environmental microbiology

Environmental Microbiology (2016) **00**(00), 00–00 doi:10.1111/1462-2920.13260

Highlight

Back to the future: function-first metagenomics returns to the fore

Thomas E. Hanson*

School of Marine Science and Policy, Department of Biological Sciences, and Delaware Biotechnology Institute, University of Delaware, Newark, DE 19716, USA.

A paper by Varaljay *et al*. from the Tabita laboratory in this issue of *Environmental Microbiology* reports an approach to retrieve functional genes from uncultivated microbes, in this case genes encoding ribulose-1,5-bisphosphate carboxylase oxygenase (RubisCO), the key enzyme of the Calvin-Benson-Bassham cycle of CO_2 fixation. RubisCO is the most abundant protein on earth and has been extensively studied to try and improve its catalytic properties, particularly the ability to discriminate between its two gaseous substrates CO_2 and O_2.

The approach relies on a strong positive selection by complementation of a mutant strain with cloned environmental DNA and should be broadly generalizable for any enzyme of interest where there exists a suitable mutant strain and complementation system. By emphasizing function first, this approach harkens back to the early days of metagenomics, all the way back to the late 1990s, when large-scale DNA sequencing was slow, expensive, and therefore had to be applied judiciously (Handelsman *et al*., 1998). Metagenomics in those days relied on cloning large pieces of environmental DNA in bacterial artificial chromosomes. Strains carrying these sections of genomes would then be screened for the production of antimicrobial compounds, enzymatic activities, or other functions of interest prior to sequencing. With the advent of next-generation sequencing technologies capable of producing hundreds of billions to trillions of basepairs of sequence information from a single sample for reasonable investments of time and money, function-first approaches have been pushed to the wayside in favor of sequence-first metagenomics.

Received 9 February, 2016; accepted 9 February, 2016. *For correspondence. E-mail: tehanson@UDel.Edu; Tel. 3028313404; Fax 3028313447.

Metagenomics is ultimately concerned with alleviating the cultivation bottleneck to help us understand how microbes operate in nature, where we have cultivated only a small fraction of microbes that can be observed by cultivation-independent approaches (Whitman *et al*., 1998). While sequence-first metagenomics does provide an incredible window into genomes of the uncultivated masses, these approaches are hampered by the fact that we currently cannot reliably assign function by computational annotation to a large fraction of genes predicted in these data sets. In concept, function-first metagenomics by complementation could identify completely novel genes encoding important functions without any *a priori* knowledge of what these genes should look like.

The paper by Varaljay *et al*. makes a good case for this concept. RubisCO is required for the growth of purple nonsulphur phototrophic bacteria under a variety of conditions. Furthermore, prior work in the Tabita laboratory has shown that growth conditions can be used to positively select for RubisCO function in a mutant strain of *Rhodobacter capsulatus* lacking both of its native RubisCO-encoding genes. For example, growth under chemolithoautotrophic conditions is a stringent selection requiring RubisCO to discriminate well between CO_2 and O_2 (Satagopan *et al*., 2009). The selection system employed here combined this *R. capsulatus* mutant strain with a plasmid containing promoters ensuring expression of inserted DNA under appropriate growth conditions irrespective of the orientation of the inserted fragment. DNA isolated from several environments was cloned into the expression vector and moved into the *R. capsulatus* mutant strain, which was subsequently asked to grow under RubisCO-requiring conditions. Cells that did not receive a functional RubisCO-encoding gene should not survive and, indeed, no clones were recovered that did not contain an intact RubisCO gene. Thus, the method reported by Varaljay *et al*. dramatically increases efficiency of relevant clone identification in metagenomics.

RubisCO genes were recovered from each environment tested and were reflective of the RubisCO genes observed in sequence-first metagenomics approaches where this data was available. Active RubisCO enzymes come in

three flavours or forms: I, II and III (Tabita *et al.*, 2007). Genes encoding both form I and II were recovered in this work, all likely of proteobacterial origin. Form III enzymes are only known in the Archaea to date, primarily methanogenic Archaea, and none of the samples examined here were expected to have large numbers of these organisms. One does wonder whether the recovery of proteobacterial RubisCO genes in a proteobacterial host is a possible limitation of the system, but foundational work has shown that *R. capsulatus* can be used to functionally express RubisCO genes from cultivated (Smith and Tabita, 2003) and uncultivated (Witte *et al.*, 2010) Cyanobacteria.

Based on this successful proof-of-principle, this specific system should allow for the recovery and characterization of RubisCO's from a wide array of environments, enabling broader exploration of the catalytic space produced by evolution and natural selection for this enzyme. Ultimately, this should enable a decoding of biological design principles for efficient RubisCO catalysis that more focused approaches have made limited progress on to date. Beyond RubisCO, function-first selection-based metagenomics could identify new genes and enzymes from nature for amino acid biosynthesis, bioremediation of organic pollutants, biomass deconstruction, escaping phage infection, and stress resistance; essentially any biological function where a strong positive selection can be established. This will require the continued development of non-traditional model systems to host relevant selections, which will ultimately increase our understanding of microbial metabolic diversity writ large.

References

Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J., and Goodman, R.M. (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* **5:** R245–R249.

Satagopan, S., Scott, S.S., Smith, T.G., and Tabita, F.R. (2009) A Rubisco mutant that confers growth under a normally 'inhibitory' oxygen concentration. *Biochemistry* **48:** 9076–9083.

Smith, S.A., and Tabita, F.R. (2003) Positive and negative selection of mutant forms of prokaryotic (cyanobacterial) ribulose-1,5-bisphosphate carboxylase/oxygenase. *J Mol Biol* **331:** 557–569.

Tabita, F.R., Hanson, T.E., Li, H., Satagopan, S., Singh, J., and Chan, S. (2007) Function, structure, and evolution of the RubisCO-like proteins and their RubisCO homologs. *Microbiol Mol Biol Rev* **71:** 576–599.

Whitman, W.B., Coleman, D.C., and Wiebe, W.J. (1998) Prokaryotes: the unseen majority. *Proc Natl Acad Sci USA* **95:** 6578–6583.

Witte, B., John, D., Wawrik, B., Paul, J.H., Dayan, D., and Tabita, F.R. (2010) Functional prokaryotic RubisCO from an oceanic metagenomic library. *Appl Environ Microbiol* **76:** 2997–3003.