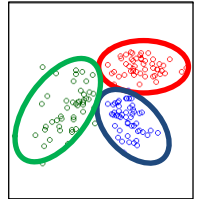
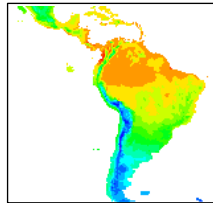
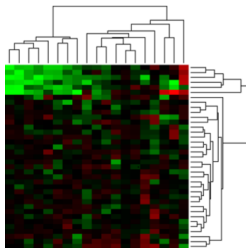
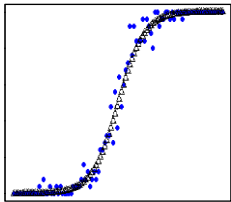
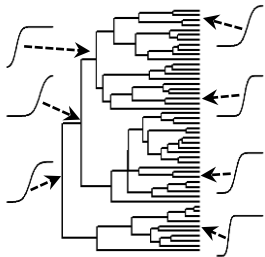


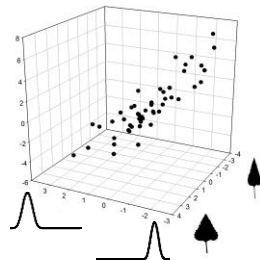
Likelihood-Based Parameter Estimation for High-Dimensional Phylogenetic Comparative Models



Overcoming the Limitations of 'Distance-Based' Methods



Eric W. Goolsby
Dept. of EEB
Brown University



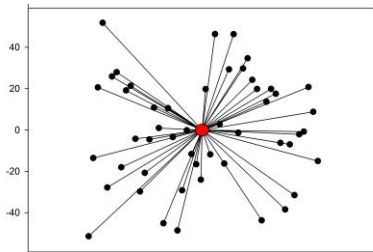
Multivariate Statistics

- As the number of traits (dimensions) increases
 - Number of covariance parameters skyrockets (at the expense of statistical power)
 - Covariance matrices approach singularity
 - Likelihood function becomes unstable

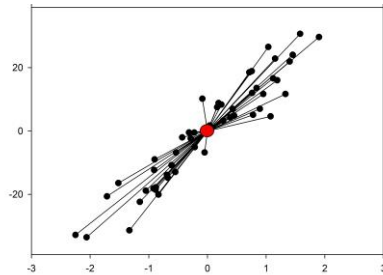
# Traits (M)	1	2	3	4	5	10	20
# Parameters	1	3	6	10	15	55	210

Distance-Based Methods

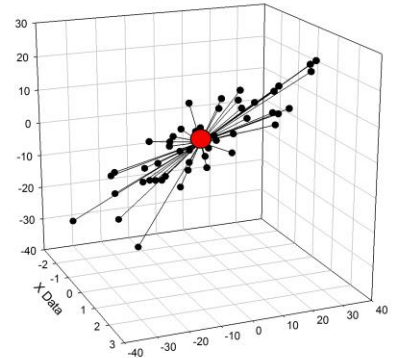
- Provide simple univariate summary statistics



1 trait



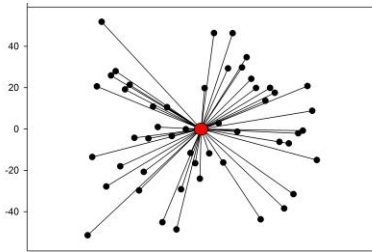
2 traits



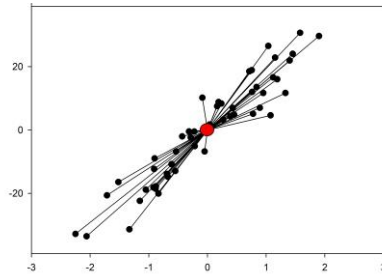
3 traits

Distance-Based Methods

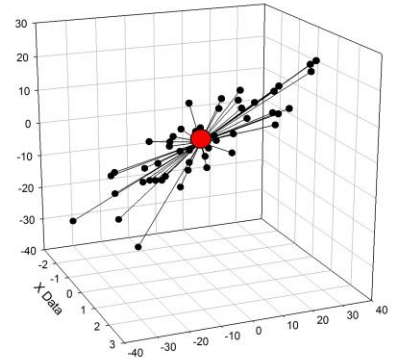
- Provide simple univariate summary statistics
- Avoid singular matrices
- Preserves statistical power*



1 trait

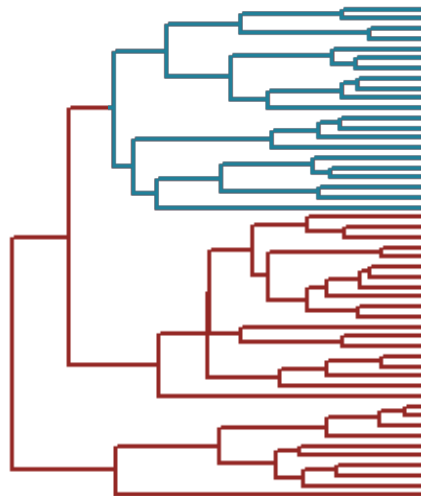


2 traits



3 traits

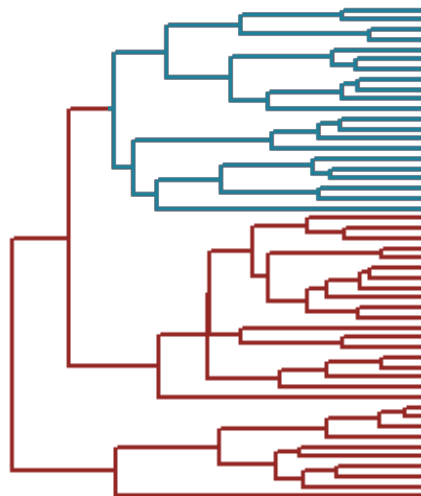
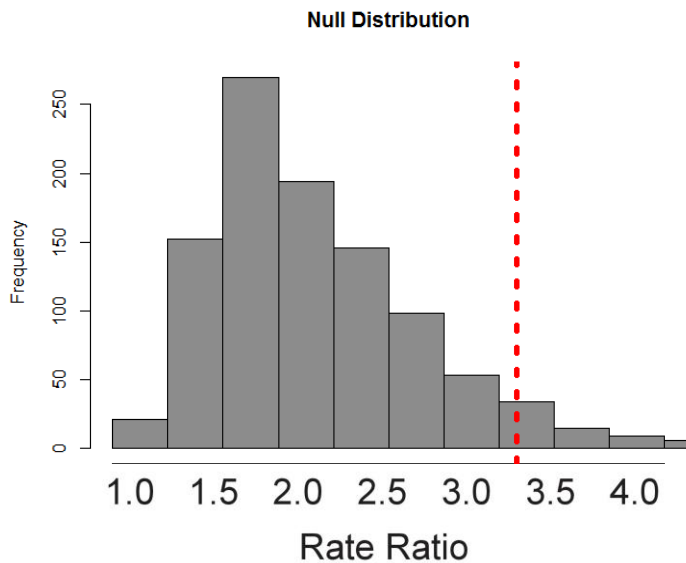
Testing for Evolutionary Rate Shifts



Adams (2014, Syst. Biol)

$$H_0: \frac{\sigma_{red}^2}{\sigma_{blue}^2} = 1$$

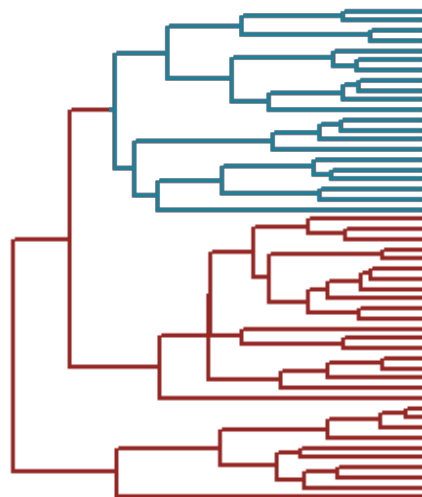
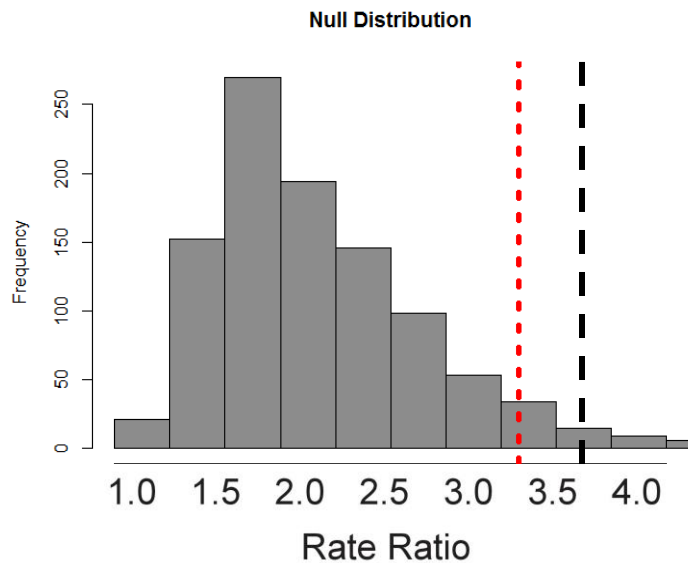
Testing for Evolutionary Rate Shifts



Adams (2014, Syst. Biol)

$$H_0: \frac{\sigma_{red}^2}{\sigma_{blue}^2} = 1$$

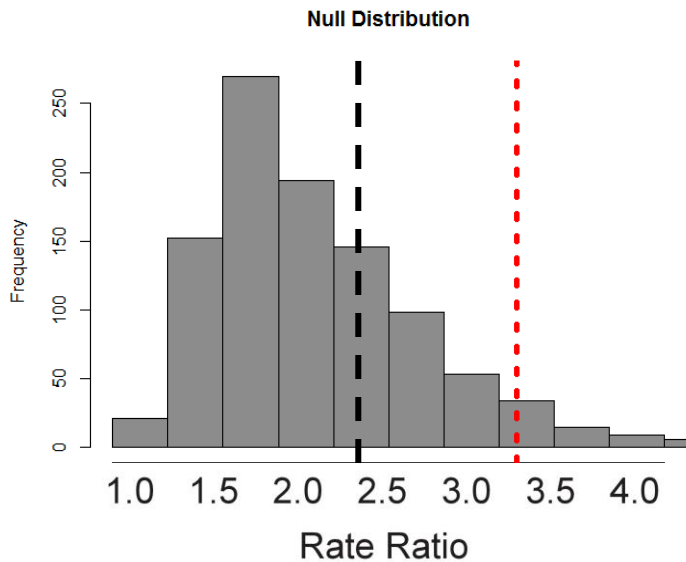
Testing for Evolutionary Rate Shifts



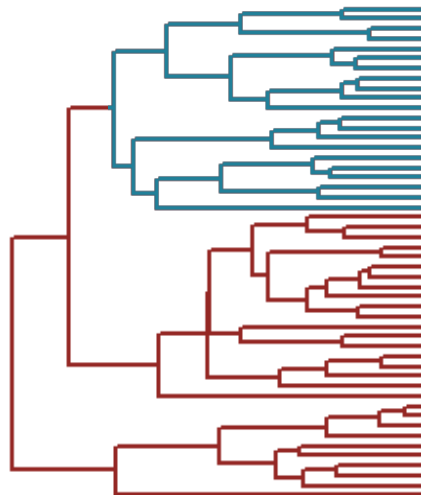
Adams (2014, Syst. Biol)

$$H_0: \frac{\sigma_{red}^2}{\sigma_{blue}^2} = 1$$

Testing for Evolutionary Rate Shifts



Adams (2014, Syst. Biol)



$$H_0: \frac{\sigma_{red}^2}{\sigma_{blue}^2} = 1$$

Other Distance-Based Methods

- Phylogenetic signal (Blomberg's K)
 - Converts mean squared distance into K

Other Distance-Based Methods

- Phylogenetic signal (Blomberg's K)
 - Converts mean squared distance into K
- Phylogenetic regression
 - Univariate predictor(s), high-dimensional response
 - Converts regression-based mean squared distance into F-ratio

Other Distance-Based Methods

- Phylogenetic signal (Blomberg's K)
 - Converts mean squared distance into K
- Phylogenetic regression
 - Univariate predictor(s), high-dimensional response
 - Converts regression-based mean squared distance into F-ratio
- Phylogenetic partial least squares
 - Two high-dimensional variables
 - Computes distance-based correlation (r)

Limitations of Distance-Based Methods

- Important information (e.g., trait covariance) covariance is lost
- Sensitive to violations of model assumptions
- No likelihood function
 - Limited to Brownian motion evolution
- Can't combine multiple hypotheses
 - e.g., phylogenetic regression with rate shift, etc.

A Pseudolikelihood-Based Alternative

- Pairwise composite likelihood
 - Product of the likelihoods for all possible pairwise combinations of variables

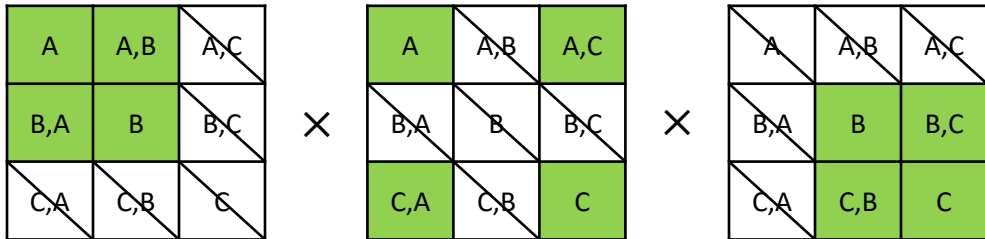
A Pseudolikelihood-Based Alternative

- Pairwise composite likelihood
 - Product of the likelihoods for all possible pairwise combinations of variables

A	A,B	A,C
B,A	B	B,C
C,A	C,B	C

A Pseudolikelihood-Based Alternative

- Pairwise composite likelihood
 - Product of the likelihoods for all possible pairwise combinations of variables



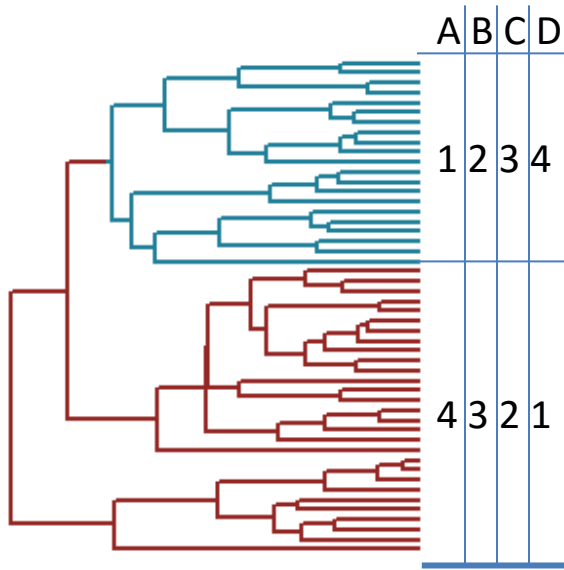
A Pseudolikelihood-Based Alternative

- Pairwise composite likelihood
 - Product of the likelihoods for all possible pairwise combinations of variables
 - Shares many properties with true likelihood
 - Consistent
 - Unbiased
 - Asymptotically normal

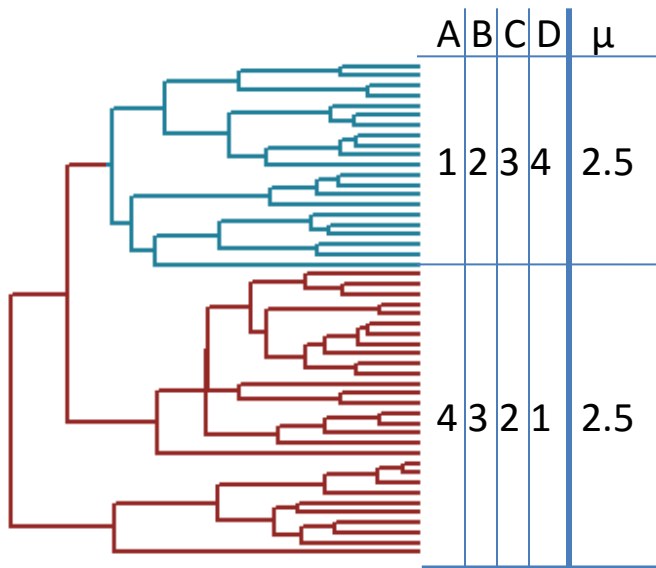
A Pseudolikelihood-Based Alternative

- Pairwise composite likelihood
 - Product of the likelihoods for all possible pairwise combinations of variables
 - Shares many properties with true likelihood
 - Consistent
 - Unbiased
 - Asymptotically normal
- Computationally stable
- Parametric bootstrapping (Boettiger et al. 2012)

Testing for Rate Shifts

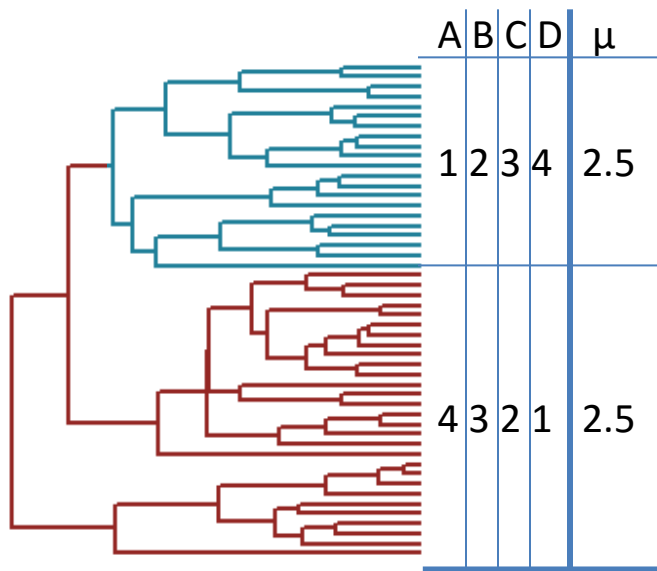


Testing for Rate Shifts

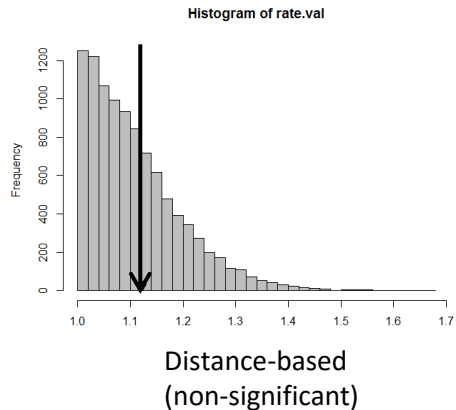


$dvar(\text{blue:red}) = 1.0$
(no shift)

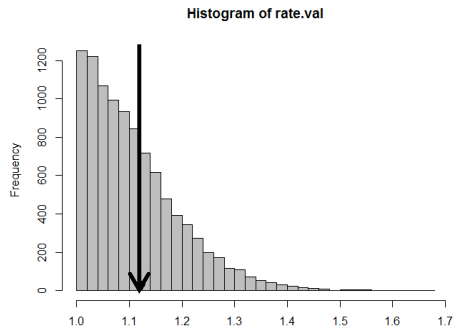
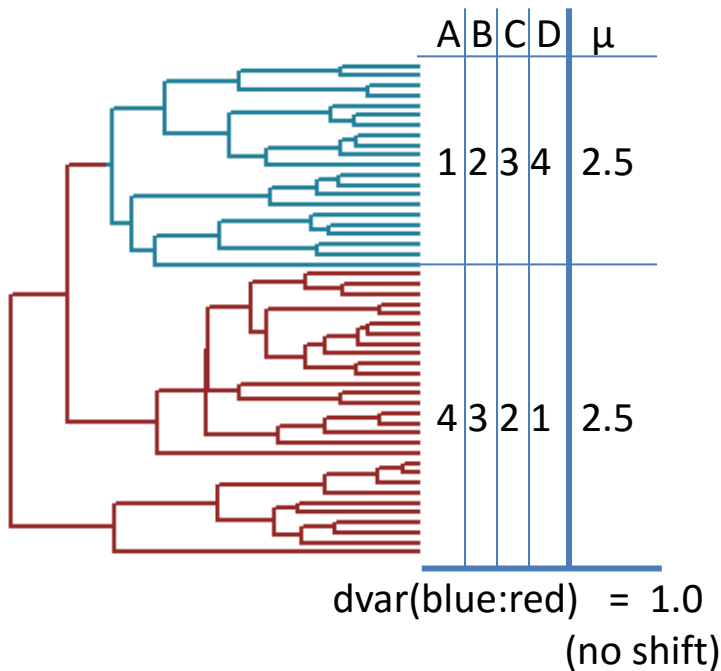
Testing for Rate Shifts



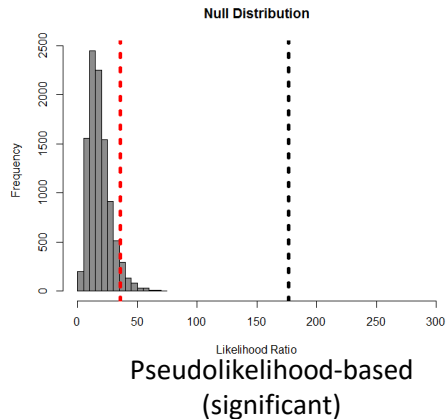
$dvar(\text{blue:red}) = 1.0$
(no shift)



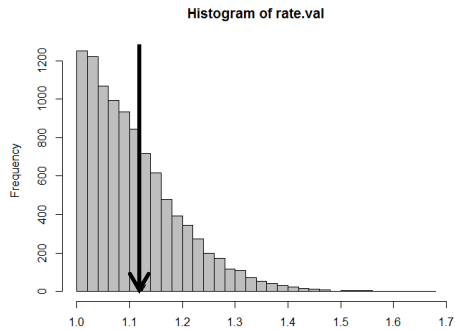
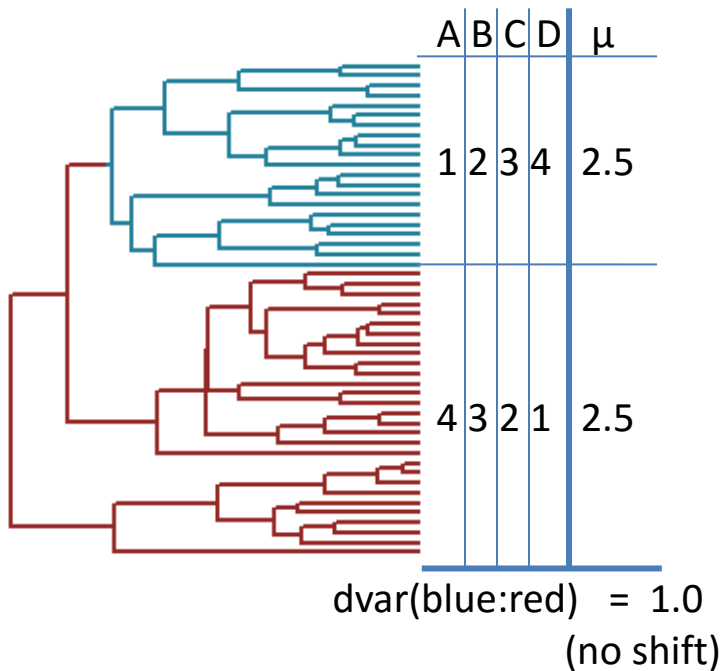
Testing for Rate Shifts



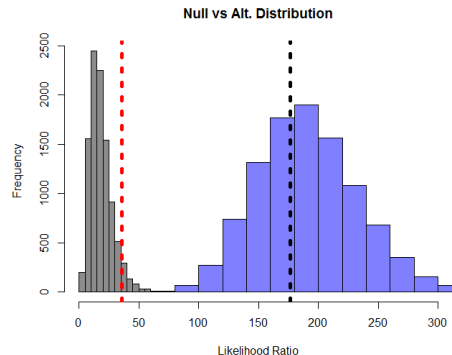
Distance-based
(non-significant)



Testing for Rate Shifts

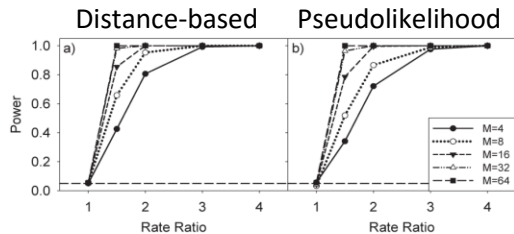


Distance-based
(non-significant)



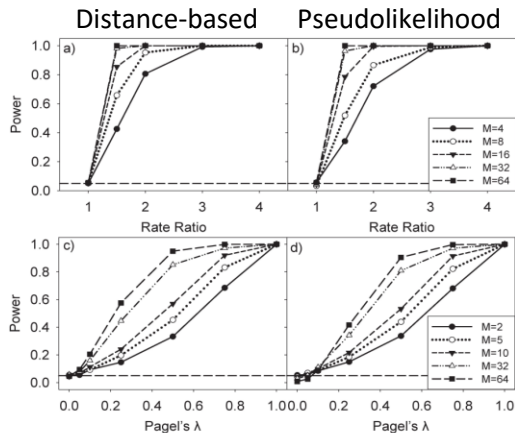
Pseudolikelihood-based
(significant)

Comparing evolutionary rates
Appropriate Type I error and power
(if distance-based assumptions aren't violated)



Comparing evolutionary rates
Appropriate Type I error and power
(if distance-based assumptions aren't violated)

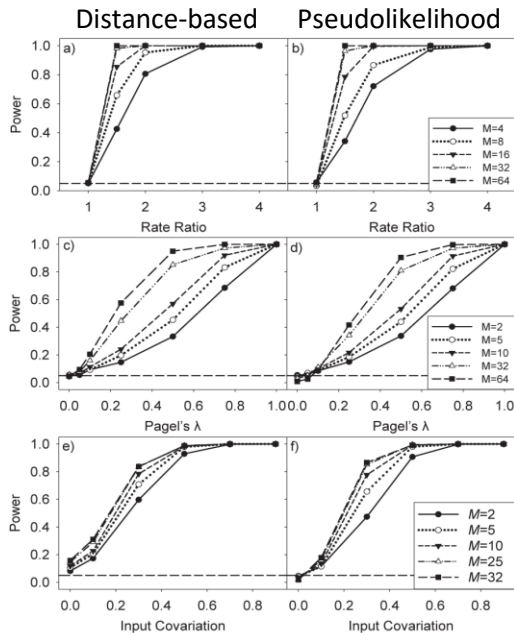
Phylogenetic signal
Appropriate Type I error and power
(if distance-based assumptions aren't violated)



Comparing evolutionary rates
Appropriate Type I error and power
(if distance-based assumptions aren't violated)

Phylogenetic signal
Appropriate Type I error and power
(if distance-based assumptions aren't violated)

Phylogenetic regression
Elevated Type I error
(distance-based)

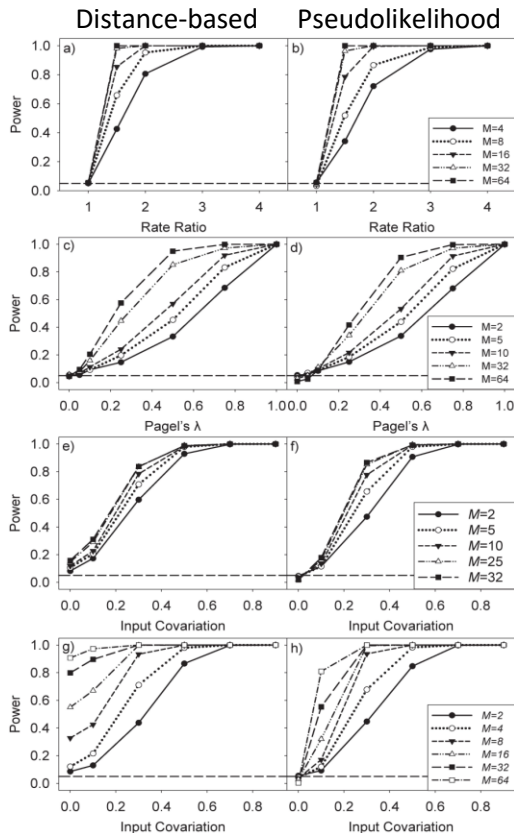


Comparing evolutionary rates
 Appropriate Type I error and power
 (if distance-based assumptions aren't violated)

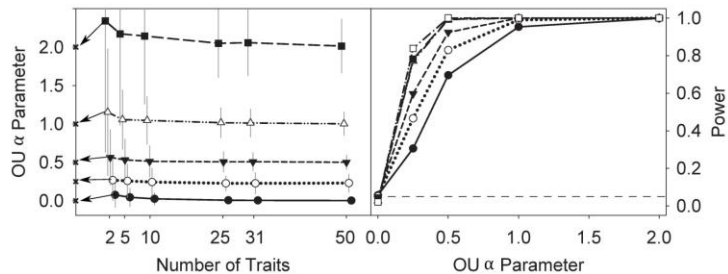
Phylogenetic signal
 Appropriate Type I error and power
 (if distance-based assumptions aren't violated)

Phylogenetic regression
Elevated Type I error
 (distance-based)

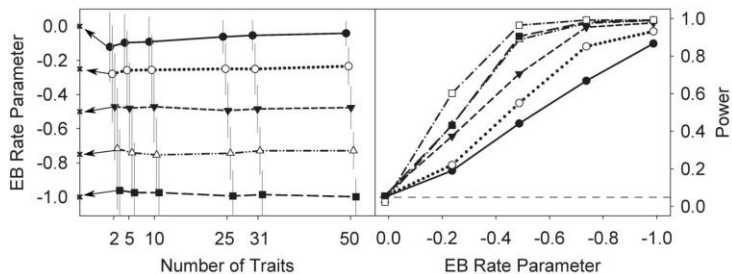
Phylogenetic partial least squares
Extremely high Type I error
 (distance-based)



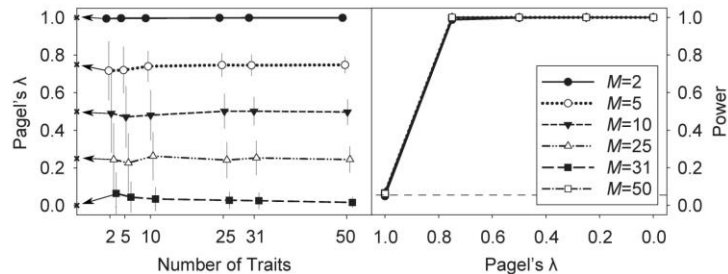
OU model



EB model



λ model



Future Directions

- Model flexibility
 - Missing data
 - Within-species variation
 - Phylogenetic uncertainty
 - Multi-optima OU

Future Directions

- Model flexibility
 - Missing data
 - Within-species variation
 - Phylogenetic uncertainty
 - Multi-optima OU
- Assess model appropriateness
 - e.g., Model adequacy (Pennell et al. 2015)

R Implementation

- R package phylocurve (www.phylocurve.org)
 - Available on CRAN
 - GitHub: <http://ericgoalsby.github.io/>
- Goolsby E.W. “Likelihood-Based Parameter Estimation for High-Dimensional Phylogenetic Comparative Models: Overcoming the Limitations of 'Distance-Based' Methods.” *Systematic Biology*. In press.
- Acknowledgements: NSF DEB-1501215