Prospects for multi-omics in the microbial ecology of water engineering

McDaniel, Elizabeth A.; Wahl, Sebastian Aljoscha; Ishii, Shun'ichi; Pinto, Ameet; Ziels, Ryan; Nielsen, Per Halkjær; McMahon, Katherine D.; Williams, Rohan B.H.

DOI
10.1016/j.watres.2021.117608

Publication date
2021

Document Version
Accepted author manuscript

Published in
Water Research

Prospects for Multi-omics in the Microbial Ecology of Water Engineering

Elizabeth A. McDaniel[1*], Sebastian Aljoscha Wahl[2], Shun'ichi Ishii[3], Ameet Pinto[4], Ryan Ziels[5], Per H Nielsen[6], Katherine D. McMahon[1,7], Rohan B.H. Williams[8*]

[1] Department of Bacteriology, University of Wisconsin – Madison, Madison, WI, USA

[2] Department of Biotechnology, Delft University of Technology, Delft, The Netherlands

[3] Super-cutting-edge Grand and Advanced Research (SUGAR) Program, Institute for Extra-cutting-edge Science and Technology Avant-garde Research (X-star), Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Yokosuka 237-0061, Japan

[4] Department of Civil and Environmental Engineering, Northeastern University, Boston, MA, USA

[5] Department of Civil Engineering, The University of British Columbia, Vancouver, BC, Canada

[6] Center for Microbial Communities, Aalborg University, Aalborg, Denmark

[7] Department of Civil and Environmental Engineering, University of Wisconsin – Madison, Madison, WI, USA

[8] Singapore Centre for Environmental Life Sciences Engineering, National University of Singapore, Republic of Singapore

* Corresponding authors: elizabethmcd93@gmail.com (E.A.M) and lsirbhw@nus.edu.sg (R.B.H.W)

ABSTRACT

Advances in high-throughput sequencing technologies and bioinformatics approaches over almost the last three decades have substantially increased our ability to explore microorganisms and their functions – including those that have yet to be cultivated in pure isolation. Genome-resolved metagenomic approaches have enabled linking powerful functional predictions to specific taxonomical groups with increasing fidelity. Additionally, related developments in both whole community gene expression surveys and metabolite profiling have permitted for direct surveys of community-scale functions in specific environmental settings. These advances have allowed for a shift in microbiome science away from descriptive studies and towards mechanistic and predictive frameworks for designing and harnessing microbial communities for desired beneficial outcomes. Water engineers, microbiologists, and microbial ecologists studying activated sludge, anaerobic digestion, and drinking water distribution systems have applied various (meta)omics techniques for connecting microbial community dynamics and physiologies to overall process parameters and system performance. However, the rapid pace at which new omics-based approaches are developed can appear daunting to those looking to apply these state-of-the-art practices for the first time. Here, we review how modern genome-resolved metagenomic approaches have been applied to a variety of water engineering applications from lab-scale bioreactors to full-scale systems. We describe integrated omics analysis across engineered water systems and the foundations for pairing these insights with modeling approaches. Lastly, we summarize emerging omics-based technologies that we believe will be powerful tools for water engineering applications. Overall, we provide a framework for microbial ecologists specializing in water engineering to apply cutting-edge omics approaches to their research questions to achieve novel functional insights. Successful adoption of predictive frameworks in engineered water systems could enable more economically and environmentally sustainable bioprocesses as demand for water and energy resources increases.

KEYWORDS

microbial ecology, metagenomics, metatranscriptomics, metaproteomics, metabolomics, metabolic modelling, water engineering

INTRODUCTION

Microorganisms and complex microbial communities are an integral and essential component of engineered water systems designed to improve public health, maintain ecosystem health, and provide optimal water infrastructure and quality. Activated sludge (AS), anaerobic digestion (AD), and drinking water systems (DWSs) are all designed to harness or manage microbial communities for desired process outcomes (**Figure 1**). AS systems are designed to enrich for specific microbial lineages to remove nutrients, mostly organic carbon, phosphorus, and nitrogen species to prevent downstream receiving water eutrophication (Ardern and Lockett, 1914). AD processes are applied to various organic waste streams to recover biogas and other valuable products as a source of renewable energy (Angelidaki et al., 2011; Weiland, 2010). While biofilms in drinking water treatment plants (DWTPs) can leverage microbial communities for pollutant removal (Kirisits et al., 2019), DWSs apply physicochemical approaches to minimize microbial abundance in tap water (Berry et al., 2006). All of these systems are the foundation of modern urban water infrastructure, and are intertwined through their dependence on adequately managing different microbial communities through a water-microbiome continuum (Raskin and Nielsen, 2019). As many of the key microorganisms underpinning these water cycles remain uncultured to date (Lloyd et al., 2018; Steen et al., 2019), our understanding of the microbial lineages involved in these engineered ecosystems have largely relied on cultivation-independent approaches.

Tools for characterizing the structure and function of microbial communities have greatly expanded since the development of methods for surveying uncultivated microorganisms based on the use of the 16S ribosomal RNA (rRNA) gene (Fox et al., 1980; Lane et al., 1986; Woese and Fox, 1977). Building on the identification of specific 16S rRNA gene sequences, uncultivated microorganisms could be directly visualized using fluorescently labelled oligodeoxynucleotide probes (DeLong et al., 1989) or microautoradiography

(Giovannoni et al., 1988). Although advances in high-throughput sequencing technologies have allowed for large-scale 16S rRNA gene amplicon surveys instead of constructing labor-intensive clone libraries (Sogin et al., 2006; Caporaso et al., 2012; Kozich et al., 2013; Thompson et al., 2017), operationally defined taxonomical units cannot be linked to functional capabilities in mixed communities with fidelity, especially for poorly characterized lineages. Metagenomics comprises a group of methods for solving this problem. The term metagenome is defined as the entire genetic content of the collective microbial genomes contained within a given environmental sample (Handelsman et al., 2002, 1998). Early metagenomic approaches cloned fragments of extracted microbial DNA into a host backbone (e.g., *E. coli)* to screen for particular sequences (Beja et al., 2000; Henne et al., 1999), functionally characterize novel enzymes through heterologous expression systems (Brady and Clardy, 2000; Rondon et al., 2000), or perform shotgun sequencing (Tyson et al., 2004; Venter et al., 2004). All of these approaches have the potential to link functional observations to phylogenetic identity, and therefore provide a powerful approach to characterize the ecology and evolution of uncultivated microorganisms.

Metagenomic sequencing approaches can be defined under two categories, read-based and assembly-based. Read-based metagenomic approaches map the resulting reads back to reference genomes to assess the taxonomic composition and functional potential of a mixed microbial sample (Quince et al., 2017b; Sharpton, 2014). However, read-based approaches heavily rely on reference databases, and therefore under-explored environments such as engineered water systems may not be well-represented in these databases. Additionally, this approach cannot directly link specific functions to individual microbial guilds or populations. Assembly-based approaches such as genome-resolved metagenomics extract individual genomes *de novo* from a mixed community through a process termed binning of assembled contigs or scaffolds (Albertsen et al., 2013; Tyson et al., 2004). The resulting genomes can be termed draft bins, metagenome-assembled genomes (MAGs), and/or population genomes, as they most often represent a mixed consensus sequence of closely related microbial cells (Bowers et al., 2017). These genomes can then be used as a foundation

for integrated multi-omics analyses using insights from metatranscriptomics, metaproteomics, and/or metabolomics that attempt to directly tie metabolic guilds and functional activities to specific microorganisms (Anantharaman et al., 2016; Woodcroft et al., 2018; Wrighton et al., 2012).

In recent years, there have been calls for microbiome science as a whole to shift away from mere descriptive studies and towards more mechanistic and predictive frameworks for designing and harnessing diverse microbial communities for desired beneficial outcomes (Alivisatos et al., 2015; Lawson et al., 2019). Successful adoption of predictive frameworks in engineered water systems could enable more economically and environmentally sustainable bioprocesses as demand for water and energy resources increases (Raskin and Nielsen, 2019; Sheik et al., 2014). To achieve this, an interdisciplinary approach combining theoretical microbial ecology and engineering principles must be applied. Genome-resolved metagenomic approaches have been widely adopted in water engineering applications to explore the diversity, metabolic potential, and spatiotemporal dynamics of microorganisms inhabiting water systems. Integrating multi-omics approaches within an eco-systems biology framework is a promising approach for creating more predictable, controllable biotechnology applications for water microbiomes (Narayanasamy et al., 2015).

Here, we review existing work and prospects for multi-omics applications in engineered water systems. We summarize pioneering work in genome-resolved metagenomic approaches over the last two decades in AS, AD, and DWDS. We then describe how mechanistic modeling and time-series forecasting of microorganisms in water systems can integrate genome-resolved metagenomics to enhance their predictive capabilities. Finally, we highlight emerging omics technologies that will pave the way for remarkable advances not only in water systems but microbiome science as a whole.

1. GENOME-RESOLVED METAGENOMIC APPROACHES FOR UNCOVERING PRINCIPLES OF MICROBIAL ECOLOGY IN WATER SYSTEMS

Numerous molecular methods have been applied to engineered water systems to explore the diversity and ecophysiology of microorganisms, including 16S rRNA gene amplicon sequencing and *in situ* hybridization approaches. However, these approaches alone are fundamentally limited in their ability to connect structure-to-function for any natural microbial community. Here, we review how early applications of molecular approaches have provided valuable insights into the microbial ecology of water systems. We then broadly summarize genome-resolved metagenomic approaches that span the methodologies of metagenomic binning, metatranscriptomics, and metaproteomics, and highlight their use in water engineering applications.

1A. Early Insights from Molecular Approaches

The demonstratable success of molecular biology in the 1970s-80s in developing new methods for isolating, sequencing, and manipulating nucleic acid sequences naturally led microbial ecologists to apply these techniques to identifying uncultivated microbes in environmental samples (see extensive contemporary reviews in Amann et al., 1995; Olsen et al., 1986). Once the small subunit rRNA was identified as a suitable marker gene for phylogenetic analysis of taxa (Woese and Fox, 1977), several landmark studies demonstrated the feasibility of surveying the composition of natural microbial communities through comparing sequences of clone libraries and amplicons (Giovannoni et al., 1990; Schmidt et al., 1991). The utility of 16S rRNA gene profiling increased dramatically with the development of second-generation sequencing technologies initially with Roche 454 and subsequently Illumina, along with the development of related bioinformatics workflows (Bolyen et al., 2019; Callahan et al., 2016; Schloss et al., 2009). 16S rRNA gene amplicon sequencing surveys have been applied to infer the diversity and dynamics of specific microbial populations in full-scale WWTP configurations over space and time, revealing contributions of core microbiome members and rare taxa in these systems (Lawson et al., 2015; Saunders et al., 2016; Wu et al., 2019; Zhang et al., 2012). Based on 16S rRNA gene sequencing surveys, the Microbial Database for Activated Sludge (MiDAS) field guide was developed to connect functional

information to manually curated taxonomical assignments for abundant microbial lineages in activated sludge (McIlroy et al., 2015), and was further updated to include anaerobic digestion systems (McIlroy et al., 2017).

The advent of 16S rRNA gene sequencing led to using radioactively-labeled oligonucleotide probes binding to rRNA structures to quantify the abundance of specific microbial populations (Giovannoni et al., 1988; Stahl et al., 1988). Panels of oligonucleotide probes were constructed for major subclasses of *Proteobacteria* and specific members of the Cytophaga-Flavobacter-Bacteroidetes (CFB) supergroup (Manz et al., 1996, 1992). Development of group-specific hybridization probes allowed for quantifying the relative abundance of methanogens in anaerobic digestion systems (Raskin et al., 1994a, 1994b), which were subsequently used to connect population dynamics to process performance parameters (Griffin et al., 1998; McMahon et al., 2001). Modifications to these methods including combining fluorescent *in situ* hybridization (FISH) with rRNA-targeted oligonucleotide probes and adding radioactively-labeled substrates through microautoradiography to visualize spatial co-localization and metabolic activities of microbial communities (Lee et al., 1999). These methods were particularly appealing for studying the dense, biofilm-like floc aggregates characteristic of activated sludge systems, such as quantifying and localizing filamentous and nitrifying bacteria within individual flocs (Mobarry et al., 1996; Wagner et al., 1994a). Subsequently, FISH-based methods combined with isotopes were used for understanding *in situ* activities (Amann, 1995; Wagner and Haider, 2012), and could also be used for verifying metabolic potentials of microorganisms made through hypotheses with reconstructed genomes (McIlroy et al., 2016).

Although both 16S rRNA gene sequencing and oligonucleotide probe applications have provided valuable insights into engineered water systems and continue to be used for novel insights, they carry significant limitations. 16S rRNA gene amplicon sequencing can be used as a high-throughput and low-cost tool to characterize microbial community dynamics through time and space; however, associated functional predictions cannot be captured *a priori* using this approach. Additionally, specific microbial traits underlying key biogeochemical

transformations have been shown to be horizontally transferred and therefore not easily inferred from 16S rRNA gene identity (Anantharaman et al., 2018; Lawson and Lücker, 2018; McDaniel et al., 2020; Palomo et al., 2018; Parsons et al., 2020). The analysis of 16S rRNA gene amplicons remains complex due to conceptual difficulties surrounding the definition of OTUs (Edgar, 2018; Stackebrandt and Goebel, 1994), the propensity of OTU-clustering to generate misleading artifacts, limitations of cross-referencing short, noisy amplicon sequences across multiple studies, and the challenges of distinguishing closely related strains harboring similar 16S rRNA gene sequences. Some of these limitations have been addressed by constructing ecosystem-specific databases that can resolve amplicons at the species level due to using full-length 16S rRNA genes from long-read sequencing technologies (Nierychlo et al., 2020).

Read-based metagenomic approaches were initially used in engineered water systems to compare *in situ* species to existing reference genomes and measure the extent of similarity and population heterogeneity (Albertsen et al., 2011; MacIlroy et al., 2013). However, read-based metagenomic approaches depend on comprehensive reference databases to ensure accuracy of results. Although improvements in metagenomic sequencing efforts have increased the amount of reference genomes from environmental ecosystems (Nayfach et al., 2020; Parks et al., 2017), the fraction of high-quality genomes from engineered water systems still remains small (Hull et al., 2019; Nayfach et al., 2020). Therefore, substantial efforts are still needed to construct high-quality reference databases of genomes from anaerobic digestion, activated sludge, and drinking water system microbiomes. Although thorough methodological reviews and best practices in metagenomics exist elsewhere (Dick, 2018; Knight et al., 2018; Quince et al., 2017b), we provide brief overviews of existing and newly developed workflows and toolkits in genome-resolved omics to provide water engineers and scientists with the most recent advances and guidelines in a field subject to very rapid development. Lastly, we outline current technologies and approaches for applying integrated multi-omics approaches to explore the dynamics and activities of microbial communities underpinning engineered water systems.

1B. Genome-Resolved Metagenomics: From Individual Genomes to Large-Scale Reference Databases

Genome-resolved metagenomic approaches apply *de novo* assembly to extract representative population genomes from a given system through metagenomic binning, and subsequently use draft bins with multi-omics approaches to gain functional insights (**Figure 2**). The first objective of any genome-resolve approach is to assemble the representative members of a given community into MAGs. This process is termed “binning”, in which short reads are *de novo* assembled into longer contigs, with groups of contigs hypothesized to arise from the same genomic context subsequently clustered based on sequence composition, differential coverage across time and/or space, or a combination of both parameters (**Figure 3**) (Albertsen et al., 2013; Alneberg et al., 2014; Dick et al., 2009; Sharon et al., 2013). These groups of “binned” contigs can be considered as working models of whole genomes, also referred to as population genomes, as the phylogenetic composition of assembled bins can be operationally defined by the minimum sequence identity of reads mapping back to assembled contigs (Jain et al., 2018; Olm et al., 2020). For example, bins are clustered and/or dereplicated at the 95% sequence identity threshold, and therefore a bin representing an “individual species” may actually be composed of multiple, closely-related strains that could not be assembled individually due to computational constraints (Jain et al., 2018; Olm et al., 2017). Draft bins can then be curated to verify uniform read coverage of all contigs (Bornemann et al., 2020; Eren et al., 2015), estimate approximate genome completion and contamination estimates by detecting universal single-copy genes (Parks et al., 2015; Seppey et al., 2019), dereplicate redundant sets of bins resulting from the use of binning results from multiple assemblies or binning algorithms (Olm et al., 2017; Sieber et al., 2018), or manually scaffold bins to close gaps and potentially recover completely circular MAGs (Chen et al., 2020; Lui et al., 2020).

After curation, taxonomical assignments and functional annotations are performed on all MAGs to connect structure to potential function. Currently, two main approaches exist for

assigning taxonomy to draft genomes; an automated approach based on whole-genome sequence similarity against publicly available genomes (Chaumeil et al., 2019; Parks et al., 2018), and approaches based on clustering of ribosomal protein sequences followed by phylogenetic reconstruction with select reference genomes (Hug et al., 2016; Lee, 2019; McDaniel et al., 2019), both achieved through concatenating alignments of single copy marker genes. After assigning taxonomical classifications to genomes, specific metabolic pathways can be investigated based on gene presence and/or absence. Numerous pipelines exist for assigning functional annotations to draft genomes, and mostly differ in the underlying database(s) from which they draw functional annotations (Aramaki et al., 2019; Hanson et al., 2014; Seemann, 2014). Recently, several pipelines have been developed for exploring specific metabolic guilds within MAG datasets based on curated databases or marker profiles (McDaniel et al., 2019; Neely et al., 2020; Shaffer et al., 2020; Zhou et al., 2019). Furthermore, ecological and evolutionary insights can be made through comparative genomics approaches, such as hypotheses of niche differentiation based on gene content (Camejo et al., 2017; Flowers et al., 2009; Koch et al., 2015; Oyserman et al., 2016a; Skennerton et al., 2015; Speth et al., 2012), or analyzing microdiversity and population heterogeneity (Albertsen et al., 2011; Leventhal et al., 2018).

Some of the first demonstrations of genome-resolved metagenomics are from laboratory-scale wastewater enrichment bioreactors. The first representative genome of the polyphosphate accumulating organism (PAO) '*Candidatus* Accumulibacter phosphatis' UW-1 was assembled from American and Australian enrichment bioreactors (Martín et al., 2006). Remarkably, both bioreactors happened to be enriched with similar Accumulibacter populations at the time of sequencing. Genomes of multiple *Ca.* Accumulibacter strains have since been assembled to date (Albertsen et al., 2016; Camejo et al., 2019; Flowers et al., 2013; Gao et al., 2019; Mao et al., 2014; Qiu et al., 2020; Skennerton et al., 2015), including high-quality genomes assembled from full-scale WWTPs (Law et al., 2016; Srinivasan et al., 2019), even though it remains to be isolated in pure-culture. The uncultivated bacterium *Kuenenia stuttgartiensis* (Strous et al., 2006) underlying anaerobic ammonium oxidation

(anammox) was assembled from a complex bioreactor community, and subsequently other anammox species within the *Planctomycetes* have been assembled (Lawson et al., 2017; Speth et al., 2016). The genome of the nitrite-oxidizing bacterium *'Candidatus* Nitrospira defluvii' was assembled from an activated sludge enrichment bioreactor (Lücker et al., 2010), followed by other *Nitrospira* and *Nitrospina* species (Koch et al., 2015; Lücker et al., 2013). Remarkably, complete ammonium oxidation (comammox) was discovered by two groups through enrichment of *Nitrospira inopinata* from the pipe of a deep oil exploration well and two distinct *Nitrospira* species from a recirculation aquaculture system biofilter, respectively (Daims et al., 2015; Van Kessel et al., 2015).

Before the development of automatic binning methods, reconstruction of genomes from uncultured microbial lineages within a mixed population was mostly limited to a few abundant members to study their metabolism in detail, as the assembly and curation of any single MAG relied on largely manual approaches (Tyson et al., 2004). Albertsen et al. first demonstrated that MAGs of low abundant microorganisms within an activated sludge enrichment bioreactor could be captured using differential-coverage and composition based binning (Albertsen et al., 2013). Since then, numerous automatic binning algorithms that implement differential coverage profiles in addition to sequence composition signatures have been developed, which include the programs CONCOCT (Alneberg et al., 2014), MetaBAT2 (Kang et al., 2015) and MaxBin (Wu et al., 2014).

Genome-resolved metagenomic surveys now routinely consist of aiming to assemble all representative species within a community for which medium- to high-quality MAGs can be reconstructed (Anantharaman et al., 2016; Bowers et al., 2017; Woodcroft et al., 2018). Recent examples of whole-community genome-resolved metagenomic analyses in engineered water systems include metagenomic assembly of *Ca.* Accumulibacter and 39 non-PAO "flanking" community members in a denitrifying enrichment bioreactor (Gao et al., 2019), investigating the impact of disinfection on microbial community structure and metabolism in a full-scale *drinking water distribution system* (DWDS) (Dai et al., 2019; Sevillano et al., 2020), nitrogen metabolism in drinking water reservoirs (Potgieter et al., 2020), and tracking the

microbial community through a pilot plant for potable water reuse of wastewater (Kantor et al., 2019). Representative genomes from full-scale systems have only recently been constructed through large-scale sequencing efforts and curation of system-specific MAG databases, such as the collection of 1,600 MAGs from anaerobic digestion biogas reactors (Campanaro et al., 2020) and 2,045 MAGs assembled from a collection of 114 activated sludge full-scale WWTPs (Ye et al., 2020).

The use of MAG-based methods has allowed for the recovery of a large number of genomes from the uncultivated majority, however, current methods have several significant limitations. The most significant relates to the fractionated nature of MAGs assembled from short read data, which is a consequence of the limited read length and the inability to reconstruct regions containing complex repeat structures or share relatively high similarity between sub-species or strains, such as in ribosomal RNA operons (Chen et al., 2020; Meziti et al., 2021). This results in genomes that are comprised between ten to hundreds of contigs, which can complicate the accurate estimation of genome completeness and contamination (Orakov et al., 2021). Furthermore, recovering genomes of individual strains from complex communities still remains a significant limitation of current methods, despite the highly sophisticated bioinformatics algorithms that have been developed for this problem (Olm et al., 2021; Quince et al., 2017a; Segata et al., 2012). All of these limitations are particularly acute in highly complex microbial communities such as those inhabiting full-scale activated sludge and anaerobic digester communities. According to the Minimum Information about a Metagenome-Assembled Genome (MIMAG), a high-quality MAG draft contains all three ribosomal rRNA genes, at least 18 tRNAs, and completion of >90% and contamination <5% calculated by the presence of single-copy marker genes (Bowers et al., 2017; Parks *et al.,* 2015). A recent large-scale assembly and binning effort of publicly available metagenomes in the Integrated Microbial Genomes (IMG) system recovered less than 20% high-quality MAGs in the ~52,000 MAG (Nayfach et al., 2020). Approximately 3,400 of these MAGs were recovered from engineered water systems (**Figure 4**), and demonstrate that the majority of these MAGs are of medium-quality according to MIMAG standards. Although a large portion

of these MAGs contain >90% completeness and <5% redundancy, they do not meet the high-quality standards due to missing rRNA and/or tRNA genes. As discussed below, obtaining high-quality MAGs is essential for ensuring accuracy of functional hypotheses and further validating these hypotheses with experiments. Using composite MAGs of medium-quality or low-quality can lead to inaccurate ecological inferences due to either inflating relative abundance calculations or missing genes conferring key biogeochemical transformations (Shaiber and Eren, 2019).

At the time of writing, there are several near-automated workflows available for recovery and evaluation of MAGs from metagenome data, for example, using MetaWRAP (Uritskiy *et al.*, 2018) and the MAG annotation workflow MAGPy (Stewart *et al.*, 2019). The anvi'o platform, an open source data analysis and visualization system, is another major development for enabling MAG based analyses, including multi-omics data (Eren *et al.*, 2015). Such analyses can be greatly facilitated by the use so-called containerized computational workflows, such as Snakemake (Koster and Rahmann, 2012) and Nextflow (https://www.nextflow.io), which enable portability across systems, extensibility and reproducible research practice. We also highlight the emerging role that machine learning methods have in MAG recovery and characterization, based on several recent publications (Ye et al., 2020; Nissen et al., 2021).

In summary, MAG recovery from short read metagenomic data, while currently is routinely and relatively easily applied, is still subject to many caveats and substantive limitations.

1C. Integrated Multi-omics Approaches in Water Systems

Integrated multi-omics analyses consist of applying genome-resolved metagenomics in combination with one or more additional omics-based approaches, such as metatranscriptomics, metaproteomics, and/or metabolomics (**Figure 2**). Although each integrated approach can enhance our understanding beyond obtaining genomic sequences alone, these methodologies are currently less developed than metagenomic binning

363 workflows, and may require more specific knowledge bases as well as system-specific
364 methodological development to be performed effectively. It is important to note that in the case
365 of metatranscriptomics and metaproteomics, the acquired data can be referenced directly
366 back to recovered genome sequences. However, this is not the case for metabolomic data,
367 which requires known or putative metabolic reaction pathways to be known, or predicted, in
368 order to link detected compounds with the corresponding genes that encode cognate enzymes
369 for producing, or utilizing, these compounds.

370 Therefore, in both genome-resolved metatranscriptomics and metaproteomics
371 approaches, short cDNA reads or peptide sequences are competitively mapped against a
372 collection of genomes to identify which genes are actually being expressed and to provide
373 estimates of relative expression levels (Shakya et al., 2019; Wilmes and Bond, 2009;
374 Woodcroft et al., 2018). Successfully conducting metatranscriptomics and metaproteomics is
375 substantially more complex than encountered in metagenomics. The extraction and handling
376 of RNA and protein are complicated by the fact that both molecules are far less stable than
377 DNA (Moran et al., 2013). Furthermore, because approximately 80-90% of the RNA in a given
378 cell is mostly ribosomal RNA (Westermann et al., 2012), then the remaining ~10-20% fraction
379 of the reads from total RNA sequencing are from protein-coding genes. At the whole
380 community level, this can result in noise-effected or false-negative calls for genes in all but the
381 most abundant genomes. Accordingly, rRNA-depletion procedures are best applied following
382 the extraction of total RNA (Culviner et al., 2020; He et al., 2010b; Stewart et al., 2010).
383 Collectively, these issues require complex bench workflows, requiring a high degree of
384 molecular biological knowledge and skill. At the data analysis level, because sequencing-
385 based gene expression estimates are generally relative, it can be challenging to understand
386 the relationship between changes in gene expression and fluctuations in taxon abundance.
387 Therefore, co-extraction of DNA, RNA, and/or protein from the same aliquot is recommended,
388 combined with recovery of MAGs from the DNA fraction, with the aim of providing accurate
389 reference sequences to support the analysis and interpretation of the corresponding RNA-Seq
390 data (Roume et al., 2013a, 2013b).

While in the case of AS and AD samples, biomass concentration is high enough that extraction procdures will yield adequate amounts of nucleic acid, in the case of DWDS, the low biomass concentration requires either size selection methods and/or DNA amplification procedures to sufficent concencentration and amount to initiate extraction procedures (Kantor et al., 2019; Vosloo *et al.* 2021*)*

Traditionally, metatranscriptomics analyses employ either gene-centric or genome-centric approaches, both of which depend on identifying highly, differentially, or co-expressed genes (Shakya et al., 2019). Gene expression dynamics of Accumulibacter were first elucidated using oligonucleotide microarrays (He et al., 2010a), with a more detailed metabolic reconstruction of this lineage generated using time-series RNA sequencing (Oyserman et al., 2016b), and ecological roles for co-existing Accumulibacter clades inferred through comparative genome-resolved metatranscriptomics (McDaniel and Moya et al., 2020; Wang et al., 2020). Genome-resolved metatranscriptomics has also been used to demonstrate activities of specific genes, such as a complete denitrification pathway in Accumulibacter clade IC (Camejo et al., 2019), and extracellular electron transfer activities of a wastewater-fed microbial full cell community (Ishii et al., 2013) and anammox bacteria (Shaw et al., 2020) . Expanding beyond the detailed study of a single microbial lineage, genome-resolved metatranscriptomic approaches have been applied to whole communities for which there is sufficient read coverage (Woodcroft et al., 2018) to infer putative interactions and connect gene expression patterns to overall ecosystem functioning. Lawson et al. (2017) applied genome-resolved metatranscriptomics to 17 assembled population genomes from an anammox bioreactor to infer interactions between heterotrophic and annamox bacteria. Scarborough et al. (2018) integrated genome-resolved metatranscriptomics and thermodynamic analysis to elucidate microbial interactions contributing to medium-chain fatty acid production of an anaerobic microbiome. Hao et al. (2020) reconstructed 182 MAGs from two full-scale anaerobic digestors and applied genome-resolved metatranscriptomics to explore gene expression programs in response to short-chain fatty acid stimulation.

Proteomics methods couple liquid chromatography (LC) with tandem mas-spectrometry (MS) to profile the protein content of a single organism or microbial community (Lipton et al., 2002; Wilmes and Bond, 2006). Overall, genome-resolved metaproteomics approaches as a whole have not been applied as frequently as metatranscriptomics. This is most likely due to both the technical difficulties in high-throughput mass spectrometry methodologies and development of appropriate bioinformatics pipelines for competitively mapping short peptide sequences to a collection of assembled genomes. However, landmark studies have applied proteomics to whole communities to examine the functional activities of dominant lineages (Lo et al., 2007; Ram et al., 2005; Roume et al., 2015a; Woodcroft et al., 2018). In the context of engineered water systems, shotgun proteomics has been performed on activated sludge enriched for Accumulibacter to test hypotheses about previously proposed metabolic models in EBPR, and associate them with strain variants within the Accumulibacter population (Wilmes et al., 2008). Shotgun proteomics, supported by shotgun metagenomics, has also been used to analyze the proteome associated with membrane-bound respiratory complexes in anammox bacteria (de Almeida et al., 2016).

Fully integrated studies across the genomic, transcriptomic, proteomic, and even metabolomic landscapes have only recently been produced for engineered water systems. Methodologies for sequential extraction of DNA, RNA, protein, and metabolites from a single microbial community sample have enhanced the reproducibility of integrated omics measurements (Roume et al., 2013a, 2013b). An integrated comparative ‘omic analysis using metagenomics, metatranscriptomics, and metaproteomics of the anoxic tank of a full-scale WWTP was carried out to construct community-wide metabolic networks (Roume et al., 2015b). Metabolomics paired with genomic, transcriptomic, and proteomic data was used to understand lipid accumulation in WWTPs and the role of *‘Candidatus* Microthrix parvicella’ (Muller et al., 2014). Genome-resolved metagenomic, metatranscriptomic, and metaproteomic approaches were applied to a partial-nitration anammox lab-scale bioreactor to explore operational parameters and microbial interactions affecting nitrogen removal (Wang et al., 2019). Recently, an integrated time-series genome-resolved metagenomics,

metatranscriptomics, metaproteomics, and metabolomics investigation of full-scale WWTP microbial community recovered 1,364 MAGs and demonstrated patterns of resistance and resilience to natural environmental perturbations (Herold et al., 2020). As technologies and bioinformatics algorithms improve, integrated multi-omics approaches have the potential to enable powerful ecological insights into engineered water microbiomes and create the foundation for more predictable and controllable systems (Muller et al., 2013).

2. PAIRING OMICS-DRIVEN INSIGHTS WITH MODELING APPROACHES

Obtaining genome sequences for uncultivated microbial lineages and integrating spatiotemporal metatranscriptomic and metaproteomic data has become increasingly commonplace in research conducted by microbial ecologists and water engineers. Looking ahead, effective prediction and control of microbial communities for desired outcomes in engineered water systems is likely to rely in part on the ability to integrate vast amounts of omics data into process modeling frameworks. In this section, we describe two overarching aims of broad modeling approaches used to *1*) interrogate metabolic networks and *2*) predict the temporal dynamics of complex communities. For metabolic modeling approaches, we summarize how emerging techniques in metabolomics and stable isotope probing are being used in engineered water systems. For modeling population dynamics, we briefly review how 16S rRNA gene sequencing approaches have been used for time-series forecasting in water engineering, and how these studies can be strengthened by applying genome-resolved multi-omics approaches.

2A. Metabolic Modeling with Pathway Flux, Stable Isotope Probing and Metabolomics

The metabolic capabilities of a microorganism depend on the available reaction networks, putative intracellular storage compounds, and environmental conditions. The past few decades have seen major developments in metabolic modeling methods, fueled by more quantitative data from the application of different omics approaches as well as increased computational power. These approaches have ranged from “simple” stoichiometric models to

complex spatiotemporal, multi-scale simulations of ecosystems (Bauer et al., 2017). Annotation and generation of genome-scale stoichiometric models is becoming a more automated process with different tools available for binning, functional annotation, and metabolic model generation (Hamilton and Reed, 2014; Machado et al., 2018). Overall, these predictions are based on 1) the metabolic reaction networks of different microorganisms, which can be derived from annotated genomes, at least within the limits of known biochemistry (Feist et al., 2009; Hamilton and Reed, 2012) 2) expression levels of the required enzymes and their activities (Hamilton et al., 2017; Molenaar et al., 2009), and 3) the kinetics of the extracellular environment and metabolite concentrations (Herold et al., 2020).

The exploration of metabolic network functionality has been enabled by stoichiometric modeling approaches, building on the availability of whole-genome sequences, namely flux balance analysis (FBA) (Orth, Thiele and Palsson, 2010) or flux variability analysis (FVA) toolkits (Gudmundsson and Thiele, 2010). These approaches were first developed for steady-state systems and were able to predict biological phenomena such as overflow metabolism in a single organism (Basan et al., 2015), as well as cross-feeding consortia (Stolyar et al., 2017; Carlson et al., 2018). Extensions to these approaches have been made to more dynamic and cyclic conditions to correctly predict storage accumulation and depletion cycles in different organisms (Reimers et al., 2017). For example, applications of FBA and FVA in engineered water systems include the study of *Ca.* Accumulibacter to highlight the flexibility of anaerobic metabolism due to intracellular accumulation of different polymers (Guedes da Silva et al., 2020). Genome-scale models can also be used to understand the gain insight into thermodynamic constraints that may pattern metabolic activity, as seen by an analysis of genome-scale metabolic network reconstructions of an anaerobic syntrophic coculture formed by the bacterium *Syntrophobacter fumaroxidans* and the methanogenic archaeon *Methanospirillum hungatei* (Hamilton et al., 2015). Weinrich et al. (2019) developed a framework to incorporate FBA into the IWA Anaerobic Digester Model No.1 to predict thousands of intracellular metabolite fluxes of individual species in anaerobic digesters, providing an opportunity to couple process modeling with multi-omics datasets. Integrated

omics approaches can also aid in improving genome-scale metabolic models and thermodynamic analyses. Scarborough et al. used genomic and transcriptomic data to understand interactions and substrates of an anaerobic microbiome producing medium-chain fatty acids (Scarborough et al., 2018). They subsequently used time-series transcriptomic data and metabolomics to understand the environmental conditions favoring medium-chain fatty acid production (Scarborough et al., 2020b) and construct unicellular and guild-based metabolic models of the community (Scarborough et al., 2020a).

Stable isotope probing (SIP) and tracing approaches have been applied to highly enriched and complex microbial communities to identify metabolically active populations (**Figure 5**). Metabolically active populations can be further taxonomically and functionally characterized by pairing heavy labelled ^{13}C isotopes with metagenomic sequencing (Dumont and Murrell, 2005; Radajewski et al., 2000) or metaproteomic sequencing (Jehmlich et al., 2008; Von Bergen et al., 2013). Incorporation of the ^{13}C isotope into a given carbon source then allows for ^{12}C and ^{13}C labelled fractions to be resolved by density-gradient centrifugation (for DNA/RNA-SIP) and separately characterized by amplicon or shotgun metagenomic sequencing approaches (Chemerys et al., 2014; Chen and Murrell, 2010; Coyotzi et al., 2016; Neufeld et al., 2007; Saidi-Mehrabad et al., 2013), or by mass-spectrometry approaches for protein-SIP (Sachsenberg et al., 2015). Genome-resolved DNA-SIP approaches can be achieved by exploiting abundance patterns between ^{12}C and ^{13}C density fractions and applying differential coverage binning to metagenomic samples (**Figure 5**). Genome-centric SIP approaches were demonstrated to elucidate populations that could degrade oleate (Ziels et al., 2018) and butyrate (Ziels et al., 2019) in anaerobic digesters. Insights into novel syntrophic acetate oxidizers in full-scale anaerobic digesters were also obtained by coupling genome-resolved metagenomics with metaproteomics-SIP (Mosbæk et al., 2016).

Isotopic tracers can also be used in metabolic flux analyses to measure the *in vivo* flux of metabolites in individual organisms (Ando and García Martín, 2019; Matsuoka and Shimizu, 2014) or cocultures (Gebreselassie and Antoniewicz, 2015). Quantifying metabolic flux changes under different environmental conditions using isotopic tracers can provide insights

into metabolic regulation (Antoniewicz, 2015; Long and Antoniewicz, 2019; Nielsen, 2003). Lawson et al. used a combination of time-series ^{13}C and ^{2}H isotopes, metabolomics, and isotopically nonstationary metabolite flux analysis to resolve the central carbon metabolism of a highly enriched anammox bioreactor community (Lawson et al., 2020). Although metabolic flux analyses are difficult to perform for complex microbial communities, advances in assigning individual peptides and metabolites to a given species has started to pave the way for metafluxomics (Beyß et al., 2019; Ghosh et al., 2014).

Recent advances in high-throughput mass spectrometry-based technologies have allowed for metabolomics approaches to shift from targeted characterization of known metabolites to untargeted surveys of numerous metabolites and other small molecules (Schrimpe-Rutledge et al., 2016). Although both targeted and untargeted strategies have their shortcomings and challenges, untargeted approaches enable discovery-based studies that can be subsequently tested with targeted approaches (Cajka and Fiehn, 2016). Metabolomics can be combined with other multi-omics and modeling approaches to compare results from the level of gene regulation to metabolite production of a given system or process (Herold et al., 2020; Lawson et al., 2020; Muller et al., 2014; Roume et al., 2015a; Scarborough et al., 2020a). Recent advances in methods for simultaneous, nondestructive extraction of DNA, RNA, and metabolites from a given sample have paved the way for these types of intrasample datasets (Roume et al., 2013a, 2013b). These multi-omics approaches integrated with metabolic modeling approaches can greatly enhance our ability to predict the environmental conditions that are favorable for desired outcomes in engineered water systems.

2B. Genome-Resolved Omics Applied to Time-Series

Time-series data have been applied to numerous ecosystems to unravel the spatiotemporal dynamics of microbial communities in the context of environmental parameters (Shade et al., 2013). For example, longitudinal studies have been used to identify marine seasonal microbial community dynamics (Gilbert et al., 2012), community succession patterns in the developing infant gut (Palmer et al., 2007), and characterize patterns of microbial community recovery

and resilience after ecosystem disturbances (Shade et al., 2012). Microbial community time-series have also been used to address important questions in engineered water systems such as community assembly in full-scale wastewater treatment plants (Griffin and Wells, 2017; Lee et al., 2015), population dynamics and immigration of anaerobic digestion communities (Griffin et al., 1998; Kirkegaard et al., 2017), and monitoring bacterial community migration dynamics in drinking water distribution systems (Boers et al., 2018; Pinto et al., 2014). However, most time-series resolved data have been generated using 16S rRNA gene amplicon sequencing surveys, which have significant shortcomings. Assigning metabolic functions to 16S rRNA gene based OTUs is not entirely intuitive for most guilds, such as heterotrophic bacteria that span diverse phyla (Frigon and Wells, 2019; Marques et al., 2017). In addition to some taxa in engineered water systems remaining unclassified, 16S rRNA gene based OTUs usually cannot be resolved more specifically than the genus-level (Kirkegaard et al., 2017).

Although long-read sequencing technologies (discussed below) can improve the resolution of resulting 16S rRNA gene based OTUs (Karst et al., 2021; Nierychlo et al., 2019), time-series data integrating metagenomics and integrated genome-resolved omics methods can be a powerful approach (Faust et al., 2015). Metagenomes collected over time can be used to improve binning efforts based on differential coverage profiles (Kang et al., 2015), and subsequently identify metabolic guilds and abundance patterns (Linz et al., 2018). If full-length 16S rRNA gene sequences are retrieved from assembled MAGs, defined microbial guilds based on metabolic reconstructions can be connected to 16S rRNA amplicon sequencing surveys of a given system (Nierychlo et al., 2019; Petriglieri et al., 2020; Singleton et al., 2020). Pérez et al. assembled 173 MAGs from a full-scale WWTP and assessed associated functional potential and the dynamics of rRNA gene operons during operational disturbances (Pérez et al., 2019). In conjunction with 16S rRNA amplicon sequencing, assembled MAGs were used to characterize functional profiles between operational periods (Pérez et al., 2019). Integrated genome-resolved omics approaches can also be applied in longitudinal analyses to understand community dynamics and functional profiles. Time-resolved omics approaches

were recently applied to a full-scale wastewater treatment plant and demonstrate the power of fully integrated omics methodologies (Herold et al., 2020; Martínez Arbas et al., 2021).

Other promising developments have combined time-series omics data with biogeochemical models to predict the fate of nutrients in environmental systems (Louca et al., 2016; Reed et al., 2014). Such ecosystem level models have been constructed based on the abundance of genes in metagenomes (Reed et al., 2014), as well as transcript and protein abundances (Louca et al., 2016). The underlying model structure of such gene-centric modeling approaches is amenable to their adaptation into existing water engineering models (e.g. IWA, ASM, and ADM). However, due to the fact that omics data are relative measurement, novel methods are needed to accurately infer relative or absolute taxa abundances so the stoichiometry can be incorporated within process models. While some recent methods show promise, such as using internal quantitation standards for metagenomes (Hardwick et al., 2018), this task is not trivial due to differences in cell size, composition, biomass concentration, and complexity of biomass morphology.

3. EMERGING TECHNOLOGIES IN OMICS AND WATER ENGINEERING

Over the last few years, advances in sequencing technologies and novel methodologies have emerged that have direct implications for improving genome-resolved metagenomic insights. Although there have been various novel, cultivation-independent inventions for broadly characterizing microbial communities, we focus on two fields that have exhibited tremendous growth and opportunity for leveraging multi-omics approaches specifically in engineered water systems. First, we summarize how improvements in the quality and substantial decrease in costs of long-read sequencing technologies have allowed for obtaining complete, closed, and high-quality reference genomes of microorganisms from mixed communities. We then highlight emerging cell-sorting techniques that can be applied in conjunction with multi-omics for novel functional insights.

3A. High-Quality Genomes Enabled by Advances in Sequencing Technologies

Although next-generation sequencing technologies have greatly enhanced high-throughput surveys of microbial communities, the emergence of so-called “third-generation” sequencing technologies using single molecule sequencing to produce longer reads are poised to revolutionize *de novo* assembly of complex genomes and metagenomes (Lee et al., 2016). Primarily, Pacific Biosciences (PacBio) and Oxford Nanopore Technologies (ONT) have emerged as the main technologies for long-read metagenomic surveys of microbial communities (Arumugam et al., 2021, 2019; Bertrand et al., 2019; Moss et al., 2020; Singleton et al., 2021; Somerville et al., 2019; Stewart et al., 2019). The greatly enhanced read length (routinely greater than 10k bp) presents substantial advantages in the process of genome assembly, by permitting the successful reconstruction of genome regions containing complex repeats, such as multiple intragenomic rRNA operons. Such repeat regions typically contribute to the high degree of fragmentation observed in assemblies constructed with short read data. In the case of gDNA sequencing from culture isolates, the use of long read technologies will typically result in the generation of a complete closed genome, represented a single, continuous sequence construct (Frank et al., 2018).

The major disadvantage of current long read methods relates to reduced sequence quality compared to short read sequencing, with error rates cited as being as high as 10-30% (Amarasinghe et al., 2020), compared to short read sequencing which holds as error rate of no more than 1%. Rapid progress in sequencing chemistry has led to reduction in these error rates, with some reports of error rates <5% error using ONT sequencing and <1% with PacBio (as discussed by Amarasinghe et al., 2020). The impact of high error rates can readily be seen in protein coding sequences, where the occurrence of frame-shift errors (insertion/deletion, or indels) can be up to 80 times higher than that observed in Illumina sequences (Huson et al., 2018). Mitigating the impact of error rates requires the application of complex data correction procedures (Morisse et al., 2020), either self-correction of the long read data themselves, approaches that make use of complementary higher-accuracy short reads, or correction via the use of sequence alignment to reference sequence databases (Huson et al., 2018). At the time of writing, the development of base-calling methods and new error-correction procedures

641 is a highly active research topic (Morisse et al., 2019). Several datasets also exist, incuding
642 mock communities, which provide co-assayed reference data from both long and short read
643 sequencing technologies, and which can be used as comparative reference data sets for
644 comparing the two different modalities (Sevim et al., 2019; Nicholls et al., 2019; Sommerville
645 et al., 2019).

646 In the context of water engineering, Arumugam and colleagues (Arumugam et al.,
647 2021, 2019) demonstrated the recovery of complete genomes (single chromosome), and in
648 some cases, circularized, closed genomes, of the most abundant community members from
649 an enrichment bioreactor (22 in total, including key PAO, GAO, and filamentous species,
650 among others; of which 10/22 were circularized). Recently Singleton et al. analyzed
651 approximately 1Tbp of Nanopore data generated from 23 full scale WWTPs in Denmark, and
652 obtained 1,083 high quality MAGs from 581 distinct prokaryotic species, including 57 with
653 circularized closed genomes (Singleton et al., 2021). Notwithstanding these advances, there
654 remain many emerging problems that will need attention with long read MAG recovery, such
655 as the need for long read specific binning procedures (Singleton et al., 2021) and the impact
656 of sequencing error rate on coding sequence quality (Arumugam et al., 2020).

657 In addition to technology advancements that directly allow for longer-reads from single
658 molecules, novel, synthetic long-read approaches such as chromosome conformation capture
659 (also referred to proximity ligation), linked-reads, and optical mapping (Amarasinghe et al.,
660 2020) also have enormous potential for applications in water systems. Chromosome
661 conformation capture (3C) methods apply formaldehyde cross-links prior to DNA lysis and
662 extraction to create interaction points between nearby loci (Dekker et al., 2002). Crosslinked
663 chromatin is then digested with specific restriction enzymes to obtain long-range interaction
664 pairs, and used to analyze the contact frequencies of these pairs (Lieberman-Aiden et al.,
665 2009; Sati and Cavalli, 2017). Originally, variants of 3C technologies were developed to study
666 the three-dimensional organization of chromosomes and genome structures (Hsieh et al.,
667 2015; Rao et al., 2014). Recently, this method has been adapted for generating genomes from
668 semi-complex and complex microbial communities, such as assembling a yeast hybrid from a

beer sample, population genomes from the human gut, and used in addition to standard Illumina shotgun sequencing to obtain ~900 genomes from the cow rumen (Burton et al., 2014; Marbouty et al., 2014; Press et al., 2017; Smukowski Heil et al., 2018; Stewart et al., 2018). Since this method produces linkages between genetic content within the same cell, this approach can also produce associations between antibiotic resistance genes (ARGs), plasmids, and mobile elements to the host genome with fidelity. This was recently demonstrated in a microbial community from a full-scale WWTP to link ARGs and plasmids to specific populations (Stalder et al., 2019), while also highlighting the challenges associated with implementing Hi-C approaches in complex and poorly studied microbial communities.

Although this review mainly covers genome-resolved approaches, long-reads have also been applied to improve amplicon sequencing surveys. Briefly, these high-throughput pipelines generate thousands to millions of full-length 16S rRNA genes (or any other amplicon of choice) from mixed microbial communities without primer-bias (Karst et al., 2021, 2018). The recently updated release of the MiDAS Field Guide with full-length 16S rRNA references allows for species-level classifications of microorganisms in activated sludge and anaerobic digestion systems (Nierychlo et al., 2019). Singleton et al. also demonstrated the ability to connect structure to function with fidelity by comparing full-length 16S rRNA sequences from large-scale microbial surveys to corresponding high-quality MAGs (Singleton et al., 2020). Full-length 16S rRNA sequences from high-quality MAGs were also used in Singleton et al. 2020 to construct novel FISH probes to visualize the dynamics of specific and understudied microorganisms in full-scale WWTPs (Singleton et al., 2020). Advances in long-read sequencing technologies will allow for the rapid recovery of high-quality genomes from complex microbial communities to use in conjunction with detailed experiments (**Figure 6**).

3B. Omics Approaches Enabled through Cell-Sorting Techniques

Cell-sorting methodologies involve the separation of cells from a complex sample into simpler subsets for various downstream purposes. Sorting can be accomplished through various strategies, such as fluorescently or physically tagging cells with specific probes or dyes,

randomly dividing groups of cells, or scanning cells for the presence of specific polymers or storage products (Malmstrom and Eloe-Fadrosh, 2019). Advances in cell-sorting technologies have allowed for both fundamental discoveries and applied practices in engineered water systems. Flow cytometry has been commonly used to monitor microbial communities in DWDS (Besmer et al., 2014; Douterelo et al., 2014), sometimes in conjunction with next-generation sequencing approaches (Prest et al., 2014). In relation to omics-based approaches, cell-sorting methods allow for downstream applications such as sequencing of simpler “mini-metagenomes” of a complex community, performing targeted cultivation of specific microbial lineages, and characterizing individual microbial cells exhibiting specific traits or physiologies (**Figure 6**).

Complex microbial communities can be subjected to cell-sorting techniques prior to DNA extraction and sequencing, leading to sequencing a “mini-metagenome” (Ji et al., 2017). This could entail randomly dividing cells by microfluidic sorting (Yu et al., 2017) or nonrandomly by fluorescently-activated cell sorting (FACS) (Schulz et al., 2018). Single amplified genomes (SAGs) are obtained through non-specific staining prior to sorting and screened for sequencing (Clingenpeel et al., 2014; Rinke et al., 2013). However, genomes from single-cell sequencing efforts are usually incomplete, contain high amounts of contamination, and can be expensive to obtain for low abundant microorganisms (Rinke et al., 2013; Swan et al., 2011; Xu and Zhao, 2018). Targeted approaches employ FISH probes to stain specific microbial lineages prior to FACS (Kalyuzhnaya et al., 2006), which can also allow access to low abundant microorganisms (Podar et al., 2007; Tan et al., 2019; Yilmaz et al., 2010). Improvements in FISH-based methods for staining specific microbial clades have led to higher quality genomes reconstructed from these efforts (Grieb et al., 2020). Intriguingly, a reverse genomics-based approach implementing engineered antibodies to capture specific microorganisms from a complex community was applied to culture members of the Saccharibacteria (TM7) (Cross et al., 2019). Live cell sorting was applied to marine nitrite-oxidizing enrichments to isolate novel *Nitrospinae* strains and investigate their nitrite affinity and metabolic differences (Mueller et al., 2020). Targeted cell sorting through fluorescent

725 labeling combined with single-cell genomics was used to enrich for low abundant Chloroflexi
726 species from a full-scale wastewater treatment plant (Dam et al., 2020).
727 In addition to targeting specific microorganisms based on phylogenetic identity,
728 microbial populations can be screened and selected based on specific metabolic activities.
729 Raman microspectroscopy methods quantify the scattering of light given off by chemical
730 bonds to generate a spectral profile of specific biomarkers at the single-cell level (Huang et
731 al., 2004; Neufeld and Murrell, 2007). Raman spectroscopy can be combined with stable
732 isotope probing (SIP) and FISH methods to connect the phylogenetic identity of cells exhibiting
733 specific metabolic activities (Huang et al., 2007). These methods have been used to retrieve
734 cells based on specific functions before sequencing (Lee et al., 2020) or confirming metabolic
735 activities of microorganisms based on genome sequences (Fernando et al., 2019). Raman
736 cell-sorting in conjunction with mini-metagenomics was applied to a mouse gut microbial
737 community to identify commensals that incorporate mucus-derived monosaccharides and
738 rationally design a probiotic microbial consortia (Pereira et al., 2020). In engineered water
739 systems, Raman microspectroscopy-based methods have been particularly used to identify
740 polyphosphate accumulating organisms (PAOs) based on quantifying intracellular storages of
741 polyphosphate (Fernando et al., 2019). In addition to the model PAO *Ca.* Accumulibacter,
742 members of the actinobacterial genus *Tetrasphaera*, multiple species of *Dechloromonas,* the
743 filamentous microorganism *Ca.* Microthrix, and the novel *Ca.* Methylophosphatis have been
744 shown to incorporate polyphosphate intracellularly using Raman microspectroscopy imaging
745 and quantification methods (Fernando et al., 2019; Petriglieri et al., 2021, 2020; Singleton et
746 al., 2020). Enzymatic tagging using click-chemistry based approaches followed by cell sorting
747 can be used to sequester low abundant populations for targeted metagenomic analyses
748 (Sakoula et al., 2021). *Bioorthogonal non-canonical amino acid tagging* (BONCAT) is one
749 such substrate analogue probing technique that labels newly translated proteins via azide-
750 alkyne click-chemistry, and can be used to selectively sort active microorganisms from
751 complex communities and environments (Couradeau et al., 2019; Hatzenpichler et al., 2016;
752 Reichart et al., 2020). Advances in cell-sorting methods provide powerful approaches to apply

identity- or function-based tagging to specific microbial populations in a complex community such as engineered water systems. These approaches can be used to construct mini-communities or synthetic consortia designed to perform a desired function, allowing for more controllable and predictable dynamics and outputs (Lawson et al., 2019; Pereira et al., 2020).

4. CONCLUSIONS

In the release of *Microbial Ecology of Activated Sludge* in 2010 (Seviour and Nielsen, 2010), with extensive field updates to the 1999 publication of *The Microbiology of Activated Sludge* (Seviour and Blackall, 1999), the editors noted that the greatest future challenge would be "to communicate with engineers to convince them that all this elegant science can be applied productively to improve how they design and operate their activated sludge systems" (Seviour and Nielsen, 2010). In the decade since, numerous breakthroughs and technological advancements in water-related microbiome science as a whole have stemmed from the desire to unravel the ecology and evolution of the microorganisms underpinning these systems. Clearly, the water engineering community is more convinced than ever that state-of-the-art molecular and omics tools can provide valuable insights into the microbiology of their systems. However, applying these insights to improve the design and operation of engineered water systems remains a grand challenge. Ultimately, successfully harnessing engineered water microbiomes for improved process design cannot be accomplished without effectively bridging engineering principles with microbial ecology theory (Lawson et al., 2019). Advances in genome-resolved multi-omics approaches provide an exciting new avenue for water engineers and microbial ecologists to tackle these interdisciplinary grand challenges.

ACKNOWLEDGEMENTS

This review resulted from materials and discussions held during the *Prospects for Multi-Omics Methods in Water Engineering* pre-conference workshop of the 8th International Water Association Microbial Ecology and Water Engineering Conference (MEWE2019, 17-20 November 2019, Hiroshima Japan). We thank Akiyoshi Ohashi, Futoshi Kurisu, and Kengo

Kubota for their support as part of the organization of MEWE2019, and the Japanese Society of Water Environment (JSWE) for financial support of the event. We thank Emilie Muller (University of Strasbourg/CNRS) and Yu Ke (Peking University, Beijing) for preliminary discussions on the workshop content.

REFERENCES

Albertsen, M., Benedicte, L., Hansen, S., Saunders, A.M., Nielsen, H., Lehmann Nielsen, K., 2011. A metagenome of a full-scale microbial community carrying out enhanced biological phosphorus removal. ISME J. 6, 1094–1106. https://doi.org/10.1038/ismej.2011.176

Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K.L., Tyson, G.W., Nielsen, P.H., 2013. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. Nat. Biotechnol. 31, 533–538. https://doi.org/10.1038/nbt.2579

Albertsen, M., McIlroy, S.J., Stokholm-Bjerregaard, M., Karst, S.M., Nielsen, P.H., 2016. "Candidatus Propionivibrio aalborgensis": A novel glycogen accumulating organism abundant in full-scale enhanced biological phosphorus removal plants. Front. Microbiol. 7, 1033. https://doi.org/10.3389/fmicb.2016.01033

Alivisatos, A.P., Blaser, M.J., Brodie, E.L., Chun, M., Dangl, J.L., Donohue, T.J., Dorrestein, P.C., Gilbert, J.A., Green, J.L., Jansson, J.K., Knight, R., Maxon, M.E., McFall-Ngai, M.J., Miller, J.F., Pollard, K.S., Ruby, E.G., Taha, S.A., 2015. A unified initiative to harness Earth's microbiomes: Transition from description to causality and engineering. Science (80-.). https://doi.org/10.1126/science.aac8480

Alneberg, J., Bjarnason, B.S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U.Z., Lahti, L., Loman, N.J., Andersson, A.F., Quince, C., 2014. Binning metagenomic contigs by coverage and composition. Nat. Methods 11, 1144–1146. https://doi.org/10.1038/nmeth.3103

AMANN, R.I., 1995. Fluorescently labelled, rRNA-targeted oligonucleotide probes in the study of microbial ecology. Mol. Ecol. https://doi.org/10.1111/j.1365-294X.1995.tb00255.x

809 Amann, R.I., Ludwig, W., Schleifer, K.H., 1995. Phylogenetic identification and in situ detection
810 of individual microbial cells without cultivation. Microbiol. Rev.
811 https://doi.org/10.1128/mmbr.59.1.143-169.1995
812 Amarasinghe, S.L., Su, S., Dong, X., Zappia, L., Ritchie, M.E., Gouil, Q., 2020. Opportunities
813 and challenges in long-read sequencing data analysis. Genome Biol.
814 https://doi.org/10.1186/s13059-020-1935-5
815 Anantharaman, K., Brown, C.T., Hug, L.A., Sharon, I., Castelle, C.J., Probst, A.J., Thomas,
816 B.C., Singh, A., Wilkins, M.J., Karaoz, U., Brodie, E.L., Williams, K.H., Hubbard, S.S.,
817 Banfield, J.F., 2016. Thousands of microbial genomes shed light on interconnected
818 biogeochemical processes in an aquifer system. Nat. Commun. 7, 13219.
819 https://doi.org/10.1038/ncomms13219
820 Anantharaman, K., Hausmann, B., Jungbluth, S.P., Kantor, R.S., Lavy, A., Warren, L.A.,
821 Rappé, M.S., Pester, M., Loy, A., Thomas, B.C., Banfield, J.F., 2018. Expanded diversity
822 of microbial groups that shape the dissimilatory sulfur cycle. ISME J. 12, 1715–1728.
823 https://doi.org/10.1038/s41396-018-0078-0
824 Ando, D., García Martín, H., 2019. Genome-scale 13C fluxomics modeling for metabolic
825 engineering of Saccharomyces cerevisiae, in: Methods in Molecular Biology. Humana
826 Press Inc., pp. 317–345. https://doi.org/10.1007/978-1-4939-8757-3_19
827 Angelidaki, I., Karakashev, D., Batstone, D.J., Plugge, C.M., Stams, A.J.M., 2011.
828 Biomethanation and its potential, in: Methods in Enzymology. Academic Press Inc., pp.
829 327–351. https://doi.org/10.1016/B978-0-12-385112-3.00016-0
830 Antoniewicz, M.R., 2015. Methods and advances in metabolic flux analysis: a mini-review. J.
831 Ind. Microbiol. Biotechnol. 42, 317–325. https://doi.org/10.1007/s10295-015-1585-x
832 Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S., Ogata, H.,
833 2019. KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive
834 score threshold. bioRxiv 602110. https://doi.org/10.1101/602110
835 Ardern, E., Lockett, W.T., 1914. Experiments on the oxidation of sewage without the aid of
836 filters. J. Soc. Chem. Ind. 33, 523–539. https://doi.org/10.1002/jctb.5000331005

837 Arumugam, K., Bağcl, C., Bessarab, I., Beier, S., Buchfink, B., Górska, A., Qiu, G., Huson,
838 D.H., Williams, R.B.H., 2019. Annotated bacterial chromosomes from frame-shift-
839 corrected long-read metagenomic data. Microbiome 7, 61.
840 https://doi.org/10.1186/s40168-019-0665-y
841 Arumugam, K., Bessarab, I., Haryono, M.A.S., Liu, X., Zuniga–Montanez, R.E., Roy, S., Qiu,
842 G., Drautz–Moses, D.I., Law, Y.Y., Wuertz, S., Lauro, F.M., Huson, D.H., Williams,
843 R.B.H., 2021. Recovery of complete genomes and non-chromosomal replicons from
844 activated sludge enrichment microbial communities with long read metagenome
845 sequencing. npj Biofilms Microbiomes 7, 1–13. https://doi.org/10.1038/s41522-021-
846 00196-6
847 Arumugam, K., Bessarab, I., Haryono, M., Liu, X., Zuniga-Montanez, R., Roy, S., Qiu, G.,
848 Drautz-Moses, D., Law, Y.Y., Wuertz, S., Lauro, F., Huson, D., Williams, R., 2020.
849 Analysis procedures for assessing recovery of high quality, complete, closed genomes
850 from Nanopore long read metagenome sequencing. Asian Sch. Environ. 639798.
851 https://doi.org/10.1101/2020.03.12.974238
852 Basan, M., Hui, S., Okano, H., Zhang, Z., Shen, Y., Williamson, J.R., Hwa, T., 2015. Overflow
853 metabolism in Escherichia coli results from efficient proteome allocation. Nature 528, 99–
854 104. https://doi.org/10.1038/nature15765
855 Bauer, E., Zimmermann, J., Baldini, F., Thiele, I., Kaleta, C., 2017. BacArena: Individual-
856 based metabolic modeling of heterogeneous microbes in complex communities. PLoS
857 Comput. Biol. 13, e1005544. https://doi.org/10.1371/journal.pcbi.1005544
858 Beja, O., Aravind, L., Koonin, E. V., Suzuki, M.T., Hadd, A., Nguyen, L.P., Jovanovich, S.B.,
859 Gates, C.M., Feldman, R.A., Spudich, J.L., Spudich, E.N., DeLong, E.F., 2000. Bacterial
860 rhodopsin: Evidence for a new type of phototrophy in the sea. Science (80-.). 289, 1902–
861 1906. https://doi.org/10.1126/science.289.5486.1902
862 Berry, D., Xi, C., Raskin, L., 2006. Microbial ecology of drinking water distribution systems.
863 Curr. Opin. Biotechnol. https://doi.org/10.1016/j.copbio.2006.05.007
864 Bertrand, D., Shaw, J., Kalathiyappan, M., Ng, A.H.Q., Kumar, M.S., Li, C., Dvornicic, M.,

865 Soldo, J.P., Koh, J.Y., Tong, C., Ng, O.T., Barkham, T., Young, B., Marimuthu, K., Chng,
866 K.R., Sikic, M., Nagarajan, N., 2019. Hybrid metagenomic assembly enables high-
867 resolution analysis of resistance determinants and mobile elements in human
868 microbiomes. Nat. Biotechnol. 37, 937–944. https://doi.org/10.1038/s41587-019-0191-2
869 Besmer, M.D., Weissbrodt, D.G., Kratochvil, B.E., Sigrist, J.A., Weyland, M.S., Hammes, F.,
870 2014. The feasibility of automated online flow cytometry for in-situ monitoring of microbial
871 dynamics in aquatic ecosystems. Front. Microbiol. 5, 265.
872 https://doi.org/10.3389/fmicb.2014.00265
873 Beyß, M., Azzouzi, S., Weitzel, M., Wiechert, W., Nöh, K., 2019. The design of FluxML: A
874 universal modeling language for 13C metabolic flux analysis. Front. Microbiol. 10, 1022.
875 https://doi.org/10.3389/fmicb.2019.01022
876 Boers, S.A., Prest, E.I., Taučer-Kapteijn, M., Knezev, A., Schaap, P.G., Hays, J.P., Jansen,
877 R., 2018. Monitoring of microbial dynamics in a drinking water distribution system using
878 the culture-free, user-friendly, MYcrobiota platform. Sci. Rep. 8, 1–8.
879 https://doi.org/10.1038/s41598-018-32987-x
880 Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A.,
881 Alexander, H., Alm, E.J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J.E., Bittinger, K.,
882 Brejnrod, A., Brislawn, C.J., Brown, C.T., Callahan, B.J., Caraballo-Rodríguez, A.M.,
883 Chase, J., Cope, E.K., Da Silva, R., Diener, C., Dorrestein, P.C., Douglas, G.M., Durall,
884 D.M., Duvallet, C., Edwardson, C.F., Ernst, M., Estaki, M., Fouquier, J., Gauglitz, J.M.,
885 Gibbons, S.M., Gibson, D.L., Gonzalez, A., Gorlick, K., Guo, J., Hillmann, B., Holmes,
886 S., Holste, H., Huttenhower, C., Huttley, G.A., Janssen, S., Jarmusch, A.K., Jiang, L.,
887 Kaehler, B.D., Kang, K. Bin, Keefe, C.R., Keim, P., Kelley, S.T., Knights, D., Koester, I.,
888 Kosciolek, T., Kreps, J., Langille, M.G.I., Lee, J., Ley, R., Liu, Y.X., Loftfield, E.,
889 Lozupone, C., Maher, M., Marotz, C., Martin, B.D., McDonald, D., McIver, L.J., Melnik,
890 A. V., Metcalf, J.L., Morgan, S.C., Morton, J.T., Naimey, A.T., Navas-Molina, J.A.,
891 Nothias, L.F., Orchanian, S.B., Pearson, T., Peoples, S.L., Petras, D., Preuss, M.L.,
892 Pruesse, E., Rasmussen, L.B., Rivers, A., Robeson, M.S., Rosenthal, P., Segata, N.,

Shaffer, M., Shiffer, A., Sinha, R., Song, S.J., Spear, J.R., Swafford, A.D., Thompson, L.R., Torres, P.J., Trinh, P., Tripathi, A., Turnbaugh, P.J., Ul-Hasan, S., van der Hooft, J.J.J., Vargas, F., Vázquez-Baeza, Y., Vogtmann, E., von Hippel, M., Walters, W., Wan, Y., Wang, M., Warren, J., Weber, K.C., Williamson, C.H.D., Willis, A.D., Xu, Z.Z., Zaneveld, J.R., Zhang, Y., Zhu, Q., Knight, R., Caporaso, J.G., 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat. Biotechnol. https://doi.org/10.1038/s41587-019-0209-9

Bornemann, T., Esser, S., Stach, T., Burg, T., Probst, A., 2020. uBin – a manual refining tool for metagenomic bins designed for educational purposes. bioRxiv 2020.07.15.204776. https://doi.org/10.1101/2020.07.15.204776

Bowers, R.M., Kyrpides, N.C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T.B.K., Schulz, F., Jarett, J., Rivers, A.R., Eloe-Fadrosh, E.A., Tringe, S.G., Ivanova, N.N., Copeland, A., Clum, A., Becraft, E.D., Malmstrom, R.R., Birren, B., Podar, M., Bork, P., Weinstock, G.M., Garrity, G.M., Dodsworth, J.A., Yooseph, S., Sutton, G., Glöckner, F.O., Gilbert, J.A., Nelson, W.C., Hallam, S.J., Jungbluth, S.P., Ettema, T.J.G., Tighe, S., Konstantinidis, K.T., Liu, W.-T., Baker, B.J., Rattei, T., Eisen, J.A., Hedlund, B., McMahon, K.D., Fierer, N., Knight, R., Finn, R., Cochrane, G., Karsch-Mizrachi, I., Tyson, G.W., Rinke, C., Kyrpides, N.C., Schriml, L., Garrity, G.M., Hugenholtz, P., Sutton, G., Yilmaz, P., Meyer, F., Glöckner, F.O., Gilbert, J.A., Knight, R., Finn, R., Cochrane, G., Karsch-Mizrachi, I., Lapidus, A., Meyer, F., Yilmaz, P., Parks, D.H., Eren, A.M., Schriml, L., Banfield, J.F., Hugenholtz, P., Woyke, T., 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat. Biotechnol. 35, 725–731. https://doi.org/10.1038/nbt.3893

Brady, S.F., Clardy, J., 2000. Long-chain N-acyl amino acid antibiotics isolated from heterologously expressed environmental DNA [20]. J. Am. Chem. Soc. https://doi.org/10.1021/ja002990u

Burton, J.N., Liachko, I., Dunham, M.J., Shendure, J., 2014. Species-level deconvolution of metagenome assemblies with Hi-C-based contact probability maps. G3 Genes,

921 Genomes, Genet. 4, 1339–1346. https://doi.org/10.1534/g3.114.011825
922 Cajka, T., Fiehn, O., 2016. Toward Merging Untargeted and Targeted Methods in Mass
923 Spectrometry-Based Metabolomics and Lipidomics. Anal. Chem.
924 https://doi.org/10.1021/acs.analchem.5b04491
925 Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., Holmes, S.P., 2016.
926 DADA2: High-resolution sample inference from Illumina amplicon data. Nat. Methods 13,
927 581–583. https://doi.org/10.1038/nmeth.3869
928 Camejo, P.Y., Oyserman, B.O., McMahon, K.D., Noguera, D.R., 2019. Integrated Omic
929 Analyses Provide Evidence that a “Candidatus Accumulibacter phosphatis” Strain
930 Performs Denitrification under Microaerobic Conditions. mSystems 4, e00193-18.
931 https://doi.org/10.1128/mSystems.00193-18
932 Camejo, P.Y., Santo Domingo, J., McMahon, K.D., Noguera, D.R., 2017. Genome-Enabled
933 Insights into the Ecophysiology of the Comammox Bacterium “Candidatus Nitrospira
934 nitrosa.” mSystems 2. https://doi.org/10.1128/msystems.00059-17
935 Campanaro, S., Treu, L., Rodriguez-R, L.M., Kovalovszki, A., Ziels, R.M., Maus, I., Zhu, X.,
936 Kougias, P.G., Basile, A., Luo, G., Schlüter, A., Konstantinidis, K.T., Angelidaki, I., 2020.
937 New insights from the biogas microbiome by comprehensive genome-resolved
938 metagenomics of nearly 1600 species originating from multiple anaerobic digesters.
939 Biotechnol. Biofuels 13, 25. https://doi.org/10.1186/s13068-020-01679-y
940 Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens,
941 S.A., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J.A., Smith, G., Knight, R.,
942 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and
943 MiSeq platforms, *ISME Journal,* **6**, 1621–1624
944 Carlson, R.P., Beck, A.E., Phalak, P., Fields, M.W., Gedeon, T., Hanley, L., Harcombe, W.R.,
945 Henson, M.A., Heys, J.J., 2018. Competitive resource allocation to metabolic pathways
946 contributes to overflow metabolisms and emergent properties in cross-feeding microbial
947 consortia. Biochem. Soc. Trans. https://doi.org/10.1042/BST20170242
948 Chaumeil, P.-A., Mussig, A.J., Hugenholtz, P., Parks, D.H., 2019. GTDB-Tk: a toolkit to

949 classify genomes with the Genome Taxonomy Database. Bioinformatics.
950 https://doi.org/10.1093/bioinformatics/btz848
951 Chemerys, A., Pelletier, E., Cruaud, C., Martin, F., Violet, F., Jouanneaua, Y., 2014.
952 Characterization of novel polycyclic aromatic hydrocarbon dioxygenases from the
953 bacterial metagenomic DNA of a contaminated soil. Appl. Environ. Microbiol. 80, 6591–
954 6600. https://doi.org/10.1128/AEM.01883-14
955 Chen, L.X., Anantharaman, K., Shaiber, A., Murat Eren, A., Banfield, J.F., 2020. Accurate and
956 complete genomes from metagenomes. Genome Res.
957 https://doi.org/10.1101/gr.258640.119
958 Chen, Y., Murrell, J.C., 2010. When metagenomics meets stable-isotope probing: Progress
959 and perspectives. Trends Microbiol. https://doi.org/10.1016/j.tim.2010.02.002
960 Clingenpeel, S., Clum, A., Schwientek, P., Rinke, C., Woyke, T., 2014. Reconstructing each
961 cell's genome within complex microbial communities - dream or reality? Front. Microbiol.
962 5. https://doi.org/10.3389/fmicb.2014.00771
963 Couradeau, E., Sasse, J., Goudeau, D., Nath, N., Hazen, T.C., Bowen, B.P., Chakraborty, R.,
964 Malmstrom, R.R., Northen, T.R., 2019. Probing the active fraction of soil microbiomes
965 using BONCAT-FACS. Nat. Commun. 10. https://doi.org/10.1038/s41467-019-10542-0
966 Coyotzi, S., Pratscher, J., Murrell, J.C., Neufeld, J.D., 2016. Targeted metagenomics of active
967 microbial populations with stable-isotope probing. Curr. Opin. Biotechnol.
968 https://doi.org/10.1016/j.copbio.2016.02.017
969 Cross, K.L., Campbell, J.H., Balachandran, M., Campbell, A.G., Cooper, S.J., Griffen, A.,
970 Heaton, M., Joshi, S., Klingeman, D., Leys, E., Yang, Z., Parks, J.M., Podar, M., 2019.
971 Targeted isolation and cultivation of uncultivated bacteria by reverse genomics. Nat.
972 Biotechnol. 37, 1314–1321. https://doi.org/10.1038/s41587-019-0260-6
973 Culviner, P.H., Guegler, C.K., Laub, M.T., 2020. A simple, cost-effective, and robust method
974 for rrna depletion in rna-sequencing studies. MBio 11.
975 https://doi.org/10.1128/mBio.00010-20
976 Dai, Z., Sevillano-Rivera, M.C., Calus, S.T., Santos, Q.M.B. los, Eren, A.M., Wielen, P.W.J.J.

van der, Ijaz, U.Z., Pinto, A.J., 2019. Disinfection exhibits systematic impacts on the drinking water microbiome. bioRxiv 828970. https://doi.org/10.1101/828970

Daims, H., Lebedeva, E. V., Pjevac, P., Han, P., Herbold, C., Albertsen, M., Jehmlich, N., Palatinszky, M., Vierheilig, J., Bulaev, A., Kirkegaard, R.H., von Bergen, M., Rattei, T., Bendinger, B., Nielsen, P.H., Wagner, M., 2015. Complete nitrification by Nitrospira bacteria. Nature 528, 504–509. https://doi.org/10.1038/nature16461

Dam, H.T., Vollmers, J., Sobol, M.S., Cabezas, A., Kaster, A.K., 2020. Targeted Cell Sorting Combined With Single Cell Genomics Captures Low Abundant Microbial Dark Matter With Higher Sensitivity Than Metagenomics. Front. Microbiol. 11, 1377. https://doi.org/10.3389/fmicb.2020.01377

de Almeida, N.M., Wessels, H.J.C.T., de Graaf, R.M., Ferousi, C., Jetten, M.S.M., Keltjens, J.T., Kartal, B., 2016. Membrane-bound electron transport systems of an anammox bacterium: A complexome analysis. Biochim. Biophys. Acta - Bioenerg. 1857, 1694–1704. https://doi.org/10.1016/j.bbabio.2016.07.006

Dekker, J., Rippe, K., Dekker, M., Kleckner, N., 2002. Capturing chromosome conformation. Science (80-.). 295, 1306–1311. https://doi.org/10.1126/science.1067799

DeLong, E.F., Wickham, G.S., Pace, N.R., 1989. Phylogenetic stains: Ribosomal RNA-based probes for the identification of single cells. Science (80-.). 243, 1360–1363. https://doi.org/10.1126/science.2466341

Dick, G., 2018. Genomic Approaches in Earth and Environmental Sciences.

Dick, G.J., Andersson, A.F., Baker, B.J., Simmons, S.L., Thomas, B.C., Yelton, A.P., Banfield, J.F., 2009. Community-wide analysis of microbial genome sequence signatures. Genome Biol. 10, R85. https://doi.org/10.1186/gb-2009-10-8-r85

Douterelo, I., Boxall, J.B., Deines, P., Sekar, R., Fish, K.E., Biggs, C.A., 2014. Methodological approaches for studying the microbial ecology of drinking water distribution systems. Water Res. https://doi.org/10.1016/j.watres.2014.07.008

Dumont, M.G., Murrell, J.C., 2005. Stable isotope probing - Linking microbial identify to function. Nat. Rev. Microbiol. 3, 499–504. https://doi.org/10.1038/nrmicro1162

1005 Edgar, R.C., 2018. Updating the 97% identity threshold for 16S ribosomal RNA OTUs.
1006 Bioinformatics 34, 2371–2375. https://doi.org/10.1093/bioinformatics/bty113
1007 Eren, A.M., Esen, Ö.C., Quince, C., Vineis, J.H., Morrison, H.G., Sogin, M.L., Delmont, T.O.,
1008 2015. Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ 3,
1009 e1319. https://doi.org/10.7717/peerj.1319
1010 Faust, K., Lahti, L., Gonze, D., de Vos, W.M., Raes, J., 2015. Metagenomics meets time series
1011 analysis: Unraveling microbial community dynamics. Curr. Opin. Microbiol.
1012 https://doi.org/10.1016/j.mib.2015.04.004
1013 Feist, A.M., Herrgård, M.J., Thiele, I., Reed, J.L., Palsson, B., 2009. Reconstruction of
1014 biochemical networks in microorganisms. Nat. Rev. Microbiol.
1015 https://doi.org/10.1038/nrmicro1949
1016 Fernando, E.Y., McIlroy, S.J., Nierychlo, M., Herbst, F.A., Petriglieri, F., Schmid, M.C.,
1017 Wagner, M., Nielsen, J.L., Nielsen, P.H., 2019. Resolving the individual contribution of
1018 key microbial populations to enhanced biological phosphorus removal with Raman–
1019 FISH. ISME J. 13, 1933–1946. https://doi.org/10.1038/s41396-019-0399-7
1020 Flowers, J.J., He, S., Malfatti, S., del Rio, T.G., Tringe, S.G., Hugenholtz, P., McMahon, K.D.,
1021 2013. Comparative genomics of two "Candidatus Accumulibacter" clades performing
1022 biological phosphorus removal. ISME J. 7, 2301–2314.
1023 https://doi.org/10.1038/ismej.2013.117
1024 Flowers, J.J., He, S., Yilmaz, S., Noguera, D.R., McMahon, K.D., 2009. Denitrification
1025 capabilities of two biological phosphorus removal sludges dominated by different
1026 "Candidatus Accumulibacter" clades. Environ. Microbiol. Rep. 1, 583–588.
1027 https://doi.org/10.1111/j.1758-2229.2009.00090.x
1028 Fox, G.E., Stackebrandt, E., Hespell, R.B., Gibson, J., Maniloff, J., Dyer, T.A., Wolfe, R.S.,
1029 Balch, W.E., Tanner, R.S., Magrum, L.J., Zahlen, L.B., Blakemore, R., Gupta, R., Bonen,
1030 L., Lewis, B.J., Stahl, D.A., Luehrsen, K.R., Chen, K.N., Woese, C.R., 1980. The
1031 phylogeny of prokaryotes. Science (80-.). 209, 457–463.
1032 https://doi.org/10.1126/science.6771870

1033 Frank, J., Lücker, S., Vossen, R.H.A.M., Jetten, M.S.M., Hall, R.J., Op Den Camp, H.J.M.,
1034 Anvar, S.Y., 2018. Resolving the complete genome of Kuenenia stuttgartiensis from a
1035 membrane bioreactor enrichment using Single-Molecule Real-Time sequencing
1036 /631/208/212/748 /631/326/41/2530 /631/326/41/2142 /631/326/171/1878 article. Sci.
1037 Rep. 8. https://doi.org/10.1038/s41598-018-23053-7
1038 Frigon, D., Wells, G., 2019. Microbial immigration in wastewater treatment systems: analytical
1039 considerations and process implications. Curr. Opin. Biotechnol.
1040 https://doi.org/10.1016/j.copbio.2019.02.021
1041 Gao, H., Mao, Y., Zhao, X., Liu, W.T., Zhang, T., Wells, G., 2019. Genome-centric
1042 metagenomics resolves microbial diversity and prevalent truncated denitrification
1043 pathways in a denitrifying PAO-enriched bioprocess. Water Res. 155, 275–287.
1044 https://doi.org/10.1016/j.watres.2019.02.020
1045 Gebreselassie, N.A., Antoniewicz, M.R., 2015. 13C-metabolic flux analysis of co-cultures: A
1046 novel approach. Metab. Eng. 31, 132–139. https://doi.org/10.1016/j.ymben.2015.07.005
1047 Ghosh, A., Nilmeier, J., Weaver, D., Adams, P.D., Keasling, J.D., Mukhopadhyay, A., Petzold,
1048 C.J., Martín, H.G., 2014. A Peptide-Based Method for 13C Metabolic Flux Analysis in
1049 Microbial Communities. PLoS Comput. Biol. 10.
1050 https://doi.org/10.1371/journal.pcbi.1003827
1051 Gilbert, J.A., Steele, J.A., Caporaso, J.G., Steinbrück, L., Reeder, J., Temperton, B., Huse,
1052 S., McHardy, A.C., Knight, R., Joint, I., Somerfield, P., Fuhrman, J.A., Field, D., 2012.
1053 Defining seasonal marine microbial community dynamics. ISME J. 6, 298–308.
1054 https://doi.org/10.1038/ismej.2011.107
1055 Giovannoni, S.J., Britschgi, T.B., Moyer, C.L., Field, K.G., 1990. Genetic diversity in Sargasso
1056 Sea bacterioplankton. Nature 345, 60–63. https://doi.org/10.1038/345060a0
1057 Giovannoni, S.J., DeLong, E.F., Olsen, G.J., Pace, N.R., 1988. Phylogenetic group-specific
1058 oligodeoxynucleotide probes for identification of single microbial cells. J. Bacteriol. 170,
1059 720–726. https://doi.org/10.1128/jb.170.2.720-726.1988
1060 Grieb, A., Bowers, R.M., Oggerin, M., Goudeau, D., Lee, J., Malmstrom, R.R., Woyke, T.,

Fuchs, B.M., 2020. A pipeline for targeted metagenomics of environmental bacteria. Microbiome 8, 21. https://doi.org/10.1186/s40168-020-0790-7

Griffin, J.S., Wells, G.F., 2017. Regional synchrony in full-scale activated sludge bioreactors due to deterministic microbial community assembly. ISME J. 11, 500–511. https://doi.org/10.1038/ismej.2016.121

Griffin, M.E., McMahon, K.D., Mackie, R.I., Raskin, L., 1998. Methanogenic population dynamics during start-up of anaerobic digesters treating municipal solid waste and biosolids. Biotechnol. Bioeng. 57, 342–355. https://doi.org/10.1002/(SICI)1097-0290(19980205)57:3<342::AID-BIT11>3.0.CO;2-I

Gudmundsson, S., Thiele, I., 2010. Computationally efficient flux variability analysis. BMC Bioinformatics 11, 489. https://doi.org/10.1186/1471-2105-11-489

Guedes da Silva, L., Olavarria Gamez, K., Castro Gomes, J., Akkermans, K., Welles, L., Abbas, B., van Loosdrecht, M.C.M., Wahl, S.A., 2020. Revealing the metabolic flexibility of Candidatus Accumulibacter phosphatis through redox cofactor analysis and metabolic network modeling. Appl. Environ. Microbiol.

Hamilton, J.J., Calixto Contreras, M., Reed, J.L., 2015. Thermodynamics and H2 Transfer in a Methanogenic, Syntrophic Community. PLOS Comput. Biol. 11, e1004364. https://doi.org/10.1371/journal.pcbi.1004364

Hamilton, J.J., Garcia, S.L., Brown, B.S., Oyserman, B.O., Moya-Flores, F., Bertilsson, S., Malmstrom, R.R., Forest, K.T., McMahon, K.D., 2017. Metabolic Network Analysis and Metatranscriptomics Reveal Auxotrophies and Nutrient Sources of the Cosmopolitan Freshwater Microbial Lineage acI. mSystems 2, e00091-17. https://doi.org/10.1128/mSystems.00091-17

Hamilton, J.J., Reed, J.L., 2014. Software platforms to facilitate reconstructing genome-scale metabolic networks. Environ. Microbiol. 16, 49–59. https://doi.org/10.1111/1462-2920.12312

Hamilton, J.J., Reed, J.L., 2012. Identification of Functional Differences in Metabolic Networks Using Comparative Genomics and Constraint-Based Models. PLoS One 7, e34670.

https://doi.org/10.1371/journal.pone.0034670

Handelsman, J., Liles, M., Mann, D., Riesenfeld, C., Goodman, R.M., 2002. Cloning the metagenome: Culture-independent access to thediversity and functions of the uncultivated microbial world. Methods Microbiol. https://doi.org/10.1016/s0580-9517(02)33014-9

Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J., 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. https://doi.org/10.1016/S1074-5521(98)90108-9

Hanson, N.W., Konwar, K.M., Wu, S.-J., Hallam, S.J., 2014. MetaPathways v2.0: A master-worker model for environmental Pathway/Genome Database construction on grids and clouds, in: 2014 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology. IEEE, pp. 1–7. https://doi.org/10.1109/CIBCB.2014.6845516

Hao, L., Michaelsen, T.Y., Singleton, C.M., Dottorini, G., Kirkegaard, R.H., Albertsen, M., Nielsen, P.H., Dueholm, M.S., 2020. Novel syntrophic bacteria in full-scale anaerobic digesters revealed by genome-centric metatranscriptomics. ISME J. 14, 906–918. https://doi.org/10.1038/s41396-019-0571-0

Hardwick, S.A., Chen, W.Y., Wong, T., Kanakamedala, B.S., Deveson, I.W., Ongley, S.E., Santini, N.S., Marcellin, E., Smith, M.A., Nielsen, L.K., Lovelock, C.E., Neilan, B.A., Mercer, T.R., 2018. Synthetic microbe communities provide internal reference standards for metagenome sequencing and analysis. Nat. Commun. 9, 1–10. https://doi.org/10.1038/s41467-018-05555-0

Hatzenpichler, R., Connon, S.A., Goudeau, D., Malmstrom, R.R., Woyke, T., Orphan, V.J., 2016. Visualizing in situ translational activity for identifying and sorting slow-growing archaeal - bacterial consortia. Proc. Natl. Acad. Sci. U. S. A. 113, E4069–E4078. https://doi.org/10.1073/pnas.1603757113

He, S., Kunin, V., Haynes, M., Martin, H.G., Ivanova, N., Rohwer, F., Hugenholtz, P., McMahon, K.D., 2010a. Metatranscriptomic array analysis of “Candidatus Accumulibacter phosphatis”-enriched enhanced biological phosphorus removal sludge.

Environ. Microbiol. 12, 1205–1217. https://doi.org/10.1111/j.1462-2920.2010.02163.x

He, S., Wurtzel, O., Singh, K., Froula, J.L., Yilmaz, S., Tringe, S.G., Wang, Z., Chen, F., Lindquist, E.A., Sorek, R., Hugenholtz, P., 2010b. Validation of two ribosomal RNA removal methods for microbial metatranscriptomics. Nat. Methods 7, 807–812. https://doi.org/10.1038/nmeth.1507

Henne, A., Daniel, R., Schmitz, R.A., Gottschalk, G., 1999. Construction of environmental DNA libraries in Escherichia coli and screening for the presence of genes conferring utilization of 4- hydroxybutyrate. Appl. Environ. Microbiol. 65, 3901–3907. https://doi.org/10.1128/aem.65.9.3901-3907.1999

Herold, M., Martínez Arbas, S., Narayanasamy, S., Sheik, A.R., Kleine-Borgmann, L.A.K., Lebrun, L.A., Kunath, B.J., Roume, H., Bessarab, I., Williams, R.B.H., Gillece, J.D., Schupp, J.M., Keim, P.S., Jäger, C., Hoopmann, M.R., Moritz, R.L., Ye, Y., Li, S., Tang, H., Heintz-Buschart, A., May, P., Muller, E.E.L., Laczny, C.C., Wilmes, P., 2020. Integration of time-series meta-omics data reveals how microbial ecosystems respond to disturbance. Nat. Commun. 11, 1–14. https://doi.org/10.1038/s41467-020-19006-2

Hsieh, T.H.S., Weiner, A., Lajoie, B., Dekker, J., Friedman, N., Rando, O.J., 2015. Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. Cell 162, 108–119. https://doi.org/10.1016/j.cell.2015.05.048

Huang, W.E., Griffiths, R.I., Thompson, I.P., Bailey, M.J., Whiteley, A.S., 2004. Raman microscopic analysis of single microbial cells. Anal. Chem. 76, 4452–4458. https://doi.org/10.1021/ac049753k

Huang, W.E., Stoecker, K., Griffiths, R., Newbold, L., Daims, H., Whiteley, A.S., Wagner, M., 2007. Raman-FISH: Combining stable-isotope Raman spectroscopy and fluorescence in situ hybridization for the single cell analysis of identity and function. Environ. Microbiol. 9, 1878–1889. https://doi.org/10.1111/j.1462-2920.2007.01352.x

Hug, L.A., Baker, B.J., Anantharaman, K., Brown, C.T., Probst, A.J., Castelle, C.J., Butterfield, C.N., Hernsdorf, A.W., Amano, Y., Ise, K., Suzuki, Y., Dudek, N., Relman, D.A., Finstad, K.M., Amundson, R., Thomas, B.C., Banfield, J.F., 2016. A new view of the tree of life.

Nat. Microbiol. 1, 16048. https://doi.org/10.1038/nmicrobiol.2016.48

Hull, N.M., Ling, F., Pinto, A.J., Albertsen, M., Jang, H.G., Hong, P.Y., Konstantinidis, K.T., LeChevallier, M., Colwell, R.R., Liu, W.T., 2019. Drinking Water Microbiome Project: Is it Time? Trends Microbiol. https://doi.org/10.1016/j.tim.2019.03.011

Huson, D.H., Albrecht, B., Bağci, C., Bessarab, I., Górska, A., Jolic, D., Williams, R.B.H., 2018. MEGAN-LR: New algorithms allow accurate binning and easy interactive exploration of metagenomic long reads and contigs. Biol. Direct 13, 6. https://doi.org/10.1186/s13062-018-0208-7

Ishii, S., Suzuki, S., Norden-Krichmar, T.M., Tenney, A., Chain, P.S.G., Scholz, M.B., Nealson, K.H., Bretschger, O., 2013. A novel metatranscriptomic approach to identify gene expression dynamics during extracellular electron transfer. Nat. Commun. 4, 1–10. https://doi.org/10.1038/ncomms2615

Jain, C., Rodriguez-R, L.M., Phillippy, A.M., Konstantinidis, K.T., Aluru, S., 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat. Commun. 9, 5114. https://doi.org/10.1038/s41467-018-07641-9

Jehmlich, N., Schmidt, F., Von Bergen, M., Richnow, H.H., Vogt, C., 2008. Protein-based stable isotope probing (Protein-SIP) reveals active species within anoxic mixed cultures. ISME J. 2, 1122–1133. https://doi.org/10.1038/ismej.2008.64

Ji, P., Zhang, Y., Wang, J., Zhao, F., 2017. MetaSort untangles metagenome assembly by reducing microbial community complexity. Nat. Commun. 8, 1–14. https://doi.org/10.1038/ncomms14306

Kalyuzhnaya, M.G., Zabinsky, R., Bowerman, S., Baker, D.R., Lidstrom, M.E., Chistoserdova, L., 2006. Fluorescence in situ hybridization-flow cytometry-cell sorting-based method for separation and enrichment of type I and type II methanotroph populations. Appl. Environ. Microbiol. 72, 4293–4301. https://doi.org/10.1128/AEM.00161-06

Kang, D.D., Froula, J., Egan, R., Wang, Z., 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ 3, e1165. https://doi.org/10.7717/peerj.1165

Kantor, R.S., Miller, S.E., Nelson, K.L., 2019. The Water Microbiome Through a Pilot Scale Advanced Treatment Facility for Direct Potable Reuse. Front. Microbiol. 10, 993. https://doi.org/10.3389/fmicb.2019.00993

Karst, S.M., Dueholm, M.S., McIlroy, S.J., Kirkegaard, R.H., Nielsen, P.H., Albertsen, M., 2018. Retrieval of a million high-quality, full-length microbial 16S and 18S rRNA gene sequences without primer bias. Nat. Publ. Gr. 36. https://doi.org/10.1038/nbt.4045

Karst, S.M., Ziels, R.M., Kirkegaard, R.H., Sørensen, E.A., McDonald, D., Zhu, Q., Knight, R., Albertsen, M., 2021. High-accuracy long-read amplicon sequences using unique molecular identifiers with Nanopore or PacBio sequencing. Nat. Methods 1–5. https://doi.org/10.1038/s41592-020-01041-y

Kirisits, M.J., Emelko, M.B., Pinto, A.J., 2019. Applying biotechnology for drinking water biofiltration: advancing science and practice. Curr. Opin. Biotechnol. https://doi.org/10.1016/j.copbio.2019.05.009

Kirkegaard, R.H., McIlroy, S.J., Kristensen, J.M., Nierychlo, M., Karst, S.M., Dueholm, M.S., Albertsen, M., Nielsen, P.H., 2017. The impact of immigration on microbial community composition in full-scale anaerobic digesters. Sci. Rep. 7, 1–11. https://doi.org/10.1038/s41598-017-09303-0

Knight, R., Vrbanac, A., Taylor, B.C., Aksenov, A., Callewaert, C., Debelius, J., Gonzalez, A., Kosciolek, T., McCall, L.I., McDonald, D., Melnik, A. V., Morton, J.T., Navas, J., Quinn, R.A., Sanders, J.G., Swafford, A.D., Thompson, L.R., Tripathi, A., Xu, Z.Z., Zaneveld, J.R., Zhu, Q., Caporaso, J.G., Dorrestein, P.C., 2018. Best practices for analysing microbiomes. Nat. Rev. Microbiol. https://doi.org/10.1038/s41579-018-0029-9

Koch, H., Lücker, S., Albertsen, M., Kitzinger, K., Herbold, C., Spieck, E., Nielsen, P.H., Wagner, M., Daims, H., 2015. Expanded metabolic versatility of ubiquitous nitrite-oxidizing bacteria from the genus Nitrospira. Proc. Natl. Acad. Sci. U. S. A. 112, 11371–11376. https://doi.org/10.1073/pnas.1506533112

Koster J., Rahmann S. 2012. Snakemake–a scalable bioinformatics workflow engine. *Bioinformatics*, **28**, 2520–2522.

Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., Schloss, P.D., 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl. Environ. Microbiol. 79, 5112–20. https://doi.org/10.1128/AEM.01043-13

Lane, J., Pace, B., Olsen, G.J., Stahl, D.A., Sogin, M.L., Pace, N.R., Lane, D.J., Stahlt, D.A., Sogint, M.L., 1986. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Evolution (N. Y). 83, 6955–6959. https://doi.org/10.1073/pnas.82.20.6955

Law, Y., Kirkegaard, R.H., Cokro, A.A., Liu, X., Arumugam, K., Xie, C., Stokholm-Bjerregaard, M., Drautz-Moses, D.I., Nielsen, P.H., Wuertz, S., Williams, R.B.H., 2016. Integrative microbial community analysis reveals full-scale enhanced biological phosphorus removal under tropical conditions. Sci. Rep. 6, 1–15. https://doi.org/10.1038/srep25719

Lawson, C.E., Harcombe, W.R., Hatzenpichler, R., Lindemann, S.R., Löffler, F.E., O'Malley, M.A., García Martín, H., Pfleger, B.F., Raskin, L., Venturelli, O.S., Weissbrodt, D.G., Noguera, D.R., McMahon, K.D., 2019. Common principles and best practices for engineering microbiomes. Nat. Rev. Microbiol. https://doi.org/10.1038/s41579-019-0255-9

Lawson, C.E., Lücker, S., 2018. Complete ammonia oxidation: an important control on nitrification in engineered ecosystems? Curr. Opin. Biotechnol. https://doi.org/10.1016/j.copbio.2018.01.015

Lawson, C.E., Nuijten, G.H.L., de Graaf, R.M., Jacobson, T.B., Pabst, M., Stevenson, D.M., Jetten, M.S.M., Noguera, D.R., McMahon, K.D., Amador-Noguez, D., Lücker, S., 2020. Autotrophic and mixotrophic metabolism of an anammox bacterium revealed by in vivo 13C and 2H metabolic network mapping. ISME J. 1–15. https://doi.org/10.1038/s41396-020-00805-w

Lawson, C.E., Strachan, B.J., Hanson, N.W., Hahn, A.S., Hall, E.R., Rabinowitz, B., Mavinic, D.S., Ramey, W.D., Hallam, S.J., 2015. Rare taxa have potential to make metabolic contributions in enhanced biological phosphorus removal ecosystems. Environ.

Microbiol. 17, 4979–4993. https://doi.org/10.1111/1462-2920.12875

Lawson, C.E., Wu, S., Bhattacharjee, A.S., Hamilton, J.J., McMahon, K.D., Goel, R., Noguera, D.R., 2017. Metabolic network analysis reveals microbial community interactions in anammox granules. Nat. Commun. 8, 15416. https://doi.org/10.1038/ncomms15416

Lee, H., Gurtowski, J., Yoo, S., Nattestad, M., Marcus, S., Goodwin, S., Mccombie, W.R., Schatz, M.C., 2016. Third-generation sequencing and the future of genomics. bioRxiv. https://doi.org/10.1101/048603

Lee, K.S., Pereira, F.C., Palatinszky, M., Behrendt, L., Alcolombri, U., Berry, D., Wagner, M., Stocker, R., 2020. Optofluidic Raman-activated cell sorting for targeted genome retrieval or cultivation of microbial cells with specific functions. Nat. Protoc. 1–43. https://doi.org/10.1038/s41596-020-00427-8

Lee, M.D., 2019. GToTree: a user-friendly workflow for phylogenomics. Bioinformatics. https://doi.org/10.1093/bioinformatics/btz188

Lee, N., Halkjaer Nielsen, P., Holm Andreasen, K., Juretschko, S., Lund Nielsen, J., Schleifer, K., Wagner, M., 1999. Combination of Fluorescent In Situ Hybridization and Microautoradiography-a New Tool for Structure-Function Analyses in Microbial Ecology, APPLIED AND ENVIRONMENTAL MICROBIOLOGY.

Lee, S.H., Kang, H.J., Park, H.D., 2015. Influence of influent wastewater communities on temporal variation of activated sludge communities. Water Res. 73, 132–144. https://doi.org/10.1016/j.watres.2015.01.014

Leventhal, G.E., Boix, C., Kuechler, U., Enke, T.N., Sliwerska, E., Holliger, C., Cordero, O.X., 2018. Strain-level diversity drives alternative community types in millimetre-scale granular biofilms. Nat. Microbiol. 3, 1295–1303. https://doi.org/10.1038/s41564-018-0242-3

Lieberman-Aiden, E., Van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., Sandstrom, R., Bernstein, B., Bender, M.A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L.A., Lander, E.S., Dekker, J., 2009. Comprehensive mapping of long-range interactions reveals folding

principles of the human genome. Science (80-.). 326, 289–293. https://doi.org/10.1126/science.1181369

Linz, A.M., He, S., Stevens, S.L.R., Anantharaman, K., Rohwer, R.R., Malmstrom, R.R., Bertilsson, S., McMahon, K.D., 2018. Freshwater carbon and nutrient cycles revealed through reconstructed population genomes. PeerJ 6, e6075. https://doi.org/10.7717/peerj.6075

Lipton, M.S., Păá-Toli, L., Anderson, G.A., Anderson, D.J., Auberry, D.L., Battista, J.R., Daly, M.J., Fredrickson, J., Hixson, K.K., Kostandarithes, H., Masselon, C., Markillie, L.M., Moore, R.J., Romine, M.F., Shen, Y., Stritmatter, E., Tolić, N., Udseth, H.R., Venkateswaran, A., Wong, K.K., Zhao, R., Smith, R.D., 2002. Global analysis of the Deinococcus radiodurans proteome by using accurate mass tags. Proc. Natl. Acad. Sci. U. S. A. 99, 11049–11054. https://doi.org/10.1073/pnas.172170199

Lloyd, K.G., Steen, A.D., Ladau, J., Yin, J., Crosby, L., 2018. Phylogenetically Novel Uncultured Microbial Cells Dominate Earth Microbiomes. mSystems 3. https://doi.org/10.1128/msystems.00055-18

Lo, I., Denef, V.J., VerBerkmoes, N.C., Shah, M.B., Goltsman, D., DiBartolo, G., Tyson, G.W., Allen, E.E., Ram, R.J., Detter, J.C., Richardson, P., Thelen, M.P., Hettich, R.L., Banfield, J.F., 2007. Strain-resolved community proteomics reveals recombining genomes of acidophilic bacteria. Nature 446, 537–541. https://doi.org/10.1038/nature05624

Long, C.P., Antoniewicz, M.R., 2019. High-resolution 13C metabolic flux analysis. Nat. Protoc. 14, 2856–2877. https://doi.org/10.1038/s41596-019-0204-0

Louca, S., Hawley, A.K., Katsev, S., Torres-Beltran, M., Bhatia, M.P., Kheirandish, S., Michiels, C.C., Capelle, D., Lavik, G., Doebeli, M., Crowe, S.A., Hallam, S.J., 2016. Integrating biogeochemistry with multiomic sequence information in a model oxygen minimum zone. Proc. Natl. Acad. Sci. U. S. A. 113, E5925–E5933. https://doi.org/10.1073/pnas.1602897113

Lücker, S., Nowka, B., Rattei, T., Spieck, E., Daims, H., 2013. The genome of Nitrospina gracilis illuminates the metabolism and evolution of the major marine nitrite oxidizer.

Front. Microbiol. 4, 27. https://doi.org/10.3389/fmicb.2013.00027

Lücker, S., Wagner, M., Maixner, F., Pelletier, E., Koch, H., Vacherie, B., Rattei, T., Damsté, J.S.S., Spieck, E., Le Paslier, D., Daims, H., 2010. A Nitrospira metagenome illuminates the physiology and evolution of globally important nitrite-oxidizing bacteria. Proc. Natl. Acad. Sci. U. S. A. 107, 13479–13484. https://doi.org/10.1073/pnas.1003860107

Lui, L.M., Nielsen, T.N., Arkin, A.P., 2020. A method for achieving complete microbial genomes and improving bins from metagenomics data. bioRxiv. https://doi.org/10.1101/2020.03.05.979740

Machado, D., Andrejev, S., Tramontano, M., Patil, K.R., 2018. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. Nucleic Acids Res. 46, 7542–7553. https://doi.org/10.1093/nar/gky537

Malmstrom, R.R., Eloe-Fadrosh, E.A., 2019. Advancing Genome-Resolved Metagenomics beyond the Shotgun. mSystems 4. https://doi.org/10.1128/msystems.00118-19

Manz, W., Amann, R., Ludwig, W., Vancanneyt, M., Schleifer, K.H., 1996. Application of a suite of 16S rRNA-specific oligonucleotide probes designed to investigate bacteria of the phylum cytophaga-flavobacter-bacteroides in the natural environment. Microbiology 142, 1097–1106. https://doi.org/10.1099/13500872-142-5-1097

Manz, W., Amann, R., Ludwig, W., Wagner, M., Schleifer, K.H., 1992. Phylogenetic Oligodeoxynucleotide Probes for the Major Subclasses of Proteobacteria: Problems and Solutions. Syst. Appl. Microbiol. 15, 593–600. https://doi.org/10.1016/S0723-2020(11)80121-9

Mao, Y., Yu, K., Xia, Y., Chao, Y., Zhang, T., 2014. Genome Reconstruction and Gene Expression of " *Candidatus* Accumulibacter phosphatis" Clade IB Performing Biological Phosphorus Removal. Environ. Sci. Technol. 48, 10363–10371. https://doi.org/10.1021/es502642b

Marbouty, M., Cournac, A., Flot, J.-F., Marie-Nelly, H., Mozziconacci, J., Koszul, R., 2014. Metagenomic chromosome conformation capture (meta3C) unveils the diversity of chromosome organization in microorganisms. Elife 3, e03318.

https://doi.org/10.7554/eLife.03318

Marques, R., Santos, J., Nguyen, H., Carvalho, G., Noronha, J.P., Nielsen, P.H., Reis, M.A.M., Oehmen, A., 2017. Metabolism and ecological niche of Tetrasphaera and Ca. Accumulibacter in enhanced biological phosphorus removal. Water Res. 122, 159–171. https://doi.org/10.1016/j.watres.2017.04.072

Martín, H.G., Ivanova, N., Kunin, V., Warnecke, F., Barry, K.W., McHardy, A.C., Yeates, C., He, S., Salamov, A.A., Szeto, E., Dalin, E., Putnam, N.H., Shapiro, H.J., Pangilinan, J.L., Rigoutsos, I., Kyrpides, N.C., Blackall, L.L., McMahon, K.D., Hugenholtz, P., 2006. Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. Nat. Biotechnol. 24, 1263–1269. https://doi.org/10.1038/nbt1247

Martínez Arbas, S., Narayanasamy, S., Herold, M., Lebrun, L.A., Hoopmann, M.R., Li, S., Lam, T.J., Kunath, B.J., Hicks, N.D., Liu, C.M., Price, L.B., Laczny, C.C., Gillece, J.D., Schupp, J.M., Keim, P.S., Moritz, R.L., Faust, K., Tang, H., Ye, Y., Skupin, A., May, P., Muller, E.E.L., Wilmes, P., 2021. Roles of bacteriophages, plasmids and CRISPR immunity in microbial community dynamics revealed using time-series integrated meta-omics. Nat. Microbiol. 6, 123–135. https://doi.org/10.1038/s41564-020-00794-8

Matsuoka, Y., Shimizu, K., 2014. 13C-metabolic flux analysis for Escherichia coli. Methods Mol. Biol. 1191, 261–289. https://doi.org/10.1007/978-1-4939-1170-7_16

McDaniel, E.A., Anantharaman, K., McMahon, K.D., 2019. metabolisHMM: Phylogenomic analysis for exploration of microbial phylogenies and metabolic pathways. bioRxiv 2019.12.20.884627. https://doi.org/10.1101/2019.12.20.884627

McDaniel, E.A., Peterson, B.D., Stevens, S.L.R., Tran, P.Q., Anantharaman, K., McMahon, K.D., 2020. Expanded Phylogenetic Diversity and Metabolic Flexibility of Mercury-Methylating Microorganisms. mSystems 5. https://doi.org/10.1128/msystems.00299-20

McDaniel, E., Moya-Flores, F., Keene Beach, N., Oyserman, B., Kizaric, M., Hoe Khor, E., McMahon, K.D., 2020. Metabolic differentiation of co-occurring Accumulibacter clades revealed through genome-resolved metatranscriptomics. bioRxiv 2020.11.23.394700. https://doi.org/10.1101/2020.11.23.394700

McIlroy, S.J., Karst, S.M., Nierychlo, M., Dueholm, M.S., Albertsen, M., Kirkegaard, R.H., Seviour, R.J., Nielsen, P.H., 2016. Genomic and in situ investigations of the novel uncultured Chloroflexi associated with 0092 morphotype filamentous bulking in activated sludge. ISME J. 10, 2223–2234. https://doi.org/10.1038/ismej.2016.14

McIlroy, S.J., Kirkegaard, R.H., McIlroy, B., Nierychlo, M., Kristensen, J.M., Karst, S.M., Albertsen, M., Nielsen, P.H., 2017. MiDAS 2.0: An ecosystem-specific taxonomy and online database for the organisms of wastewater treatment systems expanded for anaerobic digester groups. Database 2017, 16. https://doi.org/10.1093/database/bax016

McIlroy, S.J., Kristiansen, R., Albertsen, M., Karst, S.M., Rossetti, S., Nielsen, J.L., Tandoi, V., Seviour, R.J., Nielsen, P.H., 2013. Metabolic model for the filamentous “Candidatus Microthrix parvicella” based on genomic and metagenomic analyses. ISME J. 7, 1161–1172. https://doi.org/10.1038/ismej.2013.6

McIlroy, S.J., Saunders, A.M., Albertsen, M., Nierychlo, M., McIlroy, B., Hansen, A.A., Karst, S.M., Nielsen, J.L., Nielsen, P.H., 2015. MiDAS: The field guide to the microbes of activated sludge. Database 2015. https://doi.org/10.1093/database/bav062

McMahon, K.D., Stroot, P.G., Mackie, R.I., Raskin, L., 2001. Anaerobic codigestion of municipal solid waste and biosolids under various mixing conditions-II: Microbial population dynamics. Water Res. 35, 1817–1827. https://doi.org/10.1016/S0043-1354(00)00438-3

Meziti, A., Rodriguez-R, L.M., Hatt, J.K., Peña-Gonzalez, A., Levy, K., Konstantinidis, K.T., 2021. How reliably do metagenome-assembled genomes (MAGs) represent natural populations? Insights from comparing MAGs against isolate genomes derived from the same fecal sample. Appl. Environ. Microbiol. 87. https://doi.org/10.1128/aem.02593-20

Mobarry, B.K., Wagner, M., Urbain, V., Rittmann, B.E., Stahl, D.A., 1996. Phylogenetic probes for analyzing abundance and spatial organization of nitrifying bacteria. Appl. Environ. Microbiol. 62.

Molenaar, D., van Berlo, R., de Ridder, D., Teusink, B., 2009. Shifts in growth strategies reflect tradeoffs in cellular economics. Mol. Syst. Biol. 5, 323.

https://doi.org/10.1038/msb.2009.82

Moran, M.A., Satinsky, B., Gifford, S.M., Luo, H., Rivers, A., Chan, L.K., Meng, J., Durham, B.P., Shen, C., Varaljay, V.A., Smith, C.B., Yager, P.L., Hopkinson, B.M., 2013. Sizing up metatranscriptomics. ISME J. https://doi.org/10.1038/ismej.2012.94

Morisse, P., Lecroq, T., Lefebvre, A., 2020. Long-read error correction: a survey and qualitative comparison. bioRxiv 2020.03.06.977975. https://doi.org/10.1101/2020.03.06.977975

Morisse, P., Marchet, C., Limasset, A., Lecroq, T., Lefebvre, A., 2019. CONSENT: Scalable self-correction of long reads with multiple sequence alignment. bioRxiv 546630. https://doi.org/10.1101/546630

Mosbæk, F., Kjeldal, H., Mulat, D.G., Albertsen, M., Ward, A.J., Feilberg, A., Nielsen, J.L., 2016. Identification of syntrophic acetate-oxidizing bacteria in anaerobic digesters by combined protein-based stable isotope probing and metagenomics. ISME J. 10, 2405–2418. https://doi.org/10.1038/ismej.2016.39

Moss, E.L., Maghini, D.G., Bhatt, A.S., 2020. Complete, closed bacterial genomes from microbiomes using nanopore sequencing. Nat. Biotechnol. 38, 701–707. https://doi.org/10.1038/s41587-020-0422-6

Mueller, A.J., Jung, M.-Y., Strachan, C.R., Herbold, C.W., Kirkegaard, R.H., Wagner, M., Daims, H., 2020. Genomic and kinetic analysis of novel Nitrospinae enriched by cell sorting. ISME J. 2020.06.09.141952. https://doi.org/10.1038/s41396-020-00809-6

Muller, E.E.L., Pinel, N., Laczny, C.C., Hoopmann, M.R., Narayanasamy, S., Lebrun, L.A., Roume, H., Lin, J., May, P., Hicks, N.D., Heintz-Buschart, A., Wampach, L., Liu, C.M., Price, L.B., Gillece, J.D., Guignard, C., Schupp, J.M., Vlassis, N., Baliga, N.S., Moritz, R.L., Keim, P.S., Wilmes, P., 2014. Community-integrated omics links dominance of a microbial generalist to fine-tuned resource usage. Nat. Commun. 5, 1–10. https://doi.org/10.1038/ncomms6603

Narayanasamy, S., Muller, E.E.L., Sheik, A.R., Wilmes, P., 2015. Integrated omics for the identification of key functionalities in biological wastewater treatment microbial

communities. Microb. Biotechnol. 8, 363–368. https://doi.org/10.1111/1751-7915.12255

Nayfach, S., Roux, S., Seshadri, R., Udwary, D., Varghese, N., Schulz, F., Wu, D., Paez-Espino, D., Chen, I.M., Huntemann, M., Palaniappan, K., Ladau, J., Mukherjee, S., Reddy, T.B.K., Nielsen, T., Kirton, E., Faria, J.P., Edirisinghe, J.N., Henry, C.S., Jungbluth, S.P., Chivian, D., Dehal, P., Wood-Charlson, E.M., Arkin, A.P., Tringe, S.G., Visel, A., Abreu, H., Acinas, S.G., Allen, E., Allen, M.A., Andersen, G., Anesio, A.M., Attwood, G., Avila-Magaña, V., Badis, Y., Bailey, J., Baker, B., Baldrian, P., Barton, H.A., Beck, D.A.C., Becraft, E.D., Beller, H.R., Beman, J.M., Bernier-Latmani, R., Berry, T.D., Bertagnolli, A., Bertilsson, S., Bhatnagar, J.M., Bird, J.T., Blumer-Schuette, S.E., Bohannan, B., Borton, M.A., Brady, A., Brawley, S.H., Brodie, J., Brown, S., Brum, J.R., Brune, A., Bryant, D.A., Buchan, A., Buckley, D.H., Buongiorno, J., Cadillo-Quiroz, H., Caffrey, S.M., Campbell, A.N., Campbell, B., Carr, S., Carroll, J.L., Cary, S.C., Cates, A.M., Cattolico, R.A., Cavicchioli, R., Chistoserdova, L., Coleman, M.L., Constant, P., Conway, J.M., Mac Cormack, W.P., Crowe, S., Crump, B., Currie, C., Daly, R., Denef, V., Denman, S.E., Desta, A., Dionisi, H., Dodsworth, J., Dombrowski, N., Donohue, T., Dopson, M., Driscoll, T., Dunfield, P., Dupont, C.L., Dynarski, K.A., Edgcomb, V., Edwards, E.A., Elshahed, M.S., Figueroa, I., Flood, B., Fortney, N., Fortunato, C.S., Francis, C., Gachon, C.M.M., Garcia, S.L., Gazitua, M.C., Gentry, T., Gerwick, L., Gharechahi, J., Girguis, P., Gladden, J., Gradoville, M., Grasby, S.E., Gravuer, K., Grettenberger, C.L., Gruninger, R.J., Guo, J., Habteselassie, M.Y., Hallam, S.J., Hatzenpichler, R., Hausmann, B., Hazen, T.C., Hedlund, B., Henny, C., Herfort, L., Hernandez, M., Hershey, O.S., Hess, M., Hollister, E.B., Hug, L.A., Hunt, D., Jansson, J., Jarett, J., Kadnikov, V. V., Kelly, C., Kelly, R., Kelly, W., Kerfeld, C.A., Kimbrel, J., Klassen, J.L., Konstantinidis, K.T., Lee, L.L., Li, W.J., Loder, A.J., Loy, A., Lozada, M., MacGregor, B., Magnabosco, C., Maria da Silva, A., McKay, R.M., McMahon, K., McSweeney, C.S., Medina, M., Meredith, L., Mizzi, J., Mock, T., Momper, L., Moran, M.A., Morgan-Lang, C., Moser, D., Muyzer, G., Myrold, D., Nash, M., Nesbø, C.L., Neumann, A.P., Neumann, R.B., Noguera, D., Northen, T., Norton, J., Nowinski, B.,

Nüsslein, K., O'Malley, M.A., Oliveira, R.S., Maia de Oliveira, V., Onstott, T., Osvatic, J., Ouyang, Y., Pachiadaki, M., Parnell, J., Partida-Martinez, L.P., Peay, K.G., Pelletier, D., Peng, X., Pester, M., Pett-Ridge, J., Peura, S., Pjevac, P., Plominsky, A.M., Poehlein, A., Pope, P.B., Ravin, N., Redmond, M.C., Reiss, R., Rich, V., Rinke, C., Rodrigues, J.L.M., Rossmassler, K., Sackett, J., Salekdeh, G.H., Saleska, S., Scarborough, M., Schachtman, D., Schadt, C.W., Schrenk, M., Sczyrba, A., Sengupta, A., Setubal, J.C., Shade, A., Sharp, C., Sherman, D.H., Shubenkova, O. V., Sierra-Garcia, I.N., Simister, R., Simon, H., Sjöling, S., Slonczewski, J., Correa de Souza, R.S., Spear, J.R., Stegen, J.C., Stepanauskas, R., Stewart, F., Suen, G., Sullivan, M., Sumner, D., Swan, B.K., Swingley, W., Tarn, J., Taylor, G.T., Teeling, H., Tekere, M., Teske, A., Thomas, T., Thrash, C., Tiedje, J., Ting, C.S., Tully, B., Tyson, G., Ulloa, O., Valentine, D.L., Van Goethem, M.W., VanderGheynst, J., Verbeke, T.J., Vollmers, J., Vuillemin, A., Waldo, N.B., Walsh, D.A., Weimer, B.C., Whitman, T., van der Wielen, P., Wilkins, M., Williams, T.J., Woodcroft, B., Woolet, J., Wrighton, K., Ye, J., Young, E.B., Youssef, N.H., Yu, F.B., Zemskaya, T.I., Ziels, R., Woyke, T., Mouncey, N.J., Ivanova, N.N., Kyrpides, N.C., Eloe-Fadrosh, E.A., 2020. A genomic catalog of Earth's microbiomes. Nat. Biotechnol. 1–11. https://doi.org/10.1038/s41587-020-0718-6

Neely, C.J., Graham, E.D., Tully, B.J., 2020. MetaSanity: an integrated microbial genome evaluation and annotation pipeline. Bioinformatics. https://doi.org/10.1093/bioinformatics/btaa512

Neufeld, J.D., Murrell, J.C., 2007. Witnessing the last supper of uncultivated microbial cells with Raman-FISH. ISME J. https://doi.org/10.1038/ismej.2007.55

Neufeld, J.D., Vohra, J., Dumont, M.G., Lueders, T., Manefield, M., Friedrich, M.W., Murrell, C.J., 2007. DNA stable-isotope probing. Nat. Protoc. 2, 860–866. https://doi.org/10.1038/nprot.2007.109

Nicholls, S.M., Quick, J.C., Tang, S.Q., Loman N.J. 2019. Ultra-deep, long-read nanopore sequencing of mock microbial community standards, *GigaScience*, **8**, giz043, https://doi.org/10.1093/gigascience/giz043

Nielsen, J., 2003. It Is All about Metabolic Fluxes. J. Bacteriol. https://doi.org/10.1128/JB.185.24.7031-7035.2003

Nierychlo, M., Andersen, K.S., Xu, Y., Green, N., Albertsen, M., Dueholm, M.S., Nielsen, P.H., 2019. Species-level microbiome composition of activated sludge - introducing the MiDAS 3 ecosystem-specific reference database and taxonomy. bioRxiv 842393. https://doi.org/10.1101/842393

Nierychlo, M., Andersen, K.S., Xu, Y., Green, N., Jiang, C., Albertsen, M., Dueholm, M.S., Nielsen, P.H., 2020. MiDAS 3: An ecosystem-specific reference database, taxonomy and knowledge platform for activated sludge and anaerobic digesters reveals species-level microbiome composition of activated sludge. Water Res. 182, 115955. https://doi.org/10.1016/j.watres.2020.115955

Nissen, J.N., Johansen, J., Allesøe, R.L. et al. 2021. Improved metagenome binning and assembly using deep variational autoencoders. *Nature Biotechnology* **39**, 555–560.

Olm, M.R., Brown, C.T., Brooks, B., Banfield, J.F., 2017. DRep: A tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J. 11, 2864–2868. https://doi.org/10.1038/ismej.2017.126

Olm, M.R., Crits-Christoph, A., Bouma-Gregson, K., Firek, B.A., Morowitz, M.J., Banfield, J.F., 2021. inStrain profiles population microdiversity from metagenomic data and sensitively detects shared microbial strains. Nat. Biotechnol. 1–10. https://doi.org/10.1038/s41587-020-00797-0

Olm, M.R., Crits-Christoph, A., Diamond, S., Lavy, A., Matheus Carnevali, P.B., Banfield, J.F., 2020. Consistent Metagenome-Derived Metrics Verify and Delineate Bacterial Species Boundaries. mSystems 5. https://doi.org/10.1128/mSystems.00731-19

Olsen, G.J., Lane, D.J., Giovannoni, S.J., Pace, N.R., Stahl, D.A., 1986. Microbial Ecology and Evolution: A Ribosomal RNA Approach. Annu. Rev. Microbiol. 40, 337–365. https://doi.org/10.1146/annurev.mi.40.100186.002005

Orakov, A., Fullam, A., Coelho, L.P., Khedkar, S., Szklarczyk, D., Mende, D.R., Schmidt, T.S.B., Bork, P., 2020. GUNC: Detection of Chimerism and Contamination in Prokaryotic

Genomes. bioRxiv. https://doi.org/10.1101/2020.12.16.422776

Oyserman, B.O., Moya, F., Lawson, C.E., Garcia, A.L., Vogt, M., Heffernen, M., Noguera, D.R., McMahon, K.D., 2016a. Ancestral genome reconstruction identifies the evolutionary basis for trait acquisition in polyphosphate accumulating bacteria. ISME J. 10, 2931–2945. https://doi.org/10.1038/ismej.2016.67

Oyserman, B.O., Noguera, D.R., del Rio, T.G., Tringe, S.G., McMahon, K.D., 2016b. Metatranscriptomic insights on gene expression and regulatory controls in Candidatus Accumulibacter phosphatis. ISME J. 10, 810–822. https://doi.org/10.1038/ismej.2015.155

Palmer, C., Bik, E.M., DiGiulio, D.B., Relman, D.A., Brown, P.O., 2007. Development of the Human Infant Intestinal Microbiota. PLoS Biol. 5, e177. https://doi.org/10.1371/journal.pbio.0050177

Palomo, A., Pedersen, A.G., Fowler, S.J., Dechesne, A., Sicheritz-Pontén, T., Smets, B.F., 2018. Comparative genomics sheds light on niche differentiation and the evolutionary history of comammox Nitrospira. ISME J. 12, 1779–1793. https://doi.org/10.1038/s41396-018-0083-3

Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarshewski, A., Chaumeil, P.-A., Hugenholtz, P., 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat. Biotechnol. 36, 996. https://doi.org/10.1038/nbt.4229

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W., 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 25, 1043–55. https://doi.org/10.1101/gr.186072.114

Parks, D.H., Rinke, C., Chuvochina, M., Chaumeil, P.-A., Woodcroft, B.J., Evans, P.N., Hugenholtz, P., Tyson, G.W., 2017. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nat. Microbiol. 2, 1533–1542. https://doi.org/10.1038/s41564-017-0012-7

Parsons, C., Stüeken, E.E., Rosen, C.J., Mateos, K., Anderson, R.E., 2020. Radiation of

nitrogen-metabolizing enzymes across the tree of life tracks environmental transitions in Earth history. Geobiology. https://doi.org/10.1111/gbi.12419

Pereira, F.C., Wasmund, K., Cobankovic, I., Jehmlich, N., Herbold, C.W., Lee, K.S., Sziranyi, B., Vesely, C., Decker, T., Stocker, R., Warth, B., von Bergen, M., Wagner, M., Berry, D., 2020. Rational design of a microbial consortium of mucosal sugar utilizers reduces Clostridiodes difficile colonization. Nat. Commun. 11, 1–15. https://doi.org/10.1038/s41467-020-18928-1

Pérez, M.V., Guerrero, L.D., Orellana, E., Figuerola, E.L., Erijman, L., 2019. Time Series Genome-Centric Analysis Unveils Bacterial Response to Operational Disturbance in Activated Sludge. mSystems 4. https://doi.org/10.1128/msystems.00169-19

Petriglieri, F., Petersen, J.F., Peces, M., Nierychlo, M., Hansen, K., Baastrand, C.E., Nielsen, U.G., Reitzel, K., Nielsen, H., 2021. Quantification of biologically and chemically bound phosphorus in activated sludge from full-scale plants with biological P-removal. bioRxiv. https://doi.org/10.1101/2021.01.04.425262

Petriglieri, F., Singleton, C., Peces, M., Petersen, J.F., Nierychlo, M., Nielsen, H., 2020. “Candidatus Dechloromonas phosphatis” and “Candidatus Dechloromonas phosphovora”, two novel polyphosphate accumulating organisms abundant in wastewater treatment systems. bioRxiv. https://doi.org/10.1101/2020.11.05.369777

Pinto, A.J., Schroeder, J., Lunn, M., Sloan, W., Raskin, L., 2014. Spatial-temporal survey and occupancy-abundance modeling to predict bacterial community dynamics in the drinking water microbiomez. MBio 5. https://doi.org/10.1128/mBio.01135-14

Podar, M., Abulencia, C.B., Walcher, M., Hutchison, D., Zengler, K., Garcia, J.A., Holland, T., Cotton, D., Hauser, L., Keller, M., 2007. Targeted access to the genomes of low-abundance organisms in complex microbial communities. Appl. Environ. Microbiol. 73, 3205–3214. https://doi.org/10.1128/AEM.02985-06

Potgieter, S.C., Dai, Z., Venter, S.N., Sigudu, M., Pinto, A.J., 2020. Microbial Nitrogen Metabolism in Chloraminated Drinking Water Reservoirs. mSphere 5. https://doi.org/10.1128/msphere.00274-20

1537 Press, M.O., Wiser, A.H., Kronenberg, Z.N., Langford, K.W., Shakya, M., Lo, C.-C., Mueller,
1538 K.A., Sullivan, S.T., Chain, P.S.G., Liachko, I., 2017. Hi-C deconvolution of a human gut
1539 microbiome yields high-quality draft genomes and reveals plasmid-genome interactions.
1540 bioRxiv 198713. https://doi.org/10.1101/198713
1541 Prest, E.I., El-Chakhtoura, J., Hammes, F., Saikaly, P.E., van Loosdrecht, M.C.M.,
1542 Vrouwenvelder, J.S., 2014. Combining flow cytometry and 16S rRNA gene
1543 pyrosequencing: A promising approach for drinking water monitoring and
1544 characterization. Water Res. 63, 179–189. https://doi.org/10.1016/j.watres.2014.06.020
1545 Qiu, G., Liu, X., Saw, N.M.M.T., Law, Y., Zuniga-Montanez, R., Thi, S.S., Ngoc Nguyen, T.Q.,
1546 Nielsen, P.H., Williams, R.B.H., Wuertz, S., 2020. Metabolic Traits of Candidatus
1547 Accumulibacter clade IIF Strain SCELSE-1 Using Amino Acids As Carbon Sources for
1548 Enhanced Biological Phosphorus Removal. Environ. Sci. Technol. 54, 2448–2458.
1549 https://doi.org/10.1021/acs.est.9b02901
1550 Quince, C., Delmont, T.O., Raguideau, S., Alneberg, J., Darling, A.E., Collins, G., Eren, A.M.,
1551 2017a. DESMAN: a new tool for de novo extraction of strains from metagenomes.
1552 Genome Biol. 18, 181. https://doi.org/10.1186/s13059-017-1309-9
1553 Quince, C., Walker, A.W., Simpson, J.T., Loman, N.J., Segata, N., 2017b. Shotgun
1554 metagenomics, from sampling to analysis. Nat. Biotechnol.
1555 https://doi.org/10.1038/nbt.3935
1556 Radajewski, S., Ineson, P., Parekh, N.R., Murrell, J.C., 2000. Stable-isotope probing as a tool
1557 in microbial ecology. Nature 403, 646–649. https://doi.org/10.1038/35001054
1558 Ram, R.J., VerBerkmoes, N.C., Thelen, M.., Tyson, G.W., Baker, B.J., Blake, R.., Shah, M.,
1559 Hettich, R.L., Banfield, J.F., 2005. Community Proteomics of a Natural Microbial Biofilm.
1560 Science (80-.). 308.
1561 Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T.,
1562 Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., Aiden, E.L., 2014. A 3D map of the
1563 human genome at kilobase resolution reveals principles of chromatin looping. Cell 159,
1564 1665–1680. https://doi.org/10.1016/j.cell.2014.11.021

Raskin, L., Nielsen, P.H., 2019. Editorial overview: Integrating biotechnology and microbial ecology in urban water infrastructure through a microbiome continuum viewpoint. Curr. Opin. Biotechnol. https://doi.org/10.1016/j.copbio.2019.05.001

Raskin, L., Poulsen, L.K., Noguera, D.R., Rittmann, B.E., Stahl, D.A., 1994a. Quantification of methanogenic groups in anaerobic biological reactors by oligonucleotide probe hybridization. Appl. Environ. Microbiol. 60, 1241–1248. https://doi.org/10.1128/aem.60.4.1241-1248.1994

Raskin, L., Stromley, J.M., Rittmann, B.E., Stahl, D.A., 1994b. Group-specific 16S rRNA hybridization probes to describe natural communities of methanogens. Appl. Environ. Microbiol. 60, 1232–1240. https://doi.org/10.1128/aem.60.4.1232-1240.1994

Reed, D.C., Algar, C.K., Huber, J.A., Dick, G.J., 2014. Gene-centric approach to integrating environmental genomics and biogeochemical models. Proc. Natl. Acad. Sci. U. S. A. 111, 1879–1884. https://doi.org/10.1073/pnas.1313713111

Reichart, N.J., Jay, Z.J., Krukenberg, V., Parker, A.E., Spietz, R.L., Hatzenpichler, R., 2020. Activity-based cell sorting reveals responses of uncultured archaea and bacteria to substrate amendment. ISME J. 14, 2851–2861. https://doi.org/10.1038/s41396-020-00749-1

Reimers, A.M., Knoop, H., Bockmayr, A., Steuer, R., 2017. Cellular trade-offs and optimal resource allocation during cyanobacterial diurnal growth. Proc. Natl. Acad. Sci. U. S. A. 114, E6457–E6465. https://doi.org/10.1073/pnas.1617508114

Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N.N., Anderson, I.J., Cheng, J.F., Darling, A., Malfatti, S., Swan, B.K., Gies, E.A., Dodsworth, J.A., Hedlund, B.P., Tsiamis, G., Sievert, S.M., Liu, W.T., Eisen, J.A., Hallam, S.J., Kyrpides, N.C., Stepanauskas, R., Rubin, E.M., Hugenholtz, P., Woyke, T., 2013. Insights into the phylogeny and coding potential of microbial dark matter. Nature 499, 431–437. https://doi.org/10.1038/nature12352

Stewart, R.D., Auffret, M.D., Snelling, T.J., Roehe, R., Watson, M. 2019. MAGpy: a reproducible pipeline for the downstream analysis of metagenome-assembled genomes (MAGs), *Bioinformatics* **35**, 2150–2152,

1593 Rondon, M.R., August, P.R., Bettermann, A.D., Brady, S.F., Grossman, T.H., Liles, M.R.,
1594 Loiacono, K.A., Lynch, B.A., MacNeil, I.A., Minor, C., Tiong, C.L., Gilman, M., Osburne,
1595 M.S., Clardy, J., Handelsman, J., Goodman, R.M., 2000. Cloning the soil metagenome:
1596 A strategy for accessing the genetic and functional diversity of uncultured
1597 microorganisms. Appl. Environ. Microbiol. 66, 2541–2547.
1598 https://doi.org/10.1128/AEM.66.6.2541-2547.2000
1599 Roume, H., El Muller, E., Cordes, T., Renaut, J., Hiller, K., Wilmes, P., 2013a. A biomolecular
1600 isolation framework for eco-systems biology. ISME J. 7, 110–121.
1601 https://doi.org/10.1038/ismej.2012.72
1602 Roume, H., Heintz-Buschart, A., Muller, E.E.L., May, P., Satagopam, V.P., Laczny, C.C.,
1603 Narayanasamy, S., Lebrun, L.A., Hoopmann, M.R., Schupp, J.M., Gillece, J.D., Hicks,
1604 N.D., Engelthaler, D.M., Sauter, T., Keim, P.S., Moritz, R.L., Wilmes, P., 2015a.
1605 Comparative integrated omics: Identification of key functionalities in microbial
1606 community-wide metabolic networks. npj Biofilms Microbiomes 1, 1.
1607 https://doi.org/10.1038/npjbiofilms.2015.7
1608 Roume, H., Heintz-Buschart, A., Muller, E.E.L., May, P., Satagopam, V.P., Laczny, C.C.,
1609 Narayanasamy, S., Lebrun, L.A., Hoopmann, M.R., Schupp, J.M., Gillece, J.D., Hicks,
1610 N.D., Engelthaler, D.M., Sauter, T., Keim, P.S., Moritz, R.L., Wilmes, P., 2015b.
1611 Comparative integrated omics: Identification of key functionalities in microbial
1612 community-wide metabolic networks. npj Biofilms Microbiomes 1, 1.
1613 https://doi.org/10.1038/npjbiofilms.2015.7
1614 Roume, H., Heintz-Buschart, A., Muller, E.E.L., Wilmes, P., 2013b. Sequential isolation of
1615 metabolites, RNA, DNA, and proteins from the same unique sample, in: Methods in
1616 Enzymology. Academic Press Inc., pp. 219–236. https://doi.org/10.1016/B978-0-12-
1617 407863-5.00011-3
1618 Sachsenberg, T., Herbst, F.A., Taubert, M., Kermer, R., Jehmlich, N., Von Bergen, M., Seifert,
1619 J., Kohlbacher, O., 2015. MetaProSIP: Automated inference of stable isotope
1620 incorporation rates in proteins for functional metaproteomics. J. Proteome Res. 14, 619–

627. https://doi.org/10.1021/pr500245w

Saidi-Mehrabad, A., He, Z., Tamas, I., Sharp, C.E., Brady, A.L., Rochman, F.F., Bodrossy, L., Abell, G.C.J., Penner, T., Dong, X., Sensen, C.W., Dunfield, P.F., 2013. Methanotrophic bacteria in oilsands tailings ponds of northern Alberta. ISME J. 7, 908–921. https://doi.org/10.1038/ismej.2012.163

Sakoula, D., Smith, G.J., Frank, J., Mesman, R.J., Kop, L.F.M., Jetten, M.S.M., Kessel, M.A.H.J. van, Lücker, S., 2021. An activity-based labelling method for the detection of ammonia and methane-oxidizing bacteria. bioRxiv.

Sati, S., Cavalli, G., 2017. Chromosome conformation capture technologies and their impact in understanding genome function. Chromosoma. https://doi.org/10.1007/s00412-016-0593-6

Saunders, A.M., Albertsen, M., Vollertsen, J., Nielsen, P.H., 2016. The activated sludge ecosystem contains a core community of abundant organisms. ISME J. 10, 11–20. https://doi.org/10.1038/ismej.2015.117

Scarborough, M.J., Hamilton, J.J., Erb, E.A., Donohue, T.J., Noguera, D.R., 2020a. Diagnosing and Predicting Mixed-Culture Fermentations with Unicellular and Guild-Based Metabolic Models. mSystems 5. https://doi.org/10.1128/msystems.00755-20

Scarborough, M.J., Lawson, C.E., Hamilton, J.J., Donohue, T.J., Noguera, D.R., 2018. Metatranscriptomic and Thermodynamic Insights into Medium-Chain Fatty Acid Production Using an Anaerobic Microbiome. mSystems 3, e00221-18. https://doi.org/10.1128/mSystems.00221-18

Scarborough, M.J., Myers, K.S., Donohue, T.J., Noguera, D.R., 2020b. Medium-chain fatty acid synthesis by “candidatus weimeria bifida” gen. nov., sp. nov., and “candidatus pseudoramibacter fermentans” sp. nov. Appl. Environ. Microbiol. 86. https://doi.org/10.1128/AEM.02242-19

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., Sahl, J.W., Stres, B., Thallinger, G.G., Van Horn, D.J., Weber, C.F., 2009. Introducing mothur: open-source, platform-

1649 independent, community-supported software for describing and comparing microbial
1650 communities. Appl. Environ. Microbiol. 75, 7537–41. https://doi.org/10.1128/AEM.01541-
1651 09
1652 Schmidt, T.M., DeLong, E.F., Pace, N.R., 1991. Analysis of a marine picoplankton community
1653 by 16S rRNA gene cloning and sequencing. J. Bacteriol. 173, 4371–4378.
1654 https://doi.org/10.1128/jb.173.14.4371-4378.1991
1655 Schrimpe-Rutledge, A.C., Codreanu, S.G., Sherrod, S.D., McLean, J.A., 2016. Untargeted
1656 Metabolomics Strategies—Challenges and Emerging Directions. J. Am. Soc. Mass
1657 Spectrom. 27, 1897–1905. https://doi.org/10.1007/s13361-016-1469-y
1658 Schulz, F., Alteio, L., Goudeau, D., Ryan, E.M., Yu, F.B., Malmstrom, R.R., Blanchard, J.,
1659 Woyke, T., 2018. Hidden diversity of soil giant viruses. Nat. Commun. 9, 1–9.
1660 https://doi.org/10.1038/s41467-018-07335-2
1661 Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30, 2068–
1662 2069. https://doi.org/10.1093/bioinformatics/btu153
1663 Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O., Huttenhower, C., 2012.
1664 Metagenomic microbial community profiling using unique clade-specific marker genes.
1665 Nat. Methods 9, 811–814. https://doi.org/10.1038/nmeth.2066
1666 Seppey, M., Manni, M., Zdobnov, E.M., 2019. BUSCO: Assessing genome assembly and
1667 annotation completeness, in: Methods in Molecular Biology. Humana Press Inc., pp. 227–
1668 245. https://doi.org/10.1007/978-1-4939-9173-0_14
1669 Sevillano, M., Dai, Z., Calus, S., Bautista-de los Santos, Q.M., Eren, A.M., van der Wielen,
1670 P.W.J.J., Ijaz, U.Z., Pinto, A.J., 2020. Differential prevalence and host-association of
1671 antimicrobial resistance traits in disinfected and non-disinfected drinking water systems.
1672 Sci. Total Environ. 141451. https://doi.org/10.1016/j.scitotenv.2020.141451
1673 Sevim, V., Lee, J., Egan, R. et al. 2019. Shotgun metagenome data of a defined mock
1674 community using Oxford Nanopore, PacBio and Illumina technologies. *Scientific Data* **6**,
1675 285. https://doi.org/10.1038/s41597-019-0287-z
1676 Seviour, R.J., Blackall, L.., 1999. The Microbiology of Activated Sludge.

https://doi.org/10.1007/978-94-011-3951-9

Seviour, R.J., Nielsen, P.H., 2010. Microbial Ecology of Activated Sludge | IWA Publishing.

Shade, A., Gregory Caporaso, J., Handelsman, J., Knight, R., Fierer, N., 2013. A meta-analysis of changes in bacterial and archaeal communities with time. ISME J. 7, 1493–1506. https://doi.org/10.1038/ismej.2013.54

Shade, A., Read, J.S., Youngblut, N.D., Fierer, N., Knight, R., Kratz, T.K., Lottig, N.R., Roden, E.E., Stanley, E.H., Stombaugh, J., Whitaker, R.J., Wu, C.H., McMahon, K.D., 2012. Lake microbial communities are resilient after a whole-ecosystem disturbance. ISME J. 6, 2153–2167. https://doi.org/10.1038/ismej.2012.56

Shaffer, M., Borton, M.A., McGivern, B.B., Zayed, A.A., La Rosa, S.L., Solden, L.M., Liu, P., Narrowe, A.B., Rodríguez-Ramos, J., Bolduc, B., Gazitúa, M.C., Daly, R.A., Smith, G.J., Vik, D.R., Pope, P.B., Sullivan, M.B., Roux, S., Wrighton, K.C., 2020. DRAM for distilling microbial metabolism to automate the curation of microbiome function. Nucleic Acids Res. 48, 8883–8900. https://doi.org/10.1093/nar/gkaa621

Shaiber, A., Eren, A.M., 2019. Composite metagenome-assembled genomes reduce the quality of public genome repositories. MBio. https://doi.org/10.1128/mBio.00725-19

Shakya, M., Lo, C.C., Chain, P.S.G., 2019. Advances and challenges in metatranscriptomic analysis. Front. Genet. https://doi.org/10.3389/fgene.2019.00904

Sharon, I., Morowitz, M.J., Thomas, B.C., Costello, E.K., Relman, D.A., Banfield, J.F., 2013. Time series community genomics analysis reveals rapid shifts in bacterial species, strains, and phage during infant gut colonization. Genome Res. 23, 111–120. https://doi.org/10.1101/gr.142315.112

Sharpton, T.J., 2014. An introduction to the analysis of shotgun metagenomic data. Front. Plant Sci. https://doi.org/10.3389/fpls.2014.00209

Shaw, D.R., Ali, M., Katuri, K.P., Gralnick, J.A., Reimann, J., Mesman, R., van Niftrik, L., Jetten, M.S.M., Saikaly, P.E., 2020. Extracellular electron transfer-dependent anaerobic oxidation of ammonium by anammox bacteria. Nat. Commun. 11, 1–12. https://doi.org/10.1038/s41467-020-16016-y

Sheik, A.R., Muller, E.E.L., Wilmes, P., 2014. A hundred years of activated sludge: Time for a rethink. Front. Microbiol. 5. https://doi.org/10.3389/fmicb.2014.00047

Sieber, C.M.K., Probst, A.J., Sharrar, A., Thomas, B.C., Hess, M., Tringe, S.G., Banfield, J.F., 2018. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nat. Microbiol. 1. https://doi.org/10.1038/s41564-018-0171-1

Singleton, C.M., Petriglieri, F., Kristensen, J.M., Kirkegaard, R.H., Michaelsen, T.Y., Andersen, M.H., Kondrotaite, Z., Karst, S.M., Dueholm, M.S., Nielsen, P.H., Albertsen, M., 2021. Connecting structure to function with the recovery of over 1000 high-quality metagenome-assembled genomes from activated sludge using long-read sequencing. Nat. Commun. 12, 2009. https://doi.org/10.1038/s41467-021-22203-2

Singleton, C.M., Petriglieri, F., Kristensen, J.M., Kirkegaard, R.H., Michaelsen, T.Y., Andersen, M.H., Kondrotaite, Z., Karst, S.M., Dueholm, M.S., Nielsen, P.H., Albertsen, M., 2020. Connecting structure to function with the recovery of over 1000 high-quality activated sludge metagenome-assembled genomes encoding full-length rRNA genes using long-read sequencing. bioRxiv 2020.05.12.088096. https://doi.org/10.1101/2020.05.12.088096

Skennerton, C.T., Barr, J.J., Slater, F.R., Bond, P.L., Tyson, G.W., 2015. Expanding our view of genomic diversity in *C andidatus* Accumulibacter clades. Environ. Microbiol. 17, 1574–1585. https://doi.org/10.1111/1462-2920.12582

Smukowski Heil, C., Burton, J.N., Liachko, I., Friedrich, A., Hanson, N.A., Morris, C.L., Schacherer, J., Shendure, J., Thomas, J.H., Dunham, M.J., 2018. Identification of a novel interspecific hybrid yeast from a metagenomic spontaneously inoculated beer sample using Hi-C. Yeast 35, 71–84. https://doi.org/10.1002/yea.3280

Sogin, M.L., Morrison, H.G., Huber, J.A., Welch, D.M., Huse, S.M., Neal, P.R., Arrieta, J.M., Herndl, G.J., 2006. Microbial diversity in the deep sea and the underexplored “rare biosphere.” Proc. Natl. Acad. Sci. U. S. A. 103, 12115–12120. https://doi.org/10.1073/pnas.0605127103

Somerville, V., Lutz, S., Schmid, M., Frei, D., Moser, A., Irmler, S., Frey, J.E., Ahrens, C.H.,

2019. Long-read based de novo assembly of low-complexity metagenome samples results in finished genomes and reveals insights into strain diversity and an active phage system. BMC Microbiol. 19, 1–18. https://doi.org/10.1186/s12866-019-1500-0

Speth, D.R., Hu, B., Bosch, N., Keltjens, J.T., Stunnenberg, H.G., Jetten, M.S.M., 2012. Comparative genomics of two independently enriched “Candidatus Kuenenia stuttgartiensis” anammox bacteria. Front. Microbiol. 3. https://doi.org/10.3389/fmicb.2012.00307

Speth, D.R., In’T Zandt, M.H., Guerrero-Cruz, S., Dutilh, B.E., Jetten, M.S.M., 2016. Genome-based microbial ecology of anammox granules in a full-scale wastewater treatment system. Nat. Commun. 7, 1–10. https://doi.org/10.1038/ncomms11172

Srinivasan, V.N., Li, G., Wang, D., Tooker, N.B., Dai, Z., Onnis-Hayden, A., Pinto, A., Gu, A.Z., 2019. Oligotyping and Genome-Resolved Metagenomics Reveal Distinct Candidatus Accumulibacter Communities in Full-Scale Side-Stream versus Conventional Enhanced Biological Phosphorus Removal (EBPR) Configurations. bioRxiv 596692. https://doi.org/10.1101/596692

Stackebrandt, E., Goebel, B.M., 1994. Taxonomic note: A place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. Int. J. Syst. Bacteriol. https://doi.org/10.1099/00207713-44-4-846

Stahl, D.A., Flesher, B., Mansfield, H.R., Montgomery, L., 1988. Use of phylogenetically based hybridization probes for studies of ruminal microbial ecology. Appl. Environ. Microbiol. 54, 1079–1084. https://doi.org/10.1128/aem.54.5.1079-1084.1988

Stalder, T., Press, M.O., Sullivan, S., Liachko, I., Top, E.M., 2019. Linking the resistome and plasmidome to the microbiome. ISME J. 13, 2437–2446. https://doi.org/10.1038/s41396-019-0446-4

Steen, A.D., Crits-Christoph, A., Carini, P., DeAngelis, K.M., Fierer, N., Lloyd, K.G., Cameron Thrash, J., 2019. High proportions of bacteria and archaea across most biomes remain uncultured. ISME J. 13, 3126–3130. https://doi.org/10.1038/s41396-019-0484-y

Stewart, F.J., Ottesen, E.A., Delong, E.F., 2010. Development and quantitative analyses of a

universal rRNA-subtraction protocol for microbial metatranscriptomics. ISME J. 4, 896–907. https://doi.org/10.1038/ismej.2010.18

Stewart, R.D., Auffret, M.D., Warr, A., Walker, A.W., Roehe, R., Watson, M., 2019. Compendium of 4,941 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery. Nat. Biotechnol. 37, 953–961. https://doi.org/10.1038/s41587-019-0202-3

Stewart, R.D., Auffret, M.D., Warr, A., Wiser, A.H., Press, M.O., Langford, K.W., Liachko, I., Snelling, T.J., Dewhurst, R.J., Walker, A.W., Roehe, R., Watson, M., 2018. Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. Nat. Commun. 9, 870. https://doi.org/10.1038/s41467-018-03317-6

Stolyar, S., Van Dien, S., Hillesland, K.L., Pinel, N., Lie, T.J., Leigh, J.A., Stahl, D.A., 2007. Metabolic modeling of a mutualistic microbial community, *Molecular Systems Biology*, **3**, 92

Strous, M., Pelletier, E., Mangenot, S., Rattei, T., Lehner, A., Taylor, M.W., Horn, M., Daims, H., Bartol-Mavel, D., Wincker, P., Barbe, V., Fonknechten, N., Vallenet, D., Segurens, B., Schenowitz-Truong, C., Médigue, C., Collingro, A., Snel, B., Dutilh, B.E., Op Den Camp, H.J.M., Van Der Drift, C., Cirpus, I., Van De Pas-Schoonen, K.T., Harhangi, H.R., Van Niftrik, L., Schmid, M., Keltjens, J., Van De Vossenberg, J., Kartal, B., Meier, H., Frishman, D., Huynen, M.A., Mewes, H.W., Weissenbach, J., Jetten, M.S.M., Wagner, M., Le Paslier, D., 2006. Deciphering the evolution and metabolism of an anammox bacterium from a community genome. Nature 440, 790–794. https://doi.org/10.1038/nature04647

Swan, B.K., Martinez-Garcia, M., Preston, C.M., Sczyrba, A., Woyke, T., Lamy, D., Reinthaler, T., Poulton, N.J., Masland, E.D.P., Gomez, M.L., Sieracki, M.E., DeLong, E.F., Herndl, G.J., Stepanauskas, R., 2011. Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean. Science (80-.). 333, 1296–1300. https://doi.org/10.1126/science.1203690

Tan, S.M., Yung, P.Y.M., Hutchinson, P.E., Xie, C., Teo, G.H., Ismail, M.H., Drautz-Moses,

D.I., Little, P.F.R., Williams, R.B.H., Cohen, Y., 2019. Primer-free FISH probes from metagenomics/metatranscriptomics data permit the study of uncharacterised taxa in complex microbial communities. npj Biofilms Microbiomes 5, 1–13. https://doi.org/10.1038/s41522-019-0090-9

Thompson, L.R., Sanders, J.G., McDonald, D., Amir, A., Ladau, J., Locey, K.J., Prill, R.J., Tripathi, A., Gibbons, S.M., Ackermann, G., Navas-Molina, J.A., Janssen, S., Kopylova, E., Vázquez-Baeza, Y., González, A., Morton, J.T., Mirarab, S., Xu, Z.Z., Jiang, L., Haroon, M.F., Kanbar, J., Zhu, Q., Song, S.J., Kosciolek, T., Bokulich, N.A., Lefler, J., Brislawn, C.J., Humphrey, G., Owens, S.M., Hampton-Marcell, J., Berg-Lyons, D., McKenzie, V., Fierer, N., Fuhrman, J.A., Clauset, A., Stevens, R.L., Shade, A., Pollard, K.S., Goodwin, K.D., Jansson, J.K., Gilbert, J.A., Knight, R., Agosto Rivera, J.L., Al-Moosawi, L., Alverdy, J., Amato, K.R., Andras, J., Angenent, L.T., Antonopoulos, D.A., Apprill, A., Armitage, D., Ballantine, K., Bárta, J., Baum, J.K., Berry, A., Bhatnagar, A., Bhatnagar, M., Biddle, J.F., Bittner, L., Boldgiv, B., Bottos, E., Boyer, D.M., Braun, J., Brazelton, W., Brearley, F.Q., Campbell, A.H., Caporaso, J.G., Cardona, C., Carroll, J.L., Cary, S.C., Casper, B.B., Charles, T.C., Chu, H., Claar, D.C., Clark, R.G., Clayton, J.B., Clemente, J.C., Cochran, A., Coleman, M.L., Collins, G., Colwell, R.R., Contreras, M., Crary, B.B., Creer, S., Cristol, D.A., Crump, B.C., Cui, D., Daly, S.E., Davalos, L., Dawson, R.D., Defazio, J., Delsuc, F., Dionisi, H.M., Dominguez-Bello, M.G., Dowell, R., Dubinsky, E.A., Dunn, P.O., Ercolini, D., Espinoza, R.E., Ezenwa, V., Fenner, N., Findlay, H.S., Fleming, I.D., Fogliano, V., Forsman, A., Freeman, C., Friedman, E.S., Galindo, G., Garcia, L., Garcia-Amado, M.A., Garshelis, D., Gasser, R.B., Gerdts, G., Gibson, M.K., Gifford, I., Gill, R.T., Giray, T., Gittel, A., Golyshin, P., Gong, D., Grossart, H.P., Guyton, K., Haig, S.J., Hale, V., Hall, R.S., Hallam, S.J., Handley, K.M., Hasan, N.A., Haydon, S.R., Hickman, J.E., Hidalgo, G., Hofmockel, K.S., Hooker, J., Hulth, S., Hultman, J., Hyde, E., Ibáñez-Álamo, J.D., Jastrow, J.D., Jex, A.R., Johnson, L.S., Johnston, E.R., Joseph, S., Jurburg, S.D., Jurelevicius, D., Karlsson, A., Karlsson, R., Kauppinen, S., Kellogg, C.T.E., Kennedy, S.J., Kerkhof, L.J., King, G.M., Kling, G.W., Koehler, A. V.,

Krezalek, M., Kueneman, J., Lamendella, R., Landon, E.M., Lanede Graaf, K., LaRoche, J., Larsen, P., Laverock, B., Lax, S., Lentino, M., Levin, I.I., Liancourt, P., Liang, W., Linz, A.M., Lipson, D.A., Liu, Y., Lladser, M.E., Lozada, M., Spirito, C.M., MacCormack, W.P., MacRae-Crerar, A., Magris, M., Martín-Platero, A.M., Martín-Vivaldi, M., Martínez, L.M., Martínez-Bueno, M., Marzinelli, E.M., Mason, O.U., Mayer, G.D., McDevitt-Irwin, J.M., McDonald, J.E., McGuire, K.L., McMahon, K.D., McMinds, R., Medina, M., Mendelson, J.R., Metcalf, J.L., Meyer, F., Michelangeli, F., Miller, K., Mills, D.A., Minich, J., Mocali, S., Moitinho-Silva, L., Moore, A., Morgan-Kiss, R.M., Munroe, P., Myrold, D., Neufeld, J.D., Ni, Y., Nicol, G.W., Nielsen, S., Nissimov, J.I., Niu, K., Nolan, M.J., Noyce, K., O'Brien, S.L., Okamoto, N., Orlando, L., Castellano, Y.O., Osuolale, O., Oswald, W., Parnell, J., Peralta-Sánchez, J.M., Petraitis, P., Pfister, C., Pilon-Smits, E., Piombino, P., Pointing, S.B., Pollock, F.J., Potter, C., Prithiviraj, B., Quince, C., Rani, A., Ranjan, R., Rao, S., Rees, A.P., Richardson, M., Riebesell, U., Robinson, C., Rockne, K.J., Rodriguezl, S.M., Rohwer, F., Roundstone, W., Safran, R.J., Sangwan, N., Sanz, V., Schrenk, M., Schrenzel, M.D., Scott, N.M., Seger, R.L., Seguinorlando, A., Seldin, L., Seyler, L.M., Shakhsheer, B., Sheets, G.M., Shen, C., Shi, Y., Shin, H., Shogan, B.D., Shutler, D., Siegel, J., Simmons, S., Sjöling, S., Smith, D.P., Soler, J.J., Sperling, M., Steinberg, P.D., Stephens, B., Stevens, M.A., Taghavi, S., Tai, V., Tait, K., Tan, C.L., Taş, N., Taylor, D.L., Thomas, T., Timling, I., Turner, B.L., Urich, T., Ursell, L.K., Van Der Lelie, D., Van Treuren, W., Van Zwieten, L., Vargas-Robles, D., Thurber, R.V., Vitaglione, P., Walker, D.A., Walters, W.A., Wang, S., Wang, T., Weaver, T., Webster, N.S., Wehrle, B., Weisenhorn, P., Weiss, S., Werner, J.J., West, K., Whitehead, A., Whitehead, S.R., Whittingham, L.A., Willerslev, E., Williams, A.E., Wood, S.A., Woodhams, D.C., Yang, Y., Zaneveld, J., Zarraonaindia, I., Zhang, Q., Zhao, H., 2017. A communal catalogue reveals Earth's multiscale microbial diversity. Nature 551, 457–463. https://doi.org/10.1038/nature24621

Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., Solovyev, V. V., Rubin, E.M., Rokhsar, D.S., Banfield, J.F., 2004. Community structure

and metabolism through reconstruction of microbial genomes from the environment. Nature 428, 37–43. https://doi.org/10.1038/nature02340

Uritskiy, G.V., DiRuggiero, J., Taylor, J. 2018. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* **6,** 158.

Van Kessel, M.A.H.J., Speth, D.R., Albertsen, M., Nielsen, P.H., Op Den Camp, H.J.M., Kartal, B., Jetten, M.S.M., Lücker, S., 2015. Complete nitrification by a single microorganism. Nature 528, 555–559. https://doi.org/10.1038/nature16459

Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J. a, Wu, D., Paulsen, I., Nelson, K.E., Nelson, W., Fouts, D.E., Levy, S., Knap, A.H., Lomas, M.W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y.-H., Smith, H.O., 2004. Environmental Genome Sshotgun Sequencing of the Sargasso Sea. Science 304, 66–74. https://doi.org/10.1126/science.1093857

Von Bergen, M., Jehmlich, N., Taubert, M., Vogt, C., Bastida, F., Herbst, F.A., Schmidt, F., Richnow, H.H., Seifert, J., 2013. Insights from quantitative metaproteomics and protein-stable isotope probing into microbial ecology. ISME J. https://doi.org/10.1038/ismej.2013.78

Vosloo, S., Huo L.-X., Anderson, C.L., Dai, Z., Sevillano, M., Pinto A. (2021). Evaluating de novo assembly and binning strategies for time-series drinking water metagenomes, *bioRxiv* 2021.07.11.451960; doi: https://doi.org/10.1101/2021.07.11.451960

Wagner, M., Amann, R., Kämpfer, P., Assmus, B., Hartmann, A., Hutzler, P., Springer, N., Schleifer, K.H., 1994. Identification and in situ Detection of Gram-negative Filamentous Bacteria in Activated Sludge. Syst. Appl. Microbiol. 17, 405–417. https://doi.org/10.1016/S0723-2020(11)80058-5

Wagner, M., Haider, S., 2012. New trends in fluorescence in situ hybridization for identification and functional analyses of microbes. Curr. Opin. Biotechnol. https://doi.org/10.1016/j.copbio.2011.10.010

Wang, Y., Gao, H., Wells, G., 2020. Integrated Omics Analyses Reveal Differential Gene

Expression and Potential for Cooperation Between Denitrifying Polyphosphate and Glycogen Accumulating Organisms. bioRxiv. https://doi.org/10.1101/2020.01.10.901413

Wang, Y., Niu, Q., Zhang, X., Liu, L., Wang, Y., Chen, Y., Negi, M., Figeys, D., Li, Y.Y., Zhang, T., 2019. Exploring the effects of operational mode and microbial interactions on bacterial community assembly in a one-stage partial-nitritation anammox reactor using integrated multi-omics. Microbiome 7, 122. https://doi.org/10.1186/s40168-019-0730-6

Weiland, P., 2010. Biogas production: Current state and perspectives. Appl. Microbiol. Biotechnol. https://doi.org/10.1007/s00253-009-2246-7

Weinrich, S., Koch, S., Bonk, F., Popp, D., Benndorf, D., Klamt, S., Centler, F., 2019. Augmenting Biogas Process Modeling by Resolving Intracellular Metabolic Activity. Front. Microbiol. 10, 1095. https://doi.org/10.3389/fmicb.2019.01095

Westermann, A.J., Gorski, S.A., Vogel, J., 2012. Dual RNA-seq of pathogen and host. Nat. Rev. Microbiol. https://doi.org/10.1038/nrmicro2852

Wilmes, P., Andersson, A.F., Lefsrud, M.G., Wexler, M., Shah, M., Zhang, B., Hettich, R.L., Bond, P.L., VerBerkmoes, N.C., Banfield, J.F., 2008. Community proteogenomics highlights microbial strain-variant protein expression within activated sludge performing enhanced biological phosphorus removal. ISME J. 2, 853–864. https://doi.org/10.1038/ismej.2008.38

Wilmes, P., Bond, P.L., 2009. Microbial community proteomics: elucidating the catalysts and metabolic mechanisms that drive the Earth's biogeochemical cycles. Curr. Opin. Microbiol. https://doi.org/10.1016/j.mib.2009.03.004

Wilmes, P., Bond, P.L., 2006. Metaproteomics: Studying functional gene expression in microbial ecosystems. Trends Microbiol. https://doi.org/10.1016/j.tim.2005.12.006

Woese, C.R., Fox, G.E., 1977. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. Proc. Natl. Acad. Sci. 74, 5088–5090. https://doi.org/10.1073/pnas.74.11.5088

Woodcroft, B.J., Singleton, C.M., Boyd, J.A., Evans, P.N., Emerson, J.B., Zayed, A.A.F., Hoelzle, R.D., Lamberton, T.O., McCalley, C.K., Hodgkins, S.B., Wilson, R.M., Purvine,

S.O., Nicora, C.D., Li, C., Frolking, S., Chanton, J.P., Crill, P.M., Saleska, S.R., Rich, V.I., Tyson, G.W., 2018. Genome-centric view of carbon processing in thawing permafrost. Nature 560, 49–54. https://doi.org/10.1038/s41586-018-0338-1

Wrighton, K.C., Thomas, B.C., Sharon, I., Miller, C.S., Castelle, C.J., VerBerkmoes, N.C., Wilkins, M.J., Hettich, R.L., Lipton, M.S., Williams, K.H., Long, P.E., Banfield, J.F., 2012. Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. Science 337, 1661–5. https://doi.org/10.1126/science.1224041

Wu, L., Ning, D., Zhang, B., Li, Y., Zhang, P., Shan, X., Zhang, Q., Brown, M., Li, Z., Van Nostrand, J.D., Ling, F., Xiao, N., Zhang, Y., Vierheilig, J., Wells, G.F., Yang, Y., Deng, Y., Tu, Q., Wang, A., Zhang, T., He, Z., Keller, J., Nielsen, P.H., Alvarez, P.J.J., Criddle, C.S., Wagner, M., Tiedje, J.M., He, Q., Curtis, T.P., Stahl, D.A., Alvarez-Cohen, L., Rittmann, B.E., Wen, X., Zhou, J., 2019. Global diversity and biogeography of bacterial communities in wastewater treatment plants. Nat. Microbiol. 1. https://doi.org/10.1038/s41564-019-0426-5

Wu, Y.-W., Tang, Y.-H., Tringe, S.G., Simmons, B.A., Singer, S.W., 2014. MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. Microbiome 2, 26. https://doi.org/10.1186/2049-2618-2-26

Xu, Y., Zhao, F., 2018. Single-cell metagenomics: challenges and applications. Protein Cell. https://doi.org/10.1007/s13238-018-0544-5

Ye, L., Mei, R., Liu, W.T., Ren, H., Zhang, X.X., 2020. Machine learning-aided analyses of thousands of draft genomes reveal specific features of activated sludge processes. Microbiome 8, 16. https://doi.org/10.1186/s40168-020-0794-3

Yilmaz, S., Haroon, M.F., Rabkin, B.A., Tyson, G.W., Hugenholtz, P., 2010. Fixation-free fluorescence in situ hybridization for targeted enrichment of microbial populations. ISME J. 4, 1352–1356. https://doi.org/10.1038/ismej.2010.73

Yu, F.B., Blainey, P.C., Schulz, F., Woyke, T., Horowitz, M.A., Quake, S.R., 2017. Microfluidic-based mini-metagenomics enables discovery of novel microbial lineages from complex

environmental samples. Elife 6. https://doi.org/10.7554/eLife.26580

Zhang, T., Shao, M.F., Ye, L., 2012. 454 Pyrosequencing reveals bacterial diversity of activated sludge from 14 sewage treatment plants. ISME J. 6, 1137–1147. https://doi.org/10.1038/ismej.2011.188

Zhou, Z., Tran, P., Liu, Y., Kieft, K., Anantharaman, K., 2019. METABOLIC: A scalable high-throughput metabolic and biogeochemical functional trait profiler based on microbial genomes. bioRxiv 761643. https://doi.org/10.1101/761643

Ziels, R.M., Nobu, M.K., Sousa, D.Z., 2019. Elucidating Syntrophic Butyrate-Degrading Populations in Anaerobic Digesters Using Stable-Isotope-Informed Genome-Resolved Metagenomics. mSystems 4. https://doi.org/10.1128/msystems.00159-19

Ziels, R.M., Sousa, D.Z., Stensel, H.D., Beck, D.A.C., 2018. DNA-SIP based genome-centric metagenomics identifies key long-chain fatty acid-degrading populations in anaerobic digesters with different feeding frequencies. ISME J. 12, 112–123. https://doi.org/10.1038/ismej.2017.143

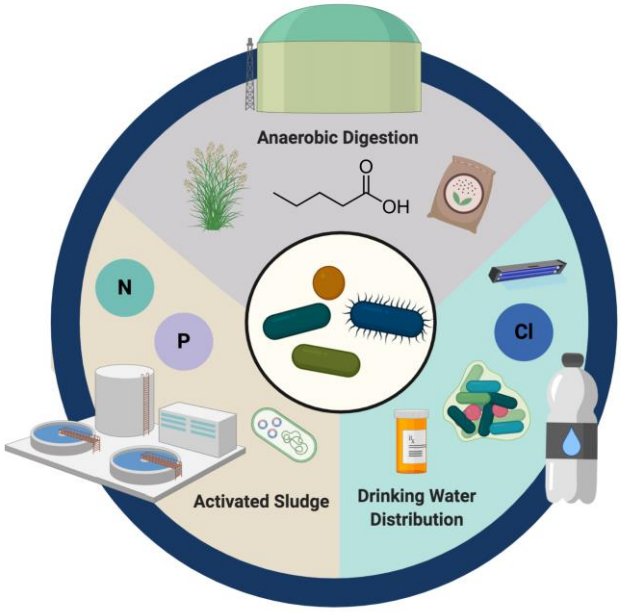

Figure 1. Engineered water systems unified through microbial ecology: engineered water systems are unified through microbial ecology in that they all aim to manage diverse microbial communities for desired outcomes. *Anerobic Digestion* (AD) systems apply microbial consortia to produce valuable bioenergy products from waste streams. *Activated Sludge* (AS) wastewater treatment systems harness the unique metabolic physiologies of specific microorganisms to remove nutrients such as nitrogen and phosphorus. *Drinking Water Distribution Systems* (DWDS) aim to manage microbial pathogens and biofilms to provide safe, clean drinking water. Created with BioRender.com.

Figure 2

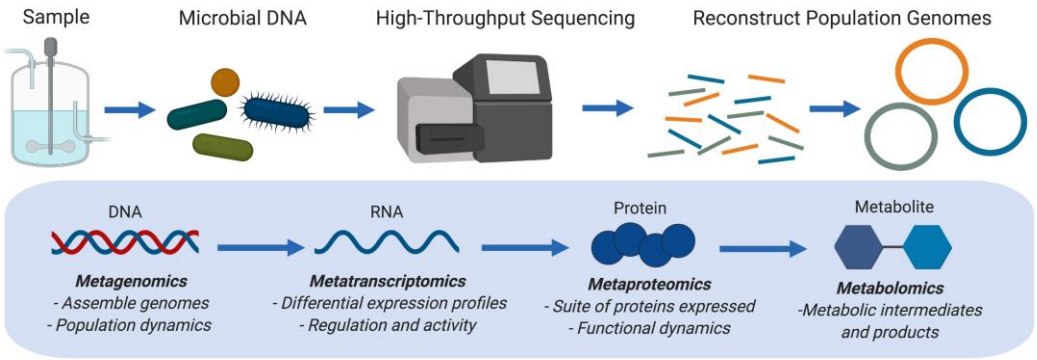

Figure 2. Genome-resolved multi-omics approaches: genome-resolved approaches can be applied individually at the genomic, transcriptomic, and proteomic levels, or in conjunction with another. Microbial DNA is extracted from the environment of interest and used for high-throughput shotgun sequencing. Short reads are assembled into longer contigs, which are used to bin into population genomes based on sequence composition and differential coverage through space and/or time. Typical workflows for binning and downstream analyses of assembled genomes are outlined in Figure 3. Population genomes from short-read data usually represent a consensus sequence of closely related strains in a sample. Created with BioRender.com.

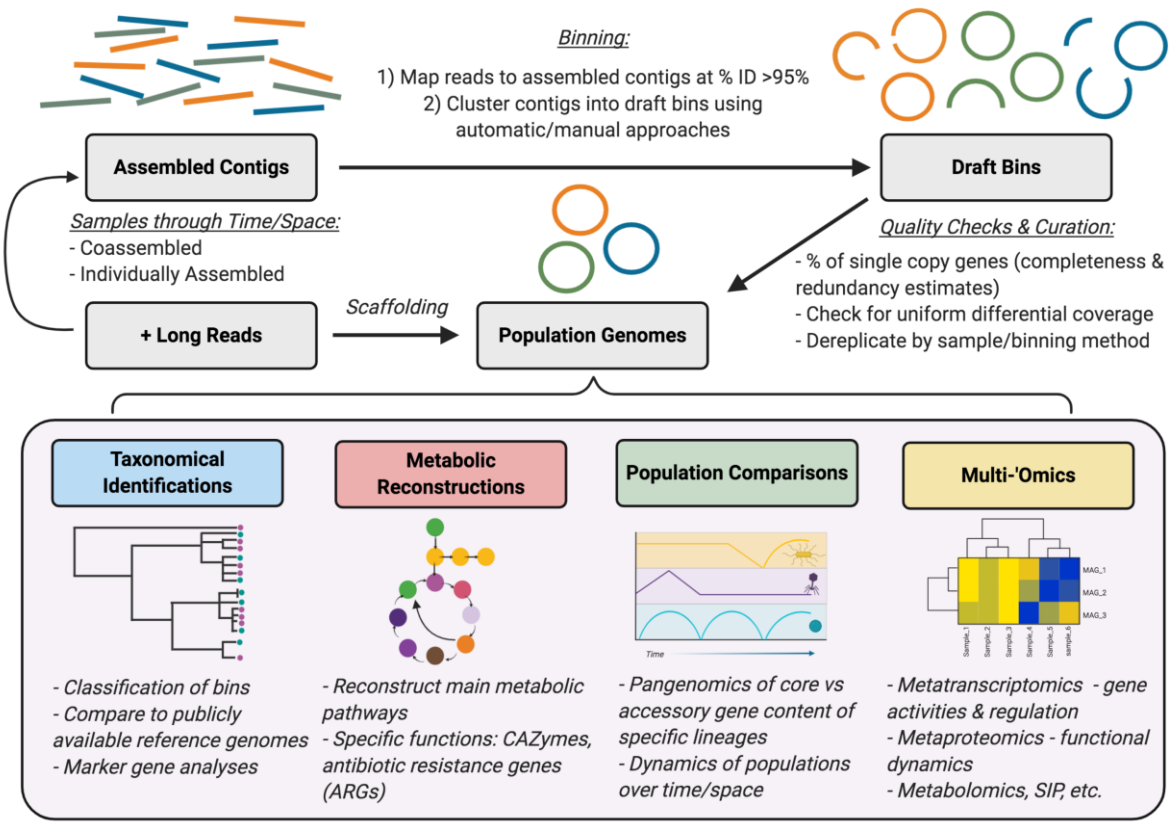

Figure 3. Workflow for binning population genomes: from sequenced metagenomes, samples are either assembled individually or coassembled together into longer contigs. These contigs are binned into draft bins by mapping reads back to assembled contigs at a percent identity >95%, and clustered by differential coverage profiles over space and/or time and/or using nucleotide signatures. These draft bins are then quality checked and curated based on the presence of single copy genes, checked for uniform differential coverage across all contigs, and dereplicating by sample or binning method. The resulting set of population genomes represent non-redundant, “species”-level bins that can be recovered from the community. Long reads can be incorporated either during initial assembly process with polishing tools to improve contigs lengths, used for scaffolding individual bins, or potentially as the basis for genome recovery (not shown). Population genomes can then be used for a variety of analyses to understand the given ecosystem – phylogenomic comparisons, metabolic reconstructions, dynamics of specific populations, and/or integrating with other multi-omics approaches. Created with BioRender.com.

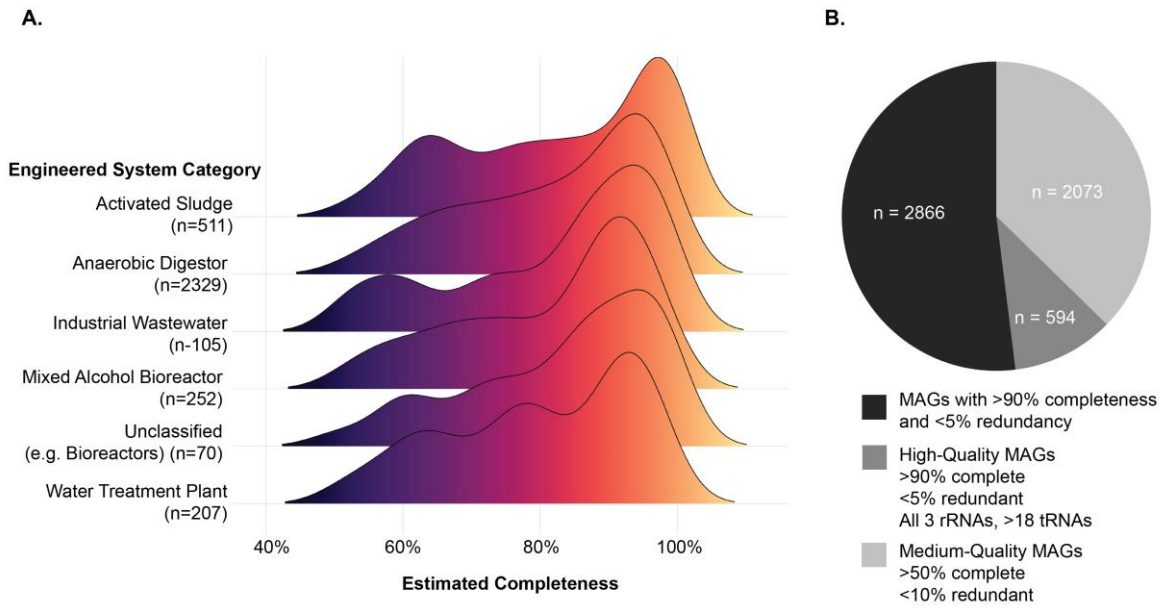

Figure 4. Quality of MAGs from engineered water systems: genome information from the Genomes from Earth's Microbiomes (GEM) catalog (Nayfach et al., 2020). From this dataset, genomes were subset by ecosystem and ecosystem category to only include genomes from engineered water systems (activated sludge, anaerobic digestion, drinking water systems, and lab-scale bioreactors simulating these processes). From 52,515 assembled MAGs, 3,474 MAGs were determined to be recovered from engineered water systems. **A.** Distribution of estimated completeness of all MAGs from different environmental categories of engineered water systems. **B.** Distribution of MAGs from engineered water systems that full under the categories high-quality, medium-quality, and genomes that are >90% complete and <5% redundant but do not contain all 3 rRNA genes and at least 18 tRNAs, as defined by the MIMAG standards (Bowers et al., 2017).

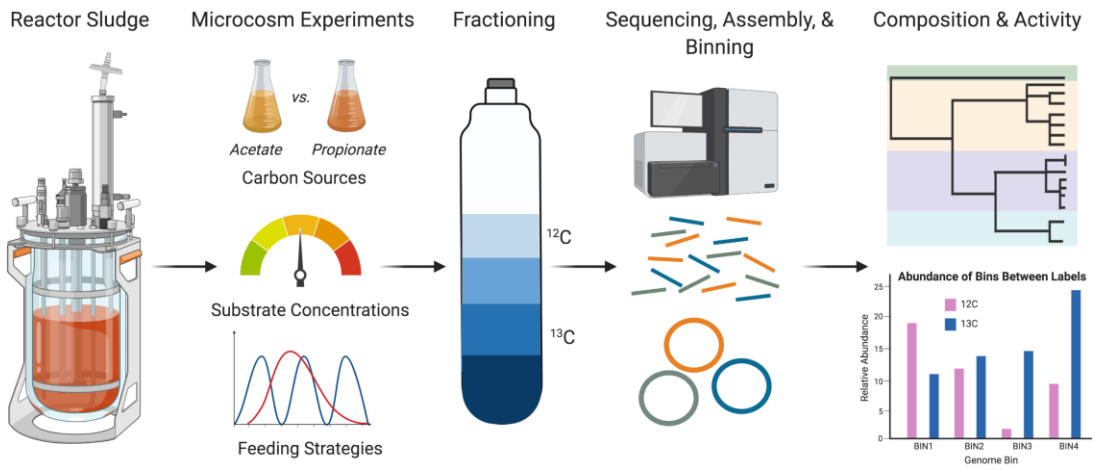

Figure 5. DNA Stable Isotope Probing (SIP) paired with genome-resolved metagenomics: DNA stable isotope probing (SIP) can identify active populations in a mixed community. Reactor sludge is inoculated into microcosms used for incubation experiments such as testing differences between carbon sources, substrate concentrations, or feeding strategies on community dynamics. Labelled DNA is density centrifuged and separated between light and heavy labelled DNA. The fractions are sequenced and assembled. Individual assembly of metagenomic contigs and differential coverage binning can then yield insights into the activities of specific populations between conditions. Created with BioRender.com.

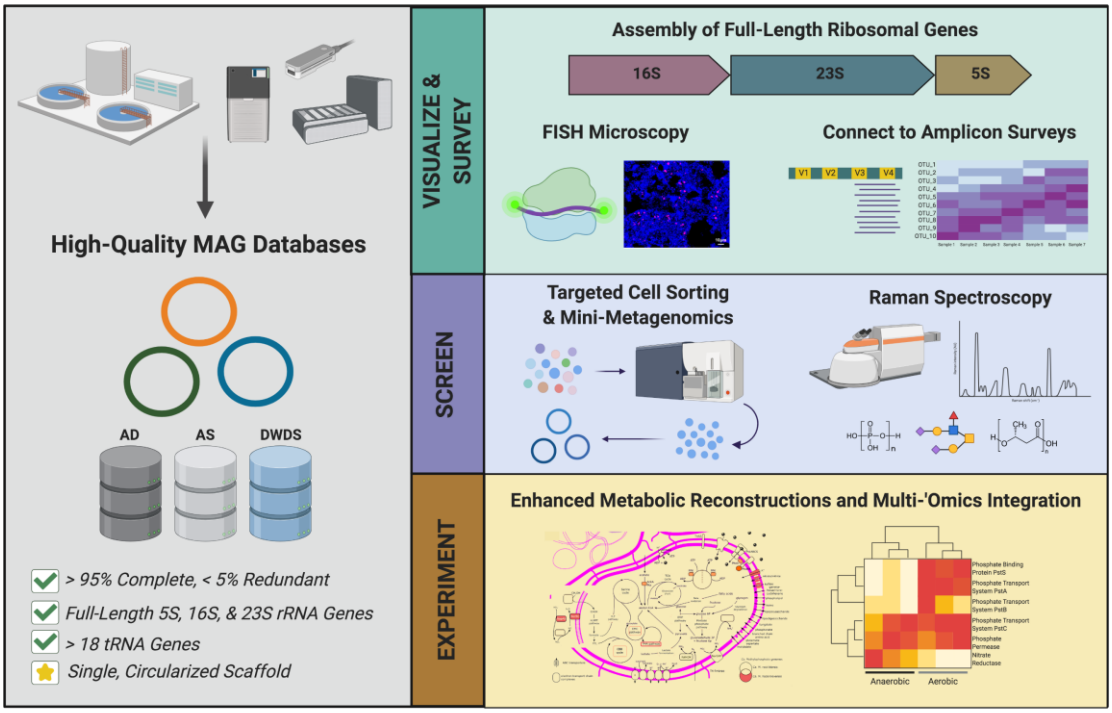

Figure 6. Emerging technologies and future prospects for multi-omics of engineered water microbiomes: improvements and decreasing costs in long-read sequencing technologies paired with novel experimental technologies have paved the way for exciting discoveries in engineered water systems. With genome-resolved metagenomic sequencing surveys of AD, AS, and DWDS ecosystems, high-quality genome databases can be constructed for downstream experiments. The assembly of full-length ribosomal genes from MAGs will allow for enhanced FISH probe construction and comparisons to amplicon sequencing studies. These genomes can then be used for the basis for screening for specific activities by targeted cell sorting or Raman spectroscopy. Multi-omics experiments can then be applied to understand the metabolism, regulation, and dynamics of key microbial guilds in these systems. We envision that these approaches will become commonplace and iterative of each other, such as demonstrated in Singleton et al. 2020. Created with BioRender.com.