



# Seeing the forest for the genes: using metagenomics to infer the aggregated traits of microbial communities

Noah Fierer<sup>1,2\*</sup>, Albert Barberán<sup>1</sup> and Daniel C. Laughlin<sup>3</sup>

<sup>1</sup> Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, USA

<sup>2</sup> Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO, USA

<sup>3</sup> Environmental Research Institute, School of Science, University of Waikato, Hamilton, New Zealand

## Edited by:

Peter Van Bodegom, VU University  
Amsterdam, Netherlands

## Reviewed by:

Angel Valverde, University of Pretoria,  
South Africa

Thulani Peter Makhalanyane,  
University of Pretoria, South Africa

## \*Correspondence:

Noah Fierer, Department of Ecology  
and Evolutionary Biology, University  
of Colorado, UCB 216, Boulder,  
CO 80309-0216, USA  
e-mail: noah.fierer@colorado.edu

Most environments harbor large numbers of microbial taxa with ecologies that remain poorly described and characterizing the functional capabilities of whole communities remains a key challenge in microbial ecology. Shotgun metagenomic analyses are increasingly recognized as a powerful tool to understand community-level attributes. However, much of this data is under-utilized due, in part, to a lack of conceptual strategies for linking the metagenomic data to the most relevant community-level characteristics. Microbial ecologists could benefit by borrowing the concept of community-aggregated traits (CATs) from plant ecologists to glean more insight from the ever-increasing amount of metagenomic data being generated. CATs can be used to quantify the mean and variance of functional traits found in a given community. A CAT-based strategy will often yield far more useful information for predicting the functional attributes of diverse microbial communities and changes in those attributes than the more commonly used analytical strategies. A more careful consideration of what CATs to measure and how they can be quantified from metagenomic data, will help build a more integrated understanding of complex microbial communities.

**Keywords: metagenomics, traits, community-aggregated traits, microbial diversity, microbial ecology**

## WHY MICROBIAL ECOLOGISTS ARE STILL LOST IN 'TERRA INCOGNITA'

Ever since van Leeuwenhoek first peered through a microscope, it has been recognized that most environments harbor diverse and complex microbial communities. Whether we are studying soil, the human gut, marine sediments, or lake waters, we find many taxa that have unknown, or at least poorly described, ecological characteristics. Our descriptions of microbial communities are littered with question marks just as maps of continents were once littered with labels of 'terra incognita' by European explorers.

Recent methodological advances, most prominently advances in DNA sequencing, have provided unique insight into the structure and function of complex microbial communities, thereby improving our ability to chart 'terra incognita.' This is particularly true for those environments, like soils and sediments, which harbor many taxa that are resistant to laboratory isolation (Whitman et al., 1998; Fierer and Lennon, 2011). It is now routine for researchers to use shotgun metagenomics, randomly sequencing from a pool of whole-community DNA extracted from environmental samples (Handelsman, 2004), to characterize complex microbial communities and their functional capabilities. Shotgun metagenomics is by no means the only approach available – but, fueled by rapid declines in sequencing costs, shotgun metagenomes are an increasingly available source of data that can be mined to characterize microbial communities.

One of the explicit goals (and promises) of many shotgun metagenomic projects is to understand the broader functional and ecological characteristics of microbial communities, insight that

cannot necessarily be gleaned from the sequencing of taxonomic or phylogenetic marker genes. These arguments are convincing because there is already evidence that such approaches can be used to better understand the function of complex microbial communities, whether those functional capabilities are related to phosphate removal from wastewater (Martin et al., 2006), carbon cycling in permafrost soils (Mackelprang et al., 2011), or arsenic geochemistry in marine sediments (Plewniak et al., 2013). While these and many other studies highlight the utility of shotgun metagenomic analyses, it is often difficult to distinguish tangible advances from the hype.

We can use metagenomic data to develop testable predictions regarding the ecological attributes of microbial communities, but the approach is no panacea and inferring the functional capabilities of communities from metagenomic data remains difficult. One reason for this is that community-level attributes are the emergent properties of a diverse array of organisms interacting directly and indirectly in a myriad of ways under environmental conditions that are rarely static. Even if we could predict the ecological attributes of all individual taxa living in a given community (a Sisyphean task in most microbial habitats), the overall functional capabilities of that community and how it responds to changes in biotic or abiotic conditions, will remain difficult to predict. If we want to know how rapidly soil communities will decompose soil organic matter or how efficiently a gut microbial community will ferment ingested polysaccharides, it is insufficient to document the genes associated with the metabolism of various organic carbon pools and their relative abundances. Even in simple communities composed of well-described microbial

taxa, predicting community-level metabolic properties from genomic or transcriptomic data can be challenging (Sieuwert et al., 2010).

If the goal is to link microbial communities to processes, a critical step is to understand the ecological attributes of whole communities, not just the attributes of individual community members as the whole is rarely the sum of its parts. Here we argue that we could often do a better job of predicting the functional capabilities of whole communities by using the concept of community-aggregated traits (CATs) to glean more useful information from the terabases of shotgun metagenomic data being generated.

## WHAT ARE COMMUNITY-AGGREGATED TRAITS AND WHY ARE THEY USEFUL?

Functional traits can be measured at various levels of organization – from the level of individual cells, to species, to whole communities. For example, one could determine the size of individual bacterial cells in a water sample, the mean cell sizes for different bacterial species found in that sample, or the mean size of cells found in the whole sample. Community-level traits can be quantified either as a “community-weighted mean trait,” where the mean trait values for all taxa in a community are weighted by their relative abundances, or as a “CAT,” where the traits are measured from a random sample of individuals irrespective of their taxonomic identities.

We can calculate community-weighted mean traits from shotgun metagenomic data by reconstructing genomes, or parts of genomes, and using this genomic information to predict the characteristics of individual community members. This ‘bottom up’ approach has already been demonstrated to be useful for describing the putative functions of undescribed microbial taxa (e.g., Walsh et al., 2009; Hug et al., 2013) or for documenting taxa likely responsible for specific microbial processes (e.g., Howard et al., 2008). Likewise, we can calculate community-weighted mean traits from *a priori* information on the traits of specific microbial taxa (or lineages). While this approach is certainly appealing (e.g., Langille et al., 2013), it has some clear disadvantages if the traits of many microbial taxa, even closely related taxa, are not already known, as would be the case in many habitat types. Likewise, for those traits that are capable of being transferred across distantly related taxa via horizontal gene transfer (e.g., antibiotic resistance; Forsberg et al., 2012), trying to determine community-levels trait solely from taxonomic or phylogenetic information would be problematic.

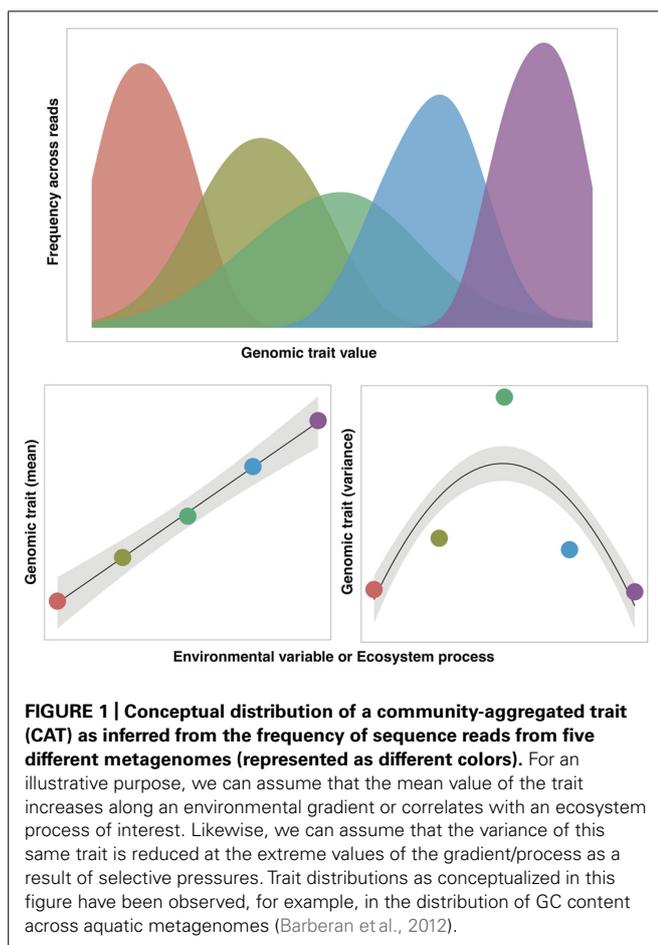
An alternative approach is to calculate CATs directly from a community of interest without collecting any information on the identities of the taxa found in a given plot. For example, leaf traits in a plant community can be determined by remote sensing of canopy spectra (Homolová et al., 2013) or through random taxon-free sampling (Gaucherand and Lavorel, 2007) without having to measure trait averages for each plant species. In a similar manner, we can calculate CATs from shotgun metagenomic data as long as we assume that our metagenomes represent a random sampling of all microbial genomes present in that sample and that the traits of interest can actually be inferred from the genomic information. Such a ‘top-down’ approach has been widely used in

microbial environments ranging from soil (Fierer et al., 2012) to marine waters (Ganesh et al., 2014) to the human gut (Greenblum et al., 2012). In all of these cases, much of the insight into community attributes came not from assigning genes to taxa, but rather by determining the relative abundances or presence/absence of genes and gene categories at the community-level of inquiry, not at the level of individual taxa.

CATs are often useful predictors of community-level properties because, according to the mass ratio hypothesis, species controls on community-level processes are in proportion to their relative abundances (Grime, 1998). The functional traits of abundant taxa will have more important influences on the functional properties of a community than the traits of subordinate taxa. There is strong empirical support for the mass ratio hypothesis from those studies that have explicitly tested the hypothesis in plant communities (reviewed in Lavorel and Grigulis, 2012). For example, Mokany et al. (2008) demonstrated that mean trait values were a good predictor of multiple ecosystem properties in grasslands, including litter decomposition rates and aboveground net primary productivity. Likewise, other studies have demonstrated the utility of using plant CATs to predict a wide range of processes and ecosystem properties from nitrification rates (Laughlin, 2011) to soil carbon dynamics (Garnier et al., 2004), and green biomass production (Lavorel et al., 2011).

In many cases, the mass ratio hypothesis should also apply to microbial communities. For example, we might expect that the efficiency by which microbes mineralize nitrogen in a composting bio-reactor would be a function of the community mean for those traits associated with nitrogen mineralization rates. The appeal of the mass ratio hypothesis to ecologists working with highly diverse microbial communities is that we do not necessarily need to know the traits of all taxa to predict a process of interest, rather we just need to know the traits of the more dominant taxa (Grime, 1998). Of course, not all community-level processes will be predictable from the mass ratio hypothesis. For example, relatively rare ‘keystone’ taxa can have a disproportionate influence on certain processes (Paine, 1995) and biotic interactions could invalidate the mass ratio hypothesis (Vile et al., 2006). Moreover, the diversity of traits (the range or variance in trait values found in a given community) could often be more relevant (Violle et al., 2012) than just the mean CAT value (**Figure 1**). For example, under fluctuating environmental conditions, like those found in wet soils exposed to wide swings in O<sub>2</sub> levels, the distribution of traits associated with O<sub>2</sub> tolerance could be more informative for predicting community responses over time than simply measuring a mean community-level O<sub>2</sub> preference (Pett-Ridge and Firestone, 2005). In theory, we should be able to use shotgun metagenomic data to calculate both means and variances of traits (**Figure 1**), based on the distribution of the relevant genes or gene categories indicative of the trait of interest within the community.

CATs have already proven useful for predicting some microbial processes. For example, soil microbial ecologists have long considered bacterial:fungal ratios to be a useful metric for understanding soil carbon dynamics because bacteria and fungi are considered to have distinct carbon use efficiencies (Bailey et al., 2002), an assumption that has been called into question



(Strickland and Rousk, 2010). Likewise, the elemental composition of planktonic cells can often be used to predict biogeochemical dynamics in aquatic systems (Finkel et al., 2009) just as C:N ratios of heterotrophic microbial communities appear to influence nitrate accumulation rates in a wide range of systems (Taylor and Townsend, 2010). Some models have incorporated community-level information on microbial traits or functional guilds to predict litter decomposition rates (Allison, 2012), soil carbon cycling (Wieder et al., 2013), and patterns in marine phytoplankton biomass (Dutkiewicz et al., 2009). Although these models did not determine CATs empirically, they do demonstrate that models which incorporate shifts in CATs (or equivalent) are useful for predicting how microbial processes, even complex microbial processes likely driven by hundreds of microbial taxa, can effectively be predicted across space or time.

### WHICH MICROBIAL CATs TO MEASURE?

If we assume that CATs provide a useful means to describe the functional attributes of communities, we then have to identify the traits that are relevant and worth quantifying. There is a long list of microbial traits that could be useful (see examples in Table 1) and the list of traits that will be most relevant will depend on the study system and the research question. Collapsing phenotypic diversity into groups with shared traits has been

one approach to simplify this complexity. For example, microbial taxa have been divided into groups based on shared life history characteristics (Fierer et al., 2007; Ho et al., 2013), groups defined by the source of carbon or energy (lithotrophs, heterotrophs, autotrophs), or divided into groups based on specific functional capabilities ( $N_2$ -fixers, photosynthesizers, methylotrophs). While pragmatic, these functional categories over-simplify trait-level variability and it is likely more appropriate to define microbial trait space as a set of continuous, quantifiable variables, where microbes, or microbial communities, sit within this multidimensional space.

There are 100s of traits that could possibly be measured for any organism – but it is often not necessary to measure every one of these traits to place species into multidimensional trait space (Laughlin, 2014). The key is to identify and measure those traits that are most relevant to the system in question and select those traits that are independent and that most effectively discriminate between taxa or communities. Such an approach has proven useful in plant ecology where a handful of plant traits can be used to accurately place species into multidimensional trait space (Laughlin et al., 2010; deBello et al., 2013) as many other plant traits are often correlated with this subset of traits. There are clearly similar trade-offs in microbial traits (Gudelj et al., 2010). For example, there are well-established trade-offs between growth rate and yield (Pfeiffer et al., 2001), between stress tolerance and the ability to compete for substrates (Ferenci and Spira, 2007), and between cell size and nutrient uptake rates (Young, 2006). Although the list of possible traits is enormous, there are inescapable morphological, physiological, and genetic constraints that narrow the list of possible trait combinations associated with different life history strategies.

Selecting CATs to study can often be done *a priori*. For example, we can assume that traits that confer tolerance to changes in water activity will likely be important if we are trying to predict soil C dynamics in arid or semi-arid systems (Lennon et al., 2012; Evans and Wallenstein, 2014). Likewise, we could assume that traits associated with nutrient uptake will have important controls on phytoplankton growth in many freshwater systems (Edwards et al., 2013). Sometimes the traits that might be relevant to predicting the community-level function are unknown or unexpected because the process itself or specific controls on the process are not well understood. For example, it would be difficult to predict *a priori* which CATs would be most relevant to understanding controls on nitrification by archaea given that physiologies remain poorly understood (You et al., 2009). In these cases, one could determine which CATs are most relevant to a process of interest by first measuring the community-level process across space and/or time and then empirically determining what CAT, or set of CATs, appears to be correlated with the measured changes in the process. In this manner, one could generate specific, testable predictions about how changes in CATs relate to changes in community-level processes and test them experimentally.

### INFERRING COMMUNITY-AGGREGATED TRAITS FROM METAGENOMIC DATA

Many microbial traits could be inferred from genomic or metagenomic data (Table 1) and there are many other examples of traits

**Table 1 | Selected examples of microbial traits and the genes or genomic characteristics that could be used to calculate community-aggregated trait (CAT) values from shotgun metagenomic data.**

Microbial trait	Selected genes, gene categories, or genomic characteristics that could be used to infer the trait value	Reference
Maximum growth rate	rRNA operon number, codon usage bias in highly expressed genes, rRNA/tRNA position	Lauro et al. (2009), Vieira-Silva and Rocha (2010)
Dormancy	Sporulation proteins, toxin–antitoxin systems, resuscitation-promoting factors	Lennon and Jones (2011)
Osmoregulation	Trehalose and peptidoglycan production	Culligan et al. (2012)
Ability to catabolize recalcitrant organic compounds	Genome size, secondary metabolite transport/metabolism	Konstantinidis and Tiedje (2004), Fierer et al. (2012)
Stress resistance (general)	Sigma factor subunits of RNA polymerases (e.g., $\sigma^B$ )	Kazmierczak et al. (2003), Marles-Wright and Lewis (2007)
Cold tolerance	Cold shock proteins, trehalose synthesis proteins	Varin et al. (2012)
Motility	Chemoreceptor, flagellar genes	Girgis et al. (2007)
Oxidative stress tolerance	Catalase, peroxidase, and polyketide synthase genes	Qin et al. (2012), Zhang et al. (2013)
Nitrogen/phosphorus affinities	Genes for membrane-bound nutrient uptake/transporters	Hewson et al. (2009), Eloë et al. (2011)
Resistance to toxic metals	COGs associated with heavy metal efflux pumps	Hemme et al. (2010), Eloë et al. (2011)
Antibiotic resistance	Genes for efflux pumps, ribosomal protection, enzymatic inactivators	Hu et al. (2013), Forsberg et al. (2014)

References include examples of studies where that trait was inferred from either genomic or metagenomic data.

that could be inferred directly from either the presence/absence of specific genes, including specific metabolic capabilities (e.g., the ability to fix CO<sub>2</sub> or oxidize elemental sulfur) or tolerances to specific environmental stressors (e.g., resistance to antibiotics or toxic metals). Other traits could be determined from genomic data but, in many cases, there will not be a single gene, or set of genes, whose presence or absence is directly associated with the trait. Instead, one would have to identify what gene abundances or gene ratios correlate with the trait of interest. For example, maximum growth rate is clearly an important trait, but there is no single gene that directly determines growth rate. However, a variety of genomic features (including number of rRNA copies, number of outer membrane proteins, presence of motility genes) have been shown to distinguish copiotrophic and oligotrophic taxa (Klappenbach et al., 2000; Lauro et al., 2009) with work by Vieira-Silva and Rocha (2010) demonstrating how such information could be used with metagenomic data to estimate microbial generation times in environmental samples. In many cases, the process of inferring specific CATs from metagenomic data will require analyzing genomes from individual taxa with well-characterized phenotypic traits or by analyzing metagenomes from communities across well-characterized gradients in CAT-space (Barberan et al., 2012). For example, there are likely genes, or combinations of genes, that could provide insight into the nutrient stoichiometries of individual taxa or whole communities, but these genes will have to be identified and their utility validated.

There are numerous caveats to consider when trying to use metagenomic data to quantify microbial CATs and nearly all of these caveats are shared by any study using metagenomes to infer the functional attributes of communities. First and foremost, traits are phenotypic characteristics of organisms and changes in

metagenomes will not necessarily equate with changes in CATs. Metagenomes allow us to make some guesses and hypotheses about changes in CATs, but these hypotheses would need to be confirmed independently. This is particularly true for traits, like cell size (Young, 2006) and nutrient stoichiometries (Scott et al., 2012), which can exhibit a high degree of plasticity depending on environmental conditions. A related concern is that CAT estimates from metagenomes will reflect both active and inactive members of the community and in many environments, like soil, a large portion of the DNA pool isolated at a given point in time could come from cells that are dormant, inactive, or even lysed (Lennon and Jones, 2011). Perhaps a larger concern is that we are often going to be restricted to estimating CATs from that portion of the metagenome that has been annotated. In environmental metagenomes, typically >50% of the genes found in metagenomes are of undetermined function (Ellrott et al., 2010) and it is a big leap of faith to assume that those annotations are thorough and accurate (Schnoes et al., 2009). Although some traits can be inferred from unannotated metagenomic data (e.g., codon usage, Roller et al., 2013), the requirement that genes are first annotated is an important limitation in most environmental samples as the large fraction of genes that fall into the ‘known unknown’ category may obscure relationships between CATs and the community-level attributes one is trying to predict.

## CONCLUSION

Microbial ecologists are inherently interested in the details and complexities of microbial communities. This is not true for most people outside the field – they want to know what microbial communities do, how they can be managed, and how they impact our health and environmental quality. Thus, there is a need for

quantifiable information on community-level functions and the relevant community-level properties. Unfortunately, the traits of most microbial taxa living in the environment remain poorly understood and trying to understand community-level properties from trait-based information on individual cells or individual taxa is often difficult, if not impossible. There is no method waiting on the horizon that will serve as a panacea to close these knowledge gaps. However, CATs represent a valuable conceptual approach that could be used by microbial ecologists to make better use of shotgun metagenomic data for predicting the functional capabilities of complex communities. The key is to determine what CATs to measure and how the relevant CATs can be quantified from metagenomic data.

## ACKNOWLEDGMENTS

We thank the graduate students at the University of Colorado who participated in the “Genomes and Traits” seminar for useful discussions on some of the topics discussed here. Support was provided by a grant from the National Science Foundation (DEB0953331) to Noah Fierer, a James S. McDonnell (JSMF) Postdoctoral Fellowship to Albert Barberán, and a grant (UOW1201) from the Royal Society of New Zealand Marsden Fund awarded to Daniel C. Laughlin.

## REFERENCES

- Allison, S. D. (2012). A trait-based approach for modelling microbial litter decomposition. *Ecol. Lett.* 15, 1058–1070. doi: 10.1111/j.1461-0248.2012.01807.x
- Bailey, V. L., Smith, J. L., and Bolton, H. (2002). Fungal-to-bacterial ratios in soils investigated for enhanced C sequestration. *Soil Biol. Biochem.* 34, 997–1007. doi: 10.1016/S0038-0717(02)00033-0
- Barberán, A., Fernandez-Guerra, A., Bohannan, B. J. M., and Casamayor, E. O. (2012). Exploration of community traits as ecological markers in microbial metagenomes. *Mol. Ecol.* 21, 1909–1917. doi: 10.1111/J.1365-294X.2011.05383.X
- Culligan, E. P., Sleator, R. D., Marchesi, J. R., and Hill, C. (2012). Functional metagenomics reveals novel salt tolerance loci from the human gut microbiome. *ISME J.* 6, 1916–1925. doi: 10.1038/ismej.2012.38
- deBello, F., Lavorel, S., Lavergne, S., Albert, C. H., Boulangeat, I., Mazel, F., et al. (2013). Hierarchical effects of environmental filters on the functional structure of plant communities: a case study in the French Alps. *Ecography* 36, 393–402. doi: 10.1111/j.1600-0587.2012.07438.x
- Dutkiewicz, S., Follows, M. J., and Bragg, J. G. (2009). Modeling the coupling of ocean ecology and biogeochemistry. *Global Biogeochem. Cycles* 23:GB4017. doi: 10.1029/2008GB003405
- Edwards, K. F., Litchman, E., and Klausmeier, C. A. (2013). Functional traits explain phytoplankton responses to environmental gradients across lakes of the United States. *Ecology* 94, 1626–1635. doi: 10.1890/12-1459.1
- Ellrott, K., Jaroszewski, L., Li, W. Z., Wooley, J. C., and Godzik, A. (2010). Expansion of the protein repertoire in newly explored environments: human gut microbiome specific protein families. *PLoS Comput. Biol.* 6:e1000798. doi: 10.1371/journal.pcbi.1000798
- Eloe, E. A., Fadrosch, D. W., Novotny, M., Zeigler, L. A., Kim, M., Lombardo, M.-J., et al. (2011). Going deeper: metagenome of a hadopelagic microbial community. *PLoS ONE* 6:e20388. doi: 10.1371/journal.pone.0020388
- Evans, S. E., and Wallenstein, M. D. (2014). Climate change alters ecological strategies of soil bacteria. *Ecol. Lett.* 17, 155–164. doi: 10.1111/ele.12206
- Ferenci, T., and Spira, B. (2007). Variation in stress responses within a bacterial species and the indirect costs of stress resistance. *Ann. N. Y. Acad. Sci.* 1113, 105–113. doi: 10.1196/annals.1391.003
- Fierer, N., Bradford, M., and Jackson, R. (2007). Toward an ecological classification of soil bacteria. *Ecology* 88, 1354–1364. doi: 10.1890/05-1839
- Fierer, N., Leff, J. W., Adams, B. J., Nielsen, U. N., Bates, S. T., Lauber, C. L., et al. (2012). Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proc. Natl. Acad. Sci. U.S.A.* 109, 21390–21395. doi: 10.1073/pnas.1215210110
- Fierer, N., and Lennon, J. T. (2011). The generation and maintenance of diversity in microbial communities. *Am. J. Bot.* 98, 439–448. doi: 10.3732/Ajb.100.0498
- Finkel, Z. V., Beardall, J., Flynn, K. J., Quigg, A., Rees, T. A. V., and Raven, J. A. (2009). Phytoplankton in a changing world: cell size and elemental stoichiometry. *J. Plank. Res.* 32, 119–137. doi: 10.1093/plankt/fbp098
- Forsberg, K. J., Patel, S., Gibson, M. K., Lauber, C. L., Knight, R., Fierer, N., et al. (2014). Bacterial phylogeny structures soil resistomes across habitats. *Nature* 509, 612–616. doi: 10.1038/Nature13377
- Forsberg, K. J., Reyes, A., Wang, B., Selleck, E. M., Sommer, M. O. A., and Dantas, G. (2012). The shared antibiotic resistome of soil bacteria and human pathogens. *Science* 337, 1107–1111. doi: 10.1126/science.1220761
- Ganesh, S., Parris, D. J., Delong, E. F., and Stewart, F. J. (2014). Metagenomic analysis of size-fractionated picoplankton in a marine oxygen minimum zone. *ISME J.* 8, 187–211. doi: 10.1038/ismej.2013.144
- Garnier, E., Cortez, J., Billès, G., Navas, M.-L., Roumet, C., Debussche, M., et al. (2004). Plant functional markers capture ecosystem properties during secondary succession. *Ecology* 85, 2630–2637. doi: 10.1890/03-0799
- Gaucherand, S., and Lavorel, S. (2007). New method for rapid assessment of the functional composition of herbaceous plant communities. *Austral Ecol.* 32, 927–936. doi: 10.1111/J.1442-9993.2007.01781.X
- Girgis, H. S., Liu, Y., Ryu, W. S., and Tavazoie, S. (2007). A comprehensive genetic characterization of bacterial motility. *PLoS Genet.* 3:e154. doi: 10.1371/journal.pgen.0030154
- Greenblum, S., Turnbaugh, P. J., and Borenstein, E. (2012). Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease. *Proc. Natl. Acad. Sci. U.S.A.* 109, 594–599. doi: 10.1073/pnas.1116053109
- Grime, J. P. (1998). Benefits of plant diversity to ecosystems: immediate, filter and founder effects. *J. Ecol.* 86, 902–910. doi: 10.1046/J.1365-2745.1998.00306.X
- Gudelj, I., Weitz, J. S., Ferenci, T., Horner-Devine, M. C., Marx, C. J., Meyer, J. R., et al. (2010). An integrative approach to understanding microbial diversity: from intracellular mechanisms to community structure. *Ecol. Lett.* 13, 1073–1084. doi: 10.1111/J.1461-0248.2010.01507.X
- Handelsman, J. (2004). Metagenomics: application of genomics to uncultured microorganisms. *Microbiol. Mol. Biol. Rev.* 68, 669–685. doi: 10.1128/MMBR.68.4.669-685.2004
- Hemme, C. L., Deng, Y., Gentry, T. J., Fields, M. W., Wu, L., Barua, S., et al. (2010). Metagenomic insights into evolution of a heavy metal-contaminated groundwater microbial community. *ISME J.* 4, 660–672. doi: 10.1038/ismej.2009.154
- Hewson, I., Paerl, R. W., Tripp, H. J., Zehr, J. P., and Karl, D. M. (2009). Metagenomic potential of microbial assemblages in the surface waters of the central Pacific Ocean tracks variability in oceanic habitat. *Limnol. Oceanogr.* 54, 1981–1994. doi: 10.4319/Lo.2009.54.6.1981
- Ho, A., Kerckhof, E.-M., Luke, C., Reim, A., Krause, S., Boon, N., et al. (2013). Conceptualizing functional traits and ecological characteristics of methane-oxidizing bacteria as life strategies. *Environ. Microbiol. Rep.* 5, 335–345. doi: 10.1111/j.1758-2229.2012.00370.x
- Homolová, L., Malenkovský, Z., Clevers, J. G. P. W., García-Santos, G., and Schaeppman, M. E. (2013). Review of optical-based remote sensing for plant trait mapping. *Ecol. Complex.* 15, 1–16. doi: 10.1016/j.ecocom.2013.06.003
- Howard, E. C., Sun, S., Biers, E. J., and Moran, M. A. (2008). Abundant and diverse bacteria involved in DMSP degradation in marine surface waters. *Environ. Microbiol.* 10, 2397–2410. doi: 10.1111/j.1462-2920.2008.01665.x
- Hu, Y., Yang, X., Qin, J., Lu, N., Cheng, G., Wu, N., et al. (2013). Metagenome-wide analysis of antibiotic resistance genes in a large cohort of human gut microbiota. *Nat. Commun.* 4:2151. doi: 10.1038/ncomms3151
- Hug, L., Castelle, C., Wrighton, K., Thomas, B., Sharon, I., Frischkorn, K., et al. (2013). Community genomic analyses constrain the distribution of metabolic traits across the Chloroflexi phylum and indicate roles in sediment carbon cycling. *Microbiome* 1:22. doi: 10.1186/2049-2618-1-22
- Kazmierczak, M. J., Mithoe, S. C., Boor, K. J., and Wiedmann, M. (2003). *Listeria* monocytogenes  $\sigma$ B regulates stress response and virulence functions. *J. Bacteriol.* 185, 5722–5734. doi: 10.1128/jb.185.19.5722-5734.2003
- Klappenbach, J. A., Dunbar, J. M., and Schmidt, T. M. (2000). rRNA operon copy number reflects ecological strategies of bacteria. *Appl. Environ. Microbiol.* 66, 1328–1333. doi: 10.1128/AEM.66.4.1328-1333.2000

- Konstantinidis, K. T., and Tiedje, J. M. (2004). Trends between gene content and genome size in prokaryotic species with larger genomes. *Proc. Natl. Acad. Sci. U.S.A.* 101, 3160–3165. doi: 10.1073/pnas.0308653100
- Langille, M. G. I., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A., et al. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* 31, 814–821. doi: 10.1038/nbt.2676
- Laughlin, D. C. (2011). Nitrification is linked to dominant leaf traits rather than functional diversity. *J. Ecol.* 99, 1091–1099. doi: 10.1111/J.1365-2745.2011.01856.X
- Laughlin, D. C. (2014). The intrinsic dimensionality of plant traits and its relevance to community assembly. *J. Ecol.* 102, 186–193. doi: 10.1111/1365-2745.12187
- Laughlin, D. C., Leppert, J. J., Moore, M. M., and Sieg, C. H. (2010). A multi-trait test of the leaf-height-seed plant strategy scheme with 133 species from a pine forest flora. *Funct. Ecol.* 24, 493–501. doi: 10.1111/j.1365-2435.2009.01672.x
- Lauro, F. M., McDougald, D., Thomas, T., Williams, T. J., Egan, S., Rice, S., et al. (2009). The genomic basis of trophic strategy in marine bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 106, 15527–15533. doi: 10.1073/pnas.0903507106
- Lavorel, S., and Grigulis, K. (2012). How fundamental plant functional trait relationships scale-up to trade-offs and synergies in ecosystem services. *J. Ecol.* 100, 128–140. doi: 10.1111/j.1365-2745.2011.01914.x
- Lavorel, S., Grigulis, K., Lamarque, P., Colace, M. P., Garden, D., Girel, J., et al. (2011). Using plant functional traits to understand the landscape distribution of multiple ecosystem services. *J. Ecol.* 99, 135–147. doi: 10.1111/J.1365-2745.2010.01753.X
- Lennon, J. T., Aanderud, Z. T., Lehmkuhl, B. K., and Schoolmaster, D. R. (2012). Mapping the niche space of soil microorganisms using taxonomy and traits. *Ecology* 93, 1867–1879. doi: 10.1890/11-1745.1
- Lennon, J. T., and Jones, S. E. (2011). Microbial seed banks: the ecological and evolutionary implications of dormancy. *Nat. Rev. Microbiol.* 9, 119–130. doi: 10.1038/Nrmicro2504
- Mackelprang, R., Waldrop, M. P., Deangelis, K. M., David, M. M., Chavarria, K. L., Blazewicz, S. J., et al. (2011). Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. *Nature* 480, 368–371. doi: 10.1038/Nature10576
- Marles-Wright, J., and Lewis, R. J. (2007). Stress responses of bacteria. *Curr. Opin. Struct. Biol.* 17, 755–760. doi: 10.1016/j.sbi.2007.08.004
- Martin, H. G., Ivanova, N., Kunin, V., Warnecke, F., Barry, K. W., McHardy, A. C., et al. (2006). Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. *Nat. Biotechnol.* 24, 1263–1269. doi: 10.1038/nbt1247
- Mokany, K., Ash, J., and Roxburgh, S. (2008). Functional identity is more important than diversity in influencing ecosystem processes in a temperate native grassland. *J. Ecol.* 96, 884–893. doi: 10.1111/j.1365-2745.2008.01395.x
- Paine, R. T. (1995). A conversation on refining the concept of keystone species. *Conserv. Biol.* 9, 962–964. doi: 10.1046/j.1523-1739.1995.09040962.x
- Pett-Ridge, J., and Firestone, M. K. (2005). Redox fluctuation structures microbial communities in a wet tropical soil. *Appl. Environ. Microbiol.* 71, 6998–7007. doi: 10.1128/aem.71.11.6998-7007.2005
- Pfeiffer, T., Schuster, S., and Bonhoeffer, S. (2001). Cooperation and competition in the evolution of ATP-producing pathways. *Science* 292, 504–507. doi: 10.1126/Science.1058079
- Plewniak, F., Koehler, S., Navet, B., Dugat-Bony, É., Bouchez, O., Peyret, P., et al. (2013). Metagenomic insights into microbial metabolism affecting arsenic dispersion in Mediterranean marine sediments. *Mol. Ecol.* 22, 4870–4883. doi: 10.1111/mec.12432
- Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., Zhang, F., et al. (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* 490, 55–60. doi: 10.1038/nature11450
- Roller, M., Lucic, V., Nagy, I., Perica, T., and Vlahovick, K. (2013). Environmental shaping of codon usage and functional adaptation across microbial communities. *Nucleic. Acids Res.* 41, 8842–8852. doi: 10.1093/Nar/Gkt673
- Schnoes, A. M., Brown, S. D., Dodevski, I., and Babbitt, P. C. (2009). Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. *PLoS Comput. Biol.* 5:e1000605. doi: 10.1371/journal.pcbi.1000605
- Scott, T., Cotner, J., and Lapara, T. (2012). Variable stoichiometry and homeostatic regulation of bacterial biomass elemental composition. *Front. Microbiol.* 3:42. doi: 10.3389/fmicb.2012.00042
- Sieuwert, S., Molenaar, D., Van Hijum, S. F. T., Beerthuyzen, M., Stevens, M. J. A., Janssen, P. W. M., et al. (2010). Mixed-culture transcriptome analysis reveals the molecular basis of mixed-culture growth in *Streptococcus thermophilus* and *Lactobacillus bulgaricus*. *Appl. Environ. Microbiol.* 76, 7775–7784. doi: 10.1128/Aem.01122-10
- Strickland, M. S., and Rousk, J. (2010). Considering fungal:bacterial dominance in soils – methods, controls, and ecosystem implications. *Soil Biol. Biochem.* 42, 1385–1395. doi: 10.1016/j.soilbio.2010.05.007
- Taylor, P. G., and Townsend, A. R. (2010). Stoichiometric control of organic carbon-nitrate relationships from soils to the sea. *Nature* 464, 1178–1181. doi: 10.1038/nature08985
- Varin, T., Lovejoy, C., Jungblut, A. D., Vincent, W. F., and Corbeil, J. (2012). Metagenomic analysis of stress genes in microbial mat communities from Antarctica and the high Arctic. *Appl. Environ. Microbiol.* 78, 549–559. doi: 10.1128/aem.06354-11.
- Veira-Silva, S., and Rocha, E. P. C. (2010). The systemic imprint of growth and its uses in ecological (meta) genomics. *PLoS Genet.* 6:e1000808. doi: 10.1371/journal.pgen.1000808
- Vile, D., Shipley, B., and Garnier, E. (2006). Ecosystem productivity can be predicted from potential relative growth rate and species abundance. *Ecol. Lett.* 9, 1061–1067. doi: 10.1111/j.1461-0248.2006.00958.x
- Violle, C., Enquist, B. J., McGill, B. J., Jiang, L., Albert, C. H., Hulshof, C., et al. (2012). The return of the variance: intraspecific variability in community ecology. *Trends Ecol. Evol.* 27, 244–252. doi: 10.1016/J.Tree.2011.11.014
- Walsh, D. A., Zaikova, E., Howes, C. G., Song, Y. C., Wright, J. J., Tringe, S. G., et al. (2009). Metagenome of a versatile chemolithoautotroph from expanding oceanic dead zones. *Science* 326, 578–582. doi: 10.1126/science.1175309
- Whitman, W., Coleman, D., and Wiebe, W. (1998). Prokaryotes: the unseen majority. *Proc. Natl. Acad. Sci. U.S.A.* 95, 6578–6583. doi: 10.1073/pnas.95.12.6578
- Wieder, W. R., Bonan, G. B., and Allison, S. D. (2013). Global soil carbon projections are improved by modelling microbial processes. *Nat. Clim. Chang.* 3, 909–912. doi: 10.1038/nclimate1951
- You, J., Das, A., Dolan, E.M., and Hu, Z. (2009). Ammonia-oxidizing archaea involved in nitrogen removal. *Water Res.* 43, 1801–1809. doi: 10.1016/j.watres.2009.01.016
- Young, K. D. (2006). The selective value of bacterial shape. *Microbiol. Mol. Biol. Rev.* 70, 660–703. doi: 10.1128/Mmbr.00001-06
- Zhang, W., Wang, Y., Lee, O. O., Tian, R., Cao, H., Gao, Z., et al. (2013). Adaptation of intertidal biofilm communities is driven by metal ion and oxidative stresses. *Sci. Rep.* 3:3180. doi: 10.1038/srep03180

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 22 September 2014; accepted: 28 October 2014; published online: 12 November 2014.

Citation: Fierer N, Barberán A and Laughlin DC (2014) Seeing the forest for the genes: using metagenomics to infer the aggregated traits of microbial communities. *Front. Microbiol.* 5:614. doi: 10.3389/fmicb.2014.00614

This article was submitted to *Systems Microbiology*, a section of the journal *Frontiers in Microbiology*.

Copyright © 2014 Fierer, Barberán and Laughlin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.