## **Supplementary Information**

**SI Data 1. gBlocks Gene Fragment Sequence for qPCR Standard.** Fragment was constructed using the 16S bac, McrA, DsrA, and RdIA genes.



**SI Figure 1.** qPCR standard curve. Sample extraction volumes were normalized in order to calculate 16S rRNA gene copies/mL.



**SI Figure 2.** A) TDS and chloride measurements plotted for this study and for similar produced waters found in the USGS PWDB (Blondes et al., 2019). Linear trendlines report high linear correlation values, displaying that chloride is the primary anion and well 11, with its lower TDS value, does not have an anomalous chloride value. The trendline excludes well 18, a sample that, based on its geochemistry and age, likely represents a mix of flowback and produced water. B) TDS and sulfate measurements plotted for produced waters in this study and for similar produced waters found in the USGS PWDB (Blondes et al., 2019). For samples in this study, there is a higher linear association (R<sup>2</sup>=0.8533) for these analytes when the youngest and spatially unique sample, well 18, is excluded. Based on measurements from this study and previous research, the elevated sulfate in well 18 could be related to residual chemicals from hydraulic fracturing operations or the product of flowback waters.



SI Figure 3. Microbial abundances for each produced water sample.



**SI Figure 4. Boxplots of alpha diversity measurements within produced water samples.** Boxplots show Chao1 Richness (top), Shannon Diversity (middle), and Pielou's Eveness (bottom) in our samples (May 2018) and those collected by Lipus et al. in March 2015, January 2015, and December 2014. All alpha diversity measurements are shown in SI Table 1 and were calculated after resampling libraries to a depth of 1942. For each group, the bars delineate the median, the hinges represent the lower and upper quartiles, the whiskers extend to the lesser of the most extreme value or 1.5 times the interquartile range, and outliers are plotted, if present.



SI Figure 5. Relative abundance of microbial families in produced water samples. Samples from our current study are shown next to samples previously collected from the same site (Lipus et al., 2018). Two of the 10 blanks that resulted in sequencing data are also shown. Samples are clustered by well number, sampling location (S: Separator, T: Tank), and sample event (My: May 2018; M: March 2015; J: January 2015; D: December 2014; O: October 2014). All families that represent ≥5% of sequences from any one sample are listed in the barplot, all other families are grouped together under "Other."



**SI Figure 6. Nonmetric multidimensional scaling (NMDS) plots** of: A) samples and blanks from our current study, B) samples and blanks from our current study as well as samples previously collected from the same site (Lipus et al., 2018), and C) samples and blanks from our current study and all samples previously collected by Lipus et al. All NMDS plots were constructed using Bray-Curtis distance calculated after sequence libraries were resampled to the depth of the sample with the fewest sequences from this experiment (1942 sequences). Constructed plots had stress values of 0.08576772, 0.07193307, and 0.07895004, respectively. Data from our current study are colored black (experimental data) or red (blanks) while data from Lipus et al. is gray. Each sample is labeled by its corresponding well number, sampling location (S: Separator, T: Tank), and sample event (My: May 2018; M: March 2015; J: January 2015; D: December 2014; O: October 2014).



**SI Figure 7.** (Left) Nonmetric multidimensional scaling (NMDS) plot of produced water samples with a stress value of 0.07194914. For this plot, sequence libraries were resampled to the depth of the sample with the fewest sequences from this experiment (1942 sequences). Vectors for geochemical parameters fit onto the plot are shown in blue. The direction of the arrow corresponds to the direction of the gradient while the length of the vector is proportional to the correlation between ordination and the environmental variable. (Right) Table containing the R<sup>2</sup> and p-values for all constructed environmental vectors.