White Paper on Genomic Sequencing

Part of the 2015 RAD-02 Goal Share:

Efficient Delivery of Laboratory Services

Authors: Leila Barker, Susanna Blunt,
and Cindy Covey
with oversight from
Bob Baumgartner, Steve Thompson,
and Dr. Ken Williamson

June 2015

White Paper on Genomic Sequencing June 2015

This paper was written in partial fulfillment of the Regulatory Affairs Department's Goal Share Measure RAD-02 (FY 2015) to address efficient delivery of laboratory services. The team (Leila Barker, Susanna Blunt, and Cindy Covey, with oversight from Bob Baumgartner, Ken Williamson, and Steve Thompson) met between February and June 2015 to research and compile findings.

Introduction

Clean Water Services (District) is interested in pursuing genomic techniques for identifying and quantifying populations of microbes in wastewater treatment plant (WWTP) processes. Understanding both the identities and metabolic capabilities of the dominant microbes present in treatment operations could allow for proactive responses to changing influent characteristics or upsets, and could improve operational efficiencies. Technologies for monitoring microbial genomics have advanced rapidly in recent years. The District's Water Quality Laboratory is investigating the future use and utility of genomic techniques as a complementary tool alongside standard analytical chemistry for monitoring the composition and activity of biological treatment processes.

The following report summarizes some tools that are commonly used for genomic analysis and provides an overview of studies on the composition and performance of wastewater bacterial communities. Additionally, the report will introduce the tools and methods needed to preserve and process genomic samples and to interpret sample data.

Background

Genetic analysis has shown that wastewater microbial communities are highly complex, and little is known about their composition and dynamics [1]. At present, analysis of biological treatment processes relies on mass balances and measurement of reactant and product compounds, with the system often regarded as something of a "black box." In recent years, advancements in genome and metagenome sequencing strategies, high-throughput sequencing platforms, and bioinformatics toolboxes have made genomic analysis of complex mixed microbial communities more accessible. The ability to identify and analyze microorganisms could enable the District to monitor treatment processes with the goal of optimizing performance and detecting early indications of changes in the communities.

Genomics is a subset of the field of genetics that deals with sequencing and analyzing the function and structure of the entire genome (the complete set of DNA within an organism). Metagenomics (also known as environmental genomics or community genomics) is the study of metagenomes, the set of genetic material recovered from environmental samples. Metagenomics is distinct from traditional genomics or microbiology in that it deals with mixed microbial communities (as opposed to pure cell cultures) and allows researchers to analyze microbial biodiversity in environmental samples. Methods that allow for cloning of specific common genes can provide representative samples of the microorganisms that make up sampled communities.

Modern genomics techniques offer an unprecedented opportunity to explore the population and gene diversity in WWTP processes. One application is for identification of species present in reactors. Genetic markers can be used to quantify populations of specific species of ammonia-oxidizing bacteria (AOBs), nitrite-oxidizing bacteria (NOBs), denitrifying bacteria, and polyphosphate-accumulating organisms

(PAOs). The populations of microbial communities in WWTP reactors have been found to vary significantly due to configuration and operating conditions as well as influent characteristics, so it is impossible to predict relative concentrations of bacterial species in a previously unstudied system.

Identification of key species and functional genes might enable the District to make operational decisions geared toward optimizing conditions for the particular species present. Gene or species quantification could identify the bacteria that are most active in a mixed community. Repeated sampling could help to establish a profile of genetic variation, and concentrations of key genes could be compared to recent and historical samples. Significant changes in gene concentrations might prove to be an early indicator of trouble, or could help pinpoint the source of observed decreases in activity.

> *"Most biokinetic estimation studies are based on mathematically approximated concentrations of the nitrifying communities in mixed culture…. There are few, if any biokinetic descriptors of nitrifying bacteria in mixed communities that are based on direct measures of AOB or NOB abundance. In the absence of such direct measurements, the estimated coefficients are mere approximations and could lead to erroneous bioreactor design, operating and monitoring strategies." (Ahn et al., 2008 [2])*

Potential Application: Biological Phosphorus Removal

Enhanced biological phosphorus removal (EBPR) is performed at the Durham and Rock Creek WWTFs. While EBPR is an economically advantageous and environmentally friendly process for achieving low effluent phosphorus concentrations, EBPR can be unstable and difficult to control, and troubleshooting is complicated due to an incomplete understanding of EBPR microbiology and community composition [3, 4]. The District is interested in improving EBPR stability. Understanding the identities of PAOs in process reactors, as well as the composition of the PAO community under different loadings, could help to promote conditions favoring phosphorus (P) removal. Genomic data could be used to determine if (and at what rate) addition of volatile fatty acids (VFAs) in the anaerobic zone is an effective tactic for promoting PAO populations. Comparing PAO genomic data to past seasons could provide information as to whether the system is likely to function as desired.

Genomic Studies on WWTP Mixed Bacterial Communities

A review of some published genomic research into WWTP microbial communities is presented below. Where relevant, papers are classified by the process on which they focused. Genomic methods are mentioned briefly for each study. Particularly salient results are highlighted in bold text.

A number of studies focused on community composition as determined through probing and sequencing of 16S rRNA genes or of genes specific to limited clades. Others quantified gene expression through mRNA measurement and compared expression to overall gene abundance or to observed kinetic or mass-balance results.

Glossary of Relevant Terms:

- 16S rRNA: The gene coding for 16S ribosomal RNA (16S rRNA) is widely used as a biomarker for studying microbial ecology. It is short and highly conserved across bacterial species, facilitating sequencing, but contains variable regions that aid in phylogenetic classification [5].
- PCR: Polymerase chain reaction. A molecular technique used to amplify a few copies of DNA over several orders of magnitude.
- qPCR: Quantitative or real-time PCR. Used to amplify and simultaneously quantify a targeted DNA sequence. Differs from classic PCR in that products are detected as they are generated.
- RT-PCR: Reverse-transcriptase PCR. Used to qualitatively study gene expression using complementary DNA transcribed from RNA.
- q-RT-PCR: Quantitative real-time reverse transcription PCR is a combined technique that uses RT-PCR and can be applied to measure RNA and used to study gene expression. When normalized against DNA abundance, this technique can indicate how much a gene is being expressed.
- DGGE: Denaturing gradient gel electrophoresis. A molecular fingerprinting method used to analyze the microbial diversity of a system based on 16S rDNA fragments without culturing or isolation. Fingerprint similarity can be analyzed to determine microbial structure differences between environments or treatments.
- FISH: Fluorescent In Situ Hybridization. A technique that utilizes staining or fluorescent nucleic acid probes to allow phylogenetic identification of microbial cells or genes within a mixed consortium. Can be used to visualize particular microbial groups and quantify them by fluorescence intensity. PCR-DGGE and FISH have been used in combination to identify and quantify dominant microorganisms in a mixture [3].
- SIP: Stable Isotope Probing. A technique used to identify certain functional groups of organisms capable of incorporating stable-isotope labeled substrates such as ^{13}C or ^{15}N.

Literature Review

Nitrification

Chandran and Love (2008) used q-RT-PCR to quantify expression of *amoA* in *Nitrosomonas europaea*, normalizing results against 16S rRNA gene abundance. FISH was performed for gene quantification. The researchers found that relative expression of *amoA* could be a viable alternative to specific oxygen uptake rate (SOUR) tests for measuring ammonia oxidation activity attributed to specific communities in uninhibited and inhibited cultures. **16S rRNA abundance was not found to correlate with physiological activity. rRNA-based FISH did not quantitatively track SOUR and *amoA* trends, and thus may not be a conclusive descriptor of *N. europaea*-based specific activity.** The study concluded that SOURs are useful in pure cultures where only one species is responsible for oxygen uptake, but **in a mixed community, studying gene expression (e.g., expression of *amoA*, normalized against gene abundance) can give an indication of how well AOBs are doing** [6].

Denitrification

Lu and Chandran (2010) used qPCR to quantify glycerol-assimilating denitrifying bacteria in mixed microbial communities in both biofilm and suspended phases of an integrated fixed-film activated sludge (IFAS) reactor. DNA SIP was used to identify carbon sources used by the bacteria. Populations of glycerol-assimilating bacteria were quantified using qPCR based on the 16S rRNA gene fraction, and

reactor performance was determined concurrently using influent nitrate and effluent nitrite, nitrate, and chemical oxygen demand (COD). Bacteria in the glycerol-fed reactor were found to be different from populations identified in a methanol-fed sequencing batch reactor (SBR) [7]. In a 2011 paper, Lu et al. used 16S rRNA concentrations as references for studying the frequency of gene expression via measurements of mRNA. Researchers found that **genes coding for methanol dehydrogenase and glycerol dehydrogenase could be used as potential biomarkers of denitrification activity in mixed-culture wastewater treatment processes** [8]**.** Similarly, Luton et al. (2002) found that the methyl coenzyme-M reductase gene (*mrcA*) was a reliable alternative to 16S rRNA for detecting methanogens [9].

Baytshtok et al. (2008) used SIP and 16S rRNA-based phylogenetic interpretation of ^{13}C labeled DNA to identify specific bacteria metabolizing methanol in a denitrifying SBR [10]. Once species were identified, qPCR primer sets were developed to determine species abundance over different modes of reactor operation. **The researchers monitored both specific denitrification rates (sDNR) and population abundance, which allowed them to see that the reduction in sDNR was mainly due to reduction in cell activity** (rather than reduction in population abundance)**.**

In a 2009 study, Baytshtok et al. identified bacterial populations of methanol- and ethanol-metabolizing bacteria using SIP of ^{13}C-labeled DNA. Populations were quantified using qPCR. When the electron donor was switched from methanol to ethanol, concentrations of *Hyphomicrobium* species decreased significantly, whereas *Methyloversatilis* species were unchanged (the former cannot metabolize ethanol while the latter can). **Researchers observed that denitrification decreased significantly as bacterial concentrations decreased.**

> *"From an engineered wastewater treatment standpoint, it is equally important to understand both the identities and metabolic capabilities of dominant methylotrophic denitrifying populations in activated sludge. Such information is critical, for instance to evaluate the feasibility of switching to a higher rate carbon source such as ethanol in lieu of methanol, which could be favored during lower winter temperatures at wastewater treatment plants, or due to the increasing price of methanol." (Baytshtok et al., 2009 [11])*

Partial Nitrification/Anammox

Ahn et al. (2008) used direct measures of bacterial concentrations (from 16S rRNA) in conjunction with activity measures and mass balances [2]. Researchers found that partial nitrification operating conditions selected for a narrow diversity of rapidly growing AOB and NOB populations (unlike conventional biological nitrogen removal reactors, which host a broader diversity of AOBs and NOBs). In a 2011 study, Ahn et al. sequenced 16S rRNA and *amoA* genes and found that **AOB populations were different during full- and partial-nitrification modes**. DGGE and sequencing was performed for *amoA.* Q-RT-PCR was used to study gene expression in both full- and partial-nitrification modes. Transition to partial nitrification resulted in spikes of N_2O and NO gas, which correlated with gene expression (*nirK* and *norB*) [12]. Park et al. (WR, 2010) also used DGGE to fingerprint anammox communities in both granular anammox and biofilm-based bioreactors, and studied AOB diversity using *amoA* [13].

Park et al. (ES&T, 2010) conducted batch inhibition assays of anammox bacteria and collected samples for gene expression measurement. Biomarkers of activity included hydrazine oxidoreductase (*hzo*) mRNA and the 16S-23S rRNA intergenic spacer region (ISR). Decreases in anammox-specific *hzo* mRNA and ISR concentrations were observed in both inhibited batches, indicating that **gene expression (as**

determined via mRNA quantification) tracked and preceded nitrogen-removal trends (as determined from batch inhibition results). Total bacterial abundance was quantified by using 16S rRNA primers. Specific fractions of AOB, NOB, and anammox were also calculated at different times. qPCR targeted anammox 16S rRNA, *amoA*, and *Nitrobacter* and *Nitrospira* 16S rRNA. Relative abundance of AOB, NOB, and anammox were calculated by assuming 2 *amoA* gene copies per AOB, 1 rRNA operon per anammox and NOB, and 4.13 rRNA operons per eubacterial cell [14].

> *"From a process monitoring and control perspective, measures of biological activity such as specific substrate consumption rates are more sensitive and rapid indicators of process upsets and recovery compared to measurement of reactant or product compounds, especially in systems operated at high solids retention times (SRTs), such as anammox reactors. However, in complex microbial communities competing for similar substrates [e.g., anammox and ammonia-oxidizing bacteria (AOB) competing for ammonia], it may not be trivial to infer microbial activities from commonly employed batch substrate (ammonia or nitrite) consumption assays. In contrast, biomarkers that target anammox metabolic processes might be more applicable to interrogate the specific activity thereof." (Park et al., 2010 [14])*

Phosphorus Removal

He et al. (2007) studied the population structure of EBPR systems using the polyphosphate kinase 1 gene (*ppk1*) as a genetic marker. 16S rRNA genes were used to reconstruct phylogenies. The relative distribution of different bacterial clades was determined using qPCR. **Relative distributions of clades were found to vary among different EBPR systems, as well as temporally within a system** [4]. Kim et al. (2013) studied communities of "*Candidatus* Accumulibacter phosphatis" clades in an SBR as it was transitioned from anaerobic-oxic (AO) to anaerobic-anoxic-oxic (A2O) conditions. Communities were compared using 16S rRNA and polyphosphate kinase genes (*ppk1*) and using FISH with specific oligonucleotide FISH probes designed to target a "*Ca.* Accumulibacter" clade. The transition of SBR operating mode was found to lead to a shift in abundances of some subpopulations, as well as an increase in abundance of other associated bacteria [15]. Zou et al. (2014) combined chemical and microbial analysis to study P release and uptake in anaerobic and anoxic conditions. Microorganisms involved in EBPR were identified via PCR-DGGE analysis, and quantitative FISH analysis was used to identify the dominant genus present in activated sludge samples from an anoxic tank [3]. Wilmes et al. (2008) employed proteomics to identify proteins in *Candidatus Accumulibacter phosphatis* populations under both anaerobic and aerobic conditions. **Protein profiles were similar in both conditions, but significant differences were seen in the abundance of different enzyme variants, suggesting that genetic diversity is essential for maintaining stable EBPR performance** [16].

Antibiotic Resistance

Kim et al. (2010) performed PCR on agar-plated WWTP bacteria to confirm presence of tetracycline resistance genes, and used qPCR to quantify samples. Tetracycline resistant bacteria were quantified before and after UV disinfection or chlorination. **Significant day-to-day variability was seen, and the average before/after fractions did not vary significantly** [17].

Nutrient Cycling

Sun et al. (2014) used GeoChip, a high-throughput microarray-based tool, to identify and quantify bacterial genes in four large-scale membrane bioreactors (MBRs). While significant variation was found between genes in the different MBRs, **results showed significant correlations between the key**

functional gene categories in reactors and corresponding influent constituents. Abundance of carbon degradation genes was correlated with influent COD, nitrogen cycling genes correlated with influent ammonium, phosphorus cycling genes correlated with influent phosphorus, and sulfur cycling genes correlated with influent sulfate [18]. Wang et al. (2014) used GeoChip to analyze activated sludge from four separate WWTPs and found significant correlations between microbial functional genes and water temperature, dissolved oxygen (DO), ammonia concentrations, and influent chemical oxygen demand [19].

Challenges for Genomic Analysis

DNA vs. mRNA Quantification

Both gene abundance (through DNA quantification) and gene transcription (through mRNA quantification) have been used to evaluate the composition and activity of mixed microbial communities. Measurements of microbial community DNA are generally accepted as measuring the metabolic potential of a community, but are less reliable than measurements of mRNA in determining functional activity [18]. While some studies have been able to link DNA abundance with environmental factors and observed activity [18, 15], others have found no significant correlations [17].

While DNA is a proxy for activity, mRNA is typically regarded as corresponding to activity because RNA transcription is a precursor of protein formation [14]. However, mRNA detection requires strict storage and rapid turnover, and is subject to interference from rRNA and tRNA [18]. This can render mRNA quantification difficult and impractical. Additionally, mRNA profiles do not always reflect activity; Yu et al. (2010) found that whole-cell responses of *Nitrosomonas europaea* to DO limitation or nitrite toxicity did not necessarily parallel transcription of the corresponding mRNA [20, 21]. Hence, mRNA-activity relationships may need to be analyzed on a case-by-case basis, taking into account known specific transcriptional responses to inhibitors or other parameters.

Variation in Microbial Diversity

Activated sludge samples vary greatly due to differences in organic loading, pH, temperature, DO, and SRT in the aeration tank. Populations have also been observed to vary with time. Long-term studies on WWTP metagenomics are lacking, and it is unknown whether community composition may be reliably compared across seasons or years.

Summary of Techniques and Associated Applications

Table 1 presents a list of some of the most commonly employed techniques used to evaluate WWTP metagenomics, along with the applications, advantages, and disadvantages of each.

Table 1. Standard Genomic Techniques and Associated Applications

Technology/Process	Application	Advantages	Disadvantages
16S rRNA gene probing, sequencing, and quantification	Probe for a wide range of bacterial genes. Follow with PCR to amplify gene sequences. Amplified sequences can be compared to libraries to identify known species.	Highly conserved sequence allows for use of universal primers [5]. Short gene length is easy to sequence. Variable regions help with classification.	Targets a wide range of bacteria; narrowing down species of interest may be time-intensive. Best used to identify species, not for quantification (concentrations of 16S rRNA genes do not necessarily correlate with concentrations of bacteria, and some genera may be significantly over- or under-estimated using 16S rRNA gene frequency [6, 22]).
qPCR (real-time PCR)	Amplify, detect, and quantify DNA in real-time. q-RT-PCR can quantify RNA.	Quick. Individual runs are inexpensive. Can target specific genes with known sequences. Can be used to quantify genes (DNA) or gene expression (mRNA). Depending on the machine, up to 7 separate targets may be simultaneously quantified during a single qPCR reaction.	Genes of interest (and their associated primers) must be identified via probing and sequencing or microarray analysis before qPCR can be performed.
GeoChip	Microarray used to detect genes through binding to specific probes included in the array.	Each microarray contains thousands of functional genes with applications for WWTP processes including C, N, P, and S cycling, antibiotic and heavy metal resistance, and stress responses [23]. Genes can be examined simultaneously without the need to design specific primers. Enables fast identification of multiple genes of interest. Not limited to one or several species or gene sequences. Array may be customized to focus on functions/processes of interest.	Only detects specific genes found in the microarray. Expensive for each test. Requires a high level of purity for DNA used in hybridization. Amplification, post-amplification, labeling, and later steps require approximately 2 days of work, regardless of the number of arrays being scanned.
FISH	Enables identification of individual microbial cells or genes in their environment.	Allows direct detection of specific genes within microorganisms. Can be used to follow spatial and temporal trends of microbial populations.	May not be a reliable indicator of activity [6]. Probes must be prepared for particular species.
DGGE	Molecular fingerprinting method that can be used to determine microbial structural differences between environments or treatments.	Relatively low cost way to compare microbial communities.	Can be difficult to obtain reproducible results. Applications may not be relevant to process questions. Prone to upstream PCR bias.

Overview of Procedures and Materials

If the District chooses to pursue genomic techniques, a number of avenues are available. The following presents an overview of common steps involved in genomic analysis, with a focus on techniques, materials, and equipment that have been successfully used by researchers working on mixed-community microbial samples from WWTP reactors.

Sample Preservation

The highest DNA yield and quality is achieved by purifying DNA from freshly harvested cells [24]. However, direct harvesting may not be practical, especially if the aim is to store samples in case they are needed for future analysis. Research on sample storage suggests that short-term storage conditions have little effect on relative abundances of microbial taxa, even when samples are kept at room temperature for 14 days [25]. Nevertheless, when DNA is not to be immediately extracted, samples should be stored under a standard set of conditions that preserve DNA integrity. Harvested cells should be pelleted by centrifugation and the supernatant removed. Freezing at -20°C or -80°C is recommended for long-term storage on the order of months or years [24]. Repeated freezing and thawing should be avoided. Kim et al. (2010) concentrated samples by centrifugation (7,000 rpm for 5 min) or filtration through 1.2 µm GF/C glass fiber filters prior to storage at -80° C [17].

DNA Extraction and Purification

Complete lysis of cell walls and membranes is required for extraction and isolation of genomic DNA; incomplete disruption will result in significantly decreased yields. Many bacterial cells can be effectively lysed using a lysis buffer and protease or proteinase K, but some bacteria also require pre-incubation with specific enzymes [24]. A number of kits are available that are specifically tailored for broad-spectrum microbial DNA extraction. Other kits have also been successfully used in studies involving wastewater microbes. Two kits used in multiple studies of wastewater bacterial communities are:

- DNeasy Blood & Tissue kit (Qiagen, Inc.) [12, 10, 2, 8, 11]
- FastDNA® SPIN Kit for Soil (MP Biomedicals, Solon, OH) [22, 26, 15, 3] (This kit was compared by Zhang et al. to 4 other kits and found to have the lowest contamination when used on samples of activated sludge [26].)

Microarrays such as GeoChip require that large fragments of genomic DNA be preserved. If microarray analysis is performed, the extraction method should be chosen accordingly and extracted DNA should be purified as soon as possible after extraction to prevent degradation [27]. Gel purification may be used. Commercial kits are also available, including the Wizard® SV Gel and PCR Clean-up System (Promega Corporation, Madison, WI) [18].

It should be noted that studies have shown differences in inferred community composition resulting from different extraction methods [28]. A consistent extraction method should therefore be maintained in order to minimize variation across samples.

DNA Quality Analysis and Integrity Verification

The concentration of extracted DNA (prior to amplification) may be determined by measuring the absorbance at 260 nm (A_{260}) in a spectrophotometer. DNA quality may also be assessed via spectrometry; pure DNA has an A_{260}/A_{280} ratio of 1.8-2.0 in 10 mM Tris buffer at pH 8.5. A low A_{260}/A_{280} ratio indicates the presence of contaminants such as proteins, while a strong absorbance at other wavelengths may indicate other specific contaminants such as phenols or solvents [24].

Fluorometry with dyes such as Hoechst 33258 or PicoGreen (Invitrogen) [4] enables accurate quantification of smaller quantities of DNA while minimizing interference from RNA [24]. Agarose gel electrophoresis may be used to quantify small amounts of DNA but is generally more appropriate for quantification of purified PCR products [24, 4].

Storage of Purified DNA

Extracted DNA may be amplified via PCR/qPCR immediately after extraction. If it is to be stored prior to amplification or other processing, however, purified DNA should be kept at -20°C or -70°C under slightly basic conditions to avoid hydrolysis [24]. Common buffers for DNA storage include Tris·Cl and TE buffer. Oxford Gene Technology recommends medium-term storage (on the order of months) at -80°C in TE buffer; for multiple years, -80°C as a precipitate under ethanol is recommended [29].

DNA Hybridization and PCR/qPCR

Primers for target genes can be selected with the aid of public databases and open-source tools.

- The PubMed Gene database (www.ncbi.nlm.nih.gov/gene) allows researchers to select appropriate primers for a gene of interest.
- NCBI/Primer-BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) is a tool for finding specific primers for target gene sequences, and for screening selected primers for potential problems [10, 13, 7, 8, 11].

Standard PCR machines (PCR thermocyclers) amplify DNA by regulating temperatures during cyclical programs. qPCR thermocyclers combine nucleic acid amplification and detection, and can be used with both DNA and RNA (in a process known as q-RT-PCR). Applications include genotyping and quantitative gene expression analysis. Most qPCR thermocyclers have an average 2-hour run time and a "fast" 40-minute run time [30]. Examples of commonly used qPCR thermocyclers include:

- iQ5 real-time PCR thermal cycler (BioRad, Hercules, CA) [13, 8, 4, 6, 11]
- LightCycler® 96 Instrument (Roche Life Science)
- Mx3005P QPCR System (Agilent Technologies)

Purification of PCR Products

PCR products may be purified using a number of commercially available kits. One used successfully in a study of WWTP bacteria is the PCRquick-spin™ PCR Product Purification Kit (iNtRON Biotechnology, South Korea) [22].

DNA Sequencing and Identification

Kits for sequencing of unknown genes include the TAQuence cycle sequencing kit (USB) used by Luton et al. [9]. Once sequences are established, multiple open-source tools can be used to identify genes. These include the following:

- GenBank (http://www.ncbi.nlm.nih.gov) [7, 8, 9] contains sequence data from a large number of microbial studies.
- dbSNP (www.ncbi.nlm.nih.gov/SNP) can be used to search small variations in gene sequences.
- The National Center for Biotechnology Information Gene Expression Omnibus (NCBI GEO, www.ncbi.nlm.nih.gov/geo/) is a public repository that stores curated gene expression DataSets, which also contain cluster tools and differential expression queries.

- ClustalW (http://www.clustal.org/clustal2/) or other NCBI online tools can be used to align related transcripts to understand exon overlap.

DNA Quantification

Extracted and purified DNA can generally be quantified using spectrophotometric or UV fluorescence.

- Spectrophotometer instruments such as the Nanodrop Lite or Nanodrop 2000 allow for quick measurement of DNA or RNA concentrations while also calculating sample purity ratios.
- The Qubit 3.0 fluorometer uses specific dyes to measure the concentrations of DNA and RNA.

Equipment for Microarray Hybridization

- GeoChip is used for DNA hybridization, targeting thousands of microbial genes involved in nutrient cycling and other processes.
- Hybridization can be performed using a HS 4800™ Pro or HS 400™ Pro Microarray Hybridization Station (Tecan).

Equipment for Microarray Scanning and Data Processing

- NimbleGen MS 200 Microarray Scanner (Roche NimbleGen) and NimbleScan software [18]
- ScanArray 5000 analysis system (Perkin-Elmer)
- ImaGene 6.0 (Biodiscovery, Inc.)

Summary

Genomic analysis shows promise for identifying the key species in biological treatment processes and for monitoring both species concentrations and metabolic activity. If gene abundance or gene expression results are shown to be reliable, genomic techniques may provide a complementary means of monitoring process performance and stability. Genomic data could also provide insights that traditional analytical chemistry cannot. For example, high gene abundance associated with low gene transcription could indicate the presence of enzymatic inhibitors. Because community composition changes relatively slowly, sampling intervals would be less frequent than sampling for most procedures, and tests could be scheduled on the order of the SRT of each process basin. Changes in genetic composition could provide an early indication of process instability before effects are seen in kinetic analysis.

Studies have shown high microbial diversity in WWTP processes. Thus, technologies that only identify species (without information on their relative abundance) may tell us little about the activity of the community of microbes. Population abundance or abundance of specific functional genes (as determined through qPCR/FISH or microarray testing, for example) is much more quantitative, and this type of data has been linked to influent characteristics [19]. Researchers have also observed decreases in activity paralleling decreases in population [11]. However, changes in activity have also been observed without significant changes in population abundance [10]. This complicates the use of population abundance as a proxy for activity, at least in the short term.

RNA and DNA concentrations cannot be tied directly to activity across the board. While correlations between gene frequency and function do exist, there is significant variation resulting from reactor configuration, influent concentration, and ambient conditions (such as pH and temperature). As such,

comparisons are likely only valid within a single reactor or multiple well-characterized reactors operated under similar conditions, and only when a long history of data collection has been established.

A number of challenges must be addressed in attempting genomic techniques for WWTP process analysis. Knowledge of which bacteria are present may be important before it is possible to probe for specific genes. Sequencing bacterial genes may be time intensive and require analysis of numerous genes. Finally, care must be taken when drawing conclusions about activity based on gene frequency, population frequency, or mRNA frequency.

Recommended Technologies

Of the technologies currently in use for metagenomics studies, the two that appear most promising for monitoring treatment plant processes are qPCR and microarray analysis. A combined qPCR/microarray approach could enable identification and quantification of numerous genes of interest in biological wastewater treatment.

qPCR provides a cost-effective tool for amplifying known gene sequences. This technology could be used to quantify genes and determine the community composition of nitrifiers, denitrifiers, and PAOs. Once genes of interest in a mixed microbial community are identified, qPCR could be used to monitor gene abundance over time and compare results week-to-week or from season to season. Gene abundance could be compared to environmental factors or operational changes to establish correlations and determine the success of interventions. q-RT-PCR could also be performed to quantify gene transcription and determine to what degree genes are expressed.

In order to perform qPCR, however, it is necessary to target a particular gene of interest. Some genes used in previous research with WWTP bacteria may be chosen. However, results of several studies indicate that WWTP bacterial communities vary between reactors of different configurations or influent characteristics, and choosing what to target without knowledge of the variety of microorganisms present in a particular process reactor may cause important genes or species to be overlooked. It is thus important to identify genes that are present in a process before choosing which DNA sequences to pursue quantitatively.

While 16S rRNA probing can be used to identify a wide range of bacteria, the sheer number of species in a mixed microbial community, such as a WWTP reactor, would make the technique time-consuming. In contrast, microarrays allow for simultaneous detection and quantification of a wide variety of genes, including those with relevance to WWTP processes. Each microarray contains thousands of functional genes (such as C, N, and P cycling genes). A distinct advantage of microarrays is that genes can be examined and quantified simultaneously without the need to design specific primers or sequence cloned DNA strands. This enables rapid identification of multiple genes of interest. Additionally, microarrays may be customized to focus on functions/processes of interest if desired [23]. The most recent GeoChip microarray contains 160,000 already-identified genes, and has been shown capable of detecting differences in DNA concentrations in WWTP reactors corresponding to varying influent parameters [18].

Each GeoChip is expensive ($750 per sample quoted on researchgate.net), rendering exclusive use of this technique impractical if regular monitoring is desired. However, GeoChip could be used as a broad-spectrum screening tool to identify the highest-frequency functional genes, which could then be targeted and quantified via qPCR for future samples. qPCR could also be used to run replicates on specific genes for statistical purposes, an application that would be cost-prohibitive with microarrays.

Table 2 summarizes the potential applications of these two techniques with observations on the logistics of implementation.

Table 2. Recommended Technologies

Technology/Process	Logistics of Implementation	Potential Use
qPCR	• Requires storage freezer, thermocycler, microcentrifuge, spectrophotometer, vortex, and hot water bath; DNA extraction kits and primers need to be purchased according to the number of samples • Turnaround: one day • The target gene or mRNA sequence must be known before qPCR can be run (primers are sequence-specific) • Up to 7 preselected targets may be amplified during a single run • After purchase of equipment (capital costs), individual tests are relatively inexpensive • Because individual tests are low-cost, replicates may be run for statistical purposes	• Quantify specific genes to determine concentrations of particular species in nitrification, denitrification, or phosphorus removal processes • Compare species concentrations to previous measurements and establish trends • Quantify expression of specific genes to determine the proportion of active cells • Findings may be studied alongside environmental factors and analytical chemistry results to identify correlations • Changes in population structure or activity level may be used to evaluate the success of interventions • Abnormal findings may indicate upsets in a process
Microarray (GeoChip)	• Requires storage freezer, vortex, hybridization station, microarray scanner; DNA extraction materials and individual microarrays must be purchased for each sample • Turnaround: two days • Each test is expensive (several hundred dollars), so it would be cost-prohibitive to run replicates • DNA extraction is more complicated than DNA extraction for qPCR • Does not require foreknowledge of genes or species present in a sample • Can be used to identify key genes for future analysis via qPCR	• Identify and quantify large numbers of functional genes present in a single sample from a process reactor (e.g., nitrification, denitrification, or phosphorus removal) • Narrow down which genes or species are most common in a sample • Enable laboratory staff to choose which genes to monitor in future samples

Equipment and Estimated Costs

Table 3 presents estimated costs of equipment for performing in-house genomic analysis via qPCR and microarray. All prices are for new equipment.

Table 3. Total Cost Estimate for Genomic Analysis Startup

DNA Extraction Equipment		**Cost**
	-80°C Freezer	$14,000
	-20°C Freezer	$7,300
	Micropipettors	$1,280
	Microcentrifuge	$2,900
	Vortex	$385
	Hot Water Bath	$395
DNA Quantification		
	Nanodrop Spectrophotometer	$5,590
	Total for Extraction Equipment	**$31,850**
qPCR		
	qPCR Thermocycler	$28,500
	Laminar Flow Hood	$5,850
	Total for qPCR	**$34,350**
Microarray		
	Microarray Scanner	$44,000
	Total for Microarray	**$44,000**
	Combined Total	**$110,200**
Reagent/Supply Cost per Sample		
	DNA Extraction	$7
	qPCR	$6
	Microarray	$200-800

Method Development

Key aspects of genomic analysis would include the following characteristics:

- Collect samples in different seasons to analyze and compare temporal microbial community dynamics.
- Establish a library of samples over the course of one year to ascertain baseline community structure and relative abundance.
- Standardize techniques for all samples (DNA extraction, etc.) to eliminate variance and error.
- Collect ancillary data to support analysis (e.g., pH, temperature, DOC, N, etc.).
- Split samples and freeze aliquots for later analysis and comparison to fresh samples.
- Determine sampling frequency based on wastewater processes. For example, samples from biological phosphorus removal may need to be taken every week, whereas samples taken from digesters will only need to be taken every 15-30 days.

Quality Assurance/Quality Control

Quality control samples would be included with each batch to assess the validity of the sample analysis. These would include duplicate samples, positive controls, negative controls, and matrix spikes.

Proposed Actions

Community metagenomics is a relatively new field, and the technology is rapidly developing and becoming more accessible. As an alternative to investing in equipment at this stage, it may be advisable to revisit this topic in the future when further research is available and materials such as microarrays are likely to be less expensive and simpler to process.

If Clean Water Services decides to pursue genomic techniques, external laboratories such as Oregon State University's Center for Genome Research and Biocomputing could be utilized initially to perform certain components of the analyses. This facility offers services such as DNA extraction, quantification and real time PCR analysis. Results could be compared and analyzed to determine optimal methods before acquiring new equipment.

Another option would be to take advantage of metagenomics tests geared toward wastewater analysis. BluCarbon Analytics, a New York-based startup, is working to develop microbial diagnostics platforms for water and wastewater treatment. If these tests become publicly available, sending samples out for processing would likely prove to be the most efficient and cost-effective means of acquiring reliable data about WWTP microbial populations. This approach would not provide immediate results, but could be used to monitor long-term trends and evaluate the utility of genomic data in comparison to traditional laboratory tests.

Metagenomics offers the potential to shine a light into the black box of biological wastewater treatment, illuminating opportunities for improvements in reactor design and operation. As the body of research grows and technologies become more accessible, genomic analytical techniques may prove to be increasingly useful in helping to understand and optimize treatment processes.

References

[1] R. J. Siezen and M. Galardini, "Genomics of biological wastewater treatment," *Microbial Biotechnology,* vol. 1, no. 5, pp. 333-340, 2008.

[2] J. H. Ahn, R. Yu and K. Chandran, "Distinctive Microbial Ecology and Biokinetics of Autotrophic Ammonia and Nitrite Oxidation in a Partial Nitrification Bioreactor," *Biotechnology and Bioengineering,* vol. 100, no. 6, pp. 1078-1087, 2008.

[3] H. Zou, X. Lu and S. Abualhail, "Characterization of Denitrifying Phosphorous Removal Microorganisms in Two-Sludge Process by Combining Chemical with Microbial Analysis," *Journal of Chemistry,* vol. 2014, 2014.

[4] S. He, D. L. Gall and K. D. McMahon, "'Candidatus Accumulibacter' Population Structure in Enhanced Biological Phosphorus Removal Sludges as Revealed by Polyphosphate Kinase Genes," *Applied and Environmental Microbiology,* vol. 73, no. 18, pp. 5865-5874, September 2007.

[5] Illumina, Inc., "16S Metagenomics Studies with the MiSeq(R) System," 2014. [Online]. Available: http://www.illumina.com/content/dam/illumina-marketing/documents/products/appnotes/appnote_16s_sequencing.pdf.

[6] K. Chandran and N. G. Love, "Physiological State, Growth Mode, and Oxidative Stress Play a Role in Cd(II)-Mediated Inhibition of Nitrosomonas europaea 19718," *Applied and Environmental Microbiology,* vol. 74, no. 8, pp. 2447-2453, 2008.

[7] H. Lu and K. Chandran, "Diagnosis and Quantification of Glycerol Assimilating Denitrifying Bacteria in an Integrated Fixed-Film Activated Sludge Reactor via 13C DNA Stable-Isotope Probing," *Environmental Science & Technology,* vol. 44, no. 23, pp. 8943-8949, 2010.

[8] H. Lu, F. Nuruzzaman, J. Ravindhar and K. Chandran, "Alcohol dehydrogenase expression as a biomarker of denitrification activity in activated sludge using methanol and glycerol as electron donors," *Environmental Microbiology,* vol. 13, no. 11, pp. 2930-2938, 2011.

[9] P. E. Luton, J. M. Wayne, R. J. Sharp and P. W. Riley, "The mcrA gene as an alternative to 16S rRNA in the phylogenetic analysis of methanogen populations in landfill," *Microbiology,* pp. 3521-3530, 2002.

[10] V. Baytshtok, S. Kim, R. Yu, H. Park and K. Chandran, "Molecular and biokinetic characterization of methylotrophic denitrification using nitrate and nitrite as terminal electron acceptors," *Water Science & Technology,* vol. 58, no. 2, pp. 359-365, 2008.

[11] V. Baytshtok, H. Lu, H. Park, S. Kim, R. Yu and K. Chandran, "Impact of Varying Electron Donors on the Molecular Microbial Ecology and Biokinetics of Methylotrohpic Denitrifying Bacteria," *Biotechnology and Bioengineering,* vol. 102, no. 6, pp. 1527-1536, 2009.

[12] J. H. Ahn, T. Kwan and K. Chandran, "Comparison of Partial and Full Nitrification Processes Applied for Treating High-Strength Nitrogen Wastewaters: Microbial Ecology through Nitrous

Oxide Production," *Environmental Science & Technology,* vol. 45, pp. 2734-2740, 2011.

[13] H. Park, A. Rosenthal, R. Jezek, K. Ramalingam, J. Fillos and K. Chandran, "Impact of inocula and growth mode on the molecular microbial ecology of anaerobic ammonia oxidation (anammox) bioreactor communities," *Water Research,* vol. 44, pp. 5005-5013, 2010.

[14] H. Park, A. Rosenthal, K. Ramalingam, J. Fillos and K. Chandran, "Linking Community Profiles, Gene Expression and N-Removal in Anammox Bioreactors Treating Municipal Anaerobic Digestion Reject Water," *Environmental Science & Technology,* vol. 44, no. 16, pp. 6110-6116, 2010.

[15] J. M. Kim, H. J. Lee, D. S. Lee and C. O. Jeon, "Characterization of the Denitrification-Associated Phosphorus Uptake Properties of "Candidatus Accumulibacter phosphatis" Clades in Sludge Subjected to Enhanced Biological Phosphorous Removal," *Applied and Environmental Microbiology,* vol. 79, no. 6, pp. 1969-1979, 2013.

[16] P. Wilmes, A. F. Andersson, M. G. Lefsrud, M. Wexler, M. Shah, B. Zhang, R. L. Hettich, P. L. Bond, N. C. VerBerkmoes and J. F. Banfield, "Community proteogenomics highlights microbial strain-variant protein expression within activated sludge performing enhanced biological phosphorus removal," *The ISME Journal,* no. 2, pp. 853-864, 2008.

[17] S. Kim, H. Park and K. Chandran, "Propensity of activated sludge to amplify or attenuate tetracycline resistance genes and tetracycline resistant bacteria: A mathematical modeling approach," *Chemosphere,* vol. 78, pp. 1071-1077, 2010.

[18] Y. Sun, Y.-x. Shen, P. Liang, J. Zhou, Y. Yang and X. Huang, "Linkages between microbial functional potential and wastewater constituents in large-scale membrane bioreactors for municipal wastewater treatment," *Water Research,* vol. 56, pp. 162-171, 2014.

[19] X. Wang, Y. Xia, X. Wen, Y. Yang and J. Zhou, "Microbial Community Functional Structures in Wastewater Treatment Plants as Characterized by GeoChip," *PLOS One,* vol. 9, no. 3, March 2014.

[20] R. Yu and K. Chandran, "Strategies of Nitrosomonas europaea 19719 to counter low dissolved oxygen and high nitrite concentrations," *BMC Microbiology,* vol. 10, no. 70, 2010.

[21] R. Yu, M. J. Kampschreur, M. C. M. Van Loosdrecht and K. Chandran, "Mechanisms and Specific Directionality of Autotrophic Nitrous Oxide and Nitric Oxide Generation during Transient Anoxia," *Environmental Science & Technology,* vol. 44, no. 4, pp. 1313-1319, 2010.

[22] L. Ye and T. Zhang, "Bacterial communities in different sections of a municipal wastewater treatment plant revealed by 16S rDNA 454 pyrosequencing," *Environmental Biotechnology,* vol. 97, pp. 2681-2690, 2013.

[23] Glomics, Inc., "GeoChip Technology," [Online]. Available: http://www.glomics.com/gch-tech.html. [Accessed 23 April 2015].

[24] Qiagen, "DNA Protocols & Applications: Considerations for isolation and quantification of both genomic DNA and plasmid DNA," QIAGEN, 2015. [Online]. Available: https://www.qiagen.com/us/resources/molecular-biology-methods/dna/. [Accessed 27 April 2015].

[25] C. L. Lauber, N. Zhou, J. I. Gordon, R. Knight and N. Fierer, "Effect of storage conditions on the assessment of bacterial community structure in soil and human-associated samples," *FEMS Microbiology Letters,* vol. 307, no. 1, pp. 80-86, 2010.

[26] T. Zhang, M.-F. Shao and L. Ye, "454 Pyrosequencing reveals bacterial diversity of activated sludge from 14 sewage treatment plants," *The ISME Journal,* vol. 6, pp. 1137-1147, 2012.

[27] J. Van Nostrand, L. Wu, Z. He and J. Zhou, "GeoChips for Analysis of Microbial Functional Communities," in *Handbook of Hydrocarbon and Lipid Microbiology*, Heidelberg, Germany, Springer-Verlag Berlin, 2010, pp. 4040-4050.

[28] M. Bahl, A. Bergstrom and L. T.R., "Freezing fecal samples prior to DNA extraction affects the Firmicutes to Bacteroidetes ratio determined by downstream quantitative PCR analysis," *FEMS Microbiology Letters,* vol. 329, no. 2, pp. 193-197, 2012.

[29] Oxford Gene Technology, "DNA Storage and Quality," Oxford Gene Technology, 22 August 2011. [Online]. Available: http://www.ogt.co.uk/resources/literature/403_dna_storage_and_quality. [Accessed 27 April 2015].

[30] Biocompare, Inc., "Real-Time Thermal Cyclers," CompareNetworks, Inc., [Online]. Available: http://www.biocompare.com/PCR-Real-Time-PCR/22353-Real-Time-PCR-Thermal-Cyclers-Thermocyclers/. [Accessed 27 April 2015].