Appendix:
Relevance of Evolutionary History For Food Web Structure

Anna Eklöf1*, Matthew R. Helmus1, M Moore1 & Stefano Allesina1,2
1Department of Ecology & Evolution, University of Chicago
2Computation Institute, University of Chicago
* Corresponding author: aeklof@uchicago.edu

October 21, 2011

Robustness to changes in data quality

The analyzed food webs were heavily simplified by the filter we used to retain only those nodes with complete taxonomic information. In order to investigate how this affected the reliability of our results, we performed two additional analysis. First, we took the two food webs which were most depleted: Stony (65.5% of the original nodes retained) and ElVerde (44.8% of the original nodes). In the original procedure, we discarded all nodes that were not taxonomically defined up to the genus level. However, several of the discarded nodes were defined up to the order or even family level. Accepting these additional nodes, we retained 95% of the original species in Stony and 87% in ElVerde. We repeated our analysis with this expanded set of nodes and links and calculated the likelihoods and marginal likelihoods up to order and/or family level (Table A1). We found the same taxonomic classification as in the original setting to be the “best” grouping.
Second, we took the largest of the analyzed webs, Weddell Sea (in which 74.6% of the original nodes were retained), and randomly removed nodes, mimicking the lack of taxonomic information. We removed 10, 20, 30, 40 and 50% of the retained nodes, and performed 10 replicates each. When we removed 10, 20 and 30% of the nodes, we still found the same taxonomic classification as in the original setting (class) to yield the highest marginal likelihood for all replicates (Table A2). For the 40 and 50% removal, the results shifted toward a lower taxonomic rank (phylum) for some of the replicates. We conclude that the filter does not affect the main conclusions of our study unless the data is too degraded to analyze.

<table>
<thead>
<tr>
<th>Food Web</th>
<th>K</th>
<th>P</th>
<th>C</th>
<th>O</th>
<th>F</th>
<th>G</th>
<th>Best</th>
</tr>
</thead>
<tbody>
<tr>
<td>ElVerde</td>
<td>-4593.4</td>
<td>-4148.9</td>
<td>-4099.6</td>
<td>-6147.2</td>
<td>-9358.1</td>
<td>-11119.3</td>
<td>C</td>
</tr>
<tr>
<td>Stony</td>
<td>-1863.3</td>
<td>-1866.4</td>
<td>-1949.6</td>
<td>-3032.3</td>
<td>-4534.5</td>
<td>-6577.7</td>
<td>K</td>
</tr>
</tbody>
</table>

Table A 1: Log marginal likelihoods from the taxonomic split into kingdom (K), phylum, class (C), order (O), family (F) and genus (G) obtained from the experiments including a larger number of nodes and links. Last column (Best) indicate the split giving the highest log marginal likelihood.
Table A 2: Mean and standard deviation (σ) of the log marginal likelihood obtained from ten replicates of the Weddell Sea where 10-50 % of the nodes were removed at random. Results shown for kingdom (K), phylum (P), class (C), order (O), family (F) and genus (G) respectively.

<table>
<thead>
<tr>
<th>%nodes removed</th>
<th>mean K</th>
<th>σK</th>
<th>mean P</th>
<th>σP</th>
<th>mean C</th>
<th>σC</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>-25847.0</td>
<td>483.3</td>
<td>-20563.4</td>
<td>487.3</td>
<td>-18658.7</td>
<td>323.1</td>
</tr>
<tr>
<td>20</td>
<td>-20890.9</td>
<td>543.5</td>
<td>-16674.7</td>
<td>478.6</td>
<td>-15401.9</td>
<td>399.1</td>
</tr>
<tr>
<td>30</td>
<td>-15552.0</td>
<td>541.7</td>
<td>-12619.2</td>
<td>423.9</td>
<td>-12175.4</td>
<td>305.1</td>
</tr>
<tr>
<td>40</td>
<td>-11461.7</td>
<td>884.3</td>
<td>-9339.1</td>
<td>616.1</td>
<td>-9301.6</td>
<td>528.2</td>
</tr>
<tr>
<td>50</td>
<td>-7937.4</td>
<td>391.2</td>
<td>-6564.1</td>
<td>348.8</td>
<td>-6821.2</td>
<td>217.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>%nodes removed</th>
<th>mean O</th>
<th>σO</th>
<th>mean F</th>
<th>σF</th>
<th>mean G</th>
<th>σG</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>-24319.5</td>
<td>454.6</td>
<td>-37492.3</td>
<td>563.8</td>
<td>-54229.6</td>
<td>271.8</td>
</tr>
<tr>
<td>20</td>
<td>-20463.0</td>
<td>479.3</td>
<td>-31403.1</td>
<td>835.1</td>
<td>-44492.6</td>
<td>975.1</td>
</tr>
<tr>
<td>30</td>
<td>-16781.7</td>
<td>439.7</td>
<td>-25288.0</td>
<td>667.3</td>
<td>-34810.1</td>
<td>623.6</td>
</tr>
<tr>
<td>40</td>
<td>-13076.6</td>
<td>494.2</td>
<td>-19711.9</td>
<td>614.1</td>
<td>-26283.4</td>
<td>484.2</td>
</tr>
<tr>
<td>50</td>
<td>-9914.3</td>
<td>209.5</td>
<td>-14465.4</td>
<td>453.1</td>
<td>-18861.6</td>
<td>349.6</td>
</tr>
</tbody>
</table>

Robustness to choice of prior

In order to analyze whether the choice of the prior distribution for the probabilities affected our result, we preformed the same analysis computing marginal likelihoods for all webs and all partitions when probabilities come from a Beta distribution rather than a uniform distribution. The Beta distribution $\mathbb{B}(\alpha, \beta)$ is quite flexible, and encompass instead of the uniform distribution as a special case $\mathbb{B}(1, 1)$.

However, the use of a Beta distribution introduces the need for the “hyper-priors” $\alpha$ and $\beta$. Because we do not have information on these hyper-priors, we searched the space choosing the values that maximize the marginal likelihood.

In a random directed graph, the marginal likelihood of the data when the probability comes from a Beta prior distribution can be written as:
\[ \mathcal{M}_{\text{Rnd}} = \int_0^1 p^L(1 - p)^Z p^{\alpha - 1}(1 - p)^{\beta - 1} \frac{1}{B(\alpha, \beta)} dp = \]
\[ = \frac{1}{B(\alpha, \beta)} \int_0^1 p^{L + \alpha - 1}(1 - p)^{Z + \beta - 1} dp = \]
\[ = \frac{B(\alpha + L, \beta + Z)}{B(\alpha, \beta)} \]

where \( p \) is the probability of any two nodes being connected, \( L \) is the number of ones in the matrix, \( Z \) is the number of zeros in the matrix and \( B(\cdot, \cdot) \) is the Beta function.

We used the \texttt{optim} package in R in order to maximize the marginal likelihood by tuning \( \alpha \) and \( \beta \). The starting values for both parameters were set to 1.0 and the maximum number of iterations was set to 5000.

We found that the choice of prior does not affect the results: the model selection process returns the same taxonomic ranks as the best model for both priors in all webs (see Table A3).

**Comparison with partitions based body mass and trophic levels**

In order compare the explanatory power of the taxonomic ranking of the species with some other relevant ranking we chose to analyze trophic level and body mass by dividing the species into groups based on these factors. Trophic levels of the species can be calculated for all food webs. We used the prey averaged trophic level [Williams and Martinez, 2004]. The species body sizes were available for three of the food webs analyzed here.

To ensure a fair comparison between the marginal likelihoods based on taxonomy and those produced by grouping in trophic level and body size, ideally one would choose to have the same complexity in all cases. Therefore, we divided the species into the same number of groups based on their trophic
Table A 3: Comparison of the marginal likelihoods for each food web and partition into kingdom (K), phylum (P), class (C), order (O), family (F) and genus (G) when an uniform respectively beta distribution is used. The columns with B shows the values when a beta distribution is used, the other clumsy show the values when a uniform distribution is used. In the last columns we report which taxonomic rank provides the best value (Best and BBest) for each web.
level/body size as used in the analysis of taxonomy. For example, in Weddell Sea we have 3 kingdoms, 20 phyla, 41 classes, 89 orders, 170 families and 258 genera. Hence, we divided the species into 3, 20, 41, 89, 170 and 258 groups respectively based on information on their trophic level/body size. In some cases there were not as many unique trophic levels/body sizes as there were groups based on taxonomy. These cases removed from further analysis. We used $k - \text{means}$ clustering [MacQueen et al., 1967] which divides $n$ observations into $k$ clusters ($k \leq n$) based on the nearest mean. Unfortunately, the method conserves the number of groups, but not their sizes, thus biasing the analysis. We still report the results, as they can be seen as an approximation of the ideal case in which we could compare the different groupings directly.

The results are inconclusive: trophic-level-based groups preform better than taxonomy in five of the nine webs and body-mass yields better results in one of the three webs tested. We conclude that taxonomy provides an explanation power comparable to that of trophic level and body size. This point should be treated more stringently in future work, for example finding a better way of partitioning the species using continuous traits, and with a larger data set.
### Table A4: Marginal likelihood for groupings based on trophic level.

The best value are in bold, and an asterisk indicates that the value is better than that for the corresponding taxonomy-based grouping. Numbers in parenthesis show the number of groups used. A dash indicates that the species could not be divided into the desired number of groups and the analysis could therefore not be preformed.

<table>
<thead>
<tr>
<th></th>
<th>Web</th>
<th>K</th>
<th>P</th>
<th>C</th>
<th>O</th>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weddell</td>
<td>-30395.7(3)</td>
<td>-34516.5(3)</td>
<td>-24195.3(20)</td>
<td>-24932.2(41)</td>
<td>-28871.5(89)</td>
<td>-28871.5(170)</td>
<td>-28871.5(170)</td>
</tr>
<tr>
<td>Caribbean</td>
<td>-7710.4(1)</td>
<td>-1115.0(3)</td>
<td>-7710.4(1)</td>
<td>-7020.0(2)</td>
<td>-6094.6(17)</td>
<td>-10715.3(60)</td>
<td>-18136.9(120)</td>
</tr>
<tr>
<td>Littlerock</td>
<td>-4315.4(3)</td>
<td>-1044.9</td>
<td>-3858.3(13)</td>
<td>3713.9(16)*</td>
<td>1055.7(9)</td>
<td>-1590.0(27)</td>
<td>-1721.4(12)</td>
</tr>
<tr>
<td>Ythan96</td>
<td>-1732.1(1)</td>
<td>-982.5(2)</td>
<td>-1522.5(7)*</td>
<td>-1721.4(12)</td>
<td>2778.7(34)</td>
<td>-3894.4(58)</td>
<td>-1721.4(12)</td>
</tr>
<tr>
<td>Ythan91</td>
<td>-1164.9(1)</td>
<td>-941.5(4)</td>
<td>-967.4(4)*</td>
<td>-1055.7(9)</td>
<td>-1590.0(27)</td>
<td>-1291.6(20)</td>
<td>-967.4(4)</td>
</tr>
<tr>
<td>Mill</td>
<td>-982.5(2)</td>
<td>-941.5(4)</td>
<td>-940.0(9)*</td>
<td>-1291.6(22)</td>
<td>6383.0(40)</td>
<td>-7505.7(56)</td>
<td>-941.5(4)</td>
</tr>
<tr>
<td>Elverde</td>
<td>-1007.0(2)</td>
<td>-550.1(6)*</td>
<td>-471.6(10)</td>
<td>-822.3(10)</td>
<td>-1209.6(20)</td>
<td>-1863.0(35)</td>
<td>-550.1(6)</td>
</tr>
<tr>
<td>Stony</td>
<td>-603.8(2)</td>
<td>-747.2(5)</td>
<td>-471.6(9)</td>
<td>-847.6(9)</td>
<td>-1406.0(22)</td>
<td>-2140.2(40)</td>
<td>-747.2(5)</td>
</tr>
</tbody>
</table>

### Table A5: As Table A4 but using body sizes instead of trophic levels.

<table>
<thead>
<tr>
<th></th>
<th>Web</th>
<th>K</th>
<th>P</th>
<th>C</th>
<th>O</th>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weddell</td>
<td>-34516.5(3)</td>
<td>-32861.4(20)</td>
<td>-30418.6(41)</td>
<td>-37880.8(89)</td>
<td>-51601.0(170)</td>
<td>-51601.0(170)</td>
<td>-51601.0(170)</td>
</tr>
<tr>
<td>Tuesday</td>
<td>-1115.0(3)</td>
<td>-796.8(9)*</td>
<td>-822.3(10)</td>
<td>-1209.6(20)</td>
<td>-1863.0(35)</td>
<td>-2368.1(47)</td>
<td>-2368.1(47)</td>
</tr>
<tr>
<td>Mill</td>
<td>-1007.0(2)</td>
<td>-747.2(5)</td>
<td>-847.6(9)</td>
<td>-1406.0(22)</td>
<td>-2140.2(40)</td>
<td>-2698.5(54)</td>
<td>-2698.5(54)</td>
</tr>
</tbody>
</table>
Figure A 1: Matrix plot describing Ythan Estuary (96) [Hall and Raffaelli, 1991] food web where the species are sorted according to their taxonomy. Red lines show the divisions between different phyla, green lines show the divisions between different classes. Black squares indicate an feeding interaction. Note that lines for phyla are of overlapping the lines for classes. Above the matrix are the names of the phyla presented, to the right the names of the classes.
References

