Ignasi Arranz, Bertrand Fournier, Nigel P. Lester, Brian J. Shuter, and Pedro R. Peres-Neto. Species compositions mediate biomass conservation: the case of lake fish communities. Ecology.

Appendix S8. Multiple combinations of species leading to different values of biomass conservation

Even though different species compositional axes (PCoA) characterize different communities in terms of their species compositions, one critical aspect is to demonstrate that a wide variety of species compositions can lead to low and high values of biomass conservation. Imagine an extreme fictional case in which lakes are dominated by only two very distinct types of communities in terms of species composition; one type with low and the other type with high average biomass slopes. In this case, only one PCoA axis would be important in describing biomass slope variation as a function of variation in species compositions. However, this fictional lake system emphasizes that we would not have enough independent degree of species variation across communities to generate strong evidence that composition drives biomass conservation. Stronger evidence would be generated if multiple different combinations of species lead to low and high degrees of biomass conservation. To tackle this issue, we conducted variation partitioning using a generalized least squares (GLS) approach in which beta-diversity was used as weights in fitting process of the model. The goal of this model is to reduce the contribution of lakes with similar species compositions in explaining variation in biomass conservation.

We start by demonstrating that many similar communities and dissimilar communities in their species composition can lead to wide differences (and similarities) in their abilities to conserve biomass. The plot below shows the variation in community composition across lakes (lake pairwise Jaccard similarity) and their absolute differences in biomass conservation. Many lakes with no species in common (Jaccard similarity equal to zero) have very similar (close to zero difference in biomass conservation) or very different biomass conservation values. Conversely, many lakes with similar (Jaccard similarity close to one) and even the exact same species (Jaccard similarity equal to one) also vary in their abilities to conserve biomass across fish size (biomass) classes. Obviously, there is a tendency that lakes with similar species compositions to have similar biomass values (negative correlation in the plot below). Otherwise, our variation partitioning results (Fig. 2) would not have been as strong (see Peres-Neto and Jackson (2001) for a discussion on how pairwise approaches, i.e., Mantel-based, and raw-data approaches). It is important to note that lakes with similar species composition do not necessary share all their species in common. For instance, the 3407 pairs of lakes with a Jaccard similarity equal or greater than 0.70 do not share similar compositions with that group. That is to say that two pairs of lakes (A-B and C-D) with a compositional similarity of 100% could share no species in common across pairs (i.e., lakes A-C and B-D). As such, there can be a wide diversity of species composition leading to a wide variation in biomass conservation values.

The GLS-based variation partitioning to reduce the contribution of lakes with similar species compositions in explaining variation in biomass conservation was conducted as follows. The total contribution of environmental variation (matrix **X1**; lakes in rows and environmental variables in columns) and species compositions (matrix **X2** containing PCoA coordinates; lakes in rows and axes in columns) to biomass conservation (**Bc**; vector of biomass slopes) was estimated with a matrix of weights (a lake pairwise matrix with the complement of the Jaccard similarity, i.e., Jaccard distance). In this way, lakes that are very similar in their species compositions contribute less to the model estimates. All matrix columns were standardized to mean equal to zero and variance equal to unity as components of variation are not affected by intercepts.

where are the predicted values of **Bc**, the transpose of the matrix containing both **X1** and **X2** (juxtaposed) and is the adjusted coefficient of determination of the model with both environment and species compositional predictors; *n* is the number of lakes and p the number of predictors in .

To produce variation partitioning estimates, we need now to estimate the models for environment and species composition, separately.

where are the predicted values of **Bc** for the model containing only environmental predictors, is the number of environmental predictors and the adjusted coefficient of determination of the model with only the environmental predictors **X1**. Finally, the model with the compositional predictors only:

where are the predicted values of **Bc** for the model containing only the compositional predictors, is the number of compositions predictors and the adjusted coefficient of determination of the model with only the compositional predictors **X2**. The unique contributions of the environment (fraction [a] here) and species composition (fraction [c] here) are calculated as:

The shared contribution (fraction [b] here) is:

The Table S1 below reports the original (Fig. 2) and the estimated values via the GLS procedure for the species compositions based on presence-absence (Fig. 2b) and biomass data (Fig. 2c). As such, **X2** above was composed of either species compositional predictors (PCoA axes) based on presence-absence or biomass. We did not consider the combination of both presence-absence and biomass (Fig. 2d) as results were equivalent.

Not only the estimates for the importance of species composition remained higher than of the environment it increased in relation to the original variation partitioning values (Table S1; original values in parenthesis). This is because the shared fraction between the environment and species compositional axes decreased in the GLM analysis in relation to the original variation partitioning. While environmental features to serve as relevant predictors of species distributions (RDA results in main text) need to be associated to particular compositions of species (an obvious premise), there can be many compositions of species that can lead to similar (or different) values of biomass conservation (Fig. S1). As such, when using weights based on compositional differences among lake communities, the shared contribution is reduced, and the contribution of species compositions increase. This demonstrates that species compositions mediate biomass conservation and that different compositions of species can lead to better or worse (“prohibit compositions”) abilities of communities to conserve biomass.

**Table S1.** Original and GLS variation partitioning results. Results for compositional predictors (presence-absence) and biomass are presented separately as in Fig. 2b and 2c. Original values (Figs. 2b and 2c) are the main values and the GLM values are in parentheses.

|  |  |  |
| --- | --- | --- |
| **fractions** | **presence-absence**  **original (GLS)** | **biomass**  **original (GLS)** |
| Environment only [a] | 0.03 (0.03) | <0.01 (<0.01) |
| Species composition only [c] | 0.20 (0.27) | 0.59 (0.50) |
| Shared variation [b] | 0.24 (0.12) | 0.27 (0.15) |

Chart

Description automatically generated with medium confidence

**Figure S1.** Lake pairwise differences in compositional similarity and their differences in biomass conservation.

References

Peres-Neto, P.R., and D.A. Jackson. 2001. How well do multivariate data sets match? The robustness and flexibility of a Procrustean superimposition approach over the Mantel test. Oecologia **129**: 169-178.