Biodiversity Comparison between Fungal Communities in Urban and Natural Ecosystems

Olivier Binette, Yunran Chen, Carol Wang

Duke University

March 24, 2020

Introduction

- **Study Design**
  - Fungal communities are sampled from 5 pairs of sites, each representing an urban area and the surrounding natural area.
  - Each site contains 3 plots representing both the core and edges.
  - From each plot, 3 replicate air and 3 replicate soil samples were taken.
  - Operational Taxonomic Units (OTUs) are identified based on the ITS2 region of each sample.
- **Question**

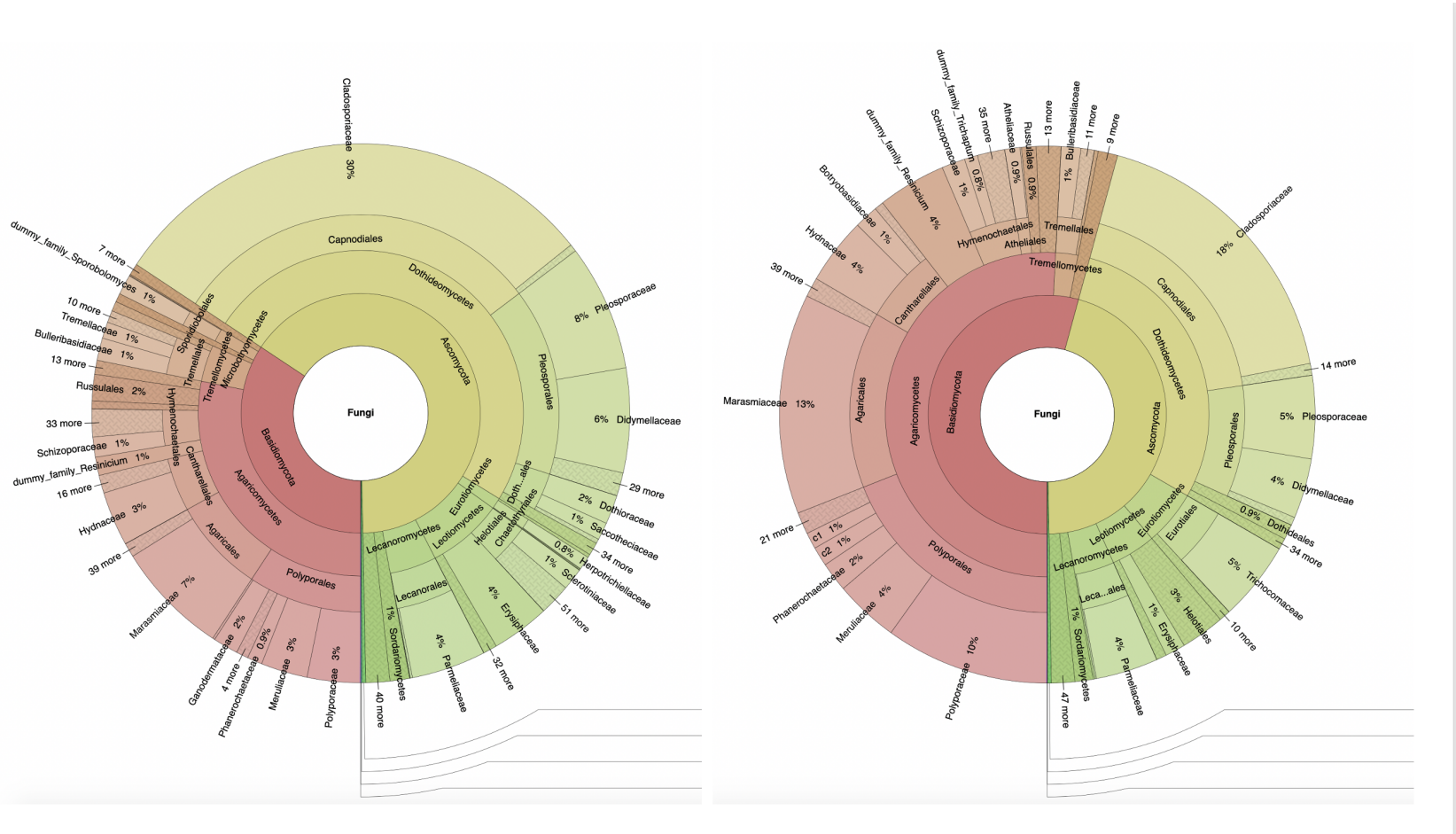
  How do natural and urban ecosystems differ in biodiversity of fungal communities?

Introduction

- **Data Formatting**
  - OTU table: 7940 OTUs and 174 samples
  - Taxonomy table: 7940 OTUs by 7 taxonomic ranks, from kingdom to species
  - Sample data: 174 samples by 9 sample variables
- **Challenges**
  - OTU counts: high-dimensional, sparse, complicated covariance structure, large within-group variability
  - Multiple scales of biodiversity
  - Bias in species richness arisen from sequencing

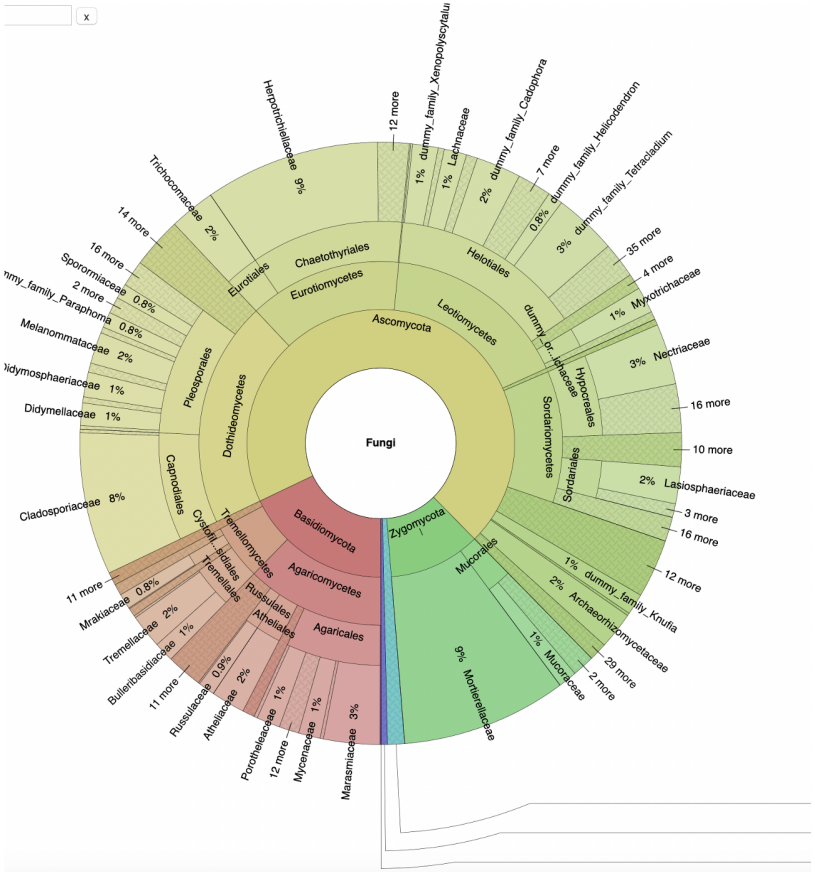
Krona Wheel

Comparison between urban and natural fungi in air (aggregated by count)



Krona Wheel

Comparison between urban and natural fungi in soil (aggregated by count)

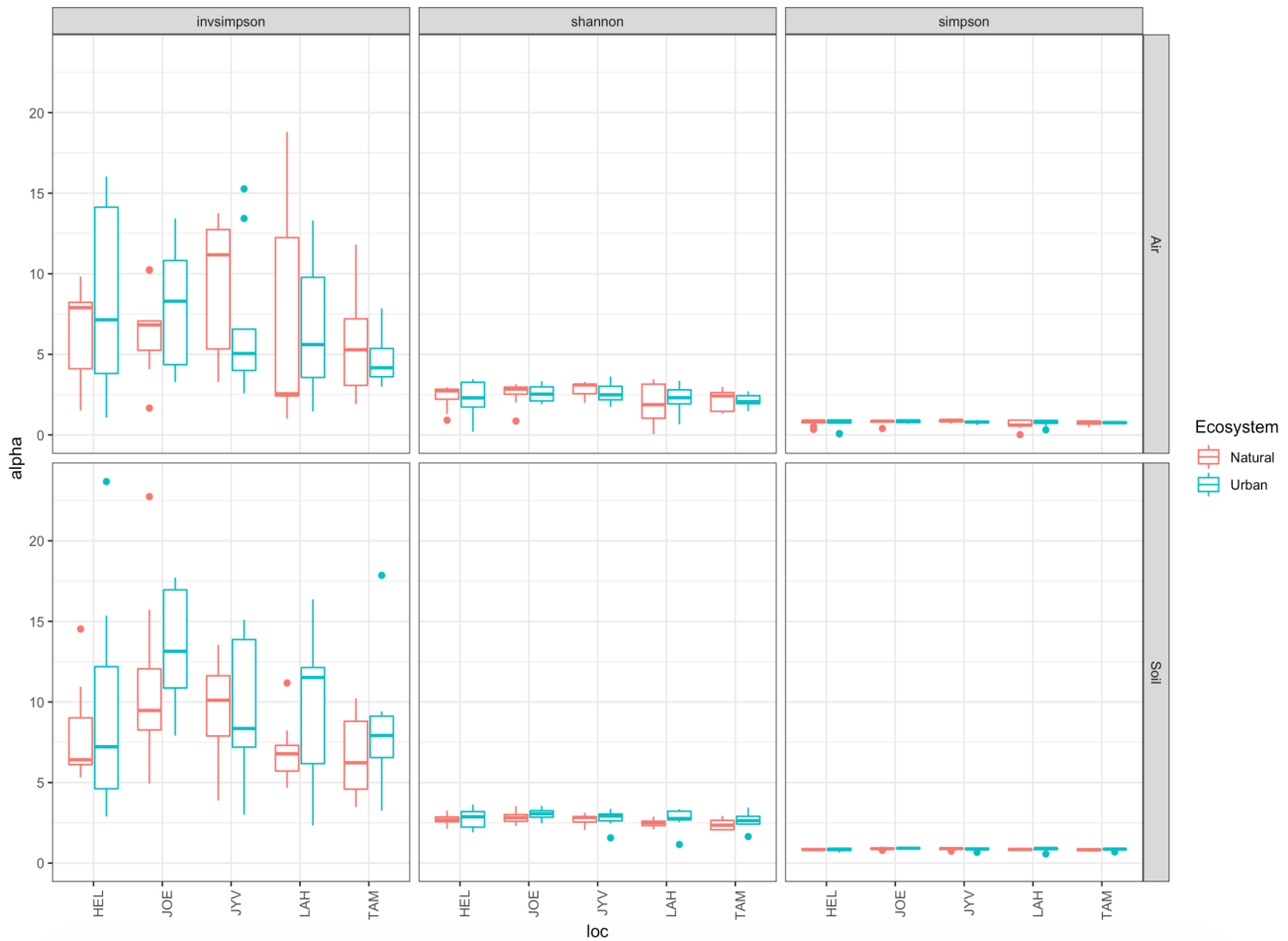


Krona Wheel

- Biodiversity of air and soil samples are significantly different.
- Biodiversity distributed different between urban and natural samples. (More evenly at the first level but more concentrated at the higher levels for natural samples.)
- Large within-group variability. [Example: HEL-N1]

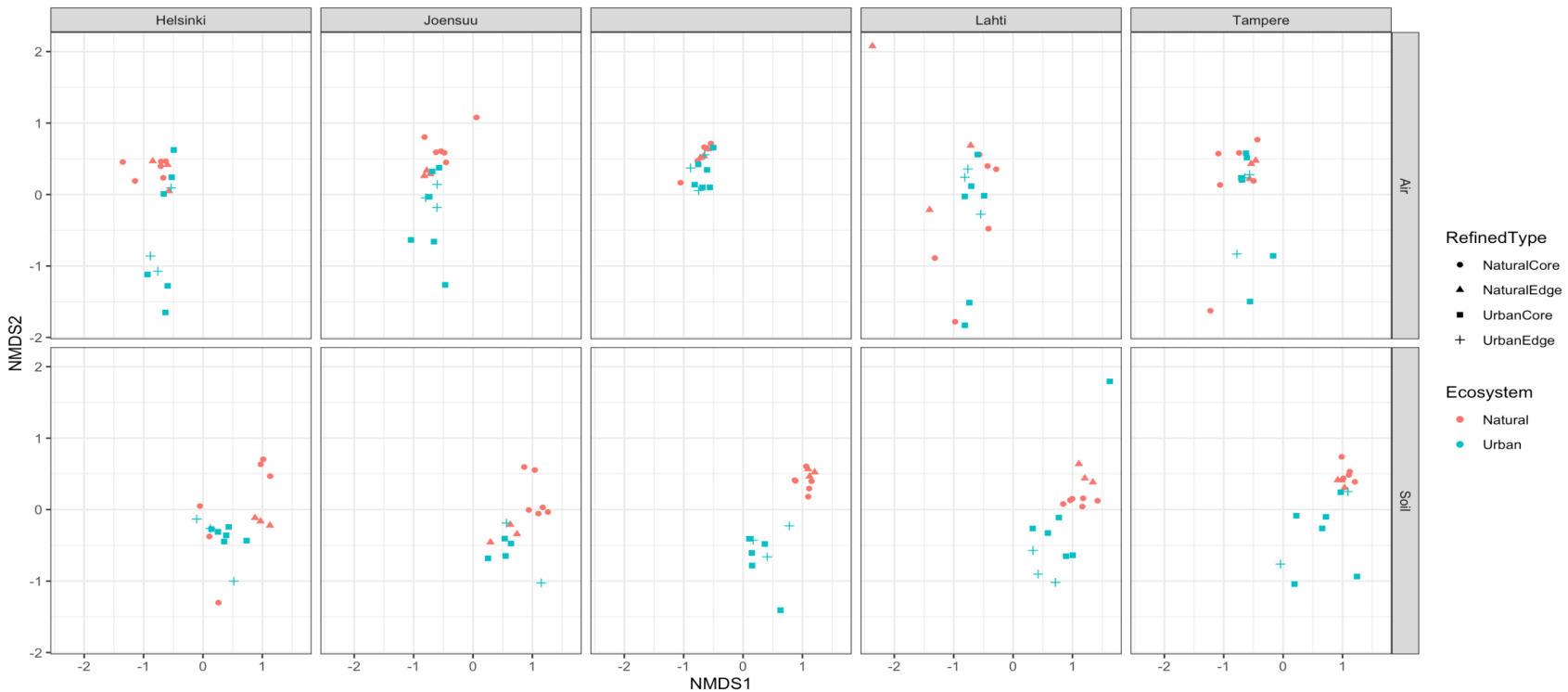
Alpha Diversity

- Overlapping.
- ANOVA test shows Type (Air vs Soil) effects are significant non-zero (<0.001).



Ordination(Nonmetric Multidimensional Scaling)

- Strong evidence supporting clear difference between natural and urban samples for soil samples.



EDA:Conclusion

- Take type (soil/air) into account.
- An intuitive measurement representing both richness and evenness or more.
- Descriptive analysis based on diversity profile differences.

Hill numbers and diversity profiles

Hill numbers:

$$ {}^qD = \left(\sum_i p_i^q \right)^{1/(1-q)} $$

with ${}^0D = \#\{i : p_i > 0\}$, ${}^1D = \exp(H(p))$, ${}^\infty D = \#\{i : p_i = \max(p)\}$.

- Intepretable rescaling of Rényi entropies as "effective number of species".
- Characterizes the distribution of p_I where $\mathbb{P}(I = i) = p_i$.
- The curve of ${}^qD$ against q is the *diversity profile* of p.

Challenge: standard packages to estimate Hill numbers are too slow in our context. We instead do direct computations with Rcpp, without anything fancy.

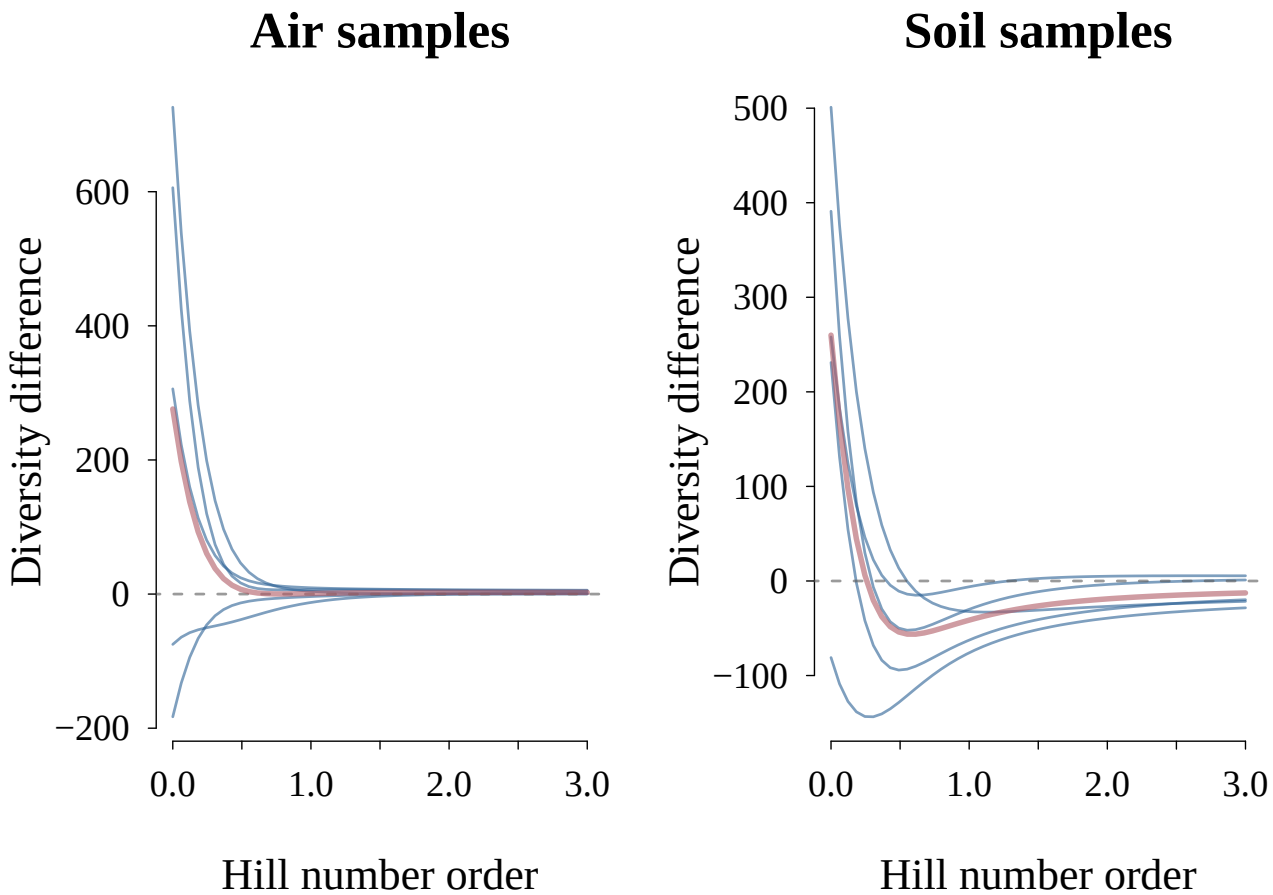
Diversity comparison

What we did:

- For each location, ecosystem and sample type (air or soil), we computed diversity profiles based on aggregating probabilities of the corresponding samples.
- We substract the profile of the **urban** ecosystem to the profile of the **natural** ecosystem.
- This can be done at each level of the phylogeny. The profiles difference showcases the **same pattern at different levels** and is **robust** to choices in data cleaning.

Diversity comparison

Differences of γ-diversity profiles between natural and urban ecosystems, at the five locations and at the fourth level of the phylogeny.



Issues and limitations

Inter-related limitations:

1. No bias correction for unseen species.
2. No uncertainty quantification.
3. Unclear how to aggregate samples.

It is easy to come up with wrong UQ, but hard to fix these things properly. Reason is:

- context of informative missingness: the data provides no information about unseen species without strong (untestable) assumptions.

We'd need expert knowledge combined with very careful modelling to properly combine samples, extrapolate to the whole population, and do (say) bootstrap uncertainty quantification.

I think it's doable, but we didn't get to it.