



The Microbial Community and Functional Potential in the Midland Basin Reveal a Community Dominated by Both Thiosulfate and Sulfate-Reducing Microorganisms

 Kara Tinker,^{a,b} Daniel Lipus,^{a,c,d} James Gardiner,^{a,b} Mengling Stuckman,^{a,b} Djuna Gulliver^a

^aNational Energy Technology Laboratory (NETL), Pittsburgh, Pennsylvania, USA

^bNETL Support Contractor, Pittsburgh, Pennsylvania, USA

^cOakridge Institute for Science and Education, Oak Ridge, Tennessee, USA

^dGFZ German Research Centre for Geosciences, Section Geomicrobiology, Potsdam, Germany

ABSTRACT The Permian Basin is the highest producing oil and gas reservoir in the United States. Hydrocarbon resources in this region are often accessed by unconventional extraction methods, including horizontal drilling and hydraulic fracturing. Despite the importance of the Permian Basin, there is no publicly available microbiological data from this region. We completed an analysis of Permian produced water samples to understand the dynamics present in hydraulically fractured wells in this region. We analyzed produced water samples taken from 10 wells in the Permian region of the Midland Basin using geochemical measurements, 16S rRNA gene sequencing, and metagenomic sequencing. Compared to other regions, we found that Permian Basin produced water was characterized by higher sulfate and lower total dissolved solids (TDS) concentrations, with a median of 1,110 mg/L and 107,000 mg/L. Additionally, geochemical measurements revealed the presence of frac hits, or interwell communication events where an established well is affected by the pumping of fracturing fluid into a new well. The occurrence of frac hits was supported by correlations between the microbiome and the geochemical parameters. Our 16S rRNA gene sequencing identified a produced water microbiome characterized by anaerobic, halophilic, and sulfur reducing taxa. Interestingly, sulfate and thiosulfate reducing taxa including *Halanaerobium*, *Orenia*, *Marinobacter*, and *Desulfohalobium* were the most prevalent microbiota in most wells. We further investigated the metabolic potential of microorganisms in the Permian Basin with metagenomic sequencing. We recovered 15 metagenome assembled genomes (MAGs) from seven different samples representing 6 unique well sites. These MAGs corroborated the high presence of sulfate and thiosulfate reducing genes across all wells, especially from key taxa including *Halanaerobium* and *Orenia*. The observed microbiome composition and metabolic capabilities in conjunction with the high sulfate concentrations demonstrate a high potential for hydrogen sulfide production in the Permian Basin. Additionally, evidence of frac hits suggests the possibility for the exchange of microbial cells and/or genetic information between wells. This exchange would increase the likelihood of hydrogen sulfide production and has implications for the oil and gas industry.

IMPORTANCE The Permian Basin is the largest producing oil and gas region in the United States and plays a critical role supplying national energy needs. Previous work in other basins has demonstrated that the geochemistry and microbiology of hydrocarbon regions can have a major impact on well infrastructure and production. Despite that, little work has been done to understand the complex dynamics present in the Permian Basin. This study characterizes and analyzes 10 unique wells and one groundwater sample in the Permian Basin using geochemical and microbial techniques. Across all wells we found a high number of classic and thiosulfate reducers, suggesting that hydrogen sulfide production may be especially prevalent in the

Editor Courtney J. Robinson, Howard University

This is a work of the U.S. Government and is not subject to copyright protection in the United States. Foreign copyrights may apply.

Address correspondence to Djuna Gulliver, Djuna.Gulliver@netl.doe.gov.

The authors declare no conflict of interest.

Received 24 March 2022

Accepted 20 May 2022

Published 13 June 2022

Permian Basin. Additionally, our analysis revealed a biogeochemical signal impacted by the presence of frac hits, or interwell communication events where an established well is affected by the pumping of fracturing fluid into a new well. This information can be utilized by the oil and gas industry to improve oil recovery efforts and minimize commercial and environmental costs.

KEYWORDS 16S RNA, Permian Basin, environmental microbiology, geomicrobiology, hydraulic fracturing, hydrocarbons, metagenomics

Hydraulically fractured oil and natural gas are one of the world's fastest growing major fuel sources and represent an essential energy source in the United States (1, 2). While efforts to minimize greenhouse gas emissions have advanced, hydraulically fractured oil and natural gas have emerged as major replacements for coal and are an important gateway toward renewable energy sources (3, 4). Advanced horizontal drilling and hydraulic fracturing technologies have contributed to increased yields as they can extract large quantities of oil and natural gas from previously impermeable shale formations (5, 6). However, these technologies are often associated with frac hits, or interwell communication events where an established well is affected by the pumping of fracturing fluid into a new well. Additionally, hydraulic fracturing operations generate vast amounts of wastewater, referred to as produced water, which is characterized by high concentrations of salt, metals, organic compounds, radionuclides, and microorganisms (7–9). Prior to reuse or disposal, these fluids need to be remediated, creating additional operational and environmental expenses (10, 11).

Issues associated with produced water management such as corrosion and fouling events, leakage and spillage of potentially hazardous substances, and hydrogen sulfide contamination highlight the need to improve and optimize current management strategies and approaches (6, 8, 12–15). Many of the key issues associated with produced water management have been linked to microbial activity (8, 16–22). Therefore, the role of microorganisms during hydraulic fracturing operations and hydraulic fracturing produced water management has become a focal point (6, 8). Currently, it is unclear if the majority of the microbial biomass found in produced fluids and the hydraulic fracturing infrastructure stems from the subsurface or is introduced and distributed through the fracturing process and by the continuous recycling of produced water (8, 23). However, research efforts to this point suggest that despite the use of a wide array of biocides, low diversity microbial populations adapted to the hypersaline and extreme conditions are able to establish themselves in the borehole and surrounding well environment including the gas separator and storage tanks. Biocides may actually play an important role in shaping the community, as shale-associated microorganisms are often resistant to biocides (18, 24, 25) and ineffective biocide treatments may result in increased antimicrobial resistance genes and/or enhanced microbial activity (18, 19, 26). In addition to the deleterious effects microbial activity can have on oil and gas operations and the surrounding environments, recent work also suggests that microbiological data from hydraulic fracturing well sites can be used to forecast geographic areas and subsurface sections, which may be particularly productive (27). These contrasting roles highlight the necessity to study the microbial community, microbial processes, and biogeochemical interactions in and around hydraulic fracturing wells so we can incorporate these findings in current produced water management strategies.

The Permian Basin is currently the largest oil-producing basin in the United States, producing nearly 3.5 million barrels of oil and 12 million cubic feet of natural gas per day (28). Therefore, obtaining insights on microbial processes potentially interfering with operations and evaluating biogeochemical interactions as part of advanced produced water management is of significant scientific and engineering interest. However, to our knowledge there is little publicly available microbiological data from the Permian Basin. A 2015 study by Kilbane et al. (29) investigated the potential for microbial sulfide production using culture-based tests and the sequencing of two

samples, one water and one biofilm. Results revealed microbial abundances of 10^3 cells/mL in water samples and 10^5 cells per cm^2 biofilm (29). *Arcobacter* was the dominant taxa present in the water sample while *Desulfovibrio* was dominant in the biofilm sample. A recent paper by Lascelles et al. (27) discusses the usefulness and necessity to analyze DNA from Permian formations, but does not provide any specific results on microbial abundance of community structure. Finally, a 2019 paper by Ursell et al. measured the presence of microbial DNA markers in produced fluid and well cuttings from different depths in the Permian Basin (30). Although they found unique DNA marker profiles at different depths, they do not report their exact methodology or the taxonomy of the microbes. These results demonstrate the overall need for robust microbial studies in the Permian Basin and suggest that there is a high potential for sulfide producing halophilic microorganisms. However, the unavailable proprietary sequencing data limits the impact of previous work and therefore additional investigation is required.

While microbiological data on Permian Basin produced waters is scarce, several studies have investigated the geochemistry and microbiology of the Marcellus Shale and other major oil and gas regions in the United States (22, 23, 31, 32). From this work, we know shale wells are characterized by high salinity and cation concentrations, with total dissolved solids (TDS) concentrations up to 40 g/L and excess Fe^{2+} , Ba^{2+} , and Mg^{2+} in these systems (22, 33–37). These conditions usually result in a halophilic microbial community with high osmotic and oxidative stress tolerance abilities (19, 36, 38). Both amplicon based and metagenomic approaches have identified several microbial taxa and reconstructed metabolic pathways that can contribute to corrosion, fouling, and souring events in hydraulic fracturing wells, pipelines, tanks, and associated equipment (17, 22–24, 31, 35). In addition, microorganisms in produced waters are known to actively express key genes involved in sulfide production and methanogenesis pathways (24). However, this work also demonstrates that oil and gas regions have unique geochemical and microbiological signatures (22, 23, 31, 32, 35, 37). For example, Bakken region produced waters have been found to exhibit greater TDS concentrations and lower biomass than Marcellus Shale or Barnett Shale produced waters, which results in a distinct ecological niche within the Bakken Shale (8, 23, 33–35). Therefore, although we can make some predictions about the Permian Basin based on previous work in other hydrocarbon regions, it is necessary to complete a comprehensive study of produced water from the Permian region in order to accurately understand and manage local resources.

Our overall goal is to assess the unique geochemical and microbiological features of hydraulically fractured shales in the Permian Basin in order to improve enhanced oil recovery efforts and minimize commercial and environmental costs. In this study, we analyzed produced water samples taken from 10 wells in the Permian region of the Midland Basin. We also analyzed one proximal groundwater sample, as groundwater is typically used as source water for fracturing fluid in this region. We characterized the microbial community of water samples using 16S rRNA gene sequencing, collected detailed geochemistry data from each sample, and used metagenomic shotgun sequencing to functional potential present in select hydraulically fractured wells. To our knowledge, this is the first extensive analysis of Permian produced water samples and represents both a significant advancement on microbial life in subsurface environments and provides valuable knowledge that can be utilized by the oil and gas industry.

RESULTS

Geochemistry of produced water from the Permian Basin. We completed geochemical analysis on 14 produced water samples and one groundwater sample in order to better understand the unique ecosystem present in the Permian Basin (Fig. S1, Table S1). Our produced water samples were collected from 10 unique well sites, with four of those sites having both paired wellhead and separator samples

TABLE 1 Geochemical measurements for produced water samples and proximal groundwater in the Permian Basin

Well no.	pH	Alkalinity (mg/L)	TDS (mg/L)	Ca ²⁺ (mg/L)	Mg ⁺ (mg/L)	Na (mg/L)	K ⁺ (mg/L)	Total Fe (mg/L)	So ₄ ²⁻ (mg/L)	Cl ⁻ (mg/L)	Br ⁻ (mg/L)
1S	6.6	522	58300	1680	601	18500	338	43	3650	32800	229
2S	6.9	737	69700	1830	506	23500	383	15	2680	39800	308
3S	6.4	366	54500	1350	517	18200	299	24	3490	30100	238
4S	6.7	337	84000	2150	509	28400	436	53	2500	49300	388
5W	7.3	378	111000	3210	593	37900	520	49	1300	66700	524
6S	7.0	368	107000	3180	565	36600	415	BDL	1090	64300	469
6W ^a	7.2	512	106500	3310	581	36300	409	67	1115	63650	470
7S	6.6	290	122000	3600	611	41700	610	58	817	74100	614
7W	6.7	388	123000	3610	615	42000	613	52	856	74500	604
8S	6.5	425	120000	3430	602	40400	542	64	864	73300	555
8W	6.7	381	118000	3420	599	40300	540	52	827	71300	554
9S	6.6	398	120000	3830	623	40400	428	87	626	73100	530
9W	6.6	464	118000	3790	617	39900	423	80	630	71300	524
10S	6.9	95	53200	1320	504	18000	302	24	3420	29300	229
Min ^b	6.4	95	53200	1320	504	18000	299	15	626	29300	229
Max ^b	7.3	737	123000	3830	623	42000	613	87	3650	74500	614
Median ^b	6.7	385	109000	3260	596	37250	426	52	1103	65500	497
Avg ^b	6.8	404	97514	2836	575	33007	447	51	1705	58111	445
Proximal groundwater	7.1	427	8432	92.9	48.5	2830	11.0	ND	2070	2960	6.23

^aValues reported for 6W are the average of two technical replicates.

^bValue was calculated using only produced water samples; the proximal groundwater sample was not included.

(Table S1). We also collected gas measurements at each wellhead, although we were unable to do so at the separator. We found that all wellheads had a measurable concentration of H₂S ranging from 111–391 ppm. We found that the produced waters display a small range in pH, from 6.4 to 7.3, with all values being circumneutral (Table 1). Alkalinity ranged from 95 to 737 mg/L, with an average of 404 mg/L and a median of 385 mg/L across all produced water samples (Table 1). Our groundwater pH and alkalinity measurements fell within these ranges, at 7.1 and 427 mg/L, respectively (Table 1). These values are consistent with previously collected geochemistry data from the Permian Basin (39, 40).

Total dissolved solids measurements for produced water samples range from 53,200 to 123,000 mg/L, with the majority of the TDS represented by high concentrations of chloride and sodium (Table 1). Notably, half of the produced water samples have a significantly lower TDS measurement than the median of 107,000 mg/L (Table 1). These lower TDS produced water samples have an elevated dissolved sulfate range (2,500–3,650 mg/L) relative to the overall dissolved sulfate median of 1,110 mg/L (Table 1). The groundwater also recorded a sulfate concentration (2,070 mg/L) higher than the median (Table 1). Since groundwater is commonly used as a major component of fracturing fluid in this region, the reduced TDS and elevated sulfate values in the produced water implies recent contact with injected groundwater. As all wells sampled were older than 6 months active, at which point produced water has a distinct chemical signature compared to the initial fracture fluid injection (8), this low TDS high sulfate signature was most likely from a nearby frac hit. It is notable that the sulfate concentrations in produced fluids with reduced TDS measurements were higher than groundwater sulfate levels. This suggests that additional reactions occurred within the reservoir after the frac hit event.

Microbial diversity and distribution of produced water microbial communities.

We found that there was a large range in all alpha diversity metrics (Table 2, Table S1, Fig. S1). Produced water from the separator site 2 (2S) had the highest number of unique amplicon sequence variants (ASVs) (66) as well as the highest recorded values for richness (77), diversity (3.03) and evenness (0.720) (Table 2, Table S1, Fig. S1). In contrast, produced water from the separator at site 9 (9S) had the lowest number of unique ASVs (7) as well as the lowest Chao1 index (7) (Table 2, Table S1, Fig. S1). Notably, the wellhead sample at site 9 (9W) had a significantly higher number of unique ASVs (29) and Chao1 index (28.1), although the diversity and evenness for 9S and 9W were comparable (Table 2, Table S1, Fig. S1). Finally, sample 1S had the lowest measurements for diversity and evenness, at 0.974 and 0.299, respectively (Table 2, Table S1, Fig. S1).

TABLE 2 Alpha diversity metrics for experimental samples^a

Sample ID	ASVs	Richness ^b	Diversity ^c	Evenness ^d
1S	26	37.3	0.974	0.299
2S	67	77	3.03	0.720
3S	22	26	1.20	0.389
4S	33	54	2.20	0.630
5W	21	35	1.26	0.414
6W	20	38	1.51	0.504
9S	7	7	1.21	0.624
9W	28	28.1	1.68	0.504
10S	47	64.5	1.73	0.450
GW	23	25.5	1.39	0.444

^aMetrics were calculated after sequence libraries were resampled to the depth of the sample with the fewest sequences from this experiment (1023 sequences).

^bChao1 Index.

^cShannon Index.

^dPielou's Evenness.

In order to visualize differences among water samples, we constructed two ordination plots using non-metric dimensional scaling (NMDS) based on Bray-Curtis dissimilarity distances calculated after sequence libraries were subsampled to the lowest sequence depth (1023 sequences). The first NMDS plot contained only produced water samples (Fig. 1), while the second plot contained all produced water samples as well as a groundwater sample (Fig. S2). For both NMDS plots, we used analysis of similarity (ANOSIM) calculations to measure possible clustering patterns. At the time of sampling, several wells were reported as frac hits by well operators (Table S2). Therefore, we specifically tested for clustering by sample origin (separator, wellhead, or groundwater), well age, and frac hit status (affected, unaffected, or groundwater). Among only the produced water samples, we found no statistically significant clustering by sample origin, well age, or frac hit status. However, when we incorporated the groundwater sample, we measured statistically significant clustering by frac hit status ($R = 0.4806$, $P < 0.05$), although there was no statistically significant clustering by sample origin or well age.

We were also interested in understanding the relationship between the microbial community present in each sample and the associated sample geochemistry. In order to measure the strength of this relationship, we completed a Mantel test. We found no statistically significant relationship among the produced water samples, however we found a statistically significant relationship when we incorporated the groundwater sample ($R = 0.4552$, $P < 0.05$). We also calculated and fit environmental vectors onto our two ordination plots in order to identify any specific geochemical measurements that had a statistically significant relationship with the microbial data. We found no statistically significant environmental vectors among the entire data set (SI Fig. 2). However, when we examined only the produced water, we found that 7 of the 11 geochemical measurements have a statistically significant correlation with the microbial data (Table S4). The R^2 of the statistically significant correlations ranged from 0.6137 to 0.7545, with TDS having an R^2 of 0.6252 (Table S4). Total dissolved solids (TDS) is a measurement of the total amount of cations and/or anions dissolved in water, therefore it is unsurprising that many of the TDS constituents also had a significant correlation with the microbial data. These include calcium, magnesium, sodium, iron, sulfate, and chloride (Table S4). In contrast alkalinity, potassium, bromide, and pH did not have a statistically significant relationship with the microbial data (Table S4).

Composition of Permian Basin produced water microbial communities. 16S rRNA gene sequencing revealed a produced water microbial community characterized by anaerobic, halophilic, and sulfur reducing taxa (Fig. 2, Fig. S3 and S4). We found that the majority of these taxa were bacteria, with archaea only representing an average of 0.47% of the sequences across all produced water samples. In contrast, the groundwater sample was primarily composed of an archaeon from the *Methanobacteriaceae* family (Fig. 2, Fig. S3). This archaeon represented 55.58% of all of the sequences from the groundwater

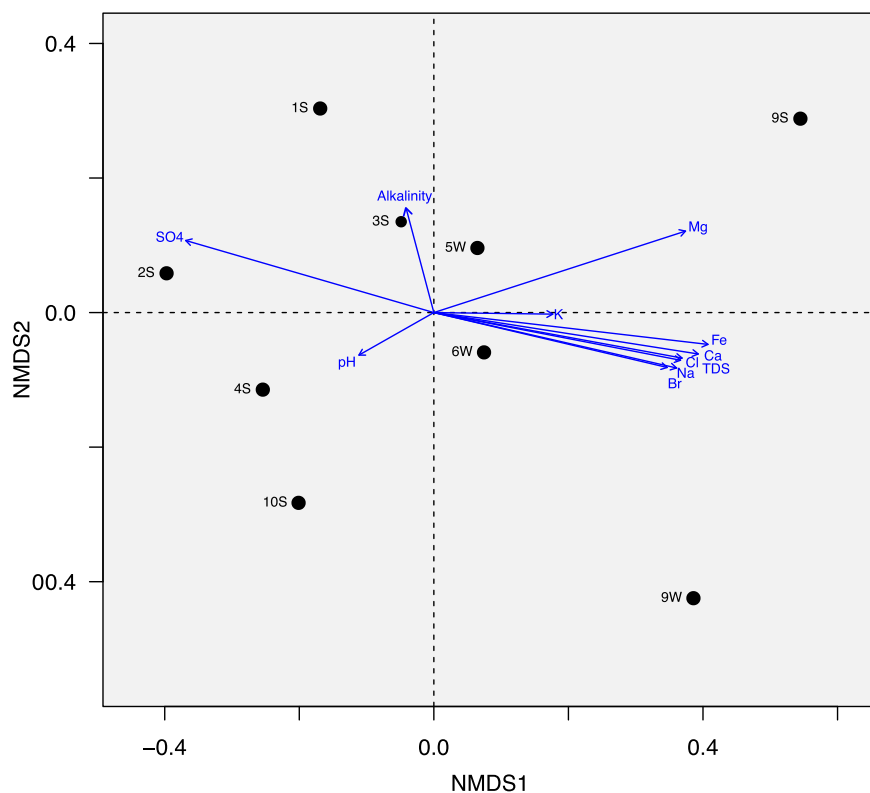


FIG 1 Nonmetric multidimensional scaling (NMDS) plot of produced water samples with a stress value of $9.83\text{E-}05$. This plot was constructed using Bray-Curtis distance calculated after sequence libraries were resampled to the depth of the sample with the fewest sequences from this experiment (1023 sequences). ANOSIM confirmed no clustering by sample origin, well age, or frac hit status and a Mantel test revealed no significant correlation between the microbial community and associated geochemical profile. Environmental vectors for geochemical parameters (Table 1) were fit onto the ordination plot, with the direction of the arrow corresponding to the direction of the gradient and the length of the vector proportional to the strength of the correlation between ordination and environmental variables. A table containing the R^2 and P -values for the corresponding vectors is located in SI Table 4, with 7 of the 11 environmental variables demonstrating a statistically significant ($P < 0.05$) correlation.

sample and was only found in one of the produced water samples, 9S, at low abundance (1.08%) (Fig. 2, Fig. S3). Our metagenomic analysis generally supported these findings, although several produced water samples are predicted to have a significantly higher abundance of archaea than that revealed by 16S rRNA gene sequencing. Specifically, metagenomic sequencing predicted a relative abundance of Archaea of up to 54% in sample 7S, 13% in sample 2S, and 30% in the 4S sample (Fig. S4).

Previous work has demonstrated that members of the Halaneroiales order are commonly found in hydraulic fracturing produced water and are often associated with thiosulfate reduction and acid production (19, 21, 22, 41). Our samples were no exception, with an average relative abundance of 33.58% Halaneroiales across all produced water samples (Fig. 2, Fig. S3 and S4). The majority of the Halaneroiales present across these samples were classified as either *Orenia* sp. or *Halanaerobium* spp. (Fig. 2). *Orenia* sp. represent 61% to 74% of the total relative abundance across samples 1S, 3S, and 5W (Fig. 2). In contrast, the majority of the Halaneroiales taxa from samples 2S, 4S, and 6W were classified as *Halanaerobium* spp., representing 20% to 44% of the total relative abundance (Fig. 2). Although both of these genera have previously been associated with hydrocarbon environments (42), *Halanaerobium* is both better characterized and more commonly associated with hydraulically fractured wells (19, 21, 22, 41).

Sulfur-reducing taxa are also frequently associated with produced water, especially those within the Desulfuromonadales and Desulfovibrionales orders (22, 35, 36, 42).

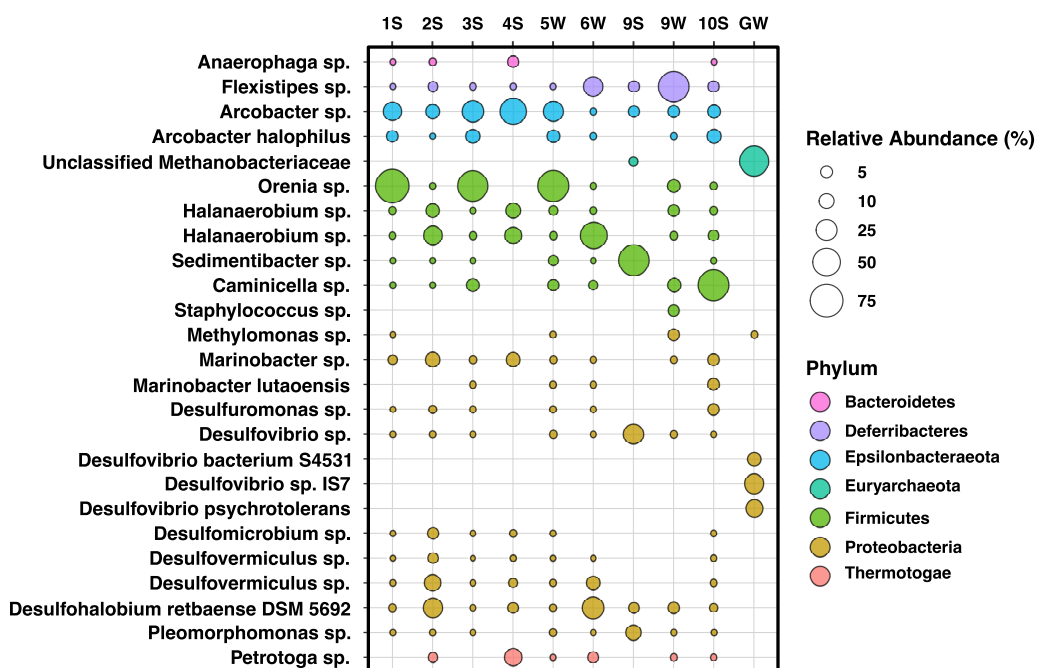


FIG 2 Bubble plot showing the relative abundance of the top 25 most abundant taxa found across the sampling sites.

Members of the Desulfuromonadales order were present across seven of the nine wells, although at a low abundance (Fig. 2, Fig. S3 and S4). Members of this order represented 1% of the total relative abundance across all produced water samples, with a maximum of 3.9% relative abundance in sample 10S (Fig. S3 and S4). In contrast, we found that members of the Desulfovibrionales order were especially enriched in samples 2S, 6W, and 9S, representing 23% to 38% of the total relative abundance (Fig. 2, Fig. S3 and S4). *Desulfotomaculum* sp. constituted the majority of the Desulfovibrionales taxa for samples 2S and 6W, at a relative abundance of 19% and 25%, respectively (Fig. 2). Sample 2S also contained a significant amount of *Desulfovermiculus* spp., at 14% of the total relative abundance (Fig. 2). Finally, the majority of the Desulfovibrionales taxa present in sample 9S were identified as members of the *Desulfovibrio* genus, at a relative abundance of 20% (Fig. 2). Interestingly, members of the Desulfovibrionales order were also highly abundant within the groundwater sample and represent 39% of the total relative abundance for that sample (Fig. 2, Fig. S3). However, the Desulfovibrionales present in the groundwater sample were identified as belonging to three distinct *Desulfovibrio* spp., none of which present in any of the produced water samples (Fig. 2).

Several members of the Campylobacterales and Deferribacterales orders were found across most produced water samples (Fig. 2, Fig. S3). Specifically, *Arcobacter* sp. was found within each produced water sample, with an average relative abundance of 12% across all samples (Fig. 2). Samples 1S, 3S, 4S, and 5W contained a particularly high amount of *Arcobacter* sp., ranging from 17% to 41% relative abundance (Fig. 2). *Arcobacter halophilus* was also present in most produced water samples, although at much lower abundances (Fig. 2). Sample 10S contained the highest amount of *Arcobacter halophilus*, with an overall relative abundance of 7% (Fig. 2). Members of the *Arcobacter* genus have previously been associated with oil and gas environments, including the Permian Basin; however, their exact metabolic role in these environments is unknown (29, 42). Similarly, members of the Deferribacterales order have previously been found in produced water samples (42). We found one member of the Deferribacterales order, *Flexistipes* sp., present across all produced water samples with an average relative abundance of 10% (Fig. 2). Notably, samples 6W and 9W contained exceptionally high amounts of *Flexistipes* sp., at a relative abundance of 19% and 61%

(Fig. 2). The Deferribacterales order is understudied and only one *Flexistipes* sp. genome has been published, from a strain isolated from the red sea (43). Thus, the ecological role of *Flexistipes* sp. in the hydrocarbon microbial community is currently unknown.

Metagenomic assembly and draft genome recovery. After quality trimming and assembling the generated metagenomic sequences, we recovered 15 metagenome assembled genomes (MAGs) from seven different samples representing 6 unique well sites (Tables S1 and S3). Completeness of the recovered MAGs ranged between 41.08% and 98.33% while contamination ranged between 1.45% and 9.83% (Table S3). Taxonomic annotation using the applied marker gene sets in CheckM and PhylopythiaS allowed assignment of the recovered MAGs to the following genera: *Alcanivorax*, *Archaeoglobus*, *Desulfohalobium*, *Geoglobus*, *Halanaerobium*, *Marinobacter*, *Methanohalophilus*, *Methanothermococcus*, *Orenia*, *Pseudomonas*, and *Ralstonia* (Table S3). Notably, several of these genera were present at high abundances with produced water samples (Fig. 2). In particular, the *Orenia* and *Halanaerobium* MAGs appear to be especially relevant, as *Orenia* sp. and *Halanaerobium* spp. constitute the majority of the microbial community for several produced water samples (Fig. 2, and Figs. S3 and S4). Understanding the functional capacity of subsurface microorganisms is especially important, as microbial activity has been linked with corrosion, fouling, and souring events in hydraulically fractured wells (24, 32–34, 44–46). These processes have been linked to metabolic potential for fermentation pathways (47), sulfur metabolism (21, 22, 48), hydrocarbon degradation (22), and biofilm formation (22) within our MAGs. We also evaluated the presence of genes involved in stress response mechanisms, as they can increase resistance to biocides used for produced water management (18).

Fermentation and methane production. In hydraulically fractured shales, subsurface microorganisms often generate energy through the fermentation of nutrients present within the shale environment, microbial produced metabolites, and/or chemical additives in injection fluid (31). While these fermentation products are beneficial for microbes, the presence of products such as acetate or ethanol can contribute to infrastructure corrosion (16, 19, 49–51). Using the DRAM product summary, we were able to confirm that all MAGs except for the *Archaeoglobus* MAG from 3S, which was only 50.25% complete, had high gene completeness over the glycolysis pathway (Fig. S5). We also used the DRAM product summary to confirm the presence of numerous genes associated with carbohydrate-active enzymes and/or short chain fatty acid (SCFA) and alcohol conversion across all MAGs (Fig. S5). We identified genes associated with mixed acid fermentation in 14 of the 15 MAGs (Fig. S6). The genes most represented across all MAGs were *ack* and *adh*, which were present in 9 and 11 of the MAGs (Fig. S6). *ack* encodes for acetate kinase, which catalyzes the phosphorylation of acetate, while *adh* encodes for an alcohol dehydrogenase, which transforms simple sugars into ethanol.

Methane is often present within shale formations. Although most of this methane is thermogenically produced, there is evidence that some of the methane is biogenically produced by methanogens (52). Methanogens have the capacity to produce methane through acetoclastic, hydrogenotrophic, or methylotrophic methanogenesis. Three of the 17 recovered MAGs were identified as methanogens; two as *Methanohalophilus* and one as *Methanothermococcus*. As expected, the DRAM product summary confirmed a high completeness for genes associated with methanogenesis in these MAGs (Fig. S5). Two of these MAGs contain the genes for the alpha, beta, and gamma subunits of methyl coenzyme M reductase (*mcr*), which catalyzes the final reaction in the methanogenesis pathway and is only found in methanogens (Fig. S6). The *Methanohalophilus* MAG from sample 2 did not contain any of the *mcr* genes, although this is likely because of its completion rate (76.2%) (Fig. S6).

Although only methanogens contain the *mcr* gene, many organisms are known to contain genes related to methane production. Certain members of the *Archaeoglobus* genus have been reported to contain genes associated with methanogenesis (53, 54). Consistent with this, the *Archaeoglobus* MAG recovered from produced water sample

4S contained three genes associated with acetoclastic methanogenesis: *cdh*, which encodes carbon monoxide dehydrogenase; and *hdrA*, which encodes for the coenzyme required by *mcr* to produce methane (Fig. S6). The *Archaeoglobus* MAG recovered from sample 4 also contained four genes associated with hydrogenotrophic methanogenesis: *ptr*, *mch*, *mtd*, and *mer* (Fig. S6). These genes encode for the enzymes responsible for converting carbon dioxide to 5-methyl-tetrahydromethanopterin, which is an important metabolic intermediate in both the acetoclastic and hydrogenotrophic methanogenesis pathway (Fig. S6).

Sulfate and thiosulfate reduction. The presence of excess sulfide in hydraulically fractured wells can lead to gas souring. Sulfide is produced by sulfate-reducing microorganisms (SRM) present in the produced water. Out of our 15 recovered MAGS, four are SRM: two *Archaeoglobus* and two *Desulfohalobium* (Fig. S6). SRM commonly reduces sulfate via the dissimilatory sulfate reduction pathway. Consistent with this, all four SRM contained genes associated with classical sulfate reduction including genes that encode for the sulfite reduction-associated complex (*DsrMKJOP*) and for dissimilatory sulfite reductase (*dsrAB*) (Fig. S6). Interestingly, the *Methanohalophilus* MAG recovered from 2S contained genes that encode for *DsrMKJOP* and *dsrAB*, although the *Methanohalophilus* MAG from samples 7S did not (Fig. S6). Although *Methanohalophilus* is not a typical sulfate reducer, recent work demonstrates that *Methanohalophilus* in hypersaline habitats have adapted to have unique sulfur metabolisms, including the genes that encode for *dsrAB* (55). The *Methanohalophilus* MAG from sample 2S may be another example of an archaeal strain uniquely adapted to a high saline environment.

Recent work has shown that in some hydrocarbon environments, sulfide production primarily occurs through thiosulfate reduction, rather than through classical sulfate reduction (16, 19, 21, 22, 56). *Halanaerobium* often dominates produced water samples and has been increasingly linked with thiosulfate-dependent sulfidogenesis (16, 19, 21, 22, 56). We recovered a *Halanaerobium* MAG with 73.66% completion from sample 3S. This MAG contains two sulfur transport genes as well as a rhodanese gene, although we did not detect the presence of any anaerobic sulfite reductase genes (*asrA*, *asrB*, *asrC*) (Fig. S6). Many of our other MAGs also contained genes associated with thiosulfate reduction. In total, 8 MAGs contain sulfur transport genes; 9 contain rhodanese or rhodanese-like genes (*rdl*), which are associated with the conversion of thiosulfate to adenylyl sulfate; 2 contain *aprAB*, which encodes for the protein responsible for converting thiosulfate to sulfite; and 3 contain at least one of the *asr* genes, which encodes for the protein that converts sulfite into sulfide (Fig. S6). Among the MAGs recovered from produced water samples 1S and 3S, there was at least one organism that contained the *mpsT*, *asr*, and/or *dsr* gene (Fig. S6). Similarly, among the MAGs recovered from produced water samples 2S, 4S, and 10S there was at least one organism that contained the *mpsT* and/or *dsr* gene (Fig. S6). Thus, the microbial community present in these sites are known to have the functional capacity to convert thiosulfate into sulfide.

Hydrocarbon degradation. Crude oil and gas contain complex mixtures of hydrocarbons, thus we were interested in examining hydrocarbon degradation capabilities present in our MAGs. We searched for the presence of 26 common hydrocarbon degradation genes associated with toluene degradation, biphenyl degradation, naphthalene degradation, catechol ortho-cleavage, and catechol meta-cleavage (Fig. S6). Overall, we found minimal presence of major hydrocarbon degradation genes in our 15 MAGS. No MAGs contained any of the selected genes associated with toluene, biphenyl, or benzoate degradation (Fig. S6). However, our *Marinobacter* MAG contained the gene for 2-hydroxychromene-2-carboxylate isomerase, which is associated with naphthalene degradation (Fig. S6). Nine of our MAGs contained one gene associated with either catechol ortho- or meta-cleavage (Fig. S6). The *Pseudomonas* MAG had two genes associated with catechol ortho-cleavage, 3-oxoadipate CoA-transferase and acetyl-CoA C-acyltransferase, and one gene associated with catechol meta-cleavage, 2-hydroxy-6-oxoheptanoate-6-semialdehyde hydrolase (Fig. S6). Overall, the *Pseudomonas* and

Marinobacter MAGs contained the most hydrocarbon degradation-associated genes, at 3 and 2 genes total (Fig. S6).

Motility and biofilm formation. Biofilm formation during or after hydraulic fracturing can lead to major operational challenges (23, 57). Biofilm growth begins when microbes adhere to one another and/or the well infrastructure. As the biofilm grows, microbes produce extracellular polymeric substances (EPS), which aid in cell attachment and can trap sulfate and other microbial products within the biofilm or between the biofilm and the adjacent well infrastructure, leading to clogging and/or increased rates of microbial induced corrosion. Established biofilms are often dense, which makes it difficult to control them with biocidal application. We examined our MAGs for flagellin and motility genes, which are thought to be important for cellular attachment and biofilm formation (45). Most recovered MAGs contained the *flg*, *fli*, *flh*, and/or *motAB* genes (Fig. S6). However, the *Geoglobus*, *Methanothermococcus*, and one of the *Methanohalophilus* MAGs did not (Fig. S6). We also found three MAGs contained the genes encoding sporulation two-component response regulator Spo0A, which is associated with surface attachment initiation, and numerous MAGs that contained the genes encoding for a glycosyl transferase group 2 family protein (*glt2*) and/or diguanylate cyclase (*adrA*), which are associated with exopolysaccharide production (58, 59). This suggests that, under the right circumstances, there is a high potential for biofilm formation within our sample sites.

Stress response. Produced water from hydraulically fractured wells is characterized by high salt concentrations and the presence of heavy metals (60–62). Thus, microbes in produced water employ stress response mechanisms to protect the cell. Halophilic microorganisms have two strategies which assist in maintaining proper osmotic pressure within the cell (63). Organisms utilizing the “salt-in” strategy accumulate molar concentrations of potassium and chloride while organisms employing the “osmolyte” strategy accumulate organic osmotic solutes (63). All recovered MAGs contained the *trkA*, *trkH*, and/or *ktrB* genes, which are associated with potassium transport and uptake (Fig. 3). Similarly, 8 recovered MAGs contained the *proX*, *proW*, and/or *opuA* genes, which are associated with glycine and betaine uptake (Fig. S6).

The presence of heavy metals can induce oxidative stress responses in nearby microorganisms. Therefore, we examined our recovered MAGs for: *perR*, a redox sensitive transcriptional regulator; *sor*, which encodes for a superoxide reductase; and a gene encoding for rubredoxin. We also searched for more general stress response genes including a periplasmic stress response gene (*ompH*), a universal stress protein (*uspA*), a heat shock protein (*grpE*), and heat shock chaperones (*groES* and *groEL*). Unsurprisingly, all our recovered MAGs contained at least one of these genes and many recovered MAGs contain multiple of these genes, including the *Halanaerobium*, *Orenia*, *Desulfohalobium*, and *Marinobacter* MAGs.

Finally, hydraulically fractured wells are often treated with biocides in an attempt to control or reduce the growth of detrimental microbes. We examined our recovered MAGs for a wide variety of genes associated with antibiotic and biocidal resistance, including multidrug resistance transporters and multidrug resistance proteins (Fig. S6). All our MAGs except for the *Archaeoglobus* from site 3S (50.25% complete) contained at least one of these genes (Fig. S6). Interestingly, the *Marinobacter* MAG contained 8 genes including several drug resistance transporters; acriflavin, bacteriocin, and polymyxin resistance proteins; and several heavy metal resistance genes (Fig. S6).

DISCUSSION

The development of hydraulic fracturing technology has led to the extraction of oil and gas from previously impermeable shale formations. Currently, the Permian Basin is the largest oil producing basin in the United States with newly fractured wells having the highest oil production rate. Despite the rapid development in the Permian region, little has been done to examine the microbial life in the Permian Basin. In this study, we address this knowledge gap by using geochemical and microbiological techniques

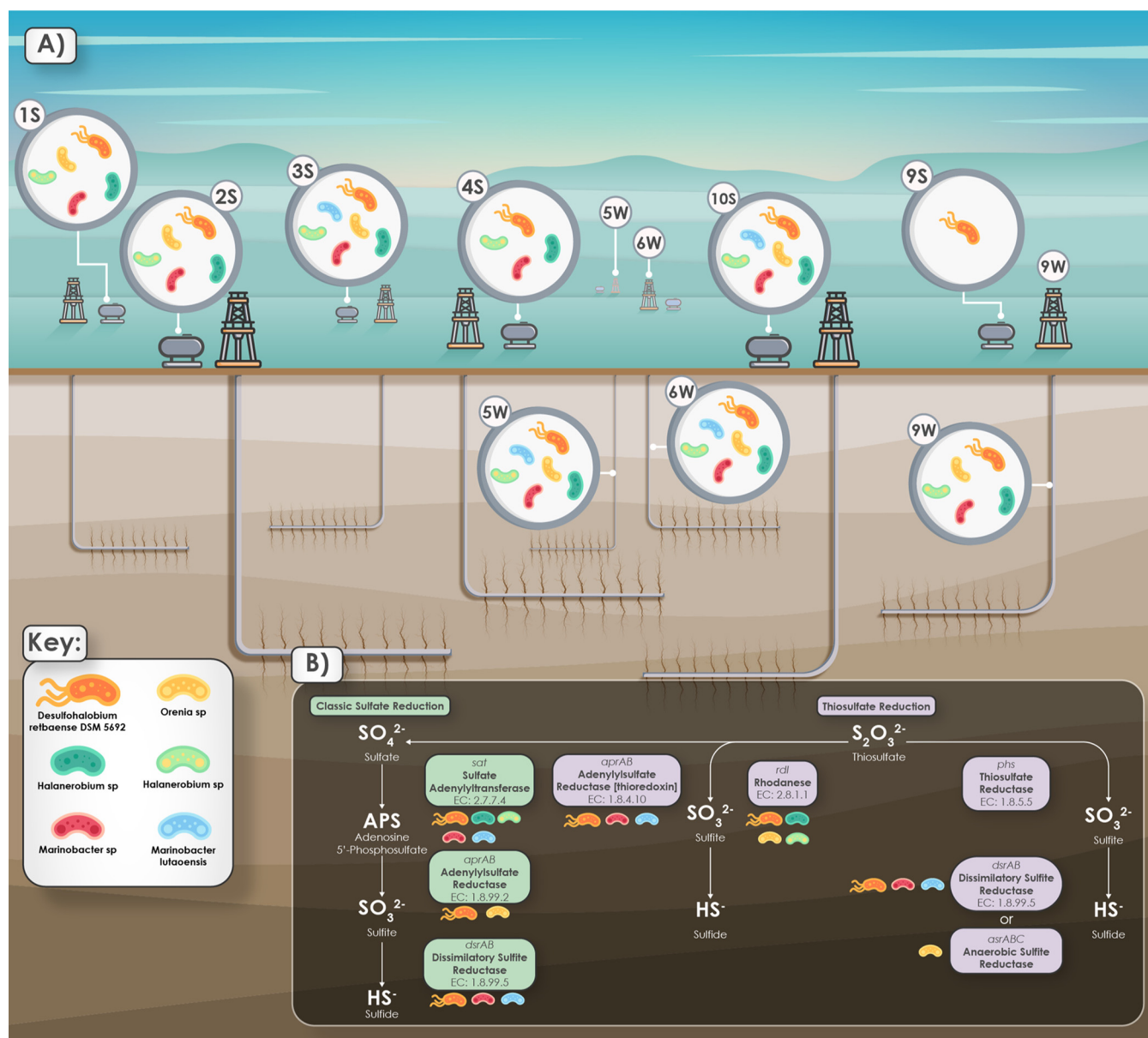


FIG 3 A, Conceptual figure demonstrating possible spatial relationships present in subsurface environments in the Permian Basin. Although the location and spatial relationships are hypothetical, the icons indicate the presence or absence of selected microorganisms with a MAG based on 16S rRNA gene sequencing data available for shown sites. B, Summary of sulfate metabolic genes present within each of the selected organisms based on the available MAG data in SI Fig. 6

to analyze produced water from 14 samples taken at 10 unique sites in the Permian region of the Midland Basin. We also recovered 15 MAGs from seven different samples representing 6 unique well sites (Table S1 and S3) in order to connect functional capabilities with specific microorganisms present within Permian Basin produced water samples.

Hydraulically fractured wells are generally characterized by high concentrations of salt, metals, and organics, with TDS concentrations as high as 345,000 mg/L (7–9). Our geochemical analysis of produced water from the Permian Basin supported these trends. We found excess Fe^{2+} , Ba^{2+} , and Mg^{2+} in these systems, which is consistent with previous work (22, 33–37). Interestingly, we also found lower TDS values in the Midland Basin compared to those in the Bakken, Marcellus, or Barnett Shales (64, 65), with sample values ranging from 53,200 to 123,000 mg/L. We found that the produced water from our sample sites clustered into two groups: those with high TDS and low

dissolved sulfate or low TDS and high dissolved sulfate, compared to the median values of 107,000 mg/L and 1,110 mg/L. Notably, even samples with low dissolved sulfate concentrations had higher values than those found within other basins, which had concentrations <200 mg/L (65). Produced water samples with low TDS and high dissolved sulfate concentrations had geochemical profiles which resembled the ground-water sample.

When a new well is drilled, fracturing fluid is injected into the subsurface and then a shut-in period occurs where the fracturing fluid remains pressurized in the hydraulically fractured well. After the shut-in period, the flow-back period begins and a large volume of water, which is chemically similar to the injection fluid, is generated. Over time, as flow-back ends and hydrocarbon production begins the chemical composition of the fluid changes to higher salinity and is referred to as produced fluid. Previous work by Mouser et al. (32) demonstrates that this process generally occurs between 3 and 6 months after the shut-in period ends. Although groundwater is commonly used as a major component of fracturing fluid in this region, because all of the wells we sampled were ≥ 6 months old, it would be highly unlikely that the produced fluid resembled the groundwater unless there was outside interference (i.e., a frac hit). Thus, this provides evidence that injection water from newly fractured wells floods nearby older wells in the Permian Basin. This finding represents, to our knowledge, the first documentation of mature wells having a change in geochemistry due to communication from an outside well. This geochemical evidence was further supported by local operators, who reported these wells as frac hits (Table S2). The sulfate concentrations in produced fluids with reduced TDS measurements were higher than typical basin produced water sulfate levels of <200 mg/L (65), suggesting that the Permian region may have a higher potential for sulfate reduction than other regions. Furthermore, as the potential fracture fluid source and the wells impacted by communication to newly drilled wells appeared to have higher sulfate levels, this suggests any frac hit could increase microbial sulfate reduction risk even in established wells.

Well-to-well communication events were self-reported by well operators; they do not take into account the amount of time since the initial frac hit event and were treated as single events in our categorical analysis. It is likely that there is repeated, ongoing well-to-well communication that is dependent on numerous factors including the productivity of each well and the unique shale fractures present in that geological formation. Because of these challenges, it is difficult to accurately determine if frac hits are associated with changes in the microbial community. However, we measured two correlations that suggest frac hits contribute to the structure of the microbial community. When we examined all of the microbial data, we measured statistically significant clustering by frac hit status, but not sample origin or well age. Additionally, 7 of the 11 geochemical measurements had a statistically significant correlation. Two of these values include TDS and sulfate concentration, which is notable as frac hits have a unique geochemical signature characterized by low TDS and high sulfate concentration.

Network analysis is a common tool used to identify and visualize microbial interactions occurring in specific environments. Previously, our group completed a network analysis using data from three unconventional hydrocarbon reservoirs: the Marcellus Shale, Bakken Formation, and Permian Basin (65). This analysis relied upon previously published data from our research group (22, 35, 36, 38) as well as the data presented in this manuscript. Keystone species are highly interactive species that have a disproportionately large effect in the environment. Two of the keystone species we identified across all three basins were *Marinobacter* and *Desulfohalobium retbaense* DSM 5692 (65). This is notable as three of the recovered, highly abundant MAGs in the Permian Basin included one *Marinobacter* and two *Desulfohalobium*. We found that *Marinobacter* sp. was present at low abundances (<10%) in all of our produced water samples except 9S (Fig. 2). *Marinobacter* has previously been associated with hydrocarbon environments and is often used in bioremediation efforts (66). *Marinobacter* has been shown to consume hydrocarbons and use them to produce and excrete EPS, which work as an effective natural surfactant (66). Although

there were minimal hydrocarbon degradation genes in our MAG, it contained numerous genes associated with biofilm formation (Fig. S6). *Desulfohalobium retbaense* DSM 5692 was also present in all of our samples, at up to 25% relative abundance (Fig. 2). Our *Desulfohalobium* MAGs contained genes associated with classical sulfate reduction, including genes that encode for the sulfite reduction-associated complex (*DsrMKJOP*) and for dissimilatory sulfite reductase (*dsrAB*). Although significant additional investigation is required, it seems likely that *Desulfohalobium* and *Marinobacter* participate in cross-feeding with each other and/or other members of the microbial community within the Permian Basin. Cross-feeding is when one species of microorganism utilizes the metabolic products of another. Cross-feeding has previously been demonstrated among *Halanaerobium*, *Geotoga*, and *Methanohalophilus* present in the Marcellus and Utica Basins (67), although it has not been investigated in the Permian Basin. In this case, *Desulfohalobium* and *Marinobacter* may be uniquely suited to utilize sulfate and sulfate-derivatives for anaerobic respiration and biosynthesis, respectively.

Established wells typically contained lower diversity microbial communities dominated by microbiota adapted to the unique reservoir conditions (32). Consistent with this, we found that our produced water samples, which were all obtained from wells ≥ 6 months old, were dominated by a smaller number of halophilic, thermophilic, sulfate-reducing, fermentative, and methanogenic taxa (Fig. 2). Interestingly, six of our 15 recovered MAGs were found to be within the top 25 most abundant taxa found (Fig. 2). These MAGs include one *Halanaerobium*, two *Orenia*, one *Marinobacter*, and two *Desulfohalobium*. Recent work has shown that in some hydrocarbon environments there is a high abundance of members of the Halanerobiales order (19, 21, 22, 41). Halanerobiales have the ability to produce sulfide through thiosulfate reduction, rather than classical sulfate reduction, and are thought to be a major contributor to microbial induced corrosion (19, 21, 22, 41). Three of our highly abundant MAGs, *Halanaerobium* and *Orenia*, belong to this order. *Halanaerobium* has previously been implicated in thiosulfate-reduction in hydraulically fractured wells (16, 21). Although *Orenia* has previously been associated with hydrocarbon environments (42), to our knowledge it has never been found in such high abundances in hydraulically fractured wells. All three of these MAGs contain genes associated with thiosulfate reduction (Fig. 3), suggesting these taxa also contribute to thiosulfate reduction in the Permian Basin. Because these organisms are closely related, it is highly possible that well-to-well communication events would foster horizontal gene transfer or introduce competing strains of bacteria (Fig. 3).

Each time a frac hit event occurs, microbial biomass is potentially exchanged between the newly fractured well and the existing well. Previous research demonstrates that newly fractured wells contain a diverse assemblage of microbes that converge over time to a low diversity microbial community dominated by halophiles uniquely adapted to the reservoir conditions (32). Established wells are often routinely treated with biocides in order to minimize microbial induced biofilm and corrosion (68, 69), therefore these unique adaptations may include biocidal resistant genes. Thus, a frac hit can act as a seeding event where highly adapted, biocidal resistant halophiles from established wells are introduced to newly fractured wells. A frac hit also acts as a disruption event for the established well, which could impact the resident microbial community composition or activity. Additionally, this may offer the opportunity for horizontal gene transfer. Although well-to-well communication is a known problem within the oil and gas industry (70), there is little published evidence investigating the impact of frac hits or other well-to-well communication events. Thus, it is challenging to know the likelihood of such scenarios. However, our microbial analysis suggests that frac hits drive the structure and functional potential of the microbial community.

This work is an important investigation into the microbial ecology of unconventional oil and gas wells. It is one of the first studies to examine the microbiology of the Permian Basin and is the first study within this region to use integrated geochemical and microbial data sets. Well-to-well communication is a known problem in the oil and gas industry. However, to our knowledge, this represents the first study to provide

geochemical and microbial evidence demonstrating occurrence of frac hits. Additionally, in this study we utilize our microbial data to provide a conceptual framework for understanding how well-to-well communication may impact in the subsurface. The Permian Basin is the largest producing oil and gas region in the United States and continues to grow. With this continued growth, frac hits will undoubtedly become more frequent. Future work should be conducted in order to investigate the impact of frac hits, on both a short and long time scale. Understanding how microbial communities are impacted by well-to-well communication will allow us to reduce risks that lead to biocidal resistance or increased hydrogen sulfide production.

MATERIALS AND METHODS

Sampling. Produced water samples were collected from 10 actively producing hydraulically fractured oil and gas wells located in the Wolfcamp Formation of the Permian Basin in Texas in April 2018 (Table S1). Proximal groundwater is typically used as fracture fluid during the hydraulic fracturing process; therefore, we also collected a groundwater sample. Our sampling method was consistent with those previously reported (8, 17, 22, 23, 35, 36, 38, 44, 71). In brief, samples were collected either from a sampling port at the wellhead (if available) or the three-phase separators. The function of the separator is to separate the hydrocarbons (oil and/or gas) from the remaining fluids (produced water). In several instances sampling from the wellhead was not possible due to missing sampling ports or high gas pressures. In those cases, the separators represented the closest available sampling location. We used the GA5000 Gas Analyzer (QED Environmental Systems, Dexter, MI) to measure the gas content produced at each wellhead, although this was not possible at the separators. Produced water samples were collected in unused plastic carboys that were pre-rinsed with sample waters and allowed between 1 and 3 h to settle into distinguishable oil and water phases in the sealed container. A portion of the water phase was collected in a sterile 1 L Nalgene bottle and immediately placed on dry ice in order to preserve the sample for microbiological analysis. Upon arrival in the lab, the frozen produced water samples were stored at -20°C until processing. The remaining water phase was passed through glass wool cartridges under gravity and then a $0.45\ \mu\text{m}$ inline filter (EnviroTech, Salt Lake City, UT) in a closed loop in order to minimize oxygen exposure. A portion of this filtered water was collected on-site and measured for pH using a Horiba multimeter (Horiba, Edison, New Jersey) and titrated for alkalinity using a Hach digital titrator (Hach, Loveland, CO). A 30-mL volume of the filtered water was acidified to 2% by volume using nitric acid and transported on ice for inductively coupled plasma optical emission spectroscopy (ICP-OES). A 15-mL volume of the filtered water was passed through a second $0.22\ \mu\text{m}$ sterile polyethersulfone filter (Millipore, Inc.) and shipped on ice for ion chromatography (IC) measurement. Finally, filtered water samples were also collected with zero-headspace in Shimadzu TOC vials and transported on ice for total organic carbon (TOC) analysis.

Chemical Analysis. Chemical analyses were measured as previously described (36). In brief, major cations and anions were detected in triplicate using ion chromatography (IC) on a ThermoFisher (ThermoFisher, Waltham, MA) ICS-5000+ with AS11-HC column for anion quantification and CS16 column for cation quantification. TDS was calculated by summing the total cations and anions in each sample. Trace metals were analyzed using U.S. EPA Method 6010D with inductively coupled plasma optical emission spectroscopy (ICP-OES) on an Optima 7300 DV (Perkin Elmer, Waltham, MA). Finally, 3-5 total organic carbon (TOC) replicates were analyzed with a Shimadzu TOC Analyzer (Model: TOC-LCSN) for non-purgeable organic carbon (NPOC) and total inorganic carbon (TIC).

DNA Extraction. Frozen produced water samples were thawed at 5°C before filtering through a $0.2\ \mu\text{m}$ polyethersulfone membrane filter (Qiagen, Hilden, Germany) in order to collect microbial biomass. For samples that could not be filtered (Table S2), biomass was pelleted by the centrifugation at $5000 \times g$. The collected biomass (from filter or centrifugation) was used for DNA extraction. To optimize DNA recovery, we utilized two different extraction kits (Table S2). We initially extracted samples using a previously described (36) modified version of the DNeasy Powersoil kit (Qiagen, Hilden, Germany) with a 60 min lysozyme digestion and 30 min of bead beating. If we were unable to obtain a sufficient amount of DNA using this protocol, we completed a second extraction using the DNA/RNA All Prep kit (Qiagen, Hilden, Germany) with 10 min of homogenization at maximum speed after the addition of the cell lysis buffer. For both methods, DNA was eluted from spin filters in $80\ \mu\text{L}$ nuclease free water. Kit blanks were also concurrently extracted to confirm that no contamination occurred.

Library preparation and sequencing. Recovered DNA from both extraction methods was amplified using universal primers targeting the V4 region of the 16S rRNA gene, as previously described (72, 73). PCR products from both extraction methods were combined and cleaned using AMPure XP beads (Beckman Coulter, Pasadena, CA), visualized on a 1% agarose gel, and quantified using Qubit (Life Technologies, Carlsbad, CA). Negative controls were also amplified, visualized, and quantified in order to confirm that no contamination occurred. Purified 16S rRNA gene libraries were pooled (with 6-8 replicates per sample), diluted to a concentration of 2 nM, and denatured using fresh 0.2 M NaOH. Libraries were further diluted according to the manufacturer's instructions and sequenced on an Illumina MiSeq (Illumina, San Diego, CA) using a 300 cycle V2 Nano kit. For several samples, extraction yielded enough DNA for metagenomic library preparation (Table S1). We used the Nextera XT library preparation kit (Illumina, San Diego, CA), which is specifically designed for low biomass samples and only required 1 ng of input template. Briefly, 1 ng of input sample DNA was tagged with Illumina primers containing

sequencing adapters and barcodes in a 12 cycle PCR step. PCR products were cleaned up using AMPure XP beads (Life Technologies, Carlsbad, CA). DNA libraries were normalized by pooling, quantified using Qubit (Life Technologies, Carlsbad, CA), and diluted to a concentration of 20–40 pM. The DNA library was denatured by heating the sample at 96°C for 2 min, cooled in an ice bath for 5 min, and sequenced using a 600 cycle V3 kit on an Illumina MiSeq sequencer (Illumina, San Diego, CA).

Produced water samples are often characterized by low biomass and the presence of various inhibitor substances (34–36, 74), which make it challenging to extract sufficient amounts of high-quality DNA and prepare sequencing libraries. In this study, 16S rRNA gene libraries were successfully prepared and sequenced from the groundwater sample and produced waters from 9 of the 14 unique sample locations (Table S1). In addition, we were able to perform duplicate extractions and amplifications for samples 1S, 4S, and 10S. For each produced water sample, the unique number denotes the collection site, and the letter indicates the sample origin (S for separator or W for wellhead) (Table S2). Illumina MiSeq sequencing generated a total of 1,261,357 sequences, with 296,651 remaining after quality control filtering (Table S2). Samples obtained from sites 6S, 7S, 7W, 8S, and 8W did not generate enough sequencing data for downstream analysis and thus were not included going forward (Tables S1 and S2). Across the remaining samples, 16S rRNA gene sequencing generated between 1,023 and 42,008 reads per sample with an average depth of 22,813 sequences per sample. Metagenomic libraries were prepared from produced water samples for which a sufficient amount and quality of DNA could be extracted (Tables S1 and S3). We successfully prepared metagenomic libraries from samples representing 8 of the 14 unique produced water sample locations, however, we were unable to generate metagenomic sequencing data for sites 5W, 6S, 7W, 8S, 8W, 9W, or the groundwater sample (Table S1). A total of 54,352,861 metagenomic reads were obtained, with an average sample library size of 6,039,206 reads after quality control filtering (Table S3).

16S rRNA gene sequencing data analysis. 16S rRNA gene sequences were analyzed using QIIME2 version 2019.10 (75) as previously described (36). In brief, sequences were imported as EMPSingleEndSequences, demultiplexed using the demux emp-single command, and processed for quality control using DADA2 (76) with the default settings and a truncation length of 250 bp. The classify-sklearn (77) command was used to classify representative sequences identified through DADA2 using a pre-trained Naive Bayes classifier trained on Silva 132 99% OTUs (78, 79) from the 515F/806R region. Finally, any sequences identified as chloroplast or mitochondria were filtered from the data.

Data generated by Qiime2 was imported into R for further analysis (80). The number of observed ASVs and Pielou's evenness were calculated in Base R (80). The Vegan package (81) was used to calculate the Chao1 index, Shannon index, and Bray–Curtis dissimilarity values. Vegan (81) was also used to calculate Bray–Curtis dissimilarity values, complete non-metric multidimensional scaling (NMDS) analyses, fit environmental parameters onto the generated ordination plot, conduct the Analysis of Similarities (ANOSIM), and complete the Mantel test. All NMDS plots were constructed with $k = 4$ dimensions. NMDS, ANOSIM, and the Mantel test were completed using Bray–Curtis distance measurements calculated after sequencing libraries were subsampled to the lowest sequence depth (1023) represented across all samples (SI Table 2). The Mantel test also relied on the Euclidean distance of the geochemical data, which was calculated in Base R. The statistical significance of the ANOSIM, the Mantel test, and the environmental vectors was based on 999 permutations of the grouping, geochemical, and environmental data, respectively. Finally, we also used the RColorBrewer package to generate colorblind accessible palettes when necessary (82).

Metagenomic data analysis. Metagenomic reads were quality trimmed (Q20, 70 bp minimum length) using CLC Genomics (Qiagen, Hilden, Germany). Taxonomy of metagenomic reads was assessed using Kaiju (83) with default parameters while 16S rRNA genes across metagenomic reads were predicted using Phyloflash (84). Metagenomic reads were then assembled into continuous sequences, or contigs, using Spades version 3.11 utilizing the metaSpades workflow (85). Assembled metagenomic contigs were binned using different binning software (Maxbin2, Vizbin and Patrick) based on tetranucleotide frequency, differential coverage, and/or marker genes (86–88) (Table S3). Final bins are referred to as metagenome assembled genomes (MAGs) and completeness and contamination of MAGs was assessed at the genus level or the next lowest available taxonomic marker set using the CheckM taxonomy_wf workflow (89). Taxonomic thresholds of bins were determined via Tetra Correlation Search (TCS) in JSpeciesWS (90). We used RNAmmer (91) to identify rRNA genes present in our MAGs and used BLAST (92) to confirm all identified genes were consistent with the taxonomy of our MAG. Recovered bins were annotated with the RASTk (Rapid Annotations using Subsystems Technology toolkit) pipeline using the SEED database (93–95). Relevant genes associated with subsurface hydrocarbon storage were identified using a detailed literature search. In instances where specific genes or pathways are minimally discussed in the literature, we also utilized the KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways database (96–98) to identify genes and pathways of interest. MAGs were also analyzed in KBase (99) with the DRAM module (100).

Data availability. The 16S rRNA gene sequences, metagenomes, and MAGs generated from this experiment were submitted to the NCBI Sequence Read Archive and are available under BioProject number [PRJNA726570](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA726570).

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 1.1 MB.

ACKNOWLEDGMENTS

We thank Randall Burt Thomas for his role as water bottle procurement specialist as well as for his assistance in sample collection and preservation.

This project was funded by the United States Department of Energy, National Energy Technology Laboratory, in part, through a site support contract. Neither the United States Government nor any agency thereof, nor any of their employees, nor the support contractor, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

REFERENCES

1. U.S. Energy Information Administration (EIA). <https://www.eia.gov/analysis/studies/usshalegas/>. Retrieved 11 May 2020.
2. Plata DL, Jackson RB, Vengosh A, Mouser PJ. 2019. More than a decade of hydraulic fracturing and horizontal drilling research. *Environ Sci Process Impacts* 21:193–194. <https://doi.org/10.1039/c9em90004g>.
3. Johnsen R, LaRiviere J, Wolff H. 2016. Estimating Indirect Benefits: Fracking, Coal and Air Pollution. Institute for the Study of Labor (IZA) Discussion Paper No 10170 51.
4. Johnsen R, LaRiviere J, Wolff H. 2019. Fracking, Coal, and Air Quality. *J the Association of Environmental and Resource Economists* 6:1001–1037. <https://doi.org/10.1086/704888>.
5. Wang Q, Chen X, Jha AN, Rogers H. 2014. Natural gas from shale formation – the evolution, evidences and challenges of shale gas revolution in United States. *Renewable and Sustainable Energy Rev* 30:1–28. <https://doi.org/10.1016/j.rser.2013.08.065>.
6. Gregory KB, Vidic RD, Dzombak DA. 2011. Water Management Challenges Associated with the Production of Shale Gas by Hydraulic Fracturing. *Elements* 7:181–186. <https://doi.org/10.2113/gselements.7.3.181>.
7. Barbot E, Vidic NS, Gregory KB, Vidic RD. 2013. Spatial and temporal correlation of water quality parameters of produced waters from devonian-age shale following hydraulic fracturing. *Environ Sci Technol* 47:2562–2569. <https://doi.org/10.1021/es304638h>.
8. Murali Mohan A, Hartsock A, Bibby KJ, Hammack RW, Vidic RD, Gregory KB. 2013. Microbial community changes in hydraulic fracturing fluids and produced water from shale gas extraction. *Environ Sci Technol* 47:13141–13150. <https://doi.org/10.1021/es402928b>.
9. Akyon B, McLaughlin M, Hernández F, Blotevogel J, Bibby K. 2019. Characterization and biological removal of organic compounds from hydraulic fracturing produced water. *Environ Sci Process Impacts* 21:279–290. <https://doi.org/10.1039/c8em00354h>.
10. Clark CE, Veil JA. 2009. Produced water volumes and management practices in the United States.
11. Boschee P. 2014. Produced and Flowback Water Recycling and Reuse: Economics, Limitations, and Technology. *Oil and Gas Facilities* 3:16–21. <https://doi.org/10.2118/0214-0016-OGF>.
12. Arthur JD, Bohm B, Coughlin BJ. 2008. Hydraulic fracturing considerations for natural gas wells of the Marcellus Shale. Cincinnati, Ohio.
13. Vidic RD, Brantley SL, Vandenbossche JM, Yoxheimer D, Abad JD. 2013. Impact of shale gas development on regional water quality. *Science* 340:1235009–1235009. <https://doi.org/10.1126/science.1235009>.
14. Estrada JM, Bhamidimarri R. 2016. A review of the issues and treatment options for wastewater from shale gas extraction by hydraulic fracturing. *Fuel* 182:292–303. <https://doi.org/10.1016/j.fuel.2016.05.051>.
15. Rosa L, Rulli MC, Davis KF, D'Odorico P. 2018. The water-energy nexus of hydraulic fracturing: a global hydrologic analysis for shale oil and gas extraction. *Earth's Future* 6:745–756. <https://doi.org/10.1002/2018EF000809>.
16. Liang R, Grizzle RS, Duncan KE, McInerney MJ, Sufita JM. 2014. Roles of thermophilic thiosulfate-reducing bacteria and methanogenic archaea in the biocorrosion of oil pipelines. *Front Microbiol* 5:89. <https://doi.org/10.3389/fmicb.2014.00089>.
17. Mohan AM, Bibby KJ, Lipus D, Hammack RW, Gregory KB. 2014. The functional potential of microbial communities in hydraulic fracturing source water and produced water from natural gas extraction characterized by metagenomic sequencing. *PLoS One* 9:e107682. <https://doi.org/10.1371/journal.pone.0107682>.
18. Vikram A, Lipus D, Bibby K. 2014. Produced water exposure alters bacterial response to biocides. *Environ Sci Technol* 48:13001–13009. <https://doi.org/10.1021/es5036915>.
19. Liang R, Davidova IA, Marks CR, Stamps BW, Harriman BH, Stevenson BS, Duncan KE, Sufita JM. 2016. Metabolic capability of a predominant Halanaerobium sp. in hydraulically fractured gas wells and its implication in pipeline corrosion. *Front Microbiol* 7.
20. An BA, Shen Y, Voordouw G. 2017. Control of sulfide production in high salinity Bakken Shale oil reservoirs by halophilic bacteria reducing nitrate to nitrite. *Front Microbiol* 8:1164. <https://doi.org/10.3389/fmicb.2017.01164>.
21. Booker AE, Borton MA, Daly RA, Welch SA, Nicora CD, Hoyt DW, Wilson T, Purvine SO, Wolfe RA, Sharma S, Mouser PJ, Cole DR, Lipton MS, Wrighton KC, Wilkins MJ. 2017. Sulfide generation by dominant Halanaerobium microorganisms in hydraulically fractured shales. *mSphere* 2:e00257-17. <https://doi.org/10.1128/mSphereDirect.00257-17>.
22. Lipus D, Vikram A, Ross D, Bain D, Gulliver D, Hammack R, Bibby K. 2017. Predominance and metabolic potential of Halanaerobium spp. in produced water from hydraulically fractured Marcellus Shale Wells. *Appl Environ Microbiol* 83:e02659-16. <https://doi.org/10.1128/AEM.02659-16>.
23. Struchtemeyer CG, Elshahed MS. 2012. Bacterial communities associated with hydraulic fracturing fluids in thermogenic natural gas wells in North Central Texas, USA. *FEMS Microbiol Ecol* 81:13–25. <https://doi.org/10.1111/j.1574-6941.2011.01196.x>.
24. Vikram A, Lipus D, Bibby K. 2016. Metatranscriptome analysis of active microbial communities in produced water samples from the Marcellus Shale. *Microb Ecol* 72:571–581. <https://doi.org/10.1007/s00248-016-0811-z>.
25. Lipus D, Vikram A, Gulliver D, Bibby K. 2019. Upregulation of peroxide scavenging enzymes and multidrug efflux proteins highlight an active sodium hypochlorite response in *Pseudomonas fluorescens* biofilms. *Biofouling* 35:329–339. <https://doi.org/10.1080/08927014.2019.1605357>.
26. Johnson K, French K, Fichter JK, Oden R. 2008. Use of microbiocides in Barnett Shale gas well fracturing fluids to control bacteria related problems. New Orleans, Louisiana.
27. Lascelles P, Wan J, Robinson L, Allmon R, Evans G, Ursell L, Scott NM, Chase J, Jablanovic J, Karimi M, Rao V. 2017. Applying Subsurface DNA Sequencing in Wolfcamp Shales, Midland BasinSPE Hydraulic Fracturing Technology Conference and Exhibition. Society of Petroleum Engineers, The Woodlands, Texas, USA. <https://doi.org/10.2118/184869-MS>.
28. U.S. Energy Information Administration (EIA). <https://www.eia.gov/petroleum/drilling/>. <https://www.eia.gov/petroleum/drilling/#tabs-summary-2>. Retrieved 3 November 2019.
29. Kilbane J, Wylde J, Williamson A. 2015. Investigation of microorganisms in a west Texas oil field using growth and genetic testing. The Woodlands, Texas, USA.

30. Ursell L, Hale M, Menendez E, Zimmerman J, Dombroski J, Hoover K, Everman Z, Liu JK, Shojaei H, Percak-Dennett EM, Ishoey T. 2019. High resolution fluid tracking from vertical and laterals using subsurface DNA diagnostics in the Permian Basin. Denver, Colorado.
31. Daly RA, Borton MA, Wilkins MJ, Hoyt DW, Kountz DJ, Wolfe RA, Welch SA, Marcus DN, Trexler RV, MacRae JD, Krzycki JA, Cole DR, Mouser PJ, Wrighton KC. 2016. Microbial metabolisms in a 2.5-km-deep ecosystem created by hydraulic fracturing in shales. *Nat Microbiol* 1:16146. <https://doi.org/10.1038/nmicrobiol.2016.146>.
32. Mouser PJ, Borton M, Darrah TH, Hartssock A, Wrighton KC. 2016. Hydraulic fracturing offers view of microbial life in the deep terrestrial subsurface. *FEMS Microbiol Ecol* 92:fiw166. <https://doi.org/10.1093/femsec/fiw166>.
33. Akob DM, Cozzarelli IM, Dunlap DS, Rowan EL, Lorah MM. 2015. Organic and inorganic composition and microbiology of produced waters from Pennsylvania shale gas wells. *Applied Geochemistry* 60:116–125. <https://doi.org/10.1016/j.apgeochem.2015.04.011>.
34. Gaspar J, Davis D, Camacho C, Alvarez PJJ. 2016. Biogenic versus thermogenic H₂S source determination in Bakken Wells: considerations for biocide application. *Environmental Science* 6.
35. Lipus D, Roy D, Khan E, Ross D, Vikram A, Gulliver D, Hammack R, Bibby K. 2018. Microbial communities in Bakken region produced water. *FEMS Microbiol Lett* 365. <https://doi.org/10.1093/femsl/fny107>.
36. Tinker K, Gardiner J, Lipus D, Sarkar P, Stuckman M, Gulliver D. 2020. Geochemistry and microbiology predict environmental niches with conditions favoring potential microbial activity in the bakken Shale. *Front Microbiol* 11:1781. <https://doi.org/10.3389/fmicb.2020.01781>.
37. Strong LC, Gould T, Kasinkas L, Sadowisky MJ, Aksan A, Wackett LP. 2014. Biodegradation in waters from hydraulic fracturing: chemistry, microbiology, and engineering. *J Environ Eng* 140. [https://doi.org/10.1061/\(ASCE\)EE.1943-7870.0000792](https://doi.org/10.1061/(ASCE)EE.1943-7870.0000792).
38. Lipus D, Vikram A, Hammack R, Bibby K, Gulliver D. 2019. The effects of sample storage conditions on the microbial community composition in hydraulic fracturing produced water. *Geomicrobiol J* 36:630–638. <https://doi.org/10.1080/01490451.2019.1599470>.
39. Engle MA, Reyes FR, Varonka MS, Orem WH, Ma L, Ianno AJ, Schell TM, Xu P, Carroll KC. 2016. Geochemistry of formation waters from the Wolfcamp and “Cline” shales: Insights into brine origin, reservoir connectivity, and fluid flow in the Permian Basin, USA. *Chem Geol* 425:76–92. <https://doi.org/10.1016/j.chemgeo.2016.01.025>.
40. Blondes MS, Gans KD, Engle MA, Kharaka YK, Reidy ME, Saraswathula V, Thordsen JJ, Rowan EL, Morrissey EA. 2019. U.S. Geological Survey National Produced Waters Geochemical Database (ver. 2.3, May 2019): U.S. Geological Survey data release, <https://doi.org/10.5066/F7J964W8>.
41. Booker AE, Hoyt DW, Meulia T, Eder E, Nicora CD, Purvine SO, Daly RA, Moore JD, Wunch K, Pfiffner SM, Lipton MS, Mouser PJ, Wrighton KC, Wilkins MJ. 2019. Deep-Subsurface Pressure Stimulates Metabolic Plasticity in Shale-Colonizing *Halanaerobium* spp. *Appl Environ Microbiol* 85:e00018-19. <https://doi.org/10.1128/AEM.00018-19>.
42. Tatar A. 2018. Microbial Enhanced Oil Recovery, p 291–508. *In* Fundamentals of Enhanced Oil and Gas Recovery from Conventional and Unconventional Reservoirs. Elsevier, Amsterdam, the Netherlands.
43. Lapidus A, Chertkov O, Nolan M, Lucas S, Hammon N, Deshpande S, Cheng J-F, Tapia R, Han C, Goodwin L, Pitluck S, Liolios K, Pagani I, Ivanova N, Huntemann M, Mavromatis K, Mikhailova N, Pati A, Chen A, Palaniappan K, Land M, Hauser L, Brambilla E-M, Rohde M, Abt B, Spring S, Göker M, Bristow J, Eisen JA, Markowitz V, Hugenholtz P, Kyrpidis NC, Klenk H-P, Woyke T. 2011. Genome sequence of the moderately thermophilic halophile *Flexistipes sinusarabici* strain (MAS10T). *Stand Genomic Sci* 5:86–96. <https://doi.org/10.4056/signs.2235024>.
44. Cluff MA, Hartssock A, MacRae JD, Carter K, Mouser PJ. 2014. Temporal changes in microbial ecology and geochemistry in produced water from hydraulically fractured Marcellus shale gas wells. *Environ Sci Technol* 48:6508–6517. <https://doi.org/10.1021/es501173p>.
45. Jan-Roblero J, Posadas A, Zavala Díaz de la Serna J, García R, Hernández-Rodríguez C. 2008. Phylogenetic characterization of bacterial consortia obtained of corroding gas pipelines in Mexico. *World J Microbiol Biotechnol* 24:1775–1784. <https://doi.org/10.1007/s11274-008-9674-x>.
46. Davis JP, Struchtemeyer CG, Elshahed MS. 2012. Bacterial Communities Associated with Production Facilities of Two Newly Drilled Thermogenic Natural Gas Wells in the Barnett Shale (Texas, USA) 13.
47. Mand J, Park HS, Jack TR, Voordouw G. 2014. The role of acetogens in microbially influenced corrosion of steel. *Front Microbiol* 5. <https://doi.org/10.3389/fmicb.2014.00268>.
48. Choudhary L, Macdonald DD, Alfantazi A. 2015. Role of Thiosulfate in the Corrosion of Steels: A Review. *CORROSION* 71:1147–1168. <https://doi.org/10.5006/1709>.
49. Chen Y, Tang Q, Senko JM, Cheng G, Zhang Newby B, Castaneda H, Ju L-K. 2015. Long-term survival of *Desulfovibrio vulgaris* on carbon steel and associated pitting corrosion. *Corrosion Science* 90:89–100. <https://doi.org/10.1016/j.corsci.2014.09.016>.
50. Lahme S, Enning D, Callbeck CM, Menendez Vega D, Curtis TP, Head IM, Hubert CRJ. 2019. Metabolites of an Oil Field Sulfide-Oxidizing, Nitrate-Reducing *Sulfurimonas* sp. Cause Severe Corrosion. *Appl Environ Microbiol* 85:e01891-18. <https://doi.org/10.1128/AEM.01891-18>.
51. Kotu SP, Mannan MS, Jayaraman A. 2019. Emerging molecular techniques for studying microbial community composition and function in microbially influenced corrosion. *International Biodeterioration & Biodegradation* 144:104722. <https://doi.org/10.1016/j.ibiod.2019.104722>.
52. Vinson DS, Blair NE, Martini AM, Larter S, Orem WH, McIntosh JC. 2017. Microbial methane from in situ biodegradation of coal and shale: A review and reevaluation of hydrogen and carbon isotope signatures. *Chem Geol* 453:128–145. <https://doi.org/10.1016/j.chemgeo.2017.01.027>.
53. Gao B, Gupta RS. 2007. Phylogenomic analysis of proteins that are distinctive of Archaea and its main subgroups and the origin of methanogenesis. *BMC Genomics* 8:86. <https://doi.org/10.1186/1471-2164-8-86>.
54. Liu Y-F, Chen J, Zaramela LS, Wang L-Y, Mbadinga SM, Hou X-Z, Wu X-L, Gu J-D, Zengler K, Mu B-Z. 2020. Genomic and Transcriptomic Evidence Supports Methane Metabolism in *Archaeoglobi*. *mSystems* 5. <https://doi.org/10.1128/mSystems.00651-19>.
55. Guan Y, Ngugi DK, Vinu M, Blom J, Alam I, Guillot S, Ferry JG, Stingl U. 2019. Comparative Genomics of the Genus *Methanohalophilus*, Including a Newly Isolated Strain From Kebrint Deep in the Red Sea. *Front Microbiol* 10:839. <https://doi.org/10.3389/fmicb.2019.00839>.
56. Ravot T, Magot M, Ollivier B, Patel BKC, Ageron E, Grimont PAD, Thomas P, Garcia J-L. 2006. *Haloanaerobium congolense* sp. nov., an anaerobic, moderately halophilic, thiosulfate- and sulfur- reducing bacterium from an African oil field. *FEMS Microbiol Lett* 147:81–88. <https://doi.org/10.1111/j.1574-6968.1997.tb10224.x>.
57. Vengosh A, Jackson RB, Warner N, Darrah TH, Kondash A. 2014. A critical review of the risks to water resources from unconventional shale gas development and hydraulic fracturing in the United States. *Environ Sci Technol* 48:8334–8348. <https://doi.org/10.1021/es405118y>.
58. Stanley NR, Lazazzera BA. 2004. Environmental signals and regulatory pathways that influence biofilm formation: Signalling in biofilm formation. *Mol Microbiol* 52:917–924. <https://doi.org/10.1111/j.1365-2958.2004.04036.x>.
59. Whiteley CG, Lee D-J. 2015. Bacterial diguanylate cyclases: Structure, function and mechanism in exopolysaccharide biofilm development. *Biotechnol Adv* 33:124–141. <https://doi.org/10.1016/j.biotechadv.2014.11.010>.
60. Kempf B, Bremer E. 1998. Uptake and synthesis of compatible solutes as microbial stress responses to high-osmolality environments. *Arch Microbiol* 170:319–330. <https://doi.org/10.1007/s002030050649>.
61. Oren A. 2000. Change of the names *Haloanaerobiales*, *Haloanaerobiaceae* and *Haloanaerobium* to *Halanaerobiales*, *Halanaerobiaceae* and *Halanaerobium*, respectively, and further nomenclatural changes within the order *Halanaerobiales*. *Int J Systematic and Evolutionary* 50:2229–2230. <https://doi.org/10.1099/00207713-50-6-2229>.
62. Touati D. 2000. Iron and Oxidative Stress in Bacteria. *Arch Biochem Biophys* 373:1–6. <https://doi.org/10.1006/abbi.1999.1518>.
63. Ma Y, Galinski EA, Grant WD, Oren A, Ventosa A. 2010. Halophiles 2010: Life in Saline Environments. *Appl Environ Microbiol* 76:6971–6981. <https://doi.org/10.1128/AEM.01868-10>.
64. Wang H, Lu L, Chen X, Bian Y, Ren ZJ. 2019. Geochemical and microbial characterizations of flowback and produced water in three shale oil and gas plays in the central and western United States. *Water Res* 164:114942. <https://doi.org/10.1016/j.watres.2019.114942>.
65. Tinker K, Lipus D, Sarkar P, Gulliver D. 2020. The Geochemistry and Microbial Ecology of Produced Waters from Three Different Unconventional Oil and Gas Regions. Proceedings of the 8th Unconventional Resources Technology Conference. American Association of Petroleum Geologists, Online.
66. Nikolova C, Gutierrez T. 2021. Biosurfactants and their applications in the oil and gas industry: current state of knowledge and future perspectives. *Front Bioeng Biotechnol* 9:626639. <https://doi.org/10.3389/fbioe.2021.626639>.

67. Borton MA, Hoyt DW, Roux S, Daly RA, Welch SA, Nicora CD, Purvine S, Eder EK, Hanson AJ, Sheets JM, Morgan DM, Wolfe RA, Sharma S, Carr TR, Cole DR, Mouser PJ, Lipton MS, Wilkins MJ, Wrighton KC. 2018. Coupled laboratory and field investigations resolve microbial interactions that underpin persistence in hydraulically fractured shales. *Proc Natl Acad Sci U S A* 115:E6585–E6594. <https://doi.org/10.1073/pnas.1800155115>.
68. Kahrilas GA, Blotvogel J, Stewart PS, Borch T. 2015. Biocides in hydraulic fracturing fluids: a critical review of their usage, mobility, degradation, and toxicity. *Environ Sci Technol* 49:16–32. <https://doi.org/10.1021/es503724k>.
69. Struchtemeyer CG, Morrison MD, Elshahed MS. 2012. A critical assessment of the efficacy of biocides used during the hydraulic fracturing process in shale natural gas wells. *International Biodeterioration & Biodegradation* 71: 15–21. <https://doi.org/10.1016/j.ibiod.2012.01.013>.
70. American Petroleum Institute (API). Wellbore Pressure and Fluid Communication Associated with Hydraulic Fracturing. Briefing Paper, Digital Media | DM2014-113 | 05.14 | PDF. Washington, DC.
71. Murali Mohan A, Hartsock A, Hammack RW, Vidic RD, Gregory KB. 2013. Microbial communities in flowback water impoundments from hydraulic fracturing for recovery of shale gas. *FEMS Microbiol Ecol* 86:567–580. <https://doi.org/10.1111/1574-6941.12183>.
72. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, Knight R. 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci U S A* 108:4516–4522. <https://doi.org/10.1073/pnas.100080107>.
73. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, Gormley N, Gilbert JA, Smith G, Knight R. 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J* 6:1621–1624. <https://doi.org/10.1038/ismej.2012.8>.
74. Lever MA, Torti A, Eickenbusch P, Michaud AB, Šantl-Temkiv T, Jørgensen BB. 2015. A modular method for the extraction of DNA and RNA, and the separation of DNA pools from diverse environmental sample types. *Front Microbiol* 6. <https://doi.org/10.3389/fmicb.2015.00476>.
75. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope EK, Da Silva R, Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, et al. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* 37:852–857. <https://doi.org/10.1038/s41587-019-0209-9>.
76. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* 13:581–583. <https://doi.org/10.1038/nmeth.3869>.
77. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D. 2011. Scikit-learn: Machine Learning in Python. *Machine Learning in PYTHON* 6.
78. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41: D590–D596. <https://doi.org/10.1093/nar/gks1219>.
79. Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, Schweer T, Peplies J, Ludwig W, Glöckner FO. 2014. The SILVA and “All-species Living Tree Project (LTP)” taxonomic frameworks. *Nucleic Acids Res* 42: D643–D648. <https://doi.org/10.1093/nar/gkt1209>.
80. R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
81. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin R, O’Hara RB, Simpson GL, Solymos M, Henry M, Stevens H, Wagner H. 2012. *Vegan: Community Ecology Package*.
82. Neuwirth E. 2014. RColorBrewer: ColorBrewer Palettes. R package version 1.1-2. <https://CRAN.R-project.org/package=RColorBrewer>.
83. Menzel P, Ng KL, Krogh A. 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat Commun* 7:11257. <https://doi.org/10.1038/ncomms11257>.
84. Gruber-Vodicka HR, Seah BKB, Pruesse E. 2019. phyloFlash — Rapid SSU rRNA profiling and targeted assembly from metagenomes. *Biorxiv* <https://doi.org/10.1101/521922>.
85. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
86. Laczny CC, Sternal T, Plugaru V, Gawron P, Atashpendar A, Margossian H, Coronado S, der Maaten L, Vlassis N, Wilmes P. 2015. VizBin – an application for reference-independent visualization and human-augmented binning of metagenomic data. *Microbiome* 3:1. <https://doi.org/10.1186/s40168-014-0066-1>.
87. Wu Y-W, Simmons BA, Singer SW. 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 32:605–607. <https://doi.org/10.1093/bioinformatics/btv638>.
88. Wattam AR, Davis JJ, Assaf R, Boisvert S, Brettin T, Bun C, Conrad N, Dietrich EM, Disz T, Gabbard JL, Gerdes S, Henry CS, Kenyon RW, Machi D, Mao C, Nordberg EK, Olsen GJ, Murphy-Olson DE, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Vonstein V, Warren A, Xia F, Yoo H, Stevens RL. 2017. Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. *Nucleic Acids Res* 45: D535–D542. <https://doi.org/10.1093/nar/gkw1017>.
89. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25:1043–1055. <https://doi.org/10.1101/gr.186072.114>.
90. Richter M, Rosselló-Móra R, Oliver Glöckner F, Peplies J. 2016. JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* 32:929–931. <https://doi.org/10.1093/bioinformatics/btv681>.
91. Lagesen K, Hallin P, Rødland EA, Staerfeldt H-H, Rognes T, Ussery DW. 2007. RNAMmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 35:3100–3108. <https://doi.org/10.1093/nar/gkm160>.
92. Agarwala R, Barrett T, Beck J, Benson DA, Bollin C, Bolton E, Boureix D, Brister JR, Bryant SH, Canese K, Cavanaugh M, Charowhas C, Clark K, Dondoshansky I, Feolo M, Fitzpatrick L, Funk K, Geer LY, Gorenkov V, Graeff A, Hlavina W, Holmes B, Johnson M, Kattman B, Khotomlianski V, Kimchi A, Kimelman M, Kimura M, Kitts P, Klimke W, Kotliarov A, Krasnov S, Kuznetsov A, Landrum MJ, Landsman D, Lathrop S, Lee JM, Leubsdorf C, Lu Z, Madden TL, Marchler-Bauer A, Malheiro A, Meric P, Karsch-Mizrachi I, Mnev A, Murphy T, Orris R, Ostell J, O’Sullivan C, Palanigobu V, NCBI Resource Coordinators, et al. 2018. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 46: D8–D13. <https://doi.org/10.1093/nar/gkx1095>.
93. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res* 42:D206–D214. <https://doi.org/10.1093/nar/gkt1226>.
94. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomason JA, Stevens R, Vonstein V, Wattam AR, Xia F. 2015. RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci Rep* 5:8365. <https://doi.org/10.1038/srep08365>.
95. Davis JJ, Wattam AR, Aziz RK, Brettin T, Butler R, Butler RM, Chlenski P, Conrad N, Dickerman A, Dietrich EM, Gabbard JL, Gerdes S, Guard A, Kenyon RW, Machi D, Mao C, Murphy-Olson D, Nguyen M, Nordberg EK, Olsen GJ, Olson RD, Overbeek JC, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomas C, VanOeffelen M, Vonstein V, Warren AS, Xia F, Xie D, Yoo H, Stevens R. 2019. The PATRIC Bioinformatics Resource Center: expanding data and analysis capabilities. *Nucleic Acids Res* 47: gkz943. <https://doi.org/10.1093/nar/gkz943>.
96. Kanehisa M, Goto S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* 28:4.
97. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. 2012. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* 40:D109–D114. <https://doi.org/10.1093/nar/gkr988>.
98. Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. 2021. KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res* 49: D545–D551. <https://doi.org/10.1093/nar/gkaa970>.
99. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, Dehal P, Ware D, Perez F, Canon S, Sneddon MW, Henderson ML, Riehl WJ, Murphy-Olson D, Chan SY, Kamimura RT, Kumari S, Drake MM, Brettin TS, Glass EM, Chivian D, Gunter D, Weston DJ, Allen BH, Baumohl J, Best AA, Bowen B, Brenner SE, Bun CC, Chandonia J-M, Chia J-M,

- Colasanti R, Conrad N, Davis JJ, Davison BH, DeJongh M, Devoid S, Dietrich E, Dubchak I, Edirisinghe JN, Fang G, Faria JP, Frybarger PM, Gerlach W, Gerstein M, Greiner A, Gurtowski J, Haun HL, He F, Jain R, Joachimiak MP, Keegan KP, Kondo S, et al. 2018. KBase: The United States Department of Energy Systems Biology Knowledgebase. *Nat Biotechnol* 36:566–569. <https://doi.org/10.1038/nbt.4163>.
100. Shaffer M, Borton MA, McGivern BB, Zayed AA, La Rosa SL, Solden LM, Liu P, Narrowe AB, Rodríguez-Ramos J, Bolduc B, Gazitúa MC, Daly RA, Smith GJ, Vik DR, Pope PB, Sullivan MB, Roux S, Wrighton KC. 2020. DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Res* 48:8883–8900. <https://doi.org/10.1093/nar/gkaa621>.