DOI: 10.1002/rem.21695

RESEARCH ARTICLE

WILEY

Lessons learned from 20 years of molecular biological tools in petroleum hydrocarbon remediation

Dora M. Taggart | **Kate Clark**

Microbial Insights, Inc., Knoxville, Tennessee, USA

Correspondence
Dora M. Taggart, Microbial Insights, Inc.
10515 Research Dr, Knoxville,
TN 37932, USA.
Email: DTaggart@microbe.com

Abstract

Stakeholders of petroleum hydrocarbon-contaminated (PHC) sites aim to meet remediation goals cost-effectively. But contaminated sites are typically complex, and when inefficient remediation strategies are implemented, time and money are wasted. Various strategies can be implemented for site remediation and almost all at some point involve a biological component. Therefore, during site characterization, seeing the complete picture by obtaining multiple lines of evidence—chemistry (concentrations of contaminants and daughter products), geochemistry (redox status, electron acceptors, electron donors), microbiology (species and their genes for catalyzing biodegradation of reactions, i.e., functional genes), and the contaminant degradation ongoing at baseline (stable isotope methods)—is critically important for selecting the best remediation strategy. Additionally, molecular biological evidence is important for monitoring treatment progress and informing decisions to retreat, change treatments, or transition to monitored natural attenuation.

KEYWORDS

hydrocarbon biodegradation, next generation sequencing, quantitative polymerase chain reaction, stable isotopes

1 | INTRODUCTION

1.1 | Increase in the use of molecular biological tools (MBTs) in the past 20 years

Over the past 20 years, for assessing biotic and abiotic degradation of petroleum hydrocarbons (petroleum hydrocarbon-contaminated [PHC]), MBTs—such as quantitative polymerase chain reaction (qPCR; and multiplex qPCR, qPCR arrays), next generation sequencing (NGS), and stable isotope methods (compound specific isotope analysis [CSIA], stable isotope probing [SIP])—have become more prominent as shown by the trends in the number of publications related to PHC remediation (Figure 1). PHC remediation-related publications mentioning SIP and PCR have increased slightly more than linearly from 2000 to 2020. However, publications mentioning CSIA started to plateau around 2010, possibly because the field maturated or the method has low sensitivity for detecting isotopic enrichment in some petroleum compounds (Blessing et al., 2008). Lastly, publications mentioning NGS were few before 2010, but then increased steeply and linearly, likely reflecting increasing availability of the method and its use in remediating sites contaminated with PHC (King et al., 2015; Tan et al., 2015). In many publications, several different MBTs are mentioned, consistent with the practice of obtaining multiple lines of evidence to improve the efficiency and cost-effectiveness of remediation.

From January 1, 2001, until April 30, 2021, in the United States alone, Microbial Insights, Inc. has processed over 116,296 groundwater

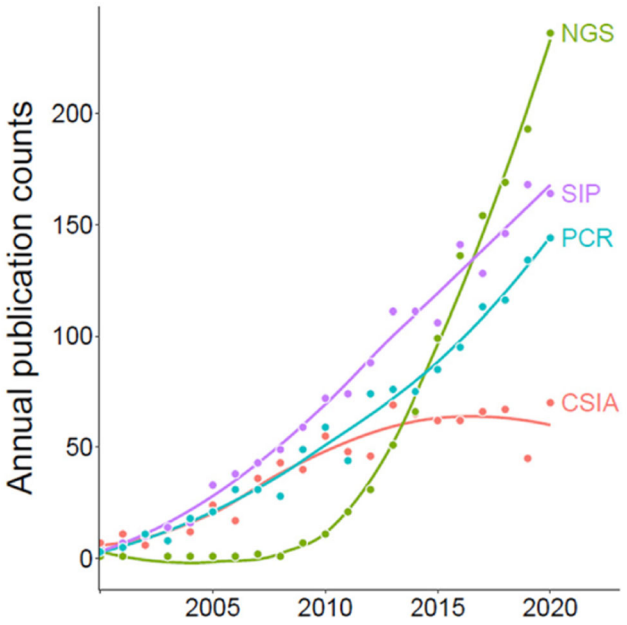

FIGURE 1 Publications mentioning petroleum remediation and MBTs. For years 2000–2020, Google Scholar was searched using the keywords: petroleum remediation AND (1) "stable isotope probing," (2) "compound specific isotope analysis," (3) "next generation sequencing," or (4) PCR AND "functional genes." The total publication count per year was tallied, plotted, and fit to a LOESS lin. CSIA, compound specific isotope analysis; LOESS, locally estimated scatterplot smoothing; MBT, molecular biological tool. NGS, next generation sequencing; PCR, polymerase chain reaction; SIP, stable isotope probing [Color figure can be viewed at wileyonlinelibrary.com]

samples from 19,365 different projects—an increase over 20 years in MBT sample analyses of nearly six-fold. Twenty percent of these MBT analyses were performed on samples from sites where the main contaminant of concern (COC) was PHC.

1.2 | MBT value along the remediation treatment train

Remediation strategies typically comprise a *treatment train* that starts with removing the source (excavation, in situ chemical oxidation [ISCO], etc.), moves to active bioremediation, and finally reaches monitored natural attenuation (MNA) or closure. At each stage, MBTs can provide actionable information for selecting remediation technologies and assessing their effectiveness.

Site characterization and remediation technology selection: Characterization of a site's chemistry (concentrations of contaminant, daughter products, etc.), geochemistry (redox status, electron acceptors, electron donors), and microbiology (qPCR, qPCR arrays, and NGS) together enable an informed selection of the most cost-effective remediation technology. Microbes carry out crucial steps in hydrocarbon degradation; therefore, the microbial genera and species and their genes capable of degrading PHCs at a site—that is, the site's biodegradative potential—are now routinely quantified using qPCR, qPCR arrays, and/or NGS.

ISCO employs oxidizing agents to chemically oxidize a wide range of volatile and semi-volatile contaminants, which can reduce their concentrations in weeks or months, without producing significant toxic degradation byproducts. After ISCO, biodegradation of the residual hydrocarbons (aerobic and anaerobic) can be enhanced (some ISCO products incorporate nutrients for just this purpose). Thus, qPCR, qPCR arrays, and NGS can be used to understand a site's capacity for these final steps in remediation.

In situ thermal remediation (ISTR) uses thermal conductive heating elements to heat the ground to temperatures above 100°C to accelerate the dissolution, desorption, volatilization/removal, and abiotic degradation of contaminants. After ISTR shutdown, bacteria can use the residual heat energy and the newly dissolved organic matter to "polish off" the residual contaminants. Additionally, lower temperature heating can stimulate biodegradation (and desorption) of contaminants (heat-stimulated bioremediation at 30°C to 40°C). These methods can be used in combination with amendments such as electron donors, electron acceptors, and bacterial cultures to promote biodegradation, and MBTs (qPCR, qPCR arrays, SIP, and NGS) can be used to assess these final remedial steps.

Natural Source Zone Depletion (NSZD) is the loss of light non-aqueous phase liquid (LNAPL) petroleum hydrocarbons from the subsurface due to dissolution, volatilization, and biodegradation. The dominant biodegradation process for NSZD is methanogenesis—the production of methane by methanogens, which currently include only Archaea (e.g., Methanomicrobia and Methanobacteria). Methanogens and methanotrophs, which can use methane as a carbon and energy source, and bacteria can live in complex syntrophic relationships. MBTs (e.g., NGS) can identify the microorganisms present down to the genus and even the species, their genes, as well as their relative proportions, informing the potential for NSZD.

Electron acceptor addition involves introducing a material that acts as an electron acceptor to enhance microbial degradation of organic contaminants, including PHCs. Microbial growth and biodegradative effectiveness can be limited by low concentrations of dissolved electron acceptors (dissolved oxygen, nitrate, and sulfate) at a site. Therefore, to stimulate bacterial growth and enhance bioremediation, an electron acceptor is often added or injected, using air (bioventing/biosparging), oxygen (oxygen infusion), or an oxygen-releasing product (e.g., ORC® [Regenesis], PermeOx® [PeroxyChem], EAS™ [EOS Remediation]). Knowledge of a site's biodegradation potential—for example, benzene, toluene, ethylbenzene, and xylene (BTEX) degraders (aerobic and anaerobic), methyl tertiary butyl ether (MTBE), and *tert*-butyl alcohol (TBA) degraders (aerobic and anaerobic), and naphthalene and other polycyclic aromatic hydrocarbon (PAH) degraders (aerobic and anaerobic)—obtained using MBTs, together with geochemistry enables selection of an appropriate electron acceptor, whether an oxygen-releasing product to stimulate aerobic bioremediation or a sulfate product to stimulate anaerobic bioremediation, and how much should be added.

MNA is a remediation strategy that relies on natural degradation processes for contaminant destruction. MNA can be a cost-effective approach; however, U.S. Environmental Protection Agency (EPA) guidance requires MNA to achieve cleanup objectives within a

reasonable time frame, making it necessary to estimate cleanup times whenever MNA is proposed as part of a cleanup strategy.

> EPA expects that MNA will be an appropriate remediation method only where its use will be protective of human health and the environment, and it will be capable of achieving site-specific remediation objectives within a timeframe that is reasonable compared to other alternatives. The effectiveness of MNA in both near-term and long-term timeframes should be demonstrated to EPA (or other overseeing regulatory authority) through: (1) sound technical analyses which provide confidence in natural attenuation's ability to achieve remediation objectives; (2) performance monitoring; and 3) contingency (or backup) remedies where appropriate. (EPA, 1999, p. 13)

Thus, it is essential to provide multiple lines of evidence, which can include the use of MBTs, in assessing the feasibility, timeframe, and cost-effectiveness of MNA.

2 | MOLECULAR BIOLOGICAL TOOLS

MBTs are used to initially characterize sites, select the remediation strategy, monitor treatment progress, and inform decisions to retreat, change treatments, or transition to MNA. Below we discuss the roles of MBTs in petroleum site remediation.

2.1 | Quantitative polymerase chain reaction

qPCR is used to quantify prokaryotic genera and their genes responsible for degrading petroleum hydrocarbons in soil or groundwater (Baldwin et al., 2010). In qPCR reactions, a target gene in a sample is located using short segments of DNA called primers. Many copies of the target gene are then generated. As each copy is made, a fluorescent marker is released, measured, and used to quantify the number of target genes present in the sample. Historically, qPCR has replaced viable bacteria counting, eliminating huge bias, since using viable plate counts only about 1%–10% of bacteria are culturable (Steen et al., 2019).

qPCR provides accurate quantification of target genes. It is also sensitive, capable of quantifying as few as 100 cells per sample. Furthermore, its dynamic range is large, over seven orders of magnitude. However, a single qPCR assay containing one primer pair and one probe can quantify only a single gene. The method is therefore most useful when the genus and the contaminant-degrading genes (functional genes) are well defined. However, this is rarely the case. Even simple contaminated sites contain complex mixtures of aliphatic, aromatic, cyclic, and heterocyclic compounds (e.g., gasoline and diesel fuel are comprised of hundreds of different compounds, including BTEX, MTBE, PAHs, and *n*-alkanes). Further complicating the picture, each contaminant is typically degraded by multiple anaerobic and aerobic pathways. Therefore, to characterize a site's biodegradative capacity, multiple genes should typically be quantified.

The design of qPCR assays requires knowledge of target genomic sequences. As of 2020, after sequencing for two decades, more than 200,000 bacterial and archaeal complete or draft genomes have been uploaded to public databases; however, they are estimated to comprise only 2% of all global prokaryotic taxa (based on 100% identities in the 16S-V4 region) (Zhang et al., 2020). Therefore, the expectation is that substantially more prokaryotic genomes sequences will be discovered. As more sequences become known, more meaningful conclusions can be drawn from PCR results. For example, in 2002, the sequences of only four bacterial strain sequences were known for the hydrocarbon-degrading gene, benzyl succinate synthase (BSS) (Beller et al., 2002). Currently, over 200 different sequences for this gene are known and qPCR assays are available. With knowledge of more sequences, evolutionarily conserved sequences are revealed that can serve as hybridization targets for primers and probes that can identify genes of additional taxa. A list of the current commercially available gene targets used in PHC-site remediation is provided in Table S1.

A common misconception in PHC-site remediation is that contaminant-degrading microbes are ubiquitous and always detected. However, in the Microbial Insights Inc. data set of 9,290 groundwater samples analyzed for two genes associated with BTEX anaerobic degradation, BSS and anaerobic benzene carboxylase (ABC), these genes were detected in only 53% and 11% of samples (median of 3.64E + 02 cells/ml and 2.00E + 01 cells/ml, respectively). Similarly, our analysis of aerobic degraders in 27,661 groundwater samples identified the phenol hydroxylase (PHE) gene in 81% of samples (median of 8.00E + 02 cells/ml) and the toluene dioxygenase (TOD) and toluene monooxygenase genes in only 44% and 49% of samples (median of 4.85E + 01 cells/ml and 1.09E + 03 cells/ml, respectively).

2.2 | qPCR arrays

qPCR arrays consist of numerous nanoliter-volume individual qPCR reactions arranged in a grid. These nanoliter reactions have accuracy and precision similar to larger-volume qPCR reactions. Importantly, arrays enable quantification of dozens of genes simultaneously and cost-effectively. Arrays are available for quantifying genes responsible for aerobic and anaerobic biodegradation of BTEX, PAHs, and a variety of short- and long-chain alkanes (Figure 2).

qPCR arrays offer increased accuracy compared to other high-throughput methods like microarrays and multiplex qPCR. Microarrays consist of multiple DNA probes attached to a solid slide. Microarray assessment is based on only a single-step hybridization. However, qPCR quantification involves primers annealing to the target genes in each of 30 or more cycles, which provides accurate quantification. Also, microarrays are mostly used in petrochemical site remediation to assess microbial genera based on the 16S ribosomal RNA (rRNA) gene, while qPCR arrays are largely used to assess biodegradative or functional genes.

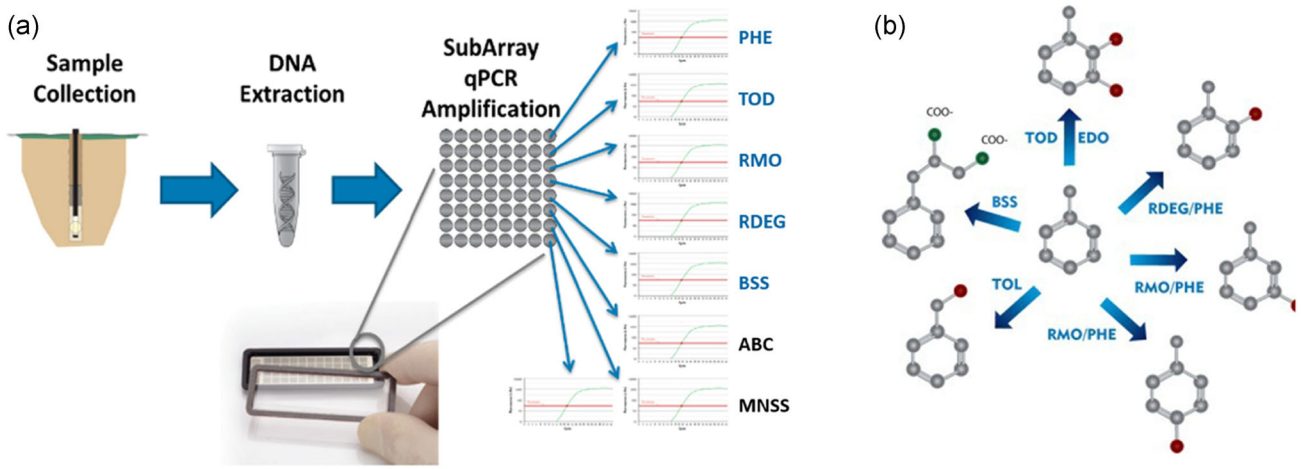

FIGURE 2 qPCR array technology for simultaneously quantifying a broad array of genes responsible for the biodegradation of petroleum hydrocarbons. (a) Illustration of the technology using toluene as an example. (b) Pathways for toluene biodegradation. The genes for enzymes in toluene degradation are indicated by three- or four-letter abbreviations. ABC, anaerobic benzene carboxylase; BSS, benzyl succinate synthase; MNSS, naphthylmethylsuccinate succinate synthase; PHE, phenol hydroxylase; qPCR, quantitative polymerase chain reaction; RDEG, ring hydroxylating toluene monooxygenase; RDEG, toluene monooxygenase 2; RMO, toluene monooxygenase; TOD, toluene dioxygenase [Color figure can be viewed at wileyonlinelibrary.com]

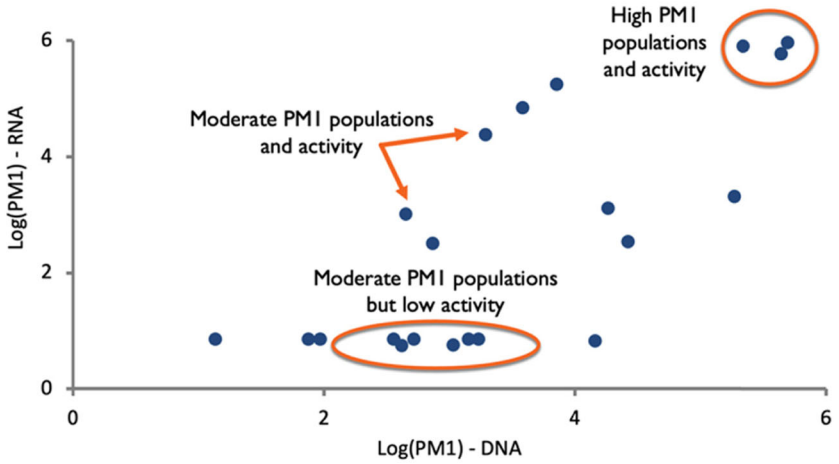

FIGURE 3 *Methylibium petroleiphilum* strain PM1: qPCR (16S DNA) versus RT-qPCR of PM1 rRNA. qPCR, quantitative polymerase chain reaction; RT-qPCR, reverse transcriptase qPCR [Color figure can be viewed at wileyonlinelibrary.com]

In multiplex qPCR, each qPCR reaction contains multiple different primer sets with each set quantifying a different gene. Like qPCR arrays, multiplex qPCR achieves some parallel quantification but is more prone to artifacts. Depending on the DNA sequences in a sample, mis-priming and artifactual amplification products can give erroneous data. This risk is reduced with qPCR arrays because each reaction uses one primer pair and is individually monitored to ensure reaction kinetics are not compromised.

qPCR array data, together with contaminant chemistry and geochemistry, can collectively provide a comprehensive view of a site's biodegradative capacity. For example, when redox conditions are variable and electron acceptor data provide only indirect evidence of microbial degradative activity, qPCR arrays can provide direct evidence of biodegradation, such as by demonstrating higher concentrations of the bacterial genera and their contaminant-degrading genes in monitoring wells within the plume versus background.

qPCR and qPCR arrays are used to assess baseline (pretreatment) concentrations of contaminant-degrader microorganisms and to evaluate a site's potential for MNA. These tools are also used in posttreatment performance monitoring to document the growth of the microorganisms in response to the treatments, providing direct evidence of treatment effectiveness.

2.3 | Reverse transcriptase qPCR (RT-qPCR)

While qPCR quantifies the DNA of genes having the potential to biodegrade contaminants, the genes may be present but not expressed. Therefore, to better understand the degradative potential at a site, two options are available. *First*, the concentration of the genes can be monitored over time to look for increases over background conditions. An increase in the genes with potential to degrade the

contaminant in the contaminated location but not in an uncontaminated location likely indicates that the organism is using the contaminant as a carbon and energy source. If higher concentrations are maintained only in the contaminated area over time and the contaminant is decreasing, then the conclusion is that organisms are utilizing the contaminant. The *second* option is to use RT-qPCR to quantify the RNA transcribed from genes involved in degradation of a contaminant to demonstrate the organism is actively expressing them.

For example, *Methylibium petroleiphilum* strain PM1 can grow on the gasoline additive, MTBE, but moderate strain concentrations, quantified by qPCR (10^3 cells/ml), have not consistently correlated with RNA expression (Figure 3). Therefore, rRNA quantification of PM1 can be a valuable tool when evaluating MTBE biodegradation, particularly under MNA because the presence of the target gene, as measured by qPCR does not necessarily indicate that it is being expressed, whereas, RT-qPCR quantification of a specific RNA transcript measures gene expression.

Another example where quantifying mRNA has shown importance is in aerobic biodegradation of BTEX compounds. TOD catalyzes the initial oxidation of benzene, toluene, and ethylbenzene, while PHE catalyzes the subsequent continued oxidation of phenol intermediates. TOD is often detected in high concentrations at sites impacted by PHCs as is PHE, which is often detected in high concentrations even during MNA. However, these genes' concentrations by qPCR do not always correlate with their mRNA concentrations by RT-qPCR (Figure 4).

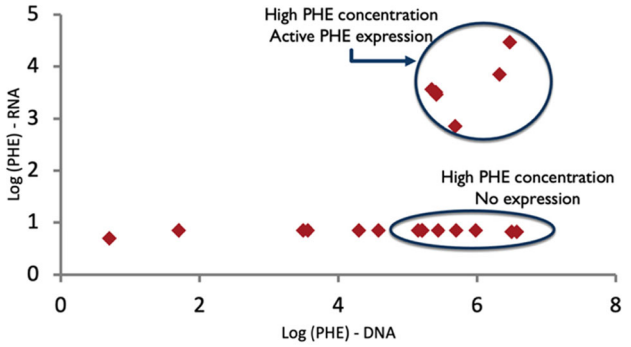

FIGURE 4 qPCR (DNA) versus RT-qPCR: PHE. PHE, phenol hydroxylase; qPCR, quantitative polymerase chain reaction; RT-qPCR, reverse transcriptase qPCR [Color figure can be viewed at wileyonlinelibrary.com]

Advances in RNA preservation and extraction now make RT-qPCR quantification of gene transcripts readily available. Whether qPCR (DNA) or RT-qPCR (RNA) should be used depends on the biodegradative process and the organisms involved. As described above, aerobic MTBE biodegradation (via *M. petroleiphilum* PM1) often does not correlate with moderate PM1 population densities quantified by qPCR of DNA; therefore, RNA analysis can be informative, particularly when evaluating lower concentrations of the methylotroph for MNA. Similarly, in evaluations of aerobic biodegradation of BTEX, the gene concentration (qPCR of DNA) and the gene expression of TOD and PHE often does not correlate; therefore, RNA analysis is recommended.

2.4 | Stable isotope tools (SIP and CSIA)

Both SIP and CSIA demonstrate biodegradation by measuring stable isotopes. SIP uses a synthetic ^{13}C-labeled contaminant of interest, called the stable isotope probe. In contrast, CSIA assesses naturally occurring heavy isotopes, typically ^{13}C or ^{2}H. Both methods assess the ratio of heavy to light isotope (e.g., ^{13}C:^{12}C) that is normalized to that of a standard (a carbonate mineral, Pee Dee Belemnite) (e.g., units are δ^{13}C or delta ^{13}C).

2.4.1 | Stable isotope probing

In SIP, the contaminant of interest is synthesized, for example, with nearly all the carbons as ^{13}C, whereas in nature ^{13}C only makes up ~1% of carbons (Dumont & Murrell, 2005; Radajewski et al., 2000). The

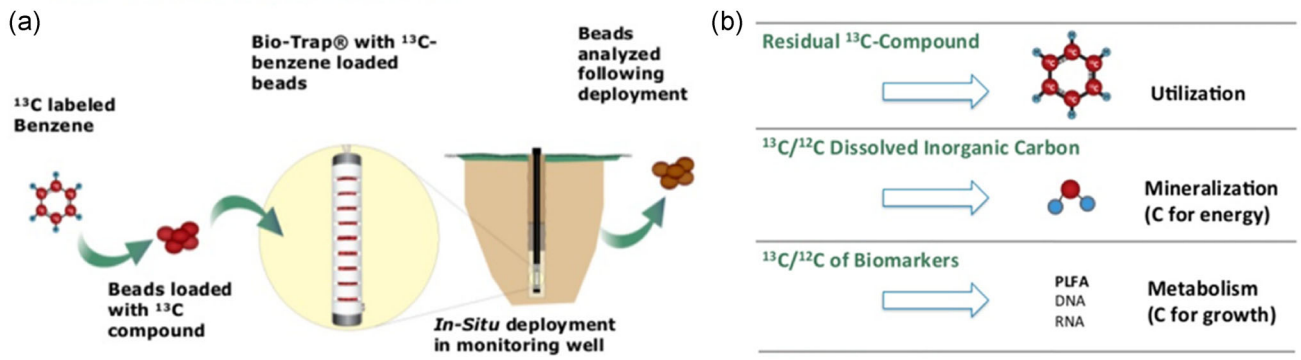

FIGURE 5 Overview of SIP for ^{13}C-benzene using Bio-Sep® beads in a Bio-Trap®. (a) ^{13}C-benzene is adsorbed to the PAC of Bio-Sep® beads (25% Nomex and 75% PAC). The porous beads loaded with ^{13}C-benzene are deployed to a monitoring well and become colonized by naturally occurring aquifer microbes. (b) After 30–45 days of incubation in the monitoring well, the beads are analyzed for residual ^{13}C-benzene and ^{13}C incorporated into dissolved inorganic carbon and PLFAs. PAC, powdered activated carbon; PLFA, phospholipid fatty acid; SIP, stable isotope probing [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 1 SIP strengths and weaknesses

Strengths	Weaknesses
No knowledge of the microbes or biodegradation pathways is needed	When used alone, identifying the responsible microbes requires isolation and analysis of ^{13}C-labeled biomolecules (e.g., DNA or PLFAs)
The contaminant can be a potential carbon/energy source or cometabolized (i.e., not incorporated into biomass)	Limited to contaminants used as carbon and energy sources or cometabolized
Inexpensive to synthesize many commonly occurring contaminants	Synthesis of labeled contaminants that are complex or large can be costly or not available
Broadly applicable to BTEX, MTBE, TBA, naphthalene, chlorobenzene, and emerging contaminants (e.g., 1,4-dioxane)	Not applicable to contaminants used as electron acceptors (e.g., chlorinated ethenes under anaerobic conditions)
Can be used with NAPL	Does not demonstrate abiotic degradation

Abbreviations: BTEX, benzene, toluene, ethylbenzene, xylene; MTBE, methyl *tert*-butyl ether; NAPL, nonaqueous phase liquid; PLFA, phospholipid fatty acid; TBA, *tert*-butyl alcohol.

^{13}C-contaminant "probe" is placed into monitoring wells to assess the ability of the microbes to degrade it. The ^{13}C-probe can be adsorbed to the powdered activated carbon (PAC) component of Bio-Sep® beads (Microbial Insights, Inc.) that have pores where microbes from the aquifer establish biofilm communities. After ~30–45 days incubation in the monitoring well, the beads are retrieved and assayed for the amount of ^{13}C-contaminant remaining (Figure 5), which is used to assess rates of contaminant degradation across the aquifer. The beads are also assayed for the incorporation of ^{13}C into biomass, typically into the microbial phospholipid fatty acids (PLFAs), but also nucleic acids (DNA and RNA), and into $^{13}CO_2$ (i.e., dissolved inorganic carbon: carbon dioxide, bicarbonate, and carbonate). The incorporation of ^{13}C from the probe into these molecules shifts their ^{13}C:^{12}C ratio to far exceed the ratio occurring naturally—providing conclusive evidence of biodegradation occurring in the aquifer under existing conditions.

In assessing MNA, analysis of the SIP data together with daughter product concentrations and numbers of contaminant-degrading microbes provides multiple lines of evidence for biodegradation. Table 1 lists the strengths and weaknesses of SIP.

2.4.2 | Compound specific isotope analysis

CSIA does not use a synthetically prepared, heavy isotope-labeled probe. Rather, it assesses naturally occurring heavy isotopes whose fraction remaining, as an organic compound degrades, changes in a predictable way (Elsner et al., 2012). For example, ^{13}C comprises ~1% of all naturally occurring carbon. As compared to ^{13}C bonds, ^{12}C bonds react slightly more quickly. Therefore, when contaminants undergo biodegradation, ^{13}C becomes enriched in the undegraded parent—that is, the δ^{13}C of the parent contaminant increases (isotope enrichment). CSIA can provide a conservative boundary on the extent of degradation.

When the δ^{13}C of the parent contaminant increases by more than two per mil (‰, parts per thousand) along the flow path from the source—that is, with time since the contaminant entered the aquifer—evidence is provided for ongoing biodegradation.

However, depending on a compound's characteristic δ^{13}C enrichment during degradation, the 2‰ criterion is met at different levels of biodegradation (Hunkeler et al., 2008). Compounds with the most negative enrichment factors (e.g., TCE undergoing reductive dechlorination) exceed 2‰ when <20% is degraded. In contrast, BTEX compounds do not reach the 2‰ threshold until almost 60% of the original contaminant mass has been degraded (Hunkeler et al., 2008).

Additionally, CSIA results can be confounded by NAPL (e.g., LNAPL) as in the source area. Several studies suggest that for compounds with small enrichment factors for carbon, such as the aromatic hydrocarbons, the larger enrichment factors of hydrogen may make two-dimensional CSIA of carbon and hydrogen the best approach to identify biodegradation (Hunkeler et al., 2008).

For a large number of pollutants, isotope enrichment factors for aerobic and anaerobic biodegradation are available that can be useful to benchmark enrichment factors found at sites (Hunkeler et al., 2008). However, because such comparisons can be confounded by the presence of NAPL, assessing downgradient wells for δ^{13}C enrichment along the flow path is important to provide biodegradation evidence. The strengths and weakness of CSIA are listed in Table 2.

When to use SIP versus CSIA

- *Is the contaminant used as an electron acceptor?*
  - If yes, SIP cannot be used so CSIA is the appropriate choice.
- *Is abiotic degradation predominating?*
  - If yes, as determined using products like zero valent iron or the occurrence of naturally occurring minerals such as iron sulfides (FeS, pyrite), iron oxides (e.g., magnetite), green rust, or iron carbonate, then CSIA is the only choice, since SIP is much less sensitive in detecting abiotic degradation.
- *Are the contaminants being used as a carbon and energy source?*
  - SIP is recommended for naphthalene and other PAHs because, for large compounds like these, lower fractionation during biodegradation reduces their CSIA detection sensitivity.
- *Are the contaminants BTEX and MTBE or TBA?*
  - For high concentrations of these contaminants, SIP is likely the best choice because it works in the presence of high concentrations of NAPL (e.g., LNAPL) as in the source area that can confound CSIA results.
  - With a long dilute plume, CSIA could be used, but probably should be done 2-dimensionally using carbon and hydrogen isotopes.

2.5 | Metagenomics (NGS)

Genera and genes that degrade contaminants can be quantified by qPCR when they are known. When they are not known, or the goal is to comprehensively understand a site's contaminant degradation and how it may change with time and treatments, then metagenomics via NGS has the potential to provide the needed information (Hidalgo et al., 2020).

2.5.1 | 16S rRNA gene sequencing (targeted taxonomic sequencing)

The prokaryotic 16S rRNA gene (1500 bp long) is a marker of taxonomy (Amann et al., 1995). The gene contains nine variable regions, interspersed between conserved regions, which are used in classifying genera and species in diverse microbial populations (Weisburg et al., 1991). This is performed using PCR with primers that hybridize to the conserved regions and amplify the variable regions so they can be sequenced and aligned to a large existing database of reference sequences to produce a frequency distribution of the genera and species in the database. These data can then be analyzed computationally to understand how microbial community members change across a site, with time, and after the addition of amendments.

Based on these data, even without the particular knowledge of the genera or genes that degrade a contaminant of interest, hierarchical clustering can be used to group samples with similar microbial composition and correlate with contaminant concentrations or geochemistry. Furthermore, samples can be grouped by microbial composition and visualized based on dimensionality-reduction algorithms such as principal component analysis. Moreover, the microbial community diversity and how it changes can be assessed (e.g., Shannon Genus Diversity Index).

Taxonomic 16S RNA gene sequencing can identify genera or species with genomes that have been sequenced and thereby identify their genes known to function in contaminant degradation.

2.5.2 | Shotgun sequencing (metagenome sequencing)

With 16S rRNA gene sequencing, a single gene is sequenced in all the organisms in a sample to which the primers hybridize but, with shotgun sequencing, all the genes in all the organisms in a sample are sequenced. This is accomplished by randomly shearing the DNA extracted from the sample and creating a library of fragments that is sequenced using high-throughput NGS. The resulting library of sequences is aligned to a genome database giving a readout of the known genes, genera, species/strains, and gene functions. While the function(s) of many genes is not known, as research advances, more and more functions are becoming known. Further differences between 16S rRNA gene sequencing and shotgun sequencing are highlighted in Table 3.

One example of the usefulness of shotgun sequencing is when species within a genus differ. For example, all *Geobacter* reduce iron but only certain species of this Proteobacteria genus reduce sulfur. However, the sulfur-reducing species are not distinguishable by taxonomic sequencing (16S RNA gene sequencing). In such cases, shotgun metagenomics (i.e., whole genome sequencing) can identify the sulfur-reducing species and their functional genes.

Also, shotgun metagenomic sequencing can be an efficient way to assess the presence of genes encoding an enzyme that is not strongly genetically conserved across genera. Acetylene hydratase is one such enzyme found in a variety of forms among 30 different

TABLE 2 CSIA strengths and weakness

Strengths
Qualitative but conclusive evidence of biotic and abiotic degradation
Can identify the source of the contamination (environmental forensics)
Relatively inexpensive
Estimates contaminant biodegradation
Useful for elucidating biotic versus abiotic and anaerobic versus aerobic biodegradation
Availability of database for comparing site enrichment data with enrichment factors of manufactured contaminants and the literature. Availability of software for generating contaminant degradation (mole fractions), dual-isotope, and modified Kuder plots
Weaknesses
Less isotopic fractionation with some large compounds (e.g., naphthalene, other PAHs, etc.), so >50%–80% biodegradation needed to conclude its occurrence
Masking of fractionation by products dissolving into the groundwater—for example, dissolution of contaminant from NAPL

TABLE 3 16S rRNA gene versus shotgun metagenomic sequencing

	16S rRNA gene sequencing	Shotgun sequencing
Bacterial coverage	More microbial species	Fewer microbial species
False positives	Low risk	Higher risk
Taxonomic resolution	Genus-species	Species-strains
Functional profiling	Based on known genomes	Can identify functional genes in unknown genomes

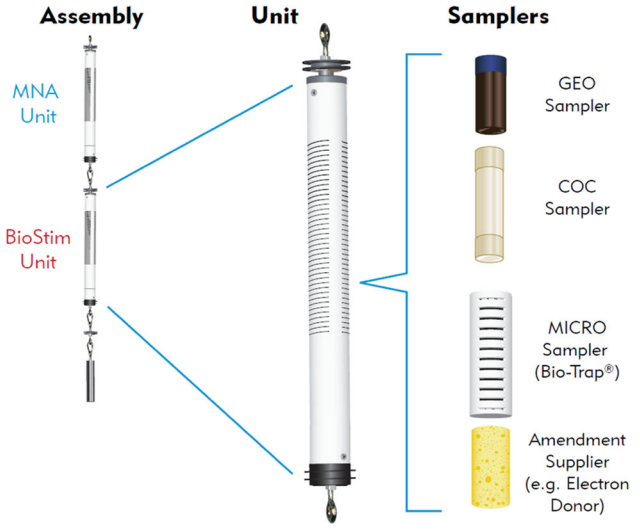

FIGURE 6 A typical two-unit ISM assembly. An assembly consisting of an *MNA Unit* and a *BioStim Unit* is shown (left). Each Unit consists of a length of slotted PVC pipe that houses three samplers and an amendment supplier. The *MNA Unit* lacks any amendment added to it and is used to assess baseline conditions. The *BioStim Unit* is amended with a nutrient or electron acceptor to assess the extent to which they stimulate microorganisms in the aquifer to degrade the contaminant. The *BioStim Unit* is shown in expanded detail and includes the following samplers: (1) the *GEO Sampler* contains nanopure water and is used to quantify geochemical parameters including electron acceptors (nitrate, sulfate, etc.), dissolved gases (methane, ethene, ethane), and chloride; (2) the *COC Sampler* is comprised of a passive diffusion bag filled with nanopure water that is used to quantify PHCs and daughter products; (3) the *MICRO Sampler* consists of a length of slotted PVC pipe containing Bio-Sep® beads (25% Nomex and 75% powdered activated carbon) that are porous with high surface area onto which microorganisms colonize; it is used to sample microorganisms for analysis of their genera and genes by qPCR array or metagenomics and can be combined with SIP by adsorbing the heavy-isotope probe to the activated carbon. The *Amendment Supplier* supplies the amendment to be tested (i.e., a nutrient or an electron acceptor) for improving geochemical, chemical, and microbial parameters, which are analyzed after incubation of the ISM in monitoring wells. COC, contaminant of concern; ISM, in situ microcosm; PVC, polyvinyl chloride; qPCR, quantitative polymerase chain reaction; SIP, stable isotope probing [Color figure can be viewed at wileyonlinelibrary.com]

genera. Identifying its genes across these genera by qPCR would require 15–20 different primer pairs. However, the acetylene hydratase gene sequences are present in the metagenome (from shotgun sequencing) and can be aligned to a reference database to assess the enzyme's relative abundance in samples.

2.6 | In situ microcosms (ISMs) for screening remediation options

ISMs are sampling units deployed in existing monitoring wells for assessing a site's subsurface conditions at baseline (MNA feasibility) and to test amendments added to the microcosm. Each ISM unit represents a different treatment condition and contains a remedial amendment in addition to contaminant, geochemical, and microbial samplers (Figure 6). An unamended MNA unit is also included to serve as a control. The units are connected together, and the assembly is deployed in a monitoring well. Rubber baffles sized to the well diameter separate the units and prevent crosstalk during incubation. ISM studies offer significant advantages over *laboratory microcosms* which have difficulty duplicating in situ conditions and *pilot field studies* which are often prohibitively expensive. ISM studies (e.g., using Bio-Traps® [Microbial Insights, Inc.]) provide microbial, chemical, and geochemical evidence for evaluating the degradation of contaminants and to screen amendments—at a fraction of the cost of lab-bench or pilot-scale studies.

ISM units can be tailored to investigate a wide variety of remedial approaches, but at petroleum-impacted sites, typically the units are configured to assess MNA and biostimulation (with nutrients or an electron acceptor such as oxygen or sulfate). After incubation of the ISM units in a monitoring well for 30–60 days, the samplers are analyzed to quantify microbial, chemical, and geochemical parameters.

Additionally, ISMs can be used in combination with SIP or CSIA. For SIP, the synthesized heavy isotope–labeled probe is adsorbed to the activated carbon of the Bio-Sep® beads and, after the incubation in the monitoring well, the ^{13}C-label is quantified (as described in Figure 5). For CSIA, the naturally occurring ^{13}C:^{12}C ratio is assessed.

3 | CASE STUDIES

The utility of MBTs is exemplified in the following four case studies that highlight real-world applications where these tools were used to evaluate remediation strategies and monitor performance, resulting in more informed site management decisions.

3.1 | Case study #1—ISMs: MNA versus anaerobic bioremediation

Case study #1 illustrates how ISMs were used to assess whether sulfate amendment would enhance PHC biodegradation relative to MNA conditions. At a petroleum storage facility where groundwater was impacted by BTEX, the subsurface conditions were highly anaerobic. The site managers wanted a cost-effective technique to assess MNA as well as to determine whether adding sulfate, an electron acceptor, to the site groundwater would stimulate the growth of anaerobic BTEX-degrading bacteria and enhance bioremediation.

ISMs were used to compare MNA to sulfate-based anaerobic bioremediation under the site conditions. Each ISM, as described in Figure 6, consisted of two units, an unamended MNA unit, and in this case, a sulfate-amended unit, that were deployed in two monitoring wells for 60 days. To assess whether sulfate stimulated BTEX-degrading bacteria, the ISMs were retrieved, and the bacteria inhabiting the Bio-Sep® beads were quantified by CENSUS qPCR (Figure 7).

The results showed that bacteria with BTEX-degrading genes were not detected in the unamended MNA units; however, the

sulfate-amended units contained 10^2–10^3 cells/bead of bacteria per bead with the genes benzene carboxylase (ABC) and BSS and 10^3–10^6 cells/bead of bacteria with the gene benzoyl coenzyme A reductase (BCR), suggesting that sulfate addition stimulated the growth of bacteria that anaerobically degrade BTEX.

The ISM test results were provided more cost-effectively than a bench-scale treatability study or a pilot-scale study and in a timelier manner. Based on these ISM results and historical groundwater monitoring results, the site managers decided to inject a sulfate-releasing product into the source area.

3.2 | Case study #2—ISMs: MNA versus aerobic bioremediation

Case study #2 summarizes an ISM employed at a petroleum-impacted site where BTEX and MTBE were the primary COCs. The subsurface conditions were aerobic, and site managers wanted a cost-effective technique to evaluate MNA and determine whether injection of an oxygen-releasing product would stimulate the growth of aerobic BTEX-degrading bacteria and enhance bioremediation.

ISMs composed of an unamended MNA unit and an oxygen-amended unit were deployed in impacted monitoring wells for approximately 35 days and, after retrieval, the Bio-Sep® beads were quantified by qPCR array (QuantArray®-Petro) for bacterial genes involved in BTEX biodegradation (Figure 8).

Overall, the concentrations of TOD, toluene monooxygenase (RMO), PHE, and other key genes involved in BTEX degradation were substantially greater in the oxygen-amended unit suggesting that the oxygen-releasing product stimulated the growth of aerobic BTEX degraders. Additionally, the total BTEX concentration sampled from the passive diffusion bags of the ISMs was 78% lower in the oxygen-amended unit. Based on these results, an oxygen-releasing product was injected in the source area.

3.3 | Case study #3—MNA assessment at a crude oil site

Case study #3 assesses the use of qPCR arrays at a petroleum storage tank farm where groundwater was impacted by the release of crude oil and condensate. BTEX were the primary COCs, but PAHs also exceeded risk-based limits. Geochemical monitoring indicated variable redox conditions but confirmed utilization of dissolved oxygen and other electron acceptors in the source area. Site managers considered MNA based on chemical and geochemical monitoring of groundwater but needed additional evidence to support the feasibility of MNA at the site.

qPCR array analysis (QuantArray®-Petro) was performed to quantify the genes responsible for aerobic and anaerobic biodegradation of BTEX and naphthalene to understand: (1) the concentrations of microbes with genes involved in the aerobic and anaerobic

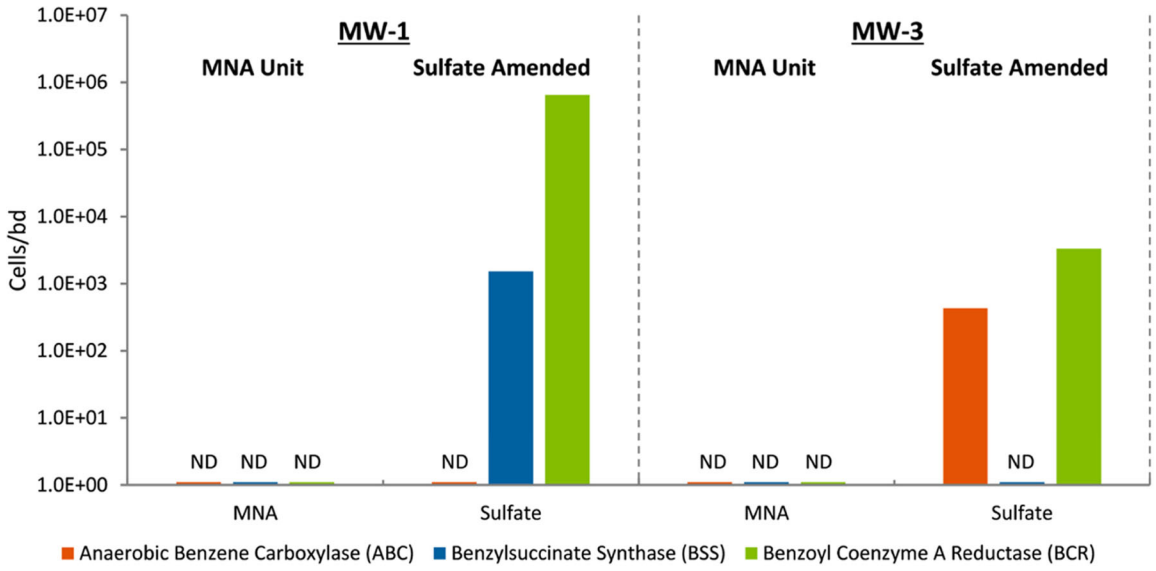

FIGURE 7 Anaerobic BTEX-biodegrading bacteria concentrations in Bio-Sep® beads of ISM deployed to MWs for 60 days. Bacteria containing the BTEX-degrading genes were quantified using CENSUS qPCR. The MNA units in MW-1 and MW-3 contained concentrations of ABC, BSS, and BCR genes below the laboratory detection limit, indicating low concentrations (cells/bead) of anaerobic BTEX degraders in the subsurface. However, in the sulfate-amended units these genes were relatively high, with BCR in MW-1 ranking at the ~80th percentile in the Microbial Insights Database. ABC, anaerobic benzene carboxylase; BCR, benzoyl coenzyme A reductase; BSS, benzyl succinate synthase; BTEX, benzene, toluene, ethylbenzene, xylene; ISM, in situ microcosm; MW, monitoring well; qPCR, quantitative polymerase chain reaction [Color figure can be viewed at wileyonlinelibrary.com]

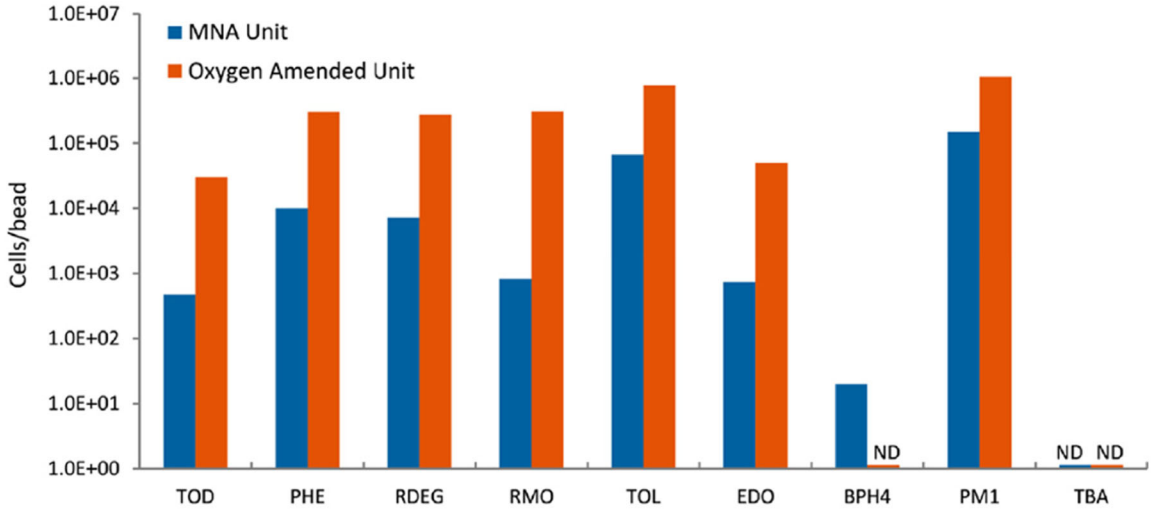

FIGURE 8 Aerobic BTEX-biodegrading genes present in Bio-Sep® beads of ISMs deployed to monitoring wells for 35 days quantified by qPCR array (QuantArray®-Petro). The concentrations of the genes and *M. petroleiphilum* strain, PM1 (quantified by 16S rRNA gene), that degrades the fuel additive, MTBE, are shown for the MNA (blue) and oxygen-amended (orange) units. The MNA unit concentrations indicate the potential for aerobic BTEX and MTBE biodegradation, but the concentrations of aromatic oxygenase genes (TOD, RMO, RDEG, PHE, TOL, and EDO) are relatively low to moderate. However, these genes in the oxygen-amended unit (orange) are on the order of 10^4 to 10^5 cells/bead—at least an order of magnitude greater than in the MNA unit—and, the concentrations of RMO, RDEG, PHE, and TOL rank near the 90th percentile in the Microbial Insights Database. BTEX, benzene, toluene, ethylbenzene, xylene; EBO, ethylbenzene/isopropylbenzene dioxygenase; ISM, in situ microcosm; MTBE, methyl tert-butyl ether; PHE, phenol hydroxylase; qPCR, quantitative polymerase chain reaction; RDEG, ring hydroxylating toluene monooxygenase; RMO, toluene monooxygenase; TOL, xylene/toluene monooxygenase; TOD, toluene/benzene dioxygenase [Color figure can be viewed at wileyonlinelibrary.com]

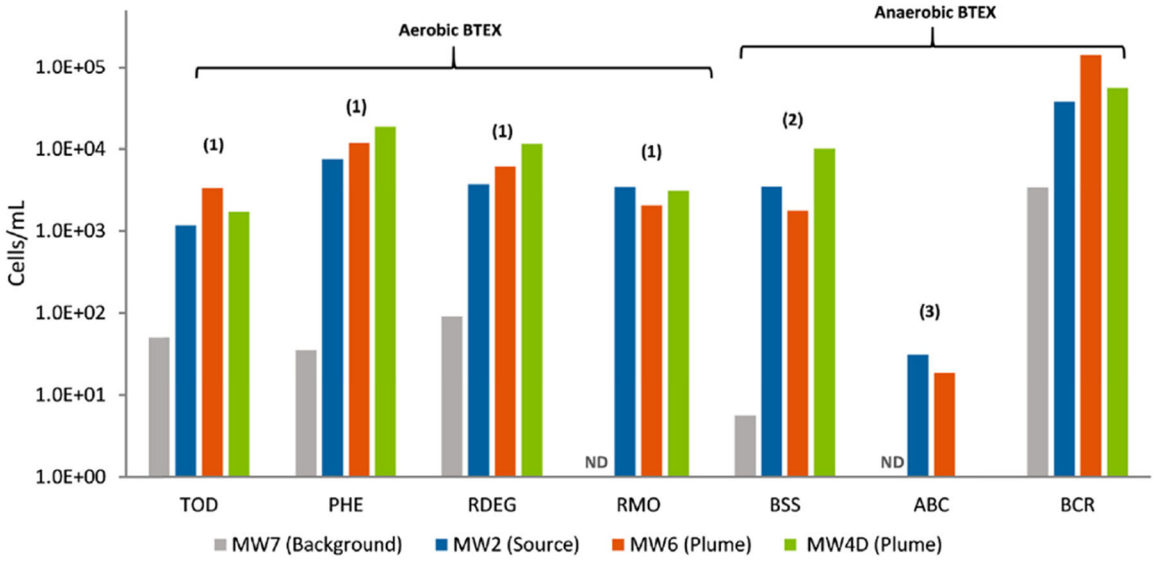

FIGURE 9 Aerobic and anaerobic BTEX-biodegrading genes present in impacted and background monitoring wells quantified by qPCR array (QuantArray®-Petro). Concentrations of TOD, PHE, and toluene/benzene monooxygenases (RMO and RDEG) genes for aerobic degradation were one to two orders of magnitude greater in impacted wells than in the background well MW7 (10^1 cells/ml). For the genes responsible for *anaerobic* BTEX biodegradation, the results were similar. Concentrations of BSS genes were two to three orders of magnitude greater in impacted wells. Additionally, ABC genes, which were below detection limits in the background well MW7, were detected in impacted wells MW2 and MW6, indicating growth of bacteria capable of anaerobic benzene biodegradation within the dissolved plume. Finally, BCR genes, which encode an enzyme involved in anaerobic metabolism of a common aromatic metabolite and are often detected at relatively high concentrations in background samples, were still detected at higher concentrations in the impacted wells. ABC, anaerobic benzene carboxylase; BCR, benzoyl coenzyme A reductase; BSS, benzyl succinate synthase; BTEX, benzene, toluene, ethylbenzene, xylene; PHE, phenol hydroxylase; qPCR, quantitative polymerase chain reaction; RDEG, ring hydroxylating toluene monooxygenase; RMO, toluene monooxygenase; TOD, toluene/benzene dioxygenase [Color figure can be viewed at wileyonlinelibrary.com]

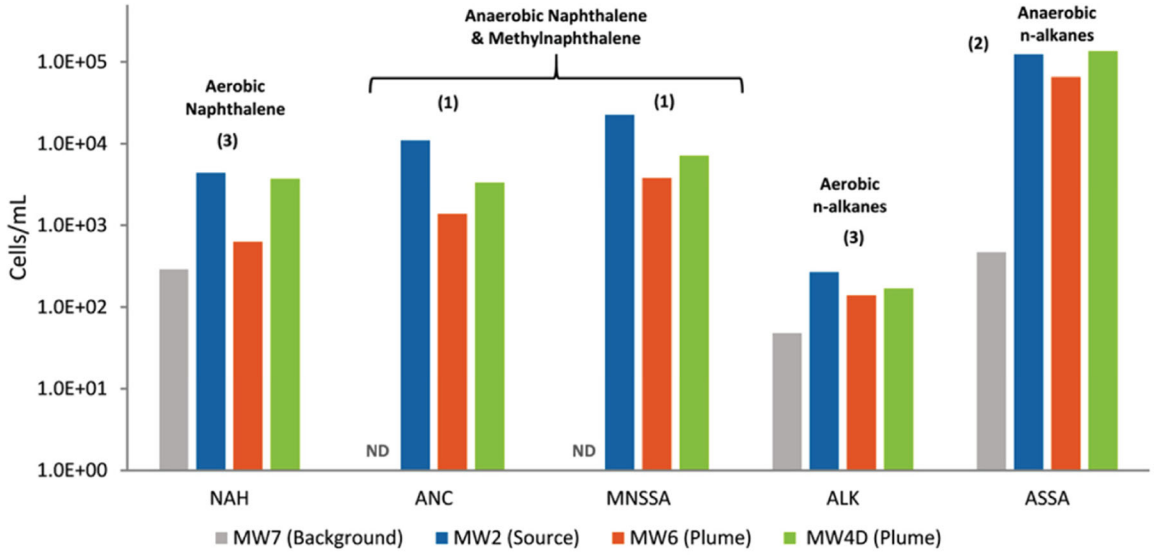

FIGURE 10 Concentrations of genes responsible for PAH and alkane biodegradation in impacted and background monitoring wells quantified by qPCR array (QuantArray®-Petro). (1) ANC and MNSSA genes, which were below detection limits in the background well MW7, were detected at high concentrations (10^3–10^4 cells/ml) in PHC-impacted wells. (2) Concentrations of ASSA genes which initiate anaerobic biodegradation of alkanes were more than two orders of magnitude greater in the impacted wells. (3) While somewhat less readily evident than for the anaerobic pathways, concentrations of NAH and ALK were also notably higher in samples from petroleum-impacted wells. ALK, alkane monooxygenase; ANC, anaerobic naphthalene carboxylase; ASSA, alkyl-succinate synthase; MNSSA, naphthylmethylsuccinate synthase; NAH, naphthalene dioxygenase; PAH, polycyclic aromatic hydrocarbon; qPCR, quantitative polymerase chain reaction [Color figure can be viewed at wileyonlinelibrary.com]

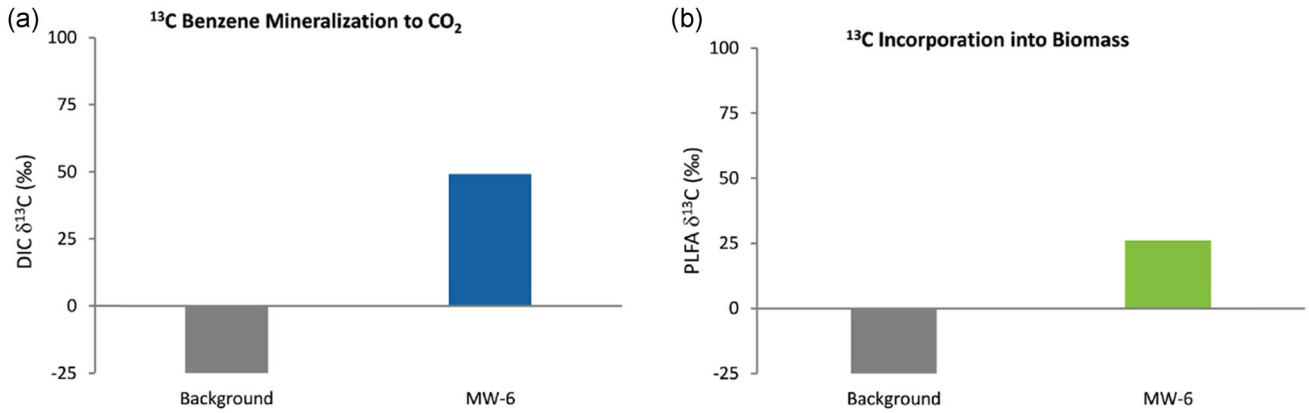

FIGURE 11 SIP results demonstrating benzene biodegradation in MW-6. (a) Incorporation of ^{13}C from ^{13}C-benzene into CO_2 (i.e., DIC: carbon dioxide, bicarbonate, and carbonate). ^{13}C-enriched DIC (blue bar, 49‰) compared to a background level (typically −25‰) conclusively demonstrated in situ mineralization of benzene. (b) Incorporation of ^{13}C from ^{13}C-benzene into PLFA (i.e., biomass). ^{13}C-enriched PLFA (green bar, 26‰) compared to a typical background level of −25‰ conclusively demonstrated incorporation of ^{13}C into biomass. DIC, dissolved inorganic carbon; PLFA, phospholipid fatty acid; SIP, stable isotope probing [Color figure can be viewed at wileyonlinelibrary.com]

degradation of these compounds under the existing conditions; and, (2) whether higher concentrations of these microbes and genes existed in the dissolved plume.

BTEX: The analysis revealed substantially greater concentrations of the genes involved in BTEX biodegradation in the impacted monitoring wells versus background (Figure 9).

PAHs: Concentrations of genes responsible for PAH and alkane biodegradation were substantially greater in impacted wells (MW2, MW6, and MW4D) than in the background well (MW7) demonstrating the growth of contaminant degraders within the dissolved plume (Figure 10).

As with BTEX degraders, the qPCR array results indicated high concentrations of aerobic as well as anaerobic naphthalene and alkane degraders, providing strong evidence for PAH and alkane biodegradation under existing site conditions. In a single analysis, the qPCR array (QuantArray®-Petro) quantified a broad spectrum of genes responsible for aerobic and anaerobic biodegradation of BTEX, PAHs, and other COCs in the dissolved plume. This evidence together with

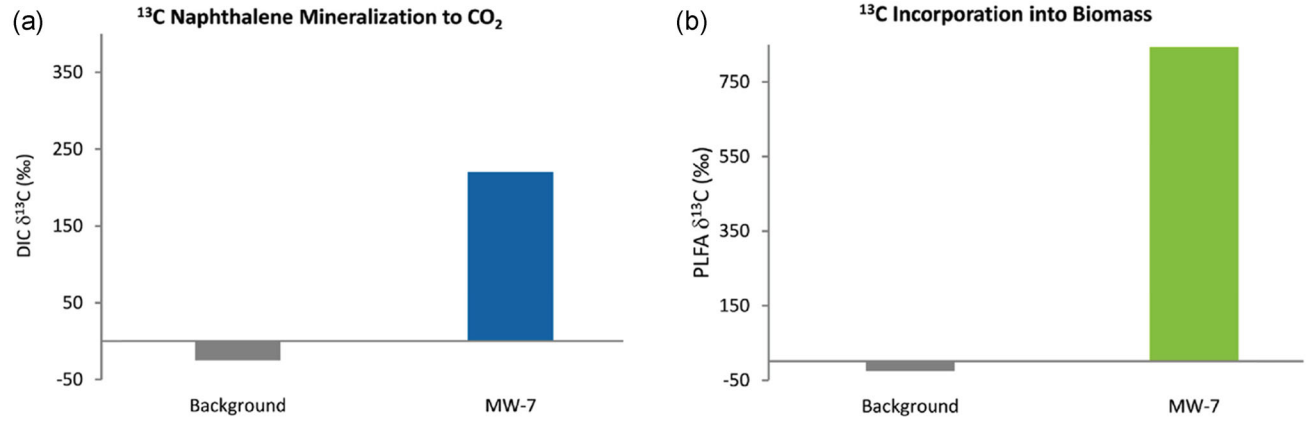

FIGURE 12 SIP results demonstrating naphthalene biodegradation in MW-7. (a) Incorporation of ^{13}C from ^{13}C-naphthalene into CO_2 (i.e., DIC: carbon dioxide, bicarbonate, and carbonate). ^{13}C-enriched DIC (blue bar, 220‰) compared to a background level conclusively demonstrated in situ mineralization of naphthalene. (B) Incorporation of ^{13}C from ^{13}C-naphthalene into PLFA (i.e., biomass). ^{13}C-enriched PLFA (green bar, 440‰) compared to background level conclusively demonstrated incorporation of ^{13}C from naphthalene into biomass. DIC, dissolved inorganic carbon; PLFA, phospholipid fatty acid; SIP, stable isotope probing [Color figure can be viewed at wileyonlinelibrary.com]

chemical evidence of stable or decreasing contaminant concentrations and geochemical evidence of microbial activity and electron acceptor availability led to selection of MNA as the site management strategy.

3.4 | Case study #4 SIP—evaluating MNA at a manufactured gas plant

Case study #4 involves a former manufactured gas plant (MGP) with benzene and naphthalene concentrations in groundwater that were stable to decreasing. Within the dissolved plume, moderate to high concentrations of BTEX and naphthalene degraders were demonstrated by qPCR array analysis (QuantArray®-Petro). Therefore, MNA was being considered. However, additional evidence was needed to conclude that benzene and naphthalene biodegradation was occurring, and more aggressive remediation was not warranted.

SIP analysis was employed for the MNA evaluation. Briefly, Bio-Traps® were amended with either ^{13}C-benzene or ^{13}C-naphthalene and placed in monitoring wells near the source area of the former MGP. The Bio-Traps were retrieved after 60 days. The results showed ^{13}C from the ^{13}C-labeled compounds was incorporated into CO_2 and PLFA, conclusively demonstrating the presence of microbes using these PHCs as an energy source and carbon source, respectively (Figures 11 and 12).

The SIP results obtained from existing monitoring wells at the site conclusively demonstrated that benzene and naphthalene were being biodegraded under existing aquifer conditions. These results supported MNA, avoiding more costly aggressive remediation.

4 | SUMMARY

MBTs are now routinely used to quantify concentrations of bacteria and their genes responsible for degrading petroleum hydrocarbons (including BTEX, MTBE)—to assess a site's biodegradative potential for supporting MNA and to evaluate biostimulation. Absolute quantification of the concentrations of specific microorganisms and functional genes encoding enzymes responsible for contaminant biodegradation provides site managers a direct line of evidence to evaluate remediation options and monitor remedy performance. Multiple lines of evidence—site chemistry (concentrations of contaminant, daughter products, etc.), geochemistry (redox status, electron acceptors, electron donors), and microbiology (genes responsible for biodegradation)—provide the most complete picture for cost effective site remediation. MBTs improve the understanding of biotic and abiotic degradation processes and thereby improve estimates of remedy effectiveness and attenuation rates, giving stakeholders greater confidence in making decisions regarding treatments, MNA, and closure.

REFERENCES

Amann, R. I., Ludwig, W., & Schleifer, K. H. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Review*, *59*(1), 143–169.

Baldwin, B. R., Biernacki, A., Blair, J., Purchase, M. P., Baker, J. M., Sublette, K., Davis, G., & Ogles, D. (2010). Monitoring gene expression to evaluate oxygen infusion at a gasoline-contaminated site. *Environmental Science & Technology*, *44*(17), 6829–6834. https://doi.org/10.1021/es101356t

Beller, H. R., Kane, S. R., Legler, T. C., & Alvarez, P. J. (2002). A real-time polymerase chain reaction method for monitoring anaerobic, hydrocarbon-degrading bacteria based on a catabolic gene. *Environmental Science & Technology*, *36*(18), 3977–3984. https://doi.org/10.1021/es025556w

Blessing, M., Jochmann, M. A., & Schmidt, T. C. (2008). Pitfalls in compound-specific isotope analysis of environmental samples. *Analytical and Bioanalytical Chemistry*, *390*(2), 591–603.

Dumont, M. G., & Murrell, J. C. (2005). Stable isotope probing—Linking microbial identity to function. *Nature Reviews Microbiology*, *3*(6), 499–504.

Elsner, M., Jochmann, M. A., Hofstetter, T. B., Hunkeler, D., Bernstein, A., Schmidt, T. C., & Schimmelmann, A. (2012). Current challenges in

compound-specific stable isotope analysis of environmental organic contaminants. *Analytical and Bioanalytical Chemistry*, *403*(9), 2471–2491.

Hidalgo, K. J., Sierra-Garcia, I. N., Dellagnezze, B. M., & de Oliveira, V. M. (2020). Metagenomic insights into the mechanisms for biodegradation of polycyclic aromatic hydrocarbons in the oil supply chain. *Frontiers in Microbiology*, *11*, 2250.

Hunkeler, D., Meckenstock, R. U., Sherwood Lollar, B., Schmidt, T. C., Wilson, J., Schmidt, T., & Wilson, J. (2008). A guide for assessing biodegradation and source identification of organic ground water contaminants using compound specific isotope analysis (CSIA). Washington, DC: Office of Research and Development, National Risk Management Research Laboratory, US Environmental Protection Agency.

King, G., Kostka, J., Hazen, T., & Sobecky, P. (2015). Microbial responses to the Deepwater Horizon oil spill: From coastal wetlands to the deep sea. *Annual Review of Marine Science*, *7*, 377–401.

Radajewski, S., Ineson, P., Parekh, N. R., & Murrell, J. C. (2000). Stable-isotope probing as a tool in microbial ecology. *Nature*, *403*(6770), 646–649.

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 Journal*, *13*(12), 3126–3130. https://doi.org/10.1038/s41396-019-0484-y

Tan, B., Ng, C. M., Nshimyimana, J. P., Loh, L.-L., Gin, K. Y.-H., & Thompson, J. R. (2015). Next-generation sequencing (NGS) for assessment of microbial water quality: Current progress, challenges, and future opportunities. *Frontiers in Microbiology*, *6*, 1027.

U.S. Environmental Protection Agency (EPA). P. (1999). *Use of monitored natural attenuation at Superfund, RCRA corrective action, and underground storage tank sites. OSWER Directive, 9200*. Author.

Weisburg, W. G., Barns, S. M., Pelletier, D. A., & Lane, D. J. (1991). 16S ribosomal DNA amplification for phylogenetic study. *Journal of Bacteriology*, *173*(2), 697–703. https://doi.org/10.1128/jb.173.2.697-703.1991

Zhang, Z., Wang, J., Wang, J., Wang, J., & Li, Y. (2020). Estimate of the sequenced proportion of the global prokaryotic genome. *Microbiome*, *8*(1), 134. https://doi.org/10.1186/s40168-020-00903-z

AUTHOR BIOGRAPHIES

Dora Taggart is President and CEO of Microbial Insights, Inc. Over the last 20 years, she has been involved in pioneering work for the development of MBTs in remediation. She is a global ambassador for the tools presenting as an invited speaker at workshops and conferences around the world.

Kate Clark, PhD is the technical director at Microbial Insights Inc. She has expertise and experience ranging from lab bench to field sites and routinely works to aid project managers in MBT assay selection, study design, and data interpretation.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Taggart, D. M., & Clark, K. (2021). Lessons learned from 20 years of molecular biological tools in petroleum hydrocarbon remediation. *Remediation Journal*, 1–13. https://doi.org/10.1002/rem.21695