

Phylogenetic diversity–ecosystem function  
relationships are insensitive to phylogenetic edge  
lengths, Marc W. Cadotte

Paper discussion by  
Karen Bisschop & Richel Bilderbeek



*Marc W. Cadotte*

# Previous research

- Marc W. Cadotte, “Experimental evidence that evolutionarily diverse assemblages result in higher productivity”, 2013, PNAS
- Research question: What is the relationship between genetic diversity and plant community biomass production? And which measure of genetic diversity is the best predictor of biomass production?



VS

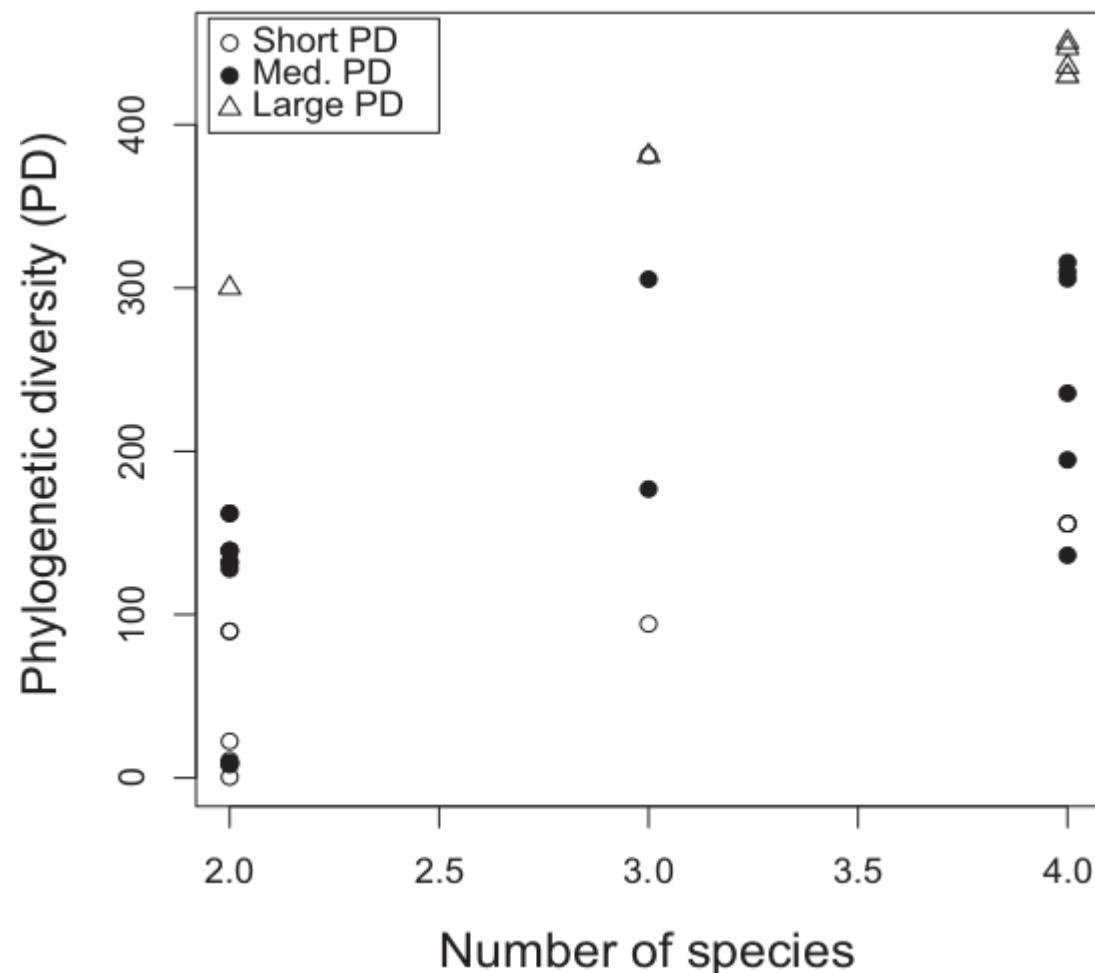


# Species used



# Previous research: experiment

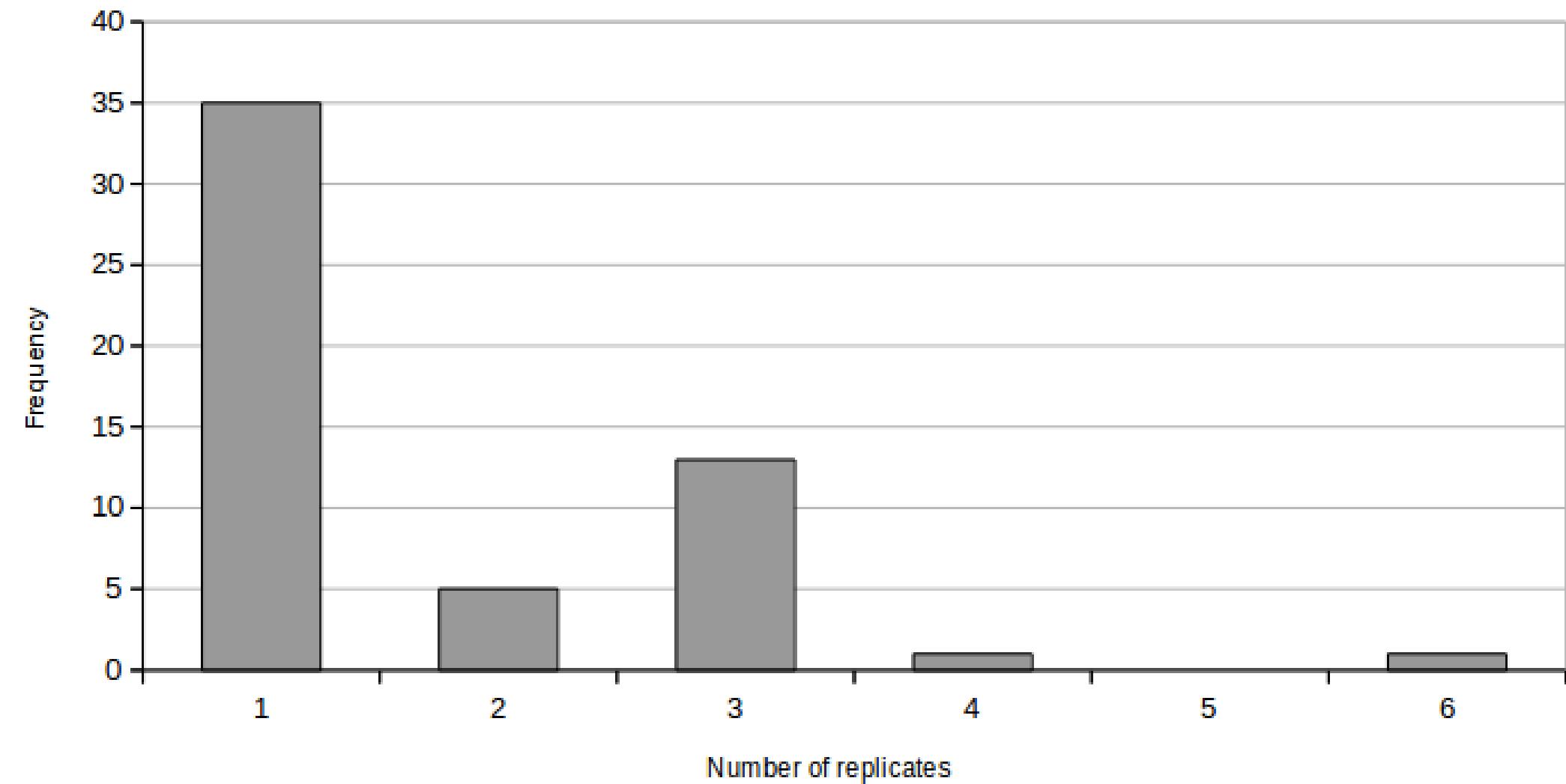
- Created 91 2x2 m plots with different plant communities of 1-4 species out of 14 species with 54 different genetic diversities
- Each community was replicated 1-6 times
- Measured above-ground biomass 2 years later



Cadotte, 2013

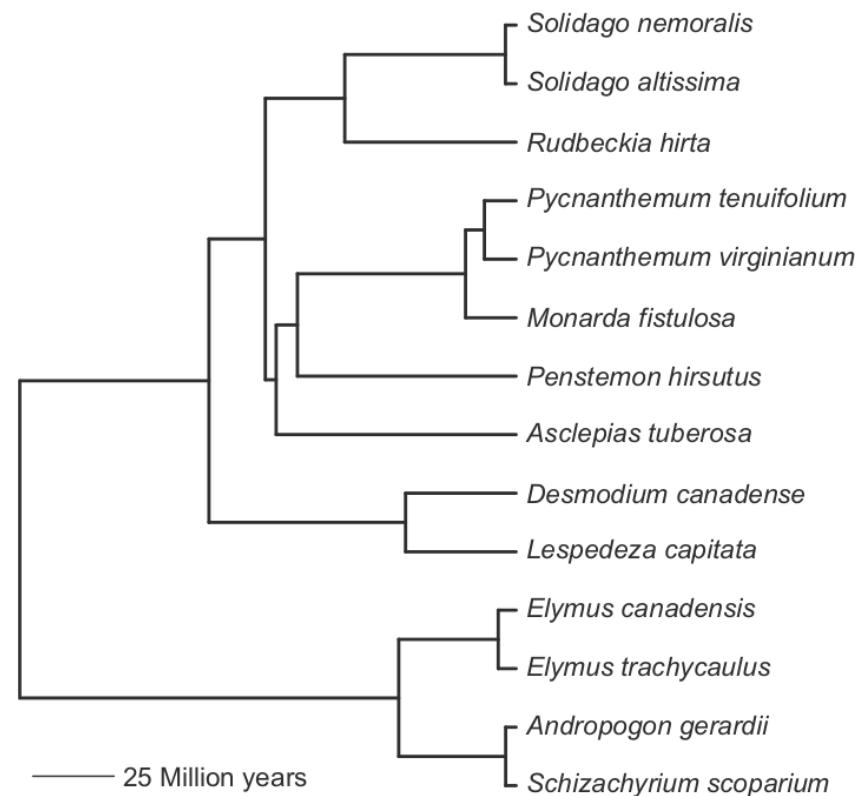
# Number of replicates

Number of replicates per community



# Previous research: phylogeny

- Use 1-5 gene sequences
- Sequence **alignment** with MUSCLE
- In **phylogeny construction**, use angiosperm lineages as outgroup and use Modeltest



Cadotte, 2013

# Genes used

Species	Gene				
	rbcl	matk	5.8s	ITS1	ITS2
<i>Monarda fistulosa</i>	Z37419	GU381745	AY943493	AY943493	AY943493
<i>Pycnanthemum virginianum</i>			DQ667319	DQ667319	DQ667319
<i>Pycnanthemum tenuifolium</i>		GU381750	GU381423	GU381423	GU381423
<i>Desmodium canadense</i>	HQ590061	HQ593266	JN407461	JN407461	JN407461
<i>Lespedeza capitata</i>			FJ409474	FJ409474	FJ409474
<i>Solidago nemoralis</i>	EU677028	EU749420	EU125357	EU125357	EU125357
 <i>Solidago altissima</i>			HQ142603	HQ142603	HQ142603
<i>Doellingeria umbellata</i>			AF477625	AF477625	AF477625
<i>Rudbeckia hirta</i>	AB530973			U72811	U72811
<i>Liatris cylindracea</i>	AY816235		AY804146	AY804146	AY804146
<i>Asclepias tuberosa</i>	EF590504	GQ248084			
<i>Penstemon hirsutus</i>			DQ471685	DQ531111	DQ471685
<i>Andropogon gerardii</i>	AJ784818	AF144577	DQ005015	DQ005015	DQ005015
<i>Schizachyrium scoparium</i>	HE577863	HE586094	DQ005071	DQ005071	DQ005071
<i>Elymus canadensis</i>			EF396960	EF396960	EF396960
<i>Elymus trachycaulus</i>	Z49839	DQ78691	FJ040168	FJ040168	FJ040168

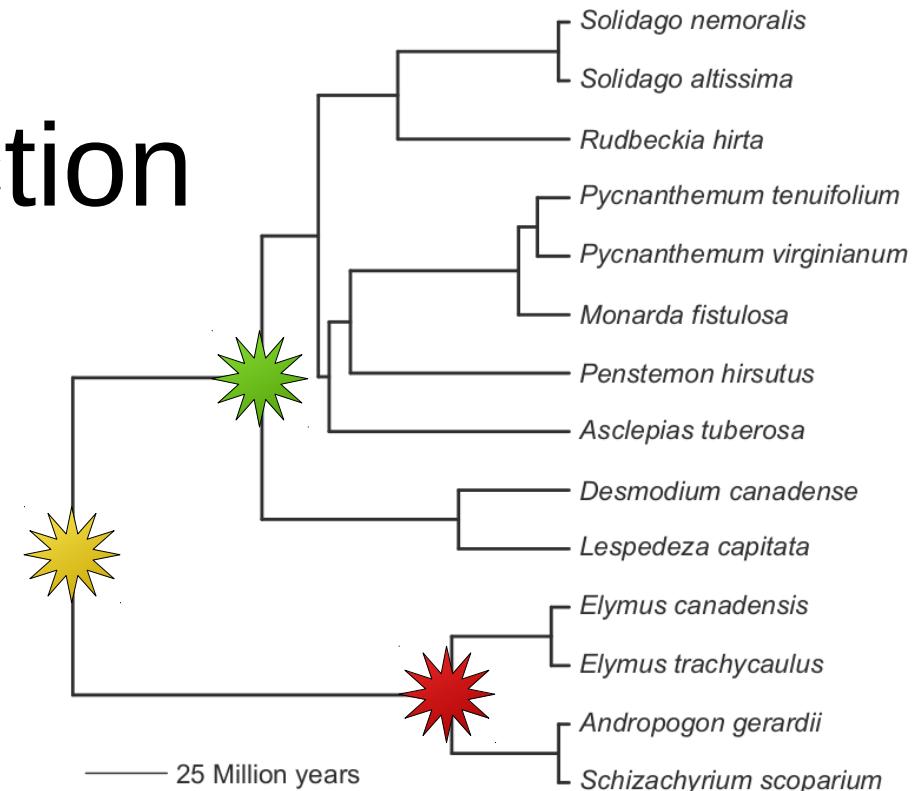
Cell codes refer to GenBank accession numbers ([www.ncbi.nlm.nih.gov/genbank](http://www.ncbi.nlm.nih.gov/genbank)). 5.8s, 5.8s ribosomal RNA; ITS1 and ITS2, internal transcribed spacers 1 and 2; matk, chloroplast maturase K gene; rbcl, ribulose-biphosphate carboxylase gene.

# Alignment

- Using MUSCLE (MUltiple Sequence Comparison by Log-Expectation)
- Comparable to CLUSTALW, MAFFT and T-COFFEE

# Phylogeny construction

- Used two other angiosperms as outgroups
- Used best-fit maximum-likelihood models of nucleotide substitution for each gene (?) using the Akaike Information Criterion with Modeltest
- (model was HKY+I+G)



Eudicots  
 Monocots  
 Angiosperms



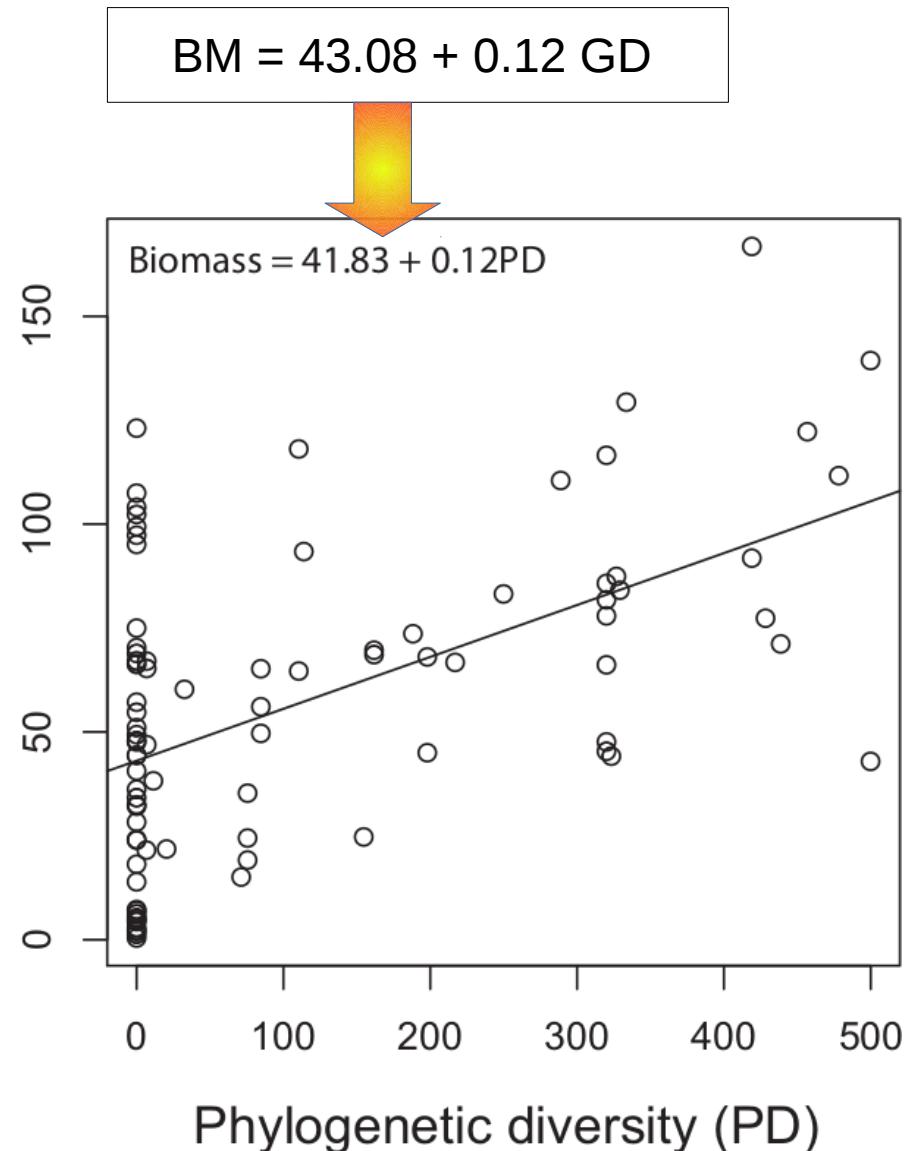
*Amborella trichopoda*  
Amborellales



*Magnolia grandiflora*  
Magnoliids

# Previous research

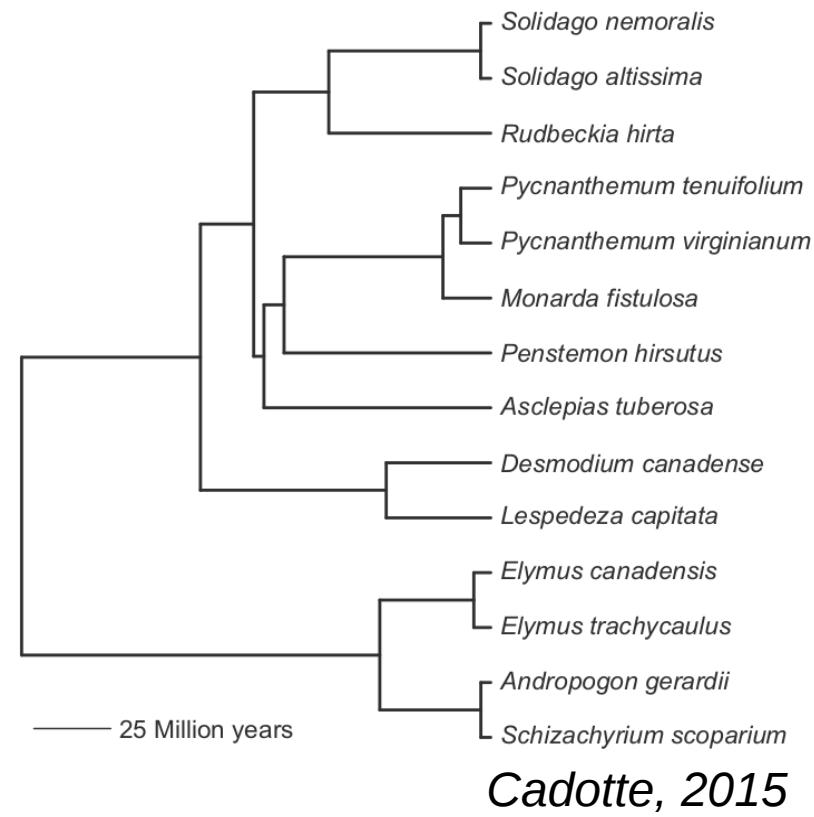
- The more genetic distance in a plot, the higher the biomass
- Genetic distance (GD) is the best predictor of biomass (BM) compared to 11 other measures of diversity



Cadotte, 2013

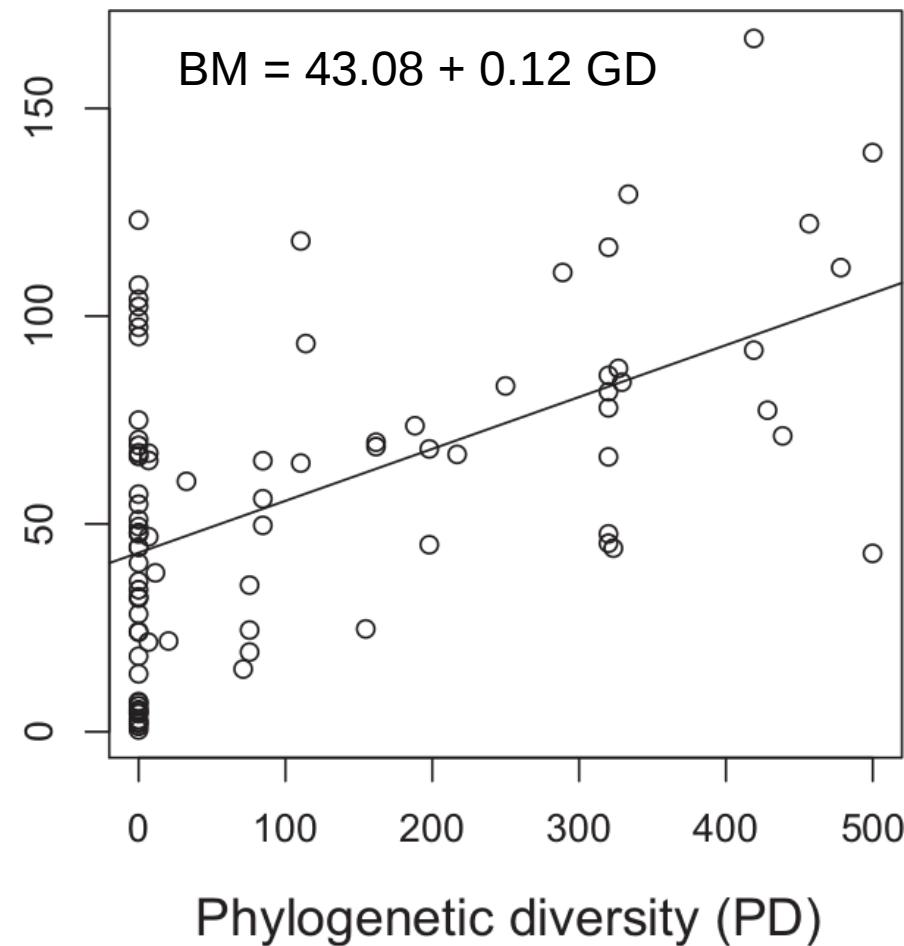
# This research

- In the previous research, constant mutation (nucleotide substitution) rate is assumed
- With different mutation rates per clade the tree will have different taxa lengths
- Does it matter if edge lengths have errors, compared to errors in topology?



# Method

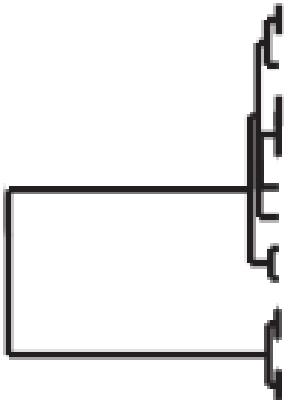
- **Randomize phylogenies**
- **Fit new PDs to real data**
- If randomized phylogenies have similar likelihoods, the data is robust, thus the treatment has little effect



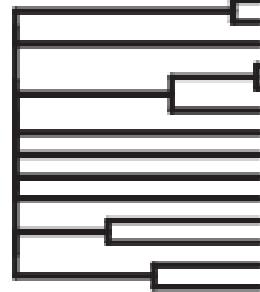
Cadotte, 2015

# Randomize

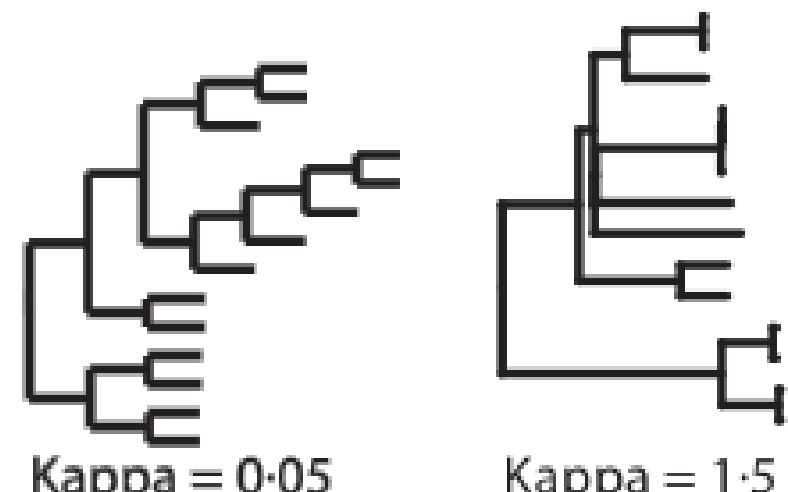
- Five treatments:
  - 1) Pagel's  $\delta$  transform on edge lengths
  - 2) Pagel's K transform on edge lengths
  - 3) Randomize edge lengths
  - 4) Randomize tips
  - 5) Randomize tips and edges



Delta = 0·1



Delta = 19



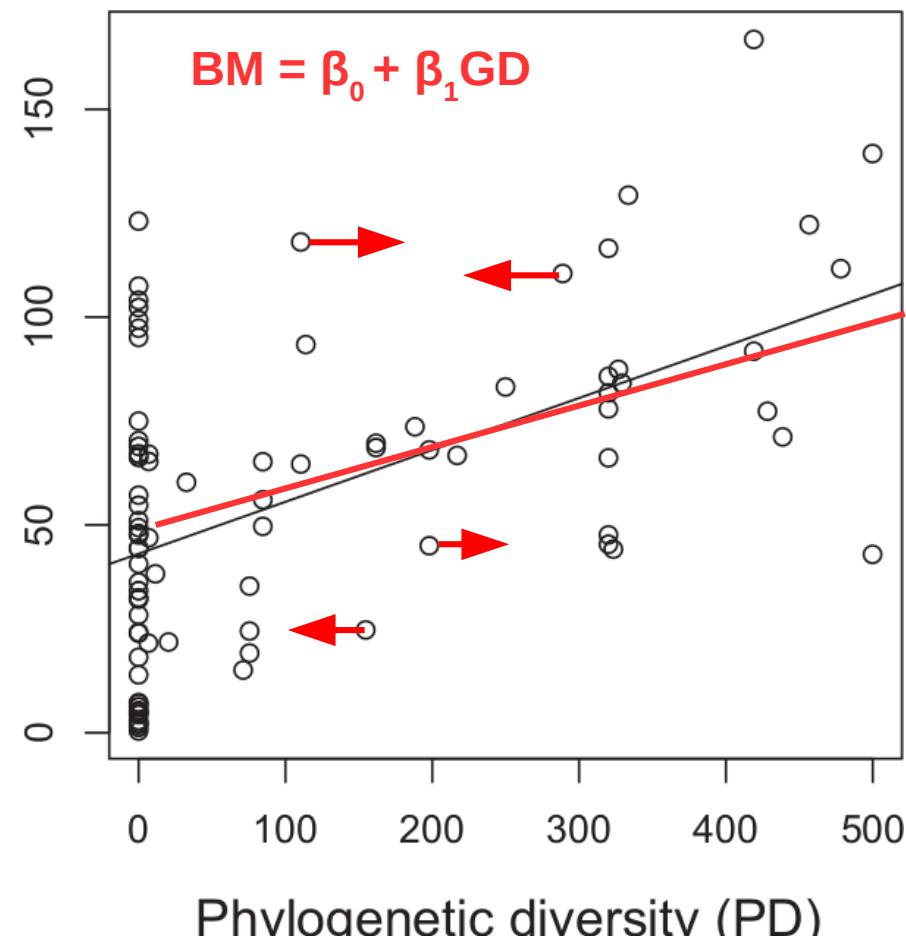
Cadotte, 2015

# Fit

- For every phylogeny
  - PD is calculated
  - New PD is fit to data with the linear model

$$EF = \beta_0 + \beta_1 PD + \varepsilon$$

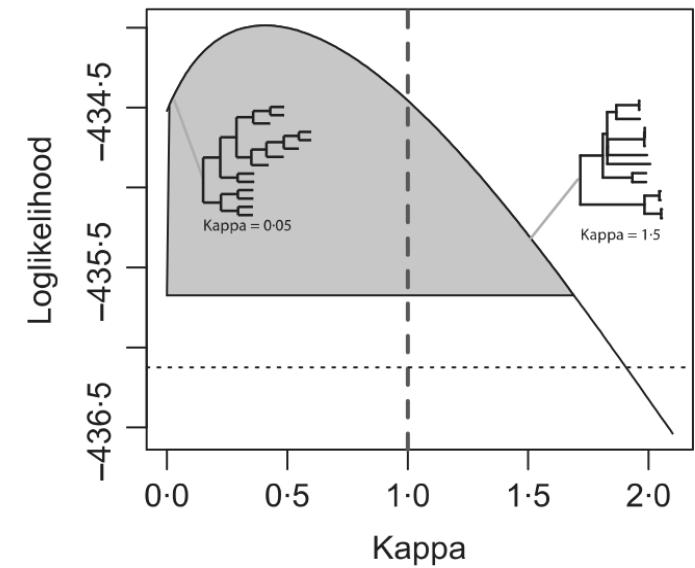
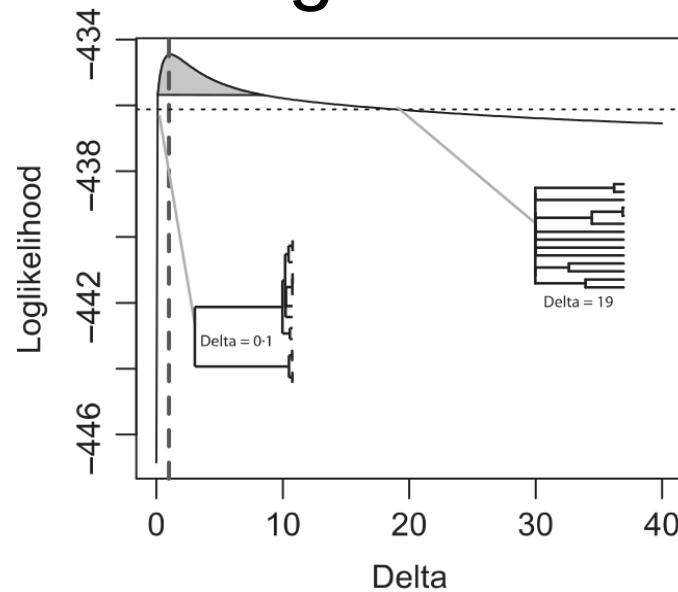
using maximum-likelihood  
(assumes error is distributed independently and identically according to a normal distribution)



Cadotte, 2015

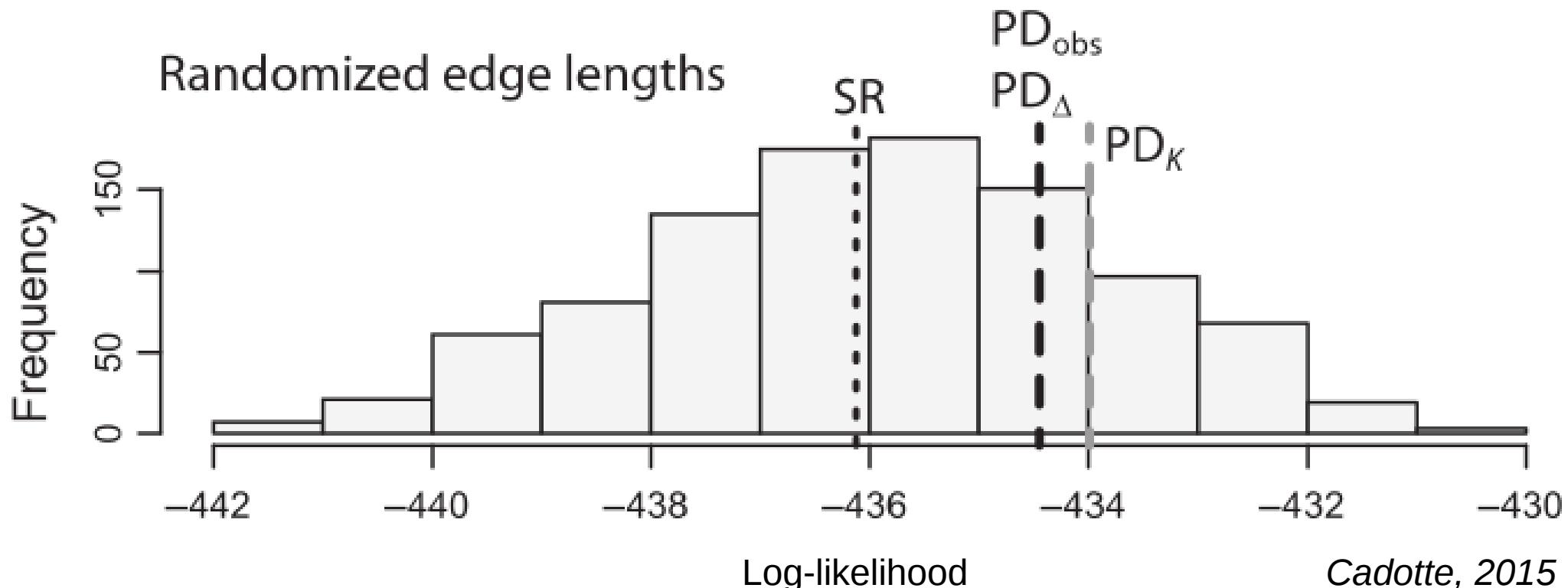
# Results

- The best Pagel's  $\delta$  transform coincidentally has highest likelihood at its original value
- The best Pagel's K transform has non-significantly different likelihood at best value compared to the original value



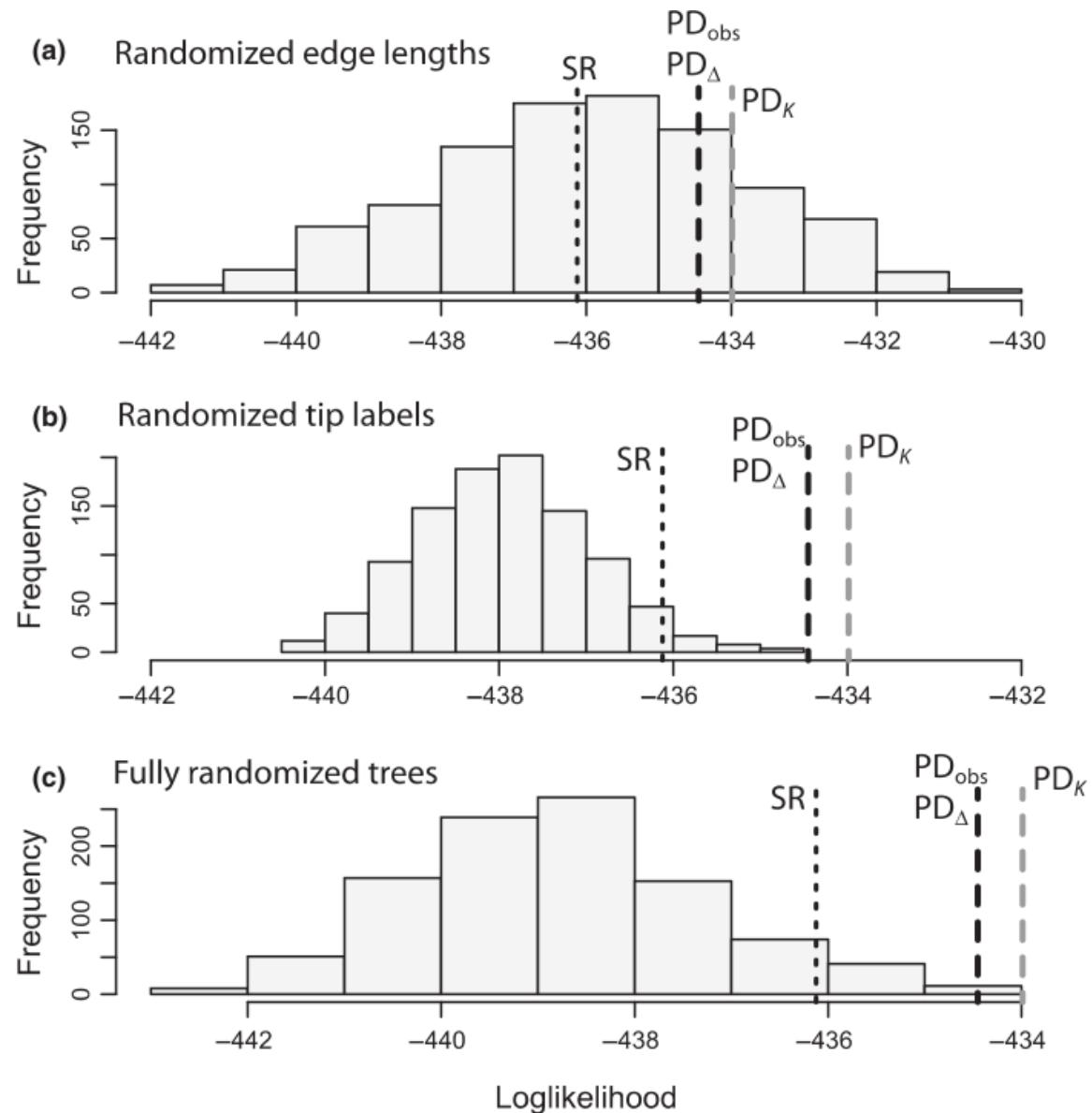
# Results

- $\text{PD}_{\text{obs}}$ : Likelihood of fit with original phylogeny
- $\text{PD}_\Delta$ : Highest likelihood of Pagel's  $\delta$  transform
- $\text{PD}_K$ : Highest likelihood of Pagel's K transform
- SR: Likelihood using species richness instead of PD



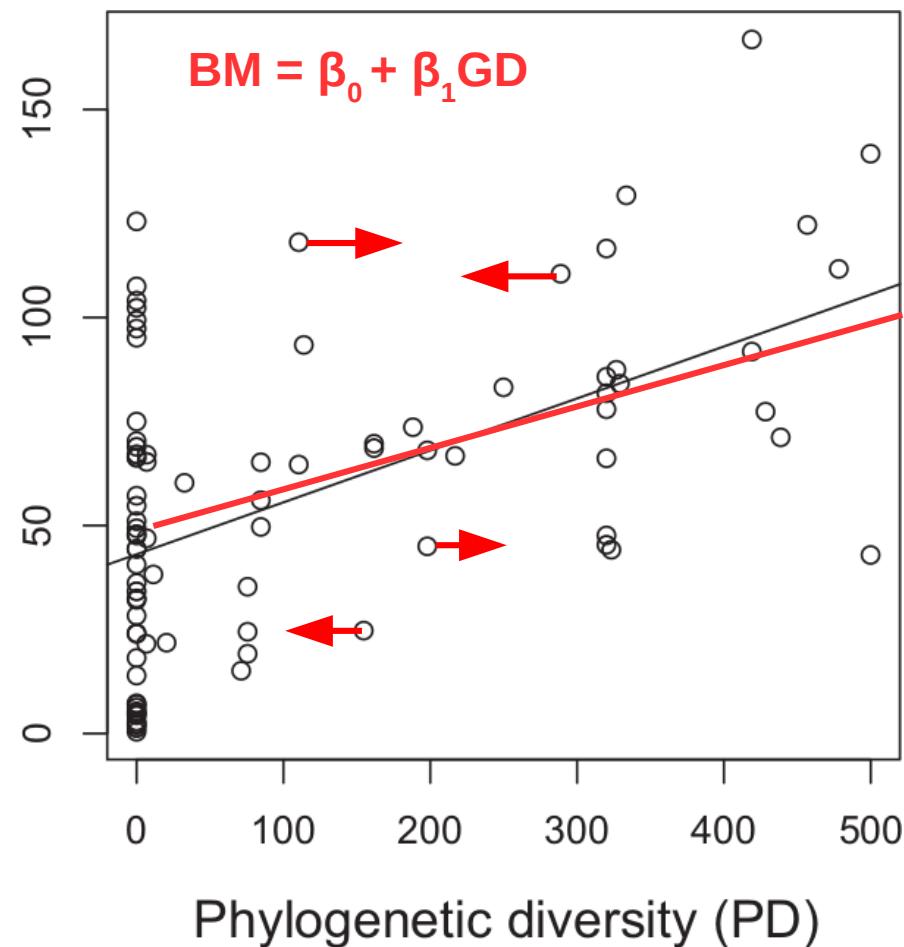
# Results

- Randomizing edge lengths has least influence, as  $\text{PD}_{\text{obs}}$  is within high frequency
- In other treatments,  $\text{PD}_{\text{obs}}$  falls outside of the random tree range



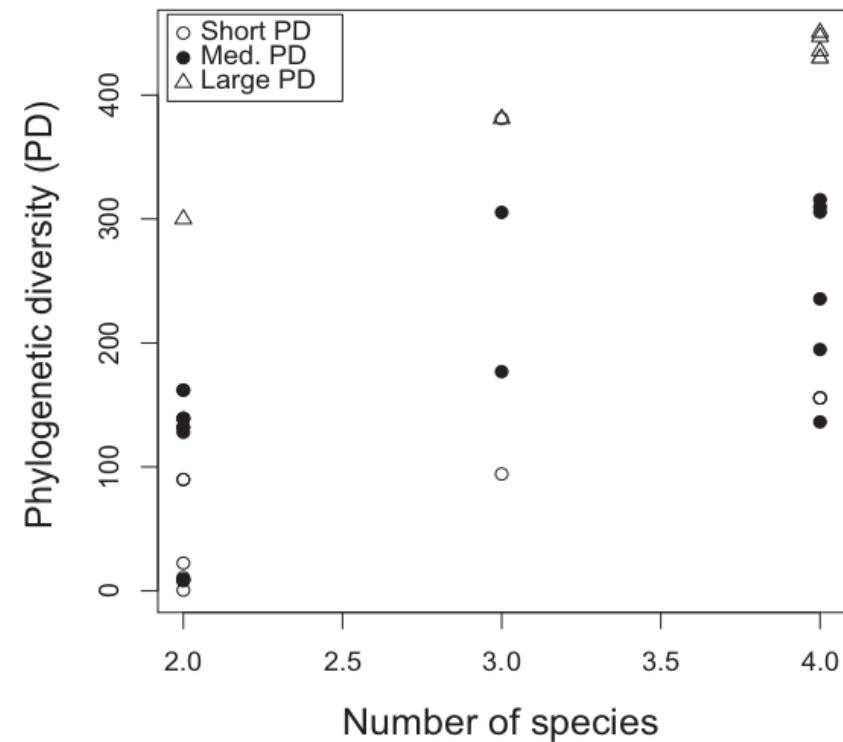
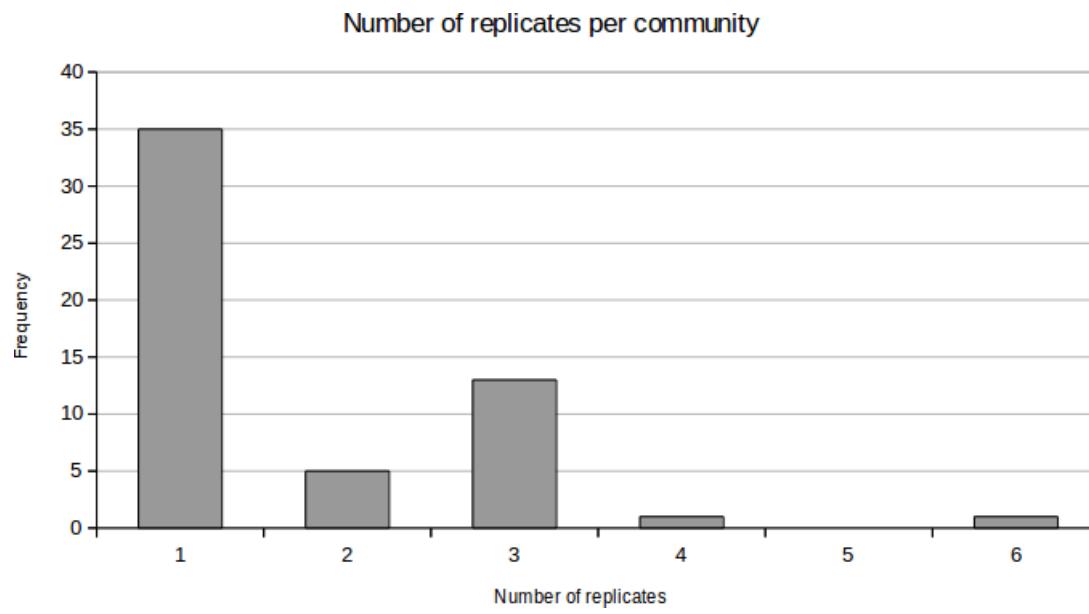
# Discussion

- Isn't it weird to use 'fitting to the data well' as a result variable?
- **Does** a better way exists?



# Discussion

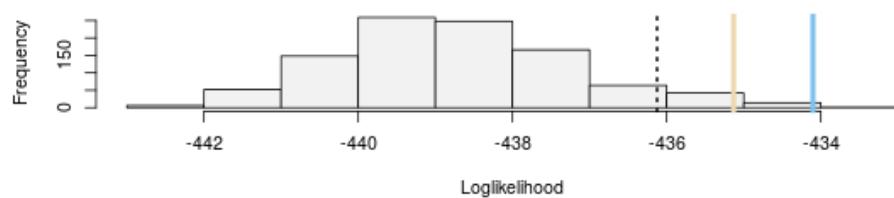
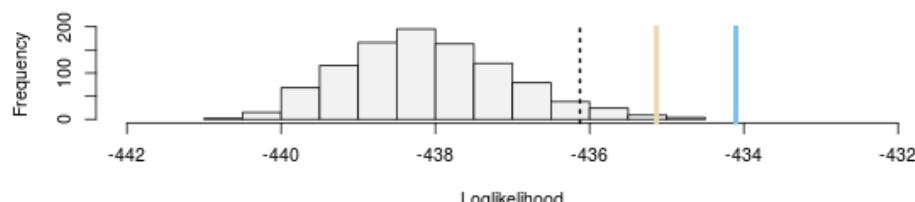
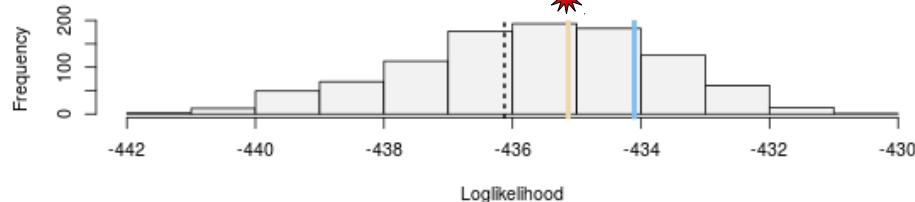
- Was the initial data of 2013 weak to begin with?
  - 35 out of 55 communities had no replicates
  - It **is** a gradient of 55 communities



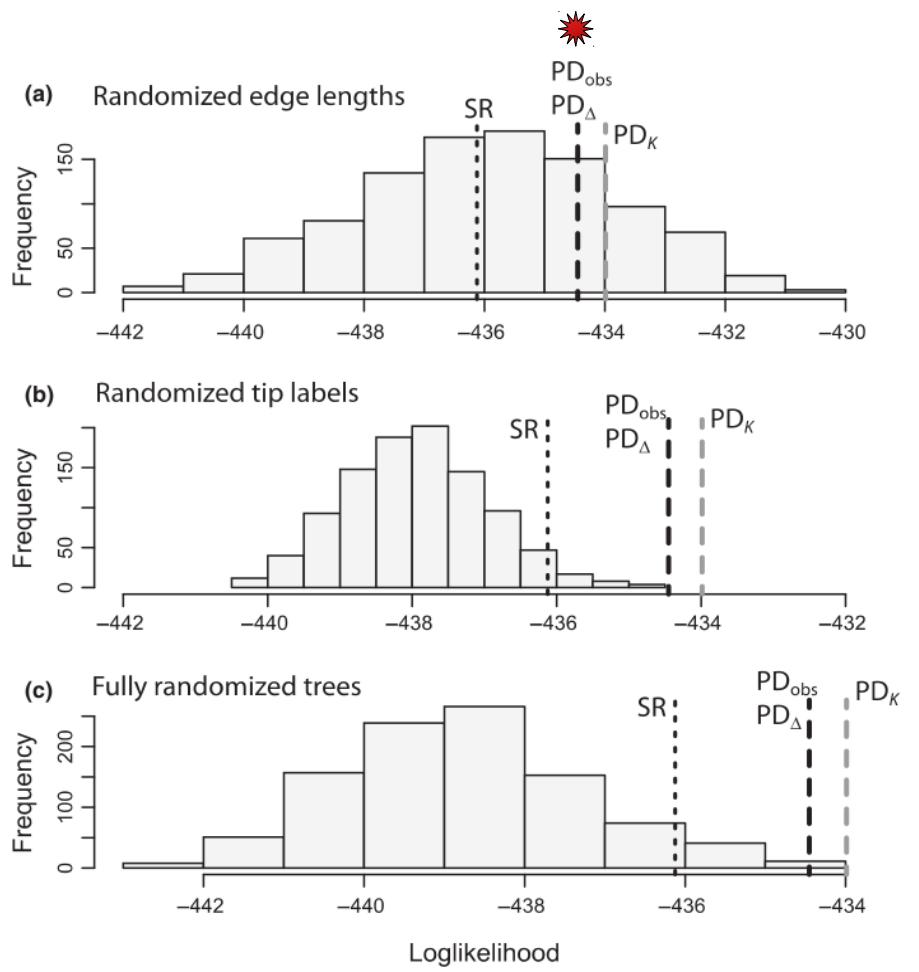
Cadotte, 2013

# Discussion

- Could not reproduce result
  - Different plot (minor detail)
  - Used random number seed



*Redone*



*Cadotte, 2015*

# Discussion

- Isn't it obvious that changing branch lengths has less impact than changing labels?
- It **is** good that this has been checked
- Should we abandon researching speciation models then?