

PRIMER

Streams of data from drops of water: 21st century molecular microbial ecology

Dave R. Clark | Robert M. W. Ferguson | Danielle N. Harris | Kirsty J. Matthews Nicholass | Hannah J. Prentice | Kate C. Randall | Luli Randell | Scott L. Warren | Alex J. Dumbrell

School of Biological Sciences, University of Essex, Wivenhoe Park, Colchester, Essex, CO4 3SQ, UK

Correspondence

Dave R. Clark, School of Biological Sciences, University of Essex, Wivenhoe Park, Colchester, Essex, CO4 3SQ, UK

Email: dclarkb@essex.ac.uk

Alex J. Dumbrell, School of Biological Sciences, University of Essex, Wivenhoe Park, Colchester, Essex, CO4 3SQ, UK.

Email: adumbrell@essex.ac.uk.

Funding information

University of Essex, School of Biological Sciences; Natural Environment Research Council, Grant/Award numbers: E15EE, W86713B, 471757, NE/M010813/1, NE/M02086X/1

Microorganisms are ubiquitous and represent a taxonomically and functionally diverse component of freshwater environments of significant ecological importance. The bacteria, archaea, and microbial eukarya in freshwater systems support a range of ecosystem processes and functions, including mediating all major biogeochemical cycles, and therefore regulate the flow of multiple ecosystem services. Yet relative to conspicuous higher taxa, microbial ecology remains poorly understood. As the anthropocene progresses, the demand for freshwater–ecosystem services is both increasing with growing human population density, and by association, increasingly threatened from multiple and often interacting stressors, such as climate change, eutrophication, and chemical pollution. Thus, it is imperative to understand the ecology of microorganisms and their functional role in freshwater ecosystems if we are to manage the future of these environments effectively. To do this, researchers have developed a vast array of molecular tools that can illuminate the diversity, composition, and activity of microbial communities. Within this primer, we discuss the history of molecular approaches in microbial ecology, and highlight the scope of questions that these methods enable researchers to address. Using some recent case studies, we describe some exemplar research into the microbial ecology of freshwater systems, and emphasize how molecular methods can provide novel ecological insights. Finally, we detail some promising developments within this research field, and how these might shape the future research landscape of freshwater microbial ecology.

This article is categorized under:

Water and Life > Methods

Water and Life > Nature of Freshwater Ecosystems

KEYWORDS

Archaea, Bacteria, Eukarya, Metagenetics, Metagenomics, Next-generation sequencing, Omics methods

1 | INTRODUCTION

Microbes are ubiquitous and the most diverse organisms on Earth. There are estimated to be more than 10^{30} microbial cells globally (Whitman, Coleman, & Wiebe, 1998), comprising several billion tons of carbon (Kallmeyer, Pockalny, Adhikari,

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2018 The Authors. *WIREs Water* published by Wiley Periodicals, Inc.

WIREs Water. 2018;e1280.

<https://doi.org/10.1002/wat2.1280>

wires.wiley.com/water

1 of 10

Smith, & D'Hondt, 2012) from more than a trillion different species (Locey & Lennon, 2016), many of which define the limits of life, and often occupy habitats originally considered lifeless. Arguably, microbes are ecologically the most important organisms, known to support a range of ecosystem processes and functions, alongside driving all major biogeochemical cycles, and thus, having a disproportionately large impact on ecosystem services (Falkowski, Fenchel, & Delong, 2008). Therefore, without understanding the diversity, composition, functionality, and activity of microbial communities a holistic view of ecosystems is unattainable.

Freshwater ecosystems are crucial to the persistence of human civilization, yet are among the most sensitive to environmental change (Dudgeon et al., 2006). Furthermore, as conduits between terrestrial and marine environments, freshwater ecosystems can be significantly altered by environmental changes in either of these aforementioned biomes. Globally increasing human population densities and associated environmental changes (e.g., climate change, eutrophication, and chemical pollution) will place a potentially unsustainable demand on freshwater–ecosystem services. Many of these services are strongly influenced by the activity and functionality of microorganisms, which ultimately determine the quality and suitability of freshwater for different human uses (Chapin et al., 1997). Thus, if we are to sustainably manage valuable ecosystem services, such as the basic provision of freshwater, understanding the ecology of freshwater microbial communities is essential. However, the study of microbial communities presents several challenges.

The fundamental need to quantify the diversity, abundance, and activity of organisms is logistically challenging for microbial ecologists working in natural ecosystems, as their study species are difficult to observe due to small sizes, difficult to count due to enormous abundances, and difficult to identify due to a lack of distinguishing morphological features. To meet these challenges, multiple molecular methods—based around studying macromolecules (primarily nucleic acids)—have been developed and are now routinely applied to the study of microbial ecology in freshwater–ecosystems. Recent advances have focused on the development of increasingly high-resolution and high-throughput approaches, arming microbial ecologists with a vast array of molecular tools. However, for the uninitiated, this toolset, the rapid turnover of molecular technologies, and the array of accompanying bioinformatic methods required to analyze these data, can appear bewildering and the associated literature difficult to penetrate.

This paper provides a basic primer for understanding molecular microbial ecology, with a specific focus on quantifying the microbial taxonomic and functional diversity present in environmental samples (e.g., sediments, biofilms, water, etc.) collected from freshwater ecosystems. First, we provide a brief history of these molecular methods, highlighting how successive development of molecular tools has overcome previous limitations. Second, we review some exemplar research that applies these tools, with a particular focus on how they have provided novel insights into freshwater ecosystems. Finally, we scan the research horizon and examine current and future developments with the potential to further advance freshwater molecular microbial ecology.

2 | A BRIEF HISTORY OF MICROBIAL ECOLOGY

Techniques used to study microbial diversity can be classified into two broad categories: culture-dependent and culture-independent. The invention of the microscope in the mid-17th century facilitated the first observations of the previously unseen microbial world (Lane, 2015). Culture-dependent methods developed rapidly thereafter, focusing on isolating microbes from the environment, observing them under the microscope and conducting biochemical tests (Brock, 1961; Caumette, Bertrand, & Normand, 2015). These approaches still provide valuable information on growth rates, metabolic pathways, nutrient demands and optimal growth conditions (Giovannoni & Stingl, 2007), which could not be gained from culture-independent methods. Culturing is therefore still an important tool in microbiology, and when used in combination with culture-independent molecular methods, can shed light on the ecology of previously unknown microorganisms (Bomar, Maltz, Colston, & Graf, 2011). However, as a standalone method, the ecological application of culturing is limited. It is often impossible to recreate the complexity of the natural environment in the laboratory, and many microorganisms have highly specific growth conditions. Consequently, most microorganisms are not readily culturable (see “*the great plate count anomaly*”) (Staley & Konopka, 1985) and so, *in situ* culture-independent molecular methods are necessary to study environmental microbial communities.

Culture-independent techniques bypass the need to isolate and culture microbes from collected environmental samples. The majority of these techniques begin with the extraction of molecular biomarkers; nowadays almost exclusively nucleic acids (DNA/RNA) contained within cells present in those samples. Once purified, a “genetic soup” comprising the genes, genomes, transcripts, and transcriptomes, from all organisms whose cells (or extracellular fragments of DNA/RNA) were present in that sample remains. A range of molecular techniques can now be applied to these samples, the choice of which is primarily dependent upon the research questions being addressed and the temporal and fiscal resources available. Frequently, such approaches target DNA, allowing researchers to study the taxonomic and functional diversity of a sample, regardless of whether the DNA originated from dead, alive, metabolically active, or dormant organisms. Alternatively, RNA can be used,

providing a clearer picture of the taxonomic and functional diversity of the metabolically active component of microbial communities (Griffiths, Whiteley, O'Donnell, & Bailey, 2000). Although this insight is advantageous, there are often significant logistical challenges associated with RNA-based approaches linked to its instability and reduced extraction efficiency from samples (Carvalhais, Dennis, Tyson, & Schenk, 2012), making them far less commonly used than DNA-based methods.

The next step in sample processing involves the amplification of the extracted nucleic acids via the polymerase chain reaction (PCR) (Hugenholtz, 2002), a technique that has revolutionized molecular biology. PCR enables the targeting (via target-specific amplification primers) of genomic/transcriptomic regions (often genes) of interest to be isolated from the “genetic soup” and amplified, generating millions of copies (amplicons) of that region to facilitate subsequent analyses. The choice of the amplicon target depends entirely on the research questions being addressed, but often within studies of freshwater microbial communities this will focus on genetic markers that provide information on the evolutionary identity of the organisms (phylogenetic markers—e.g., 16S or 18S rRNA genes) or their functional capabilities (functional markers—e.g., *amoA* genes targeting ammonia oxidizers, or *mcrA* genes targeting methanogens). Additionally, with appropriate primer design, it is possible to target specific microbial taxa, making PCR approaches an effective method to screen water samples for the presence of pathogenic microorganisms (Bernhard & Field, 2000). With a slight modification of the PCR technique used to that of real-time quantitative PCR (qPCR), the quantification of the gene copy numbers as they are amplified can be achieved, thus, providing an additional and complementary enumeration of taxonomic or functional groups of microorganisms from complex communities (Lansdown et al., 2016). Regardless of whether qPCR has been used as well, the post-PCR sample from a standard PCR will now only contain the target region of interest, but the amplicons present will have come from across all organisms originally present in the sample as long as they have the phylogenetic or functional marker of interest. Thus the sample still comprises a “genetic soup,” but now it is focused on millions of copies of a specific phylogenetic or functional marker.

The challenge in evaluating the microbial diversity (either phylogenetic or functional) is now to separate the “genetic soup” of amplicons into its constituent parts, identifying the species present based on differences across the amplicons in DNA sequence composition. Techniques such as denaturing (or temperature) gradient gel electrophoresis, terminal restriction fragment length polymorphism and automated ribosomal intergenic spacer analysis reflect early solutions to this challenge. These techniques exploit differences in molecular weight or nucleotide composition of amplicons, generating a “fingerprint” comparable between samples (Muyzer & Smalla, 1998), thus enabling quantification of both α - and β -diversity (within, and between sample diversity, respectively). However, they do not easily yield taxonomic identification of the microorganisms present. To do this, the amplicons are separated and the DNA sequenced. Initially, this process involved the insertion of PCR amplicons into plasmid vectors (a small piece of DNA, used to house the amplicon), which solved the issue by single plasmids taking up single amplicons. Transformation of these plasmids (containing the amplicon) into bacterial hosts, usually *Escherichia coli*, would then be used to create a clone library (i.e., each *E. coli* cell takes up one plasmid and thus one amplicon, and multiple *E. coli* colonies are grown effectively separating the amplicons out). *E. coli* colonies that have taken up a vector would then be individually selected for Sanger sequencing. However, this process was costly, time consuming and lacked high-throughput capabilities required for large sample volumes (Leigh, Taylor, & Neufeld, 2015).

The rapid development of sequencing technologies and computational power over the last 20 years has since transformed the field of molecular ecology (Xu, 2006). While the use of clone libraries coupled with the first-generation of sequencing technologies (Sanger sequencing) allowed only a small number of amplicons to be sequenced in a serial manner, Next Generation Sequencing (NGS) technologies now allow millions of amplicons to be sequenced in parallel. A major breakthrough provided by NGS approaches, is that the entire “genetic soup” of PCR amplicons from multiple samples can be pooled together and sequenced in parallel, with the NGS platforms directly separating out the different sequence types present while ensuring the sample they originate from is also recorded—a process commonly termed either metagenetics, metabarcoding, or amplicon sequencing. Metagenetics can be likened to a biodiversity survey (Figure 1); with the identification of species based around matching different DNA sequences to reference collections (DNA sequence databases) and their relative abundances estimated based on the number of identical DNA sequences recorded. These metagenetics approaches can provide almost exhaustive enumeration of the organisms present, allowing characterization of the extensive microbial “rare biosphere” (Shendure & Ji, 2008; Sogin et al., 2006). By targeting specific microbial taxa, metagenetics approaches are also effective at screening for and identifying specific microorganisms, which in the context of freshwater ecosystems, is particularly important for monitoring potential sources of contamination for example, by detecting fecal coliforms from sewage (Bernhard & Field, 2000), or for detecting antimicrobial resistance genes in environmental microbial communities (Taylor, Verner-Jeffreys, & Baker-Austin, 2011). Furthermore, when applied to environmental DNA (*eDNA*; DNA often existing outside of the cell), metagenetics can be used to quickly determine the presence of (micro-, or macro-) organisms that may otherwise elude observation (Thomsen & Willerslev, 2015).

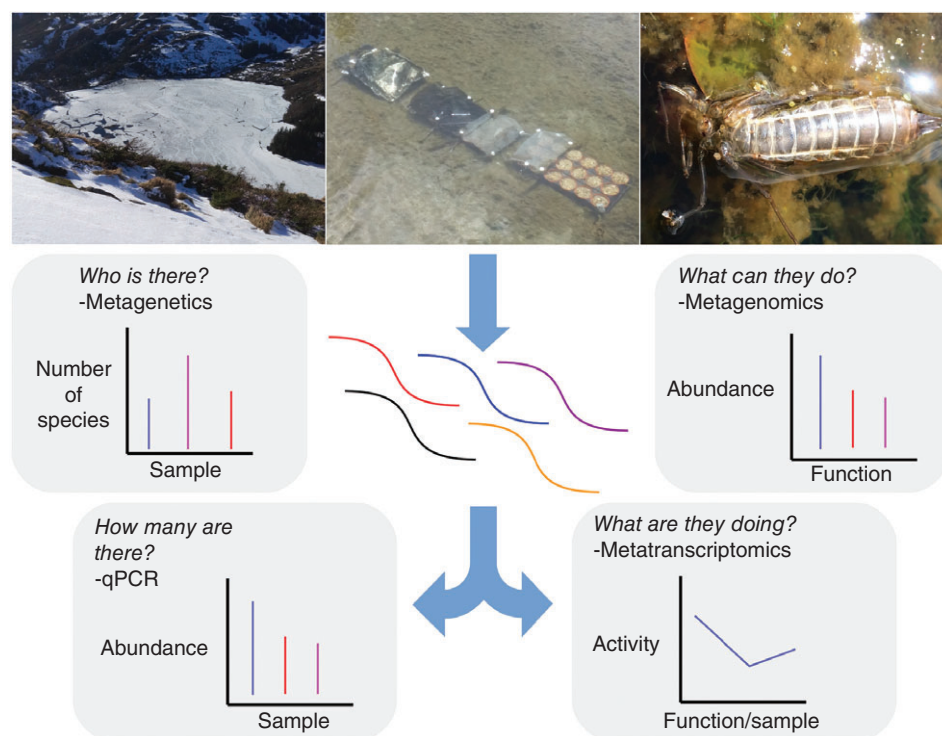


FIGURE 1 A schematic diagram of some of the most frequently used molecular and “omics” methods, along with the types of research questions they are suitable to address. Whether the study is observational, experimental, or routine biomonitoring, all approaches require the extraction and purification of biological molecules, such as DNA or RNA from the environment. Approaches on the left require amplification of specific genes via PCR, whereas approaches on the right do not require a target, and are conducted on the entire pool of nucleic acids

Once sequence data are obtained, rigorous bioinformatic analyses are required to quality control the sequence data, thereby ensuring that ecologically robust conclusions are reached (Dumbrell, Ferguson, & Clark, 2016). After removal of sequencing artifacts and low-quality data, it is common to assign taxonomy to sequence data by comparison to a relevant sequence database. To this end, a number of phylogenetic and functional gene sequence databases are available for the researcher to annotate their sequence data. Phylogenetic databases include the Ribosomal Database Project (Cole et al., 2013), SILVA (Quast et al., 2012), UNITE (Abarenkov et al., 2010), and PR2 (Guillou et al., 2012) databases, all of which vary in the range of taxa archived, marker gene identities, and level of curation and updates. Thus, careful consideration and comparison of such databases is required in order to make the most of phylogenetic sequence data (Werner et al., 2012). In comparison, functional gene databases are comparatively scarce (e.g., FUNgene; Fish et al., 2013), and coverage can vary considerably across different functional genes.

The advent of NGS technologies also provided easy access to a separate approach that is not reliant on the sequencing of PCR amplicons, termed metagenomics. Instead of amplifying phylogenetic or functional markers from a sample of extracted DNA, the DNA comprising genomes from all species present is sheared in to smaller fragments and sequenced directly (Figure 1). By bioinformatically assembling the resulting sequence fragments, longer sequences (referred to as contigs) can be assembled, and if sufficient sequences are obtained, entire genomes may be reconstructed (hence “*meta*” genomics) (Simon & Daniel, 2011). The advantage of this approach is the more accurate identification of the microorganisms present, and the ability to predict the complete functional profile of a community by examining the functional and metabolic attributes of the sequenced genomes. Additionally, metagenomic approaches avoid the known biases associated with PCR amplification, including primer bias (where primers miss certain taxa; Hong, Bunge, Leslin, Jeon, & Epstein, 2009), or chimera formation (where artificial “hybrid” sequences may be created during PCR amplification; Ashelford, Chuzhanova, Fry, Jones, & Weightman, 2005). Therefore, bioinformatically extracting phylogenetic or functional marker genes from metagenomic sequence data may offer a less biased method of surveying microbial diversity than metagenetic approaches (Guo, Cole, Zhang, Brown, & Tiedje, 2016). The compromise is that as metagenomics is not targeted, the number of sequences covering any given gene is usually far lower than could be obtained via metagenetic approaches, meaning that far deeper sequencing is required for metagenomics to be effective for such applications as biodiversity surveys. In turn, the financial and computational costs associated with acquiring and analyzing these data are often prohibitory. Nevertheless, metagenomics holds great promise as a versatile technique for studying many aspects of microbial community ecology (Neufeld, 2017).

As discussed previously, by examining RNA instead of DNA, the same (metagenomic) approach can be employed to quantify which genes are actively expressed at a given time (referred to as metatranscriptomics, Carvalhais et al., 2012, Figure 1). Thus, metatranscriptomics can provide a detailed picture of the functional and metabolic activity of a community at a given moment, and can be used to examine almost instantaneous changes to this in response to environmental perturbations or existing gradients. In subtle contrast, metagenomics quantifies the functional and metabolic potential of a community,

integrating data over a far longer temporal scale (Simon & Daniel, 2011). In addition to both these approaches, alternative “*omics*” methods can provide insights into the next level of biological organization, by studying the proteins (metaproteomics) (Wilmes & Bond, 2006) and metabolites (metabolomics) present within a community (Bundy, Davey, & Viant, 2009).

Over the last decade, the number of sequences obtained from a single NGS run has risen by several orders of magnitude (Loman et al., 2012; Quail et al., 2012). Historically, Roche’s 454 Pyrosequencing platform was the most widely used high-throughput sequencing technology in molecular microbial ecology (Dumbrell et al., 2016), yet was discontinued in 2016. In its place, Illumina’s MiSeq and HiSeq platforms now largely dominate the field, comparing favorably to other platforms in terms of number of sequences and cost to run (Loman et al., 2012; Quail et al., 2012). With the growth of sequence data sets from thousands to millions, or even billions, of sequences per run, the computational cost of analyzing molecular data sets has also grown considerably. The sheer size of modern sequence data sets has meant that bioinformatic analyses are now best conducted on high-performance computers, where the large volume of hard-disk space and random-access memory (RAM) enable researchers to work efficiently with their data. Furthermore, by taking advantage of the multiple computing cores available in modern computer processors (and across computer clusters), current bioinformatics tools are able to speed up analyses considerably by “parallelising” some tasks between cores. However, the availability of sufficient computing resources should still be a key consideration when undertaking molecular analyses via high-throughput platforms (Dumbrell et al., 2016). These platforms, coupled with the continued development of high-performance computing and bioinformatic methods (Dumbrell et al., 2016), mean that molecular ecologists now have the potential to explore the structure and function of microbial communities in detail. The ever-decreasing costs, along with the suitability of these methods to examine ecological hypotheses means that “*omics*” methods are now the de facto standard, applied widely across the field.

3 | CASE STUDIES

3.1 | qPCR—A catastrophic pesticide spill in the river Kennet

In aquatic systems, microorganisms comprise the base of food webs, and often show rapid changes in response to environmental perturbations. Therefore, understanding how microbial communities respond to environmental change illuminates how ecological impacts propagate through food webs, alongside providing information about functional changes of the microbial community itself (Thompson et al., 2016). For example, Thompson et al. (2016) demonstrated this in response to a sudden and catastrophic pesticide spill within the River Kennet (UK). The River Kennet is a major tributary of the River Thames and designated a site of special scientific interest. In the Summer of 2013, a catastrophic spill of the pesticide (insecticide) chlorpyrifos was detected within the Kennet, extinguishing all invertebrate life over a 15-km stretch of the river. By sampling microbial communities, as well as invertebrates and other macrofauna, Thompson et al. (2016) sought to understand the potential impacts of this spill across levels of biological organization: from genes to the entire ecosystem. qPCR analysis of microbial communities upstream and downstream of the pollution event, showed the pesticide had both direct and indirect effects on microbes. By quantifying changes in the abundance of the organophosphate hydrolase gene (*opd*; involved in the degradation of chlorpyrifos) across control and impacted sites (Figure 2a), Thompson et al. showed that bacteria capable of degrading the pesticide had increased sevenfold following the spill—and importantly, also highlighted that the residency time of chlorpyrifos is longer in sediments than in the water column. Quantification of the ammonia monooxygenase gene (*amoA*, involved in the oxidation of ammonia), revealed even larger changes in abundance (approximately 30-fold) of ammonia oxidizing microorganisms. The authors attributed this to the increased mass of decaying invertebrates, which would result in increased concentrations of ammonia. Thus, by targeting these two microbial functional groups, the authors were able to examine the direct and indirect effects of chlorpyrifos on microbial communities; and provide explanations for observed changes in ecosystem functioning (Thompson et al., 2016).

3.2 | Metagenetics—Anammox as an important nitrogen loss pathway in freshwaters

NGS methods have also helped to reveal the presence of various functional and metabolic pathways in freshwater microbial communities. For example, Lansdown et al. (2016) studied an alternative pathway of N_2 production—Anaerobic ammonium oxidation (anammox)—which is known to be important in marine environments, but relatively understudied in freshwaters. Anammox is the process of converting ammonia to N_2 gas in the absence of oxygen and without undergoing denitrification and nitrification, the processes that represent the canonical nitrogen cycle (Jetten et al., 1998). Previously, this process was believed to be limited to deep marine sediments where anaerobic conditions predominate (Kuypers et al., 2003), and only recently was this process observed in freshwaters (Trimmer, Nicholls, & Deflandre, 2003). Lansdown et al. (2016) sought to understand whether the base geology of freshwater streams influenced anammox, and in addition how this was underpinned

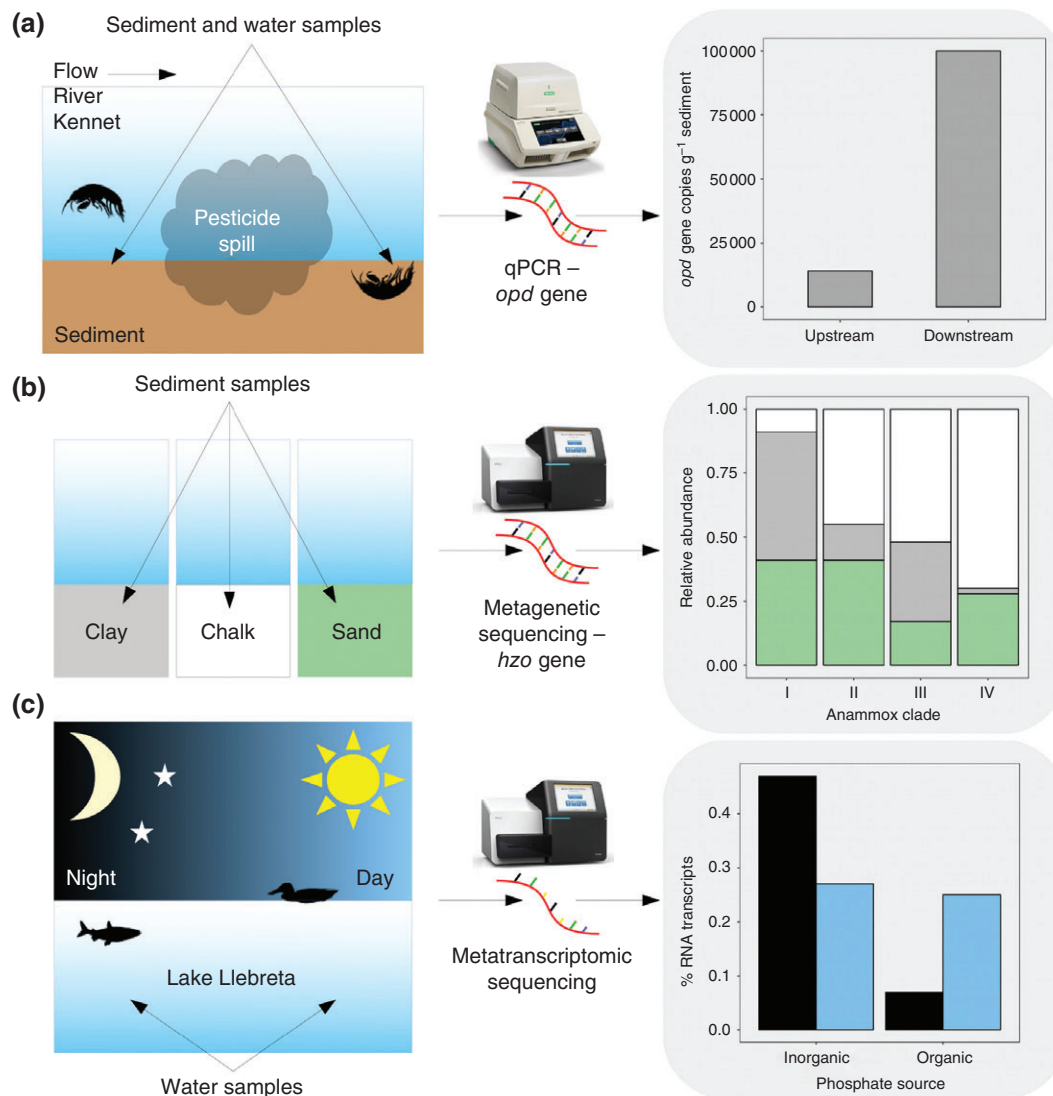


FIGURE 2 A graphical abstract for each of the case-studies presented here. (a) Thompson et al. (2016) used qPCR analysis of the *opd* gene to show that the abundance of pesticide degrading microorganisms increased sevenfold in contaminated sediments, relative to control sediments. (b) Lansdown et al. (2016) conducted metagenetic sequencing of anammox (*hzo*) genes from riverbed sediments of different underlying geology. They discovered four distinct phylogenetic clades of anammox bacteria that differed markedly in relative abundance across the different riverbed geologies. (c) Vila-Costa, Sharma, Moran, and Casamayor (2013) performed metatranscriptomic sequencing of lake plankton across a diel cycle. Notably, they found that phosphate uptake genes were differentially expressed according to the source of phosphate, and time of day

by the abundance and activity of anammox bacteria. They studied sedimentary microorganisms originating from three different geology types—chalk, greensand, and clay—using NGS metagenetic methods (Figure 2b). Known genera of anammox bacteria were not detected in 16S rRNA metagenetic data sets, suggesting they are relatively rare in comparison to other bacterial taxa. However, metagenetic sequencing of the hydrazine oxidoreductase gene (*hzo*; required for anammox) revealed the presence of four distinct clades of anammox bacteria, which changed in relative abundance between the different geologies (Figure 2b). These results supported anammox process-rate measurements, and revealed anammox as the major pathway of N₂ loss in oxic chalk streams (Lansdown et al., 2016).

3.3 | Metatranscriptomics—Diel functional diversity of freshwater microbes

Although metagenetics is highly informative, research questions addressing broad-scale patterns of microbial functional diversity are often better answered via metagenomic or metatranscriptomic approaches. For example, Vila-Costa et al., (2013) provided a comprehensive assessment of diel shifts in ecological strategies to uptake and metabolize the essential nutrient phosphorus (P), among other key functions, via a metatranscriptomic analysis of Lake Llebrete—an oligotrophic (nutrient poor) freshwater lake located in the Spanish Pyrenees (Vila-Costa et al., 2013). They detected an increased expression of genes for photoheterotrophic processes during the day relative to the night, in addition to a higher gene expression

related to the uptake of inorganic P (Figure 2c). Conversely, during the night, this shifted to increased expression of genes involved in the uptake and metabolism of organic P (Figure 2c), thus suggesting different ecological strategies of bacterial P metabolism operating throughout a 24-hr period (Vila-Costa et al., 2013). In addition, transcripts belonging to the bacterial classes Bacteroidetes and Betaproteobacteria were the most abundant regardless of sampling time, providing insight into the profile of the active freshwater bacterial community. Currently, a limitation of metatranscriptomics can be the low annotation and characterization of specific proteins, particularly within freshwater microbial communities. This is primarily due to a scarcity of freshwater reference databases, which needs considerable researcher investment for improvement, but no doubt will do so as the field continues to progress.

4 | THE FUTURE OF MOLECULAR MICROBIAL ECOLOGY

The development of DNA/RNA sequencing technologies in the last two decades has been exceptionally rapid. For example, as aforementioned, Roche's 454 pyrosequencers—a high-throughput sequencing platform used in many early NGS studies—was discontinued after only a decade of use; being superseded by newer and operationally cheaper NGS platforms with higher throughput capabilities (Goodwin, McPherson, & McCombie, 2016; Roesch et al., 2007). Thus, within the next few decades, the operational landscape of molecular microbial ecology may be almost unrecognizable. Here, we highlight some of the most promising developments likely to drive the field forward over the next few years.

Molecular tools are most likely going to provide the foundations of the “*Next Generation*” of biomonitoring methods, potentially revolutionizing the routine monitoring and management of freshwater ecosystems (Bohan et al., 2017; Jackson et al., 2016). These approaches borrow heavily from methods originally developed to quantify microbial diversity (notably metagenetics—see above), but are extended to detect all taxonomic groups simultaneously; from micro- to macro-organisms. This extension of approaches is conceptually very simple; the “genetic soup” of nucleic acids isolated from an environmental sample will also contain environmental DNA (eDNA) originating from feces, mucous, shedded tissue, or decaying matter, among other sources, of all organisms present within that ecosystem (Thomsen & Willerslev, 2015). Thus minor modifications to PCR amplification protocols (e.g., using primers that target all domains of life, and not just microbial taxa) can quantify all species present within a sample. The potential for these approaches is vast; overcoming issues associated with the availability of taxonomic expertise, and the nonstandardization of current biomonitoring approaches (Hopkins & Freckleton, 2002). Consequently, the use of eDNA methods to quantify biodiversity is rapidly growing and driving continued methodological improvements (Thomsen & Willerslev, 2015), and will shortly become a routine approach in freshwater habitats (Rees, Maddison, Middleditch, Patmore, & Gough, 2014; Thomsen et al., 2012).

While molecular tools offer a huge potential in the monitoring of freshwater biodiversity, examining just the microbial component of the food web can provide a rapid and real-time measure of contemporary ecosystem health. As aforementioned (see Case Studies), microbial communities respond almost instantaneously to environmental change, thus providing environmental management authorities with an “early warning” system. Indeed, the route of entry into the food web of many chemical stressors is via basal microbial biofilms. These are not novel concepts, and microbial eukaryotes, such as diatoms are routinely used in biomonitoring. For example, the trophic diatom index (TDI) indicates the trophic status of a water body, and is based on the presence of certain diatom taxa associated with waters of different trophic status (Kelly et al., 2008). However, the TDI requires time-consuming morphological identification of a challenging group of organisms, and attempts to update this via molecular methods have been hindered by a disconnect between morphological and molecular taxonomy (Apothéoz-Perret-Gentil et al., 2017; Visco et al., 2015). Yet, as technologies develop, we foresee overcoming current limitations, and the rapid detection of novel stressors in freshwater ecosystems via routine monitoring of microbial communities.

Future developments may also reduce the need for dedicated molecular laboratories, by providing portable field-based technologies (Bohan et al., 2017). For example, the Nanopore MinION already provides researchers with in-field NGS capabilities, although improvements in sequence quality are required (Mikheyev & Tin, 2014). While, the miniaturization of PCR technologies currently allows in-field nucleic acid amplification, for both metagenetics and qPCR (Marx, 2015). These developments, coupled with easily programmable micro-computers (e.g., Raspberry Pi) (Cressey, 2017), bring the intriguing prospect of fully automated, remote molecular biomonitoring closer to reality (Bohan et al., 2017; Jackson et al., 2016). This has the potential to provide faster identification of environmental perturbations, and thus quicker implementation of remedial actions, while at the same time reducing sample handling and storage issues that could influence microbial community composition (Lauber, Zhou, Gordon, Knight, & Fierer, 2010; Rochelle, Cragg, Fry, Parkes, & Weightman, 1994). This latter point is particularly important for RNA-based approaches (e.g., metatranscriptomics), due to its short half-life and instability outside of dedicated cold (below -80°C) storage.

Not all developments will be field-based. Until recently, due to a lack of taxonomically annotated functional gene databases, linking phylogenetic and functional gene data has been challenging, prompting bioinformatic approaches that attempt

to “predict” functional traits from phylogenetic sequence data, such as PICRUSt (Langille et al., 2013). Subsequently, the only way to comprehensively link functional genes with phylogenetically informative genes was to isolate the organism of interest, and sequence its entire genome, which is limited by the (lack of) culturability of most microorganisms. However, “Emulsion, Paired Isolation, and Concatenation PCR” (epicPCR) can overcome this limitation by capturing individual microbial cells within emulsion droplets before isolating nucleic acids, allowing the co-amplification of phylogenetic and functional genes (Spencer et al., 2016). Alternatively, single-cell genomics approaches are able to isolate individual microbial cells from mixed communities (e.g., by flow cytometry) before genome sequencing, thereby negating the need to isolate cells in pure-culture. Single-cell “omics” approaches are already revealing the previously unknown functionality of ecologically important microorganisms in freshwater systems (Ghylin et al., 2014). The coupling of cutting edge high-throughput workflows to single-cell “omics” introduces the tantalizing prospect of gaining enormous insight into the function and phylogeny of aquatic microbial communities (Lan, Demaree, Ahmed, & Abate, 2017). These approaches provide novel, and potentially high-throughput, methods of simultaneously assessing microbial taxonomic and functional diversity; arguably achieving the gold-standard of ecologically meaningful data.

5 | CONCLUDING REMARKS

The rapid advancement of molecular techniques has instigated a “golden age” in microbial ecology; providing the tools to study freshwater microbial communities in unprecedented detail. As a result, a more holistic understanding of freshwater habitats is emerging, alongside the critical roles microorganisms fulfill within them. Currently, developments in molecular microbial ecology are being utilized across the wider ecological research community; refining biomonitoring, redefining food webs, and examining ecological networks that would have previously omitted microbial taxa. The continued increase in multiple stressors acting on freshwater ecosystems leads to growing challenges in environmental management, but with the latest generation of molecular tools at our disposal, the role microbial communities play in meeting and understanding these challenges will never be overlooked again.

ACKNOWLEDGMENTS

We are grateful to various funding sources. This work was supported by the Natural Environment Research Council (NERC) via the following grants and studentships: “Impacts of global warming in sentinel systems: from genes to ecosystems” (NE/M02086X/1 – AJD, DH, HJP, KCR); “Rapid monitoring of bioaerosols in Urban, Agricultural and Industrial Environments” (NE/M010813/1 – AJD, RMWF); quota award studentships (471757 – DC; W86713B – SLW); EnvEast DTP studentship (E15EE – LR). Research was also supported by a PhD studentship provided from the University of Essex, School of Biological Sciences (KJMN).

CONFLICT OF INTEREST

The authors have declared no conflicts of interest for this article.

RELATED WIREs ARTICLE

[Impacts and indicators of change in lotic ecosystems](#)

REFERENCES

- Abarenkov, K., Henrik Nilsson, R., Larsson, K. H., Alexander, I. J., Eberhardt, U., Erland, S., ... Sen, R. (2010). The UNITE database for molecular identification of fungi—Recent updates and future perspectives. *New Phytologist*, 186(2), 281–285.
- Apothélos-Perret-Gentil, L., Cordonier, A., Straub, F., Iseli, J., Esling, P., & Pawlowski, J. (2017). Taxonomy-free molecular diatom index for high-throughput eDNA biomonitoring. *Molecular Ecology Resources*, 17(6), 1231–1242.
- Ashelford, K. E., Chuzhanova, N. A., Fry, J. C., Jones, A. J., & Weightman, A. J. (2005). At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. *Applied and Environmental Microbiology*, 71(12), 7724–7736.
- Bernhard, A. E., & Field, K. G. (2000). Identification of nonpoint sources of fecal pollution in coastal waters by using host-specific 16S ribosomal DNA genetic markers from fecal anaerobes. *Applied and Environmental Microbiology*, 66(4), 1587–1594.
- Bohan, D. A., Vacher, C., Tamaddon-Nezhad, A., Raybould, A., Dumbrell, A. J., & Woodward, G. (2017). Next-generation global biomonitoring: Large-scale, automated reconstruction of ecological networks. *Trends in Ecology & Evolution*, 32(7), 477–487.
- Bomar, L., Maltz, M., Colston, S., & Graf, J. (2011). Directed culturing of microorganisms using metatranscriptomics. *MBio*, 2(2), e00012-11.
- Brock, T. D. (1961). Milestones in microbiology. *Academic Medicine*, 36(7), 847.
- Bundy, J. G., Davey, M. P., & Viant, M. R. (2009). Environmental metabolomics: A critical review and future perspectives. *Metabolomics*, 5(1), 3–21.

- Carvalho, L. C., Dennis, P. G., Tyson, G. W., & Schenk, P. M. (2012). Application of metatranscriptomics to soil environments. *Journal of Microbiological Methods*, 91(2), 246–251.
- Caumette, P., Bertrand, J. C., & Normand, P. (2015). Some historical elements of microbial ecology. In *Environmental microbiology: Fundamentals and applications* (pp. 9–24). Dordrecht, The Netherlands: Springer.
- Chapin, F. S., Walker, B. H., Hobbs, R. J., Hooper, D. U., Lawton, J. H., Sala, O. E., & Tilman, D. (1997). Biotic control over the functioning of ecosystems. *Science*, 277(5325), 500–504.
- Cole, J. R., Wang, Q., Fish, J. A., Chai, B., McGarrell, D. M., Sun, Y., ... Tiedje, J. M. (2013). Ribosomal database project: Data and tools for high throughput rRNA analysis. *Nucleic Acids Research*, 42(D1), D633–D642.
- Cressey, D. (2017). The DIY electronics transforming research. *Nature News*, 544(7648), 125–126.
- Dudgeon, D., Arthington, A. H., Gessner, M. O., Kawabata, Z. I., Knowler, D. J., Lévêque, C., ... Sullivan, C. A. (2006). Freshwater biodiversity: Importance, threats, status and conservation challenges. *Biological Reviews*, 81(2), 163–182.
- Dumbrell, A. J., Ferguson, R. M., & Clark, D. R. (2016). Microbial community analysis by single-amplicon high-throughput next generation sequencing: Data analysis—from raw output to ecology. In *Hydrocarbon and lipid microbiology protocols* (pp. 155–206). Berlin, Germany: Springer.
- Falkowski, P. G., Fenchel, T., & DeLong, E. F. (2008). The microbial engines that drive Earth's biogeochemical cycles. *Science*, 320(5879), 1034–1039.
- Fish, J. A., Chai, B., Wang, Q., Sun, Y., Brown, C. T., Tiedje, J. M., & Cole, J. R. (2013). FunGene: The functional gene pipeline and repository. *Frontiers in Microbiology*, 4(291), 1–14.
- Ghylin, T. W., Garcia, S. L., Moya, F., Oyserman, B. O., Schwientek, P., Forest, K. T., ... Sczyrba, A. (2014). Comparative single-cell genomics reveals potential ecological niches for the freshwater actinobacteria lineage. *The ISME Journal*, 8(12), 2503–2516.
- Giovannoni, S., & Stingl, U. (2007). The importance of culturing bacterioplankton in the 'omics' age. *Nature Reviews Microbiology*, 5(10), 820–826.
- Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: Ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, 17(6), 333–351.
- Griffiths, R. I., Whiteley, A. S., O'Donnell, A. G., & Bailey, M. J. (2000). Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA-and rRNA-based microbial community composition. *Applied and Environmental Microbiology*, 66(12), 5488–5491.
- Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., ... Del Campo, J. (2012). The Protist ribosomal reference database (PR2): A catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Research*, 41(D1), D597–D604.
- Guo, J., Cole, J. R., Zhang, Q., Brown, C. T., & Tiedje, J. M. (2016). Microbial community analysis with ribosomal gene fragments from shotgun metagenomes. *Applied and Environmental Microbiology*, 82(1), 157–166.
- Hong, S., Bunge, J., Leslin, C., Jeon, S., & Epstein, S. S. (2009). Polymerase chain reaction primers miss half of rRNA microbial diversity. *The ISME Journal*, 3(12), 1365–1373.
- Hopkins, G. W., & Freckleton, R. P. (2002). Declines in the numbers of amateur and professional taxonomists: Implications for conservation. *Animal Conservation*, 5, 245–249.
- Hugenholtz, P. (2002). Exploring prokaryotic diversity in the genomic era. *Genome Biology*, 3(2), reviews0003-1.
- Jackson, M. C., Weyl, O. L. F., Altermatt, F., Durance, I., Friberg, N., Dumbrell, A. J., ... Leadley, P. W. (2016). Chapter twelve—recommendations for the next generation of global freshwater biological monitoring tools. *Advances in Ecological Research*, 55, 615–636.
- Jetten, M. S., Strous, M., Van de Pas-Schoonen, K. T., Schalk, J., van Dongen, U. G., van de Graaf, A. A., ... Kuenen, J. G. (1998). The anaerobic oxidation of ammonium. *FEMS Microbiology Reviews*, 22(5), 421–437.
- Kallmeyer, J., Pockalny, R., Adhikari, R. R., Smith, D. C., & D'Hondt, S. (2012). Global distribution of microbial abundance and biomass in seafloor sediment. *Proceedings of the National Academy of Sciences*, 109(40), 16213–16216.
- Kelly, M., Juggins, S., Guthrie, R., Pritchard, S., Jamieson, J., Rippey, B., ... Yallop, M. (2008). Assessment of ecological status in UK rivers using diatoms. *Freshwater Biology*, 53(2), 403–422.
- Kuypers, M. M., Sliemers, A. O., Lavik, G., Schmid, M., Jørgensen, B. B., Kuenen, J. G., ... Jetten, M. S. (2003). Anaerobic ammonium oxidation by anammox bacteria in the Black Sea. *Nature*, 422(6932), 608–611.
- Lan, F., Demaree, B., Ahmed, N., & Abate, A. R. (2017). Single-cell genome sequencing at ultra-high-throughput with microfluidic droplet barcoding. *Nature Biotechnology*, 35, 640–646.
- Lane, N. (2015). The unseen world: Reflections on Leeuwenhoek (1677) 'concerning little animals'. *Philosophical Transactions of the Royal Society B*, 370(1666), 20140344.
- Langille, M. G., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A., ... Beiko, R. G. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature Biotechnology*, 31(9), 814–821.
- Lansdown, K., McKew, B. A., Whitby, C., Heppell, C. M., Dumbrell, A. J., Binley, A., ... Trimmer, M. (2016). Importance and controls of anaerobic ammonium oxidation influenced by riverbed geology. *Nature Geoscience*, 9(5), 357–360.
- Laufer, C. L., Zhou, N., Gordon, J. I., Knight, R., & Fierer, N. (2010). Effect of storage conditions on the assessment of bacterial community structure in soil and human-associated samples. *FEMS Microbiology Letters*, 307(1), 80–86.
- Leigh, M. B., Taylor, L., & Neufeld, J. D. (2015). Clone libraries of ribosomal RNA gene sequences for characterization of microbial communities. In *Hydrocarbon and lipid microbiology protocols* (pp. 127–154). Berlin, Germany: Springer.
- Locey, K. J., & Lennon, J. T. (2016). Scaling laws predict global microbial diversity. *Proceedings of the National Academy of Sciences*, 113(21), 5970–5975.
- Loman, N. J., Misra, R. V., Dallman, T. J., Constantinidou, C., Gharbia, S. E., Wain, J., & Pallen, M. J. (2012). Performance comparison of benchtop high-throughput sequencing platforms. *Nature Biotechnology*, 30(5), 434–439.
- Marx, V. (2015). PCR heads into the field. *Nature Methods*, 12(5), 393–397.
- Mikheyev, A. S., & Tin, M. M. (2014). A first look at the Oxford Nanopore MinION sequencer. *Molecular Ecology Resources*, 14(6), 1097–1102.
- Muyzer, G., & Smalla, K. (1998). Application of denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) in microbial ecology. *Antonie Van Leeuwenhoek*, 73(1), 127–141.
- Neufeld, J. D. (2017). Migrating SSU rRNA gene surveys to the metagenomics era. *Environmental Microbiology Reports*, 9(1), 23–24.
- Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., ... Gu, Y. (2012). A tale of three next generation sequencing platforms: Comparison of ion torrent, Pacific biosciences and Illumina MiSeq sequencers. *BMC Genomics*, 13(1), 341.
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., ... Glöckner, F. O. (2012). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, 41(D1), D590–D596.
- Rees, H. C., Maddison, B. C., Middleditch, D. J., Patmore, J. R., & Gough, K. C. (2014). The detection of aquatic animal species using environmental DNA—A review of eDNA as a survey tool in ecology. *Journal of Applied Ecology*, 51(5), 1450–1459.
- Roche, P. A., Cragg, B. A., Fry, J. C., Parkes, R. J., & Weightman, A. J. (1994). Effect of sample handling on estimation of bacterial diversity in marine sediments by 16S rRNA gene sequence analysis. *FEMS Microbiology Ecology*, 15(1–2), 215–225.

- Roesch, L. F., Fulthorpe, R. R., Riva, A., Casella, G., Hadwin, A. K., Kent, A. D., ... Triplett, E. W. (2007). Pyrosequencing enumerates and contrasts soil microbial diversity. *The ISME Journal*, 1(4), 283–290.
- Shendure, J., & Ji, H. (2008). Next-generation DNA sequencing. *Nature Biotechnology*, 26(10), 1135–1145.
- Simon, C., & Daniel, R. (2011). Metagenomic analyses: Past and future trends. *Applied and Environmental Microbiology*, 77(4), 1153–1161.
- Sogin, M. L., Morrison, H. G., Huber, J. A., Welch, D. M., Huse, S. M., Neal, P. R., ... Herndl, G. J. (2006). Microbial diversity in the deep sea and the underexplored “rare biosphere”. *Proceedings of the National Academy of Sciences*, 103(32), 12115–12120.
- Spencer, S. J., Tamminen, M. V., Preheim, S. P., Guo, M. T., Briggs, A. W., Brito, I. L., ... Alm, E. J. (2016). Massively parallel sequencing of single cells by epicPCR links functional genes with phylogenetic markers. *The ISME Journal*, 10(2), 427–436.
- Staley, J. T., & Konopka, A. (1985). Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annual Reviews in Microbiology*, 39(1), 321–346.
- Taylor, N. G., Verner-Jeffreys, D. W., & Baker-Austin, C. (2011). Aquatic systems: Maintaining, mixing and mobilising antimicrobial resistance? *Trends in Ecology & Evolution*, 26(6), 278–284.
- Thompson, M. S., Bankier, C., Bell, T., Dumbrell, A. J., Gray, C., Ledger, M. E., ... Trimmer, M. (2016). Gene-to-ecosystem impacts of a catastrophic pesticide spill: Testing a multilevel bioassessment approach in a river ecosystem. *Freshwater Biology*, 61(12), 2037–2050.
- Thomsen, P., Kielgast, J. O. S., Iversen, L. L., Wiuf, C., Rasmussen, M., Gilbert, M. T. P., ... Willerslev, E. (2012). Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular Ecology*, 21(11), 2565–2573.
- Thomsen, P. F., & Willerslev, E. (2015). Environmental DNA—An emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation*, 183, 4–18.
- Trimmer, M., Nicholls, J. C., & Deflandre, B. (2003). Anaerobic ammonium oxidation measured in sediments along the Thames estuary, United Kingdom. *Applied and Environmental Microbiology*, 69(11), 6447–6454.
- Vila-Costa, M., Sharma, S., Moran, M. A., & Casamayor, E. O. (2013). Diel gene expression profiles of a phosphorus limited mountain lake using metatranscriptomics. *Environmental Microbiology*, 15(4), 1190–1203.
- Visco, J. A., Apothéloz-Perret-Gentil, L., Cordonier, A., Esling, P., Pillet, L., & Pawlowski, J. (2015). Environmental monitoring: Inferring the diatom index from next-generation sequencing data. *Environmental Science & Technology*, 49(13), 7597–7605.
- Werner, J. J., Koren, O., Hugenholtz, P., DeSantis, T. Z., Walters, W. A., Caporaso, J. G., ... Ley, R. E. (2012). Impact of training sets on classification of high-throughput bacterial 16S rRNA gene surveys. *The ISME Journal*, 6(1), 94–103.
- Whitman, W. B., Coleman, D. C., & Wiebe, W. J. (1998). Prokaryotes: The unseen majority. *Proceedings of the National Academy of Sciences*, 95(12), 6578–6583.
- Wilmes, P., & Bond, P. L. (2006). Metaproteomics: Studying functional gene expression in microbial ecosystems. *Trends in Microbiology*, 14(2), 92–97.
- Xu, J. (2006). Invited review: Microbial ecology in the age of genomics and metagenomics: Concepts, tools, and recent advances. *Molecular Ecology*, 15(7), 1713–1731.

How to cite this article: Clark DR, Ferguson RMW, Harris DN, et al. Streams of data from drops of water: 21st century molecular microbial ecology. *WIREs Water*. 2018;e1280. <https://doi.org/10.1002/wat2.1280>