20 YEARS & COUNTING

THE CONTINUING EVOLUTION OF THE MOLECULAR LIMNOLOGIST'S TOOLBOX

by **Steven W. Wilhelm, Helena L. Pound, and Robbie M. Martin**

It has been two decades since the tools of modern molecular biology—primarily DNA and RNA sequencing—have been commonly available to limnologists. As DNA sequences provide the blueprint for all living organisms and many viruses, researchers can use sequence information to determine which organisms are present in a sample and the metabolic potential these organisms may have. From viruses (Stough et al. 2017) to fish (Thomsen and Willerslev 2015), and whether in sediments (DeBruyn et al. 2009) or open waters (Steffen et al. 2017), these molecular tools can be employed in the assessment of any biotic community, its members, or its processes. With careful interpretation, this information has incredible power and still unrealized potential.

Today, DNA-sequence-based molecular methods generally fall into one of two broad categories. **Targeted** approaches typically use the **polymerase chain reaction** to amplify marker genes whose sequences contain key information about the identity or functional potential of an organism or community. A classic example includes the 16S rRNA gene as it is highly conserved among bacteria and archaea; sequencing this gene gives researchers an idea of which prokaryotic species are present in a microbial community. Alternatively, **shotgun** approaches attempt to sequence all genes in a sample to assess that community's functional potential. DNA sequencing (i.e., **genomics**) can provide a blueprint that indicates what cells *may* try to do. This blueprint helps researchers to predict possible responses of a microbial community to perturbations in its environment. On the other hand, RNA sequencing (i.e., **transcriptomics** for single species or **meta**transcriptomics for an entire community) can reveal what cells *are* trying to do. Given the rapid turnover of RNA in cells, researchers can view actual community responses that occur in minutes to hours.

Sequencing technology has evolved at a rapid pace during the last two decades. While typical sequencing in the 1990s commonly topped out at 50–100 sequences per sample, shotgun sequencing can now routinely generate billions of individual DNA or RNA **reads** per project,

continued

Definitions and jargon for modern molecular ecology in lakes

Targeted: any methods that address a specific gene (or set of genes) in a sample

Polymerase chain reaction (PCR): a method in molecular biology to make thousands to millions of copies of a specific DNA segment

Shotgun: any methods that address random DNA or RNA targets

Genomics: the field of biology focusing on the structure, function, evolution, mapping, and editing of genomes

Transcriptomics: the field of biology focusing on the examination of whole RNA molecules expressed by organisms across biological conditions

Meta-XXX: the prefix "meta" describes approaches used in mixed populations or samples

Reads (aka sequences; aka transcripts in the context of RNA): specific copies of DNA or RNA that have been analyzed using sequencing approaches

Bioinformatics: methods and software tools for understanding biological data, or an interdisciplinary field of science that combines biology, computer science, information engineering, mathematics, and statistics to analyze and interpret biological data (see examples below)

- Assembly: aligning and merging short nucleic acid fragments to form a longer sequence in an effort to reconstruct the original sequence
- Annotation: identifying the locations of genes and coding regions in a nucleic acid sequence to determine what those genes do
- Recruitment: matching short reads from sequencing efforts to longer assemblies or genomes to determine the relative occurrence of specific sequences in a sample

Proteomics: the large-scale study of the protein complement of an organism or community of organism

Metabolomics: the large-scale study of small molecules, commonly known as metabolites, within cells, tissues, or organisms

with 30–50 million reads being common for every sample (e.g., Tang et al. 2018). This deeper exploration of the genetic content that modern molecular techniques enable is particularly pertinent to studies of the biogeochemical processes in lakes. Bacteria are a major driver of these processes, and, as they persist ubiquitously at abundances of 105 to 106 per mL of lake water (DeBruyn et al. 2004), their abundance and distribution make them excellent targets for molecular assessment as sufficient bacterial material can be collected for community analyses from just a few cups of water. One example of such a molecular assessment is the estimation of the quantity of toxin-encoding genes in a cyanobacterial bloom (pictured). This metric provides a relative measure of the bloom's toxin potential and can be used in conjunction with nutrient and temperature data to help biologists and water managers understand when high toxin loads might occur in a given water body (Rinta-Kanto et al. 2009). Increased ability to analyze genetic content using molecular tools, accompanied by parallel advances in **bioinformatics**, allows for the simultaneous characterization of nearly all biological processes for each member of a microbial community. By analyzing samples collected to have spatial and temporal variability, researchers can tease apart how the millions of microbes in every milliliter of water work together to shape lake function.

While modern molecular techniques allow assessment of much about the microbial community, we note that some of these approaches are only semi-quantitative. Shotgun sequencing, for example, provides information in terms of relative, not absolute, numbers in a sample; as such, it allows for powerful comparisons of shifts in function or community member activity, but it does not provide complete quantification. However, targeted approaches performed for an entire community can allow for absolute quantification. Evolving technological adaptations can further improve the quantification precision of these approaches.

Both technological adaptations of existing molecular tools as well as new molecular technologies are on the horizon. Examples of technological adaptations include digital PCR, which is a more precise and sensitive alternative to conventional and real-time quantitative PCR, and **proteomics**, which uses liquid chromatography coupled to sophisticated mass spectrometry to identify specific proteins in a sample. Interpretation of proteomics data can provide for a robust, semi-quantitative assessment of proteins, which are both longer-lived and, in many cases, more indicative of actual metabolic function than are RNA transcripts. In contrast to these existing tools, **metabolomics** is a still-emerging molecular technology that assesses small molecules in a biological sample and uses this information to infer the active biochemical pathways within communities. Metabolomics can be particularly powerful for detecting major perturbations in a system, but, as



Photo courtesy Steven W. Wilhelm

A toxic cyanobacterial bloom near the Taihu Laboratory for Lake Ecosystem Research in China. Microcystis blooms have become a global problem in recent decades, and the tools of the modern molecular limnologist are shedding light on the factors that promote and constrain such events.

homeostasis is something all cells work toward (Steffen et al. 2015), minor changes in a community may not be detected by this method.

Ultimately, how modern limnologists use molecular tools is a function of both the research question being asked and the limitations of each tool. DNA-based approaches can only be used to interpret community potential, while RNA-based and protein-based assessments indicate actual activity. Metabolomics is an appealing approach to investigating large-scale system disruptions, but any application of this tool must consider that it may not be able to detect minor system disturbances. Finally, the style applied to many of these approaches matters: while a shotgun-styled approach allows for assessment of a broad spectrum of the community, it produces results that are generally only relative in nature; in contrast, absolute quantification is achievable with targeted approaches, but the necessarily narrower scope of targeted assessments can make them less than desirable for some studies. Ultimately, and in spite of their limitations, the tools of the modern molecular biologist hold great promise for limnologists in the future.

Helena Pound is a doctoral student and Robbie Martin a postdoctoral fellow in the Department of Microbiology at the University of Tennessee. Steven Wilhelm is the Kenneth & Blaire Mossman Professor at the University of Tennessee and a faculty member in the Department of Microbiology. The Wilhelm lab has had an active research program with interests in viruses, bacteria, and harmful algal blooms in large lakes around the world since 1998.