Model-based Assessment of Functional and Phylogenetic Diversity

Shaoqian Huang

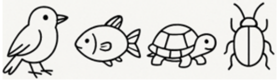
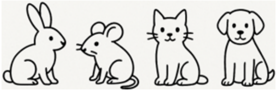
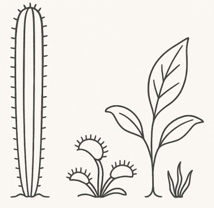
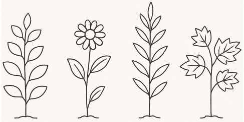
Supervisor: Prof. David Warton
School of Mathematics and Statistics
University of New South Wales

27 November 2025

Overview

1. An Introduction of Basic Concepts
2. From Distance-based Measures to Model-based Assessment
3. Model-based Assessment of PD and FD
4. The Future Work

Phylogenetic Diversity (PD) vs Functional Diversity (FD)

	Definition	High Diversity	Low Diversity
PD	The extent to which species differ in their **evolutionary history**.	 Fig. 1	 Fig. 2
FD	The extent to which species differ in their **functional traits**.	 Fig. 3	 Fig. 4

Rao's Quadratic Entropy (Rao's Q)

$$Q = \mathbf{p}^\top \mathbf{D} \mathbf{p} \tag{1}$$

- $\mathbf{D}$: distance matrix, elements d_{ij} = pairwise evolutionary/functional distances
- $\mathbf{p}$: vector of relative abundances, $\sum_i p_i = 1$

$$Q = \sum_i \sum_j p_i p_j d_{ij} \tag{2}$$

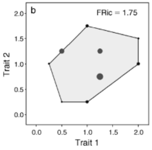
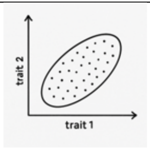
- **A mean dissimilarity measure.**

Distance-based Measures for PD/FD

Sub-category	Method	Formula	PD?	FD?
Mean dissimilarity	Rao's Q	$Q = \sum_i \sum_j p_i p_j d_{ij}$	Yes	Yes
	MPD	$\text{MPD} = \frac{\sum_{i<j} d_{ij}}{\binom{m}{2}}$	Yes	Yes
	FDis	$\text{FDis} = \frac{\sum_i p_i \lvert x_i - \bar{x} \rvert}{\sum_i p_i}$	No	Yes
Dendrogram	Faith's PD	$\text{Faith's PD} = \sum \text{branch lengths}$	Yes	Yes

References: Botta-Dukát (2005), Webb et al. (2002),Laliberté and Legendre (2010), Faith (1992), Petchey and Gaston (2002).

Multivariate Space - based Measures for PD/FD

Method	Representation	PD?	FD?
Binary hypervolume		No	Yes
Probabilistic hypervolume		No	Yes

References: Villéger et al. (2008), Cornwell et al. (2006), Blonder (2016).

From Distance-based Measures to Model-based Assessment

- Traditional measure (e.g. Rao's Q): mean dissimilarity.
    - Hard to evaluate how diversity changes along environmental gradients.
    - Highly sensitive to sampling intensity (i.e., how much effort is invested in data collection at each site).
- Model-based assessment (e.g. GLM):
    - Estimate compositional changes along environmental gradients.
    - Analyze their consequences for diversity changes.

Model Abundance

Abundance matrix

$$Y = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1m} \\ y_{21} & y_{22} & \cdots & y_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nm} \end{bmatrix}$$

Columns: Species 1...m

Rows: Site 1...n

Probabilistic distribution

$$y_{ij} \sim F(\mu_i, \phi_j)$$

GLM / Model-based assessment

Link abundance to environmental gradients via

$$g(\mu_{ij}) = \underbrace{\alpha_{i0} + \alpha_{0j} + x_i^T \alpha}_{\alpha\textbf{-diversity}} + \underbrace{x_i^T \beta_{\cdot j}}_{\beta\textbf{-diversity}}$$

$$i = 1, \ldots, n, \quad j = 1, \ldots, m$$

How to measure changes in PD/FD along gradients?

A New Diversity Change Measure

1. Rao's Diversity Change

$$\Delta\mathrm{Div}_{\mathrm{Rao}} = \left| p_2^T D p_2 - p_1^T D p_1 \right| = \left| p_2^T S p_2 - p_1^T S p_1 \right| ,$$

where D is a distance matrix and $S > 0$ is a similarity matrix.

A New Diversity Change Measure

1. Rao's Diversity Change

$$\Delta\mathrm{Div}_{\mathrm{Rao}} = \left| p_2^T D p_2 - p_1^T D p_1 \right| = \left| p_2^T S p_2 - p_1^T S p_1 \right|,$$

where D is a distance matrix and $S > 0$ is a similarity matrix.

2. The Proposed Diversity Change

Let the relative abundance change be

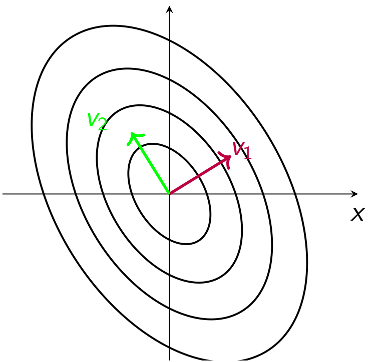
$$\beta = p_2 - p_1, \quad \sum_i \beta_i = 0.$$

The resulting diversity change is

$$\Delta\mathrm{Div}_{\mathrm{Model}} = \beta^T S \beta.$$

Why the leading eigenvectors matter

Fig. 5: Contour of $\beta^\top S\beta$



The main idea

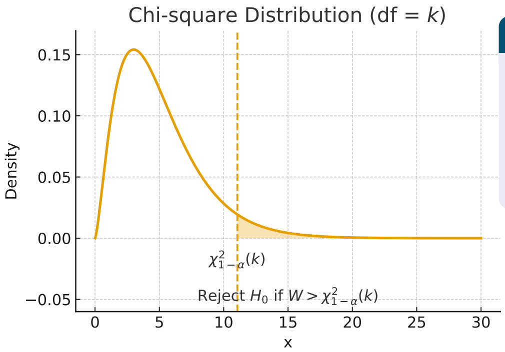
- If $\beta \parallel v_1$, PD/FD change.
- If $\beta \perp v_1$, No PD/FD change.

Hypothesis Testing for PD/FD Change

Hypothesis

$$H_0: \quad \text{No PD/FD change} \quad \Leftrightarrow \quad V_k^T \boldsymbol{\beta} = \mathbf{0}.$$

Here, $V_k = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$ is formed by the first k leading eigenvectors.



Wald Test Statistic

$$W = \left(V_k^T \hat{\boldsymbol{\beta}}\right)^T \left[\mathrm{Var}\left(V_k^T \hat{\boldsymbol{\beta}}\right)\right]^{-1} \left(V_k^T \hat{\boldsymbol{\beta}}\right),$$

which follows a χ^2 distribution with k degrees of freedom under H_0.

$$p = 1 - F_{\chi^2_k}(W).$$

Potential Rules for Determining k

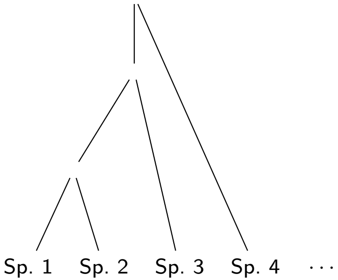
$$V_k = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k)$$

denotes the matrix composed of the first k leading eigenvectors of S, corresponding to the eigenvalues $\lambda_1, \ldots, \lambda_k$.

Ways to determine k:

1. Pre-specify k (e.g., $k = 1$ or $k = 2$).
2. Choose k such that the proportion of variance explained exceeds a threshold τ (Proportion bigger than τ.):
$$\frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{m} \lambda_i} \geq \tau.$$
3. Select k such that λ_i is greater than the mean eigenvalue (Bigger than averaged eigenvalues.):
$$\lambda_i > \frac{1}{m} \sum_{j=1}^{m} \lambda_j.$$

Simulation Design



$$y_{ijh} \sim \text{Binomial}(1, \mu_{ij}), \quad \mu_{ij} = \text{expit}(x)$$

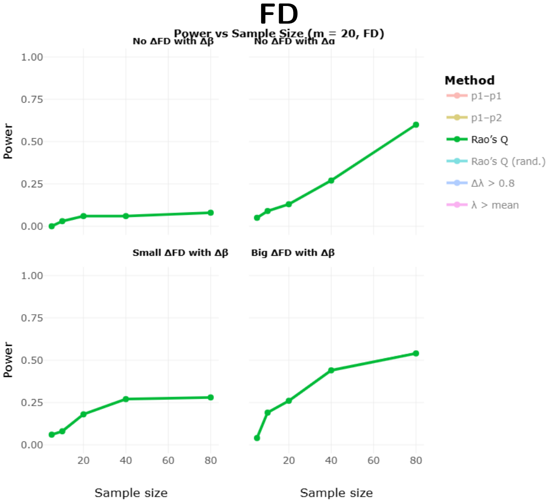
We vary sample size r,
 # species m,
and # mean swaps s.

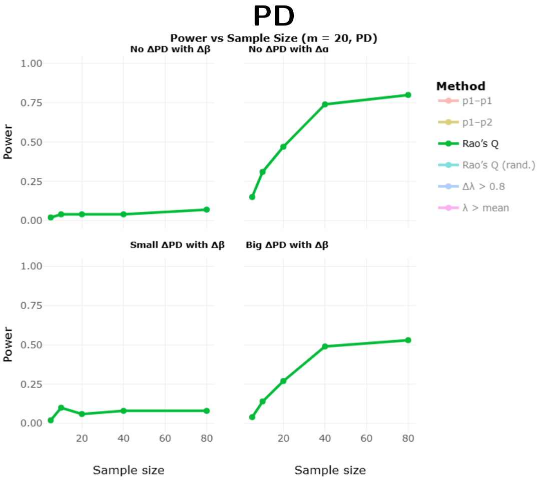
$$\textbf{Power} = \frac{\#\{\text{simulation with } p\text{-value} < 0.05\}}{\#\text{ simulations}}$$

	Sp. 1	Sp. 2	Sp. 3	Sp. 4	$\cdots$	Sp. m	$\Delta\beta$	$\Delta\alpha$	ΔPD/ΔFD
Com 1	expit($-\beta$)	expit(β)	expit($-\beta$)	expit(β)	$\cdots$	expit($\pm\beta$)		Reference	
Com 2	expit(β)	expit($-\beta$)	expit($-\beta$)	expit(β)	$\cdots$	expit($\pm\beta$)	$4\beta \cdot s$	No	No
Com 3	expit($-\beta$)	expit($-\beta$)	expit(β)	expit(β)	$\cdots$	expit($\pm\beta$)	$4\beta \cdot s$	No	Small
Com 4	expit(β)	expit(β)	expit($-\beta$)	expit($-\beta$)	$\cdots$	expit($\pm\beta$)	$4\beta \cdot s$	No	Big
Com 5	expit($\alpha-\beta$)	expit($\alpha+\beta$)	expit($\alpha-\beta$)	expit($\alpha+\beta$)	$\cdots$	expit($\alpha\pm\beta$)	No	$\alpha \cdot m$	No

Note: $\text{expit}(x) = \frac{e^x}{1+e^x}$; s = # mean swaps; Com = community; Sp = species.

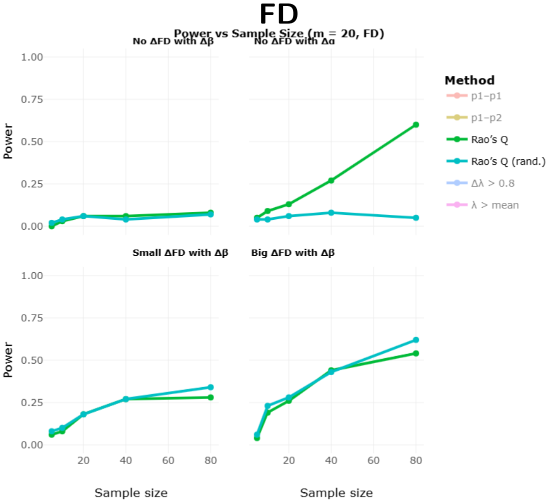
Simulation Results

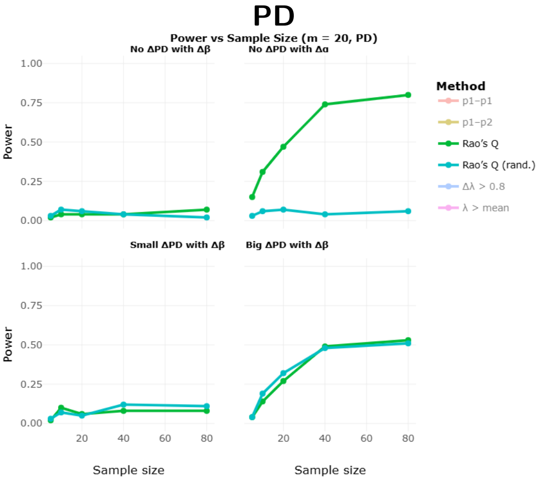




Rao's Q

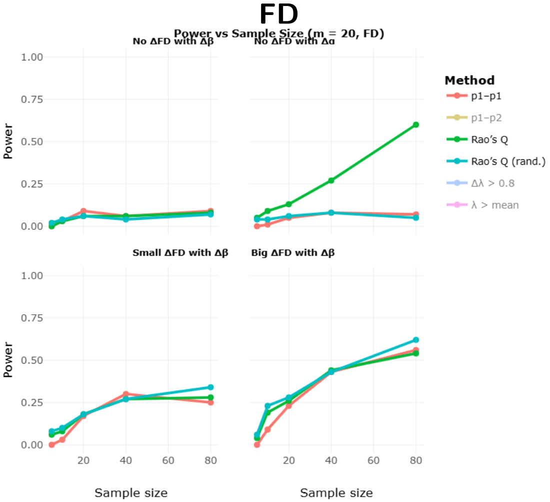
Simulation Results

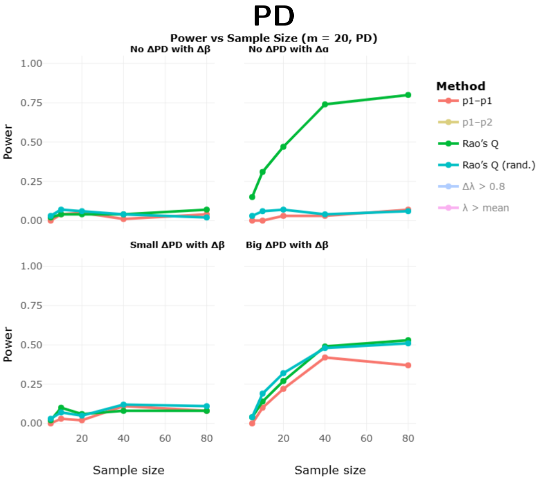




Rao's Q $\rightarrow$ Rao's Q (rand.)

Simulation Results

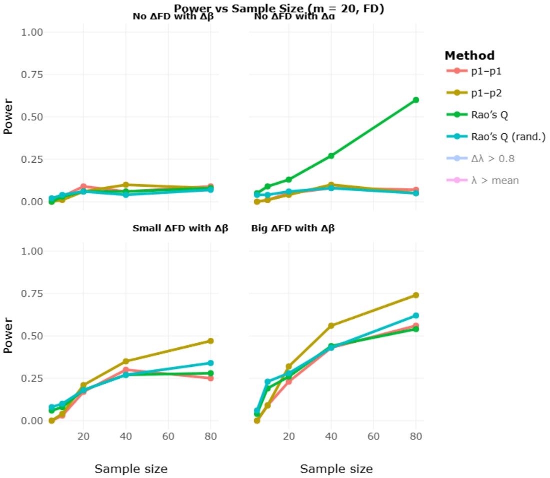




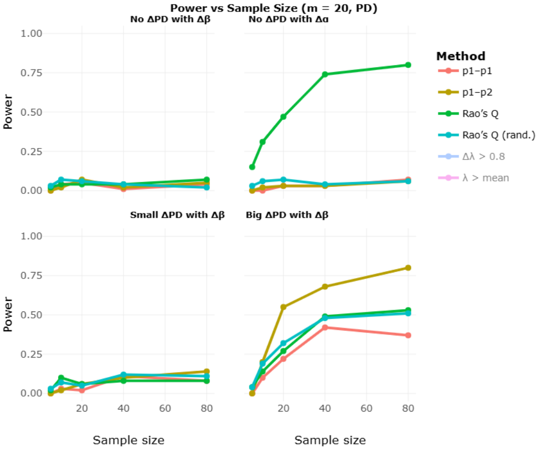
Rao's Q $\rightarrow$ Rao's Q (rand.) $\rightarrow$ **p1**

Simulation Results

FD

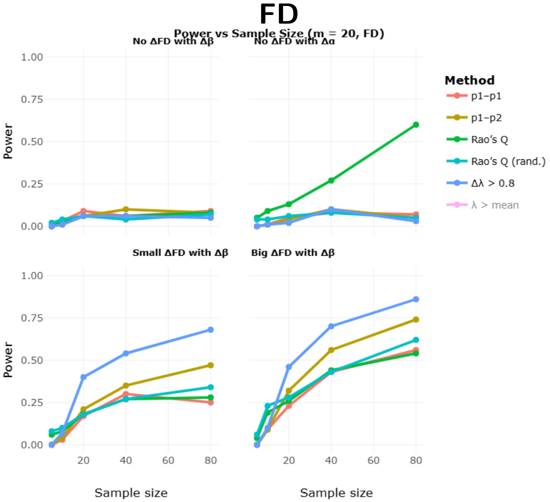


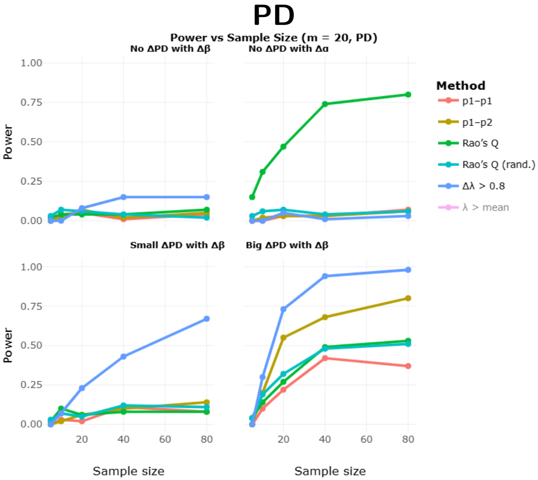
PD



Rao's Q $\rightarrow$ Rao's Q (rand.) $\rightarrow$ **p1** $\rightarrow$ **p1,p2**

Simulation Results

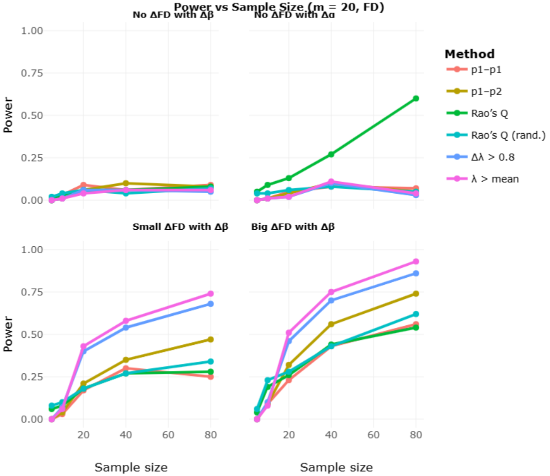




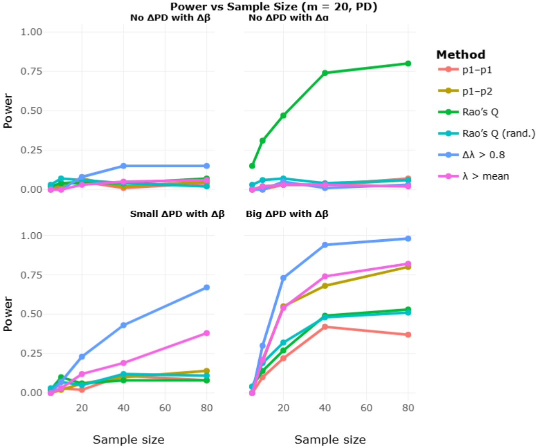
Rao's Q $\rightarrow$ Rao's Q (rand.) $\rightarrow$ **p1** $\rightarrow$ **p1,p2** $\rightarrow$ $\Delta\lambda >$ **0.8**

Simulation Results

FD



PD



Rao's Q $\rightarrow$ Rao's Q (rand.) $\rightarrow$ **p1** $\rightarrow$ **p1,p2** $\rightarrow$ $\Delta\lambda >$**0.8** $\rightarrow$ $\lambda >$ **mean**

The Future Work

- Implement large-scale simulations, including correlations between abundances and different distributions (e.g., Beta, Poisson)
- Apply the model to real-world case studies
- Develop a user-friendly software tool for FD and PD calculation

References

Blonder, B. (2016). Do hypervolumes have holes? *The American Naturalist*, 187(4):E93–E105. PMID: 27028084.

Botta-Dukát, Z. (2005). Rao's quadratic entropy as a measure of functional diversity based on multiple traits. *Journal of Vegetation Science*, 16(5):533–540.

Cornwell, W. K., Schwilk, D. W., and Ackerly, D. D. (2006). A trait-based test for habitat filtering: Convex hull volume. *Ecology*, 87(6):1465–1471.

Faith, D. P. (1992). Conservation evaluation and phylogenetic diversity. *Biological Conservation*, 61(1):1–10.

Laliberté, E. and Legendre, P. (2010). A distance-based framework for measuring functional diversity from multiple traits. *Ecology*, 91(1):299–305.

Petchey, O. L. and Gaston, K. J. (2002). Functional diversity (fd), species richness and community composition. *Ecology Letters*, 5(3):402–411.

Villéger, S., Mason, N. W. H., and Mouillot, D. (2008). New multidimensional functional diversity indices for a multifaceted framework in functional ecology. *Ecology*, 89(8):2290–2301.

Webb, C. O., Ackerly, D. D., McPeek, M. A., and Donoghue, M. J. (2002). Phylogenies and community ecology. *Annual Review of Ecology, Evolution, and Systematics*, 33(Volume 33, 2002):475–505.

The End