

Environmental proteomics, biodiversity statistics and food-web structure

Nicholas J. Gotelli¹, Aaron M. Ellison² and Bryan A. Ballif^{1,3}

¹ Department of Biology, University of Vermont, 109 Carrigan Drive, Burlington, VT 05405, USA

² Harvard Forest, Harvard University, 324 North Main Street, Petersham, MA 01366, USA

³ Vermont Genetics Network Proteomics Facility, University of Vermont, 109 Carrigan Drive, Burlington, VT 05405, USA

Pioneering studies in environmental proteomics have revealed links between protein diversity and ecological function in simple ecological communities, such as microbial biofilms. In the near future, high-throughput proteomic methods will be applied to more complex ecological systems in which microbes and macrobes interact. Data structures in biodiversity and protein surveys have many similarities, so the statistical methods that ecologists use for analyzing biodiversity data should be adapted for use with quantitative surveys of protein diversity. However, increasing quantities of protein and bioinformatics data will not, by themselves, reveal the functional significance of proteins. Instead, ecologists should be measuring changes in the abundance of protein cohorts in response to replicated field manipulations, including nutrient enrichment and removal of top predators.

Surveying the protein diversity of communities and ecosystems

High-throughput methods for studying and characterizing large numbers of proteins of uncultured biological samples (metaproteomics; see [Glossary](#)) now allow for cataloguing the proteins of component species of assemblages (community proteomics) or of parts of ecosystems, without precise knowledge of the organism that produced the protein (environmental proteomics) [1,2]. Pioneering studies have surveyed relatively simple, microbial-dominated assemblages (reviewed in [3,4]), sampled from habitats such as acid-mine drainages [5], microfilm surfaces of leaves [6], seawater [7] and soil [8] (Box 1).

These methods and results should be of great interest to all community and ecosystem ecologists, because patterns of differential protein production and expression reflect physiological responses to changing or stressful environmental conditions [9], including climatic change [10] and the presence of predators [11]. In addition, shifts in protein abundance and composition, which are best characterized through analyses of microbial processes [12], can indicate changes in magnitude or rates of material and energetic fluxes within and between ecosystems [13].

Although pioneering studies of microbes suggest directions for future work in environmental proteomics, it is

unclear how easily results from microbial assemblages can be scaled to more complex ecological systems. The structural complexity of microbial systems appears to be simplified relative to that of most terrestrial and aquatic ecosystems. Biofilm and microbial systems generally lack the photosynthetic primary producers, detritivores and higher trophic levels that characterize typical 'green' and 'brown' food webs [14] of interacting multicellular ('macrobial') eukaryotes and prokaryotes [15]. Technical challenges, including the identification of proteins from species without sequenced genomes, variation in the physical and biological conformation of proteins, *in situ* activity of isolated proteins and the difficulty of reliably extracting proteins from complex media such as seawater and soil [16], also hinder the application of proteomic (and genomic) approaches to macrobial assemblages and food webs [3]. Nevertheless, as these challenges are overcome with new technologies and bioinformatics tools, proteomic surveys of a range of ecological communities and ecosystems will become an important complement to parallel metagenomic, metatranscriptomic and metametabolomic surveys [2]. All of these methods generate complementary data, although from our perspective, proteins are the most desirable unit of study because they are most closely related to the functioning of ecosystems and are a more direct measure of the 'molecular phenotype' [17].

Glossary

Community proteomics: proteomic surveys of the taxa that comprise a local assemblage or community.

Environmental proteomics: proteomic surveys of an assemblage or community without specification of the particular taxon that a protein originated from.

Gene ontology database: a relational database categorizing genes and proteins by function, expression level, biochemical activity, metabolic process, subcellular localization, or similar attributes for bioinformatic analyses.

Metagenomics: the large-scale characterization of genetic material recovered from a biological or environmental source.

Metametabolomics: the large-scale characterization of small-molecule metabolites from a biological or environmental source.

Metaproteomics: the large-scale characterization of proteins recovered from a biological or environmental source.

Metatranscriptomics: the large-scale characterization of transcribed RNA molecules from a biological or environmental source.

Molecular phenotyping: the large-scale composite characterization of biological samples by surveying differences in transcript or protein levels.

Tryptic peptide: the proteolytic product following the digestion of a protein with trypsin. Trypsin is a reliable and economical enzyme that cuts after basic residues in proteins, yielding peptides that are highly compatible with mass spectrometry applications.

Corresponding author: Gotelli, N.J. (ngotelli@uvm.edu).

Box 1. 'Omic characterization of community structure and function

Acid-rock and acid-mine drainages (AMDs) can be found wherever water moves through pyrite (iron sulfide: FeS_2); examples include mineral-rich deposits in caves and piles of tailings from locations where sulfide minerals (including uranium and coal) are mined. AMDs have very low pH (often reaching 0.5–1.0), but can host highly productive ecosystems consisting of diverse bacteria, Archaea and eukaryotes. Many of these taxa are chemoautotrophs that have evolved adaptations to heavy-metal environments [69]. The chemical isolation of these communities from external sources of carbon and nitrogen, and their physical segregation from higher pH assemblages in specific drainages makes them an ideal self-contained, micro-ecosystem for assessing proteomic diversity and function in a microbial system.

Banfield *et al.* [12] sampled the AMD biofilm community at six sites along an environmental gradient (defined by distance from ore deposits and seepage flow) at the Richmond Iron Mountain Mine Complex in northern California. Diversity of microbes from water samples and biofilm scraped from rock surfaces was determined using a combination of *in situ* hybridization with fluorescent probes screened with epifluorescence and confocal microscopy, followed by direct DNA sequencing of entire genomes. The sequence data were then used to identify proteins from particular dominant organisms. Unknown sequences were screened against proteomic databases to identify common proteins. Functions of novel proteins were characterized further with biological observation, mass spectrometry, biochemical assays (including microtiter plates that can assay approximately 100 000 samples at a time for multiple biochemical functions) and protein-structural modeling.

Perhaps the most important practical result of ongoing technological advances is that ecologists will be able to survey proteomes efficiently and repeatedly and describe their variability in natural ecosystems, rather than ignoring variation and classifying species, communities and their proteomes using fixed typologies. For example, Figure 1 shows three proteomic 'fingerprints' of samples from three replicate food webs inhabiting water-filled leaves of the pitcher plant *Sarracenia purpurea*. Each pitcher supports an independent, intact food web consisting of a resource base of captured insect prey (mostly ants and flies), a sub-web of bacteria, protozoa, algae and fungi, and a suite of aquatic invertebrate larvae (flies, midges and mosquitoes) that interact as filter feeders, detritivores, omnivores and top predators [18]. The obvious variability in protein profiles among replicate food webs in the microbial component of each web (Figure 1) probably reflects differences in species composition and abundance of invertebrates and microbes within each leaf, along with differences in the quantity and composition of captured insect prey. At the same time, consistent protein bands suggest repeatable patterns and protein structure among replicate food webs. In an associated community proteomics survey [19], species-specific proteins and peptide signatures of the three most abundant macroinvertebrates were revealed, a first step towards proteomic profiling of an entire microbe–macrobe ecosystem.

Understanding the functional significance of protein diversity

In just two decades, cost-effective methods have been developed to increase throughput and comprehensive identification of proteins [20,21]; accurately quantify protein relative abundance [22,23]; enrich rare protein types and

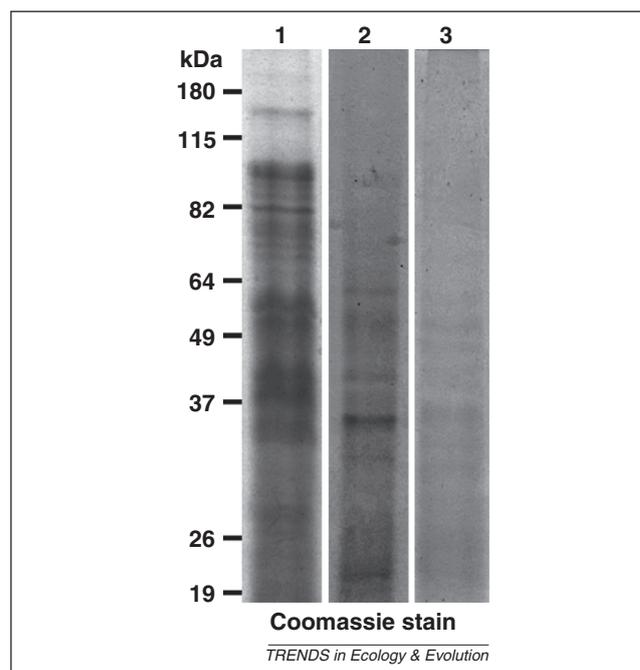


Figure 1. SDS-PAGE separation of proteins from the microbial fraction of material within three individual *Sarracenia purpurea* ecosystems. The pitcher-plant liquid was extracted from each plant and filtered using a Bio-Rad polyprep column with a porous 30- μm polyethylene frit bed and gentle suction; 2 ml of the filtered liquid was subjected to centrifugation at approximately $13\,000 \times g$. The pellet was resuspended in 100- μl denaturing protein sample buffer, heated to 95°C for five minutes, and 20 μl was subjected to SDS-PAGE. Gels were stained with Coomassie blue. Approximate molecular weights are indicated in kilodaltons (kDa). Some protein bands appear to be common to all samples (e.g. approximately 55 kDa), whereas others (e.g. approximately 82 kDa and approximately 140 kDa in Sample 1) appear to be unique to a particular sample.

protein modifications [24,25]; and expand the capacity and precision of bioinformatic tools that enhance the functional interpretation of proteomic data sets [26,27]. Although some of these methods are still in their infancy, rapid technological and statistical advances daily are increasing the quantity and quality of information about proteins that can be sampled from functioning ecosystems. However, by itself, this flood of bioinformatics data is unlikely to reveal which particular proteins are important causes and consequences of community structure and ecosystem function.

An analogous disconnection between data density and functional inference plagued community ecology in the initial stages of documenting the diversity of species, their assembly and their interactions within food webs. Pioneering naturalists and taxonomists, such as Rumphius, Linnaeus, Darwin, Wallace and von Humboldt, among many others, collected, catalogued and named species from around the world. Beginning during the late 1800s, ecologists assembled detailed records of food-web structure (extensive lists of 'who eats whom') in plant and animal assemblages. Now, species diversity and food-web structure are both encapsulated in summary statistics and indices, respectively [28] and as network diagrams with nodes representing species and links representing trophic interactions including predation and parasitism [29].

However, pattern identification and summary statistics are not the same as identifying causal processes, which

requires experimental manipulations and statistical modeling of the results. In studies of biodiversity [30,31] and ecological food webs [32,33], there are long dialectical traditions of thesis, antithesis and synthesis. The same dialectic could be applied profitably to proteomics research in ecology: characterizing and summarizing the diversity of proteins and their potential interactions, analyzing proteomic assemblages of experimentally manipulated food webs in the field and synthesizing the resulting data with statistical modeling.

Characterizing and summarizing proteomic diversity

Just as the first step to working with an ecosystem is identifying its constituent organisms, a first step towards understanding the functional significance of protein diversity of an ecosystem will be identifying and enumerating its distinct proteins and estimating their relative abundances. However, just as it is impossible to detect all of the species in a microbial or macrobial assemblage [34,35], so too it is unlikely that it will be possible to detail an entire ecological proteome in the foreseeable future. Estimates of proteomic diversity in microbial ecosystems range from 10^4 to 10^9 expressed proteins [16]. There are approximately 6000–60 000 protein-encoding genes in the genome of a single prokaryote or eukaryote species [36], and even a single tissue sampled from a single multicellular species contains hundreds or even thousands of distinct proteins [37]. In ecological assemblages that contain both microbes and macrobes, proteomic diversity will probably exceed 10^{10} expressed proteins because, unlike the genome, the proteome of a multicellular organism is not constant during its lifetime, but changes ontogenetically [38] and in response to changing biotic and abiotic conditions [9].

For proteins, as for species, ecologists must remain satisfied with sampling only a small fraction of the potential diversity. The resulting data from surveys of both ecosystems and proteomes are lists of the component species (some of which might be taxonomically ambiguous) or proteins (some of which might be incompletely characterized because a single tryptic peptide might map to several, typically related, proteins) and an estimate of relative abundances (usually from counts of individual organisms [31] or counts of peptides from different proteins [39]). Statistical distributions of both species and proteins are also similar: usually, there is a small number of relatively common species (or proteins), a large number of relatively rare species (or proteins) and an unknown number of species (or proteins) present, but that are not detected in the sample. The observed distribution of relative abundances of species is approximately log-normal in shape [40], although the precise statistical characterization continues to be controversial [41]. However, the observed distribution of the relative abundance of proteins has not yet been analyzed with the same statistical tools that ecologists have used to analyze the relative abundance of species.

Over the past several decades, ecologists have developed a unified toolbox of non-parametric statistical methods for analyzing such data. Rarefaction methods interpolate diversity estimates of small random subsamples of data, allowing investigators to compare multiple, standardized

data sets on the basis of common sampling effort [42]. Asymptotic estimators can extrapolate diversity data to estimate the minimum number of undetected species (or proteins) in a sample, based on the frequencies of the rare species that are present [43]. The amount of sampling effort that would be needed to reveal these undetected elements also can be calculated [44]. Finally, sample variances and confidence intervals characterize the uncertainty associated with such interpolations and extrapolations [45].

These statistical tools have not yet been adapted for protein studies; Koziol *et al.* [46] is an exemplary study, and the only published example we know of in which biodiversity sampling models were applied to proteomic data (Figure 2). Although it might seem premature to discuss statistical methods for data that have not yet been collected at the ecosystem and community level, we think that it is crucial to look forward and anticipate methods of analysis for proteomic data that will emerge in future studies. Such analyses might reveal, for example, whether the relationship between protein diversity and ecosystem function mirrors that of species diversity and ecosystem function.

However, the analysis of proteomic data differs from analysis of biodiversity samples in one crucial way. To characterize a proteome, the proteins are proteolytically digested into smaller peptides before analysis; this is analogous to taking the species sampled, shredding them into small pieces and then reassembling, identifying and counting the reconstituted organisms. Considerable effort in bioinformatics is devoted towards statistically reconstructing the original protein(s) from these peptide fragments.

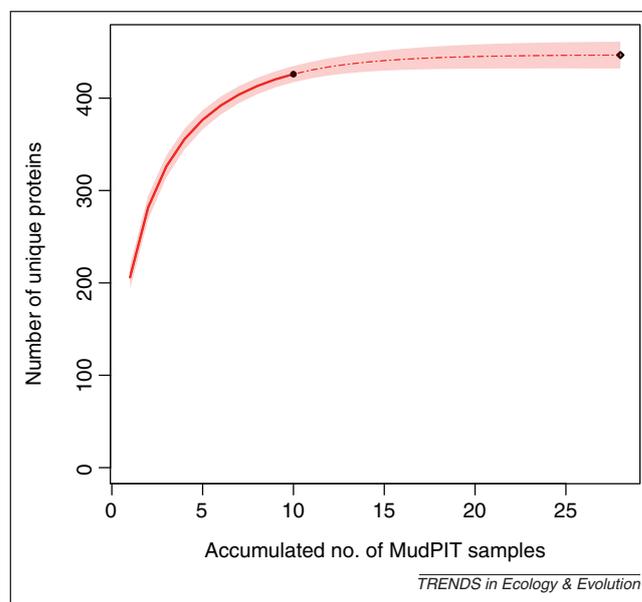


Figure 2. Sample-based rarefaction analysis of proteomic data. The x-axis is the number of multidimensional protein identification technology (MudPIT) samples of rat endothelial tissue (original data from [67]). The y-axis is the number of unique proteins discovered. The set of 10 MudPIT samples yielded a total of 426 unique proteins (left solid circle). The unbroken red line is the rarefaction curve interpolated from the original data to smaller numbers of samples. The shaded funnel is a symmetric parametric 95% confidence interval, based on 1000 replicate draws [44]. The broken red line is the extrapolation out to the asymptotic minimum estimator (right solid circle) of 446 proteins [68]. Using the estimator of Chao *et al.* [44], an additional 18 MudPIT samples would be needed to reach this asymptote. Koziol *et al.* [46] describe other sampling models with these data.

This added twist makes it difficult to apply traditional analyses of random samples of 'individuals' to the estimation of relative abundances of different protein types. In particular, standard rarefaction of peptide frequency data will be biased: assuming a similar distribution in the size and ionization efficiency of peptides from each protein, the frequency of large proteins will be overestimated and the frequency of small proteins will be underestimated. However, using the number of amino acids in each identified protein, a simple modification to rarefaction is possible that will help to correct such biases (Figure 3).

Proteomics data, similar to biodiversity data, are increasingly available over the Internet from institutional data repositories, such as the Proteomics Identifications database (PRIDE; <http://www.ebi.ac.uk/pride>). However, unlike ecological data, proteomic data come in only a few, highly standardized forms (mass-spectroscopy spectra and protein identification supported by peptide identifications) that are routinely (and mandatorily) submitted to data archives as part of the manuscript submission and publication process. Proteomic data repositories require detailed metadata about the origins, processing and analysis of samples, and they have controlled vocabularies (ontologies) for annotating data and metadata that can be searched with predefined computational algorithms and structured work

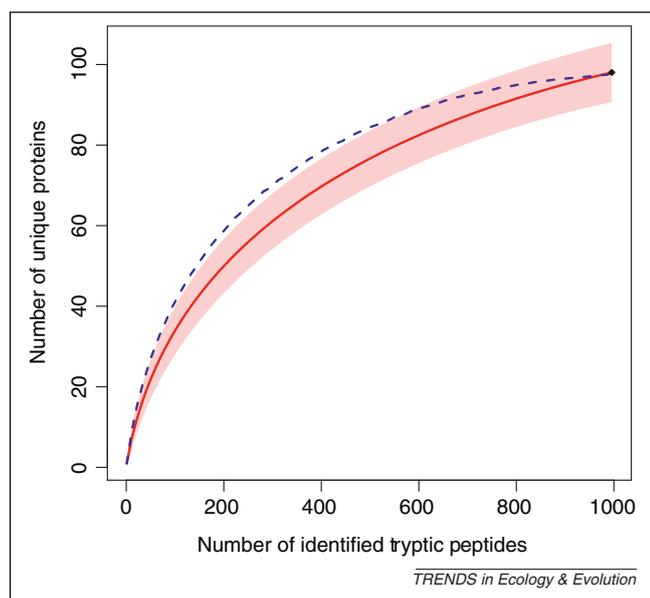


Figure 3. Proteomic sampling curve for the pitcher-plant mosquito larva *Wyeomyia smithii*. The unbroken red line is a rarefaction curve illustrating the number of unique proteins identified as a function of the number of identified tryptic peptides. A total of 995 tryptic peptides were mapped to 98 unique proteins (closed circle). The rarefaction curve gives the expected number of proteins for a smaller random subsample of tryptic peptides. The shaded red funnel is a 95% confidence interval, based on 1000 replicates. The standard rarefaction analysis assumes that all tryptic peptides have an equal probability of detection. However, even if protein abundances were equal, proteins with high amino acid numbers would be over-represented in this analysis because they would be more likely to be cut into several tryptic peptides than would proteins with low amino acid number. To adjust for these differences, the broken blue line illustrates the expected rarefaction curve in which the probability of sampling a protein is proportional to $1/W$, where W is the number of amino acids that comprise it. Amino acid numbers in this data set ranged from 217 (Nacalpa-PB_FBgn0086904) to 8648 (bt-PC_FBgn0005666). This analysis suggests that protein diversity is higher than would be estimated from a standard rarefaction sampling curve. Data modified from Gotelli *et al.* [19].

flows [47]. Thus, these data repositories provide an additional, as yet untapped, resource for exploring patterns of proteomic diversity.

In summary, there are remarkable similarities between the biodiversity data and environmental and community proteomics data, in the constraints and challenges of sampling, in the form of the data that result from such sampling, and in the availability of archived data and work flows for analyzing them. With adjustments to account for bias in protein size, the statistical framework developed by ecologists to characterize and summarize biological diversity will be applicable to protein diversity data.

Experimental environmental proteomics

Summary metrics and statistical analyses are only the first step forward with proteomic data. We are more interested in how proteins interact in an assemblage of organisms or in a complete ecosystem, and expect that an understanding of these interactions will provide new insights into the workings of ecological processes. As noted earlier, the sheer number of proteins in even a single tissue suggests that trying to identify each and every function for each and every protein in an ecological system will be a long time coming, and might not be informative for ecologists working with typical food webs of macrobes and microbes. Mechanistic studies of food webs have grappled with the analogous question: is there a need to know the identity and role of every species in a food web, or can the inherent complexity of food webs be reduced by grouping species into trophic categories (consumers, detritivores and predators), functional feeding groups (shredders, scrapers and filter-feeders), or taxonomic and/or ecological guilds (tube-building polychaetes or seed-eating finches) that still provide robust insights into mechanisms driving ecological dynamics? When these simplified groupings have been organized into abstracted webs that have few nodes and links [48], they have proven to be more amenable to mechanistic modeling [49]. The predictions of these simplified models have been tested directly by adding or removing species at different trophic levels [50], manipulating nutrients or basal resources [51], and quantifying the responses of the rest of the assemblage to such perturbations [52].

A similar approach might profitably be used to understand the functional significance of proteins in ecological systems. This would be achieved most easily by grouping proteins (or their homologs) by class using gene ontology databases in which proteins are classified into functional categories based on similar biological processes, biochemical activities and subcellular localization [53,54]. Indeed, one of the key reasons that proteomics might yield new ecological insights is that many species, particularly of microbes, might be functionally equivalent because they produce similar proteins with similar functions in similar environments [55,56]. This kind of ecological redundancy is a key concept in the study of ecosystem function [57] that already allows ecologists to simplify the vast diversity that characterizes natural assemblages [58].

Thus, rather than trying to characterize all of the uncommon proteins in an assemblage, ecologists should

first concentrate on the small number of identified proteins that change in abundance either from common to rare or rare to common in response to experimental manipulations. By adding or removing species and species groups, ecologists will discover which kinds of protein are involved in the response. This molecular phenotyping [17] of entire assemblages should provide insight into how these proteins function, and repeated sampling might also reveal temporal dynamics. Molecular phenotyping and recognizing the functions of proteins do not require that functionally important proteins be common in natural assemblages. When proteins that are absent or rare in control assemblages become abundant in response to experimental perturbations, there is good reason to suspect that they have an important function in the ecosystem.

Moreover, as recombinant DNA technology and protein expression systems become widely available to ecologists, entirely new avenues of ecological experiments will become possible, at least for secreted proteins. Rather than manipulating species or trophic groups, ecologists will be able to add synthesized proteins directly to experimental mesocosms. There is a long tradition of experimentally (or unintentionally) enriching food webs with nitrogen, phosphorous and other crucial nutrients that limit the growth of phytoplankton, algae and terrestrial plants [59,60]. This idea can be extended to enriching ecosystems with potentially important proteins and metabolites, and then measuring food-web responses. Measuring the ecological proteome in response to species removals and additions, and measuring the biodiversity response to the addition of synthetic proteins should lead ecologists to understand the functions of proteins in ecosystem organization.

Of course, these kinds of experiment will generate new statistical challenges. For example, if 500 proteins are surveyed in an ecological field experiment with a control and a predator removal treatment, a traditional frequentist analysis would potentially generate 500 *t*-tests and associated *P*-values. Even if there are no effects of the top predator on the ecological proteome, at least some of these comparisons will be statistically significant by chance alone [61]. However, promising new analyses based on empirical Bayesian [62] and other approaches [63] have been effective in screening large data sets, such as those generated by proteomic surveys [64], microarrays [65] and species occurrence matrices [66].

Concluding remarks

Pioneering studies of environmental proteomics have recently revealed the functional significance of proteins in simple biofilm communities. However, it is not clear how well the results can be generalized to more typical communities of interacting microbes and macrobes in structurally complex environments. Technological advances and improvements in bioinformatics and data archiving will allow for proteomic characterization of a variety of terrestrial and aquatic habitats. However, protein diversity is vast, and there will always be undetected rare proteins in biological material. Rather than trying to sample exhaustively this rare tail of protein diversity, ecologists should concentrate on measuring more common

proteins in replicated samples and quantifying the variability inherent in natural food webs.

If the relative abundances of different protein types from a biological sample can be estimated, the resulting data are remarkably similar in form to counts of individuals and species that already are familiar to ecologists who analyze patterns of biodiversity. For such analyses, a statistical tool box of rarefaction methods for interpolation, and asymptotic estimators for extrapolation can be applied to summarize protein diversity data effectively and will allow ecologists to standardize samples for meaningful comparisons.

However, as the history of food-web studies has illustrated, statistical methods alone usually cannot reveal important functional roles. Therefore, ecologists should begin using proteomics in combination with traditional experimental field manipulations, such as nutrient enrichment, species additions and removals, and the modification of habitat complexity. In such experiments, proteins that change status from 'common' to 'rare' or 'rare' to 'common' are good candidates for functional importance to food-web structure, although in some cases they might simply mirror changes in the relative abundance of constituent species. More innovative experiments would involve the addition to food webs of synthetic proteins first identified by proteomic analysis.

Although there are still serious methodological challenges to extracting proteins from complex media such as soil and seawater, proteomics will eventually become an important tool for community and ecosystem ecologists. The analysis of environmental proteomic data, in combination with statistical biodiversity methods and experimental field manipulations, has the potential to increase greatly the understanding of the role of proteins in the organization and function of food webs and ecosystems.

Acknowledgments

We thank Rachel Brooks for generating the data in Figure 1. NJG was supported by the US National Science Foundation (US-NSF; DEB-136703 and DEB-0541936) and the US Department of Energy (US-DoE; DE-FG02-08ER64510), and a University of Vermont Research Award to NJG and BAB. AME was supported by grants from the US-NSF (DEB-136703, DEB-0541680, DEB 06-20443, and DBI-1003938), and the US-DoE (DE-FG02-08ER64510). BAB was also supported by the US-NSF (DEB-136703) and the Vermont Genetics Network through US National Institutes of Health grant P20 RR16462 from the INBRE Program of the NCRR.

References

- 1 Maron, P.A. *et al.* (2007) Metaproteomics: a new approach for studying functional microbial ecology. *Microb. Ecol.* 53, 486–493
- 2 Schneider, T. and Riedel, K. (2010) Environmental proteomics: analysis of structure and function of microbial communities. *Proteomics* 10, 785–798
- 3 Vandenkoornhuys, P. *et al.* (2010) Integration of molecular functions at the ecosystemic level: breakthroughs and future goals of environmental genomics and post-genomics. *Ecol. Lett.* 13, 776–791
- 4 VerBerkmoes, N.C. *et al.* (2009) Functional analysis of natural microbial consortia using community proteomics. *Nat. Rev. Microbiol.* 7, 196–205
- 5 Ram, R.J. *et al.* (2005) Community proteomics of a natural microbial biofilm. *Science* 308, 1915–1920
- 6 Delmotte, N. *et al.* (2009) Community proteogenomics reveals insights into the physiology of phyllosphere bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 106, 16428–16433
- 7 Kan, J. *et al.* (2005) Metaproteomic analysis of Chesapeake Bay microbial communities. *Saline Syst.* 1, 7

- 8 Schulze, W.Z. *et al.* (2005) A proteomics fingerprint of dissolved organic carbon and of soil particles. *Oecologia* 142, 335–343
- 9 Nesatyy, V.J. and Suter, M.J.F. (2007) Proteomics for the analysis of environmental stress responses in organisms. *Environ. Sci. Technol.* 41, 6891–6900
- 10 Tomanek, L. and Zuzov, M.J. (2010) The proteomic response of the mussel congeners *Mytilus galloprovincialis* and *M. trossulus* to acute heat stress: implications for thermal tolerance limits and metabolic costs of thermal stress. *J. Exp. Biol.* 213, 3559–3574
- 11 Vaitkevicius, K. *et al.* (2006) A *Vibrio cholerae* protease needed for killing of *Caenorhabditis elegans* has a role in protection from natural predator grazing. *Proc. Natl. Acad. Sci. U.S.A.* 103, 9280–9285
- 12 Banfield, J.F. *et al.* (2005) Proteogenomic approaches for the molecular characterization of natural microbial communities. *OMICS* 9, 301–333
- 13 Schultze, W. (2004) Environmental proteomics – what proteins from soil and surface water can tell us: a perspective. *Biogeosci. Discuss.* 1, 195–218
- 14 Grossert, H.-P. (2010) Ecological consequences of bacterioplankton lifestyles: changes in concepts are needed. *Environ. Microbiol. Rep.* 2, 706–714
- 15 Moore, J.C. *et al.* (2004) Detritus, trophic dynamics and biodiversity. *Ecol. Lett.* 7, 584–600
- 16 Wilmes, P. and Bond, P.L. (2006) Metaproteomics: studying functional gene expression in microbial ecosystems. *Trends Microbiol.* 14, 92–97
- 17 Biron, D.G. *et al.* (2006) Population proteomics: an emerging discipline to study metapopulation ecology. *Proteomics* 6, 1712–1715
- 18 Gotelli, N.J. and Ellison, A.M. (2006) Food-web models predict species abundances in response to habitat change. *PLoS Biol.* 4, 1869–1873
- 19 Gotelli, N.J. *et al.* (2011) Proteomic characterization of the major arthropod associates of the carnivorous pitcher plant *Sarracenia purpurea*. *Proteomics* 11, 2354–2358
- 20 Beck, M. *et al.* (2011) Comprehensive proteomics. *Curr. Opin. Biotechnol.* 22, 3–8
- 21 Thakur, S.S. *et al.* (2011) Deep and highly sensitive proteome coverage by LC-MS/MS without prefractionation. *Mol. Cell. Proteomics* 10, DOI: 10.1074/mcp.M110.003699
- 22 Wilm, M. (2009) Quantitative proteomics in biological research. *Proteomics* 9, 4590–4605
- 23 Walther, T.C. and Mann, M. (2010) Mass spectrometry-based proteomics in cell biology. *J. Cell Biol.* 190, 491–500
- 24 Groen, A.J. and Lilly, K.S. (2010) Proteomics of total membranes and subcellular membranes. *Expert Rev. Proteomics* 7, 867–878
- 25 Goswami, T. and Ballif, B.A. (2011) Methods for the isolation of phosphoproteins and phosphopeptides for mass spectrometry analysis: toward increased functional phosphoproteomics. In *Sample Preparation in Biological Mass Spectrometry* (Ivanov, A. and Lazarev, A., eds), pp. 627–655, Springer
- 26 Chen, C. *et al.* (2011) Protein bioinformatics databases and resources. *Methods Mol. Biol.* 694, 3–24
- 27 Jacob, R.J. (2010) Bioinformatics for LC-MS/MS-based proteomics. *Methods Mol. Biol.* 658, 61–91
- 28 Jost, L. (2007) Partitioning diversity into independent alpha and beta components. *Ecology* 88, 2427–2439
- 29 Digel, C. *et al.* (2011) Body sizes, cumulative and allometric degree distributions across natural food webs. *Oikos* 120, 503–509
- 30 Rosenzweig, M. (1995) *Species Diversity in Space and Time*, Cambridge University Press
- 31 Magurran, A.E. and McGill, B.J., eds (2011) *Biological Diversity: Frontiers in Measurement and Assessment*, Oxford University Press
- 32 Cohen, J.E. (1978) *Food Webs and Niche Space*, Princeton University Press
- 33 Pascual, M. and Dunne, J.A., eds (2005) *Ecological Networks: Linking Structure to Dynamics in Food Webs*, Oxford University Press
- 34 Hughes, J.B. *et al.* (2001) Counting the uncountable: statistical approaches to estimating microbial diversity. *Appl. Environ. Microbiol.* 67, 4399–4406
- 35 Longino, J.T. *et al.* (2002) The ant fauna of a tropical rain forest: estimating species richness three different ways. *Ecology* 83, 689–702
- 36 Van Straalen, N.I. and Roelofs, D. (2006) *Introduction to Ecological Genetics*, Oxford University Press
- 37 Shen, Y.F. *et al.* (2004) Ultra-high-efficiency strong cation exchange LC/RPLC/MS/MS for high dynamic range characterization of the human plasma proteome. *Anal. Chem.* 76, 1134–1144
- 38 Guercio, R.A.P. *et al.* (2006) Ontogenetic variations in the venom proteome of the Amazonian snake *Bothrops atrox*. *Proteome Sci.* 4, DOI: 10.1186/1477-5956-4-11
- 39 Lundgren, D.H. *et al.* (2010) Role of spectral counting in quantitative proteomics. *Expert Rev. Proteomics* 7, 39–53
- 40 McGill, B.J. (2011) Species abundance distributions. In *Biological Diversity: Frontiers In Measurement & Assessment* (Magurran, A.E. and McGill, B.J., eds), pp. 105–122, Oxford University Press
- 41 Connolly, S.R. *et al.* (2009) Testing species abundance models: a new bootstrap approach applied to Indo-Pacific coral reefs. *Ecology* 90, 3138–3149
- 42 Gotelli, N.J. and Colwell, R.K. (2011) Estimating species richness. In *Biological Diversity: Frontiers In Measurement & Assessment* (Magurran, A.E. and McGill, B.J., eds), pp. 39–54, Oxford University Press
- 43 Colwell, R.K. and Coddington, J.A. (1994) Estimating terrestrial biodiversity through extrapolation. *Philos. Trans. R. Soc. Lond. Ser. B: Biol. Sci.* 345, 101–118
- 44 Chao, A. *et al.* (2009) Sufficient sampling for asymptotic minimum species richness estimators. *Ecology* 90, 1125–1133
- 45 Colwell, R.K. *et al.* (2012) Models and estimators linking individual-based and sample-based rarefaction, extrapolation, and comparison of assemblages. *J. Plant Ecol.* 5, 3–21
- 46 Koziol, J.A. *et al.* (2006) Application of capture–recapture models to estimation of protein count in MudPIT experiments. *Anal. Chem.* 78, 3203–3207
- 47 Michener, W.K. and Jones, M.B. (2012) Ecoinformatics: supporting ecology as a data-intensive science. *Trends Ecol. Evol.* 27, 85–93
- 48 Holyoak, M. and Sachdev, S. (1998) Omnivory and the stability of simple food webs. *Oecologia* 117, 413–419
- 49 Allesina, S. *et al.* (2008) A general model for food web structure. *Science* 320, 658–661
- 50 Wootton, J.T. (1992) Indirect effects, prey susceptibility, and habitat selection – impacts of birds on limpets and algae. *Ecology* 73, 981–991
- 51 Becks, L. *et al.* (2005) Experimental demonstration of chaos in a microbial food web. *Nature* 435, 1226–1229
- 52 Eveleigh, E.S. *et al.* (2007) Fluctuations in density of an outbreak species drive diversity cascades in food webs. *Proc. Natl. Acad. Sci. U.S.A.* 104, 16976–16981
- 53 Lacerda, C.M.R. *et al.* (2007) Metaproteomic analysis of a bacterial community response to cadmium exposure. *J. Proteome Res.* 6, 1145–1152
- 54 Matsuoka, S. *et al.* (2007) ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science* 316, 1160–1166
- 55 Nagy, R. *et al.* (2005) The characterization of novel mycorrhiza-specific phosphate transporters from *Lycopersicon esculentum* and *Solanum tuberosum* uncovers functional redundancy in symbiotic phosphate transport in solanaceous species. *Plant J.* 42, 236–250
- 56 Schloss, P.D. and Handelsman, J. (2008) A statistical toolbox for metagenomics: assessing functional diversity in microbial communities. *BMC Bioinform.* 9, 34
- 57 Rosenfield, J.S. (2002) Functional redundancy in ecology and conservation. *Oikos* 98, 156–162
- 58 Mouchet, M.A. *et al.* (2010) Functional diversity measures: an overview of their redundancy and their ability to discriminate community assembly rules. *Funct. Ecol.* 24, 867–876
- 59 Hecky, R.E. and Kilham, P. (1988) Nutrient limitation of phytoplankton in fresh-water and marine environments – a review of recent evidence on the effects of enrichment. *Limnol. Oceanogr.* 33, 796–822
- 60 Smith, V.H. *et al.* (1999) Eutrophication: impacts of excess nutrient inputs on freshwater, marine, and terrestrial ecosystems. *Environ. Pollut.* 100, 179–196
- 61 Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate – a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B: Methodol.* 57, 289–300
- 62 Efron, B. (2010) The future of indirect evidence. *Stat. Sci.* 25, 145–157

- 63 Storey, J.D. and Tibshirani, R. (2003) Statistical significance for genome-wide studies. *Proc. Natl. Acad. Sci. U.S.A.* 106, 9440–9445
- 64 Huttlin, E.L. *et al.* (2007) Prediction of error associated with false positive rate determination for peptide identification in large-scale proteomics experiments using a combined reverse and forward peptide sequence database strategy. *J. Proteome Res.* 6, 392–398
- 65 Dudoit, S. *et al.* (2003) Multiple hypothesis testing in microarray experiments. *Stat. Sci.* 18, 71–103
- 66 Gotelli, N.J. and Ulrich, W. (2010) The empirical Bayes approach as a tool to identify non-random species associations. *Oecologia* 162, 463–477
- 67 Durr, E. *et al.* (2004) Direct proteomic mapping of the lung microvascular endothelial cell surface *in vivo* and in cell culture. *Nat. Biotechnol.* 22, 985–992
- 68 Chao, A. (1987) Estimating the population size for capture–recapture data with unequal catchability. *Biometrics* 43, 783–791
- 69 Baker, B.J. and Banfield, J.F. (2003) Microbial communities associated with acid mine drainage. *FEMS Microb. Rev.* 44, 139–152