Elephants in the room: protists and the importance of morphology and behaviour

Patrick J. Keeling, Canadian Institute for Advanced Research, Botany Department, University of British Columbia, 3529-6270 University Boulevard, Vancouver, BC, Canada V6T 1Z4.

A couple of years ago I found out I was not a microbiologist after all. I always thought I was, and even told strangers that is what I did, if they ever asked. But at a meeting of the American Society for Microbiology, I learned that my definition of a 'microbe' was not particularly representative. This is because I work on protists. Protists are microbial eukaryotes (more or less – we cannot quite decide on a definition), they are found in most of the environments you would expect to find other kinds of microbes (which is to say, everywhere), they are abundant, extraordinarily diverse, and (among my friends, anyway) generally considered to be ecologically important. They do come up sometimes in conversation, or even arguments, such as 'who is the most important primary producer?', or 'are viruses or grazers more important for nutrient cycling?'. But protists are too often excluded from microbial ecosystem models or assessments of their composition; even studies that assess a complete 'microbiome' more often than not ignore the microbial eukaryotes.

Before I am written off as a whinging specialist who is feeling marginalized, let me state that there are good reasons for this gap in our knowledge; they reflect interesting reasons that go back to fundamental differences in biology. Indeed, the problems associated with a thorough understanding of microbial eukaryotic ecology are so stark, that my prediction for the next year is not that we will solve these problems, or even make progress. My prediction (or perhaps wishful thinking) is that the 'eukaryotic question' will increasingly emerge as an elephant in the room, which is an elegant idiom to describe our failure to grasp the role of so many large microbes that are right under our noses.

Bigger yes, but also different

I would like to discuss two reasons why protists have not entered the mainstream of conventional high-throughput environmental microbiology. The first of these is trivial and well understood: their genomes are bigger and organized differently. We know that new sequencing technologies have had a major impact in our understanding of the diversity and ecological roles of bacteria, archaea and viruses, for example, by allowing whole-community metagenomic surveys. To include protists in these surveys is easy – simply do not filter them out! However, we also know that nuclear genome sizes would require epic sequencing and analysis budgets that are simply not practical. Moreover, we cannot accrue the same benefits for protists, even if we could sequence enough, because their genomes are fragmented, repeat-rich, and lack functionally related gene clustering, all of which limit the inferences we can make about individual genomes and metabolic networks from metagenomics by limiting our ability to link genes to other genes in a genome.

But there is another less discussed, but infinitely more interesting problem. Bacterial and archaeal diversity is substantially manifested at the level of metabolism. Accordingly, the sequence of a bacterial or archaeal genome can go a long way to describing what that organism 'does' in the community, because we have developed reasonable ways to translate the information in a genome into predictions about that organism's metabolic actions in the environment. This is not the case for eukaryotes: although microbial eukaryotes harbour a sizable metabolic diversity, they are distinguished from other microbial life in that they manifest a great deal more diversity at the levels of *morphology* and *behaviour*. Indeed, morphology and behaviour have a much greater effect on what most protists 'do' in the environment than do their metabolic capacities (photosynthesis being an obvious exception). Unfortunately, the manifestation of these properties is much more complex than a straightforward gene–protein correspondence, and we are accordingly much worse at translating the information in a genome into predictions about what an organism looks like or how it behaves.

To illustrate this problem, imagine four dinoflagellate protists living in the same marine environment: one is a free-living benthic autotroph, one is an intracellular parasite of gastropods, one is an obligate photosynthetic symbiont of cnidarians, and one is a heterotrophic grazer feeding on bacteria and eukaryotic algae. Now imagine we have sequenced whole genomes and whole transcriptomes for all four of these organisms. How easy would it be to reconstruct these interactions? The answer is, it would be virtually impossible, even with these miraculous quantities of molecular data. We could recognize that two were photosynthetic, but this might even mislead us to assume they shared a similar niche, when in reality the two forming intracellular relationships with invertebrates might share more in common. This failure is because the most important characteristics that distinguish these organisms and their activities are derived from poorly understood coordinated actions of thousands of gene products, and worse still, subtleties of regulation and epigenetics relating to thousands of genes.

Organisms DO matter – how do we study them?

They say that if you have a hammer, everything looks like a nail, and right now our biggest hammer is sequencing. Getting more sequence data from eukaryotes at the environmental level is a technical problem that can, and soon will be, solved. The most revolutionary solution will be the arrival of routine single-cell genomics and transcriptomics. Despite all we have learned through metagenomic approaches, cells *do* matter in the final analysis because biological activities are compartmentalized and how the metabolism of a community is partitioned makes a difference; a community is not just the sum of its enzymes, and seeing how functions are distributed across a community will change how we interpret them. Single-cell genomics will therefore be a boon to all environmental microbiology. And for eukaryotes, single-cell transcriptomics in particular will give us a first inroad to their otherwise intractable genomes when it can be automated across natural communities.

How we interpret environmental sequence data from eukaryotes is another problem altogether. If the predictive power of even genome-wide sequence data is critically limited by our inability to infer characteristics of morphology and behaviour from it, then how do we integrate protists into a detailed picture of a microbial community that is primarily based on such data? Certainly being able to predict what an organism is like based on its close relatives will continue to be important, but requires a lot of 'model' systems scattered around the tree of eukaryotes to be truly effective. The real answer likely lies in a re-emergence, and indeed a reinvention, of arts like cultivation, ultrastructural characterization, identification and observation of live cells within their natural community, and field microscopy – some of which are badly under-appreciated at present. Our challenge is therefore not to put away our hammer, but to place more emphasis on the need for other tools too (in fact, I once watched a graduate student hammering a screw, so perhaps there is even greater depth to this need). It is not always obvious how these tools will be as adapted to a high-throughput approach as genomic methods were, but advances in imaging and cell sorting open a host of possibilities. So, to some extent, the way forward involves integrating existing methods rather than inventing new ones (e.g. linking high-throughput imaging with single-cell sorting would allow morphology to be linked with genomic data).

In summary then, it is my hope that in the coming years microbial eukaryotes emerge a bit from the shadows of their smaller cousins. Luring them out into the open will require more than protists simply 'catching up' with existing methods: we must improve the integration of protists with our understanding of other members of microbial communities by coordination and deliberate efforts to reconstruct entire microbiomes, including all members and their interactions. The genomic revolution has allowed astonishing advances, but perhaps this only means that it needs to be grounded in biology more than ever.

Adopting modularity of metabolism as a guiding paradigm may lead to better accounting and understanding of the unseen majority of life: exercised with focus on the nitrogen cycle

Martin G. Klotz, Evolutionary and Genomic Microbiology Laboratory, Department of Biology, The University of North Carolina, Charlotte, NC 28223, USA.

Obligate aerobic, chemolithotrophic and predominantly autotrophic ammonia-oxidizing bacteria ('AOB') cluster within two distant monophyletic groups: the betaproteobacterial family Nitrosomonadaceae and the purple sulfur bacterial genus *Nitrosococcus* of the Gammaproteobacteria. Yet, these two distant groups seemingly live identical catabolic lifestyles, posing challenging evolutionary questions that have awaited answers for several decades. Long generation times of the AOB and their infamous recalcitrance to transformation, as well as cloning and recombinant expression of their genes, have prevented extensive molecular genetic experimentation to verify their catabolic pathways. Thus, the opportunity in 1999 to sequence and annotate the genome of a bacterium once thought to be the ultimate representative for aerobic nitrogen biology created a lot of buzz and expectations; however, it took almost 4 years from the isolation of 'pure enough' genomic DNA to reporting the results (Chain *et al.*, 2003). Aside from the exhilarating experience of finding all the genes necessary to make a living cell and the previously implicated inventory for it being an AOB, little could be gleaned from the genome to answer pressing questions on the evolution of nitrification as a process or the obligate nature of the ammonia-oxidizing lifestyle. This initial genome analysis was soon followed by additional sequencing projects, including other AOB and obligate aerobic chemolithotrophic nitrite-oxidizing bacteria ('NOB'), that were facilitated by the then fully established DOE Joint Genome Institute (JGI) and initially coordinated by a group of Principal Investigators (PIs) supported by funding from the US National Science Foundation for a Research Coordination Network. The outcome of this endeavour was tremendous: Principal Investigators with different interests and expertise as well as at different levels of advancement in their careers came together and witnessed the power of genuine collaboration, which included the immersion of postdocs, graduate and even undergraduate students (http://nitrificationnetwork.org).