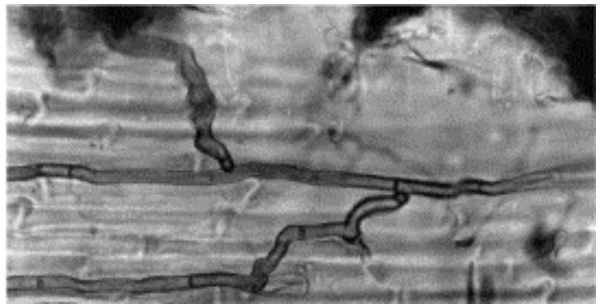
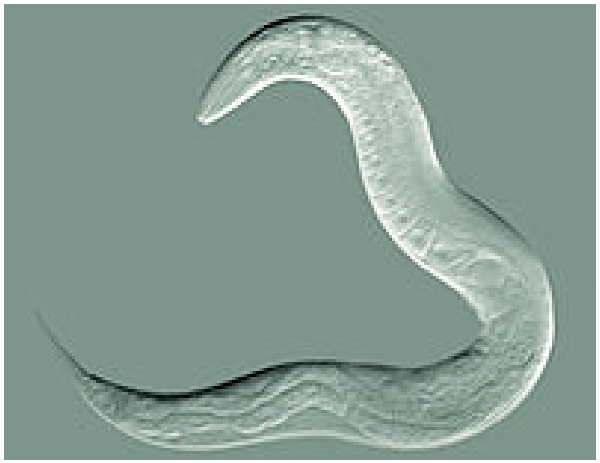
Exploring taxon-specific barcode attributes

Zachary Foster

June 6, 2014

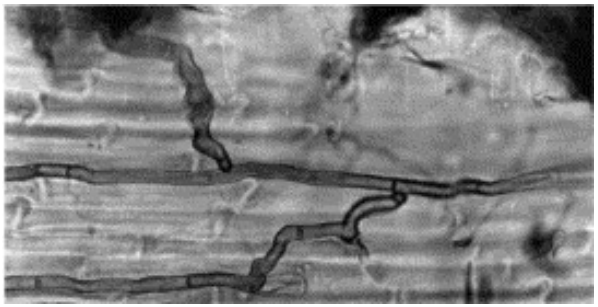
Benefits of Metabarcoding

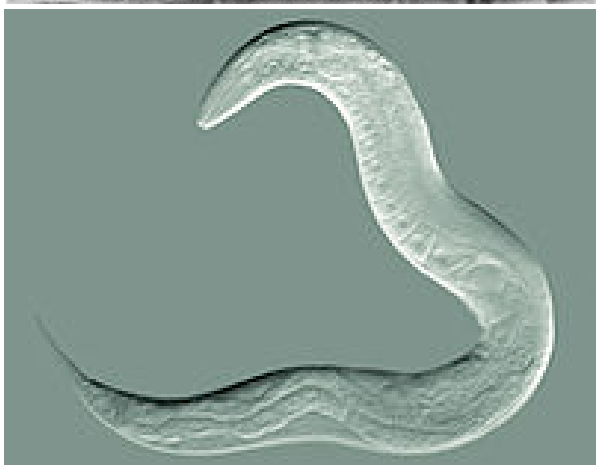




- Targeted, yet culture independent
- Samples at much greater depth than shotgun metagenomics
- Established and curated databases available
- Relatively easy to analyze results
- Relatively east to construct reference databases using standard barcoding

Biases of Metabarcoding





- Relies on reference databases for sequence identification.
- PCR primer specificity
- Small sequence size
- Barcode locus characteristics

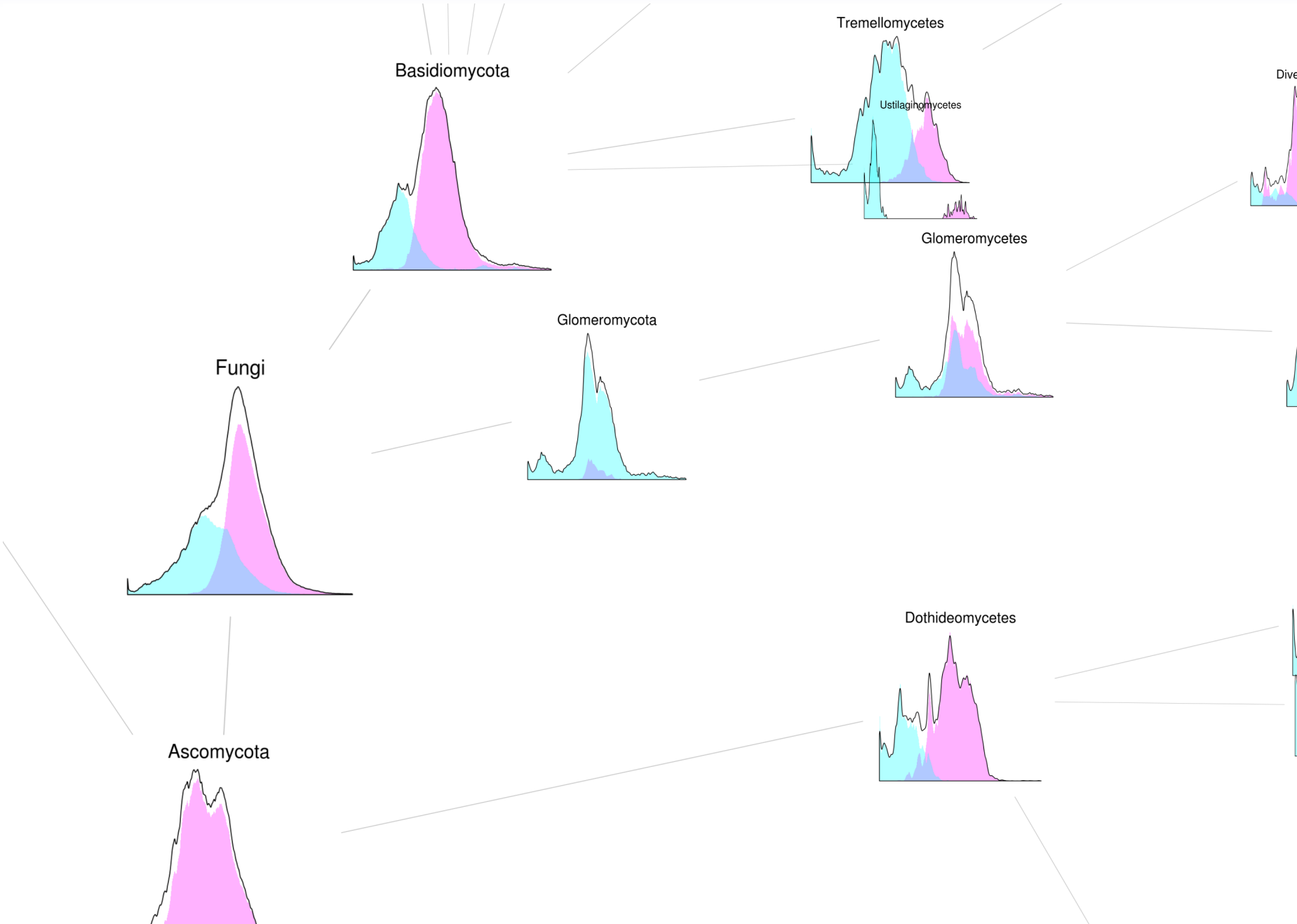
Implied assumptions of OTU clustering and diversity estimation

1. Evolution rate of the barcode locus is constant throughout group
2. Taxonomic levels correspond to divergences of similar degree
3. Different barcode loci yield comparable results
4. Different primers yield comparable results

Process overview

1. Obtain multiple sequence alignment with taxonomy information from RDP
2. Sub sample multiple sequence alignment based on taxonomy information
3. Use fdnadist to calculate a pair-wise distance matrix
4. Sub sample distance matrix for each taxon at each taxonomic level
5. Calculate taxon-specific statistics

Taxon intra vs inter distance distributions



Barcoding Gap and Clustering Threshold Calibration

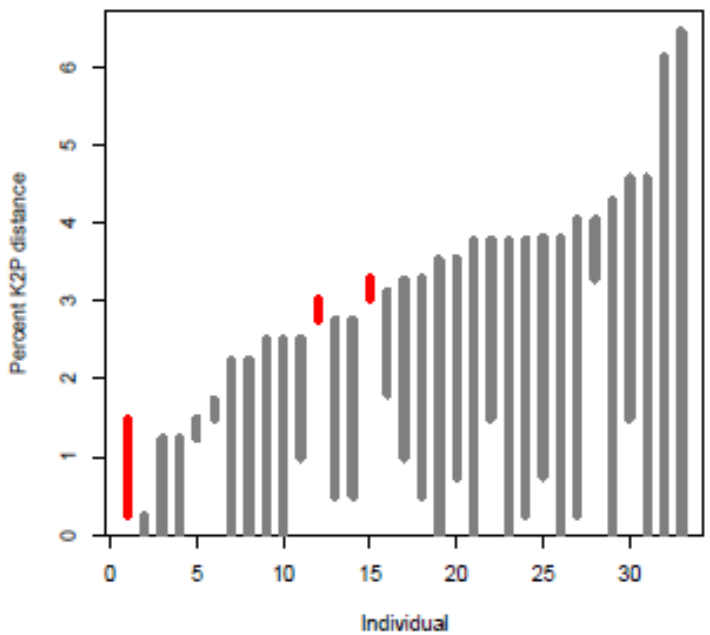


Figure 1: Barcode gap analysis (Brown et al. 2012)

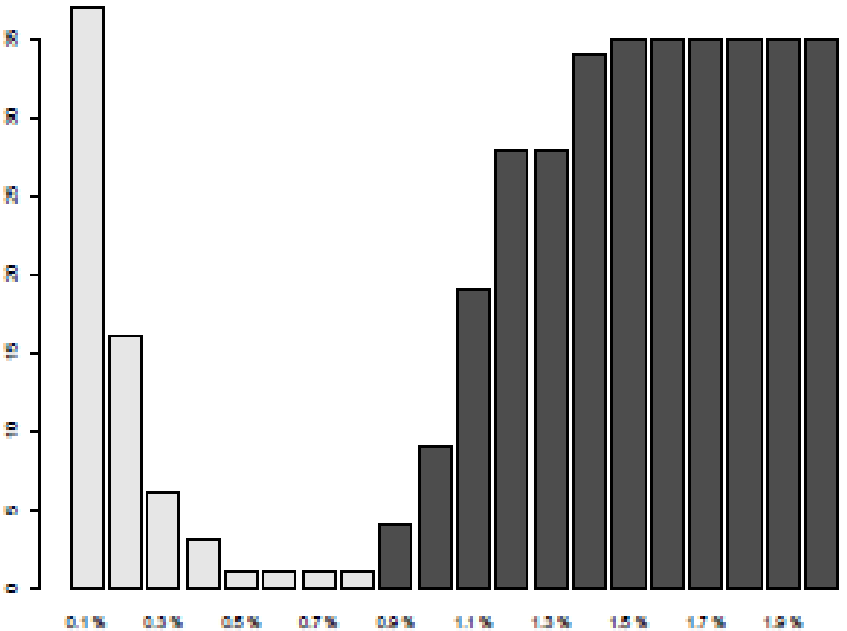


Figure 2: Clustering threshold optimization (Brown et al. 2012)

Looking for Patterns with NMS and PCA

- Dimension reduction can be used to visualize cryptic associations between species abundance and sample characteristics.
- Principal component analysis (PCA)
- Non-metric multidimensional scaling (NMS)

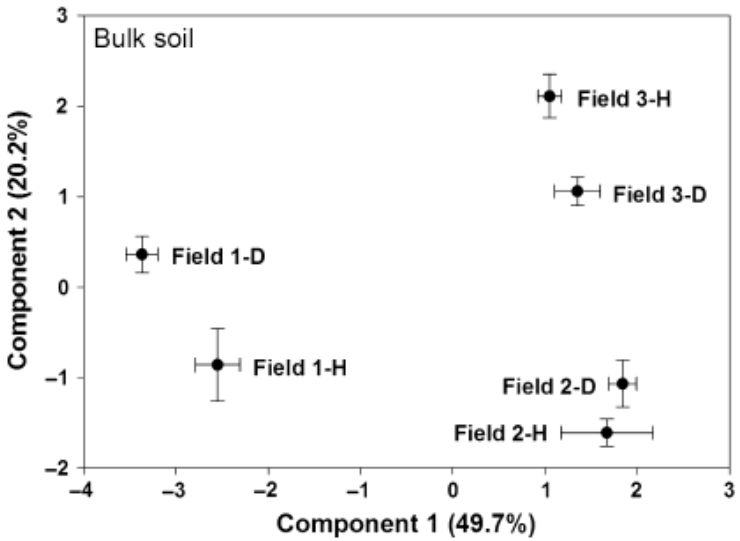


Figure 3: PCA of pea fungal diversity (Xu et al. 2012)

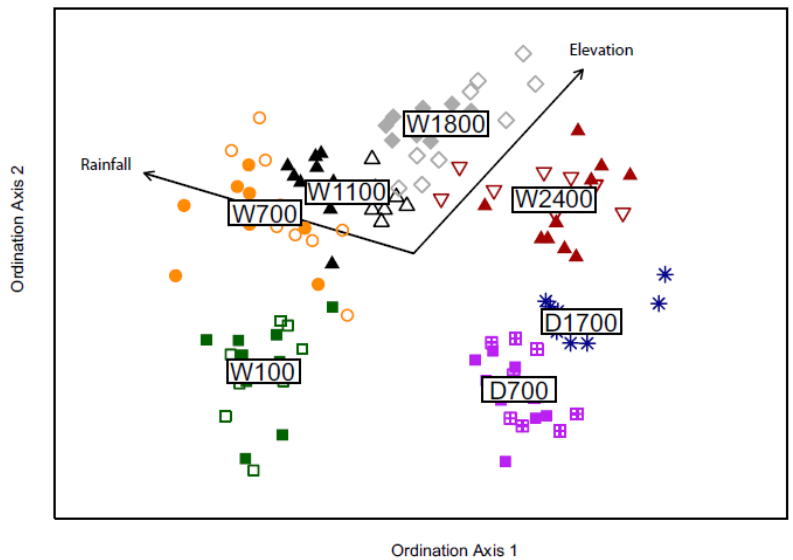


Figure 4: NMS analysis of foliar fungal diversity (Zimmerman and Vitousek 2012)

Taking into account Primers with in silico PCR Amplification

Primer choice is crucial to minimize taxa-specific bias.

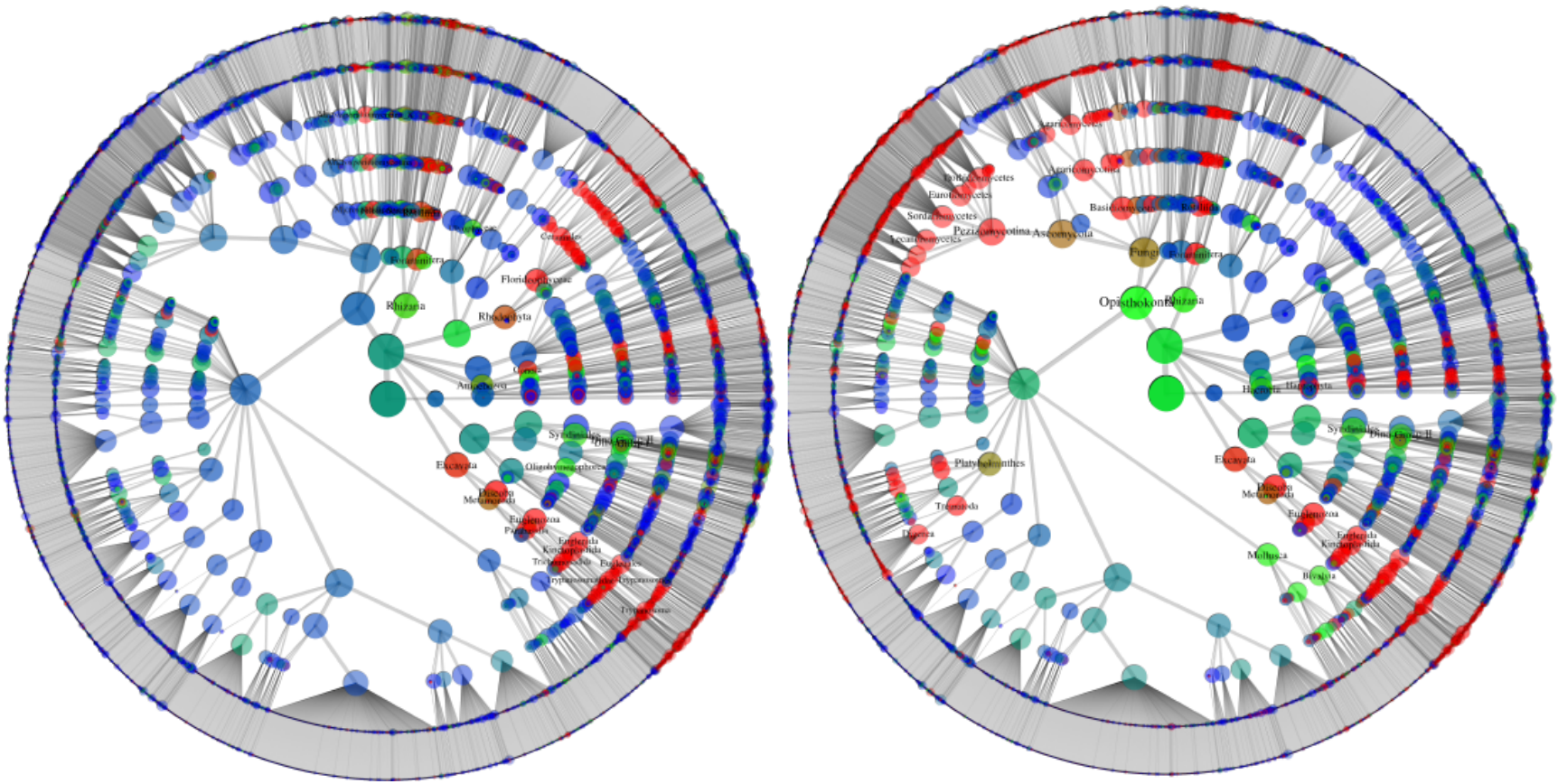


Figure 5: Results of in silico PCR of a potential primer pair against the PR2 reference database for protists

References I

[1] Samuel D. J. Brown et al. "Spider: An R package for the analysis of species identity and evolution, with particular reference to DNA barcoding". en. In: *Molecular Ecology Resources* 12.3 (2012), 562565. ISSN: 1755-0998. DOI: 10.1111/j.1755-0998.2011.03108.x. URL: http://onlinelibrary.wiley.com/doi/10.1111/j.1755-0998.2011.03108.x/abstract (visited on 11/20/2013).

[2] Lihui Xu et al. "Linking fungal communities in roots, rhizosphere, and soil to the health status of Pisum sativum". en. In: *FEMS Microbiology Ecology* 82.3 (2012), 736745. ISSN: 1574-6941. DOI: 10.1111/j.1574-6941.2012.01445.x. URL: http://onlinelibrary.wiley.com/doi/10.1111/j.1574-6941.2012.01445.x/abstract (visited on 07/18/2013).

[3] Naupaka B Zimmerman and Peter M Vitousek. "Fungal endophyte communities reflect environmental structuring across a Hawaiian landscape". eng. In: *Proceedings of the National Academy of Sciences of the United States of America* 109.32 (Aug. 2012). PMID: 22837398, pp. 13022–13027. ISSN: 1091-6490. DOI: 10.1073/pnas.1209872109.