available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/funeco

The Leinster and Cobbold indices improve inferences about microbial diversity

Stavros D. VERESOGLOU[a,b,*], *Jeff R. POWELL*[c], *John DAVISON*[d], *Ylva LEKBERG*[e], *Matthias C. RILLIG*[a,b]

[a]*Freie Universität Berlin, Institut für Biologie, Berlin, Germany*
[b]*Berlin-Brandenburg Institute of Advanced Biodiversity Research (BBIB), D-14195 Berlin, Germany*
[c]*Hawkesbury Institute for the Environment, University of Western Sydney, Australia*
[d]*Department of Botany, Institute of Ecology and Earth Sciences, University of Tartu, Tartu, Estonia*
[e]*MPG Ranch and College of Forestry and Conservation, University of Montana, Missoula, MT, USA*

ARTICLE INFO

Article history:
Received 3 July 2013
Revision received 2 February 2014
Accepted 7 March 2014
Available online
Corresponding editor:
Gareth W Griffith

Keywords:
Arbuscular mycorrhizal fungi
Bootstrapping
Hill numbers
Phylogenetic taxa similarity

ABSTRACT

An increasing number of ecological studies compare the diversity of microbial taxa along environmental gradients or between imposed treatments. Estimates are often based on analysis-of-variance of taxon-richness inferred from pyrosequencing data. We conducted a reanalysis of three 454-pyrosequencing studies on arbuscular-mycorrhizal-fungal diversity to evaluate the suitability of using the Leinster and Cobbold diversity-indices (LCdis) to assess diversity. We expected that the potential of LCdis to consider phylogenic relationships could resolve problems arising from ambiguous species-delineation in microbial-systems. Our reanalysis showed that comparisons between studies differing considerably in sequencing depth may be risky. Moreover, we show that LCdis not only reproduce the results of analyses of variance but can also resolve issues connected to variation in sequence read number, while additionally representing a less conservative metric of diversity than analysis-of-variance of taxa-richness. Based on these results we advocate the use of inclusive diversity indices in ecological studies targeting microbial communities.

Introduction

Recent decades have increasingly seen ecological theory being tested using microbial systems (Poisot et al., 2013). There are several reasons for this, the most obvious being that shifts in community composition and structure can be observed over small temporal and spatial scales (Jessup et al., 2004). The development of pyrosequencing technologies have greatly contributed to the accumulation of such studies, and will continue to do so, due to their cost-efficiency (per sequence; Rothberg and Leamon, 2008) and their ability to detect rare individuals that cannot be adequately sampled using earlier approaches (Öpik et al., 2009). Although it is common for such studies to incorporate microbial community information in analyses of α- and β- diversity after pooling data, in many studies the unit for analyses of diversity is the replicated sample.

The most common way to summarize diversity information is to apply analysis of variance (ANOVA) to estimates of species richness (e.g. Öpik et al., 2009; Lekberg et al., 2011), which is the only diversity metric robust to potential PCR-related biases in quantitative-community matrices.

* *Corresponding author*. Freie Universität Berlin, Institut für Biologie, Berlin, Germany. Tel.: +49 3083853172.
E-mail addresses: sveresoglou@zedat.fu-berlin.de, seby31@zeroone.net (S.D. Veresoglou).
1754-5048/$ – see front matter
http://dx.doi.org/10.1016/j.funeco.2014.03.006

Community richness represents one extreme in diversity assessment in which rare species are assigned the same weight as abundant species. The other extreme is the inverse of the Berger and Parker index (Berger and Parker, 1970), in which only abundant species are considered. Both of these extreme approaches for assessing community diversity may be problematic. The former may be misleading since it does not consider shifts in evenness among taxa, a characteristic of communities that has been demonstrated to predict functional resilience (Wittebolle et al., 2009). The latter ignores rare or elusive taxa that may become important under certain environmental conditions, such as in the lead-up to toxic algal blooms (Burkholder et al., 1992; Hooper et al., 2005). Therefore, a robust representation of species abundance is necessary in order to accurately represent diversity in ecological studies of microbes.

Leinster and Cobbold (2012) proposed the use of diversity profiles (a series of multiple diversity estimates that differ in the relative weighting of abundant vs. rare taxa), as opposed to point estimates of diversity (e.g. richness and Shannon diversity index), to visually represent the importance of rare taxa in making comparisons among sample groups. They also used an extension of Hill numbers (Hill, 1973) to account for gradients in similarity among species; diversity profiles may then compare naive diversity profiles to profiles that explicitly account for genetic or functional variation among species within communities. A major novelty of these indices is their ability to consider phylogenetic relationships between taxa; according to the non-naive version of the indices a site with a given number of species that belong to a single genus is typically less diverse than another site with an equal number of species that belong to different genera. This attribute may be of particularly high applied value in microbial ecology, where considerable uncertainty surrounds approaches to species delineation (e.g. Stockinger et al., 2010). Here we use the term "inclusive diversity indices" to describe metrics that address diversity on the basis of diversity profiles, i.e., considering the entire range of weighting applied to rare (richness) and abundant (evenness) species. The result is a generalized comparison of diversity between communities that requires no *a priori* assumptions about the importance of species abundance.

Here, we use the approach of Leinster and Cobbold (2012) to reanalyze data from published studies that have applied next-generation sequencing. We had three objectives. First, we highlight the importance of standardizing sequencing depth in pyrosequencing studies before conducting assessments of diversity. While the sample size dependency of diversity metrics such as species richness has been known for a long time (e.g. Smith and van Belle, 1984) this is a component of bioinformatics that has not been sufficiently stressed even within the latest pyrosequencing user guides (e.g. Lindahl et al., 2013). Then we assess whether the results of recent next-generation sequencing studies are comparable to earlier studies that were based on lower numbers of sequences and which consequently detected a much smaller fraction of each community. This objective is important as it will allow researchers to determine whether the inferences made from Sanger sequencing studies are comparable to those made from pyrosequencing studies. Finally, we assess whether the conclusions that had been reached through implementation of traditional comparison techniques such as ANOVAs would be robust to consideration of alternative diversity metrics. Specifically we consider the diversity profiles that are obtained from inclusive diversity indices when rare species are not weighted equally to abundance species, i.e., we ask whether attributing variable importance to abundant species might change our view of diversity responses in some cases.

Materials and methods

We focus here on ecological studies of arbuscular mycorrhizal fungi (AMF), ubiquitous obligate symbionts associating with the roots of most terrestrial plant species (Smith and Read, 2008). We chose AMF as a model for our purposes because this system provides a clear sampling criterion delineating a community of interacting individuals (within the root system of a single host plant). This criterion is difficult to satisfy for many ecological studies of microbes, where issues of scale always exist and are dependent on the quantity and properties of the substratum sampled (e.g. Kang and Mills, 2006), which can lead to confused definitions of α- and β- diversity (Whittaker et al., 2001). In addition, a recurrent problem in ecological studies of microbes is the recognition of 'species' and 'individuals' and the delineation of clusters of DNA sequences into these groups (Smith et al., 1992; Powell, 2012). This problem has been shown to have consequences for our understanding of the ecological characteristics of AMF (Krüger et al., 2011; Powell et al., 2011) and other microorganisms (Koeppel et al., 2008). Finally, AMF represent a group of ecologically important microbes that has been studied extensively in the past (Öpik et al., 2010). We did not consider multiple microbial groups as we wanted our datasets to be sufficiently homogenous to be comparable. However, we acknowledge that this limits our ability to generalize our results to other microbial groups.

We reanalyzed two studies that utilized next-generation sequencing to assay AMF communities from plant roots (Lekberg et al., 2011; Becklin et al., 2012). To expand our dataset we further used a study that assayed AMF communities from soil (Davison et al., 2012); at the time when the project was initiated these were the only published next-generation sequencing studies on AMF that both reported abundances and included replicated designs. In all three studies phylogenetic information had been partitioned into operational taxonomic units – OTUs; in our reanalysis we adopted these OTU definitions as a means of delineating species – a surrogate of a species in plants and animals.

Lekberg et al. (2011) used eleven replicates per treatment of plants (*Plantago lanceolata*): (i) subjected to limited disturbance; (ii) disturbed with recolonization from the surrounding AMF community prevented; (iii) disturbed with recolonization from the surrounding AMF community possible; and (iv) mycorrhizal plants adjacent to these units; the native *P. lanceolata* treatment was dropped to generate a balanced dataset (Table 1).

Becklin et al. (2012) included 4–8 replicates per treatment (*Taraxacum ceratophorum*, *T. officinale* and *P. viscosum* harvested from open meadow plants and willow understory habitats) (Table 1).