

Sampling from the Environment: The Problem of Hyperdiversity

Brendan Bohannon

University of Oregon

Overview

I. Hyperdiversity and undersampling: what are they and why are they a problem?

II. Reducing undersampling through brute force and finesse.

III. Estimating “unseen taxa” in undersampled environments: three approaches.

IV. The future of diversity estimation.

I. Hyperdiversity and undersampling

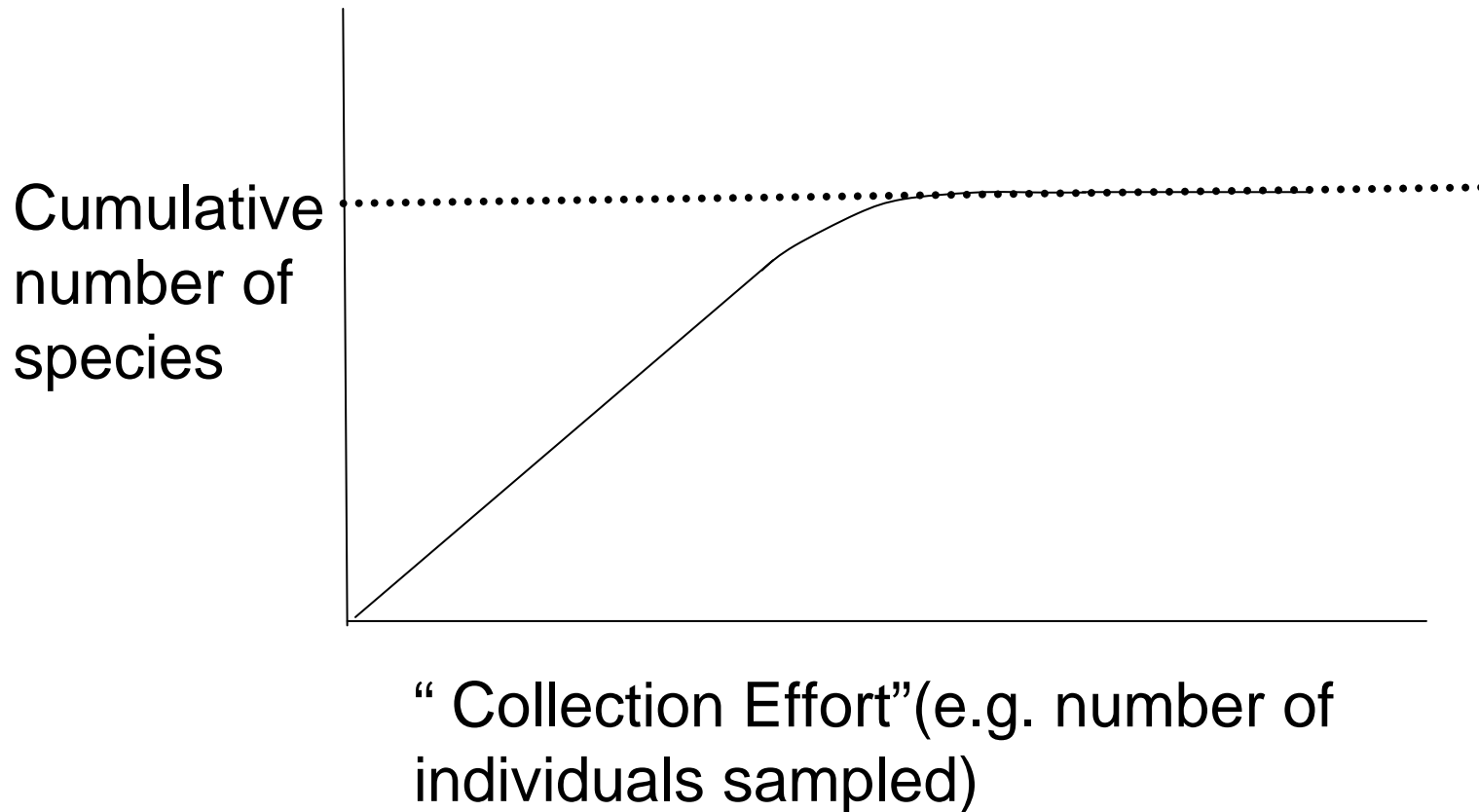
What is hyperdiversity?

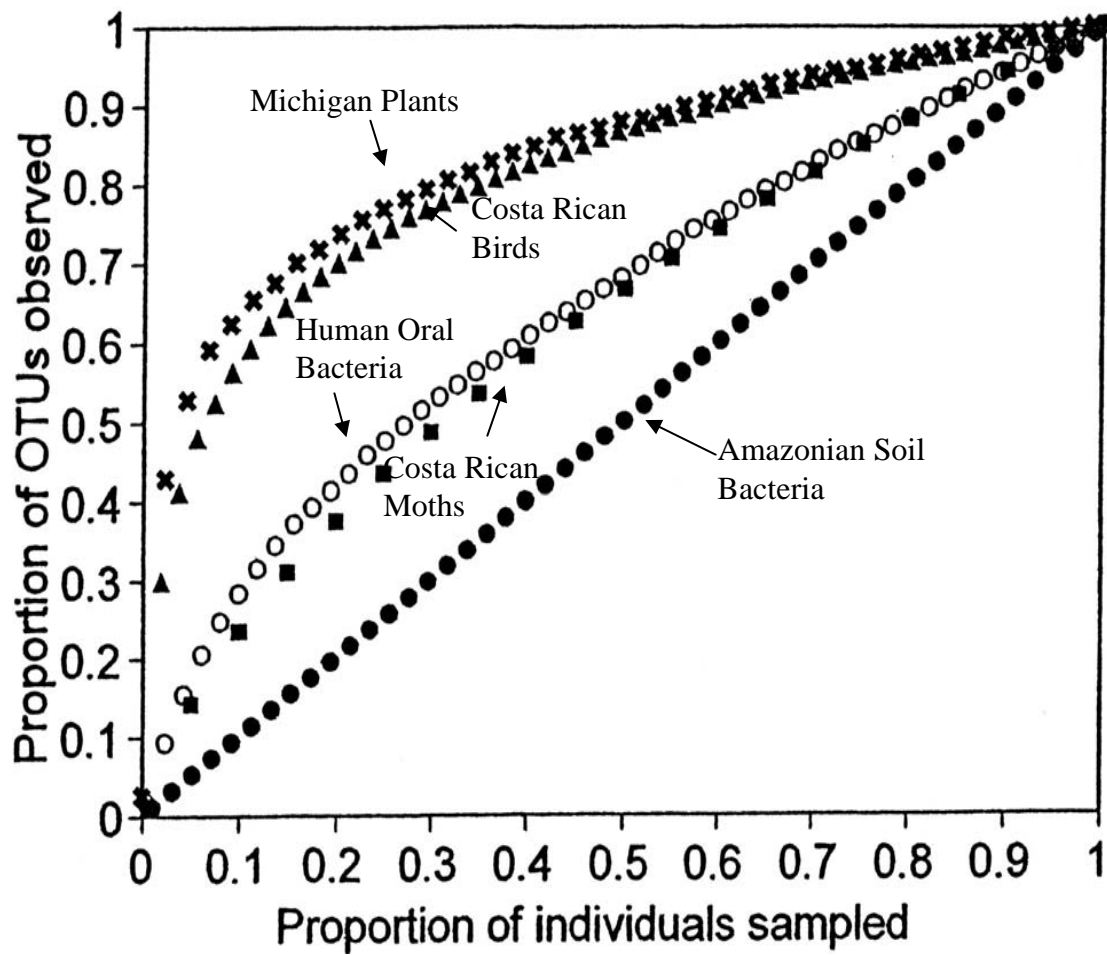
1. More taxa in an environment than one expects (based on other surveys of diversity, e.g. of plants and animals).

2. More taxa in an environment than one can exhaustively sample.

Are fungi hyperdiverse?

The *collector's curve* or *accumulation curve*





Are fungi hyperdiverse?

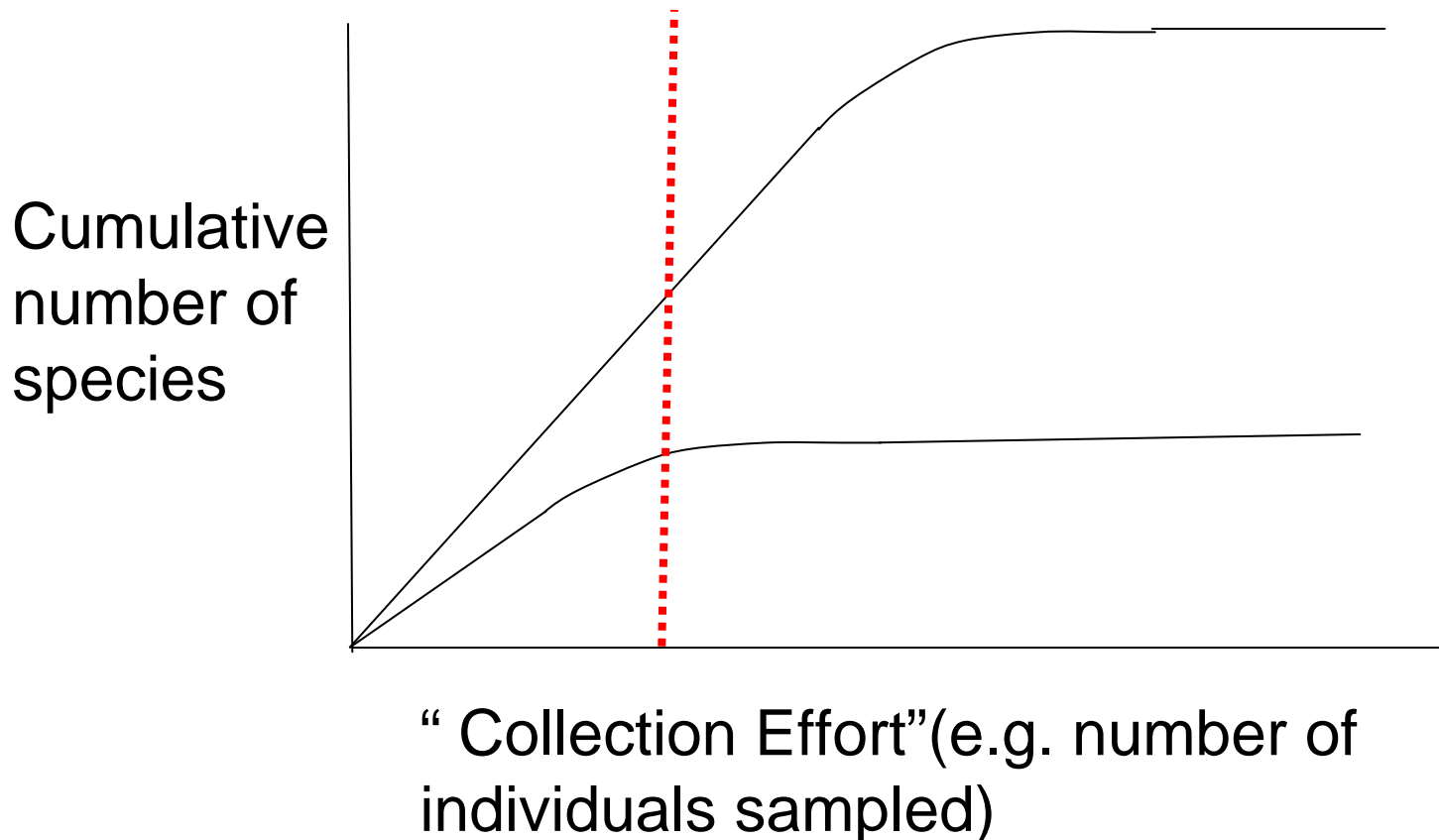
Fungal accumulation curves

QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.

(O'Brien et al. 2005)

Hyperdiversity leads to *undersampling*

Imperfect detection of taxa, often due to sampling only the most abundant taxa in an environment.



Why does undersampling matter?

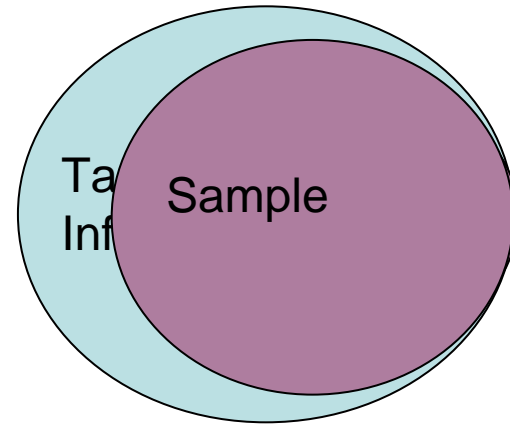
1. Incomplete knowledge of diversity.
2. Abundant taxa are often widespread, leading to severe underestimation of endemic taxa.
3. Endemic taxa may be those most sensitive to environmental change.
4. Etc. Etc.

What is the solution?

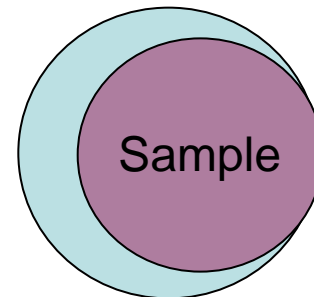
II. Reducing undersampling

Two ways to do this:

1. **Increasing sample size.**



2. **Decreasing the “target of Inference”.**



Increasing sample size

**High-throughput methods (e.g.
Illumina, 454 Life Sciences, etc.).**

QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.

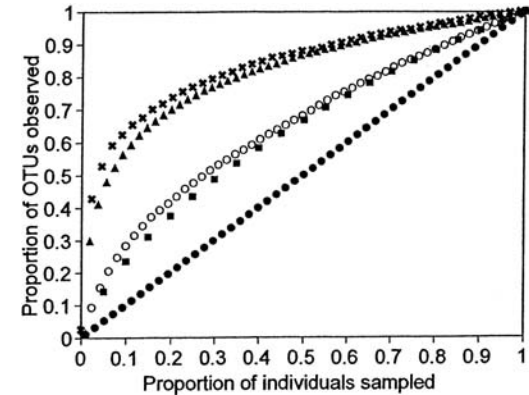
Decreasing target of inference

1. **Increasing taxonomic focus:** e.g. EMF rather than soil fungi.
2. **Decreasing taxonomic resolution:** e.g. genus or family rather than species or strain.
3. **Constraining area or volume of target:** e.g. choosing a endophytes in a single leaf as the target rather than endophytes across an entire plant.

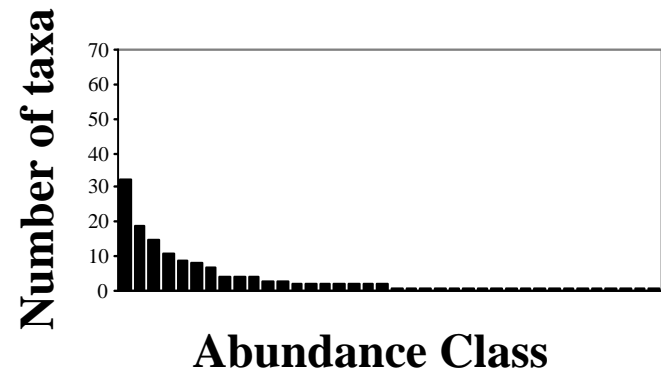
III. Estimating “unseen taxa”

3 major sources of information:

1. The shape of accumulation curves.



2. The shape of taxa-abundance relationships.



3. The proportion of rare taxa.

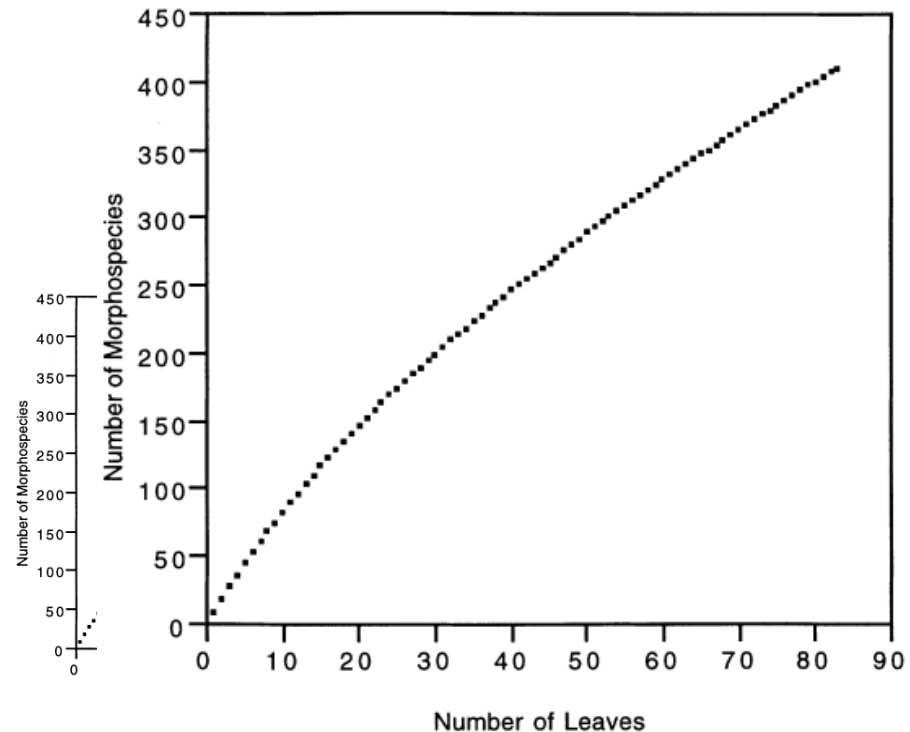
Accumulation curves

How used:

Can fit curves to the data and extrapolate actual richness

Limitations:

1. Microbial curves often do not plateau.
2. Not always obvious what function to fit to data.
3. Curves can cross (e.g. if communities vary in evenness).



Fungal example

QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.

(Girlanda et al. 2001)

Shape of taxa-abundance relationships: limitations

1. Requires an understanding of microbial community assembly.
2. Usually not one “best fit” distribution.

Proportion of rare taxa

How used: via “nonparametric estimators”.

Examples include:

1. Good’s coverage estimator

$$C = 1 - (s/n)$$

where s is number of singletons and n is number of individuals

2. Chao1 estimator

$$S_{chao} = S_{obs} + s/2d$$

where s is number of singletons and d is number of doubletons.

3. Abundance-based coverage estimators (ACE)

Fungal example

QuickTime™ and a
TIFF (Uncompressed) decompressor
are needed to see this picture.

(O'Brien et al. 2005)

Proportion of rare taxa: limitations

1. Usually only provides a lower bound on richness (e.g. Colwell & Coddington 1994, Hughes et al. 2001, Bohannan & Hughes 2003, Hong et al. 2006).

2. Requires reliable estimates of relative taxon abundance.

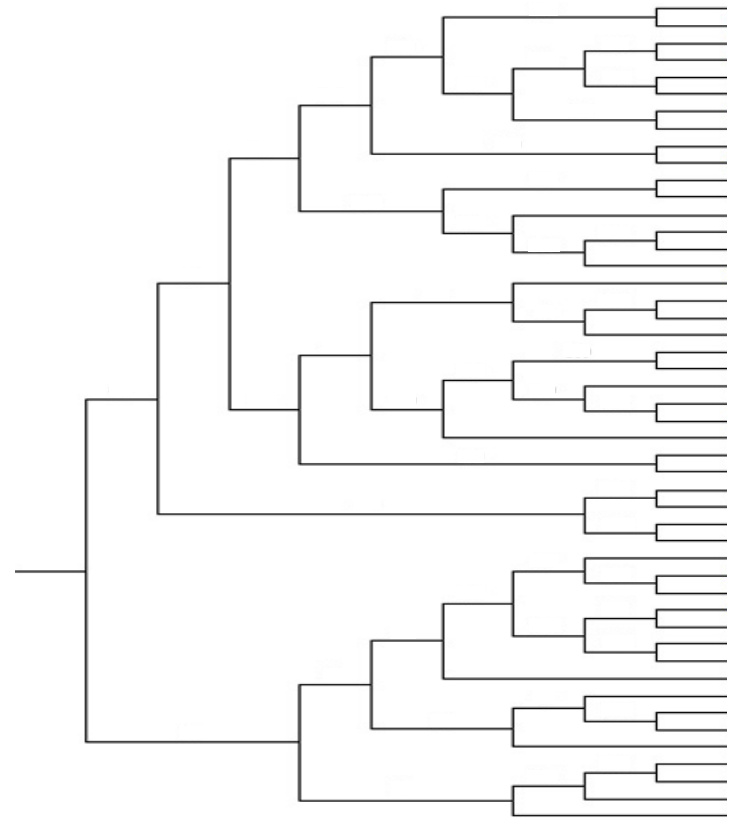
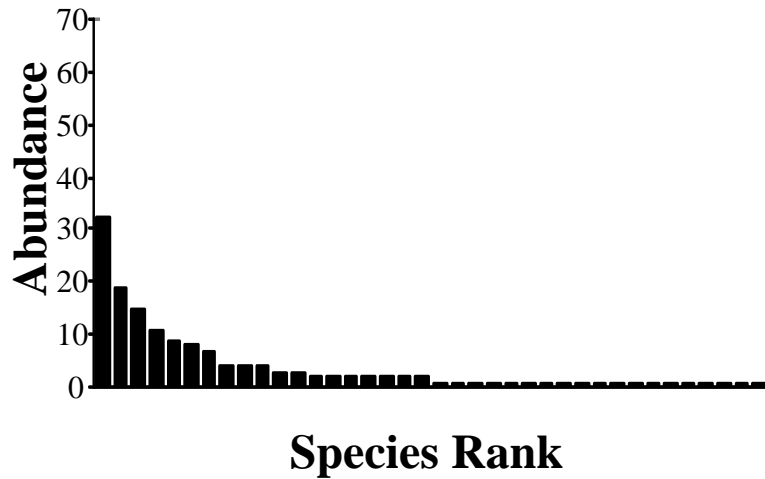
Recommendations

To confront the problem of hyperdiversity, we need to:

1. Reduce undersampling (through both increasing sample sizes and decreasing our target of inference).
2. Use multiple approaches to estimate “unseen species” (extrapolation, and both parametric and nonparametric estimation).
3. Ask questions about relative rather than absolute diversity.

IV. The future of microbial diversity estimation?

Molecules are more than markers: estimators of phylogenetic diversity



The ESA Microbial Ecology Section

“Supporting science at the interface of microbial
and general ecology”

Sponsored Symposium: “Have microbes read the
book? Theoretical ecology for microbial
communities” Tuesday, August 5, 2008, 8:30 AM-
11:15 AM.

Business meeting: Tuesday, August 5, 2008, 11:30
AM-1:15 PM.

Do richness estimators “work” for microbial communities?

Yes...they are an improvement over comparing observed richness in undersampled communities.

No...they cannot accurately estimate absolute richness in grossly undersampled communities.

2. Trade-off extensive sampling for intensive sampling.

