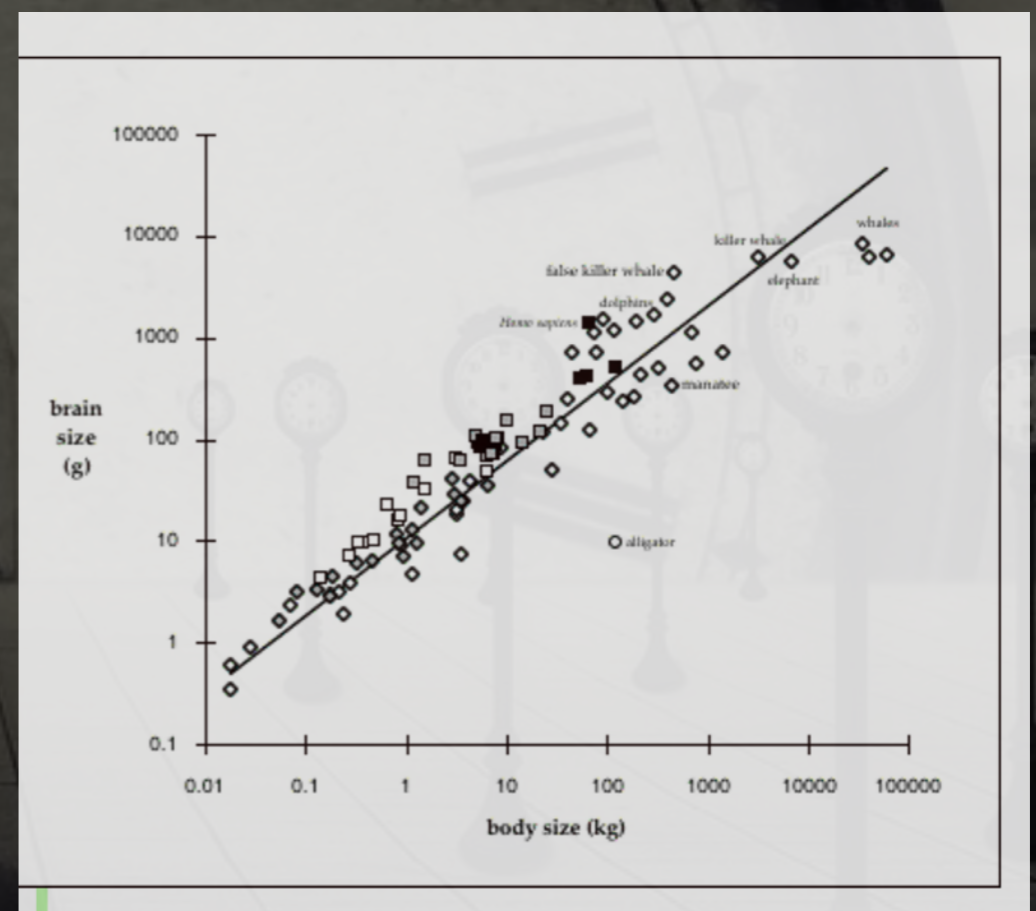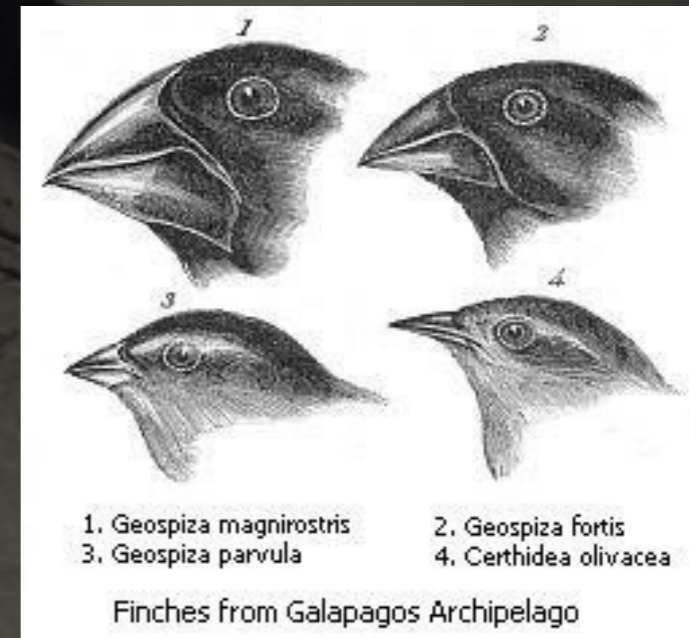# Reformulating phylogenetic mixed models to improve flexibility and speed

Michael Li
McMaster University

International Statistical Ecology Conference
2018

Cross-Species Comparative analysis

- Basic formula (phenotype vs phenotype, or phenotype vs environmental factor)

- Simplest method is do some kind of regression

- Phylogenetic comparative analysis = Basic formula + phylogenetic relationships across species

1. Geospiza magnirostris 2. Geospiza fortis
3. Geospiza parvula 4. Certhidea olivacea

Finches from Galapagos Archipelago

(Martin 1981)

Flaw of the simple regression approach

- Biological problem: independence, star phylogeny

- species are part of a hierarchically structured phylogeny
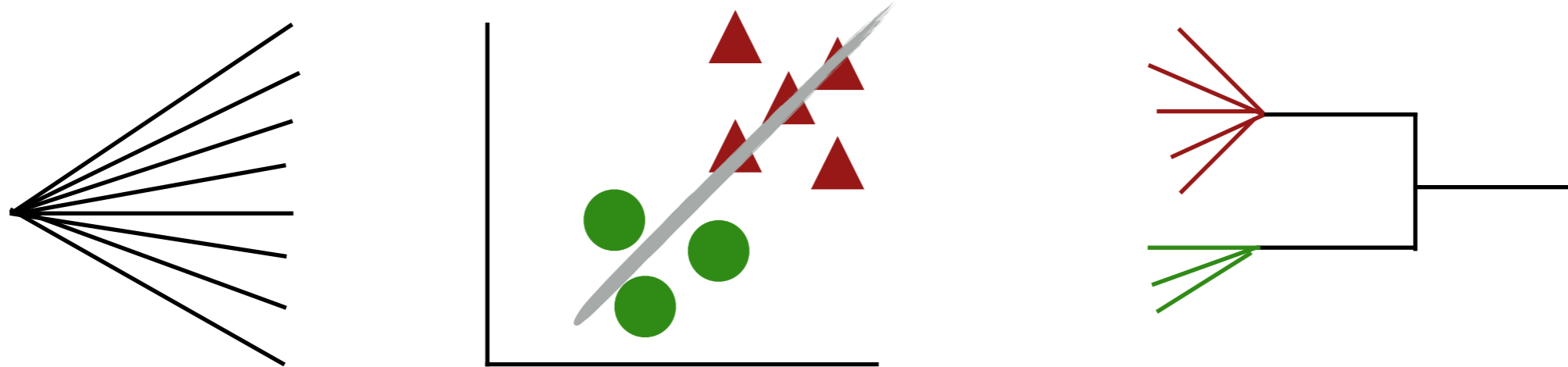
- Statistical problem: decrease accuracy and power

- inflate type I error rates and can lead to false conclusions (Li and Ives, 2017)

State of the Art of PCM

- Felsenstein's phylogenetic independent contrast (IC)

  1 parameter

- Phylogenetic generalize least square (PGLS, statistically equivalent under BM evolution assumption)

- Pagel's λ - scaling parameter

- Blomberg's K - scaled ratio

  2 parameters

- Phylogenetic generalized linear model

- Phylogenetic mixed models

  $\geq$ 2 parameters

Limitations of existing methods
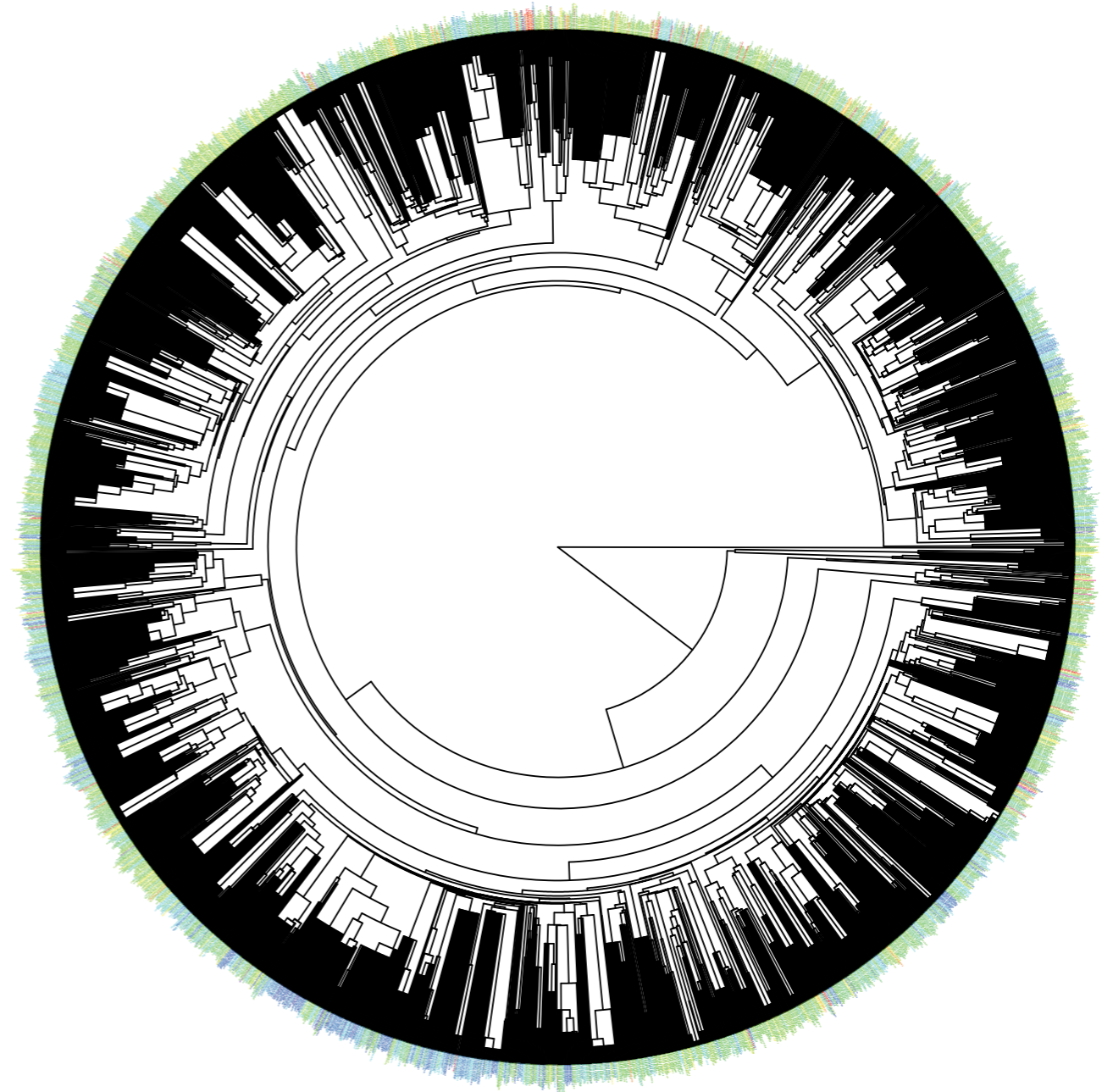
Phylogenetic signal
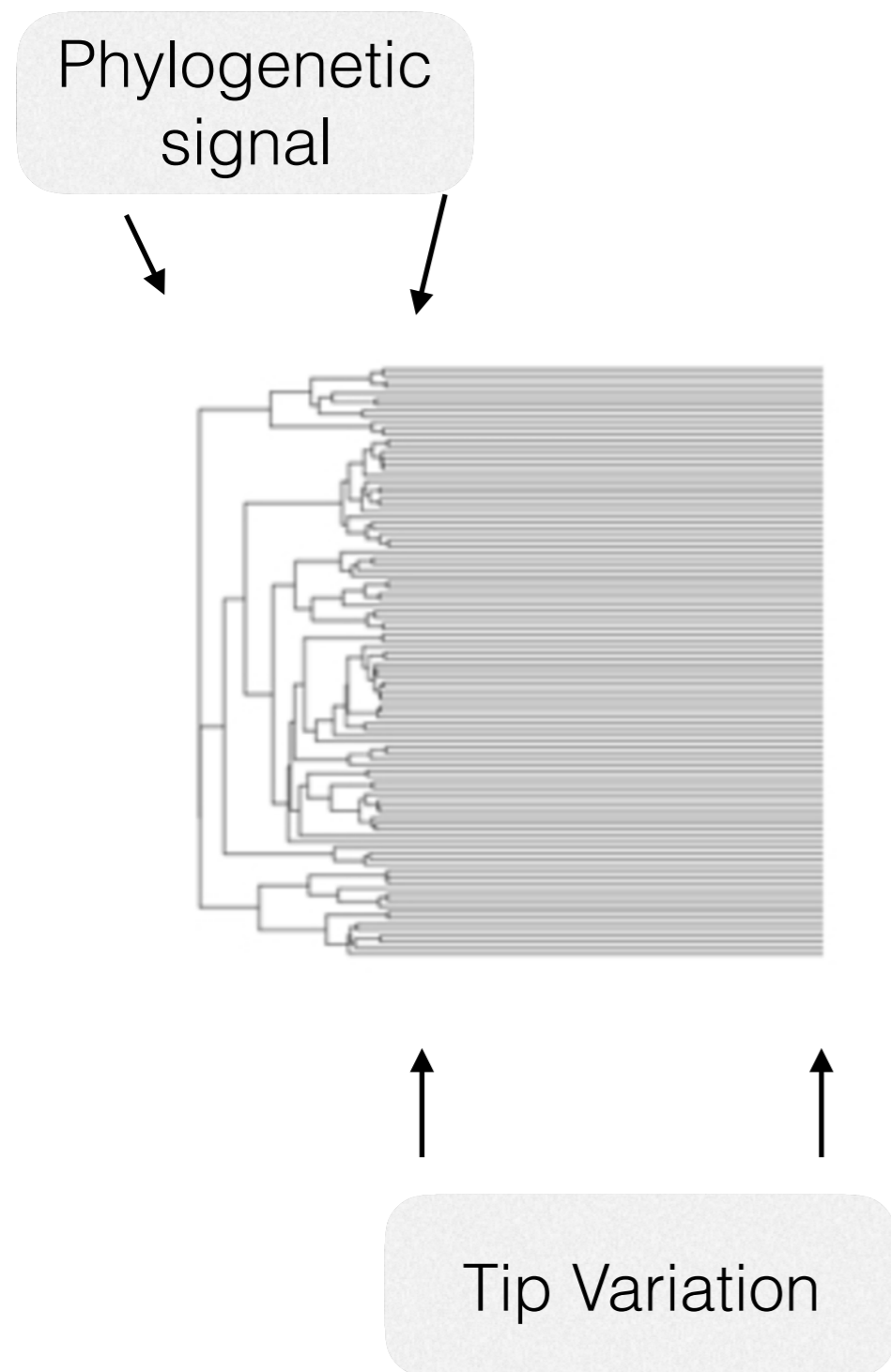
Tip Variation

- IC/PGLS assumes phylogenetic signal only

- BM phylolm, Blomberg's K, Pagel's Lambda can partition two types of variation

- What about multiple/ repeated observation of the same species?

Two challenges in linking phylogenic correlations to statistical frameworks

1. Incorporate phylogenetic relationships among species, tip variation and within species variation

2. Extend (1) to common features of statistical models, such as variation among multiple grouping variables or variation in effects (random slopes) as well as correlated residual variation.
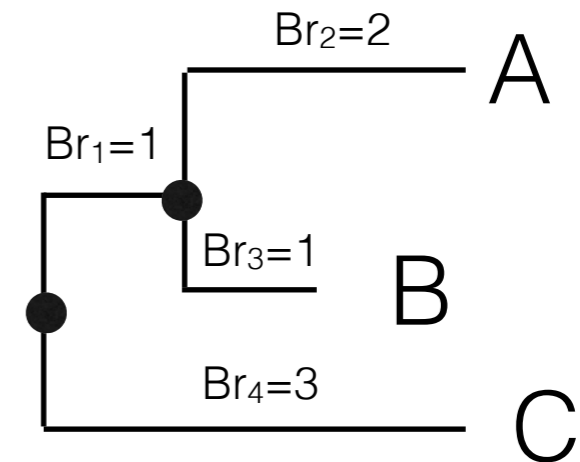
Limitations of existing phyloglmm methods/platforms

- Phylogenetic conservatism in plant phenology (Davies et al, 2015)

- Climate change models, ranging 5 to 184 year time series, 11-1822 plant species per site (23 sites in NH) …

- pez (Pearse et al. 2015) for community ecology (cannot handle multiple observations)

- Bayesian methods (MCMCglmm, brms)

- Not practical with existing software

Branch-Species (BS) matrix approach

- The phylogenetic variability of a particular observation can be written as the sum of the evolutionary changes occurred on all the branches in the phylogeny in its past as independent errors.

- Another statistically equivalent method but much more efficient

- lme4, glmmTMB machinery can do single/multiple sites/groups, unbalance designs, multiple observations

$Br_2 = 2$

A

$Br_1 = 1$

$Br_3 = 1$

B

$Br_4 = 3$

C

$$
\begin{array}{cc}
 & \begin{array}{cccc} Br_1 & Br_2 & Br_3 & Br_4 \end{array} \\
\begin{array}{c} A \\ B \\ C \end{array} &
\begin{bmatrix}
1 & 2 & . & . \\
1 & . & 1 & . \\
. & . & . & 3
\end{bmatrix}
\end{array}
$$

$$Y_{ij} = \beta X_i + \epsilon_{ij}$$

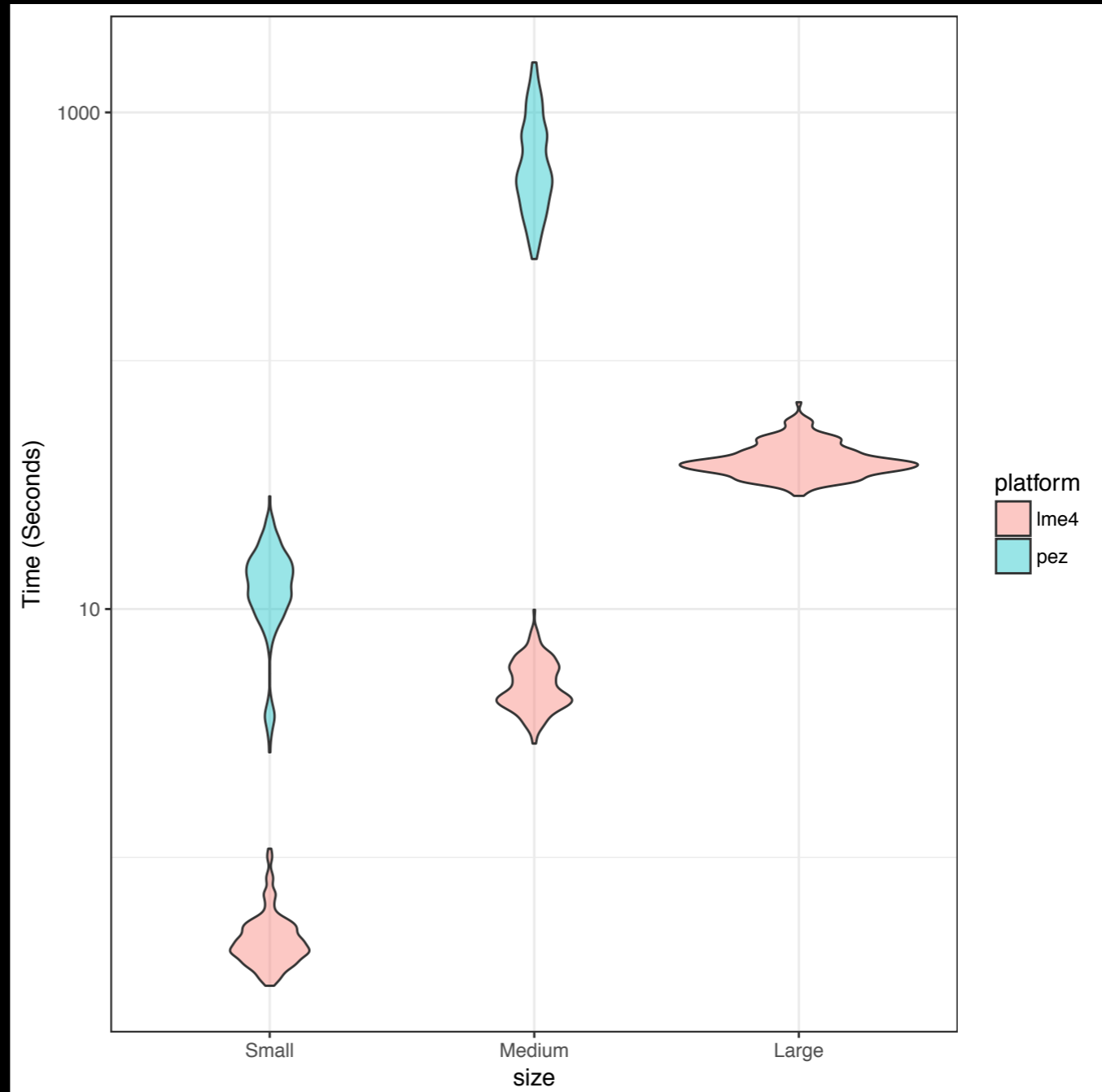$$\epsilon_{ij} = \sum \epsilon_{branches} + \epsilon_{ij}$$

Simulation world

- Phylogenetic signal in response trait (B_0)

- Phylogenetic signal in regression coefficient (B_X)

- Correlation of phylogenetic signal in B_0 and B_X

- Tip variation (residual)

- Species by Site interaction
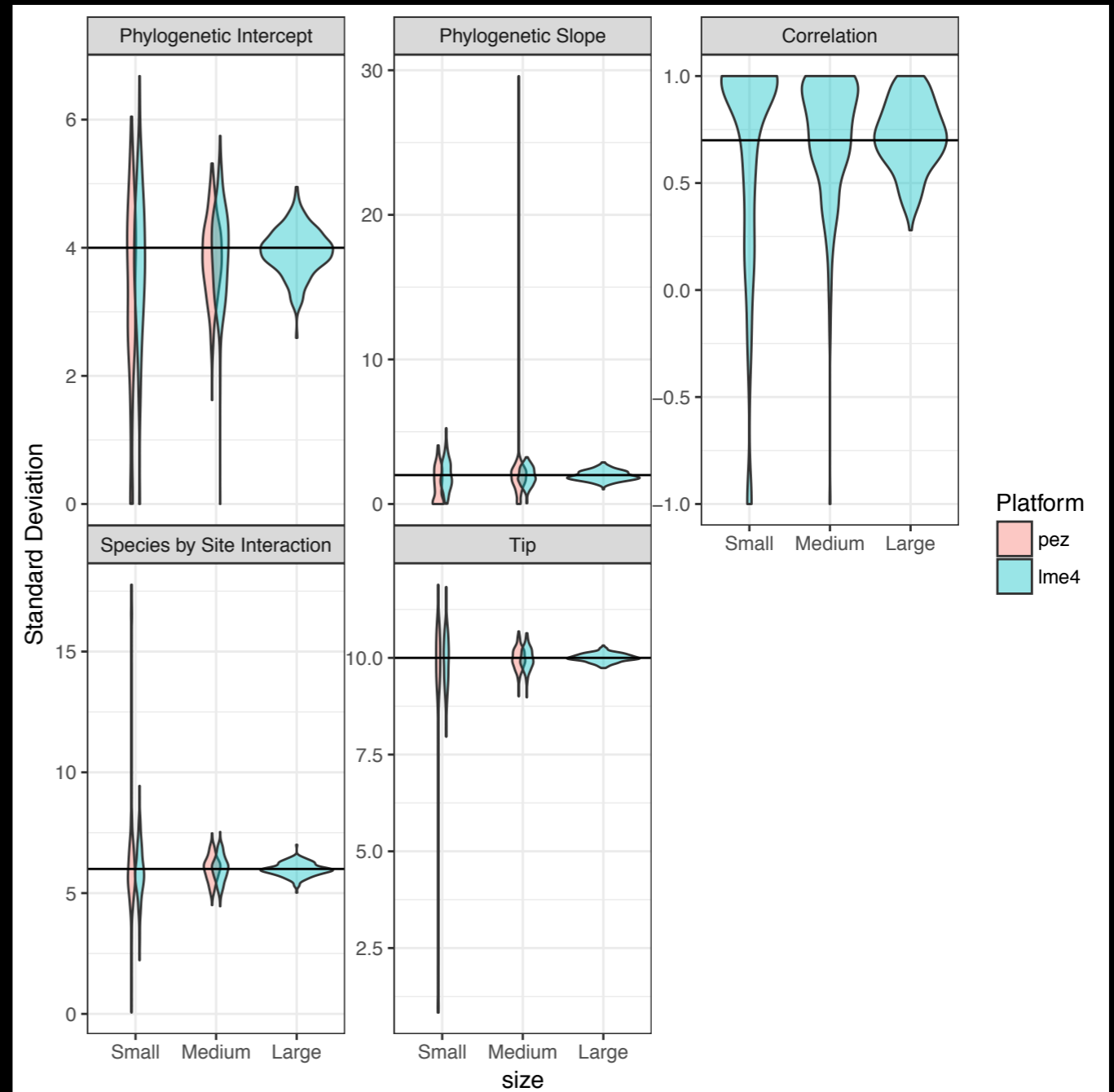
Simulation world

- Orders of magnitude faster

Remarks

- Felsenstein's Difficulties (Phylogeny assumption, ignore phylogeny)

- Complexity tradeoff

- New applications/ideas (PhylogImm in disease modelling)

Acknowledgments

- Ben Bolker

- Jonathan Dushoff

- Morgan Kain and the rest of Bolker-Dushoff lab.