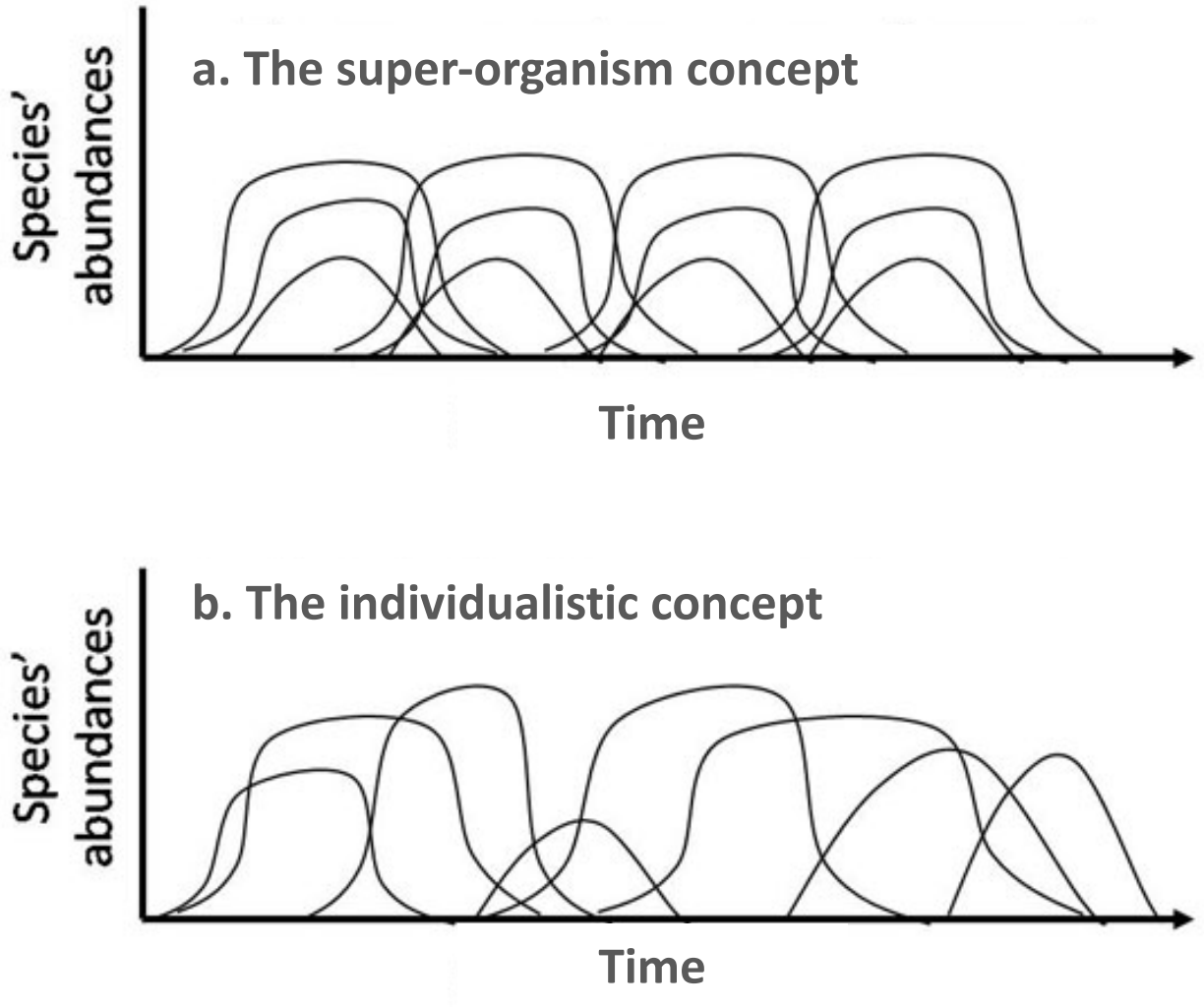
Henry Gleason (1917)

- "Each species is an independent entity"

  -distribution of each species is due to genetics, physiology, life history, and relations to abiotic and biotic factors

  -communities are just a consequence of independent interactions among species and their environment

Introduction: Clements and Gleason debate



(A) Groups of species are tightly associated, and are supplanted by other groups of tightly associated species. (B) Individual species independently respond to environmental conditions. Each curve on the graphs represents the abundance of a single species.

Emery, S. (2010)

What influences how communities are formed?

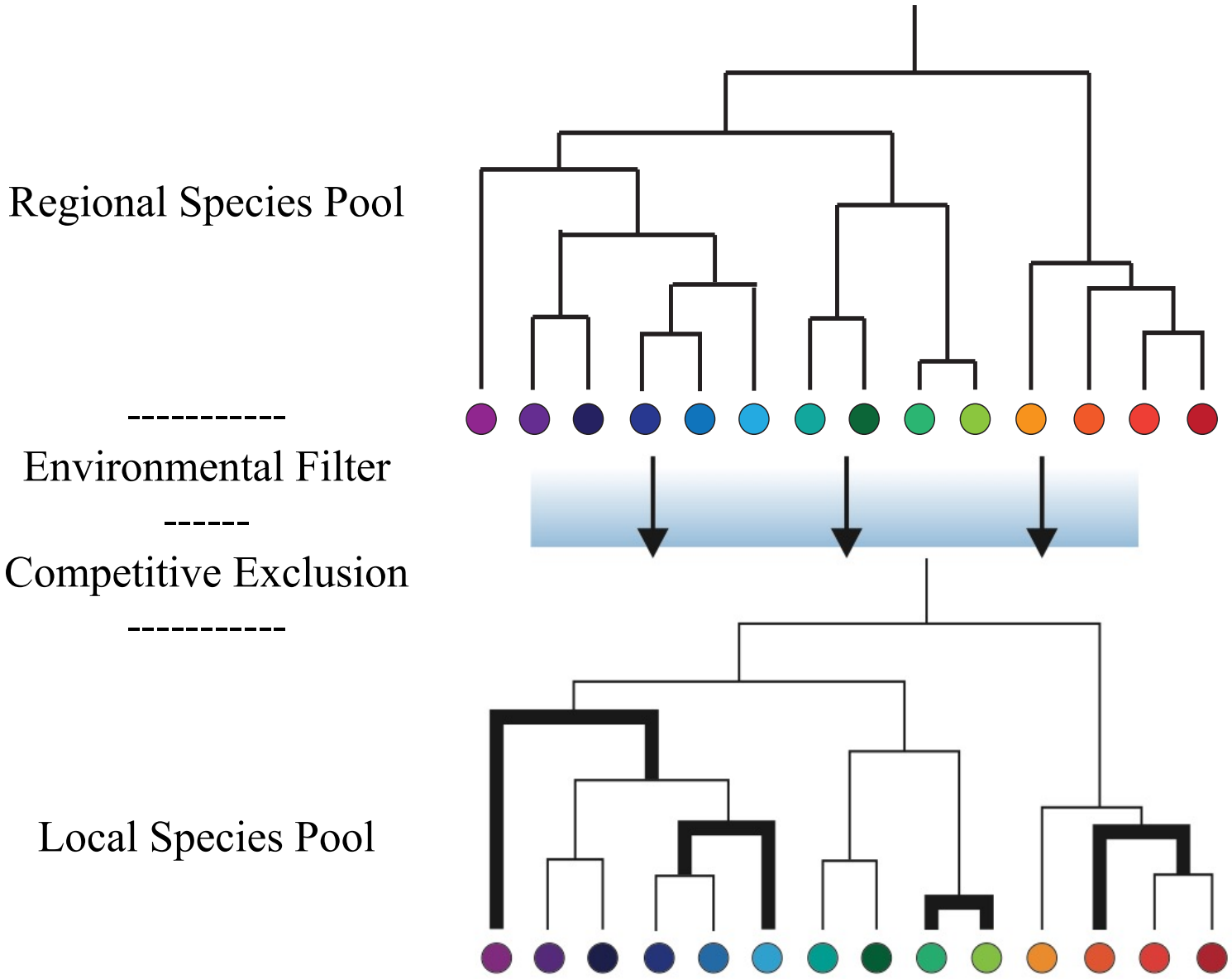
Clements – the environment

Habitat Filtering – species fail to establish in community due to incompatibility with relevant environmental factors

Gleason – the species

Competitive Exclusion – species fail to establish in a community when they cannot out-compete another species for their desired niche.

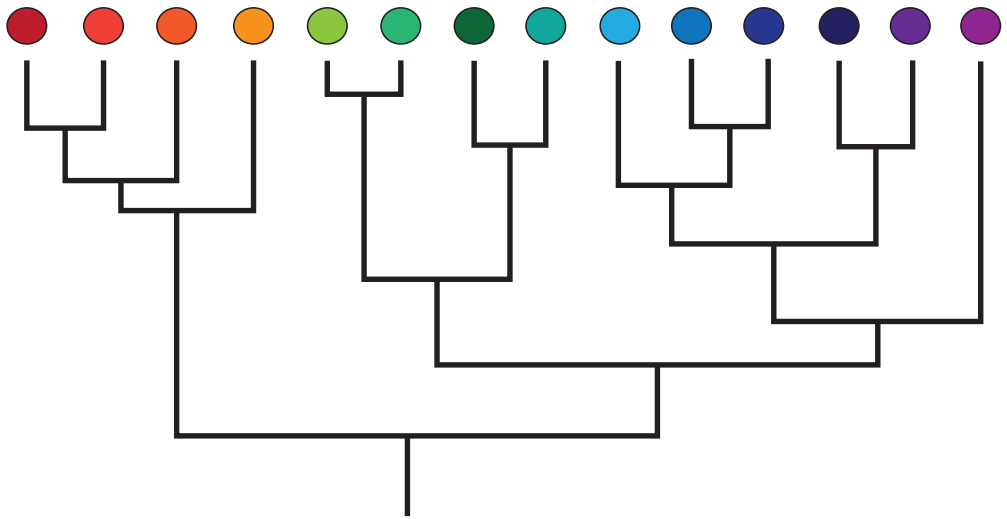
Community Assembly



How are community assembly models quantified?

Phylogenetic dispersion metrics are commonly use to detect signal for habitat filtering and competitive exclusion.

-Mean Phylogenetic Distance (MPD)
-Mean Nearest Taxon Distance (MNTD)

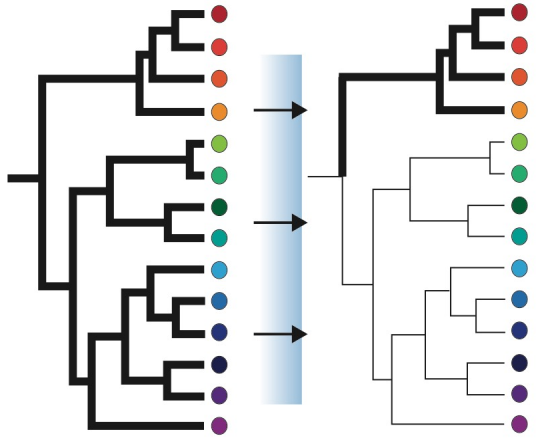


Assumption: functional traits of importance are phylogenetically conserved.

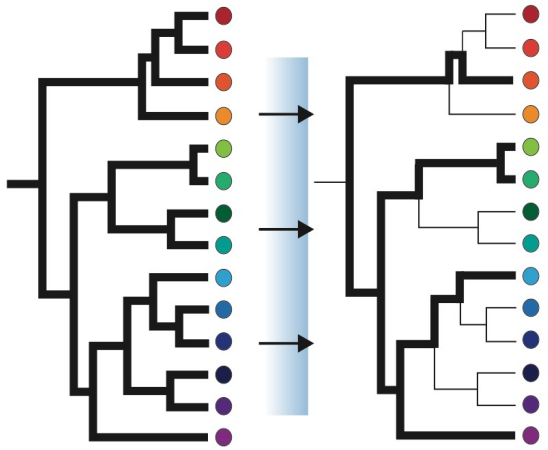
Introduction: Community Assembly

Traits are phylogenetically conserved

Habitat Filtering = phylogenetic underdispersion

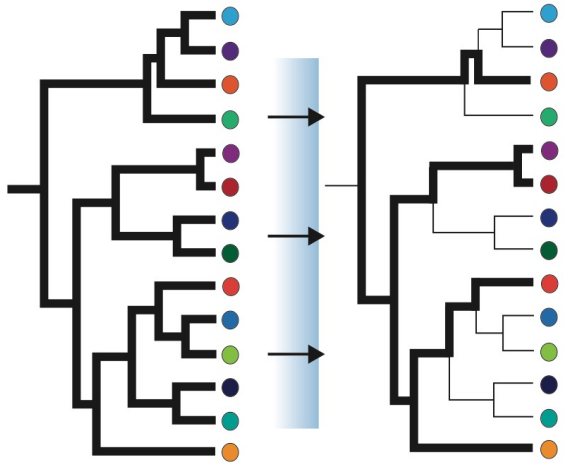


Competitive Exclusion = phylogenetic overdispersion

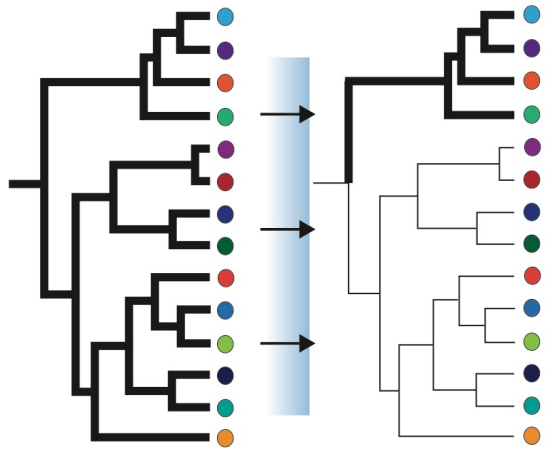


Traits are phylogenetically convergent

Habitat Filtering = phylogenetic overdispersion



Competitive Exclusion = phylogenetic underdispersion



Cavender-Bares et al. 2009

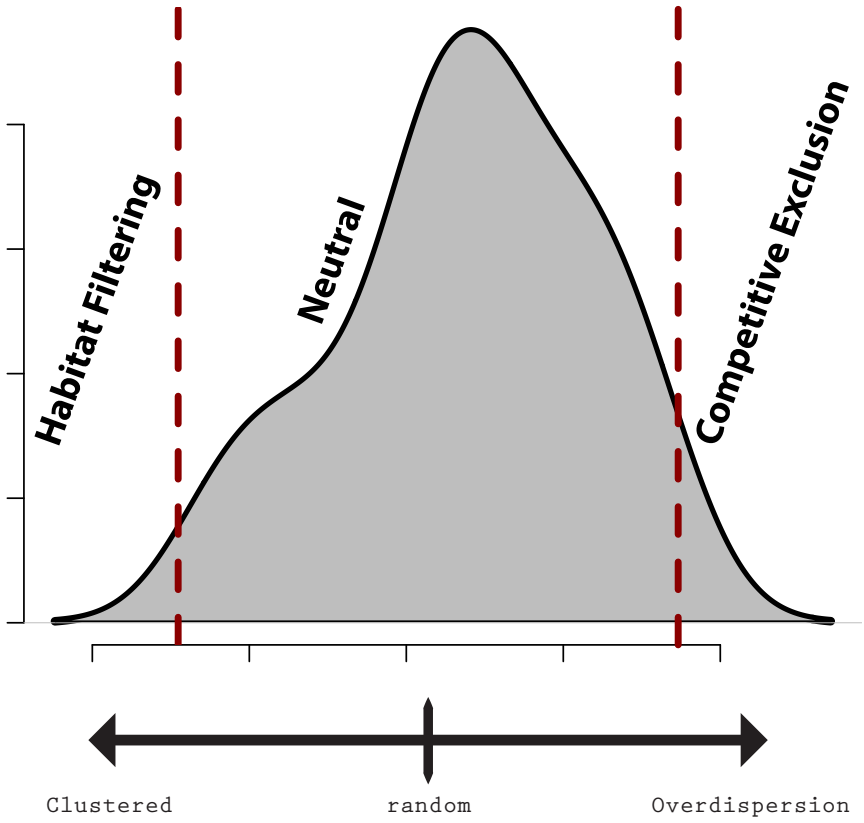
Community Assembly Inference; dispersion metrics

- Phylogenetic Community Structure

  Grandcolas 1998, Webb et al. 2000, 2002, Kraft et al. 2007

- Trait Community Structure

  Weiher et al. 1999, McGill et al. 2006, Cornwell et al. 2006, Kraft et al. 2007



Community Assembly Inference; dispersion metrics

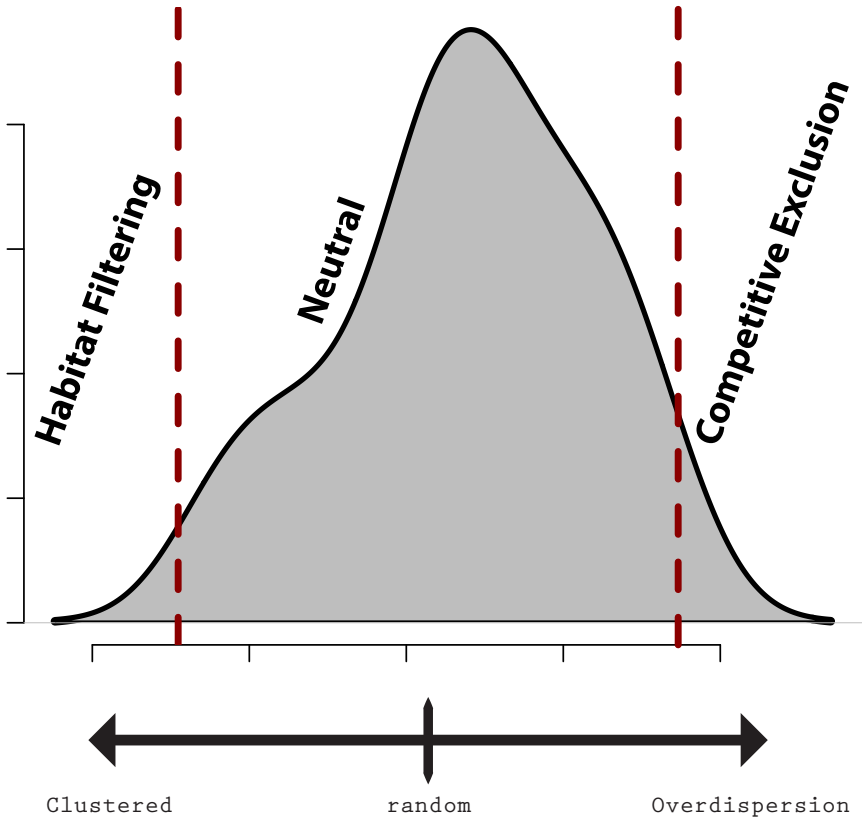
- Phylogenetic Community Structure

  Grandcolas 1998, Webb et al. 2000, 2002, Kraft et al. 2007

- Trait Community Structure

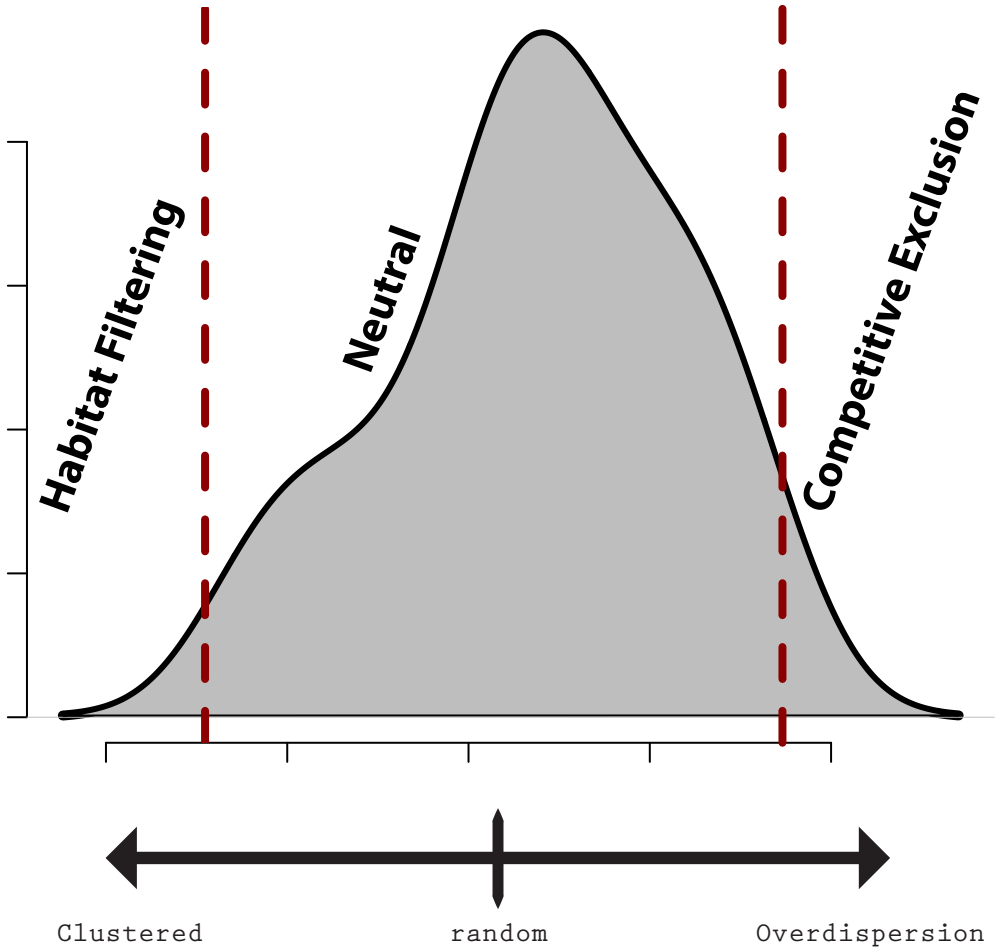
  Weiher et al. 1999, McGill et al. 2006, Cornwell et al. 2006, Kraft et al. 2007

Weiher and Keddy 1995, Kraft et al. 2007, 2010, Cavender-Bares et al. 2009; Kemble 2009; Mayfield and Levine 2010; Gerhold et al. 2015



Current Community Assembly Inference Methods

- No uncertainty associated with the inference
- Not comparable across communities



CAMI development objectives

1. *Implement a model-based inference procedure*

CAMI development objectives

1. *Implement a model-based inference procedure*
   - *Compare support for neutral and non-neutral models simultaneously*

CAMI development objectives

1. *Implement a model-based inference procedure*
   - *Compare support for neutral and non-neutral models simultaneously*

   ***Cannot write down a likelihood, so we will need to use an approximate approach based on simulations*

CAMI development objectives

1. *Implement a model-based inference procedure*
   - *Compare support for neutral and non-neutral models simultaneously*

   ***Cannot write down a likelihood, so we will need to use an approximate approach based on simulations*

2. *Parameterize the strength of the non-neutral assembly processes*

CAMI development objectives

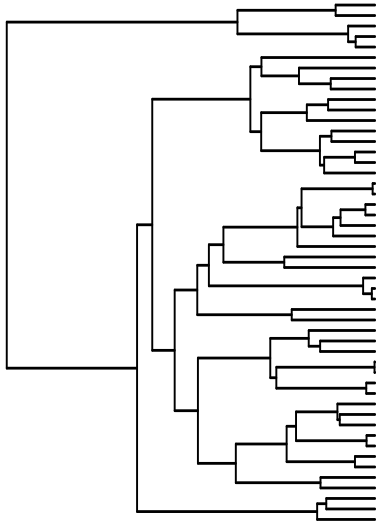
1. ***Implement a model-based inference procedure***
   - *Compare support for neutral and non-neutral models simultaneously*

   ***Cannot write down a likelihood, so we will need to use an approximate approach based on simulations*

2. *Parameterize the strength of the non-neutral assembly processes*

Objective 1: Implement a model-based inference procedure

1.1 Regional community phylogeny: N, λ, μ

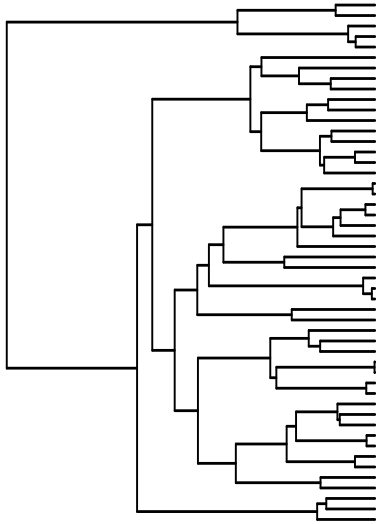


N	regional species pool
λ	speciation rate
μ	extinction rate
σ^2	rate of character change
α	strength of constraints (OU only)
n	local species pool
t_E	effect of environmental filtering
t_C	effect of competitive exclusions

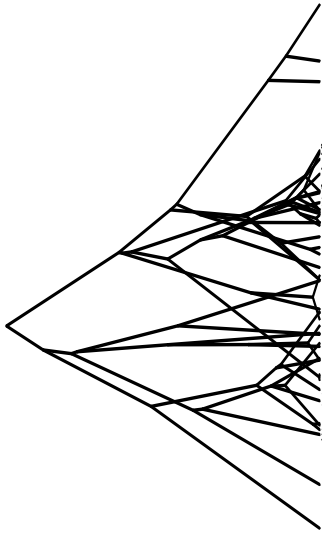
Ruffley *et al.* 2019 *Ecol. & Evol. Accepted*

Objective 1: Implement a model-based inference procedure

1.1 Regional community phylogeny: N, λ, μ



1.2 Trait evolution: σ^2, α

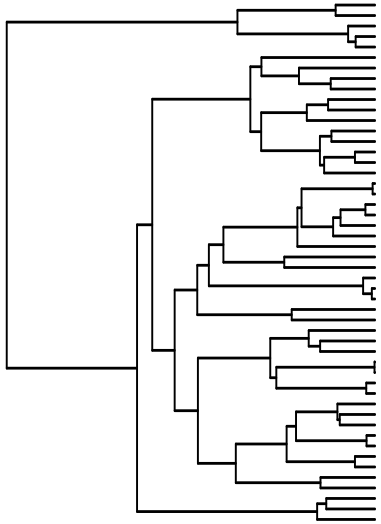


N	regional species pool
λ	speciation rate
μ	extinction rate
σ^2	rate of character change
α	strength of constraints (OU only)
n	local species pool
t_E	effect of environmental filtering
t_C	effect of competitive exclusions

Ruffley *et al.* 2019 *Ecol. & Evol. Accepted*

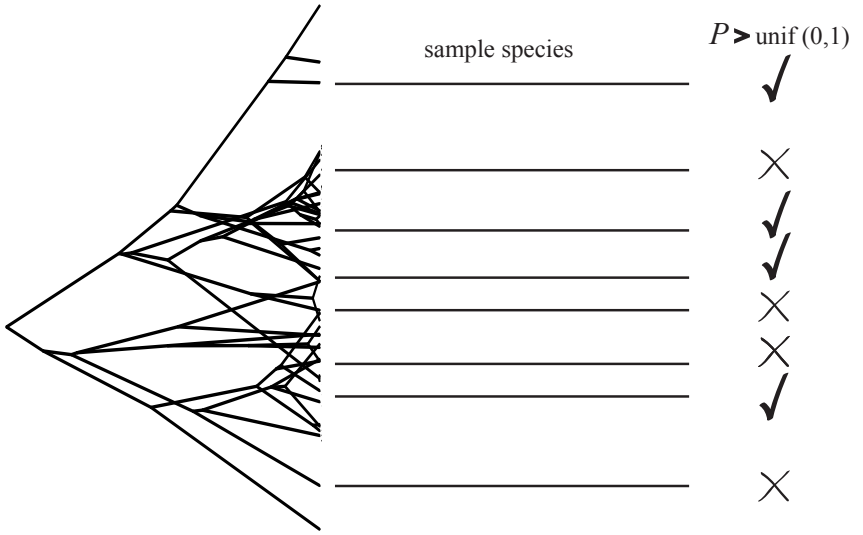
Objective 1: Implement a model-based inference procedure

1.1 Regional community phylogeny: N, λ, μ



1.2 Trait evolution: σ^2, α

1.3 Local community assembly: n, t_E, t_C



N	regional species pool
λ	speciation rate
μ	extinction rate
σ^2	rate of character change
α	strength of constraints (OU only)
n	local species pool
t_E	effect of environmental filtering
t_C	effect of competitive exclusions

- Neutral
- Environmental Filtering
- Competitive Exclusion

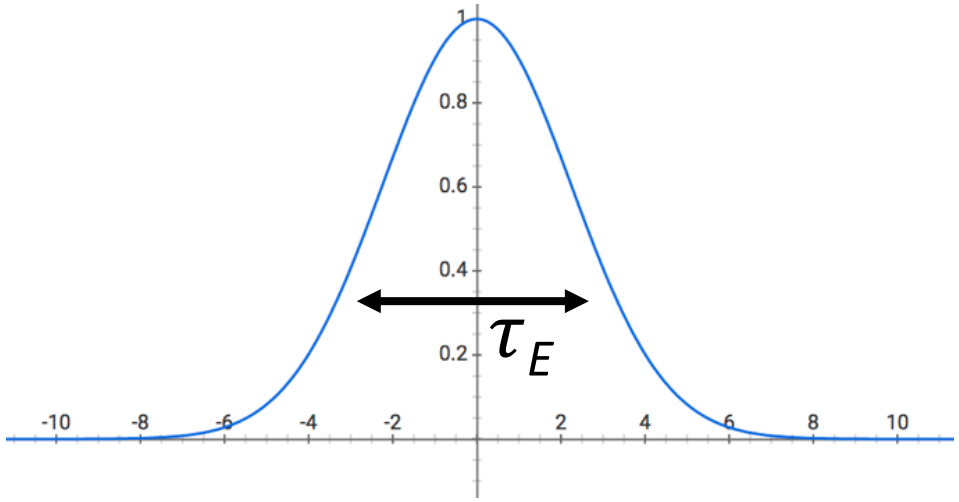
Ruffley *et al.* 2019 *Ecol. & Evol. Accepted*

Objective 2: Parameterize the strength of the non-neutral assembly processes

Objective 2: Parameterize the strength of the non-neutral assembly processes

Environmental Filtering

$$P = e^{-\frac{(Zi-Zj)^2}{\tau}}$$

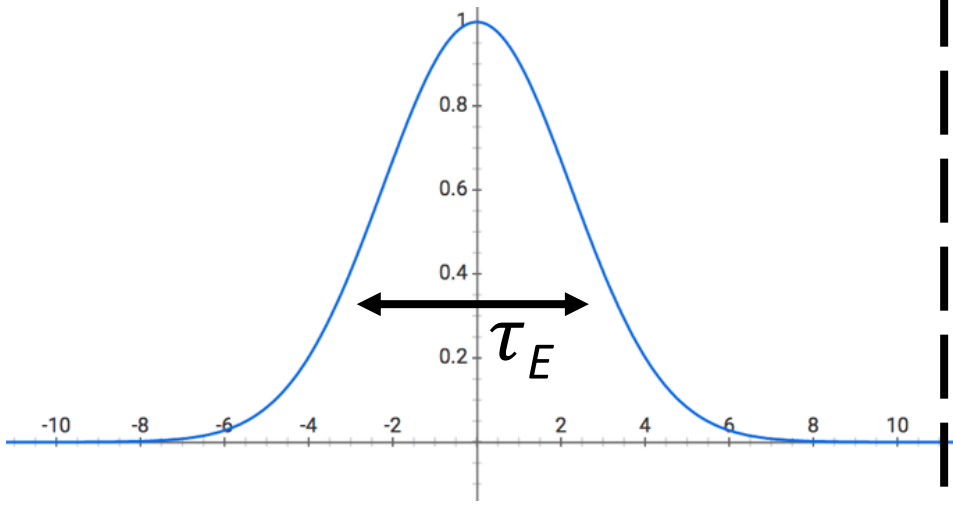


Ruffley *et al.* 2019 *Ecol. & Evol. Accepted*

Objective 2: Parameterize the strength of the non-neutral assembly processes

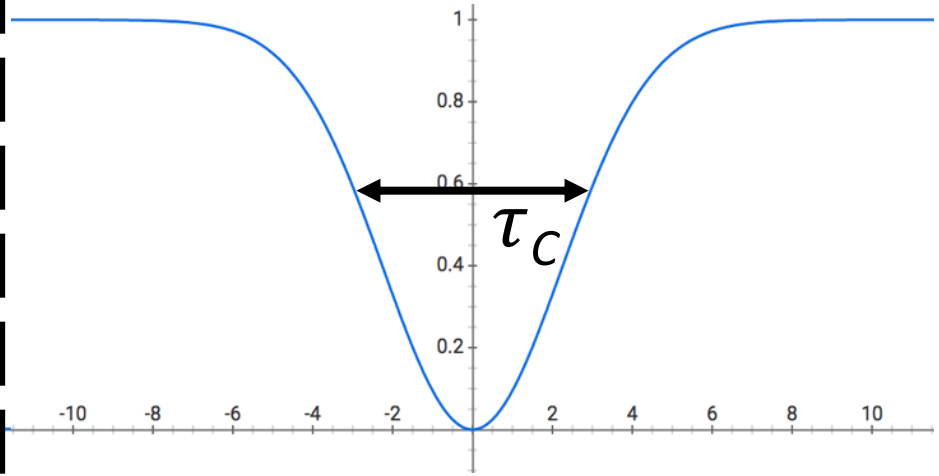
Environmental Filtering

$$P = e^{-\frac{(Zi-Zj)^2}{\tau}}$$



Competitive Exclusion

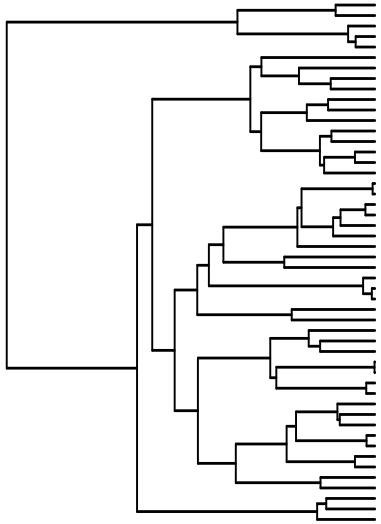
$$P = 1 - e^{-\frac{(Zi-Zj)^2}{\tau}}$$



Ruffley *et al.* 2019 *Ecol. & Evol. Accepted*

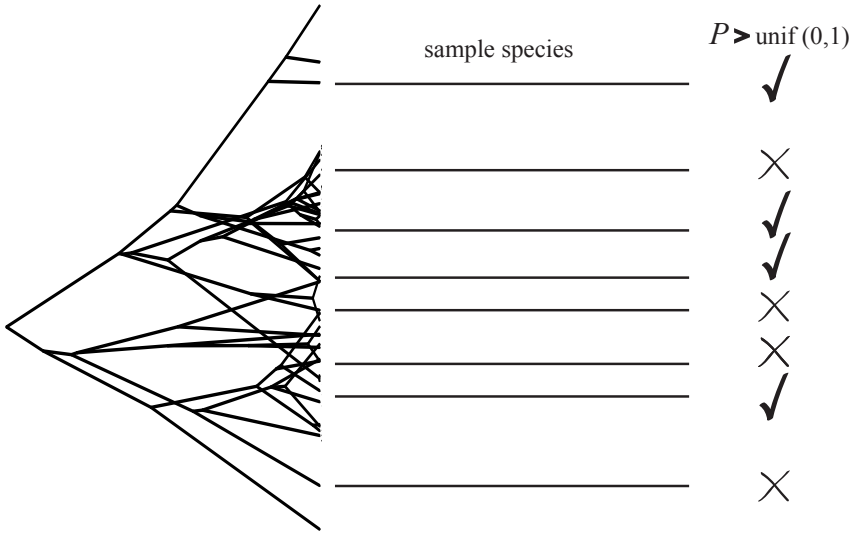
Objective 1: Implement a model-based inference procedure

1.1 Regional community phylogeny: N, λ, μ



1.2 Trait evolution: σ^2, α

1.3 Local community assembly: n, t_E, t_C

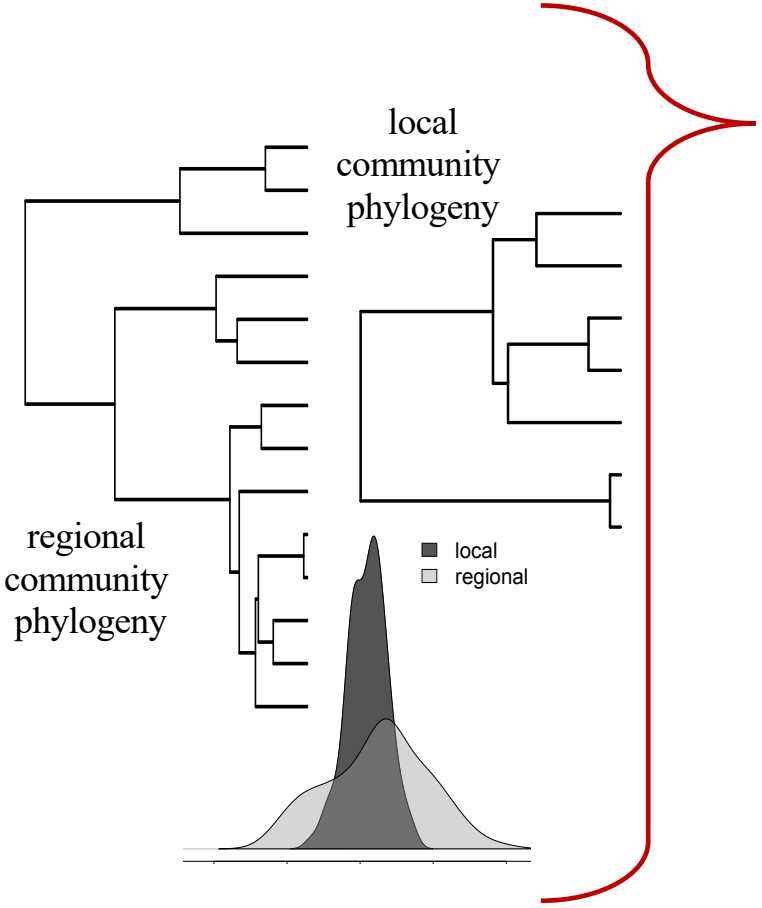


N	regional species pool
λ	speciation rate
μ	extinction rate
σ^2	rate of character change
α	strength of constraints (OU only)
n	local species pool
t_E	effect of environmental filtering
t_C	effect of competitive exclusions

- Neutral
- Environmental Filtering
- Competitive Exclusion

Ruffley *et al.* 2019 *Ecol. & Evol. Accepted*

Objective 1: Implement a model-based inference procedure



Simulation #	*SS.1*	*SS.2*	*SS.3*	*…..*	*SS.30*	*Model*
1	29.3201204	0.047795881	2.202369339	0.024039567	-0.551807157	**Neutral**
2	58.01018705	0.05683677	4.296006276	0.014001377	-0.141285853	**Neutral**
3	28.35028897	0.057748823	2.565336481	0.02472052	0.543737366	**Neutral**
4	54.4519145	0.058161893	4.462506196	0.006577184	-0.297716652	**Neutral**
5	29.52415799	0.0449925	1.332280042	-0.027307941	-0.424597011	**Neutral**
6	13.8002455	0.043015337	3.827085961	0.052868925	0.00038215	**Filtering**
7	68.51513963	0.05221236	3.773466742	0.006122832	-0.373019012	**Filtering**
8	6.409247373	0.067539483	1.27370363	0.056550081	-0.419361409	**Filtering**
9	7.869379132	0.044893904	1.621625795	0.005944114	-0.225941885	**Filtering**
10	56.74215972	0.042376631	3.870139564	-0.010551762	1.612683176	**Filtering**
11	32.21489305	0.039088122	4.436192932	-0.005017495	-0.608370405	**Competition**
12	35.31406132	0.055897014	1.802319164	0.00665565	0.575928644	**Competition**
…	15.17747744	0.052302964	2.833102254	-0.007315425	-0.343133625	**Competition**
10,000	176.8646799	0.046759559	3.404803136	0.003965272	-0.492954288	**Competition**

Ruffley *et al.* 2019 *Ecol. & Evol. Accepted*

Objective 1: Implement a model-based inference procedure

Model-selection using approximate approaches

- randomForests (*Breiman 2001, Breiman and Cutler 2007*)
- Approximate Bayesian Computation (ABC; *Beaumont, et al. 2002, Csilléry, et al. 2012)*

Simulation #	*SS.1*	*SS.2*	*SS.3*	*…..*	*SS.30*	*Model*
1	29.3201204	0.047795881	2.202369339	0.024039567	-0.551807157	**Neutral**
2	58.01018705	0.05683677	4.296006276	0.014001377	-0.141285853	**Neutral**
3	28.35028897	0.057748823	2.565336481	0.02472052	0.543737366	**Neutral**
4	54.4519145	0.058161893	4.462506196	0.006577184	-0.297716652	**Neutral**
5	29.52415799	0.0449925	1.332280042	-0.027307941	-0.424597011	**Neutral**
6	13.8002455	0.043015337	3.827085961	0.052868925	0.00038215	**Filtering**
7	68.51513963	0.05221236	3.773466742	0.006122832	-0.373019012	**Filtering**
8	6.409247373	0.067539483	1.27370363	0.056550081	-0.419361409	**Filtering**
9	7.869379132	0.044893904	1.621625795	0.005944114	-0.225941885	**Filtering**
10	56.74215972	0.042376631	3.870139564	-0.010551762	1.612683176	**Filtering**
11	32.21489305	0.039088122	4.436192932	-0.005017495	-0.608370405	**Competition**
12	35.31406132	0.055897014	1.802319164	0.00665565	0.575928644	**Competition**
…	15.17747744	0.052302964	2.833102254	-0.007315425	-0.343133625	**Competition**
10,000	176.8646799	0.046759559	3.404803136	0.003965272	-0.492954288	**Competition**

Ruffley *et al.* 2019 *Ecol. & Evol. Accepted*

Results

1. *Power Analysis*

2. *Parameter Estimation*

3. *Empirical study*

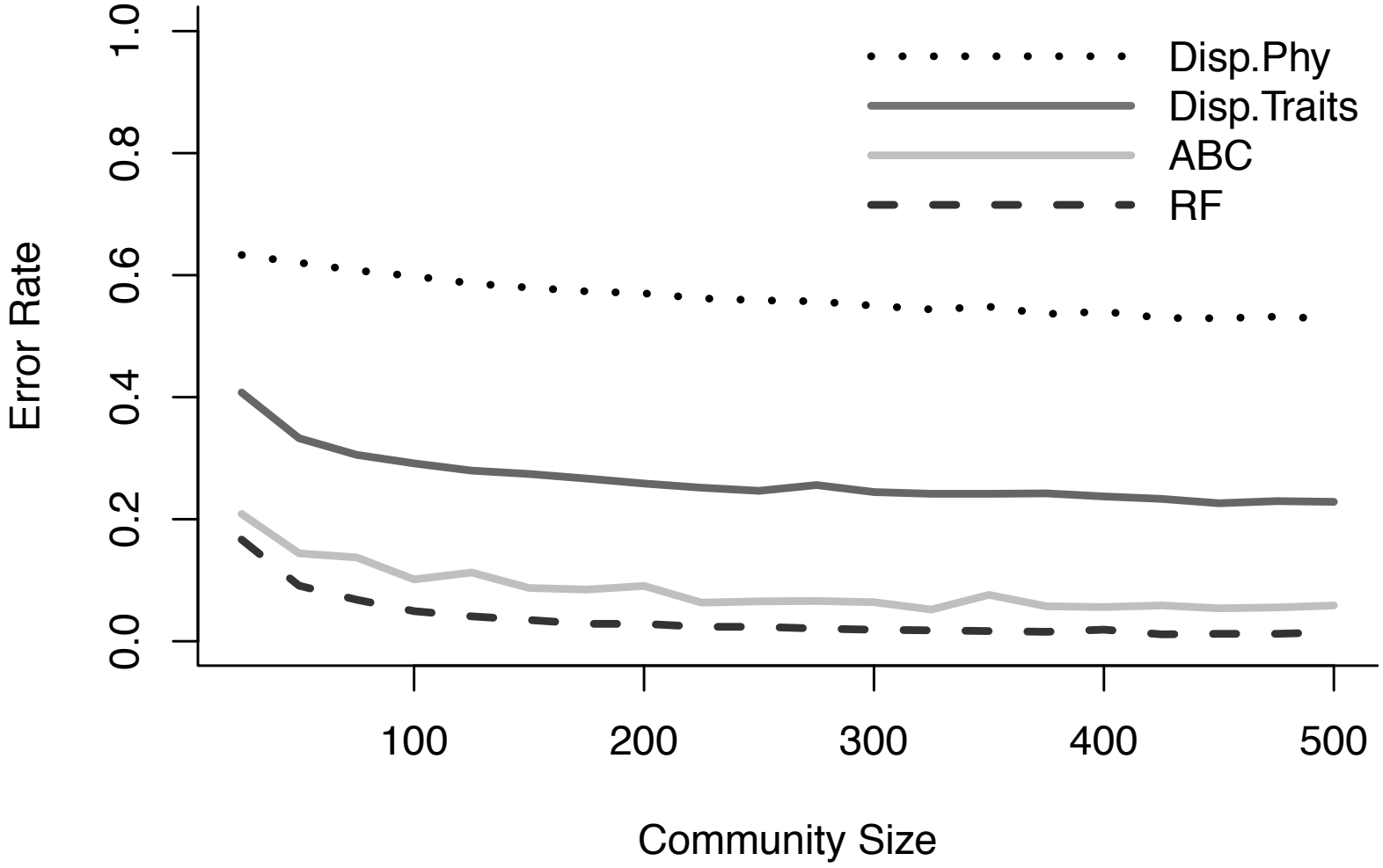
Results

1. Power Analysis

- ***Dispersion Metrics***
  - *Phylogenetic data*
  - *Functional trait data*

- ***randomForest (RF)***
- ***ABC***

Results: Power Analysis



Ruffley *et al.* 2019 *Ecol. & Evol. Accepted*

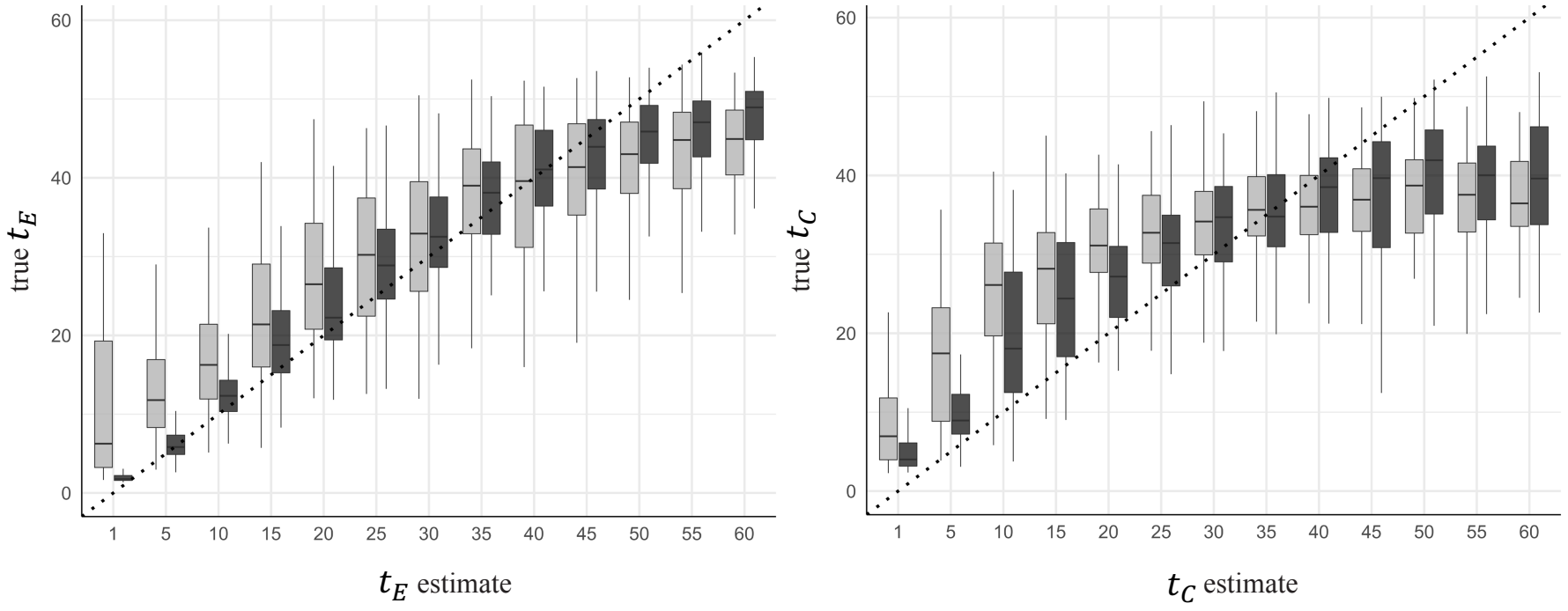
Results

1. *Power Analysis*

2. ***Parameter Estimation***
   -Estimate τ_E and t_C

3. *Empirical studies*

Results: Parameter Estimation



Ruffley *et al.* 2019 *Ecol. & Evol. Accepted*

- DO first tutorial and introduction to CAMI,
    - Load it
    - Simlate some data
    - Build and RF