Supplementary Methods

$Chagnon, PL^{1*}, Magain, N^{2}, Miadlikowska, J^{2}, Lutzoni, F^{2}$

1. Université de Montréal, Montréal, QC H1X 2B2, CANADA
2. Duke University, Durham, NC 27708, USA

* Corresponding author: pierre-luc.chagnon@umontreal.ca

1 Modularity analysis

Many different approaches exist to measure and optimize modularity [1–3]. Here, we used the algorithm developed by Dormann & Straus [2] for quantitative networks, as this method has been shown, using benchmark tests, to be highly sensitive and accurate. This algorithm uses a dendrogram-based approach to split the network into modules (which has the advantage over several other methods to detect modules nested into larger modules), and optimizes the modularity level following a simulated annealing procedure.

1.1 Simulated annealing

Simulated annealing is a widely used optimization technique that is based on a Markov Chain Monte Carlo (MCMC) procedure. Briefly, it starts with a random state (here, distribution of species into different modules), swaps to an alternative state (the swap is selected randomly among all possible swaps), and the modification is accepted or not based on a given probability. This probability will be proportional to the cost it entails to the optimization procedure. Here, since we want to increase the modularity level of our network module configuration, a swap that decreases modularity would get a low probability of acceptance. However, the important property of simulated annealing is that the algorithm is less constrained in the earlier stages of the optimization process: this is allowed by progressively decreasing a parameter routinely called "temperature" during the algorithm. In the earlier stages, temperature is high, and virtually any swap will be accepted. This will allow the algorithm to broadly explore all potential states of the system (here, allocation of all species in the network to various modules). Progressively, the temperature will go down, and the algorithm will start accepting swaps that decrease modularity with a very low probability, up to a point where it will only accept swaps that increase modularity.

This method is very computationally intensive, but the advantage of Dormann and Strauss's algorithm [2] is that it is coded in C++, which allows hundreds of thousands of swaps to be done within a few seconds. This is in sharp contrast with other approaches coded in R, which can take several minutes for a single network [4].

2 Beta-diversity decomposition and nestedness analyses

2.1 Beta-diversity decompositions

Beta-diversity decompositions have been introduced in the ecological literature to tease apart distinct sources of variation among sites in a meta-community (e.g., [5]). The same reasoning can be applied to interaction networks, where

two given species within a guild can be different because (1) they have a different number of interactions or (2) they may have partners of different identity. Podani et al. [6] have developed a mathematical framework to distinguish these two sources of variation, and to display it in a 2D system, the ternary triangular plot. In this framework, for each pair of species that we compare within a guild, we calculate three distinct quantities that will sum to 1 (thus allowing to plot them in a ternary triangle): S is the Ruzicka similarity between two species, D is the relativized abundance difference measure (i.e., how species vary in their number of interactions) and R is the relativized abundance replacement measure (i.e., how species distribute their interactions with different partners). For comparing species j and k, interacting with a range of n potential partners, these quantities are calculated as follows:

$$S_{Ruz(j,k)} = \frac{\sum_{i=1}^{n} min(x_{ij}, x_{ik})}{\sum_{i=1}^{n} max(x_{ij}, x_{ik})} \tag{1}$$

$$D_{rel(j,k)} = \frac{|\sum_{i=1}^{n} x_{ij} - \sum_{i=1}^{n} x_{ik}|}{\sum_{i=1}^{n} max(x_{ij}, x_{ik})} \tag{2}$$

$$R_{rel(j,k)} = \frac{\sum_{i=1}^{n} |x_{ij} - x_{ik}| - |\sum_{i=1}^{n} x_{ij} - \sum_{i=1}^{n} x_{ik}|}{\sum_{i=1}^{n} max(x_{ij}, x_{ik})} \tag{3}$$

where x_{ij} represents the number of interactions between species j and partner i (and likewise for x_{ik}). Podani et al. [6] present in their original paper idealized matrix scenarios to help visualize where they would fall on the ternary triangle.

2.2 Nestedness

The D_{rel} component of SDR analyses is closely related to the concept of nestedness in meta-communities or interaction networks. Thus, as a way to validate our SDR analyses, we quantified the quantitative (i.e., abundance-based) nestedness of our interaction network based on the $wNODF$ index developed by Almeida-Neto & Ulrich [7]. This index also compares pairs of species one at a time, but here the matrix (i.e., network) is ordered prior to the comparisons, and the comparisons only goes from one side to the other in the matrix. For example, the most typical way of ordering the matrix, when no *a priori* gradient is available, is to sort rows or columns by their total number of non-zero cells (i.e., number of different partners in an interaction network). Then, we look whether the partner distribution of a species with less non-zero cells is nested within that of a species with more non-zero cells. So, in a lichen network, for

example, for a pair of mycobiont species i and j, with species i having more partners (more non-zero cells in the matrix) than species j

$$wNODF_{(i,j)} = k_{ij}/N_j \tag{4}$$

where k_{ij} is the number of photobionts with whom species i interacts more frequently than species j, and N_j is the total number of photobiont partners for species j. It is then possible to compute the overall nestedness for mycobionts as

$$wNODF_{myc} = \sum_{i=1}^{n-1} \sum_{i=j+1}^{n} \frac{k_{ij}}{N_j} \tag{5}$$

where n represents the total number of mycobiont species in the network. Of course, the same reasoning can be applied to evaluate whether different photobionts have a nested distribution of mycobiont partners.

3 Null model

To evaluate the statistical significance of our modularity and nestedness indices, we needed to compare our observed values to random expectations. Since no *a priori* distribution can be expected to reflect the probability distribution of our measured network indices under a random scenario, we needed to generate one using randomizations of our initial data with a given null model. Here, the choice of the null model was crucial: an overly liberal null model would tend to produce matrices drastically different from the original data, such that the null hypothesis (i.e., absence of a significant network-level pattern) would always be rejected (type I error). On the other hand, an overly conservative null model would be very inefficient at exploring the null space, and thus would tend to produce matrices very similar to the original one, such that the null hypothesis could never be rejected (type II error). Here, we chose a null model that was fairly conservative, in that it preserved from the initial matrix/network the connectance (i.e., total number of non-zero cells) and also, as much as possible, the columns and rows marginal totals (i.e., the species' total number of interactions). This makes biological sense, since our null model thus preserves the existence of more or less generalist taxa in our dataset, and also preserves the total number of interactions recorded, which is likely to be influenced by our sampling effort (and thus needs to be controlled for in the randomizations). By using a conservative null model, we are thus less prone to reject the null hypothesis while it is true (i.e., falsely detecting significant patterns in our network) [8].

To generate our null matrices, we used the *vaznull* function implemented in the R package *bipartite* [9]. This function uses a probabilistic approach to fill the cell of the null matrices, with a celll having a higher probability of receiving an interaction if the product of the marginal totals of its corresponding row and column in the original dataset was high. In this view, the function fails at *perfectly* preserving rows and columns marginal totals, but in our case,

this approach was still better than a swap-based approach. Indeed, while the latter strictly preserves rows and columns marginal totals (thus being even more conservative), it often fails to converge to a solution in a reasonable time frame, as it was the case with our dataset. Moreover, since our network connectance was very low, an overly conservative null model such as the swap-based approach would have even more drastically increased type II error rates [10].

4 Interaction symmetry

Interaction symmetry was evaluated following Vázquez et al. [11]. This approach allows evaluating the reciprocity of a species' dependence on its partners. We first define the effect of a species i on its partner j (s_{ij}), as the proportion of all interactions of species j that are realized with species i. Then, for this species pair, we can calculate the difference (d) in the effect of species i on species j and vice versa: $d_{ij} = s_{ij} - s_{ji}$. Thus, for a given mycobiont i, we can calculate its d coefficients for all its photobiont partners, and express its asymmetry coefficient (A_i) as the mean of these. A_i is close to 1 if species i exerts strong effects on its partners while experiencing little reciprocal effects from them, and close to -1 in the reverse situation.

5 Rarefaction curves

Rarefaction curves were generated to ensure that our sampling effort was sufficient to uncover relevant network patterns. To do so, we used a bootstrap approach to re-sample our dataset (250 thalli) with given sampling efforts. For each re-sampling, we re-assembled the interaction matrix and evaluated the following network parameters: connectance (i.e., proportion of non-empty cells in a matrix), modularity and nestedness ($wNODF$). For modularity and nestedness, we evaluated their their standardised scores, that is, not their crude value but rather how they compare to random expectations, just like we did for the analysis of the whole dataset. So for each re-sampling, we calculated its observed modularity or nestedness, and generated 1000 random matrices using the same null model as for the whole network analysis. We then calculated modularity and nestedness $z - scores$ using the following formula $z = obs - mean_{null}/SD_{null}$ where z is the z-score, obs is the observed score (i.e., modularity or nestedness) and $mean_{null}$ and SD_{null} are mean and standard deviation of the scores calculated for the random matrices.

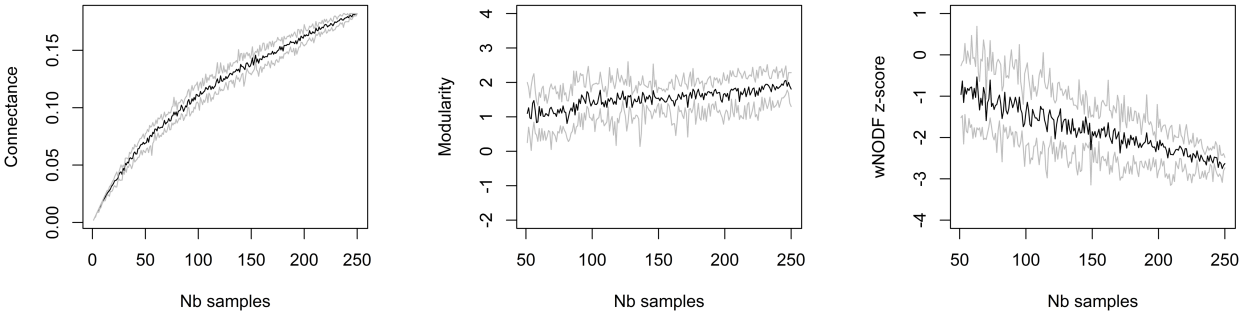

Figure 1: Rarefaction curves

For each curve, black lines indicate the mean $z-score$ for a given re-sampling intensity, and the grey lines indicate the 95% confidence interval. As we can see in Figure 1, all 3 parameters (connectance, modularity and nestedness) covary with sampling effort. However, it is clear from the plots that additional sampling would only have strenghtened the trends that we found. Modularity starts getting significant with 220 thalli, and anti-nestedness was significant with as little as 100 thalli (significance being evaluated with a one-tailed Z-test). Thus, our sampling effort seems to be sufficient to address our questions.

References

[1] R. Guimerà and L. A. Nunes Amaral, "Functional cartography of complex metabolic networks," *Nature*, vol. 433, no. 7028, pp. 895–900, 2005.

[2] C. F. Dormann and R. Strauss, "A method for detecting modules in quantitative bipartite networks," *Methods in Ecology and Evolution*, vol. 5, no. 1, pp. 90–98, 2014.

[3] J.-B. Leger, J.-J. Daudin, and C. Vacher, "Clustering methods differ in their ability to detect patterns in ecological networks," *Methods in Ecology and Evolution*, vol. 6, pp. 474–481, 2015.

[4] P.-L. Chagnon, R. L. Bradley, and J. N. Klironomos, "Using ecological network theory to evaluate the causes and consequences of arbuscular mycorrhizal community structure," *New Phytologist*, vol. 194, pp. 307–312, 2012.

[5] J. C. Carvalho, P. Cardoso, P. A. V. Borges, D. Schmera, and J. Podani, "Measuring fractions of beta diversity and their relationships to nestedness: a theoretical and empirical comparison of novel approaches," *Oikos*, vol. 122, pp. 825–834, 2013.

[6] J. Podani, C. Ricotta, and D. Schmera, "A general framework for analyzing beta diversity, nestedness and related community-level phenomena based on abundance data," *Ecological Complexity*, vol. 15, pp. 52–61, 2013.

[7] M. Almeida-Neto and W. Ulrich, "A straightforward computational approach for measuring nestedness using quantitative matrices," *Environmental Modelling and Software*, vol. 26, no. 2, pp. 173–178, 2011.

[8] W. Ulrich and N. J. Gotelli, "Pattern detection in null model analysis," *Oikos*, vol. 122, no. 1, pp. 2–18, 2013.

[9] C. F. Dormann, J. Frund, N. Bluthgen, and B. Gruber, "Indices, Graphs and Null Models: Analyzing Bipartite Ecological Networks," *The Open Ecology Journal*, vol. 2, no. 1, pp. 7–24, 2009.

[10] P.-L. Chagnon, "Characterizing topology of ecological networks along gradients: The limits of metrics' standardization," *Ecological Complexity*, vol. 22, pp. 36–39, 2015.

[11] D. P. Vázquez, C. J. Melian, N. M. Williams, N. Blüthgen, B. R. Krasnov, and R. Poulin, "Species abundance and asymmetric interaction strength in ecological networks," *Oikos*, vol. 116, no. 7, pp. 1120–1127, 2007.