Toward a function-first framework to make soil microbial ecology predictive

Hicks, Lettice C.; Frey, Beat; Kjøller, Rasmus; Lukac, Martin; Moora, Mari; Weedon, James T.; Rousk, Johannes

Published in:
Ecology

DOI:
10.1002/ecy.3594

Publication date:
2022

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY

Citation for published version (APA):
Toward a function-first framework to make soil microbial ecology predictive

LETTICE C. HICKS,¹,² BEAT FREY,² RASMUS KJOLLER,³ MARTIN LUKAC,⁴,⁵ MARI MOORA,⁶ JAMES T. WEEDEON,⁷ AND JOHANNES ROUSK ¹

¹Section of Microbial Ecology, Department of Biology, Lund University, Ecology Building, Lund 22362 Sweden
²Forest Soils and Biogeochemistry, Swiss Federal Research Institute WSL, Birmensdorf 8903 Switzerland
³Department of Biology, Terrestrial Ecology Section, University of Copenhagen, Universitetsparken 15, Copenhagen 2100 Denmark
⁴School of Agriculture, Policy and Development, University of Reading, Reading RG6 6AR United Kingdom
⁵Faculty of Forestry and Wood Sciences, Czech University of Life Sciences Prague, Prague 16500 Czech Republic
⁶Department of Botany, Institute of Ecology and Earth Sciences, University of Tartu, Lai 40, Tartu 51005 Estonia
⁷Systems Ecology, Department of Ecological Science, Vrije Universiteit Amsterdam, Amsterdam 1081 HV The Netherlands


Abstract. Soil microbial communities perform vital ecosystem functions, such as the decomposition of organic matter to provide plant nutrition. However, despite the functional importance of soil microorganisms, attribution of ecosystem function to particular constituents of the microbial community has been impeded by a lack of information linking microbial function to community composition and structure. Here, we propose a function-first framework to predict how microbial communities influence ecosystem functions. We first view the microbial community associated with a specific function as a whole and describe the dependence of microbial functions on environmental factors (e.g., the intrinsic temperature dependence of bacterial growth rates). This step defines the aggregate functional response curve of the community. Second, the contribution of the whole community to ecosystem function can be predicted, by combining the functional response curve with current environmental conditions. Functional response curves can then be linked with taxonomic data in order to identify sets of “biomarker” taxa that signal how microbial communities regulate ecosystem functions. Ultimately, such indicator taxa may be used as a diagnostic tool, enabling predictions of ecosystem function from community composition. In this paper, we provide three examples to illustrate the proposed framework, whereby the dependence of bacterial growth on environmental factors, including temperature, pH, and salinity, is defined as the functional response curve used to interlink soil bacterial community structure and function. Applying this framework will make it possible to predict ecosystem functions directly from microbial community composition.

Key words: biogeochemistry; community ecology; predictive ecology; soil carbon; soil microorganisms; structure and function.

INTRODUCTION

Understanding of how organisms regulate ecosystem functions is relatively mature in the study of macrobiota (Brown et al. 2004). For example, variation in plant litter quality is tree species specific, clearly connecting tree community composition to global patterns of litter decomposition (Verheijen et al. 2015). This is in sharp contrast to the rudimentary state of knowledge relating soil microbiota to ecosystem function (Prosser 2013, Bier et al. 2015). Recent progress in microbial ecology has been fueled by the rapid development of molecular approaches. Indeed, microbial DNA, RNA, and protein profiles can be correlated to ecosystem processes, a development often viewed as a path toward a mechanistic understanding of ecosystem function (Raes and Bork 2008, Bier et al. 2015). However, while these powerful “omic” approaches are moving closer to elucidating the in situ functioning of microbial communities, they have
three major shortcomings for inferring ecosystem function. First, observed abundances of microbial genes, transcripts, proteins, and enzymes can only provide the potential for function and translating functional potentials into realized in situ microbial process rates remains a considerable challenge (Rocca et al. 2015). Second, omics-based predictions of ecosystem functions on the basis of microbial community composition rely on knowledge of individual microbial taxa and their function. This taxonomic information is often gained from isolates grown in pure culture. Problematically, the vast majority of the community members are difficult or impossible to cultivate and progress is slow (Raes and Bork 2008, Bier et al. 2015). Third, microbial taxa do not function in isolation. Rather, they operate in multitrophic consortia that form complex food webs. Crucially, most of these within- and between-consortia interactions still need to be described. Thus, even if we had a complete parts list of phylogenetic and functional genes from a given environment, combined with full functional information from pure culture isolates, it is doubtful whether this would provide sufficient information to predict overall community-level ecosystem functions. Thus, a truly predictive omics-based framework is likely to remain elusive into the foreseeable future, unless new strategies are implemented.

In order to overcome the shortcoming associated with the established omics-based approach, where information on individual pieces is put together to construct the function puzzle (e.g., Fierer et al. 2014, Martiny et al. 2015, Treseder and Lennon 2015, Douglas et al. 2018), we propose a function-first framework by first determining aggregate community-level functional response curves and then relating these to taxonomic information (Fig. 1). The proposed approach therefore establishes a foundation for function in the structure–function relationship that is based on measured in situ process rates of the microbial community considered in aggregate. This approach affords several advantages: process rates measured on the whole microbial community are at the scale most relevant to biogeochemical modeling; the rates are determined on communities in their natural environmental contexts, meaning all relevant interactions among different taxa are implicitly included, even if they are not all described; and traits of the community can be directly compared to measured rates. Thus, we use techniques previously applied in functional ecology where, e.g., plant traits have been used to predict responses to changing resource availability (Lavorel and Garnier 2002). Our proposed framework also makes use of advances in theoretical ecology (Grime 1977, Southwood 1977, McGill et al. 2006, Suding et al. 2008, Webb et al. 2010), which are beginning to be integrated into microbial ecology. For example, Allison (2012) developed a trait-based model to predict litter decomposition rates based on the enzymatic trait distribution of a microbial community.

DEFINING AGGREGATE COMMUNITY-LEVEL FUNCTIONAL RESPONSE CURVES AND ESTIMATING ECOSYSTEM FUNCTIONS

A microbial community is composed of a diverse assembly of different taxa, each characterized by its own set of traits (Violle et al. 2007, Suding et al. 2008, Webb et al. 2010). Traits are defined as the phenotypic characteristics of an organism, which determine how an organism interacts with other organisms and the environment. For example, each bacterial taxon has a specific pH range, temperature dependence, tolerance to salinity, etc. The distribution of all individual response traits within a community therefore holds information that can be used to determine its dependence on environmental factors (Norberg et al. 2001). While phenotypic plasticity may destabilize the genotype–phenotype relationship, resulting in acclimation of multicellular eukaryotic organisms, such as plants, to altered environmental conditions (Hochachka and Somero 2002), phenotypic plasticity within bacterial strains is much smaller than phenotypic variation between strains (Beier et al. 2015). Thus, the effect of phenotypic plasticity will be practically negligible for the populous and diverse bacterial communities in environmental samples (Barraclough 2015), where the signal from shifts in microbial community composition will dominate over the phenotypic plasticity of individual taxa.

Some traits, such as the life-strategy or growth form of an organism, require detailed assessments of individuals or at least populations (Violle et al. 2007), making the use of such traits very challenging. However, there are many traits for which an aggregate characterization can be achieved at a higher organizational level (Norberg et al. 2001, Suding et al. 2008). Examples include the dependence on environmental factors of gross primary production (proxied as the uptake of inorganic C into biomass), microbial soil C formation (proxied as microbial growth), or microbial decomposition of organic matter (proxied as microbial respiration or litter mass loss); these relationships can be captured at the whole-community level by measurements of CO2 fixation, microbial growth and respiration, or mass loss, respectively (sensu Norberg et al. 2001, McGill et al. 2006, Suding et al. 2008, Wallenstein and Hall 2012). Crucially, in these cases, the link between the community aggregated distribution of “traits” and ecosystem function is clear (Suding et al. 2008). As such, just as the response of a plant population to environmental change has been described by defining its response traits (e.g., Keddy 1992), the functional response of whole microbial communities to environmental change can be predicted from its aggregate functional response curve (Norberg et al. 2001, Suding et al. 2008), as long as the controlling factors for the response curve are known.

Whole-community-level functional response curves can be determined by using environmental samples containing natural communities. Important functional response curves include relationships with pH, salinity,
By building on recent methodological advances in microbial ecology, it is now possible to establish estimates of the dependences of microbial functions on environmental parameters through quantifying dose–response relationships of microbial growth and respiration (e.g., Rath et al. 2016). These dose–response relationships describe relative changes in rates \((y)\) linked to an independent factor \((x)\), analogous to rate modifiers describing the effect of, e.g., temperature on microbial rates used in current biogeochemical models (Sierra et al. 2015). These relationships describe the community-level functional response curve (e.g., the dependence of growth rate on temperature; Fig. 2), and characterize how communities are shaped by their environment and respond to environmental change (Figs. 2, 3), as described by Wallenstein and Hall (2012). Once the community’s performance has been defined as a function of an environmental factor (e.g., the response curve for temperature has been established; Fig. 2a), this can be combined with environmental data to estimate the ecosystem function (e.g., soil organic matter decomposition determined by microbial respiration or soil carbon sequestration determined by microbial growth; Fig. 2b). Validation of community aggregated functional response curves determined in short-term laboratory assays by comparing with in situ measurements will be important, as will investigations into how covarying environmental factors interact. For some factor combinations, interactions are expected (e.g., moisture and salinity), while for others they are not. For example, a recent study designed to test for interactions between factors found that the dependence of microbial growth and respiration on temperature and moisture were independent (Cruz-Paredes et al. 2021).

**Identification of Indicator Taxa**

In a final step to predict ecosystem function from community composition, information from community-level functional response curves can be linked to taxonomic information from the same environmental samples. Thus, we aim to not only identify keystone or indicator taxa (e.g., Banerjee et al. 2008), but to also link these taxa with differences in ecosystem function. DNA-sequencing based assessments of taxonomic composition of communities may be most appropriate for this purpose, as the turnover of microbial communities and DNA are closely tracked (Oliverio et al. 2017, Papp et al. 2018). Although horizontal gene transfer may weaken the composition–function relationship, a change in the DNA profile will capture changes in the microbial community, providing an opportunity to link microbial community composition to the community aggregated functional response curve (e.g., Rath et al. 2019). This link between microbial community structure and function may be achieved by comparative analysis of information from environmental gradients (e.g., temperature, soil moisture, pH, salinity) combined with observations from manipulative experiments. These comparative

---

**Fig. 1.** The new conceptual framework visualized. The community aggregated functional response curve describes the dependence of microbial functions on environmental conditions. (1a) Environmental change will influence microbial community composition and thus the aggregate functional response curve. (1b) In parallel, the environmental conditions also constrain the community responses curve (i.e., defines the x-axis value on the environment–response curve). The output of this combination of the environment and the current microbial community is the ecosystem function. (2) By linking microbial community composition to differences in community aggregated functional response curves, we can identify "biomarker" taxa that characterize the community’s ability to provide functions. (3) Indicator sets of biomarker taxa combined with environmental data can then be used to predict ecosystem functions directly from microbial community composition.
studies may identify the “common denominator” taxa (sets of sequences) that capture community-level functional response curves. In a first correlative step, putative “biomarkers” or “indicator taxa” can be ranked according to their degree of correlation with the community aggregated functional response curves, using a suitable index such as the temperature minimum or pH optimum extracted from each functional response curve. After reducing overall microbial taxa data to a smaller set of classifiers (e.g., logistic regression or other statistical approaches such as ordination to reduce many dimensions down to a manageable few), combinations of biomarkers should have sufficient power to predict aggregate community functional response curves that, combined with environmental data, can be used to estimate ecosystem functions in independent data sets (e.g., The Earth Microbiome Project; Thompson et al. 2017). These biomarkers will not causally link particular members of the microbial community with resulting microbial functions, but rather serve as a diagnostic tool to enable predictions of ecosystem functions from community composition information, analogous to the use of biomarkers in clinical diagnosis (e.g., Segata et al. 2011). Identified biomarkers may, however, be targeted as candidate taxa driving functions, providing testable hypotheses for causal links between specific microbial community members and functions. For example, the identified biomarker taxa may be characterized in culture, with the expectation that, e.g., their temperature traits should align with specific environmental conditions. It will also be possible to investigate whether combinations of biomarker taxa are associated with the same or different functional responses, whereby functional redundancy would result in a weaker relationship between the microbial composition and community-level functions. For some functions that involve a specific physiological pathway or are carried out by a phylogenetically constrained group, such as nitrification, we expect there to be stronger links between community composition and function, while for more physiologically “broad” functions, the link may be weaker (Schimel and Schaeffer 2012).

**Literature Cases: Temperature, Soil pH, and Salinity**

Here, we illustrate our proposed approach by applying it to previously published examples. By determining community aggregated functional response curves and estimating microbial functions for key environmental factors such as temperature, soil pH, and salinity, ecosystem functions may be better predicted by taxonomic information about the microbial community.

**Temperature**

Microbial communities exhibit functional responses to temperature aligned with their thermal environment (Li et al. 2021). For example, a recent meta-analysis revealed a positive relationship between the native thermal environment of microorganisms and the temperature optimum for microbial activity (Alster et al. 2018), demonstrating that microbial responses to warming may be predicted by determining community aggregated functional response curves in environmental samples. Rousk et al. (2012) assessed the effects of 5°C experimental warming on soil bacteria in a temperate hardwood forest, hypothesizing that warming would induce a change in the performance of soil bacterial communities, resulting in a change in the microbial dependence on temperature. Relationships between temperature and the growth of bacterial communities were measured and represent an example of the community functional response curve. Upon warming, the response curve of bacterial communities shifted to become more warm adapted, illustrating that the microbial dependence on temperature had adapted to the new thermal regime. It has been shown that temperature change below the thermal optimum does not induce a shift in functional response curves under stable laboratory conditions for months (Fig. 3; Bárcenas-Moreno et al. 2009) or in the field within seasons (Birgander et al. 2018). Hence, in this case, the shift in functional response curve was likely induced when the 5°C warming treatment in the field increased soil temperature above the original optimum (i.e., at peak temperatures in the summer). In a second step, with the knowledge that these response curves would reflect the dependence of bacterial growth on temperature at all temperatures below the optimum, it was possible to determine the functional relevance of these community shifts (field warming treatment vs. control). To do so, the measured functional response curves were combined with soil temperatures measured in situ, to compute the estimated bacterial carbon production as an ecosystem function. This enabled a comparison of the functional consequences of warming, where community changes (shift in community-level functional response curve) could be differentiated from the direct effect of temperature (higher rates for all processes at warmer temperatures). Different microbial taxa are known to vary markedly in their temperature sensitivity (Wang et al. 2021). Thus, although shifts in community-level functional response curves could occur due to phenotypic acclimation or genotypic adaptation of taxa within the microbial community, the shift was most likely caused by a change in microbial community composition ("species sorting"), where taxa already better adapted to the warmer temperature regime outcompeted other less-well-adapted taxa (Allison and Martiny 2008, Donhauser et al. 2020). For example, several taxa from the families Koribacteraceae, Acidobacteriaceae, and Solibacteres were found to be responsive to changing temperature across a broad range of soil types (Oliverio et al. 2017), illustrating the potential to link these responsive indicator taxa to associated shifts in microbial functions under environmental change.
FIG. 2. (a) The community aggregated functional response curve of a bacterial community. In this example, it is the bacterial growth dependence on temperature. Bacterial growth is measured instantaneously at a range of different temperatures, then the rates are plotted against those same temperatures to define the fundamental “temperature niche” of the studied community, characterizing the community aggregated functional response curve. The determined relationship provides precise predictive information about how temperature influences the bacterial growth rates of the studied community. (b) An estimate of microbial function from the community-level functional response curve combined with environmental data (e.g., dashed arrow shows the bacterial growth rate predicted by the environmental temperature). In this example, the dependence of bacterial growth on temperature (i.e., the functional response curve, as defined in panel (a)) is combined with in situ temperatures to estimate a microbial function (bacterial growth, i.e., the bacterial production of carbon in soil). Panel (a) is adapted from Cruz-Paredes et al. (2021) and some results from panel (b) are adapted from Rousk et al. (2012).
Bárcenas-Moreno et al. (2016) tested the functional implications of an alignment between microbial community characteristics and its environment. This was achieved by sterilizing soil samples from along a pH gradient (pH 4–8), and subsequent reciprocal inoculation in a full-factorial design. The dependence of bacterial growth on pH (community-level functional response curve) was quantified through response functions (respiration and bacterial growth). Soil samples inoculated with a community sourced from soils of a different pH developed bacterial pH response curves “misaligned” with the new environment. The resulting microbial functions, bacterial growth rates and respiration, were still determined by the pH of the soil sample, but the misalignment of the bacterial community with the new environment led to an associated impairment of function. In this case, the misalignment resulted in a 25% reduction in the decomposer function (i.e., respiration) compared to the functioning of the bacterial community, which was aligned with its environment. This example further demonstrates that microbial response curves can be resistant to change, likely constrained by the rate of community turnover (Wallenstein and Hall 2012). Just as for temperature, it will, however, be possible to...
identify the environmental conditions and time required for a shift in functional response curves occur and to link these functional responses to associated differences in taxa.

Salinity

Rath et al. (2019) used a community-level approach to assess how salinity altered microbial tolerance to salt, by measuring how the growth of bacterial communities from two salinity gradients responded to changes in salinity. In line with expectations, the authors found that more saline soils exhibited a higher community-level tolerance to salt. Moreover, the authors were able to identify consistent shifts in bacterial community composition accompanying the change in salt tolerance. In this case, a high community tolerance to salt was positively correlated with taxa belonging to the family Flavobacteriaceae (Bacteroidetes), including Gillisia, Gramella, and Saliminicrobium, and Gammaproteobacteria, including Salinisphaera and Alkanibacter. These identified taxa could therefore be targeted as indicators for salt-affected communities, demonstrating the potential to use community-level data to predict functional responses (also see Oliverio et al. [2017] testing a similar approach for temperature). Changes in indicator taxa will also capture changes in community-level functional response curves, circumventing the need to identify the conditions under which a shift in response curves occurs (as noted in the previous examples), and enabling predictions of microbial function from taxonomic information coupled with environmental conditions (Evans and Wallenstein 2013). To facilitate progress, identified biomarkers could be cross-checked against databases of sequence information and environmental data that are already available, for example as part of The Earth Microbiome Project (Thompson et al. 2017), or acquired using Seqenv (Sinclair et al. 2016).

SYNTHESIS

The proposed function-first framework can be used to predict how ecosystems will respond to environmental change and the resulting biogeochemical consequences (Fig. 1). Here we used the dependence of bacterial growth on environmental factors as an example of a community-level functional response. We showed that, by defining community aggregated functional response curves in natural environments, environment–response relationships can be used to predict microbial ecosystem functioning (Fig. 2). Moreover, our examples highlighted that community-level functional responses can also contain information about (1) how a community will respond to environmental change (Suding et al. 2008, Allison 2012) and (2) how that change will affect ecosystem functioning. Conversely, we were also able to determine the functional consequences of a “misalignment” of the community’s response curve with its environment (Fig. 3). Together, these examples demonstrate the scope for this framework to also be applied to estimate other functions (e.g., nitrogen transformations). Additionally, although soil bacteria are unique in their high diversity and abundance, we foresee that this framework could be extended to also include other soil microorganisms and other systems.

Systematically linking microbial community composition to community-level response curves holds the potential to identify sets of biomarkers that can be used to infer ecosystem function (e.g., Rath et al. 2019). Thus, microbial composition would become informative in the prediction of ecosystem functions. Currently, there are databases that characterize functional traits for strains in isolation (e.g., functional traits for fungi in FUNGuild; Nguyen et al. 2016), but there are no databases that collate functional response curve data for microbial communities in natural environments. While the chance of finding spurious correlations between microbial community composition and function will be high in small data sets, spurious correlations will “fall-out” when more data sets are combined, underlining the need to compile information in databases. To facilitate progress, this information could be collected in a public database analogous to the TRY-database for plants (Kattge et al. 2020). A variety of observations will accumulate detailed information about how environmental factors (pH, temperature, soil moisture, salinity, etc.) modulate community aggregated functional response curves and community composition. The framework proposed here could subsequently draw on the database information to identify biomarkers that are able to provide prediction about a community’s contribution to ecosystem functions, and its susceptibility to environmental change. Ultimately, these biomarkers can be used as a diagnostic tool to predict ecosystem functions from taxonomic surveys available from, e.g., the global catalogue of microbial taxonomic data generated by The Earth Microbiome Project (Thompson et al. 2017). This function-first framework may thus provide a complementary approach to the assembly of individual traits defined for strains in pure culture or inferred from genomic information (e.g., PICRUSt [Douglas et al. 2018], Tax4Fun [Asshauer et al. 2015]) to derive function from structure, reinvigorating the capabilities of microbial sequence data to predict ecosystem functions.

ACKNOWLEDGMENTS

We thank Bruce Hungate for incisive comments that helped us clarify our perspective. We would also like to acknowledge networking support by WG2 of the COST Action FP1305 “BioLink.” This work was supported by grants from The Swedish Research Council Vetenskapsrådet (2020-03858 and 2020-04083), The Swedish Research Council Formas (2018-01315), The European Regional Developmental Fund (Centre of Excellence Ecolchange), The Estonian Research Council (PRG 1065), and The Knut and Alice Wallenberg Foundation (KAW 2017.0171). The authors declare that they have no conflict of interest.


