Shale gas development has limited effects on stream biology and geochemistry in a gradient-based, multiparameter study in Pennsylvania

Adam C. Mumford[a,1], Kelly O. Maloney[b], Denise M. Akob[a], Sarah Nettemann[c], Arianne Proctor[d], Jason Ditty[d], Luke Ulsamer[d], Josh Lookenbill[e], and Isabelle M. Cozzarelli[a]

[a]Water Mission Area, US Geological Survey, Reston, VA 20192; [b]Leetown Science Center, US Geological Survey, Kearneysville, WV 25430; [c]Applied Geology, Institute of Geosciences, Friedrich Schiller University Jena, Burgweg 11, 07749 Jena, Germany; [d]Bureau of Forestry, Pennsylvania Department of Conservation and Natural Resources, Harrisburg, PA 17101; and [e]Division of Water Quality, Pennsylvania Department of Environmental Protection, Harrisburg, PA 17101

Edited by Andrea Rinaldo, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, and approved December 28, 2019 (received for review July 8, 2019)

The number of horizontally drilled shale oil and gas wells in the United States has increased from nearly 28,000 in 2007 to nearly 127,000 in 2017, and research has suggested the potential for the development of shale resources to affect nearby stream ecosystems. However, the ability to generalize current studies is limited by the small geographic scope as well as limited breadth and integration of measured chemical and biological indicators parameters. This study tested the hypothesis that a quantifiable, significant relationship exists between the density of oil and gas (OG) development, increasing stream water concentrations of known geochemical tracers of OG extraction, and the composition of benthic macroinvertebrate and microbial communities. Twenty-five headwater streams that drain lands across a gradient of shale gas development intensity were sampled. Our strategy included comprehensive measurements across multiple seasons of sampling to account for temporal variability of geochemical parameters, including known shale OG geochemical tracers, and microbial and benthic macroinvertebrate communities. No significant relationships were found between the intensity of OG development, shale OG geochemical tracers, or benthic macroinvertebrate or microbial community composition, whereas significant seasonal differences in stream chemistry were observed. These results highlight the importance of considering spatial and temporal variability in stream chemistry and biota and not only the presence of anthropogenic activities in a watershed. This comprehensive, integrated study of geochemical and biological variability of headwater streams in watersheds undergoing OG development provides a robust framework for examining the effects of energy development at a regional scale.

Marcellus Shale | hydraulic fracturing | water quality | microbiology | macroinvertebrates

The development of unconventional oil and gas (OG) resources over the past decade has led to a drastic increase in the importance of natural gas as an energy resource. This change has perhaps been most notable for the Marcellus Formation in Pennsylvania, where production has increased from 193 billion cubic feet (BCF) per year in 2006 to nearly 5,500 BCF per year in 2017 and now accounts for nearly 20% of total natural gas production in the United States (1). In Pennsylvania, these increases in production have come with an increase in the number of horizontally drilled, hydraulically fractured wells, which have increased from 3 in 2006 to 7,977 in 2017 (1). Numerous activities associated with OG development and production, including unintentional releases of hydraulic fracturing fluid and waste water, erosion, and sedimentation due to the construction and operation of well pads, pipelines, and unpaved roads, as well as increased road use and releases in the course of waste disposal, can pose potential risks to surface water (2–7). A growing body of studies has investigated the ecological effects of OG production; however, no clear consensus has emerged, and the comparability of the studies is limited, due to variations in study design; the confounding effects of regional differences in land use, geology, and regulation; and a lack of long-term baseline data. The effects of regional and temporal heterogeneity were recently highlighted by Knee and Masker (8), whose comparative study of water geochemistry in southwestern Pennsylvania and western Maryland was unable to definitively link observed changes in geochemical composition to OG development in the face of differences among historic and current land use and potential seasonal differences.

Hydraulic fracturing fluid is a complex mixture of water, sand, hydrocarbons, surfactants, biocides, and a wide array of proprietary compounds designed to enhance fracturing of and production from the hydrocarbon bearing formation (9–12). Produced water is a complex mixture of reservoir formation brines and injected fluids which return to the surface over the lifetime of the OG well (13–17). In general, and in the Marcellus Shale in particular, produced

Significance

This investigation provides a comprehensive evaluation of the geochemical and biological effects of shale gas development on 25 small watersheds over the course of 2 y. Sampling headwater streams seasonally over two consecutive years yielded no statistically significant relationships between the intensity, presence, or absence of shale gas development and any signal in a comprehensive set of chemical constituents (including those recognized as oil and gas geochemical tracers) or any changes in microbial or benthic macroinvertebrate community composition. This work provides a framework for investigations of anthropogenic effects stemming from natural resource development, and highlights the importance of conducting studies which control for regional and temporal variability.

Author contributions: A.C.M., K.O.M., A.P., J.D., L.U., and I.M.C. designed research; A.C.M., K.O.M., S.N., L.U., and J.L. performed research; A.C.M., K.O.M., D.M.A., S.N., J.L., and I.M.C. analyzed data; and A.C.M., K.O.M., D.M.A., and I.M.C. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Data deposition: Sequence data have been deposited as a BioProject in the NCBI GenBank database under accession PRJNA544240. Sediment chemistry data are available from USGS ScienceBase at https://doi.org/10.5066/P9GJTRYR. Water chemistry data are available from the USGS National Water Information System, https://waterdata.usgs.gov/nwis, under the ID numbers listed in the manuscript. Macroinvertebrate community data are available from the Pennsylvania Department of Environmental Protection.

[1]To whom correspondence may be addressed. Email: amumford@usgs.gov.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1911458117/-/DCSupplemental.

First published February 3, 2020.

waters are highly saline and have a distinct elemental composition in comparison to produced water from younger formations (16, 18–23). Marcellus Shale-produced waters in north-central Pennsylvania are described as Na–Ca–Cl brines characteristic of evaporated seawater (15, 16), have concentrations of total dissolved solids (TDS) in excess of 100,000 mg/L, and contain high concentrations of Na, Ca, Cl, Li, B, Ba, Sr, and Br, which are useful as geochemical tracers for the presence of produced water (24–29). Streambed sediments have been shown to retain indicators of OG wastes including Ba, Sr, and Ca when there are few detectable geochemical tracers in stream water (25, 30); however, few studies have examined bed sediments at a regional scale.

Releases of produced water into surface water have led to accumulation of metals in freshwater mussel shells (30) and are reported to cause changes in streambed microbial community structure (29, 31, 32). Benthic macroinvertebrate communities are sensitive to changes in water quality, and changes in benthic macroinvertebrate community structure have occurred in streams in Arkansas within watersheds undergoing extraction of natural gas from the Fayetteville Shale (33). Exposure to elevated salinity is known to be toxic to benthic macroinvertebrates (34, 35). Similarly, an increase in salinity, as would occur following a release of produced waters, can have lethal effects on fish (36, 37). Case studies following releases of produced water have suggested the potential for such releases to effect microbial community structure and function (29, 38); several studies detailing experimental exposures to components of hydraulic fracturing fluid and produced water have demonstrated this potential for alterations to microbial communities (39–41). Furthermore, produced water and produced water impoundment ponds have distinct microbial populations (13, 42–47), suggesting that microbial community profiles may have a potential utility as biological tracers for produced water. A group of studies focusing on northwestern Pennsylvania have demonstrated changes in microbial community structure in watersheds with OG activity when contrasted to similar watersheds without activity, and report that microorganisms capable of using components of produced water and methane are found in higher abundance in surface water and sediments in watersheds with OG activity than in those without (31, 32, 48).

Shale gas development and production in the Pennsylvania State Forests (PASF) provides an opportunity to study the effects of shale gas production on water quality and stream biotic communities of headwater streams in an environment with few other potential anthropogenic stressors. OG development is a component of state forest management, with 74 OG lease sales resulting in more than 2,000 OG wells being drilled in the PASF since 1947. Of these, ∼500 are Marcellus Shale gas wells completed since 2008 (49, 50). Many headwater streams within the PASF system are designated as protected use High Quality (HQ) or Exceptional Value (EV) in state regulations (51). To manage and protect the state forests and their water resources, the Bureau of Forestry of the Pennsylvania Department of Conservation and Natural Resources (BOF) uses a robust and comprehensive lease agreement combined with the Guidelines for Administering Oil and Gas Activity on State Forest Land to manage this development (50). Further, the BOF established a shale gas monitoring program in 2011 that consists of an integrated monitoring team, on-the-ground management activities, and research collaborations with external partners (49), and the findings from this program inform the development of best management practices (BMPs).

While the studies undertaken, to date, have improved our understanding of the potential effects of OG development on water quality and stream biota, none have coupled detailed trace geochemical analyses with biological responses across a large geographical area with consistent land use or accounted for temporal variability over multiple years. This has limited the ability to define quantitative linkages between OG development and biological effects. Furthermore, it is challenging to extrapolate the findings from case studies of extraordinary spill events to predict potential regional impacts. This study hypothesizes a quantifiable relationship between the density of OG development and the water quality and biota of headwater streams in the PASF system. We test this hypothesis by examining and integrating stream water chemistry, sediment chemistry, benthic macroinvertebrate community composition, and streambed microbial community structure across a gradient of OG development intensity over the course of 2 y.

Study Design and Site Selection

The PASF system was chosen as the study area because it contains a gradient of shale gas development under a consistent regulatory structure while having only limited land use other than recreation, timber harvesting, and OG development since the early to mid 20th century (52). To select sites, we used the 12-digit Hydrologic Unit Code (HUC12; https://water.usgs.gov/GIS/huc.html) watershed vulnerability scores from Entrekin et al. (6). Entrekin et al. (6) computed vulnerability scores from indices developed to describe watershed sensitivity and exposure to natural and anthropogenic disturbances, including two measures of shale gas development (well density and proximity to nearest stream), for six shale plays across the contiguous United States. To categorize the HUC12 watersheds within the PASF, we took the Entrekin et al. (6) vulnerability index for each HUC12 watershed within the PASF system, binned them by quintiles, and assigned each watershed to one of five categories (Highest, High, Medium, Low, and None). Five streams with similar surficial geology were then selected at random from each category (Fig. 1, and see *SI Appendix*, Table S1 and Fig. S1 for the distribution of conventional and shale gas wells). After the first year of the study, a finer-scaled (1:24,000 catchments), more regionally focused standardized Disturbance Intensity Index (sDII) was published for the Pennsylvania portion of the Upper Susquehanna River Basin (PAUSRB) (7), which improved upon the Entrekin et al. (6) 2015 index by including upstream drainage information on 17 measures of OG that incorporate all steps of the development process (infrastructure, gas and waste production, notice of violations [including spills], and water withdrawals) (refs. 7 and 53 and *SI Appendix*, Tables S2–S4). For each catchment in the PAUSRB, each OG metric was assigned a rank based on the underlying distribution of the stressor in the study, which was then multiplied by a weighting score that ranged from one to three based on potential impact to a stream (7). The land use and OG parameter data for the watersheds in this study can be found in *SI Appendix*, Tables S2 and S3. The sDII was then calculated for each catchment as the sum of weight-adjusted scores, and standardized from 0 to 100 by dividing by the study-wide maximum sDII. The sites were recategorized (highest, high, medium, low, lowest) based on this finer-grained index of development intensity in individual drainages rather than at the HUC12 scale. These sDII scores and categories did not change year to year (*SI Appendix*, Table S1) and were used for all statistical analyses. The complete dataset used to construct the sDII is available from US Geological Survey (USGS) ScienceBase (53).

Abandoned mine drainage (AMD) is a known stressor to water quality in Pennsylvania (54); therefore, two streams with sources of AMD in their watersheds (55), Bark Camp Run (National Water Information System [NWIS] ID 01542613) and Boone Run (NWIS ID 015162658), were included to provide a geochemical and biological signature of AMD stress. These AMD-influenced streams were used to provide an initial screen and reference for AMD but were not included in any other statistical analyses, as their geochemical and biological composition could not be considered comparable.

ENVIRONMENTAL SCIENCES

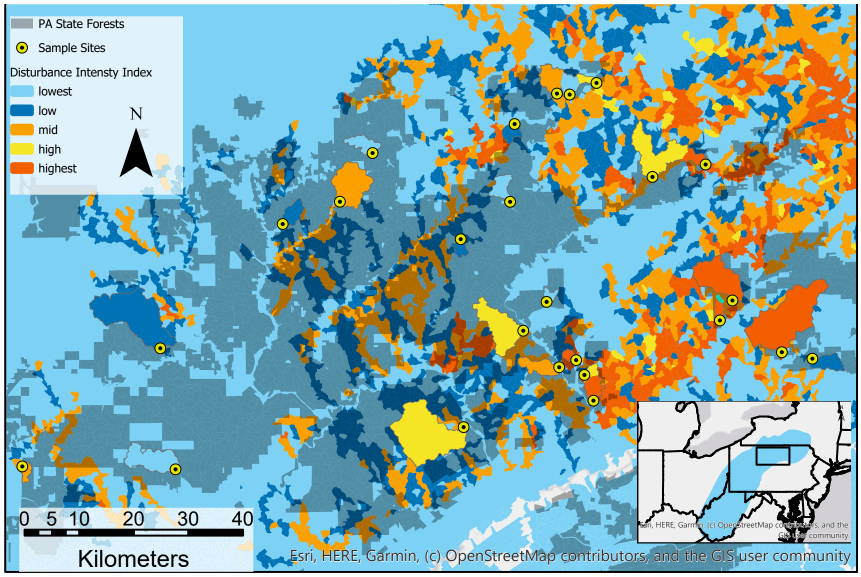

Fig. 1. Location of study sites within the PASF system. The boundaries of the Marcellus Shale are shaded in blue, PASF land is shaded in gray, and 1:24,000 drainages are colored by sDII category. Sample locations are marked with yellow bull's-eyes. *Inset* shows the extent of the Marcellus Shale in blue; the study area is outlined in black. Image courtesy of Esri, HERE, Garmin © OpenStreetMap contributors, and the GIS user community.

Sampling Design. Streams were sampled during the spring and fall of 2016 and 2017. Samples were taken in early spring to evaluate water quality, benthic macroinvertebrate communities prior to emergence, and microbial communities during spring snowmelt and runoff. Fall samples were taken to evaluate water quality and microbial community structure under base flow conditions. Sampling for geochemical parameters and sediment microbial community structure was carried out as described in refs. 28 and 29 (*SI Appendix, Supplemental Item 1* and Tables S5 and S6). Analytical procedures for determination of alkalinity, anions, cations/trace metals, and nonvolatile dissolved organic carbon are described in *SI Appendix, Supplemental Item 1*. Trace light hydrocarbons in stream water, a measure of stray gas from leaking well casings (56), were sampled, analyzed, and reported in Haase et al. (57). Sediment samples were dried at 60 °C, sieved to <2 mm, and digested as described by Environmental Protection Agency method 3051A (58) prior to analysis for trace metal composition by inductively coupled plasma–mass spectrometry (*SI Appendix*, Table S6). Macroinvertebrates were sampled, subsampled, and identified according to the Pennsylvania Department of Environmental Protection (PADEP) data collection protocols (59) and assessment methods (60). While this method does not generate quantitative macroinvertebrate census data, it has been demonstrated to provide robust relative abundance data at a sample-to-sample level (61). Sensitive Ephemeroptera, Plecoptera, and Trichoptera (EPT) taxa metrics were calculated using taxa with Pollution Tolerance Values of 0 to 4 as defined by PADEP (59, 60).

Microbial DNA was extracted from site sediments using the Qiagen DNeasy PowerSoil Kit according to the manufacturer's instructions (Qiagen) prior to Illumina MiSeq sequencing of the 16S ribosomal RNA gene (*SI Appendix, Supplemental Item 1*).

Data Analysis. Data analysis was performed in R, using CORE components and the VEGAN, PHYLOSEQ, and DPLYR packages (62–65) (*SI Appendix, Supplemental Item 4*). Geochemical tracer data were subjected to three-way ANOVA to separate the effects of OG development from seasonal and annual variation (*SI Appendix, Supplemental Item 4* and Table S4). Initial quality control, alignment, and taxonomic assignment of microbial sequence data were performed using MOTHUR v1.39.5, based on the Silva 128 nonredundant database (66–68) using the USGS Advanced Research Computing (ARC) Yeti high-performance computing facility (66, 69) (*SI Appendix, Supplemental Item 2*). Maximum likelihood trees were constructed using RaxML and exaML (70, 71) (*SI Appendix, Supplemental Item 3*). To account for the relatedness of microbial taxa, weighted UniFrac distance matrices (72) were used instead of Euclidean distance matrices for microbial community analysis.

Data Availability. Microbial sequence data are available from the National Center for Biotechnology Information (NCBI) GenBank database under BioProject PRJNA544240 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA544240) (73). Water chemistry data are available from the USGS NWIS, https://waterdata.usgs.gov/nwis, using the NWIS identification numbers provided in *SI Appendix*, Table S1 (74). Sediment chemistry data are available as a USGS data release at https://doi.org/10.5066/P9GJTRYR (75). Macroinvertebrate data are available from PADEP.

Results and Discussion

Aqueous and Sediment Geochemistry. No significant relationships were found among any of the measured parameters and the sDII categories, although significant seasonal variation was observed. Forty physical and chemical parameters were measured for this study (*SI Appendix*, Table S5), and, of these, specific conductance (as a proxy for TDS) and dissolved Cl, Br, Ca, Na, Li, B, Ba, and Sr are known as useful geochemical tracers of produced water (24, 25, 28, 29, 76, 77). The pH was measured as a potential indicator for acids used in shale gas well completion and for potential effects on microbial and macroinvertebrate communities (32, 33, 48, 78).

Seasonality was identified as the primary driver of variability in water chemistry and produced water geochemical tracers, with little variability observed between years or across the sDII gradient. Streamflow was significantly higher in spring than in fall (spring median 0.31 ± 0.41 m^3/s, fall median 0.01 ± 0.06 m^3/s, t test $P < 0.05$), and specific conductance values were significantly lower in spring than in fall (spring median 35.0 ± 12.2 µS, fall median 62.7 ± 29.2 µS, t test $P < 0.05$). These data suggest that precipitation and snowmelt comprised a larger proportion of the stream water in spring (Fig. 2 and *SI Appendix*, Table S4). Water composition (*SI Appendix*, Table S7) did not significantly differ across sDII categories (ADONIS, $P = 0.71$) or years (ADONIS, $P = 0.23$) once the significant seasonal variation (ADONIS, $P < 0.05$) was accounted for, suggesting that OG development did not influence stream chemistry in the study area. Analysis of water chemistry data by nonmetric multidimensional scaling (NMDS) (Fig. 3) illustrates that the composition was less variable during the spring than in the fall, presumably due to the greater influence of dilution with precipitation during the spring compared to the base flow conditions found in the fall.

Ratios of Ca, Cl, Li, B, Br, Ba, and Sr concentrations have been described as useful tools for identifying the presence of produced waters in surface waters (24, 25, 28, 29, 76, 77). In this study, the utility of these geochemical tracers was limited by low occurrence above quantification limits across the study area. No sample had a Br concentration above the reporting limit of 0.3 mg/L, B was measured above the reporting limit of 10 μg/L at only three sites in the fall and none in the spring, and Li was measured above the reporting limit of 1 μg/L at seven sites during fall and five sites during spring. While this limited occurrence precludes the use of elemental ratios involving Br, B, and Li to provide regional evidence for the presence of produced water, it provides a strong indication that produced waters have not influenced the streams in the study area despite the presence of OG development. Chloride concentration did not differ between years or seasons (Fig. 2*D*); however, it was significantly higher in the low sDII category than in the high sDII category (ANOVA, Tukey's honest significant difference [HSD] $P = 0.02$). The highest average stream Cl concentration, 6.9 ± 1.6 mg/L, was observed at Slate Run (NWIS ID 01548615), which has no OG development activity in the watershed but has more residential inholdings and greater road development, which may

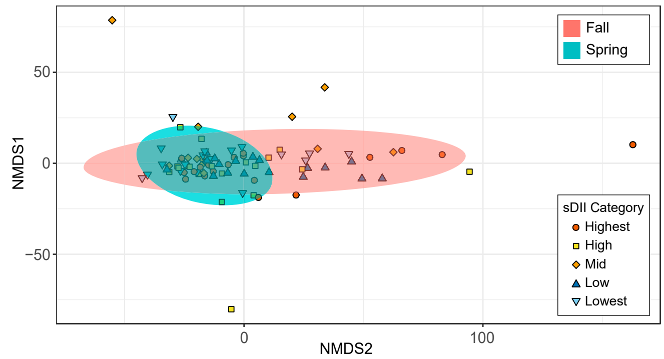

Fig. 3. NMDS of Euclidean distances of aqueous chemical profiles. Ellipses represent 95% CIs for spring and fall samples. While more variability was observed in the fall than in the spring, no significant differences were observed among sDII categories. We hypothesize that the increased variability seen along NMDS2 is related to seasonal differences and local differences in geology and groundwater contribution to streamflow.

provide a source of Cl (79). We found no relationship between Ba/Cl ratios and sDII categories (ANOVA, Tukey's HSD $P > 0.05$; *SI Appendix*, Fig. S2), and Ba/Cl ratios in all sDII categories were significantly lower (ANOVA, Tukey's HSD $P < 0.05$) than those reported for Marcellus-produced water from this region (*SI Appendix*, Fig. S2 and ref. 16). Produced water from this region of the Marcellus Shale is reported to have higher Ba/Cl ratios than elsewhere in the Marcellus Shale (16, 80). When the very low concentrations of Cl and Ba in the study streams are taken into consideration (Fig. 2 *D* and *F*), we would expect even a small contribution of produced waters to result in measurable change. That we found no significant changes in Ba/Cl ratios provides

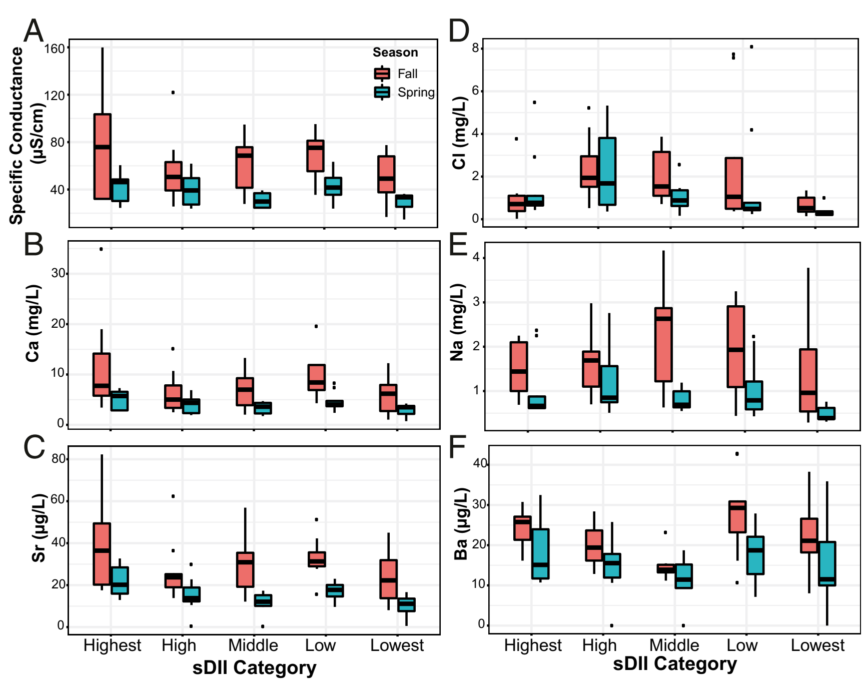

Fig. 2. Comparisons of aqueous geochemical indicators of OG wastewater between spring and fall, averaged across the 2-y study. (*A*) Specific conductance (microsiemens per centimeter), (*B*) Ca (milligrams per liter), (*C*) Sr (micrograms per liter), (*D*) Cl (milligrams per liter), (*E*) Na (milligrams per liter), and (*F*) Ba (micrograms per liter). The lower and upper hinges correspond to the 25th and 75th percentile, respectively, and the whiskers correspond to 1.5 × the interquartile range. Specific conductance and concentrations of Ca, Na, Sr, and Ba were significantly higher in fall than in spring, presumably due to dilution of groundwater with precipitation and snowmelt. No significant differences among sDII categories or years were observed. Cl concentration did not significantly vary by year or season but was significantly higher in the low sDII category, presumably due to residential and road development in one watershed.

further indication that Marcellus-produced waters are not influencing stream chemistry.

Heilweil et al. (56) describe trace light hydrocarbons as a tracer for gas migration from faulty well casings, and were measured at our sites during the spring and fall of 2016. The results from these samples are reported in ref. 57. Concentrations above atmospheric equilibrium were not observed in any of the samples (57), indicating that gas from the Marcellus Formation is not actively entering the streams within 1 km of the sampling points and suggesting that the study streams are not being influenced by faulty well casings. These findings are consistent with those of Barth-Naftilan et al. (81), which reported no changes in groundwater methane concentrations that could be linked to shale gas development.

The pH did not differ among sDII categories with or without controlling for season or year, indicating OG development did not have any significant effect on stream pH within our study. The lowest pH value in the study (5.0) was observed in spring 2017 at Sebring Branch of Mill Run (NWIS ID 01548770), a site in the lowest sDII category with no OG or historic mining activity in its watershed and with the lowest observed specific conductance (14.7 μS/cm), suggesting that the low pH is a function of local geology, hydrology, and/or precipitation. This contrasts the findings of Chen See et al. (31), Trexler et al. (32), and Ulrich et al. (48), which suggested that changes in pH resulting from acids used in shale gas well completion were significantly related to shale gas development in a similar geologic setting.

We evaluated the concentration of geochemical OG tracers in streambed sediment (*SI Appendix*, Table S8), because they have been shown to retain indicators of OG wastes when there are few detectable geochemical tracers in stream water (25, 30). No significant relationships among sDII categories and OG geochemical tracer concentrations in streambed sediment were observed, and the concentrations of OG geochemical tracers in streambed sediments followed similar trends as those observed in stream water (*SI Appendix*, Table S8). As was observed for water chemistry, no significant differences were observed among sDII, year, or season for sediment concentrations of Ba and Sr (ANOVA, Tukey's HSD $P > 0.05$), but significant seasonal differences were observed for Li (ANOVA, Tukey's HSD $P = 0.012$). Sediment B concentrations were higher in fall than in spring (ANOVA, Tukey's HSD $P < 0.05$), and higher in spring 2016 than in spring 2017. The concentration of B in streambed sediment did not differ among sDII categories after accounting for season and year. The lack of any significant difference in geochemical OG tracers between the lowest sDII category sites with very little or no current or historic OG development and those higher along the sDII gradient suggests that these are the naturally occurring concentrations for these elements, and not indicative of OG activity.

Sediment Microbial Communities. No significant changes in streambed microbial diversity were observed in relation to sDII category, season, or year, as measured by the Chao1 richness and Shannon diversity metrics (Fig. 4). This finding is consistent with ref. 31, which reports no observed changes in microbial alpha diversity with proximity to OG activity in northeastern Pennsylvania (31). No significant variation in microbial community structure was observed among sDII categories (Fig. 5*A*), and no clear trends emerged to indicate changes to community structure in relation to presence or density of shale gas activity (*SI Appendix*, Table S9). Annual differences were not significant (ADONIS, $P = 0.88$), nor were significant differences among sDII categories observed when controlling for seasonal variation (ADONIS, $P = 0.8$).

Several aqueous constituents were found to have a relationship to site-level differences in streambed microbial community structure during the spring (Fig. 5*B* and *SI Appendix*, Tables S10 and S11). Silver (Ag) and molybdenum (Mo) concentrations

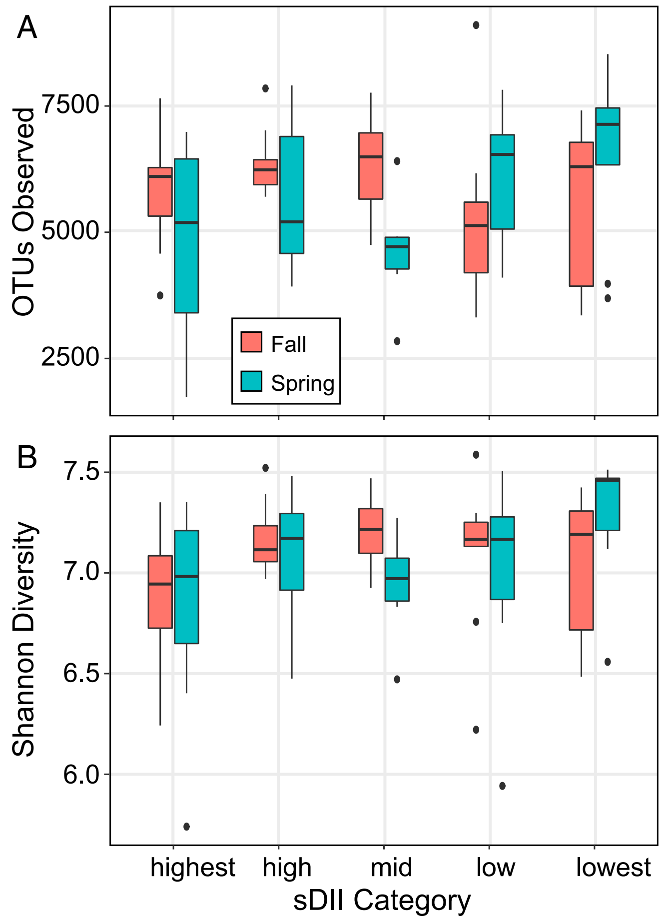

Fig. 4. (*A*) Chao1 richness estimate and (*B*) Shannon diversity of streambed microbial communities. The lower and upper hinges correspond to the 25th and 75th percentile, respectively, and the whiskers correspond to 1.5 × the interquartile range. Significant seasonal differences were observed, but no significant differences in microbial community richness or diversity were observed among sDII categories when accounting for seasonality. OTUs, operational taxonomic units based on 97% similarity.

were significantly related to microbial community structure; however, this may be a statistical artifact, as Ag and Mo were only present above the reporting limit in two and eight samples, respectively (*SI Appendix*, Table S11). Ag and Mo concentrations did not covary with OG-related parameters and did not have any relationship to sDII categories (Fig. 5*B*), suggesting that they are not reliable indicators of OG activity. In contrast to the relationship of pH to changes in microbial community structure described in ref. 32, pH did not play a significant role in differences in microbial community structure (ENVFIT, $P > 0.05$). No sediment constituents were significantly related to variations in microbial community structure.

Of the variables directly related to shale gas production used in constructing the sDII (*SI Appendix*, Tables S2 and S4 and ref. 7), the mass of drill cuttings, volumes of gas and water produced, and volumes of drilling fluid and fracturing fluid used had significant site-level relationships to community structure during the spring (Fig. 5*B*; ENVFIT, $P < 0.05$). In addition, significant relationships were identified between changes in spring microbial community structure and Environmental Health and Safety Violations and Pennsylvania Clean Stream Law Violations (*SI Appendix*, Table S8). The number and density of pipeline crossings (*SI Appendix*, Table S4) had a significant relationship to changes in microbial community structure during the spring

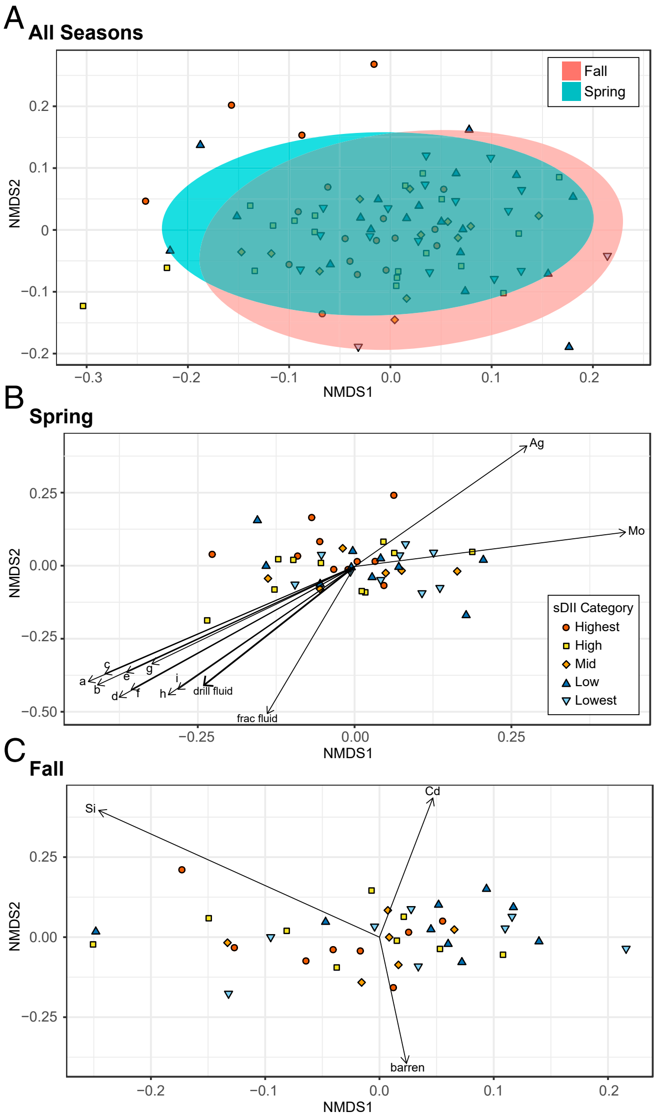

Fig. 5. (*A*) NMDS of weighted Unifrac distances among streambed microbial communities. Ellipses represent 95% CIs between samples from spring and fall. (*B*) NMDS of spring streambed sediment microbial community structure with significant (ENVFIT, $P < 0.05$) landscape and chemical vectors fitted. Vectors in the lower left side of the figure are labeled with letters as follows: a, Environmental Health and Safety Notices of Violation; b, Waters of Pennsylvania Notices of Violations; c, volume of gas produced; d, number of pipe crossings; e, sDII score; f, mass of drill cuttings; g, well count; h, pipe crossing density; I, volume of produced water; Ag, silver (micrograms per liter); and Mo, molybdenum (micrograms per liter). (*SI Appendix*, Tables S10 and S11). (*C*) NMDS of fall streambed sediment microbial communities with significant vectors fitted.

(Fig. 5*B* and *SI Appendix*, Table S11; ENVFIT, $P < 0.05$) but not the fall (Fig. 5*C*). Runoff and sedimentation are higher during spring than fall in this region (3), and we hypothesize that the significant effects observed only during the spring may be a proxy for increased runoff and sedimentation from well pads, associated roads, road stream crossings, and pipeline stream crossings, given that we observed no indication of any geochemical or biological tracer of produced water in stream water or sediments. Additional research is required to determine the validity of this hypothesis and to determine the role, if any, of other unassessed factors, such as road use related to non-OG activities.

Findings from Macroinvertebrate Community Analysis. Macroinvertebrate community structure differed between 2016 and 2017 (ADONIS, $P < 0.05$); however, no significant differences in macroinvertebrate community structure were observed among sDII categories (Fig. 6). Despite the differences in community structure between years, total richness (Fig. 7*A*), richness of sensitive EPT taxa (as defined by ref. 60) (Fig. 7*B*), Shannon Diversity (Fig. 7*C*), and Index of Biotic Integrity (IBI) scores as defined in ref. 60 did not differ between years or among sDII categories. Based on these findings, no stream in our study would be rated as nonattaining according to ref. 82.

Johnson et al. (33) suggest that changes in sedimentation as a result of shale gas development may lead to changes in functional feeding group (FFG) distributions, but we found no significant relationships among sDII categories and FFG classifications after accounting for annual differences (*SI Appendix*, Table S13).

Conclusions

This study hypothesized the existence of a quantifiable relationship between the intensity of disturbance from Marcellus Shale gas development and changes in water chemistry, microbial community structure, and macroinvertebrate community composition in headwater streams in the PASF system. No quantifiable relationships were identified between the intensity of OG development, water composition, and the composition of benthic macroinvertebrate and microbial communities. No definitive indications that hydraulic fracturing fluid, flowback water, or produced water have entered any of the study streams were found. However, the role of sedimentation related to increased traffic from OG development on unpaved roads, pipelines, and well pads as a stressor of stream microbiota was identified as an important relationship for further investigation.

Shale gas development in the PASF is closely overseen by the Pennsylvania Department of Conservation and Natural Resources (PADCNR) and PADEP, who have implemented a program of frequent inspection to ensure permit requirements and lease provisions, and BMPs (50) are followed. The BMPs employed by PADCNR include holistically evaluating the potential effects of development plans at the landscape level rather than by individual well pad, increasing setback distances to streams and wetlands to 60 m (90 m for EV/HQ streams as defined by ref. 51) from the 30 m required by ref. 83, and defining practices for pad construction, road improvement, and waste management (50). The BOF monitoring program (49) quantifies the effects of shale gas development on PASF, and uses the findings from this program to continually evaluate and update BMPs (49, 50). We note that ~32% of the spills of OG-related materials across the state of Pennsylvania occurred within 100 m of a stream (84), and future work applying the methods described here to multiple streams

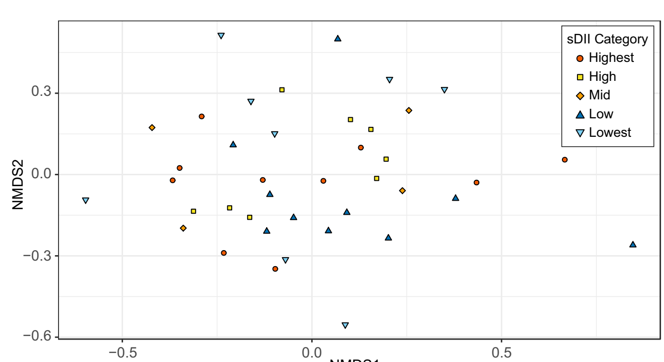

Fig. 6. Plot of NMDS analysis of Bray–Curtis distances among macroinvertebrate community structures based on relative abundance data. No significant differences were observed among the sDII categories.

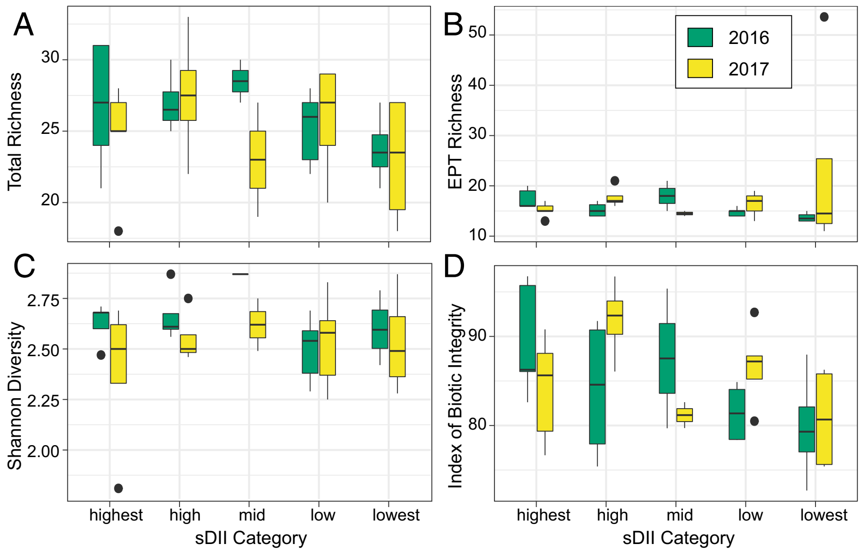

Fig. 7. Macroinvertebrate community composition metrics based on relative abundance. (*A*) Total richness, (*B*) sensitive EPT taxa richness (60), (*C*) Shannon diversity, and (*D*) IBI as described by ref. 60. No significant differences were observed in any metrics among the sDII categories.

outside the PASF and in close proximity to spills is warranted. The applicability of the findings of this study and the utility of the sDII to predict impact from OG development to regions under different regulatory regimes and with more varied land uses will require further study.

This region-scale study integrates major and trace stream water and bed sediment geochemistry, measures of microbial community structure, and measures of benthic macroinvertebrate community structure across multiple seasons and years. Previous studies have shown significant effects downstream from discharges from municipal and industrial wastewater treatment plants treating produced water (26, 30) and shale gas wastewater disposal facilities (29), and large spill events (24, 28, 29), but the regional effects of shale gas development on stream environments remain unclear. Previous work has also shown an inability to confidently identify an OG signal when faced with other land uses (8), which are more likely to cooccur in a watershed as size increases.

Our unique study incorporates a stratified random sampling design to assure sufficient representation of a gradient of disturbance calculated from a wide range of parameters, and combines a broad suite of geochemical and biological metrics. It provides a comprehensive, spatially rigorous, temporally controlled geochemical and biological evaluation of headwater watersheds undergoing shale gas development. Our results highlight the importance of understanding the role of spatial and temporal heterogeneity and the need to account for other land uses in explaining variability in water chemistry and biological measures of ecosystem health in regions undergoing energy development.

ACKNOWLEDGMENTS. This study was funded by the USGS Toxic Substances Hydrology Program, the USGS Environmental Health Mission Area, the PADCNR Shale Gas Monitoring Program, and PADEP. Invaluable field support was provided by Anthony Haynie, Dale Gower, Shawn Lehman, Sharon Morris, Laurie Nau, Brett Pifer, Adam Benthem, Trout Unlimited Volunteers, and the staff of the USGS Northern Appalachian Research Laboratory. Critical analytical support was provided by Jeanne Jaeschke, Kalla Fleger, and Mike Doughten, and the USGS ARC Yeti Team. The manuscript was greatly improved by comments and suggestions from Jenna Shelton (USGS) and two anonymous reviewers. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the US government.

1. US Energy Information Administration, "The distribution of U.S. oil and natural gas wells by production rate" (US Energy Information Administration, Washington, DC, 2018).
2. M. Zoback, S. Kitasei, B. Copithorne, *Addressing the Environmental Risks from Shale Gas Development* (Worldwatch Institute, 2010).
3. S. Entrekin, M. Evans-White, B. Johnson, E. Hagenbuch, Rapid expansion of natural gas development poses a threat to surface waters. *Front. Ecol. Environ.* **9**, 503–511 (2011).
4. D. J. Rozell, S. J. Reaven, Water pollution risk associated with natural gas extraction from the Marcellus Shale. *Risk Anal.* **32**, 1382–1393 (2012).
5. A. Vengosh, R. B. Jackson, N. Warner, T. H. Darrah, A. Kondash, 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 (2014).
6. S. A. Entrekin *et al*., Stream vulnerability to widespread and emergent stressors: A focus on unconventional oil and gas. *PLoS One* **10**, e0137416 (2015).
7. K. O. Maloney *et al*., A detailed risk assessment of shale gas development on headwater streams in the Pennsylvania portion of the Upper Susquehanna River Basin, U.S.A. *Sci. Total Environ.* **610–611**, 154–166 (2018).
8. K. L. Knee, A. E. Masker, Association between unconventional oil and gas (UOG) development and water quality in small streams overlying the Marcellus Shale. *Freshwater Sci.* **38**, 113–130.
9. J. Fichter, K. Johnson, K. French, R. Oden, "Use of microbiocides in Barnett Shale gas well fracturing fluids to control bacteria related problems" in *NACE–International Corrosion Conference Series* (National Association of Corrosion Engineers, New Orleans, LA, 2008).
10. T. J. Gallegos, B. A. Varela, "Data regarding hydraulic fracturing distributions and treatment fluids, additives, proppants, and water volumes applied to wells drilled in the United States from 1947 through 2010" (Data Ser. 868, US Geological Survey, 2015).
11. G. E. King, "Hydraulic fracturing 101: What every representative, environmentalist, regulator, reporter, investor, university researcher, neighbor and engineer should know about estimating frac risk and improving frac performance in unconventional gas and oil wells" in *SPE Hydraulic Fracturing Conference* (Society of Petroleum Engineers, Woodlands, TX, 2012).
12. W. T. Stringfellow, J. K. Domen, M. K. Camarillo, W. L. Sandelin, S. Borglin, Physical, chemical, and biological characteristics of compounds used in hydraulic fracturing. *J. Hazard. Mater.* **275**, 37–54 (2014).
13. D. M. Akob, I. M. Cozzarelli, D. S. Dunlap, E. L. Rowan, M. M. Lorah, Organic and inorganic composition and microbiology of produced waters from Pennsylvania shale gas wells. *Appl. Geochem.* **60**, 116–125 (2015).
14. Blondes MS, *et al*., U.S. Geological Survey National Produced Waters Geochemical Database, v2.3, US Geological Survey data release. https://doi.org/10.5066/F7J964W8. Accessed 5 June 2019.
15. T. T. Phan *et al*., Factors controlling Li concentration and isotopic composition in formation waters and host rocks of Marcellus Shale, Appalachian Basin. *Chem. Geol.* **420**, 162–179 (2016).
16. E. L. Rowan *et al*., Geochemical and isotopic evolution of water produced from Middle Devonian Marcellus shale gas wells, Appalachian basin, Pennsylvania. *AAPG Bull.* **99**, 181–206 (2015).
17. M. A. Engle, I. M. Cozzarelli, B. D. Smith, "USGS investigations of water produced during hydrocarbon reservoir development" (Fact Sheet 2014-3104, US Geological Survey, 2014).
18. M. E. Blauch, R. R. Myers, T. Moore, B. A. Lipinski, N. A. Houston, "Marcellus Shale post-frac flowback waters-where is all the salt coming from and what are the implications?" in *SPE Eastern Regional Meeting* (Society of Petroleum Engineering, Charleston, WV, 2009).

19. E. C. Chapman *et al*., Geochemical and strontium isotope characterization of produced waters from Marcellus Shale natural gas extraction. *Environ. Sci. Technol.* **46**, 3545–3553 (2012).
20. L. O. Haluszczak, A. W. Rose, L. R. Kump, Geochemical evaluation of flowback brine from Marcellus gas wells in Pennsylvania, USA. *Appl. Geochem.* **28**, 55–61 (2013).
21. B. D. Lutz, A. N. Lewis, M. W. Doyle, Generation, transport, and disposal of wastewater associated with Marcellus Shale gas development. *Water Resour. Res.* **49**, 647–656 (2013).
22. B. W. Stewart *et al*., Origin of brines, salts and carbonate from shales of the Marcellus Formation: Evidence from geochemical and Sr isotope study of sequentially extracted fluids. *Appl. Geochem.* **60**, 78–88 (2015).
23. P. F. Ziemkiewicz, Y. Thomas He, Evolution of water chemistry during Marcellus Shale gas development: A case study in West Virginia. *Chemosphere* **134**, 224–231 (2015).
24. N. E. Lauer, J. S. Harkness, A. Vengosh, Brine spills associated with unconventional oil development in North Dakota. *Environ. Sci. Technol.* **50**, 5389–5397 (2016).
25. N. R. Warner *et al*., New tracers identify hydraulic fracturing fluids and accidental releases from oil and gas operations. *Environ. Sci. Technol.* **48**, 12552–12560 (2014).
26. N. R. Warner, C. A. Christie, R. B. Jackson, A. Vengosh, Impacts of shale gas wastewater disposal on water quality in western Pennsylvania. *Environ. Sci. Technol.* **47**, 11849–11857 (2013).
27. N. R. Warner *et al*., Geochemical evidence for possible natural migration of Marcellus Formation brine to shallow aquifers in Pennsylvania. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 11961–11966 (2012).
28. I. M. Cozzarelli *et al*., Environmental signatures and effects of an oil and gas wastewater spill in the Williston Basin, North Dakota. *Sci. Total Environ.* **579**, 1781–1793 (2017).
29. D. M. Akob *et al*., Wastewater disposal from unconventional oil and gas development degrades stream quality at a West Virginia injection facility. *Environ. Sci. Technol.* **50**, 5517–5525 (2016).
30. T. J. Geeza, D. P. Gillikin, B. McDevitt, K. Van Sice, N. R. Warner, Accumulation of Marcellus Formation oil and gas wastewater metals in freshwater mussel shells. *Environ. Sci. Technol.* **52**, 10883–10892 (2018).
31. J. R. Chen See *et al*. Bacterial biomarkers of Marcellus Shale activity in Pennsylvania. *Front. Microbiol.*, **9**, 1697 (2018).
32. R. Trexler *et al*., Assessing impacts of unconventional natural gas extraction on microbial communities in headwater stream ecosystems in Northwestern Pennsylvania. *Front. Microbiol.* **5**, 522 (2014).
33. E. Johnson *et al*., Stream macroinvertebrate communities across a gradient of natural gas development in the Fayetteville Shale. *Sci. Total Environ.* **530-531**, 323–332 (2015).
34. A. M. Farag, D. D. Harper, A review of environmental impacts of salts from produced waters on aquatic resources. *Int. J. Coal Geol.* **126**, 157 (2014).
35. D. R. Mount, D. D. Gulley, J. R. Hockett, T. D. Garrison, J. M. Evans, Statistical models to predict the toxicity of major ions to *Ceriodaphnia dubia, Daphnia magna* and *Pimephales promelas* (fathead minnows). *Environ. Toxicol. Chem.* **16**, 2009–2019 (1997).
36. D. M. Papoulias, A. L. Velasco, Histopathological analysis of fish from Acorn Fork Creek, Kentucky, exposed to hydraulic fracturing fluid releases. *Southeast. Nat.* **12**, 92 (2013).
37. N. Wang, J. L. Kunz, D. Cleveland, J. A. Steevens, I. M. Cozzarelli, Biological effects of elevated major ions in surface water contaminated by a produced water from oil production. *Arch. Environ. Contam. Toxicol.* **76**, 670–677 (2019).
38. N. L. Fahrenfeld *et al*., Shifts in microbial community structure and function in surface waters impacted by unconventional oil and gas wastewater revealed by metagenomics. *Sci. Total Environ.* **580**, 1205–1213 (2017).
39. A. C. Mumford, D. M. Akob, J. G. Klinges, I. M. Cozzarelli, Common hydraulic fracturing fluid additives alter the structure and function of anaerobic microbial communities. *Appl. Environ. Microbiol.* **84**, e02729-17 (2018).
40. P. J. Mouser *et al*., Redox conditions alter biodegradation rates and microbial community dynamics of hydraulic fracturing fluid organic additives in soil–groundwater microcosms. *Environ. Eng. Sci.* **33**, 827–838 (2016).
41. D. Kekacs, B. D. Drollette, M. Brooker, D. L. Plata, P. J. Mouser, Aerobic biodegradation of organic compounds in hydraulic fracturing fluids. *Biodegradation* **26**, 271–287 (2015).
42. P. J. Mouser, M. Borton, T. H. Darrah, A. Hartsock, K. C. Wrighton, Hydraulic fracturing offers view of microbial life in the deep terrestrial subsurface. *FEMS Microbiol. Ecol.* **92**, fiw166 (2016).
43. M. A. Cluff, A. Hartsock, J. D. MacRae, K. Carter, P. J. Mouser, Temporal changes in microbial ecology and geochemistry in produced water from hydraulically fractured Marcellus shale gas wells. *Environ. Sci. Technol.* **48**, 6508–6517 (2014).
44. A. Murali Mohan, A. Hartsock, R. W. Hammack, R. D. Vidic, K. B. Gregory, Microbial communities in flowback water impoundments from hydraulic fracturing for recovery of shale gas. *FEMS Microbiol. Ecol.* **86**, 567–580 (2013).
45. A. Murali Mohan *et al*., Microbial community changes in hydraulic fracturing fluids and produced water from shale gas extraction. *Environ. Sci. Technol.* **47**, 13141–13150 (2013).
46. C. G. Struchtemeyer, J. P. Davis, M. S. Elshahed, Influence of the drilling mud formulation process on the bacterial communities in thermogenic natural gas wells of the Barnett Shale. *Appl. Environ. Microbiol.* **77**, 4744–4753 (2011).
47. C. G. Struchtemeyer, M. S. Elshahed, Bacterial communities associated with hydraulic fracturing fluids in thermogenic natural gas wells in North Central Texas, USA. *FEMS Microbiol. Ecol.* **81**, 13–25 (2012).
48. N. Ulrich *et al*., Response of aquatic bacterial communities to hydraulic fracturing in Northwestern Pennsylvania: A five-year study. *Sci. Rep.* **8**, 5683 (2018).
49. Pennsylvania Department of Conservation and Natural Resources Bureau of Forestry, "Shale gas monitoring report" (Pennsylvania Department of Conservation and Natural Resources, Harrisburg, PA, 2018).
50. Pennsylvania Department of Conservation and Natural Resources Bureau of Forestry, *Guidelines for Administering Oil and Gas Activity on State Forest Lands* (Pennsylvania Department of Conservation and Natural Resources, ed. 4, Harrisburg, PA, 2016).
51. Designated water uses and water quality criteria, 25 Pennsylvania Code §93.9 (2017).
52. Pennsylvania Department of Conservation and Natural Resources Bureau of Forestry "Penn's Woods—Sustaining our forests" (Pennsylvania Department of Conservation and Natural Resources, Harrisburg, PA, 2002).
53. Maloney KO (2018) Shale gas data used in development of the Disturbance Intensity Index for the Pennsylvania portion of the Upper Susquehanna River basin in Maloney et al. 2018. https://doi.org/10.5066/F7Z036NF. Accessed 3 June 2019.
54. C. A. Cravotta, Dissolved metals and associated constituents in abandoned coal-mine discharges, Pennsylvania, USA. Part 1: Constituent quantities and correlations. *Appl. Geochem.* **23**, 166–202 (2008).
55. Pennsylvania Department of Environmental Protection, Abandoned mine land inventory sites 2019. http://www.pasda.psu.edu/uci/DataSummary.aspx?dataset=460. Accessed 3 June 2019.
56. V. M. Heilweil *et al*., A stream-based methane monitoring approach for evaluating groundwater impacts associated with unconventional gas development. *Ground Water* **51**, 511–524 (2013).
57. K. B. Haase *et al*., Dataset of trace dissolved hydrocarbons in surface water and groundwater in North Dakota, Pennsylvania, Virginia, and West Virginia between 2014 and 2017. https://doi.org/10.5066/P9RDPWXO. Accessed 5 June 2019.
58. US Environmental Protection Agency, "Method 3051A: Microwave assisted acid digestion of sediments, sludges, soils and oils" (US Environmental Protection Agency, 2007).
59. D. S. Shull, M. J. Lookenbil, *Water Quality Monitoring Protocols for Streams and Rivers* (Pennsylvania Department of Environmental Protection, Harrisburg, PA, 2018).
60. D. S. Shull, M. Pulket, *Assessment Methodology for Rivers and Streams* (Pennsylvania Department of Environmental Protection, Harrisburg, PA, 2018).
61. D. R. Shull, M. J. Lookenbill, Assessing the expansion of wadeable benthic macroinvertebrate collection methods in large semiwadeable rivers. *Freshw. Sci.* **36**, 683–691 (2017).
62. R Core Team, R: A Language and Environment for Statistical Computing (Version 3.3.0, R Foundation for Statistical Computing, Vienna, Austria, 2016).
63. P. J. McMurdie, S. Holmes, phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* **8**, e61217 (2013).
64. J. Oksanen *et al*., vegan: Community Ecology Package, R package Version 2.5-6. https://CRAN.R-project.org/package=vegan. Accessed 21 January 2019.
65. H. Wickham, R. François, L. Henry, K. Mülle, dplyr: A Grammar of Data Manipulation, R package Version 0.8.3. https://CRAN.R-project.org/package=dplyr. Accessed 21 January 2019.
66. P. D. Schloss *et al*., Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**, 7537–7541 (2009).
67. C. Quast *et al*., The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2013).
68. P. Yilmaz *et al*., The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Res.* **42**, D643–D648 (2014).
69. E. Pruesse *et al*., SILVA: A comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* **35**, 7188–7196 (2007).
70. A. M. Kozlov, A. J. Aberer, A. Stamatakis, ExaML version 3: A tool for phylogenomic analyses on supercomputers. *Bioinformatics* **31**, 2577–2579 (2015).
71. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
72. C. Lozupone, M. E. Lladser, D. Knights, J. Stombaugh, R. Knight, UniFrac: An effective distance metric for microbial community comparison. *ISME J.* **5**, 169–172 (2011).
73. A. C. Mumford, D. M. Akob, Multivariate analysis of shale gas development on the chemical and biological health of headwater streams. NCBI GenBank Database. https://www.ncbi.nlm.nih.gov/bioproject/544240. Deposited 22 May 2019.
74. US Geological Survey, USGS Water Data for the Nation, National Water Information System: Web Interface. http://waterdata.usgs.gov/nwis/. Accessed 28 June 2019.
75. A. C. Mumford, Sediment composition data from northern Pennsylvania: US Geological Survey data release. https://doi.org/10.5066/P9GJTRYR. Accessed 4 October 2019.
76. S. L. Brantley *et al*., Water resource impacts during unconventional shale gas development: The Pennsylvania experience. *Int. J. Coal Geol.* **126**, 140–156 (2014).
77. M. A. Engle, E. L. Rowan, Interpretation of Na–Cl–Br systematics in sedimentary basin brines: Comparison of concentration, element ratio, and isometric log-ratio approaches. *Math. Geosci.* **45**, 87–101 (2013).
78. B. M. Weigel *et al*., Relative influence of variables at multiple spatial scales on stream macroinvertebrates in the Northern Lakes and Forest ecoregion, U.S.A. *Freshw. Biol.* **48**, 1440–1461 (2003).
79. R. D. Vidic, S. L. Brantley, J. M. Vandenbossche, D. Yoxtheimer, J. D. Abad, Impact of shale gas development on regional water quality. *Science* **340**, 1235009 (2013).
80. E. Barbot, N. S. Vidic, K. B. Gregory, R. D. Vidic, Spatial and temporal correlation of water quality parameters of produced waters from Devonian-age shale following hydraulic fracturing. *Environ. Sci. Technol.* **47**, 2562–2569 (2013).
81. E. Barth-Naftilan, J. Sohng, J. E. Saiers, Methane in groundwater before, during, and after hydraulic fracturing of the Marcellus Shale. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 6970–6975 (2018).
82. Water quality standards, 25 Pennsylvania Code §93 (2017).
83. Well location restrictions, 58 Pennsylvania Code §3215 (2012).
84. K. O. Maloney *et al*., Unconventional oil and gas spills: Materials, volumes, and risks to surface waters in four states of the U.S. *Sci. Total Environ.* **581–582**, 369–377 (2017).