

DISSERTATION

MACROINVERTEBRATE AND MICROBIAL COMMUNITY RESPONSES TO METAL
STRESS AS MEASURES OF ECOLOGICAL RECOVERY OF A RESTORED HEADWATER
STREAM

Submitted by

Brian A. Wolff

Graduate Degree Program in Ecology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2020

Doctoral Committee:

Advisor: William H. Clements

Co-advisor: Edward K. Hall

Boris C. Kondratieff

Travis S. Schmidt

Copyright by Brian A. Wolff 2020

All Rights Reserved

ABSTRACT

MACROINVERTEBRATE AND MICROBIAL COMMUNITY RESPONSES TO METAL STRESS AS MEASURES OF ECOLOGICAL RECOVERY OF A RESTORED HEADWATER STREAM

Metal contamination from mining or natural weathering is a common feature of surface waters in the American west. The upper Arkansas River is a high elevation Rocky Mountain stream that was added to the EPA Priorities ('Superfund') list in 1983 due to chronic metals pollution from the California Gulch site near Leadville, Colorado. Water treatment infrastructure and removal of mine tailings reduced metals contamination in the late-1990s and total abundance and richness of aquatic macroinvertebrate communities subsequently recovered to reference site conditions. Despite improvements in abundance and richness, other metrics (e.g., community composition and resilience to metals stress) suggest that the upper Arkansas River may not be fully recovered, and these communities may have undergone a regime shift to an alternative stable state set of ecological conditions.

In Chapter 1, we report results of a long-term (28 year) observational study and a series of stream mesocosm experiments that investigated a potential regime shift in a mining-contaminated watershed. We tested the hypothesis that establishment of a novel, metal-tolerant benthic invertebrate community in a previously contaminated stream impeded recovery of sensitive invertebrate species, despite significant improvements in water quality over the past two decades. Although individual abundance and species richness recovered following

reductions in metal concentrations, we observed persistent changes in community composition and trophic structure downstream from a former source of metal contamination. Mesocosm experiments conducted with reference and downstream invertebrate communities demonstrated that novel communities retained their tolerance to metals but were significantly more sensitive to other stressors. We suggest that the failure of downstream communities to recover represented a contaminant-induced regime shift that resulted from the sustained dominance of metal-tolerant species. Stream restoration ecologists generally consider increased species diversity or abundance of recreationally/commercially important species as indicators of restoration success. However, few stream restoration projects have quantified ecological resilience, and none have included a significant experimental component. Our finding that novel communities in a previously metal-contaminated watershed retained their tolerance to metals but were more sensitive to other stressors has important implications for how restoration success is defined. In systems where restoration to pre-disturbance conditions is unlikely, a better understanding of the ecological resilience of novel communities may be critical for assessing the restoration success. We suggest that resilience to novel stressors is an important indicator of restoration effectiveness that may be applicable in other aquatic ecosystems.

Traditionally, stream macroinvertebrate community metrics have been used for stream quality assessments. Advances in microbial analyses have created the potential for routine sampling of aquatic microbiomes as a tool to assess the quality of stream habitat. In Chapter 2, we sought to determine if microbiome diversity and membership were affected by metal contamination in a manner similar to what has been observed for stream macroinvertebrates, and if so, identify candidate microbial taxa to be used to indicate metal stress in stream ecosystems. To address this, we evaluated microbiome membership from sediments at multiple sites within

the principal drainage of an EPA superfund site near the headwaters of the Upper Arkansas River, Leadville, CO. From each sample, we extracted DNA and sequenced the 16S rRNA gene amplicon on the Illumina MiSeq platform. We used the remaining sediments from each sample to simultaneously evaluate environmental metal concentrations. To control for confounding factors in the field, we also conducted an artificial stream mesocosm experiment using sediments collected from two of the observational study sites. The mesocosm experiment had a 2x2 factorial design: 1) location (upstream or downstream of contaminating tributary), and 2) treatment (metal exposure or control). We found that stream microbiomes showed no difference in alpha diversity between upstream and downstream sites in the field. Similarly, in the mesocosm experiments, alpha diversity changed very little following experimental metal exposure. However, microbiome membership differed between upstream and downstream locations and experimental metal exposure changed microbiome membership in a manner that depended on origin of the sediments used in each mesocosm. Our results suggest that microbiomes can be reliable indicators of ecosystem metal stress even when surface water chemistry and other metrics used to assess ecosystem health do not indicate stream stress.

Lastly, a large-scale restoration project designed to improve habitat for Brown trout (*Salmo trutta*) populations was completed in the upper Arkansas River watershed. In Chapter 3, we examined how aquatic macroinvertebrate communities responded to this habitat restoration since they are a key prey resource to brown trout in this system. Over an eight-year period (2009 – 2017), we measured benthic macroinvertebrate communities, export of aquatic macroinvertebrates from the river (emergence traps), inputs from riparian habitats (pan traps), and diets of Brown trout in response to physical habitat restoration efforts. Among these sampling techniques we found very similar effects – an initial decrease in abundances

immediately following/during restoration (2014), but abundances generally recovered and remained similar to pre-restoration conditions. However, fish diets had a notable trend increasing over baseline conditions in the last two years of monitoring (2016 and 2017). Additionally, we found that overall community composition remained unchanged between pre- and post-restoration treatments. Our results suggest that physical habitat restoration provided limited benefit in terms of abundance and community composition of macroinvertebrate communities, and that other factors such as water quality or continued presence of metal-tolerant competitors may still be important drivers in this system.

ACKNOWLEDGMENTS

This dissertation research has only been possible because of support of my committee and family. Both of my advisors have been exceptionally helpful and patient throughout my time as a Ph.D. student. I admired Will Clements' research – and general attitude toward life – so much that he was my only pick an advisor for a Ph.D. I am incredibly honored that Will agreed to accept me as a student, and he has continued to support me personally and professionally ever since. When you are in Will's lab it truly feels like being in a second family and I will cherish the times that we had working together. Will also allowed me to incorporate microbial ecology into my research – something that both of us knew very little about at the time. However, I eventually convinced Ed Hall – my current co-advisor – to help me develop that research. I credit Ed for constantly challenging me to pursue greatness in all of my efforts, and he has shaped much of my outlook to research and life in general. I am very grateful to my advisors for providing me with numerous opportunities through professional workshops and meetings that expanded both my education and professional network. I also want to acknowledge the other members of my committee for their tireless support. Boris Kondratieff was always willing to lend a hand without hesitation and he spent hours and hours helping me insect identifications. Without Boris' help I am certain that I would still be looking at a microscope! Travis Schmidt was tremendously helpful with providing ideas and he also generously shared much of his work from the USGS that will undoubtedly contribute greatly to ecotoxicological science. Last, but certainly not least, I would like to thank my loving wife, Megan, and our two wonderful children, Hallie and Landon. Pursing a Ph.D. with children is extremely difficult, but their love and support were some of the strongest motivating factors to keep me pushing to the finish.

TABLE OF CONTENTS

ABSTRACT.....	ii
ACKNOWLEDGMENTS	vi
CHAPTER 1	1
INTRODUCTION.....	1
METHODS.....	4
<i>Study site</i>	4
<i>Long-term monitoring</i>	4
<i>Mesocosm experiments</i>	5
<i>Statistical analyses</i>	7
RESULTS.....	8
<i>Field observations in the Arkansas River</i>	8
<i>Mesocosm experiments</i>	9
DISCUSSION.....	11
<i>Has the Arkansas River undergone a regime shift?</i>	13
<i>What is an appropriate criterion for restoration success?</i>	15
TABLES AND FIGURES	18
LITERATURE CITED	26
CHAPTER 2	34
INTRODUCTION.....	34
METHODS.....	37
<i>Study Site</i>	37
<i>Experimental Mesocosm</i>	38
<i>DNA preparation and 16S rRNA amplicon sequencing</i>	39
<i>Metals preparation</i>	40
<i>Biomonitoring statistics</i>	41
<i>Mesocosm Statistics</i>	42
RESULTS.....	42
<i>Observational Study</i>	42
<i>Mesocosm Study</i>	45
<i>Genera-level responses between observational and mesocosm studies</i>	46
DISCUSSION.....	47
TABLES AND FIGURES	56
LITERATURE CITED	65
CHAPTER 3	71
INTRODUCTION.....	71

METHODS.....	74
<i>Study Site</i>	74
<i>Benthic macroinvertebrates</i>	75
<i>Aquatic insect emergence</i>	75
<i>Riparian inputs</i>	76
<i>Brown Trout Diets</i>	77
<i>Statistical Analyses</i>	77
RESULTS.....	78
DISCUSSION.....	80
TABLES AND FIGURES	86
LITERATURE CITED	95
APPENDIX.....	101
CHAPTER 1 SUPPLEMENTAL TABLES AND FIGURES.....	101

CHAPTER 1

RESILIENCE AND REGIME SHIFTS: NOVEL BENTHIC COMMUNITIES IMPEDE ECOLOGICAL RECOVERY IN A HISTORICALLY METAL-CONTAMINATED STREAM

Introduction

The likelihood that degraded ecosystems can recover following the removal of a stressor and the length of time required for these systems to return to pre-disturbance conditions are critical questions in applied ecology (Millennium Ecosystem Assessment 2005). Stream restoration projects that are designed to improve water or habitat quality offer unique opportunities to address these questions (Young et al. 2005). However, despite investments exceeding \$1 billion (U.S.) per year, our ability to quantify the effectiveness of stream restoration projects remains limited (Bernhardt et al. 2005). Poor study designs (Bernhardt et al. 2005), limited spatiotemporal scale (Bernhardt et al. 2007), the focus on habitat instead of water quality (Palmer et al. 2010) and failure to consider ecological theory (Palmer et al. 1997; Lake et al. 2007) impede our ability to learn from these potential management experiments (Gunn et al. 2010). In addition, there remains considerable uncertainty regarding which specific ecological measures are most appropriate for quantifying restoration effectiveness. While increased species diversity or abundance of recreationally important species are considered indicators of restoration success, few projects also measure ecosystem processes (e.g., decomposition; export of riparian subsidies). Although there is an assumed relationship between species diversity and ecological processes in restored terrestrial systems (Ruiz-Jaen and Aide 2005), with few exceptions (e.g., Cardinale et al. 2000; Cardinale and Palmer 2002), this relationship has not

been rigorously tested in freshwater ecosystems. Finally, there is debate among ecologists regarding restoration goals in novel or hybrid communities, defined as new combinations of species resulting from anthropogenic stressors (Hobbs et al. 2006; Hobbs 2007; Hobbs et al. 2009). Returning novel communities to reference conditions may not be possible after crossing a critical threshold and shifting to an alternative state.

Measures of species richness and abundance are among the most commonly employed indicators of ecological condition in stream bioassessments (Simon 2003). However, ecologists also recognize that trophic composition or changes in community structure in space and time (e.g., beta diversity) may be more sensitive indicators of anthropogenic perturbations than the number of species. For example, taxonomic richness may be an ineffective indicator of disturbance because species replacement can mask spatial and temporal changes the total number of species (De Laender et al. 2012; Dornelas et al. 2014). Furthermore, disturbed communities are often dominated by cosmopolitan species (Magurran 2016), resulting in biotic homogenization, reduced functional diversity and loss of ecological resilience (Folke 2004; Cadotte 2011; Mori et al. 2018). Shifts in community composition and loss of ecological resilience have long-term implications for ecosystem recovery following the removal of a stressor. Studies in some systems (e.g., coral reefs) suggest that regime shifts may be very difficult to reverse, even after a stressor has been removed (Bellwood et al. 2004; Hughes et al. 2005). New combinations of species resulting from regime shifts may provide very different ecosystem services (Hobbs et al. 2006; Hobbs et al. 2009; Oliver et al. 2015) and therefore have important implications for how ecologists define targets for restoration.

Long-term observational studies have the potential to identify regime shifts in disturbed ecosystems; however, experimental approaches may be necessary to demonstrate cause-and-

effect relationships (Scheffer and Carpenter 2003; Schroder 2005). Although some investigators have identified regime shifts in stream ecosystems in response to a stressor (Rocha et al. 2015), none have included a substantial experimental component. In this research we report results of a long-term (28 year) observational study complemented by a series of stream mesocosm experiments that investigated a potential regime shift in a mining-contaminated watershed. We focus on metal pollution associated with current and historical mining operations, which is a significant global environmental problem (Hudson-Edwards 2016). Although catastrophic discharges of mining wastes such as the 2015 Minas Gerais (Brazil) tailings dam failure and the Animas River (Colorado, USA) spill have significant effects on individual watersheds, elevated levels of metals from abandoned mines often have much broader regional impacts (Jones et al. 2013). For example, a survey of 95 randomly selected sites in Colorado's southern Rocky Mountain ecoregion showed that 23% of these streams were affected by metals from abandoned mines (Clements et al. 2000).

The primary goal of this research was to test the hypothesis that benthic communities in a previously metal-polluted watershed (Arkansas River, CO, USA) have undergone a regime shift. A second objective was to quantify responses of these communities to a novel anthropogenic stressor (a simulated diesel fuel spill). Building on previous work (Courtney and Clements 2000; Kashian et al. 2007; Clements et al. 2008), we hypothesized that ecological resilience, defined here as the ability of a system to withstand disturbance (Holling 1973), is an appropriate indicator of restoration effectiveness in streams recovering from long-term exposure to contaminants.

Methods

Study site

The field study was conducted in the headwaters of the Arkansas River, located near the town of Leadville and approximately 100 km west of Denver, Colorado (USA). Mining in the watershed began in the mid-1800s, resulting in significant degradation of water quality, habitat and aquatic communities (Clements et al. 2010). In 1983 California Gulch, a tributary to the Arkansas River, was listed as a U.S. EPA Superfund Site. A large-scale restoration project completed in 2013 was designed to improve water quality and restore a productive brown trout (*Salmo trutta*) fishery to the watershed. From 1989 to 2017 we monitored responses of benthic macroinvertebrate communities, the primary prey of brown trout, to several restoration treatments. By 2000, significant improvements in water quality, macroinvertebrate communities and brown trout populations were observed following the completion of water treatment facilities, removal of waste tailings, and bank stabilization (Clements et al. 2010). The final phase of the project, which focused on improvements in stream habitat and riparian vegetation, was finished in 2013. In 2014 this reach of the Arkansas River was identified as a “gold medal” trout stream, a designation intended to recognize the highest quality fisheries in Colorado.

Long-term monitoring

Details of the field sampling methods and study sites have been published previously (Clements et al. 2010). Benthic macroinvertebrate samples were collected from sites upstream and downstream from California Gulch (CG), the primary source of metals to the system. Although upstream sites historically received some input of metals from the Leadville Mine Drainage Tunnel (LMDT), this source of contamination was eliminated in 1992. Replicate (n = 5

per site), quantitative samples were collected annually in fall and spring from 1989 to 2017 using a Hess sampler (0.1 m²). Samples were rinsed through a 350 µm sieve in the field and materials retained were preserved in 80% ethanol. In the laboratory, samples were sorted and identified to the lowest practical level of taxonomic resolution (genus for most aquatic insects; family for chironomids). Benthic data were analyzed across 3 time periods: before (1989-2009) and after completion of water treatment facilities (2000-2008) and habitat restoration (2009-2017). Water samples for determination of Zn, the primary contaminant in this system, were collected on each sampling occasion, filtered through a 0.45 µm filter and preserved in nitric acid. Water samples were analyzed using flame or furnace atomic adsorption spectrophotometry.

Mesocosm experiments

To compare the responses of benthic communities from sites upstream and downstream of historical metal inputs, five stream mesocosm experiments were conducted between 2009 and 2017. Four of these experiments exposed communities to metals and a 5th experiment simulated a spill of diesel fuel, a novel stressor in this system. Details of the experimental approach and the mesocosm facility have been described previously (Clements 2004). Experiments were conducted using natural benthic communities that colonized plastic trays (10 x 10 x 6 cm) filled with pebble and small cobble substrate. Our previous experiments have shown that trays are colonized by a diverse assemblage of aquatic insects (25-40 taxa per tray) and that communities on trays are quite similar to those collected from the natural substrate (Courtney and Clements 2002; Clements et al. 2013). For each experiment, trays were deployed at an upstream reference site and a site downstream from the former CG Superfund Site. The upstream and downstream sites (AR1 and AR5, Clements et al. 2010) were separated by approximately 12 km. Colonized

trays were collected after 30-40 d, placed in individual coolers filled with ambient stream water (4 trays per cooler), and transported to the Stream Research Laboratory (SRL) at Colorado State University, USA. The contents of each cooler were randomly assigned to one of 18 stream mesocosms. Water supplied to the mesocosms originated from a deep, mesotrophic reservoir and was delivered to each 20 L stream at a rate of 1.0 L/min. Paddlewheels maintained a constant current velocity of 0.35 m/s.

For each experiment we used a 2 X 2 factorial design with site (upstream versus downstream) and metal treatment (control versus metals) as the main effects. Control streams were replicated four times and treatment streams were replicated 4-5 times. Peristaltic pumps delivered stock solutions of metals from 20-L carboys at a rate of 10-mL/min to maintain target concentrations of 25 µg Cu/L and 650 µg Zn/L in treated streams. Routine physicochemical measurements (temperature, dissolved oxygen, pH and specific conductance) were measured every 2-3 d using hand held meters (models 550A and 63; YSI Incorporated, Yellow Springs, Ohio). Water hardness and alkalinity, factors that determine metal bioavailability, were measured in the laboratory using standard titration procedures. Filtered (0.45-µm) water samples were collected from each stream on 2-3 occasions to verify concentrations of metals and analyzed as described above. After 10 d the four trays in each mesocosm were removed, rinsed through a 350 µm sieve and organisms were preserved in 80% ethanol. Benthic samples were processed in the laboratory as described above. Mesh nets covering the standpipes prevented emigration of organisms from the flow-through mesocosms and therefore differences between control and treated streams at the end of the experiment were attributed to mortality.

To quantify resilience of benthic communities to a novel stressor, an additional mesocosm experiment was conducted in fall 2016 that simulated a diesel fuel spill. We were

interested in the effects of diesel exposure because of a recent spill that occurred in West Creek, a tributary to the Dolores River in western Colorado (Duggan et al. 2018). We used the same study design (2 sites and 2 treatments) and collected organisms from the same sites as described above, but the mesocosms were modified to accommodate diesel exposure. Mesocosms for this experiment were constructed from stainless steel pans (53 x 32 x 15 cm) in a flow-through design. Current velocity (0.25 m/s) and stream depth (11 cm) were maintained by stainless steel paddle wheels and stand pipes, respectively. The rate of diluent flow to these streams was 0.3 L/min, resulting in water residence time of approximately 40 minutes. In a 2 X 2 factorial design, replicate mesocosms (n = 4) with communities from either the upstream or downstream station were randomly assigned to one of two treatments: 0 or 600 mg/L diesel. After 48 hours of acclimation, diluent flow was briefly terminated and a single pulse of diesel was added to each mesocosm to achieve the target concentration. After 7 days exposure, trays in each stream were removed and the remaining organisms were preserved and processed as described above.

Statistical analyses

In the four mesocosm experiments with metals we used general linear models (PROC GLM; SAS 9.3) to analyze the main effects of source community (upstream versus downstream), treatment and year on benthic communities. The primary goal of this analysis was to test the hypothesis that effects of metals differed between upstream and downstream communities and therefore we were especially interested in the site x treatment interaction term. Because of differences in community composition and abundance of dominant taxa among years, analyses were mostly restricted to abundance of major aquatic insect orders (Ephemeroptera, Plecoptera, Trichoptera and Diptera), total abundance, number of taxa and abundance of functional feeding

groups (e.g., grazers, predators). We used canonical discriminant analysis (PROC CANDISC) to examine separation of macroinvertebrate communities between sites and treatments based on abundance of the 18 dominant taxa. In the mesocosm experiment with diesel, we used a 2-way ANOVA to measure the effects of source community and treatment. Where necessary, abundance values were $\ln(x+1)$ transformed to satisfy assumptions of parametric statistics.

Results

Field observations in the Arkansas River

Changes in Zn concentrations measured at upstream and downstream stations showed considerable annual variation, particularly at the downstream station (**Fig. 1**). Zn concentrations in the upstream reach were reduced in 1992 after completion of the LMDT treatment facility and generally remained below the U.S. EPA chronic criterion value (dashed horizontal line). Zn concentrations were approximately 2-10 times greater at the downstream station and consistently exceeded the EPA chronic criterion value before 2001. Metal concentrations decreased significantly after the water treatment facilities were completed at the CG Superfund Site and were only slightly higher than those at the upstream station after 2008. Between 2009 and 2017, the time period when our mesocosm experiments were conducted, Zn concentration exceeded the U.S. EPA criterion value on only one occasion (2011). These results demonstrate the significant improvements in water quality and overall success of remediation with respect to reducing metal concentrations at the CG Superfund Site.

Benthic communities in the Arkansas River responded to these improvements in water quality. Long-term variation in taxonomic richness and total macroinvertebrate abundance at upstream and downstream sites were similar across the 3 time periods (**Fig. 2**). Mean number of

taxa and abundance increased significantly ($p < 0.0001$) in both reaches after completion of the water treatment facilities in 2000. Total abundance and number of taxa at both stations were greater than values reported in a survey of 108 uncontaminated reference streams in Colorado that used the same sampling methods (Schmidt et al. 2010).

In contrast to the patterns for abundance and taxonomic richness, functional feeding group composition of benthic communities differed markedly between upstream and downstream sites (**Fig. 3**). Differences in the composition of functional feeding groups resulted primarily from the failure of several species of mayflies (Ephemeroptera) and other grazers to recover downstream. These organisms were replaced by filter-feeding caddisflies (Trichoptera) and non-insects (primarily oligochaetes), which accounted for approximately 29% of the macroinvertebrate community at the downstream reach (versus < 10% at the reference site).

Mesocosm experiments

Routine physicochemical characteristics (pH, specific conductance, temperature and dissolved oxygen) in mesocosms did not differ among treatments (**Table S1**) and were similar to those measured in uncontaminated Rocky Mountain streams (Clements et al. 2000; Schmidt et al. 2010). Concentrations of metals measured during the four experiments were similar between treated mesocosms, but were 12-20% lower than target values (**Table S2**). Differences between measured and target concentrations likely resulted from uptake and adsorption of metals during the experiments. Concentrations of metals in treated mesocosms across all experiments were approximately 18 times the U.S. EPA hardness-adjusted criterion value for Cu and Zn.

Analysis of benthic communities showed significant effects of metals on total macroinvertebrates abundance, number of taxa and abundance of mayflies, stoneflies and

dipterans (**Fig. 4**). Results of GLM analyses showed that site, metal treatment, year and the associated interaction terms explained 82-95% of the total variation in abundance (**Table 1**). More importantly, effects of metals were consistently greater on communities from the upstream reference site, as indicated by the significant site x metals interaction term for mayflies, stoneflies and total abundance. Although treatment effects on dipterans (primarily Chironomidae) were also highly significant, responses to metals were similar between upstream and downstream communities. There were no effects of metals on caddisflies, which were dominated by three filter-feeding taxa (*Arctopsyche*, *Hydropsyche* and *Brachycentrus*).

Exposure to metals shifted the composition of most functional feeding groups, but the magnitude of these effects also differed between upstream and downstream communities (**Fig. 5; Table 1**). In particular, we observed a highly significant site x metal interaction effect on grazers ($p = 0.0002$), which were reduced by 72% at the upstream site compared to 43% at the downstream site. No other functional feeding group showed a significant site x treatment interaction, indicating that effects of metals on these groups were similar at upstream and downstream reaches.

Results of multivariate analyses based on abundance of the 18 dominant taxa demonstrated that effects of metals on upstream reference communities were greater than on downstream communities (**Fig. S1**). Although upstream and downstream communities were separated along canonical variable 1, canonical variable 2 significantly ($p < 0.0001$) separated control and metal-treated streams. Separation between controls and treated streams was greater for communities from the upstream station than the downstream station. Separation along canonical variable 2 resulted primarily from greater abundance of mayflies (*Baetis* and

Paraleptophlebia) in control streams and greater abundance of caddisflies (*Arctopsyche* and *Rhyacophila*) in treated streams.

Exposure of Arkansas River benthic communities to diesel fuel, a novel stressor in this system, significantly reduced abundance of all macroinvertebrate groups (**Fig. 6; Table S3**), including those unaffected by metals (e.g., Trichoptera). Interestingly, we observed the greatest effects of diesel exposure on dipterans (primarily chironomids), which were reduced by 85-90% in mesocosms across both sites. Mayflies were the only macroinvertebrates that showed a significant site x treatment interaction ($p = 0.0149$), indicating that the effects of diesel varied between sites. However, in contrast to experiments conducted with metals, mayflies from the downstream site were slightly more sensitive to diesel than those from the upstream reference site. Abundance of mayflies from the downstream site was reduced by 74% compared to a 62% reduction at the upstream site.

Discussion

Results of four separate mesocosm experiments conducted between 2009 and 2017 indicated that benthic communities located downstream from a former U.S. EPA Superfund Site were more tolerant to metals compared to those from the upstream reference site. We observed significantly greater reductions in total macroinvertebrate abundance and greater shifts in community structure and functional feeding group composition at the upstream site after exposure to metals. Shifts in community composition observed in treated mesocosms, particularly the loss of grazing mayflies and increased relative abundance of caddisflies, were consistent with responses observed in the field. Differences in metal tolerance between communities persisted throughout the study, despite significant improvements in water quality

over the past 20 years. Although several macroinvertebrate taxa from the upstream site were more sensitive to metals, the greatest difference between sites was observed for grazing mayflies, organisms known to be especially sensitive to metals (Clements et al. 2000; Mebane et al. 2015; Herbst et al. 2018).

There are at least two potential explanations for the greater tolerance of downstream communities to metals, both of which are related to the historical effects of mining pollution. First, persistent alterations in community structure and trophic composition at the downstream site, especially the dominance of filter-feeding caddisflies, have resulted in greater densities of metal-tolerant taxa. These metal-tolerant organisms have replaced several species of grazing mayflies, organisms which failed to recover despite improvements in water quality (Clements et al. 2010). A second explanation for the different responses between upstream and downstream communities is the greater tolerance of *Baetis tricaudatus* collected below CG. Effects of metals on *B. tricaudatus*, the dominant mayfly in the Arkansas River, were significantly greater on upstream populations (site x metals interaction, $p < 0.0001$). Across the four mesocosm experiments, survival of *B. tricaudatus* from the upstream site was reduced by 88% compared to a 48% reduction at the downstream site. These differences in sensitivity between populations were a result of either physiological factors that regulate metal tolerance (e.g., increased production of metal-binding proteins) and/or genetic differences between mayfly populations. *Baetis* is multivoltine, highly mobile, ubiquitous and the dominant mayfly species in many western streams (Ward et al. 2002; Merritt et al. 2008). Because of these life history traits, it is likely that *B. tricaudatus* can rapidly acclimate or adapt to metals exposure. However, given the potential cost of tolerance to metals (Wilson 1998; Kashian et al. 2007; Zuellig et al. 2007; Vigneron et al. 2015) and the rapid rate at which organisms lose this tolerance after metal levels

have declined (Levinton et al. 2003), it is surprising that these population-level differences have persisted following restoration. We have previously reported that reduced genetic diversity and alterations in metal-responsive genes in *B. tricaudatus* resulted from a population bottleneck (Beaty et al. 1998) downstream from the CG Superfund Site. Although we are unable to distinguish between physiological acclimation and genetic changes in these populations, results suggest that both community-level and population differences influenced responses to metals.

Has the Arkansas River undergone a regime shift?

There are several potential explanations for the failure of downstream communities in the Arkansas River to recover following improvements in water quality. First, it is possible that metal concentrations have not been sufficiently reduced to allow grazing mayflies and other sensitive taxa to recolonize. Although Zn concentrations in the downstream reach declined significantly after 2001 and reached nontoxic concentrations, they occasionally exceeded the established water quality criterion value (Fig. 1). Water quality criteria for metals, defined as the highest concentration that should not to pose a significant risk to most species (U.S. EPA 2002), are established to protect all components of aquatic ecosystems. Although aquatic insects are not well represented in the database used to establish criteria (Brix et al. 2011; Clements et al. 2013; Poteat et al. 2014), these values are considered to be relatively conservative. Our previous mesocosm experiments have shown that the water quality criterion value for Zn is protective of most species, including those known to be sensitive to metals (Clements 2004). Therefore, it is unlikely that the modest exceedances of criterion values for metals over the last 9 years are responsible for the continued alteration of downstream communities.

A second explanation for differences between reaches is that there has not been sufficient time for downstream communities to fully recover. However, macroinvertebrate communities in the upstream reach responded rapidly (within 2 years) to improvements in water (Clements et al. 2010). Furthermore, other investigators have reported rapid recovery of benthic communities in mining polluted streams following reductions in metal inputs (Hornberger et al. 2009; Mebane et al. 2015; Herbst et al. 2018). Therefore, we believe that 9 years is more than sufficient time for downstream communities in the Arkansas River to recover.

A third explanation for the failure of downstream communities to recover after improvements in water quality was the result of a contaminant-induced regime shift and the continued dominance by metal-tolerant species. There is evidence from both aquatic and terrestrial systems that exposure to anthropogenic stressors can reduce community resilience and functional redundancy, thereby increasing the likelihood that the system will shift to an alternative state (Beisner et al. 2003; Scheffer 2003; Bellwood et al. 2004; Folke 2004). The specific mechanism responsible for the failure of grazers to recover in the Arkansas River are not known; however, previous experiments conducted with benthic communities in coldwater streams have demonstrated strong competitive interactions between grazers and filter-feeders (Dudley et al. 1990; Kohler 1992). Identifying regime shifts in natural systems is important because reducing or even eliminating stressors that initially caused the transition may not be sufficient to return a system to pre-disturbance conditions. We hypothesize that the long-term reduction in abundance of metal-sensitive species, their replacement by metal-tolerant species and the shift in trophic composition has produced a novel community that may not return to pre-disturbance conditions.

The concept of novel or hybrid communities (Hobbs et al. 2006) is an emerging paradigm in ecology with important implications for how we define restoration success (Hobbs et al. 2009). While total abundance and number of species may be similar in reference and disturbed systems, species replacement results in novel communities that may provide very different ecosystem services (Dornelas et al. 2014). If a majority of species in a disturbed community have similar responses to environmental perturbations, these communities are likely to show lower resilience to other stressors (Folke 2004; Oliver et al. 2015). A loss of resilience can even occur in systems where species richness is unchanged if communities become homogenized and dominated by cosmopolitan species (Mori et al. 2018). We believe this situation has occurred in the Arkansas River. If restoration of the Arkansas River to pre-disturbance conditions is unachievable because this system has undergone a regime shift, a better understanding of the stability and ecological resilience of this novel community is critical.

What is an appropriate criterion for restoration success?

In situations where study designs allow investigators to quantify restoration effectiveness, the determination of success is often dependent on which specific measures are being evaluated. Reductions in contaminant concentrations and improvements in habitat quality are clear indicators of success (Gunn et al. 2010). Biological responses to these improvements, such as increased macroinvertebrate abundance and diversity, are also appropriate measures of restoration effectiveness, in part because of their potential contribution to community stability (Palmer et al. 2010). The increase in brown trout populations and the designation of the Arkansas River as a Gold Medal fishery in 2014 will have long-lasting socioeconomic benefits. However, despite these improvements, persistent alterations in community composition and greater

sensitivity of downstream communities to novel stressors suggest that this system has not fully recovered. Shifts in the composition of functional feeding groups, such as the elimination of grazers and their replacement by filter-feeders, have the potential to significantly alter ecosystem processes (Wallace et al. 1989). Although there is some evidence of functional redundancy for processes in streams (Cardinale and Palmer 2002), the relationship between species diversity and ecosystem function has not been well established in these systems (Palmer et al. 1997). A meta-analysis of restoration effectiveness in aquatic ecosystems showed that stream restoration generally improved biodiversity, but not ecosystem services (Benayas 2009). These results suggest that maintaining ecosystem processes may require restoring the full complement of species within different functional groups (Lake et al. 2007). Because of the direct connection between functional groups and ecosystem processes, we believe that recovery of functional diversity may be a more useful indicator of restoration effectiveness than improvements in species diversity (Cadotte 2011).

In addition to returning the structural and functional integrity of a community to pre-disturbance conditions, restoring ecological resilience will be especially important under future scenarios of land use alterations, climate change, chronic pollution and other anthropogenic disturbances (Paine et al. 1998). Although ecologists recognize the importance of ecological resilience, quantifying resilience in natural systems is challenging (Oliver et al. 2015; Angeler et al. 2018), largely because of a lack of experimental studies (Scheffer 2003; Schroder 2005). Consequently, most studies simply measure changes of more tractable measures (e.g., species richness) and assume a relationship with community resilience. We believe that experimental approaches are of critical importance for demonstrating cause-and-effect relationships among stressors, community resilience and the potential for regime shifts.

In the current study, mesocosm experiments coupled with long-term biomonitoring provided an opportunity to quantify differences in community resilience. Benthic communities downstream from a historical source of metals pollution retained their tolerance to metals, despite significant reductions in metals exposure and improvements in water quality. We hypothesize that the failure of downstream communities to recover and the continued dominance by metal-tolerant organisms is evidence of a regime shift in this system. More importantly, these novel communities were also more sensitive to other stressors. Results of the current experiments with diesel fuel and previous studies showed that the effects of several biotic and abiotic stressors, including stonefly predation, acidification and UV-B radiation, were consistently greater on metal-tolerant communities (Kiffney 1996; Clements 1999; Courtney and Clements 2000; Kashian et al. 2007; Zuellig et al. 2007; Clements et al. 2008). Thus, in addition to the traditional measures of abundance and species richness, we suggest that experimental quantification of community resilience and sensitivity to novel stressors are appropriate measures of restoration effectiveness in contaminated ecosystems.

Tables and Figures

Table 1. Results of ANOVA testing for effects of metals, year, site and the interaction terms for each main effect in the model (none of the 3-way interaction terms were significant) in stream mesocosms. Table shows F-values and p-values (in parentheses) for abundance of major macroinvertebrate orders, total abundance, total richness and abundance of the five functional feeding groups.

Variable	R ²	Year x					
		Metals	Year	Site	Site x Metals	Year x Metals	Site
Ephemeroptera	0.95	208.34 (<0.0001)	106.35 (<0.0001)	120.14 (<0.0001)	6.94 (0.011)	4.14 (0.0103)	84.59 (<0.0001)
Plecoptera	0.87	3.32 (0.0741)	25.62 (<0.0001)	103.65 (<0.0001)	3.7 (0.0597)	0.5 (0.685)	55.94 (<0.0001)
Trichoptera	0.90	0.83 (0.3671)	94.19 (<0.0001)	34.96 (<0.0001)	0.11 (0.7451)	0.77 (0.5137)	52.71 (<0.0001)
Diptera	0.89	98.64 (<0.0001)	64.3 (<0.0001)	47.38 (<0.0001)	0.58 (0.4483)	0.48 (0.6962)	29.28 (<0.0001)
Total abundance	0.93	186.33 (<0.0001)	121.77 (<0.0001)	88.44 (<0.0001)	6.85 (0.0115)	0.26 (0.8531)	27.69 (<0.0001)
Total taxa	0.82	4.55 (0.0376)	12.31 (<0.0001)	124.86 (<0.0001)	0.03 (0.8571)	4.04 (0.0115)	22.63 (<0.0001)
Grazers	0.94	195.23 (<0.0001)	125.16 (<0.0001)	24.04 (<0.0001)	16.14 (0.0002)	3.05 (0.0362)	85.64 (<0.0001)
Filterers	0.90	0.07 (0.7996)	85.8 (<0.0001)	48.76 (<0.0001)	0.04 (0.8406)	1.75 (0.1683)	47.34 (<0.0001)
Shredders	0.79	0.02 (0.8921)	25.09 (<0.0001)	3.67 (0.0607)	0.61 (0.4389)	1.24 (0.3054)	40.7 (<0.0001)
Gatherers	0.90	115.12 (<0.0001)	67.85 (<0.0001)	100.59 (<0.0001)	0.24 (0.6294)	0.42 (0.7423)	21.21 (<0.0001)
Predators	0.87	0.01 (0.9033)	40.00 (<0.0001)	144.29 (<0.0001)	2.25 (0.1394)	1.37 (0.2628)	34.08 (<0.0001)

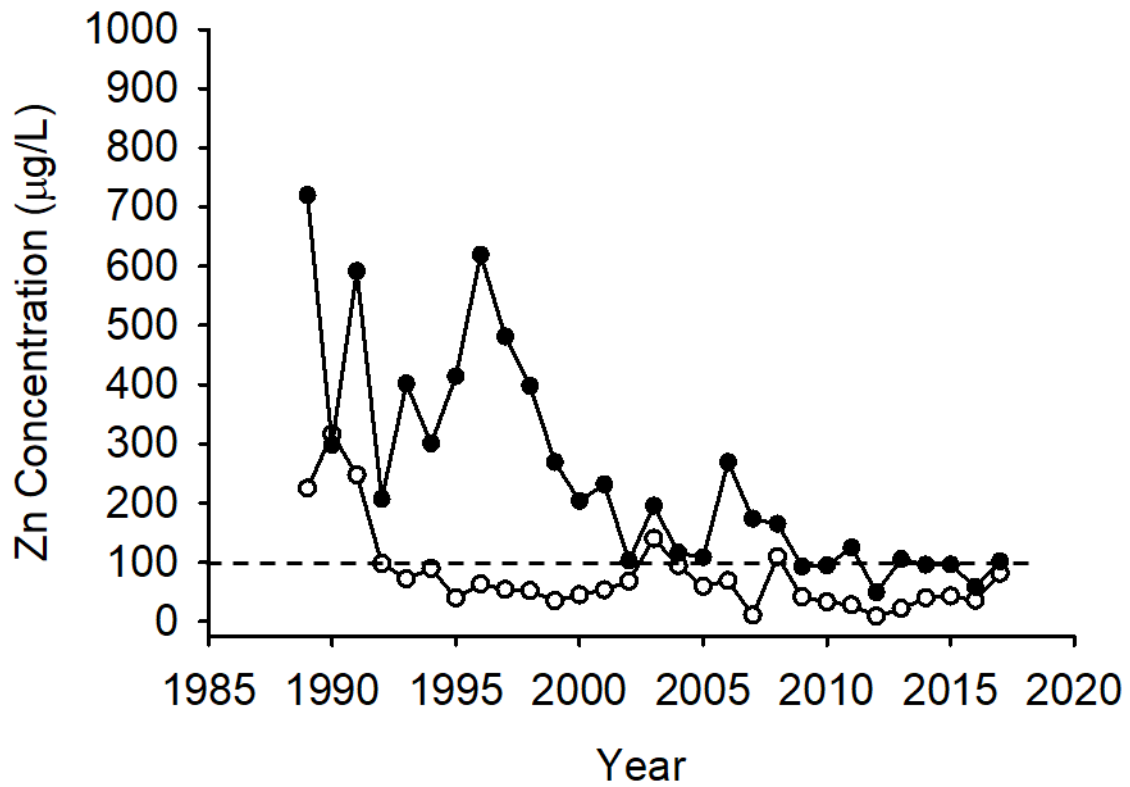


Figure 1. Long-term (1989-2017) changes in Zn concentration measured at upstream (open circles) and downstream (closed circles) stations in the Arkansas River. Dashed horizontal line represents the U.S. EPA chronic criterion value for Zn.

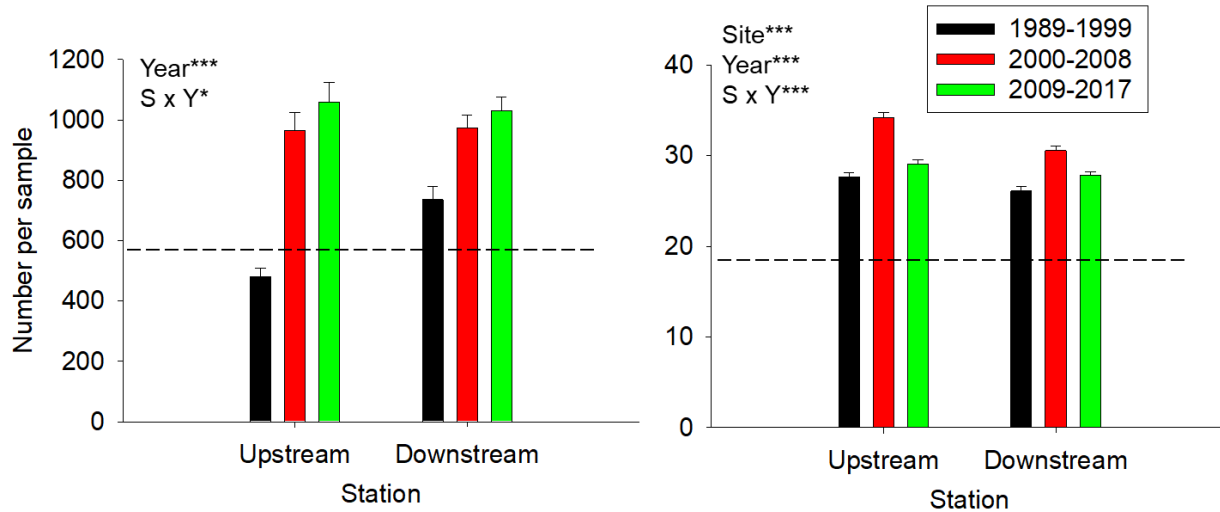


Figure 2. Long-term (1989-2017) changes mean (+s.e.) macroinvertebrate abundance (left panel) and number of taxa (right panel) at upstream and downstream reaches in the Arkansas River. Data summarize 3 time periods: before (1989-2009) and after completion of water treatment facilities (2000-2008) and habitat restoration (2009-2017). Results of 2-way ANOVA testing for the effects of year, site and the site x year interaction are shown (* $p < 0.05$; *** $p < 0.0001$). Horizontal lines represent mean abundance and number of taxa reported from a survey of 108 Colorado reference streams that employed the same sampling methods as in the present study (Schmidt et al. 2010).

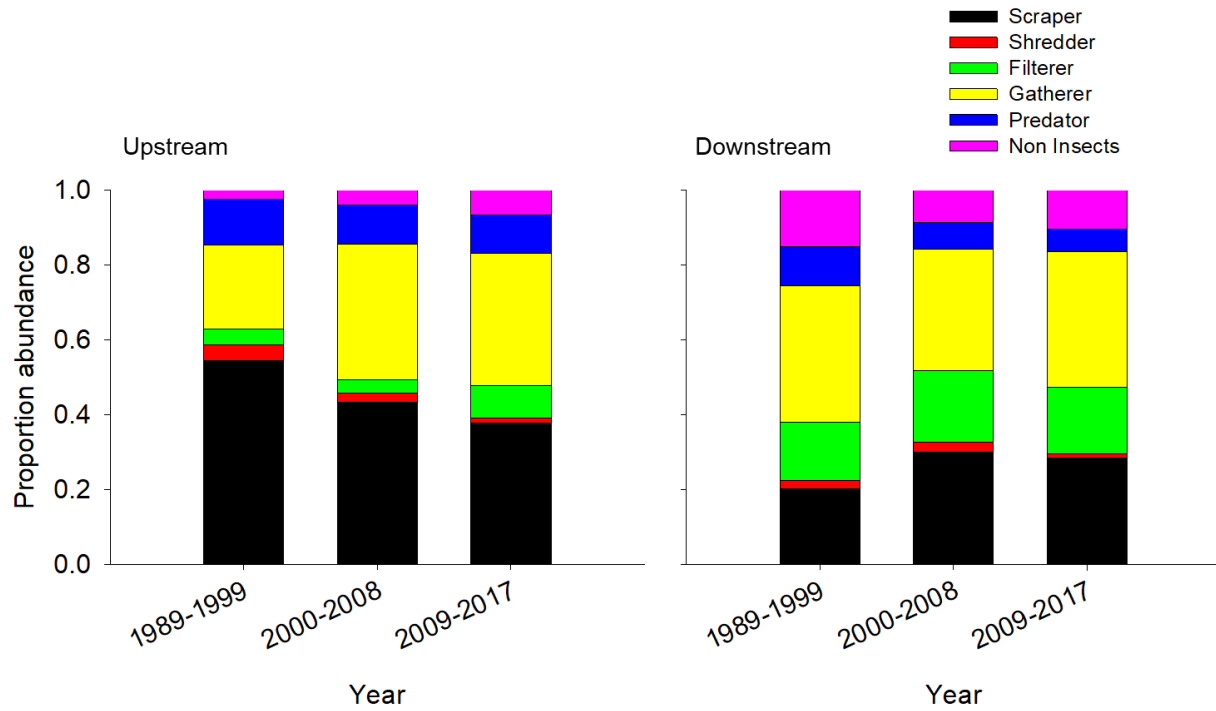


Figure 3. Long-term (1989-2017) changes in the proportion of the major functional feeding groups and non-insects at upstream and downstream reaches in the Arkansas River. Data summarize 3 time periods: before (1989-2009) and after completion of water treatment facilities (2000-2008) and habitat restoration (2009-2017).

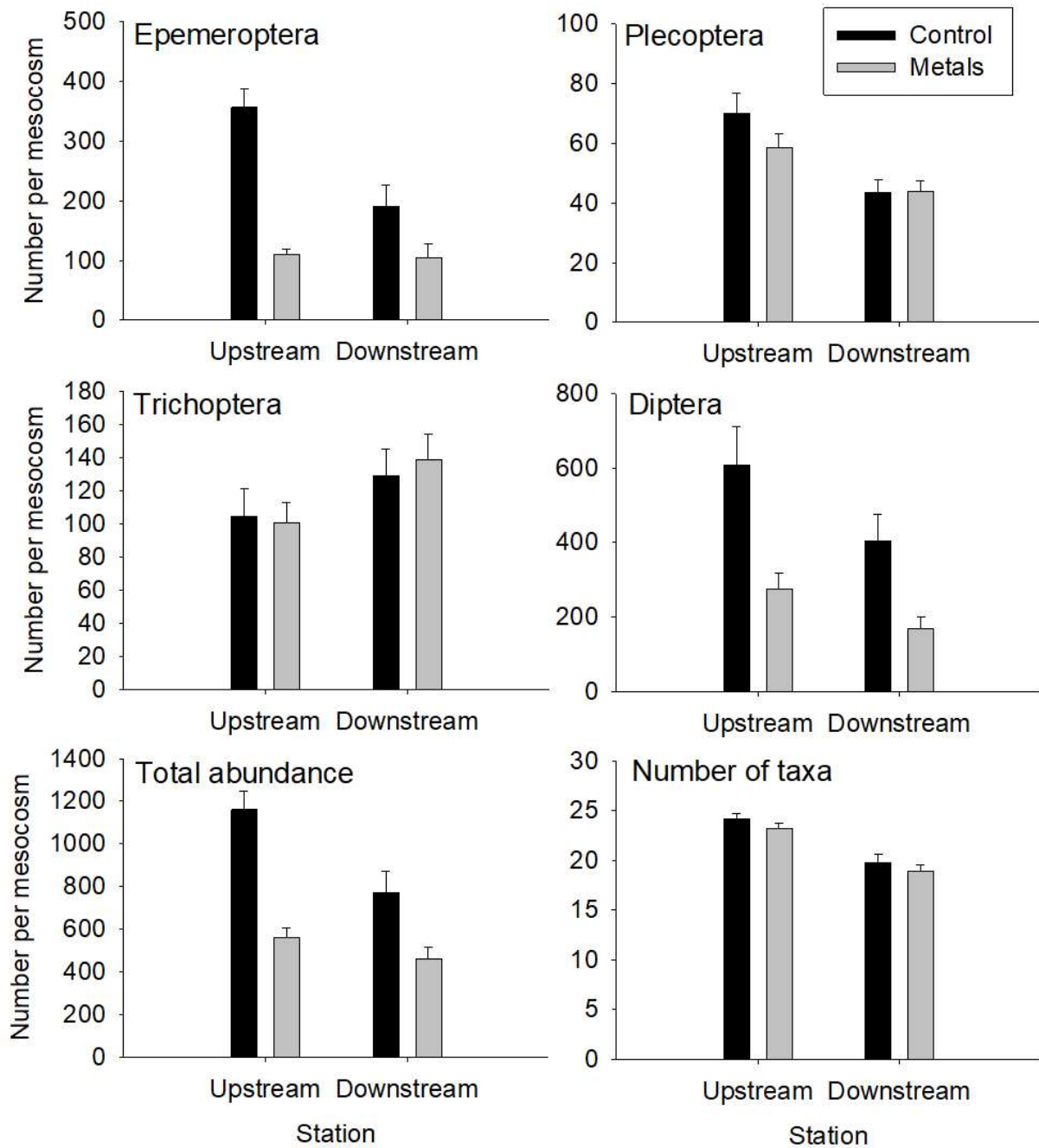


Figure 4. Results of mesocosm experiments showing effects of metals on benthic communities collected upstream and downstream from the former California Gulch Superfund Site. Figure shows mean (+s.e.) abundance of the 4 major macroinvertebrate orders, total abundance and number of taxa in control and metal-treated streams. Results of statistical analyses are shown in Table 1.

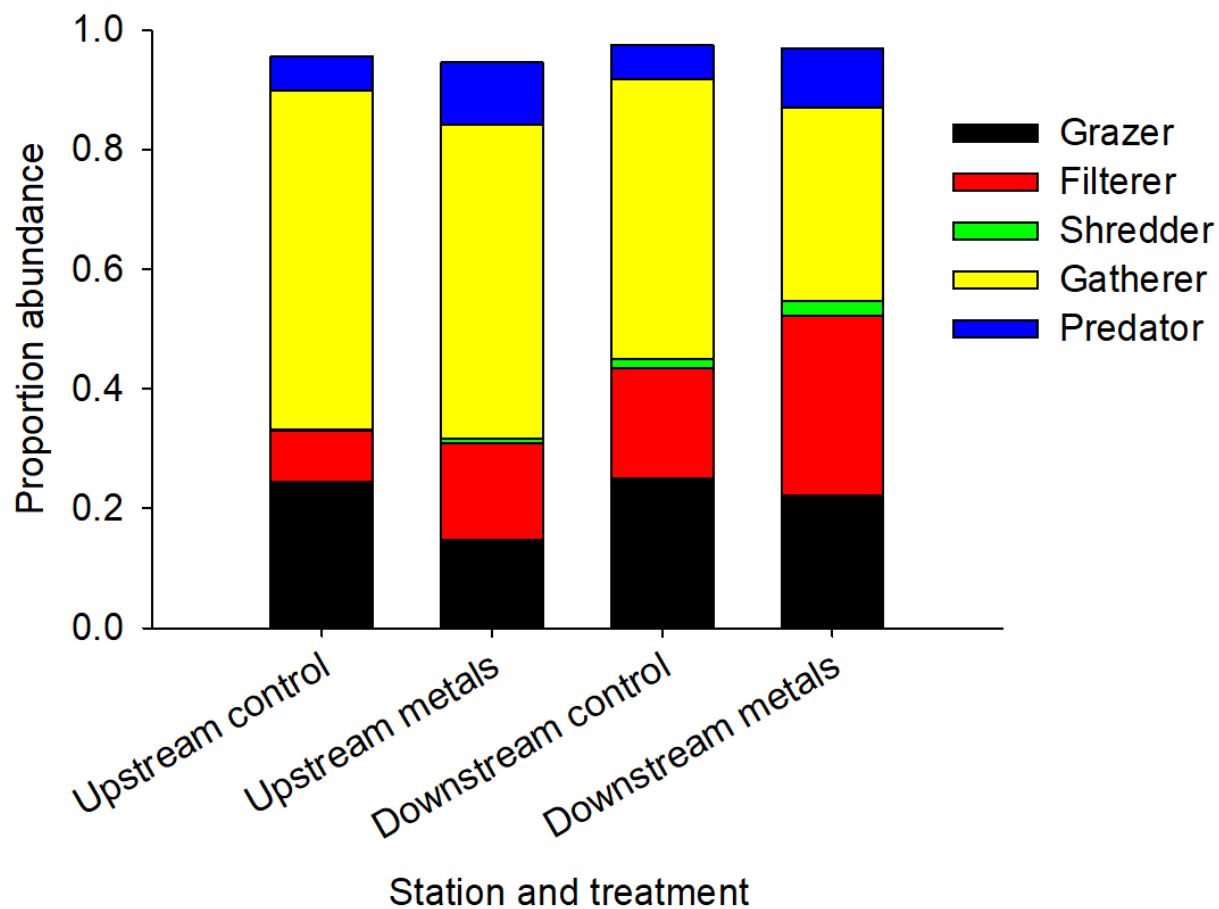


Figure 5. Results of mesocosm experiments showing effects of metals on abundance of the 5 major feeding groups collected upstream and downstream from the former California Gulch Superfund Site. Figure shows mean abundance of the 5 major functional feeding groups in control and metal-treated streams. Results of statistical analyses are shown in Table 1.

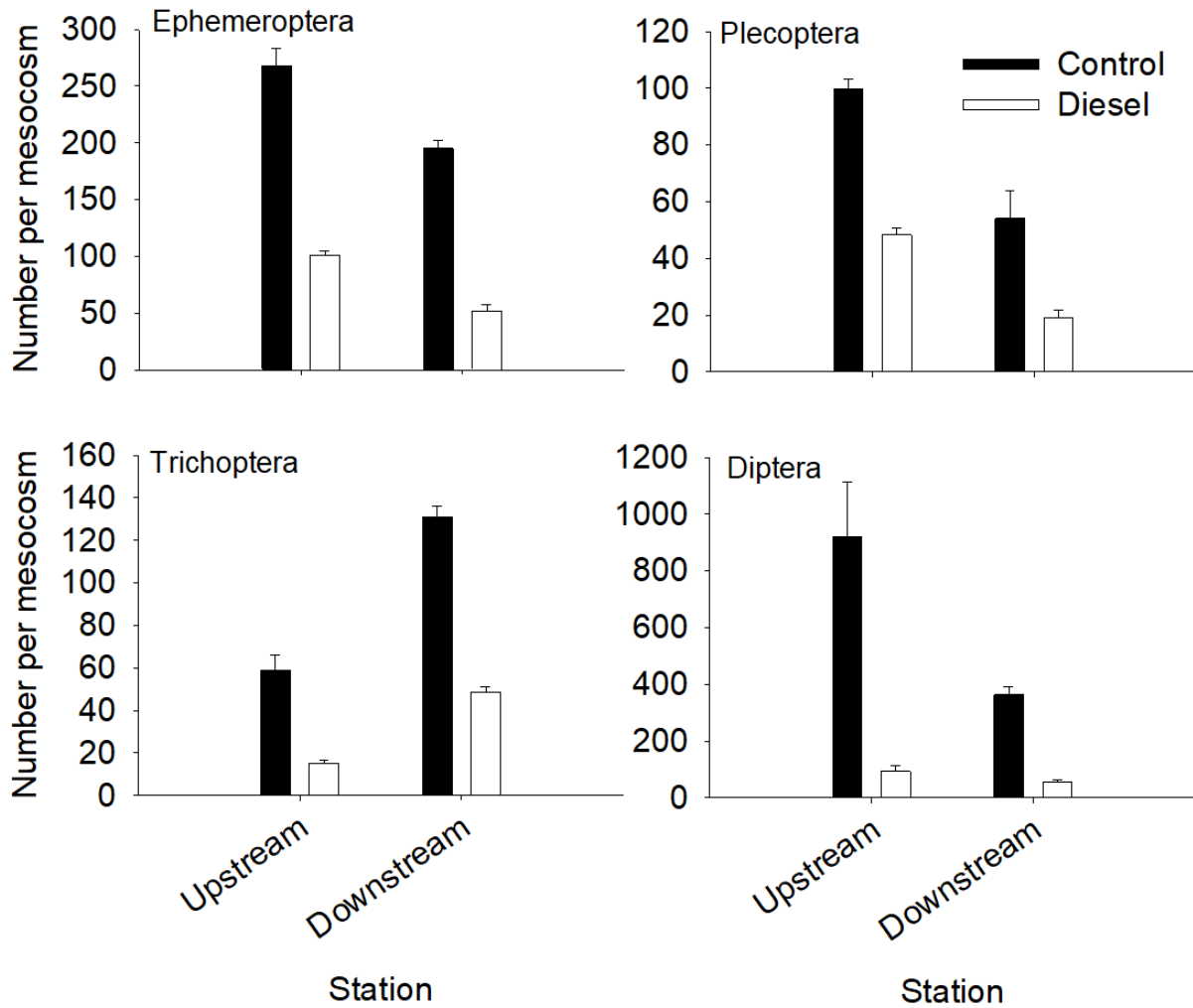


Figure 6. Results of mesocosm experiments showing effects of diesel fuel on abundance of the 4 major macroinvertebrate orders collected upstream and downstream from the former California Gulch Superfund Site. Results of ANOVA testing for the effects of diesel treatment, site and the treatment x site interaction are shown Table S3.

Literature Cited

- Angeler, D. G., and C. R. Allen. (2016) Quantifying resilience. *Journal of Applied Ecology* 53:617–624.
- Beatty B.J., Black W.C., Carlson J.O., Clements W.H., Duteau N., Harrahy E., ... Rayms-Keller A. (1998) Molecular and genetic ecotoxicologic approaches to aquatic environmental bioreporting. *Environmental Health Perspectives*, **106**, 1395–1407.
- Beisner B.E., Haydon D.T. & Cuddington K. (2003) Alternative stable states in ecology. *Frontiers in Ecology and the Environment*, **1**, 376–382.
- Bellwood D.R., Hughes T.P., Folke C. & Nystrom M. (2004) Confronting the coral reef crisis. *Nature*, **429**, 827–833.
- Benayas J.M.R., Newton A.C., Diaz A. & Bullock J.M. (2009) Enhancement of biodiversity and ecosystem services by ecological restoration: a meta-analysis. *Science*, **325**, 1121–1124.
- Bernhardt E.S., Palmer M.A., Allan J.D., Alexander G., Barnas K., Brooks S., ... Sudduth E. (2005) Synthesizing U.S. river restoration efforts. *Science*, **308**, 636–637.
- Bernhardt E.S., Sudduth E.B., Palmer M.A., Allan J.D., Meyer J.L., Alexander G., ... Pagano L. (2007) Restoring rivers one reach at a time: results from a survey of US river restoration practitioners. *Restoration Ecology*, **15**, 482–493.
- Brix K.V., Deforest D.K. & Adams W.J. (2011) The sensitivity of aquatic insects to divalent metals: A comparative analysis of laboratory and field data. *Science of the Total Environment*, **409**, 4187–4197.
- Cadotte M.W., Carscadden K. & Mirotnick N. (2011) Beyond species: functional diversity and the maintenance of ecological processes and services. *Journal of Applied Ecology*, **48**, 1079–1087.

- Cardinale B.J., Nelson K. & Palmer M.A. (2000) Linking species diversity to the functioning of ecosystems: on the importance of environmental context. *Oikos*, **91**, 175–183.
- Cardinale B.J. & Palmer M.A. (2002) Disturbance moderates biodiversity-ecosystem function relationships: Experimental evidence from caddisflies in stream mesocosms. *Ecology*, **83**, 1915–1927.
- Clements, W.H. 2004. Small-scale experiments support causal relationships between metal contamination and macroinvertebrate community responses. *Ecological Applications*, **14**, 954-967.
- Clements, W.H. 1999. Metal tolerance and predator-prey interactions in benthic macroinvertebrate stream communities. *Ecological Applications* **9**:1073–1084.
- Clements, W.H., Brooks, M.L., D.R. Kashian, R.E. Zuellig (2008) Changes in dissolved organic material determine exposure of stream benthic communities to UV-B radiation and heavy metals: implications for climate change. *Global Change Biology* **14**: 2201–2214.
- Clements W.H., Cadmus P. & Brinkman S.F. (2013) Responses of aquatic insects to Cu and Zn in stream microcosms: understanding differences between single species tests and field responses. *Environmental Science & Technology*, **47**, 7506–7513.
- Clements W.H., Carlisle D.M., Lazorchak J.M. & Johnson P.C. (2000) Heavy metals structure benthic communities in Colorado mountain streams. *Ecological Applications*, **10**, 626–638.
- Clements W.H., Hickey C.W. & Kidd K.A. (2012) How do aquatic communities respond to contaminants? It depends on the ecological context. *Environmental Toxicology and Chemistry*, **31**, 1932–1940.

- Clements W.H., Kashian D.R., Kiffney P.M. & Zuellig R.E. (2016) Perspectives on the context-dependency of stream community responses to contaminants. *Freshwater Biology*, **61**, 2162–2170.
- Clements W.H., Vieira N.K.M. & Church S.E. (2010) Quantifying restoration success and recovery in a metal-polluted stream: a 17-year assessment of physicochemical and biological responses. *Journal of Applied Ecology*, **47**, 899–910.
- Courtney, L.A. and W.H. Clements (2000) Sensitivity to acidic pH in benthic invertebrate assemblages with different histories of metal exposure. *Journal of the North American Benthological Society* 19:112-127.
- De Laender F., Verschuren D., Bindler R., Thas O. & Janssen C.R. (2012) Biodiversity of freshwater diatom communities during 1000 years of metal mining, land use, and climate change in Central Sweden. *Environmental Science & Technology*, **46**, 9097–9105.
- Dornelas M., Gotelli N.J., McGill B., Shimadzu H., Moyes F., Sievers C. & Magurran A.E. (2014) Assemblage time series reveal biodiversity change but not systematic loss. *Science*, **344**, 296–299.
- Dudley, T. L., C. M. Dantonio, and S. D. Cooper (1990) Mechanisms and consequences of interspecific competition between 2 stream insects. *Journal of Animal Ecology* **59**,849-866.
- Duggan, S.B., Kotalik, C.J. and Clements, W.H. (2018) Integrating results of field biomonitoring and mesocosm experiments to validate post-spill impacts of petroleum hydrocarbons on stream benthic communities. *Environmental Science & Technology*, 52, 13584–13590.

- Folke C., Carpenter S., Walker B., Scheffer M., Elmqvist T., Gunderson L. & Holling C.S. (2004) Regime shifts, resilience, and biodiversity in ecosystem management. *Annual Review of Ecology Evolution and Systematics*, **35**, 557–581.
- Gunn J., Sarrazin-Delay C., Wesolek B., Stasko A. & Szkokan-Emilson E. (2010) Delayed recovery of benthic macroinvertebrate communities in Junction Creek, Sudbury, Ontario, after the diversion of acid mine drainage. *Human and Ecological Risk Assessment*, **16**, 901–912.
- Herbst, D. B.; Medhurst, R. B.; Black, N. J. (2018) Long-term effects and recovery of streams from acid mine drainage and evaluation of toxic metal threshold ranges for macroinvertebrate community reassembly. *Environmental Toxicology and Chemistry* **37**: 2575-2592.
- Hobbs R.J. (2007) Setting effective and realistic restoration goals: Key directions for research. *Restoration Ecology*, **15**, 354–357.
- Hobbs R.J., Arico S., Aronson J., Baron J.S., Bridgewater P., Cramer V.A., ...Zobel M. (2006) Novel ecosystems: theoretical and management aspects of the new ecological world order. *Global Ecology and Biogeography*, **15**, 1–7.
- Hobbs R.J., Higgs E. & Harris J.A. (2009) Novel ecosystems: implications for conservation and restoration. *Trends in Ecology & Evolution*, **24**, 599–605.
- Holling, C. S. 1973. Resilience and stability of ecological systems. *Annual Review of Ecology and Systematics*, **4**, 1-24.
- Hornberger, M. I.; Luoma, S. N.; Johnson, M. L.; Holyoak, M. (2009) Influence of remediation in a mine-impacted river: metal trends over large spatial and temporal scales. *Ecological Applications*. 19:1522-1535.

- Hudson-Edwards K. (2016) Tackling mine wastes. *Science*, **352**, 288–290.
- Hughes T.P., Bellwood D.R., Folke C., Steneck R.S. & Wilson J. (2005) New paradigms for supporting the resilience of marine ecosystems. *Trends in Ecology and Evolution*, **20**, 380–386.
- Jones A., Rogerson M., Greenway G., Potter H.a.B. & Mayes W.M. (2013) Mine water geochemistry and metal flux in a major historic Pb-Zn-F orefield, the Yorkshire Pennines, UK. *Environmental Science and Pollution Research*, **20**, 7570–7581.
- Kashian, D.R., R.E. Zuellig¹, K.A. Mitchell and W.H. Clements. 2007. The cost of tolerance: sensitivity of stream benthic communities to UV-B and metals. *Ecological Applications* 17: 365-375.
- Kiffney, P. M. 1996. Main and interactive effects of invertebrate density, predation, and metals on a Rocky Mountain stream macroinvertebrate community. *Canadian Journal of Fisheries and Aquatic Sciences* 53:1595–1601.
- Kohler S. L. (1992) Competition and the structure of a benthic stream community. *Ecological Monographs* **62**, 65-188.
- Lake P.S., Bond N. & Reich P. (2007) Linking ecological theory with stream restoration. *Freshwater Biology*, **52**, 597–615.
- Levinton J.S., Suatoni E., Wallace W., Junkins R., Kelaher B. & Allen B.J. (2003) Rapid loss of genetically based resistance to metals after the cleanup of a Superfund Site. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 9889–9891.
- Magurran A.E. (2016) How ecosystems change. *Science*, **351**, 448–449.
- Mebane C.A., Eakins R.J., Fraser B.G., Adams W.J. (2015) Recovery of a mining-damaged stream ecosystem. *Elementa* 3:000042.

- Merritt R.W., Cummins K.W. & Berg M.B. (2008) An introduction to the aquatic insects of North America. p. 1158. Kendall Hunt Publishing, Debuque, Iowa.
- Millennium Ecosystem Assessment, 2005. Ecosystems and human well-being: wetlands and water. World Resources Institute, Washington, DC.
- Mori A.S., Isbell F. & Seidl R. (2018) beta-Diversity, community assembly, and ecosystem functioning. *Trends in Ecology & Evolution*, **33**, 549–564.
- Oliver T.H., Heard M.S., Isaac N.J.B., Roy D.B., Procter D., Eigenbrod F., ... Bullock J.M. (2015) Biodiversity and resilience of ecosystem functions. *Trends in Ecology & Evolution*, **30**, 673–684.
- Paine R.T., Tegner M.J. & Johnson E.A. (1998) Compounded perturbations yield ecological surprises. *Ecosystems*, **1**, 535–545.
- Palmer M.A., Ambrose R.F. & Poff N.L. (1997) Ecological theory and community restoration ecology. *Restoration Ecology*, **5**, 291–300.
- Palmer M.A., Menninger H.L. & Bernhardt E. (2010) River restoration, habitat heterogeneity and biodiversity: a failure of theory or practice? *Freshwater Biology*, **55**, 205–222.
- Poteat M.D. & Buchwalter D.B. (2014) Four reasons why traditional metal toxicity testing with aquatic insects is irrelevant. *Environmental Science & Technology*, **48**, 887–888.
- Rocha, J. C., G. D. Peterson, and R. Biggs. 2015. Regime shifts in the anthropocene: drivers, risks, and resilience. *Plos One* 10:e0134639-e0134639.
- Ruiz-Jaen M.C. & Aide T.M. (2005) Restoration success: How is it being measured? *Restoration Ecology*, **13**, 569–577.
- Scheffer M., Carpenter S., Foley J.A., Folke C. & Walker B. (2001) Catastrophic shifts in ecosystems. *Nature*, **413**, 591–596.

- Scheffer M. & Carpenter S.R. (2003) Catastrophic regime shifts in ecosystems: linking theory to observation. *Trends in Ecology & Evolution*, **18**, 648–656.
- Schmidt T.S., Clements W.H., Mitchell K.A., Church S.E., Wanty R.B., Fey D.L., Verplanck P.L. & San Juan C.A. (2010) Development of a new toxic-unit model for the bioassessment of metals in streams. *Environmental Toxicology and Chemistry* **29**, 2432–2442.
- Schroder A., Persson L. & De Roos A.M. (2005) Direct experimental evidence for alternative stable states: a review. *Oikos*, **110**, 3–19.
- Simon T.P. (2003) Biological response signatures. Indicator patterns using aquatic communities. CRC Press, Boca Raton, FL.
- U.S. EPA (2002) National recommended water quality criteria, U.S. Environmental Protection Agency, Office of Water. EPA-822-R-02-047, Washington, D.C.
- Vigneron A., Geffard O., Coquery M., Francois A., Queau H. & Chaumot A. (2015) Evolution of cadmium tolerance and associated costs in a *Gammarus fossarum* population inhabiting a low-level contaminated stream. *Ecotoxicology*, **24**, 1239–1249.
- Wallace J.B., Lughart G.J., Cuffney T.F. & Schurr G.A. (1989) The impact of repeated insecticidal treatments on drift and benthos of a headwater stream. *Hydrobiologia*, **179**, 135–147.
- Ward J.V., Kondratieff B.C. & Zuellig R.E. (2002) *An illustrated guide to the mountain streams insects of Colorado*, University Press of Colorado, Boulder, CO.
- Wilson J.B. (1988) The cost of heavy-metal tolerance: an example. *Evolution*, **42**, 408–413.
- Young T.P., Petersen D.A. & Clary J.J. (2005) The ecology of restoration: historical links, emerging issues and unexplored realms. *Ecology Letters*, **8**, 662–673.

Zuellig R.E., Kashian D.R., Brooks M.L., Kiffney P.M. & Clements W.H. (2007) Metal exposure history and the influence of ultraviolet-B radiation on benthic communities in Colorado Rocky Mountain streams. *J. North Am. Benthol. Soc.*, **27**, 120–134.

CHAPTER 2

METALS ALTER MEMBERSHIP BUT NOT DIVERSITY OF A HEADWATER STREAM MICROBIOME

Introduction

Streams in the western United States are frequently impaired from elevated metal concentrations due to a combination of historical mining activities and to a lesser extent natural weathering processes. In Colorado, there are approximately 23,000 abandoned mines (Graves 2017) resulting in approximately 23% of Colorado streams qualifying as impaired (Clements et al. 2000). One metric routinely used to evaluate stream water quality and ecosystem health is stream macroinvertebrate community composition. Various protocols that use macroinvertebrates continue to be standard practice for stream biomonitoring (e.g., De Pauw and Vanhooren 1983; Barbour et al. 1999; Carlisle et al. 2008). Macroinvertebrates have a series of characteristics that have proven useful for stream bioassessments including ubiquity, high diversity, restricted ranges, short generation times, small size, and are important food sources for aquatic and terrestrial consumers alike (e.g., Baxter et al. 2005). However recent studies have found that macroinvertebrate membership may be a better indicator than diversity for how contaminants affect stream ecosystems (Wolff et al. 2019). Because microorganisms have many of these same characteristics and because analyses of microbiome characteristics have become more routine, we investigated whether microbiomes had dynamic responses to metal exposure in metal-contaminated ecosystems that would make them an appropriate indicator of stream ecosystem health.

We hypothesized that microbiomes may potentially be better indicators of water quality than macroinvertebrates because they are even more ubiquitous and dynamic and thus may report even subtler differences in water quality. For instance, typical bacterial generation times (i.e., doubling time) occur over hours or days (Gibson et al. 2018) compared to weeks to months for macroinvertebrates (Merritt and Cummins 2019). The spatial scale at which microbiomes operate is also much smaller than for stream macroinvertebrates, creating potential to identify small pockets of contamination in heterogeneous stream ecosystems. We now know that microbial biofilms are formed by complex, non-random assemblages of algae, bacteria, and fungi (Flemming et al. 2016) and that these diverse microbiomes can be shaped by physical properties like stream velocity (Battin et al. 2003) and substrate type (reviewed in Zeglin 2015) as well as chemical properties such as pH (Fierer et al. 2007, Lear et al. 2009). It is also clear that metals affect the function of microbiomes, including evidence for metals decreasing stream nitrification (Niyogi et al. 2003), and reducing rates of microbial respiration (Carlisle and Clements 2005). Metals also affect microbiome membership, including evidence where specific sub-phyla increased (γ -proteobacteria) or decreased (β -proteobacteria) with metal exposure (Feris et al. 2003).

To test the potential for microbiomes to act as indicators of metal contamination we evaluated the stream microbiome of the upper Arkansas River near Leadville, Colorado, USA. The upper Arkansas River has been impaired by metal pollution due to historical mining since the mid-1800s (Clements et al. 2010). By the late 1990s, implementation of water treatment facilities and removal of floodplain mine tailings resulted in significant improvements in water quality including decreased dissolved metals – principally cadmium (Cd), copper (Cu), and zinc (Zn) – downstream from where California Gulch enters the mainstem of the upper Arkansas

River (Clements et al. 2010). Despite improved water quality, macroinvertebrate community membership has remained different between upstream reference sites and sites downstream of California Gulch (Clements et al. 2010). Although species richness has remained similar between upstream and downstream locations, community membership has continued to differ among sites (Wolff et al. 2019).

To assess how microbiomes were affected by metal exposure in the upper Arkansas River we chose to focus on the bacterial component of the stream microbiome because: A) sediment biofilms are primarily composed of bacterial biomass (from 90-99%) (Nealson 1997; Bar-On et al. 2018), B) many stream macroinvertebrates spend significant portion of their lifecycle grazing on biofilm in sediments (Boeker et al. 2016) so changes in microbiome may have effects on higher trophic levels, and C) 16S rRNA gene sequences have better developed sequence libraries compared to analogous phylogenetic markers for other groups, such as the 18S rRNA gene for eukaryotic microbes (Hugerth et al. 2014). We collected samples at locations upstream and downstream of California Gulch during both Spring and Fall seasons. From each sample, we used 16S rRNA amplicon sequencing of the sediment biofilms on the Arkansas River to determine how metals influenced microbiome diversity and membership in the upper Arkansas River. We complemented our field observations with experiments that exposed microbial communities sampled from both upstream and downstream of California Gulch to elevated metal concentrations. The purpose of this study was to examine: (1) how microbiome diversity and membership differed in an ecosystem that has elevated metal levels, (2) if these differences could be attributed to exposure to metals and (3) if certain microbiome genera were consistently enriched or depleted in response to metal exposure and therefore may be candidates for indicators of stream water quality. The last goal is an important step toward identifying

mechanistic responses of individual bacteria and aid in the development in using certain groups as sensitive “indicators” of metals stress.

Methods

Study Site

We conducted our observational study on the upper Arkansas River, located near the town of Leadville, approximately 100 km west of Denver, Colorado. This area of the upper Arkansas has been monitored since 1989 and the site conditions are well characterized in previous studies (e.g., Clements et al. 2010). Briefly, this area is approximately 2,820 meters above sea level, and typically receives ~30 cm of precipitation annually. The Arkansas River has a snowmelt driven hydrograph, with peak discharges in May or June, normally reaching base flow conditions by September. Variable run-off alters streamflow and contributions of solutes (including metals) from the watershed, resulting in higher metal concentrations recorded during Spring (i.e., during snowmelt runoff) compared to the Fall (i.e., at baseflow, Clements et al. 2010). In the reach of the Upper Arkansas we evaluated the stream substrate was primarily composed of medium to large cobble in a matrix of gravel and sand. Most riparian vegetation was composed of sagebrush (*Artemisia* spp.), grasses, and willow (*Salix* spp.) trees.

Observational Study We sampled sediment bacteria communities in the main stem of the Arkansas River at three locations: 2 sites upstream (AR1 and AR2), 4 sites downstream (AR3, AR4, AR4G, and AR5), and 1 site within the principal metal contributing tributary, California Gulch (Figure 1). At each site we collected samples to be analyzed for metal concentration and 16S amplicon sequencing in Spring (first week of May) and Fall (first week of October) of 2017. For each sampling event we collected four independent sediment samples in riffle habitat (with a

water column depth of ca. 0.25 m) at each of the 7 sites. For each sample, we removed a large cobble (~ 0.3 m diameter) and scooped underlying sediments into separate 50 ml Falcon™ tubes.

Experimental Mesocosm

To more explicitly test the effects of metals on stream microbiomes, we designed a common garden experiment using samples from an upstream and downstream location. Specifically, we tested if experimentally manipulated metal concentrations would result in similar effects on the microbiome as seen from the metal gradient in the upper Arkansas River. The observational study was conducted in Spring and Fall, however the mesocosm experiments were conducted only in the Fall. The design and parameters of the mesocosm experiments have been described elsewhere (e.g., Clements et al. 2019). Briefly, biofilms for the experiments were collected by placing plastic trays containing clean (scrubbed and air-dried) cobble in the river for 31 days (09/05/2017 – 10/06/2017) allowing microbial biofilms to colonize the cobble in each tray. Trays were deployed at one reference site upstream of California Gulch (AR1; hereafter “Upstream”) and one site downstream of California Gulch (AR5, hereafter “Downstream”). Upon retrieval, 4 colonized trays were collected from each site and placed into individual coolers filled with ambient stream water then immediately transported to CSU’s Stream Research Laboratory (~ 3 hours from the sampling site). The 4 trays from each cooler were then placed into an individual experimental “racetrack” stream which after an equilibration period (~ 24 hours) were randomly assigned to 2 treatments: metals or no metals (control).

Each artificial stream received source water from the hypolimnion of a mesotrophic reservoir (Horsetooth Reservoir) that was delivered at a rate of 1.0 L min⁻¹, resulting in a residence time of approximately 20 minutes for each mesocosm. Characteristics of the source

water (e.g., pH, conductivity, temperature, dissolved oxygen) were typical of non-polluted mountain streams in Colorado (Cadmus et al. 2016). We implemented a 2 X 2 (location X treatment) factorial experimental design: 1) control-upstream, 2) control-downstream, 3) metals-upstream, and 4) metals-downstream. Each control and treatment were replicated four times for a total of 16 experimental streams. We started metal additions after a 24 hr acclimation period. Peristaltic pumps delivered stock solutions of a metal mixture from a 20 L concentrated carboy at a rate of 10 ml min⁻¹ to obtain a targeted concentration of 25 µg L⁻¹ Cu and 650 µg L⁻¹ Zn for each treatment. Paddlewheels provided a constant flow of 0.35 m s⁻¹ to each mesocosm. Metals concentrated in the carboys were refreshed daily during the 10 day experiment. We checked water and peristaltic pump flows twice daily to ensure consistent delivery of metal solutions among treatments. We measured ambient metals concentrations from each mesocosm by filtering (0.45 µm) 15 mL water samples on Day 2, Day 4, and Day 10 of the experiment. On Day 10, all trays from each stream were collected, sieved (350 µm) into a clean, plastic bucket. Buckets were then decanted and the remaining material (e.g., sediments and periphyton floc) were transferred into 50 ml Falcon tubes and frozen at -80 °C until DNA extraction and metals analysis.

DNA preparation and 16S rRNA amplicon sequencing

We extracted DNA from each sample with a MoBio Power Soil DNA extraction kit and amplified and sequenced the V4 region of the 16S rRNA gene on an Illumina MiSeq sequencing platform to identify microbiome membership from each sample. Sequence reads were analyzed using MOTHUR (version 1.40.2; Schloss et al. 2009) and OTUs counts defined at a 97% similarity of the sequence using the OptiClust algorithm. Generated OTUs were then aligned to a SILVA reference file (version 132; Quast et al. 2013). After sequences were processed through

the Mothur pipeline, we then imported the data in R studio (R Core Team 2017) for statistical analyses and visualization. Within R, subsequent analyses were performed utilizing the package Phyloseq. Sequences were pre-processed to remove Operational Taxonomic Units (OTUs) that were not counted at least 3 times in 20% of the samples.

For most analyses, raw OTU counts were transformed to relative abundances within each sample to reduce issues that can arise from count data from samples with varying library size. However, for DESeq2 analyses, we did not normalize count data to relative abundances because DESeq2 algorithms require raw sequence count data inputs. We also aggregated all OTUs that shared the same genera before performing DESeq2 analysis. We visualized DESeq2 results with Log2-fold change plots analyses (SRA Accession # [in progress]).

Metals preparation

We measured metal concentrations using material remaining from sediment samples that had previously been sub-sampled for DNA preparation. We dried sediments in a drying oven at 60 °C for at least 24 hours with periodic weighing of each sample until no more mass was lost and the sample remained at a constant weight. A small amount of sediment (0.14 – 0.25 g) was then weighed and transferred into 15 mL Falcon® tubes. Next, 1 mL trace-element grade nitric acid (HNO₃) was added to each sample. Samples were vortexed and then placed in a hot water bath at 90 °C for 4 – 6 hours. Samples were then cooled outside of the hot water bath and ca. 0.2 mL hydrogen peroxide (H₂O₂) was then added. Samples were vortexed again and then placed back in the hot water bath for an additional 4 – 6 hours. After this period, samples were cooled, and 8.8 mL of Milli-Q water was added to ensure that all samples were diluted to a total of 10 mL. Samples were vortexed for a final time, centrifuged at 2500 rpms for 5 minutes, and the supernatant was extracted into clean 15 mL falcon tubes for quantification of metal

concentration. We used dry weight and dilution volume (10 mL) to calculate the concentration of metals in each sample ($\mu\text{g g}^{-1}$). Metal concentrations were quantified using a flame Atomic Absorption Spectrometer at the Colorado State University's Aquatic Ecotoxicology Laboratory. From each sample we measured copper (Cu), cadmium (Cd; observational samples only), and zinc (Zn). These metals have been determined from previous studies to be the principal metals contaminating the upper Arkansas River (Clements et al. 2010).

Biomonitoring statistics

We estimated microbiome diversity using number of observed OTUs (richness) and the Shannon Index (Shannon and Weaver 1949), where the higher the number indicates greater alpha diversity for an individual site for both metrics. Whereas observed OTUs is an index of alpha diversity that captures the number of unique OTUs present in each sample, other indices (e.g., the Shannon Index employed here) of alpha diversity also account for the distribution of those OTUs (evenness). To assess the effect of metals on membership we grouped upstream sites (AR1 and AR2) together (subsequently referred to as "Upstream"), and we grouped the four downstream sites (AR3, AR4, AR4G, and AR5) together (subsequently referred to as "Downstream"). To test for differences in microbiome membership between upstream and downstream communities we used a Bray-Curtis Similarity Index and visualized the similarities in community membership using Principle Coordinates Analysis (PCoA plot). To determine if clusters from each location/season were statistically different from each other we used a PERMANOVA model. For the *in-situ* Arkansas River bacterial communities, we tested the effect of location (upstream vs. downstream), season (Spring vs. Fall), and their interaction. PERMANOVA tests were also used for pairwise comparisons (e.g., Upstream-Spring vs. Downstream-Spring; Upstream-Fall vs. Downstream-Fall, etc.).

After evaluating whole community membership differences, we used log₂-plots to visualize what, if any, microbiome genera were significantly enriched or depleted in the upstream or downstream locations. We used a very low alpha value ($\alpha \leq 0.0005$) to protect against type-I error when identifying genera that are differentially enriched between upstream and downstream locations. If certain genera were more enriched at downstream locations (positive log₂-fold value) we considered them as potential indicators of metal-tolerant bacteria. Similarly, genera more enriched at upstream sites (negative log₂-fold value) we considered to be indicators of metal-sensitive bacteria. To aid in comparison with mesocosm results, we only included samples sourced directly from AR1 (upstream) and AR5 (downstream) locations for Log₂-fold plots.

Mesocosm Statistics

For the mesocosm experiment samples, we used a 2x2 factorial design that tested the effect of metals treatment (control vs. treatment), and location (upstream vs. downstream). We used many of the same statistical and visual analyses as described previously for the observational study. One notable difference is that for mesocosm log₂-fold plots, we compared metal-treated groups (positive log₂-fold values) to control groups (negative log₂-fold values) separately by each location.

Results

Observational Study

To assess the effect of metals on the upper Arkansas River microbiome during the Spring and Fall of 2017 we analyzed *in situ* pH, sediment metal concentration, and 16S rRNA amplicons from California Gulch and sites upstream and downstream of California Gulch in the main stem of the upper Arkansas River. For all sampling locations during both seasons

California Gulch had the highest sediment metal concentrations for all three metals, followed by downstream sites, with upstream sites consistently having the lowest sediment metal levels (Table 1). At downstream sites there was a statistically significant increase of Cu and Zn in the Spring compared to Fall, but not for Cd (Table 1). However, we did not observe statistically different metal concentrations between Fall and Spring for any of the metals at the upstream sites (consistently low) or in California Gulch (consistently high, Table 1).

We also examined surface water pH at the time of sediment sample collection since the addition of metals can lower pH of the receiving waters, and more importantly, lower pH can make metals more bioavailable to aquatic organisms. While differences in pH among locations were minimal (they ranged between 7.12 – 7.75), in general, samples with higher metal concentrations had lower pH (Table 1). There was also a statistically significant season by location interaction for pH ($p=0.002$). Spring pH values were different among all locations with upstream sites having the highest pH values followed by downstream sites, and then California Gulch. In the Fall, pH at upstream sites was significantly higher than California Gulch and all downstream sites, but there was no statistical difference between California Gulch and downstream locations (Table 1). For all locations, pH was lower in the Spring than in the Fall.

At each site we assessed the 16S rRNA amplicons from all sediment samples. The number of observed OTUs were significantly different among locations and between seasons (Table 1). California Gulch had the lowest observed OTUs in both Spring and Fall (Table 1), however observed OTUs were statistically indistinguishable between upstream and downstream locations in either season (Table 1). At downstream locations, observed OTUs were significantly lower ($p=0.0365$) in the Spring than the Fall, but we observed no significant differences in microbiome observed OTUs between seasons at the upstream sites or in California Gulch (Table

1). Like observed OTUs, Shannon Index values were lowest at California Gulch in both Spring and Fall (Table 1), and there was no statistical difference in alpha diversity between upstream and downstream locations (Table 1). However, unlike the raw estimates of observed OTUs, we did not see any difference in Shannon Index values between seasons, or any significant season by location interactions (Table 1).

We also evaluated microbiome membership to see if metals altered the composition of the microbiome even though indices of alpha diversity may not have differed. We found that microbiome membership was significantly different among locations ($p=0.001$) and between seasons ($p=0.015$). The membership of the California Gulch microbiome was different than the membership of the microbiome at upstream and downstream locations for all sampling dates (Figure 2). Because the difference in membership between California Gulch and either location in the mainstem of the Arkansas River was so pronounced we also performed a PERMANOVA that excluded California Gulch from the dataset to focus on differences between upstream and downstream sites. We found a significant difference in membership ($p=0.003$) between locations and a significant within site difference between seasons ($p=0.016$). However, the season by location (i.e., upstream vs. downstream) effect was not significant ($p\text{-value} = 0.670$), suggesting that microbiome membership differed between seasons, but changes occurred at both locations. A constrained analysis of principles coordinates (CAP) illustrated that differences in microbiome membership between locations were primarily driven by higher Cu and Zn sediment concentrations in downstream sites, which also likely reduced pH at the downstream sites (Figure 3).

Mesocosm Study

We did not find significant effects of location, metals, or an interaction between location and metals, on the number of observed OTUs in any of the mesocosms after the 10-day exposure period (Table 2). However, alpha diversity (as assessed by the Shannon Index) was consistently higher for the microbiomes sourced from upstream sites compared to the microbiomes sourced from downstream sites for both the control and metal enriched treatments. When we assessed the effect of the metal treatment on alpha diversity within a location, we found that alpha diversity was significantly lower in the metal-treated samples for the downstream location, but we did not see a similar change in alpha diversity in response to metals for the upstream location (Table 2).

Interestingly, in the metal-treated samples more Cu and Zn were retained in the downstream sediments compared to the upstream sediments over the course of the experiment (Table 2). Although, downstream sediments likely started with greater metal concentrations (inferred from the observational study), we did not see a similar difference in concentrations in the control samples between upstream and downstream sediments suggesting that the metal treated sediments retained metals during the course of the experiment. This also mirrored the response of the stream microbiome, where the effect of metals depended on the location from which sediments were sourced. Over the course of the incubation, microbiomes from the downstream site showed a more pronounced change in membership to metal exposure than the upstream microbiome (Figure 4). This result was supported by a PERMANOVA that identified significant differences in membership between location ($p=0.001$), treatment ($p=0.001$), and the location by treatment interaction term ($p=0.007$).

Genera-level responses between observational and mesocosm studies

In order to assess if microbiome membership was altered by the presence of metals similarly in our observational and experimental studies, we used Log₂-fold plots to evaluate changes in genera (all OTUs identified to a common genus were aggregated) among the two components of the study. From the observational study we focused on the upstream (AR1) and downstream (AR5) sites that were used to seed the mesocosm experiments. The downstream site was significant ($p \leq 0.0005$) enriched in Cyanobacteria and Verrucomicrobia relative to upstream site (Figure 6). The upstream site was enriched in Latescibacteria, Acidobacteria, Proteobacteria, and Rokubacteria relative to the downstream site (Figure 6).

The microbiomes of the mesocosms sourced with upstream sediment (Figure 5) had a more pronounced shift in membership in response to the addition of metals compared to the mesocosms sourced with downstream sediment (Figure 5). However, microbiomes from each location did have genera that were enriched in the metal treated mesocosms compared to the control treatments. Both locations were enriched in Bacteroidetes in the metal treatment compared to control mesocosms (Figure 6). Genera from the Patescibacteria, Planctomycetes, Acidobacteria, Armatimonadetes, and Cyanobacteria were significantly enriched in control mesocosms (Figure 6). Some genera from the Proteobacteria were significantly enriched in both metal-treated while other genera from the same phylum were significantly enriched in control mesocosms.

When we compared our observational and mesocosms studies we did not find representatives from any genera that were consistently enriched at the downstream site in the Upper Arkansas River and in the metal-treated mesocosms. We also did not find OTUs that were enriched from the upstream site in the Upper Arkansas River and in the control mesocosms.

However, there was a representative of an unclassified genus from the Family *Verrucomicrobiales* (Phylum *Verrucomicrobia*; Class *Verrucomicrobiae*) that was enriched in downstream sites and metal-treated mesocosms.

Discussion

In the observational study of the upper Arkansas River, observed OTUs and alpha diversity microbiome were not different between sites upstream and downstream of California Gulch despite consistently higher sediment metal concentrations of Cd, Cu, and Zn at downstream sites. However, the California Gulch microbiome did have significantly lower observed OTUs and Shannon Index values than either the upstream or downstream locations. These results suggest that metal concentrations may need to exceed a certain threshold before having significant effects on microbiome diversity. Previous research has also shown that mining activities or metal impaired streams have very little impact on microbiome diversity, particularly when the metal contamination does not have a pronounced impact on pH (e.g., Reis et al. 2013; Bier et al. 2015), as was the case in this study. Similar effects of metal exposure on macroinvertebrate diversity have been reported from these same locations in the upper Arkansas River. Specifically, overall species richness for macroinvertebrates upstream and downstream of California Gulch were found to be similar, both within and between seasons (Clements et al. 2010). However, membership for the macroinvertebrate community was different between upstream and downstream locations (Wolff et al. 2019) with the metals disproportionately affecting some taxa more than others.

When we examined membership of the river microbiome at each location, we found significant differences in membership between upstream and downstream locations (Figure 3). Because metal delivery to the stream is most pronounced during Spring run-off, we hypothesized

that differences in microbiome membership between locations would be greatest in the springtime, and downstream membership would change the most between seasons. However, site-specific microbiome membership did not change between seasons. Between seasons membership of microbiomes within upstream and downstream sites were from each other (Figure 3). When we examined microbiome membership among sites the majority (39.6%) of the variance in membership explained by the first two principal coordinates was driven by the differences between the California Gulch microbiome and the two main stem locations. Although subtle, membership of the downstream microbiome was more similar to the California Gulch microbiome and this similarity was more pronounced in Spring than Fall, consistent with the idea of metals exerting an influence on the membership of downstream microbiomes. The differences in microbiome membership between upstream and downstream sites (Figure 1) are notable because concentration of metals in the surface water downstream of California Gulch are typically below EPA chronic aquatic life criteria (US EPA 2002).

Previous investigations have shown that experimental metal exposure resulted in much greater effect on composition from macroinvertebrate communities sourced from sites upstream of California Gulch (Wolff et al. 2019). In this mesocosm experiment we expected to see greater changes in microbiome membership of upstream microbiomes in response to the metal treatments, since this site has historically had lower metal exposure and we anticipated the microbes would be more sensitive to metal stress. However, we observed a greater change in downstream microbiome membership in response to metal exposure, in contrast to our expectations. One potential explanation for the differences in treatment effect between locations is that the samples from the downstream location had metal-resistant bacteria present within the microbiome whereas the upstream samples did not or had fewer. Thus, the downstream

microbiome had a more rapid response to metal exposure than the upstream community after the 10-day treatment. This mechanism is supported by the lower evenness in samples sourced from the downstream location following metal exposure compared to upstream microbiomes, suggesting an enrichment of metal tolerant organisms altered the rank abundance of the downstream microbiome. A recent study examining the effects of the antimicrobial drug Ciprofloxacin also reported more pronounced differences in microbiome membership from experimental exposure to Ciprofloxacin along a gradient of urbanized streams in New York (Rosi et al. 2017). The greatest difference in microbiome membership were observed in stream reaches with the highest ambient concentrations of Ciprofloxacin (Rosi et al. 2017). We posit that the discrepancy between community responses in macroinvertebrates versus microbiomes in response to metal exposure was due to the relative timescale of our study. For instance, over a 10-day period of metal exposure, observed differences of macroinvertebrate membership are by driven by mortality of the original community, whereas, microbiomes may experience multiple generations during that same time period. Thus, microbiome membership was likely not only altered by mortality but also by enrichment of metal tolerant taxa, which was more pronounced at the downstream compared to the upstream sites.

We also observed differences in the amount of metals retained in microbial biomass between microbiomes sourced from different locations. In the metal-treated samples, the downstream microbiomes had approximately 5-8x greater Cu and Zn compared to the mesocosms with sediments sourced from the upstream site (Table 2). In contrast, metal concentrations in microbial biomass were very similar between locations in the control treatments. One potential mechanism for this may be increased tolerance of downstream microbiomes through greater production of Extracellular Polymeric Substance (EPS). EPS from

stream biofilms have been shown to that retain metals in proteins, polysaccharides, and humic components (Flemming and Wingender 2000). Therefore, microbiome metal-tolerance may further exacerbate metal exposure to higher trophic levels by retaining metals in their biomass – either internally (e.g., metals compartmentalized within cell vacuole) or externally (e.g., metals bound to EPS). Recent studies have shown that that macroinvertebrates derive much of their metals from diet and not just from aqueous exposure (Croteau and Luoma 2008; Xie et al. 2009). Additionally, metal-resistant populations of oligochaetes in Foundry Cove, New York increased metal exposure to higher trophic levels by production of metal-binding proteins in their tissues (Levinton et al. 2003). If the downstream microbiome did produce more EPS in response to metal exposure (thus retaining more metals) this would suggest a mechanism for how the microbial response to environmental stress (increased EPS production) altered the diets of the next trophic level. Dietary exposure to metals, or decrease in resource quality, provides a mechanism that could explain the differences in macroinvertebrate membership between upstream and downstream locations previously reported for this same location (Wolff et al. 2019).

Similarity in response between the observational and mesocosm studies

When we compared the membership of microbiomes in the field to those incubated in the experimental streams for ~10 days we did not observe strong association between samples from observational and mesocosm studies at the genera-level. Whereas the mesocosm experiment was designed to isolate the effect of metals on the stream microbiome our experimental design likely introduced other confounding factors. For example, comparison between the field observations and the mesocosm experiments were complicated by differences in source water chemistry.

Mesocosms received natural water inoculum from the hypolimnion of a large reservoir (Horsetooth Reservoir) and not upper Arkansas River water. Therefore, it is possible that differences in membership between the mesocosm and the field samples may be due in part to differences in source water and the microorganisms that were associated with the water from each ecosystem. In addition, the sites in the upper Arkansas River are open canopy and the downstream site is located downstream from the town of Leadville, CO (pop. ~3,000), whereas the reservoir water is comparatively lower in nutrients and sourced from the aphotic hypolimnion. Thus, differences in light environment and water chemistry between the field and the laboratory may also contribute to enrichment of certain genera due to other factors that were not influenced in the same way in the mesocosm studies. Additionally, California Gulch is likely enriched in ammonia and other nutrients from the wastewater treatment process that may be responsible for some observed differences. For instance, we found that *Cyanobacteria* were more enriched in the downstream site (i.e., possibly indicating metal-tolerance), but in the mesocosms *Cyanobacteria* were only significantly enriched in controls. Because we did not see similar enrichment of members of the Cyanobacteria in response to metal exposure in the mesocosm experiments, this would also suggest that differences in Cyanobacteria between locations in the field were not entirely driven by metals but perhaps were due to some of the confounding factors that were present in the observational study.

We observed consistent enrichment of members of a single phylum between both observational and experimental studies and enrichment of members of some genera between both as well. For example, Proteobacteria were prevalent in both upstream and downstream observational samples. This is not surprising given the extremely high diversity within the proteobacteria, but also further highlights the idea that comparisons among different phyla are

likely too broad for applied microbial ecology research, as has been previously suggested (Newton et al. 2011). It is unlikely that metal sensitivity or resistance to metals is a trait that is conserved at the level of the phyla. However, we did find that members of Bacteroidetes were consistently enriched in downstream and metal-treated mesocosms compared to other phyla. This is consistent with previous research that has shown Bacteroidetes to be significantly enriched at other metal polluted sites (e.g., Jordaan et al. 2019). Three other genera that were significantly enriched in metal-treated mesocosms from both the upstream and downstream locations: *oc32* (Phylum Proteobacteria; Class Gammaproteobacteria; Order Burkholderiales; Family Nitrosomonadaceae), *Parasediminibacterium* (Phylum Bacteroidetes; Class Shingobacteriia; Order Sphingobacteriales; Family Chitinophagaceae) and an OTU from the unclassified genus in the Family Microscillaceae (Phylum Bacteroidetes; Class Cytophagia; Order Cytophagales). An OTU from an unclassified genus in the Family Flavobacteriales was found to be enriched in control mesocosms from both upstream and downstream locations. These four genera (e.g., 3 metal-tolerant and 1 metal-sensitive) represent the best candidates from our study to assess the impact of metals on stream ecosystems. Using only the observational study we may have assumed that other genera were also good candidates of metal contamination, however the comparison with the mesocosm experiment excluded these candidates. Since many of the metal-treated genera that were enriched came from the Phylum Bacteroidetes this group may be a logical starting point for more directed research into using microbiome membership as an indicator of metal contamination.

The effects of low metal exposure to stream ecosystems

The effect of metals on microbiomes upstream and downstream of the contaminant site had similar diversity but significant differences in membership that appeared to be caused by

exposure to heavy metals. Other studies have documented changes in microbiome composition associated with a range of heavy metal exposure in the field from diverse geologic sources such as mountaintop mining (Bier et al. 2015), acid mine drainage (Feris et al. 2003), and streams influenced by urban runoff (Ancion et al. 2013). In our study, the use of complimentary field observations and experimental mesocosms allowed us to assess which constituents of natural microbiomes were likely to be consistently affected by metal exposure.

We conclude that the microbiomes in the Upper Arkansas River downstream of California Gulch are still responding to metals stress. Even though diversity metrics suggest similarities between upstream and downstream microbiomes, differences in membership indicate that metals are impacting the upper Arkansas River even after extensive restoration efforts have lowered surface water metals below current US EPA criteria levels (US EPA 2002). Interestingly, typical metal concentrations downstream of California Gulch are not directly toxic to macroinvertebrate communities (Clements 2004). However, the patterns we observed between upstream and downstream microbiomes (i.e., comparable alpha diversity but distinct membership) were similar to patterns found for the macroinvertebrate communities at the same site (Wolff et al. 2019). We also note that in the mesocosm study the microbiomes sourced from the downstream site accumulated more metals during the experiments than those sourced from the upstream site. We did not see a similar difference in metal content between the upstream and downstream sourced mesocosms for the control experiment. Taken together these results suggest that dietary exposure to metals or changes in microbial biomass that decrease nutritive quality (e.g. generation of excess EPS to metal exposure [e.g., Ha et al. 2010; Jittawuttipoka et al. 2013; Ding et al. 2018], or both may cause shifts in the macroinvertebrate community composition. This mechanism is further supported by the much lower abundance of functional feeding groups

in the downstream communities that would be indicative of a dietary shift from biofilm to seston feeders. A previous study (Wolff et al. 2019) found that upstream macroinvertebrates were enriched in mayflies and other “scrapers” (scraper is the common name given to insects of the functional feeding type that “scrape” biofilms from rocks as a food source) and downstream communities were enriched in Caddisflies and other seston-feeding taxa (Wolff et al. 2019).

Conclusions

In this current study we cannot conclusively link the response of the microbiome to metals to changes in diet quality of their primary consumers, stream macroinvertebrates. However, several results presented in this study are consistent with the idea that a microbial response to metals at the base of the food web may be affecting consumers one trophic level above. If this is indeed the case then it suggests that the current criteria that uses chronic exposure of aquatic macroinvertebrates to assess stream health (a threshold below which is thought to be protective of ~95% of the aquatic community) is insufficient to assess the impact of metals on stream ecosystems. It is becoming increasingly evident that dietary exposure is as important as direct exposure to aquatic life and should be considered when assessing the impact of metals on stream ecosystems (e.g., Croteau and Luoma 2008; Xie et al. 2009; Kim et al. 2012). One challenge presented here is that the metal content of macroinvertebrate diets are much harder to measure than the metal content of the surface water. However, if the dietary exposure is mediated through shifts in the microbiome, then microbial metrics, as presented here, may provide a better alternative to assess the impact of metals on stream ecosystems. Our research suggests current best practice guidelines of stream water quality (e.g., EPA aquatic life criteria) may miss impacts of metal contamination on the community that form the base of the stream ecosystems and

additional factors (e.g., dietary exposure, microbial metrics) should be included as these standards are improved.

.

Tables and Figures

Table 2. Descriptive statistics for the observational study in the upper Arkansas River. Different letters refer to a statistically significant difference ($\alpha = 0.05$) among locations (upstream, California Gulch [CalGulch], and downstream samples) in the Spring (lower case) and in the Fall (upper case). Asterisks refer to statistically significant difference ($\alpha = 0.05$) between Spring and Fall for each location.

Location	Season	n	pH		Sediment Cd		Sediment Cu		Sediment Zn		Observed OTUs		Shannon Index	
			Mean	StdDev	Mean	StdDev	Mean	StdDev	Mean	StdDev	Mean	StdDev	Mean	StdDev
Upstream	Spring	8	7.63 a*	0.21	1.59 a	0.41	4.29 a	1.15	220.24 a	92.16	4825.38 a	1131.69	6.56 a	0.35
CalGulch	Spring	4	7.12 b*	-	23.66 b	10.15	331.30 b	113.47	6300.8 b	1673.60	2414.00 b	1402.84	5.84 b	0.25
Downstream	Spring	13	7.41 c*	0.05	6.29 c	4.09	53.56 c*	69.58	1145.10 c*	590.00	4095.31 a*	1148.54	6.49 a	0.33
Upstream	Fall	8	7.75 A**	0.53	1.74 A	0.58	4.19 A	2.28	167.17 A	111.22	5288.50 A	1010.73	6.70 A	0.23
CalGulch	Fall	4	7.50 B**	-	33.45 B	16.58	382.49 B	174.56	6488.64 B	1059.45	2729.50 B	388.95	5.65 B	0.06
Downstream	Fall	14	7.49 B**	0.07	5.14 C	2.81	18.13 C**	8.24	767.12 C**	395.35	5033.29 A**	826.92	6.49 A	0.41

Table 3. Descriptive statistics for the mesocosm study. Different letters refer to a statistically significant difference ($\alpha = 0.05$) between upstream and downstream control samples (lower case), and between upstream and downstream metal-treated samples (upper case). Asterisks refer to statistically significant difference ($\alpha = 0.05$) between control and metal-treated samples for each location.

Location	Season	Treatment	n	Water Cu ($\mu\text{g/L}$)		Water Zn ($\mu\text{g/L}$)		Sediment Cu ($\mu\text{g/g}$ dry)		Sediment Zn ($\mu\text{g/g}$ dry)		Observed OTUs		Shannon Index	
				Mean	StdDev	Mean	StdDev	Mean	StdDev	Mean	StdDev	Mean	StdDev	Mean	StdDev
Upstream	Fall	Control	4	2.35 a	2.65	0.05 a	0.17	43.07 a	17.85	1311.91 a	229.62	4715.50 a	2706.24	6.78 a	0.12
Downstream	Fall	Control	4	1.33 a	0.92	0.23 a	0.77	75.93 b	19.63	1600.16 b	36.47	3728.50 a	599.13	6.30 b*	0.01
Upstream	Fall	Metals	4	17.21 A	6.31	588.85 A	57.08	162.71 A	68.09	2451.97 A	645.75	5668.25 A	3493.11	6.72 A	0.11
Downstream	Fall	Metals	4	18.04 A	4.94	613.53 A	49.32	656.28 B	161.64	7280.31 B	1578.52	3608.00 A	1007.13	6.11 A**	0.13

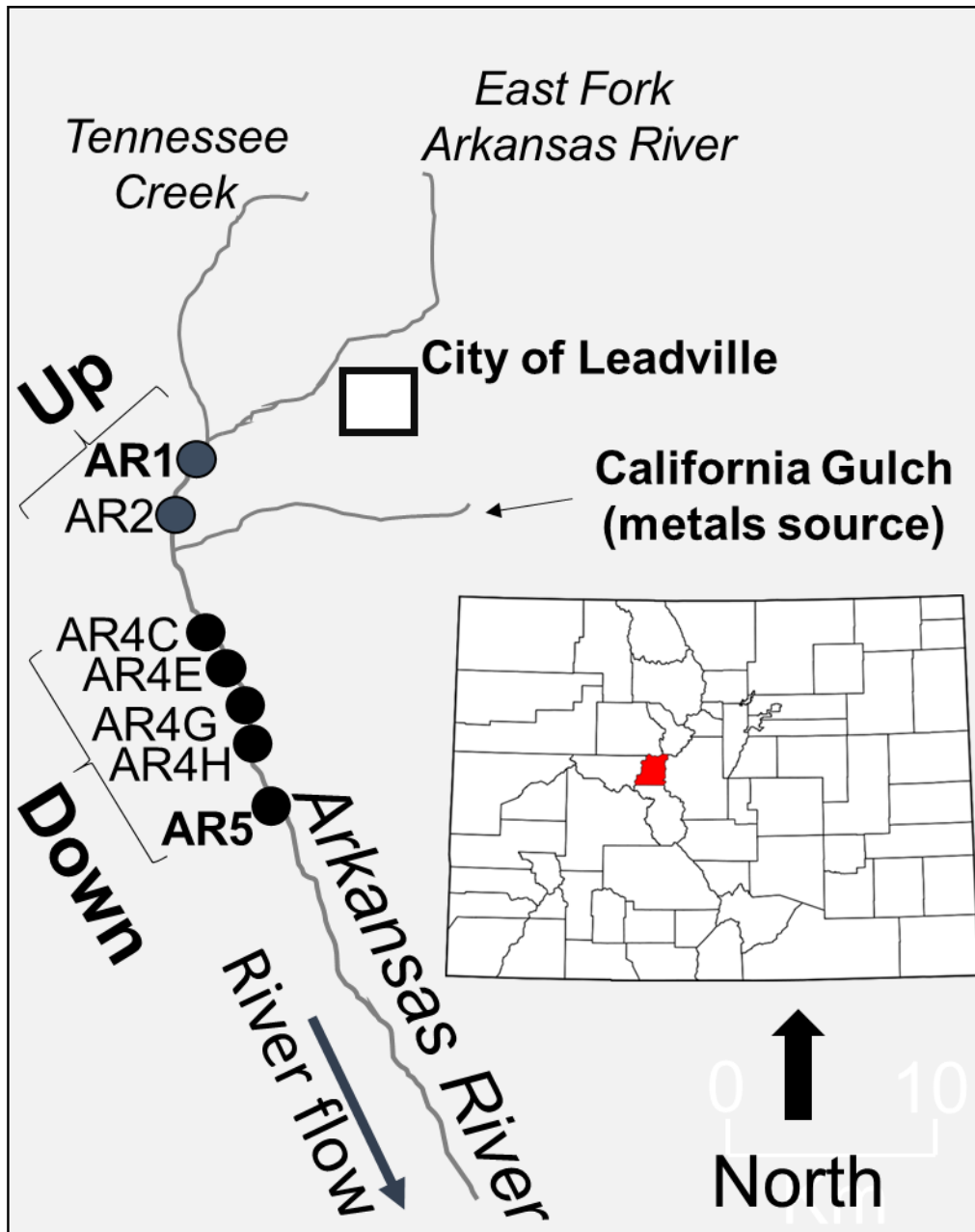


Figure 7. Map of the study area in the upper Arkansas River, Colorado, USA.

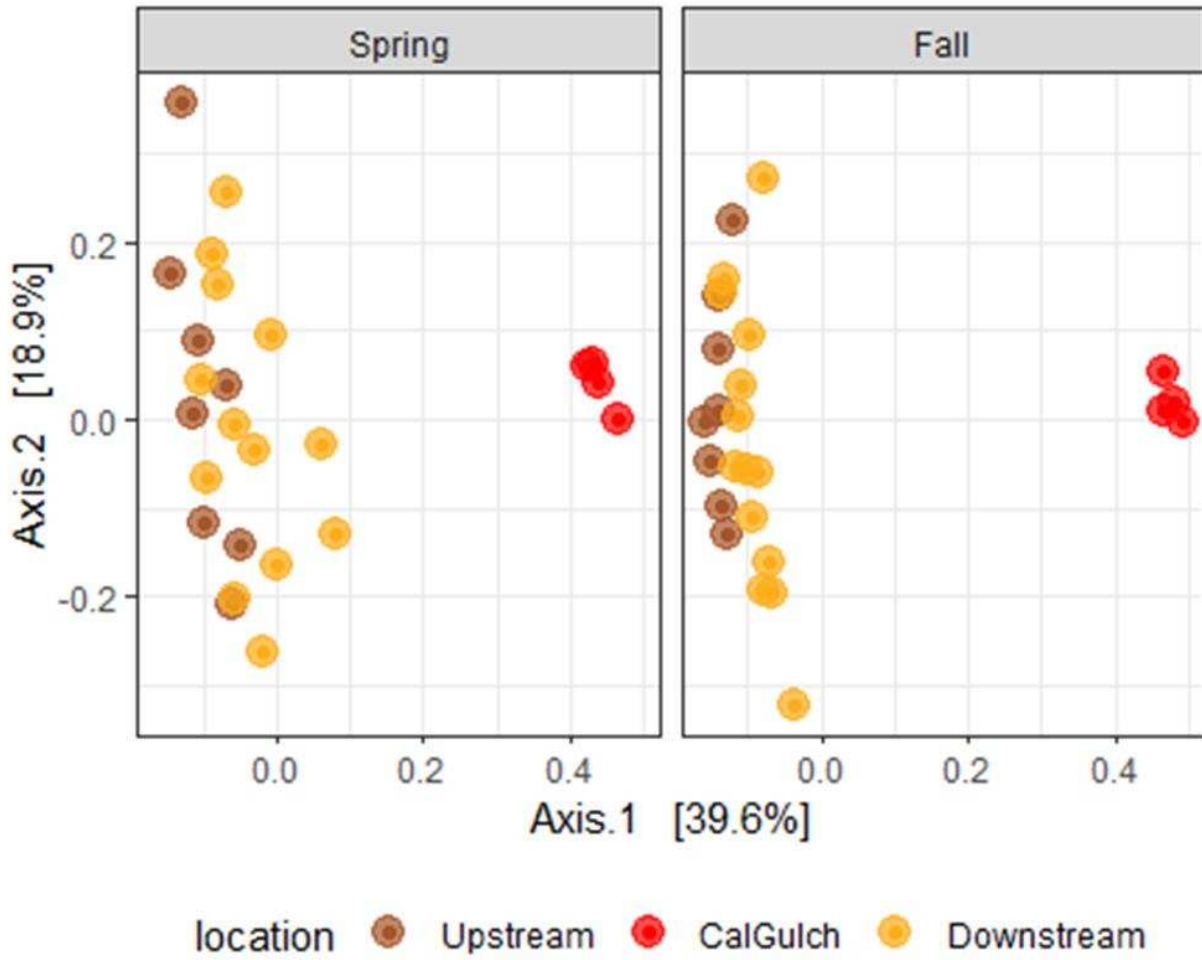
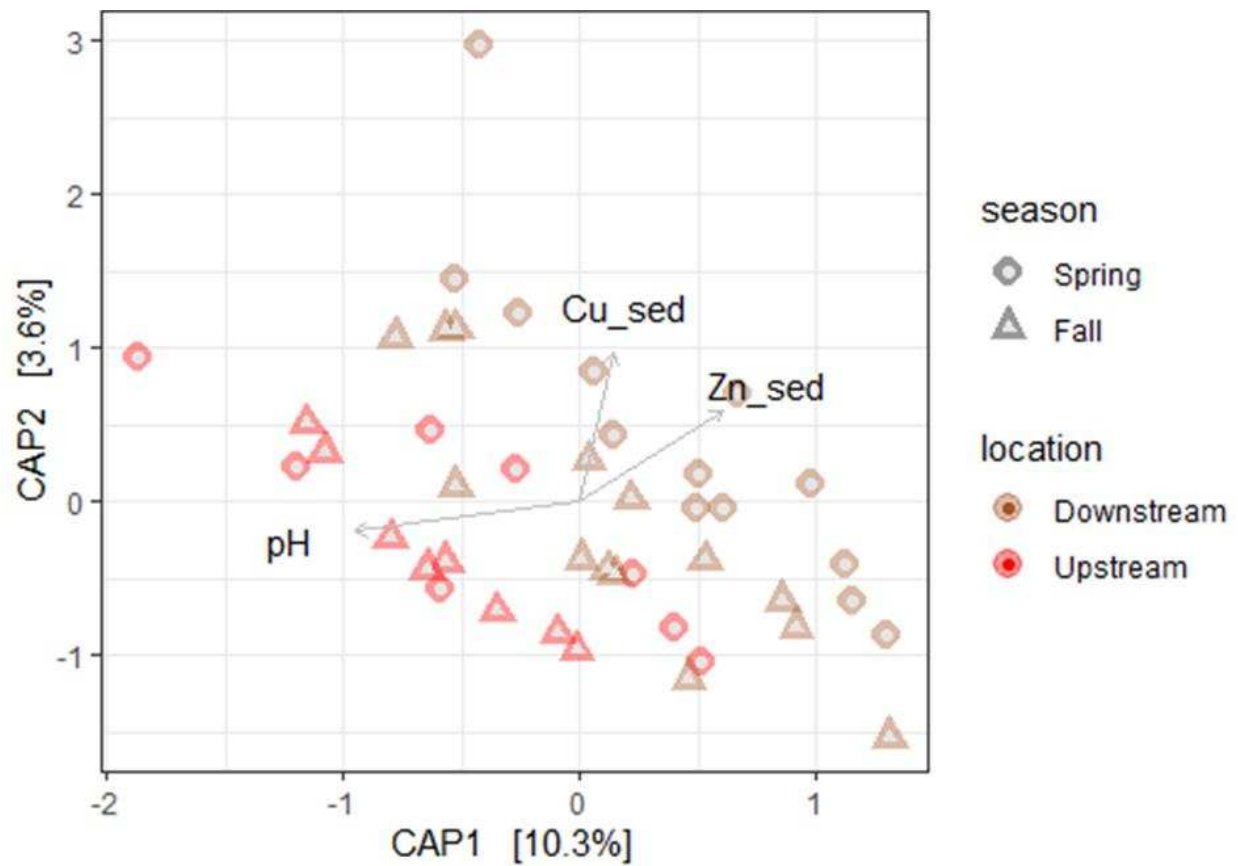


Figure 8. Principle Coordinates Analysis (PCoA) of microbiomes among locations both with California Gulch (CalGulch).



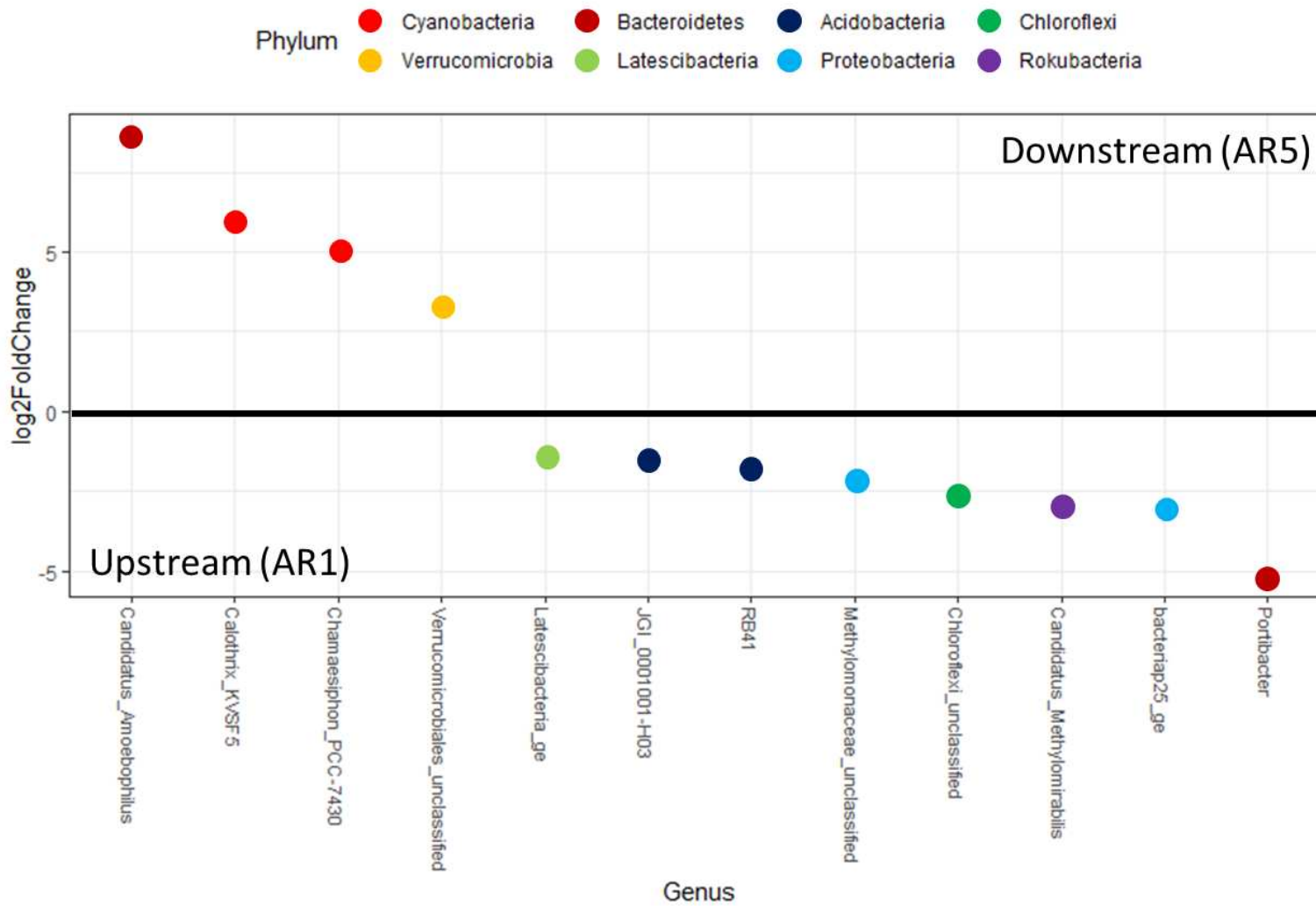


Figure 10. Log₂-fold change plots of genera between upstream (AR1) and downstream (AR5) samples. Samples with a positive value are more enriched at upstream sites, and negative values more enriched at downstream sites. The color of each dot is the phylogenetic Order.

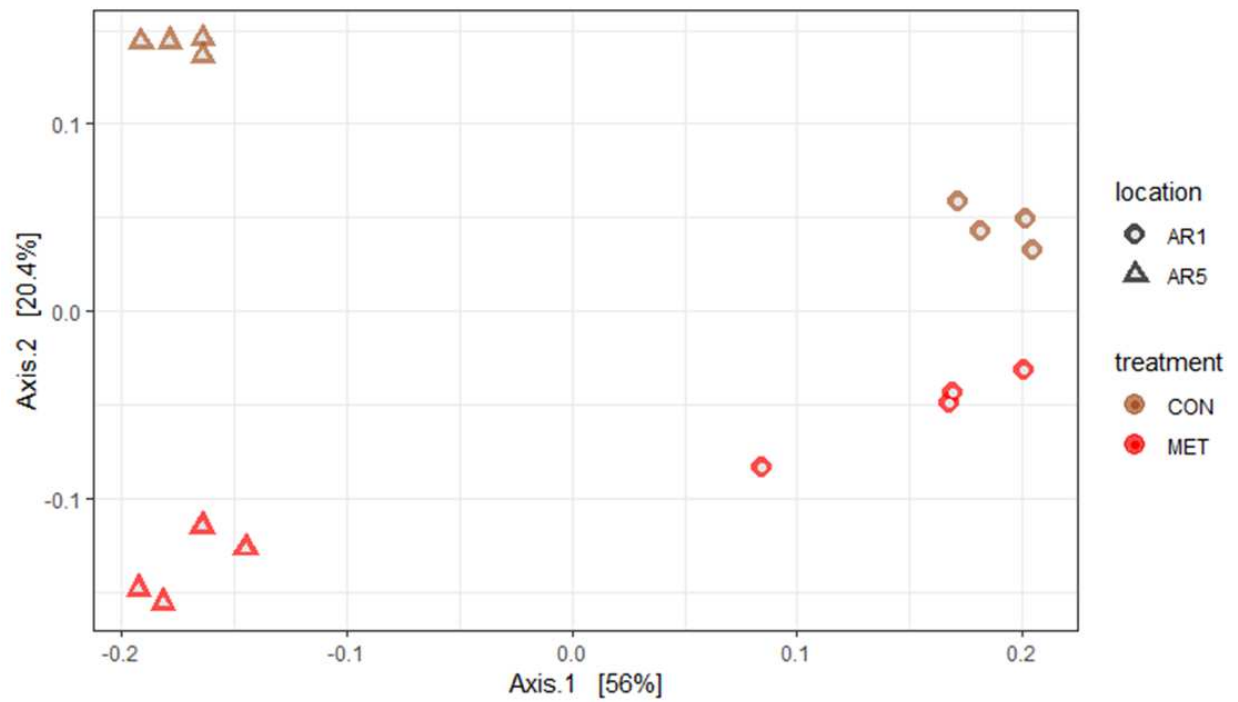


Figure 11. Principle Coordinates Analysis (PCoA) of community membership of upstream (circles) vs. downstream (triangles) sediment samples. Samples treated with metals are in red and non-treated controls are in brown.

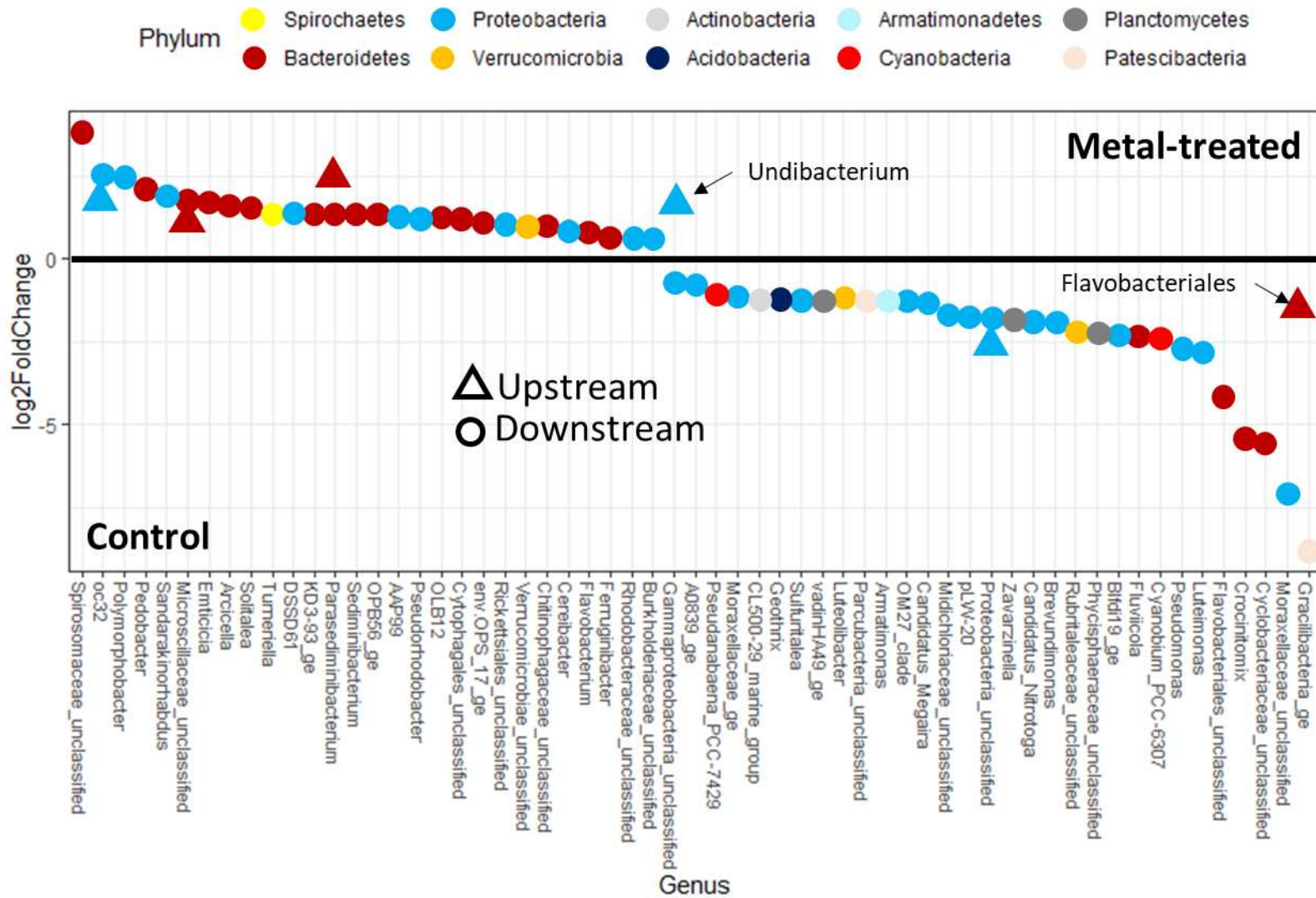


Figure 12. Log2-fold change plot between metal-treated vs. control samples from upstream (triangles) and downstream (circles) communities. Note, Unibacterium and Flaovbacteriales were drawn in the figure and do not correspond to the taxa listed on the x-axis.

Literature Cited

- Ancion, P., G. Lear, A. Dopheide, and G.D. Lewis. 2013. Metal concentrations in stream biofilm and sediments and their potential to explain biofilm microbial community structure. *Environmental Pollution* 173: 117-124.
- Bar-On, Y.M., R. Phillips, and Ron Miloš. 2018. The biomass distribution on Earth. *Proceedings of the National Academy of Sciences* 115: 6506-6511.
- Barbour, M.T., J. Gerritsen, B.D. Snyder, and J.B. Stribling. 1999. *Rapid Bioassessment Protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates, and fish*. EPA 841-B-99-002, 2nd Edition. U.S. Environmental Protection Agency. Office of Water, Washington, D.C.
- Battin, T.J., L.A. Kaplan, J.D. Newbold, X. Cheng, and C. and Hansen. 2003. Effects of current velocity on the nascent architecture of stream microbial biofilms. *Applications in Environmental Microbiology* 69: 5443-5452.
- Baxter, C.V., K.D. Fausch, and W.C. Saunders. 2005. Tangled webs: reciprocal flows of invertebrate prey link streams and riparian zones. *Freshwater Biology* 50: 201-220.
- Bier, R.L., K.A. Voss, and E.S. Bernhardt. 2015. Bacterial community responses to a gradient of alkaline mountaintop mine drainage in Central Appalachian streams. *The ISME Journal* 9: 1378-1390.
- Cadmus, P., W.H. Clements, J.L. Williamson, J.F. Ranville, J.S. Meyer, and M.J.G. Ginés. 2016. The use of field and mesocosm experiments to quantify effects of physical and chemical stressors in mining-contaminated streams. *Environmental Science and Technology* 50: 7825-7833.

- Carlisle, D.M., and W.H. Clements. 2005. Leaf litter breakdown, microbial respiration and shredder production in metal-polluted streams. *Freshwater Biology* 2: 380-390.
- Carlisle, D.M., C.P. Hawkins, M.R. Meador, M. Potapova, and J. Falcone. 2008. Biological assessments of Appalachian streams based on predictive models for fish, macroinvertebrate, and diatom assemblages. *Journal of the North American Benthological Society*, 27(1): 16-37.
- Clements, W.H. 2004. Small-scale experiments support causal relationships between metal contamination and macroinvertebrate community responses. *Ecological Applications*, 14, 954-967.
- Clements, W.H., D.M. Carlisle, J.M. Lazorchak, and P.C. Johnson. 2000. Heavy metals structure benthic communities in Colorado mountain streams. *Ecological Applications* 10: 626-638.
- Clements, W.H., N.K.M. Vieira, and S.E. Church. 2010. Quantifying restoration success and recovery in a metal-polluted stream: a 17-year assessment of physiochemical and biological responses. *Journal of Applied Ecology* 47: 899-910.
- Clements, W.H., H.P. Cadmus, C.J. Kotalik, and B.A. Wolff. 2019. Context-dependent responses of aquatic insects to metals and metal mixtures: a quantitative analysis summarizing 24 yr of stream mesocosm experiments. *Environmental Toxicology and Chemistry* 38: 2486-2496.
- Croteau, M.N., and S.N. Luoma. 2008. A biodynamic understanding of dietborne metal uptake by a freshwater invertebrate. *Environmental Science and Technology* 42: 1801–1806.
- De Pauw, N, and G. Vanhooren. 1983. Method for biological quality assessment of watercourses in Belgium. *Hydrobiologia*. 100: 153-168.

- Ding, P., W. Song, Z. Yang, and J. Jian. 2018. Influence of Zn(II) stress-induction on component variation and sorption performance of extracellular polymeric substances (EPS from *Bacillus vallismortis*. *Bioprocess and Biosystems Engineering* 41: 781-791.
- Feris, K., P. Ramsey, C. Frazar., J.N. Moore, J.E. Gannon, and W.E. Holben. 2003. Differences in Hyporheic-Zone Microbial Community Structure along a Heavy-Metal Contamination Gradient. *Applied and Environmental Microbiology* 69: 5563-5573.
- Fierer, N., J.L. Morse, S.T. Berthrong, E.S. Bernhardt, and R.B. Jackson. 2007. Environmental controls on the landscape--scale biogeography of stream bacterial communities. *Ecology* 88: 2162-2173.
- Flemming, H-C., and J. Wingender. 2000. The biofilm matrix. *Nature reviews microbiology* 8: 623-633.
- Flemming, H-C., J. Wingender, U. Szewzyk, P. Steinberg, S.A. Rice, and S. Kjelleberg. 2016. Biofilms: an emergent form of bacterial life. *Nature Reviews Microbiology* 14: 563-575.
- Gibson, B., D.J. Wilson, E. Feil, and A. Eyre-Walker. 2018. The distribution of bacterial doubling times in the wild. *Proceedings. Biological Sciences* 285: 20180789.
- Graves, J. 2017. Abandoned mines: progress, monitoring, law and regulations. *Water Resources Review Committee Presentation*. August 2017. Colorado Department of Natural Resources. Division of Reclamation, Mining and Safety Inactive Mine Reclamation Program.
- Ha, J.Y., A. Gelabert, A.M. Spormann, and G.E. Brown. 2010. Role of extracellular polymeric substances in metal ion complexation on *Shewanella oneidensis*: batch uptake, thermodynamic modeling, ATR-FTIR, and EXAFS study. *Geochimica et Cosmochimica Acta* 74: 1e15.

- Hugerth, L.W., E.E.L. Muller, Y.O.O. Hu, L.A.M. Lebrun, H. Roume, D. Lundin, P. Wilmes, and A.F. Andersson. 2014. Systematic design of 18S rRNA gene primers for determining Eukaryotic diversity in microbial consortia. *PLoS One* 9: e95567.
- Jittawuttipoka, T., M. Planchon, O. Spalla, K. Benzerara, F. Guyot, C. Cassier-Chauvat, and F. Chauvat. 2013. Multidisciplinary evidences that *Synechocystis* PCC6803 exopolysaccharides operate in cell sedimentation and protection against salt and metal stresses. *PloS one* 8; e55564.
- Jordaan, K., A.M. Comeau, D.P. Khasa, and C.C. Bezuidenhout. 2019. An integrated insight into the response of bacterial communities to anthropogenic contaminants in a river: a case study of the Wonderfonteinspruit catchment area, South Africa. *PLoS One* 14: e0216758.
- Kim, K.S., D.H. Funk, and D.B. Buchwalter. Dietary (periphyton) and aqueous Zn bioaccumulation dynamics in the mayfly *Centroptilum triangulifer*. *Ecotoxicology* 21: 2288-2296.
- Lear, G., D. Niyogi, J. Harding, Y. Dong, and G. Lewis. 2009. Biofilm bacterial community structure in streams affected by Acid Mine Drainage. *Applied and Environmental Microbiology* 75:3455-3460.
- Levinton, J.S., E. Suatoni, W. Wallace, R. Junkins, B. Kelaher, and B.J. Allen. 2003. Rapid loss of genetically based resistance to metals after the cleanup of a Superfund site. *PNAS* 100: 9889-9891.
- Merritt, R.W., and K.W. Cummins. 2019. An introduction to the aquatic insects of North America. Kendall/Hunt Publishing Company. Debuque, Iowa.
- Nealson, K.H. 1997. Sediment bacteria: Who's there, what are they doing, and what's new? *Annual Review of Earth and Planetary Sciences* 25: 403-434.

- Newton, R.J., S.E. Jones, A. Eiler, K.D. McMahon, and S. Bertilsson. 2011. A guide to the natural history of freshwater lake bacteria. *Microbiology and Molecular Biology Reviews*. 75: 14-49.
- Niyogi, D.K., W.M. Lewis, and D.M. McKnight. 2003. Direct and indirect effects of mine drainage on bacterial processes in mountain streams. *Journal of the North American Benthological Society* 22: 276-291.
- Quast, C., E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, and F.O. Glöckner. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* 41: D590-D596.
- R Core Team. 2017. R: A language and environment for statistical computing. R Foundation of Statistical Computing, Vienna, Austria.
- Reis, M.P, F.R. Barbosa, E. Chartone-Souza, and A.M.A. Nascimento. 2013. The prokaryotic community of a historically mining-impacted tropical stream sediment is as diverse as that from a pristine stream sediment. *Extremophiles* 17: 301-309.
- Rosi, E.J., H.A. Bechtold, D. Snow, M. Rojas, A.J. Reisinger, and J.J. Kelly. 2018. Urban stream microbial communities show resistance to pharmaceutical exposure. *Ecosphere* 9: e02041. [10.1002/ecs2.2041](https://doi.org/10.1002/ecs2.2041).
- Schloss, P.D., S.L. Westcott, T. Ryabin, J.R. Hall, M. Hartmann, E.B. Hollister, R.A. Lesniewski, B.B. Oakley, D.H. Parks, C.J. Robinson, J.W. Sahl, B. Stes, G.G. Thallinger, D.J. Van Horn, and C.F. Weber. 2009. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applications in Environmental Microbiology* 75:7537-41.

- Shannon, C.E., and W. Weaver. 1949. The mathematical theory of communication. The University of Illinois Press, Urbana, Illinois.
- US EPA (U.S. Environmental Protection Agency). 2002. National recommended water quality criteria, U.S. Environmental Protection Agency, Office of Water. EPA-822-R-02-047, Washington, DC.
- Wolff, B.A., W.H. Clements, and S.B. Duggan. 2019. Resilience and regime shifts: do novel benthic communities impede ecological recovery in a historically metal-contaminated stream? *Journal of Applied Ecology*. DOI: 10.1111/1365-2664.13503.
- Xie, L., D.H. Funk, and D.B. Buchwalter. 2009. Trophic transfer of Cd from natural periphyton to the grazing mayfly *Centroptilum triangulifer* in a life cycle test. *Environmental Pollution* 158: 272-277.
- Zeglin, L.H. 2015. Stream microbial diversity in response to environmental changes: review and synthesis of existing research. *Frontiers in Microbiology* 6: 454.

CHAPTER 3

PHYSICAL HABITAT RESTORATION HAS LIMITED INFLUENCE ON MACROINVERTEBRATE COMMUNITIES IN A HISTORICALLY METAL- CONTAMINATED STREAM

Introduction

The total cost for remediation of mine-impacted streams in the U.S. has been estimated to reach \$72 billion (Buxton et al. 1997). A report to the U.S. Office of Management and Budget concluded that despite spending \$500 million on water pollution abatement programs over the past 30 years, the effectiveness of these investments in meeting objectives of the Clean Water Act has not been evaluated. Further, nearly 20% of US restoration projects had no listed goals for purposes of restoration, and the inability to quantify the success of restoration in aquatic ecosystems is largely a result of the limited monitoring, where only 10% of the US restoration projects implemented pre- or post-monitoring (Bernhardt et al. 2005).

The Upper Arkansas River Basin is in the mineralized belt of the Rocky Mountains, near Leadville, Colorado, USA. The area was significantly disturbed by mining activities from the mid 1800's through the late 20th century. As a result of these activities, water quality in the Arkansas River was significantly impaired and limited the native fishery (Clements 1994; Clements and Rees 1997). The impacts were severe enough that area was added to the US EPA Priorities List (i.e., "Superfund Site") in 1983.

There were initially two metal sources entering the upper Arkansas River: the Leadville Mine Drainage Tunnel and California Gulch (Fig. 1). Prior to the start of remediation, Clements (1994) reported patterns of reduced invertebrate abundance, species richness, and differences in

community composition downstream of both metal sources. Community composition shifted from relatively metal-sensitive species (e.g., mayflies and stoneflies) at the upstream reference sites to metal-tolerant species (e.g., dipterans and caddisflies) at downstream metal-polluted sites. Although some of these effects could have been confounded with natural variation along longitudinal stream gradients (Vannote et al. 1980; Clements and Kiffney 1995), changes in community indices (e.g., abundance and richness) were consistent with metals sources and not with distance, stream geomorphology, or other physiochemical differences between sites (e.g., temperature, dissolved oxygen, hardness, or alkalinity).

Water treatment facilities began operating at the Leadville Mine Drainage Tunnel in 1992 (LMDT; upstream of California Gulch and flows into the East Fork Arkansas River) and Yak Tunnel in 1994 (flows into California Gulch] (Clements et al. 2010). Removal of mine tailings (150,000 m³) from the floodplain was completed by 1999 (Clements et al. 2010). These efforts resulted in significant improvements in water quality which improved overall benthic macroinvertebrate communities and fish populations (Clements et al. 2010). Metals concentrations in water decreased immediately at sites downstream of the LMDT water treatment facility, and water samples below California Gulch were generally below US EPA water quality criteria (e.g., hardness-adjusted Zn, Cu, and Cd. Benthic macroinvertebrate abundance and richness (particularly sensitive mayflies) also increased quickly (within a couple years) at sites upstream of California Gulch. Improvements in abundance and species richness took longer (significant increases not observed until 1999) at sites below California Gulch despite water quality improvements at the Yak tunnel quality treatment facility. It has been hypothesized that invertebrate communities upstream of California Gulch responded faster due to recolonization from nearby undisturbed communities (e.g., Tennessee Creek and East Fork

Arkansas River) (Clements et al. 2010). Despite increases in abundance and species richness, community composition has remained different between sites upstream and downstream of California Gulch. Clements et al. (2010) reported that between 2000 and 2006 the total invertebrate community was dominated by metals-sensitive taxa at sites upstream of California Gulch, whereas the downstream sites comprised up to 66% metals-tolerant taxa.

The upper Arkansas River has since been a popular angling destination for a self-sustaining non-native Brown trout, and habitat restoration has recently been implemented as part of a Natural Resources Damage Assessment (NRDA) settlement to further enhance the brown trout fishery. Habitat enhancements included planting riparian vegetation, bank stabilization, and various in-stream structures (e.g., pools, riffles, and gravel bars). The NRDA restoration goal was to provide more suitable habitats to increase trout population size and improve body condition. Physical in-stream habitat restoration as well as bank stabilization was initiated in the Fall of 2013 and was completed by Winter 2014. We stress the importance that this restoration project was designed to benefit fish populations.

To address the effects of restoration efforts on fish populations, we focused on responses of aquatic macroinvertebrates before and after habitat restoration because aquatic and terrestrial insects are the major prey resource for brown trout in this system, and we were interested in how restoration would influence this resource. Some advantages of using macroinvertebrates include limited dispersal (i.e., represent on-site conditions), variable life-stages (e.g., aquatic and terrestrial), ease of collections, variable sensitivities to stressors, and long-term integration of environmental factors not possible using only physiochemical measurements (Metcalf 1989).

Since macroinvertebrates are rarely the primary consideration of many stream restoration goals, it is difficult to evaluate how these communities typically respond to habitat

enhancements. However, a review by Palmer et al. (2010) found that only 2 of 78 independent restoration projects reported increases in invertebrate richness. Reasons for a lack of response of invertebrate communities were likely due to poor statistical design (i.e., controls and replication of assessment of restoration efforts were insufficient), lack of adequate monitoring duration, insufficient durability of structures, or differential responses of macroinvertebrates compared with fishes (Palmer et al. 2010). Although invertebrate richness may not be as important as abundance and/or biomass to support increased fish populations, it highlights the general lack of proper invertebrate monitoring.

We sought to overcome the limitations in statistical design of this analysis by using a before-after control-impact (BACI) design (Stewart-Oaten et al. 1986; Stewart-Oaten et al. 1992) over a significant period (2009 – 2017). Because BACI designs account for both temporal (before versus after) and spatial (reference versus treated sites) variation, it allows researchers to attribute responses directly to treatments and therefore makes a stronger case for cause-and-effect relationships associated with restoration efforts. The primary objective of this research was to assess how restoration affects macroinvertebrates at control and treated stream sites through a variety of measures: 1) abundance and richness of the benthic community, 2) abundance of emerging adult insects, 3) input of riparian insects, 4) prey abundance in Brown trout diets.

Methods

Study Site

The Upper Arkansas River is a 5th order stream located in northern Colorado, USA (Figure 1). The study area is located at approximately 2,820 meters above sea level, and receives an average precipitation of 29.5 cm, mostly in the form of snow during the winter. Substrate at

all sites was dominated by medium to large cobble in a matrix of gravel and sand. The primary riparian vegetation includes sagebrush (*Artemisia* spp.), grasses, and willow (*Salix* spp.) trees. The Arkansas River has a hydrograph typical of a snowmelt fed stream, with peak discharges in late Spring to early Summer, and reaching base flow conditions by late Summer or early Fall. Four study sites (AR4C, AR4E, AR4G, AR4H) were monitored along a 4 km reach of the Upper Arkansas River that was restored as part of a NRDA (Natural Resources Damages Assessment) settlement. One additional site downstream from the restoration area (AR5) that did not receive habitat treatment was used as a control. The treatment sites were located approximately 5 km downstream from California Gulch, the historical source of metals to the river, and included one site (AR5) that has been sampled 2-3 times annually since 1989 (Clements et al. 2010).

Benthic macroinvertebrates

We collected five quantitative benthic macroinvertebrate samples from riffle habitats at each site using a modified Hess sampler (0.1 m²) fitted with a 350 µm mesh net in mid-August in 2012 – 2017. Samples were washed through a 350 µm sieve and organisms were preserved in 80% ethanol and transported to Colorado State University. Samples were sorted in the laboratory and identified to the lowest practical taxonomic level – generally genus or species for most taxa. Chironomids were identified to subfamily or tribe.

Aquatic insect emergence

We assessed aquatic macroinvertebrate emergence using three floating emergence traps at each site. Emergence traps were placed at scour pools along the stream margin for two 24-hour periods each August. Previous work has indicated that pools have higher rates of emergence

(Iwata 2007, Malison et al. 2010). Floating emergence traps were modified from Malison et al (2010). This modified emergence trap design has been described previously (Cadmus et al. 2016). Briefly, emergence traps had the same base dimensions (0.33 m²) as a Malison et al. (2010) design, but the shape was changed to a pyramid with an ethanol-filled collection bottle at the top. In the laboratory, aquatic larvae were typically identified to the following 5 categories: Ephemeroptera, Plecoptera, Trichoptera, Aquatic Diptera, and Aquatic Other.

Riparian inputs

We collected adult aquatic and terrestrial samples from the riparian area at each site using plastic pan traps (100 cm X 41 cm X 15 cm deep) placed along the stream bank. We deployed the pan traps (n = 3) for two 24-hour intervals in August at the same time emergence nets were deployed. We chose this time period because abundance of terrestrial inputs into the river were hypothesized at their highest level (Saunders and Fausch 2012). We filled each plastic pan with stream water and approximately 5 mL of unscented, biodegradable surfactant to reduce surface tension. To incorporate the diversity of different riparian habitats, at each site we selectively deployed a pan trap at each of one of the three dominant habitats: willow, grass, and gravel bars. Traps were placed in areas that had homogenous habitat for at least 5 m upstream, downstream, and lateral of the pan placement.

Macroinvertebrates were removed from the pan traps using a small aquarium net and preserved in 80% ethanol. In the laboratory, aquatic larvae were typically identified to the following 7 categories: Ephemeroptera, Plecoptera, Trichoptera, Aquatic Diptera, Aquatic Other, Terrestrial Diptera, and Terrestrial Other. Taxa that utilize aquatic habitats for any part of their

life history were considered aquatic, while taxa that complete their entire life history in the terrestrial environment were considered terrestrial (Saunders and Fausch 2007).

Brown Trout Diets

Brown trout (*Salmo trutta*) were collected for diet analyses from four sites (AR4C, AR4E, AR4G, and AR4H) in mid-August using electroshocking. Stomach sampling coincided with benthic macroinvertebrate and pan trap sampling. Trout (n=17-25 per site) were sampled for 2 years before habitat restoration began (2012 – 2013) to quantify natural annual variability in feeding habits, and for four years following the restoration (2014 – 2017). Stomach contents were recovered from each trout using gastric lavage (Waters et al. 2004; Saunders and Fausch 2007). We selected trout generally between 120 – 350 mm total length because larger fish tend to be piscivorous, and smaller fish are difficult to sample effectively without causing harm (Saunders and Fausch 2007). Stomach contents were sieved through a 350 μ m screen and preserved in ethanol in the field. Trout recovered in a live well before being returned to the stream. In the laboratory, aquatic larvae were typically identified to the following categories: Ephemeroptera, Plecoptera, Trichoptera, Aquatic Diptera, Aquatic Other, Terrestrial Diptera, and Terrestrial Other.

Statistical Analyses

Abundance data were $\log_{10}(x+1)$ transformed in order to satisfy assumptions of parametric statistics. Analysis of variance (ANOVA) was used to test for differences in the abundances of benthic macroinvertebrates, emerging insects, terrestrial and aquatic insects in pan traps and brown trout prey items. Within each sampling type (benthic, emergence, pan traps, and

fish diets) we tested the effects of restoration (pre vs. post), treatment (4 restored sites vs. AR5 control site), year (2009 – 2017), habitat (pan traps only; grass, gravel, and willow), season (benthic samples only; Spring, Summer, and Fall), and a restoration x treatment interaction term. Feeding habits of Brown trout were not evaluated at the control site. While effects of year, site, habitat, and season were incorporated to better explain variance in the models, we were primarily interested in the restoration x treatment term (i.e., determining if restoration at the treatment sites were different than at the control site). We only evaluated changes in richness from benthic samples since taxa were generally identified to the genus or species, whereas taxonomic resolution was much more generalized in emergence, pan trap, and fish diet samples.

We excluded data collected in 2014 from the statistical analyses because ongoing restoration caused a large decline in abundance during that period (Figure 2). Most statistical analyses were performed in R (R Core Development Team 2010). Benthic macroinvertebrate community composition was examined using Permutational ANOVA (PERMANOVA) and Non-metric Multi-Dimensional Scaling (NMDS) plots using Primer-e (Clarke and Gorley 2006).

Results

Although we found significant effects of restoration on abundance for most taxonomic groups (e.g., all groups except Ephemeroptera and Aquatic Other), the restoration x treatment interaction was not significant among any of the taxonomic groups (Table 1). This suggests that benthic abundance differed significantly before and after restoration, but these differences were similar between control and treatment sites (Figure 3). In fact, most of the variance explained in the models was driven primarily by habitat and year.

Results of benthic richness ANOVAs closely mirrored the abundance results. The effect of restoration on richness was significant for most groups, but the restoration x treatment interaction term was only significant for Plecoptera, and most of the variance explained in the models was driven primarily by habitat and year (Table 1). Therefore, except for Plecoptera, differences in richness were indistinguishable between control and treatment sites between pre- and post-restoration.

We found that abundance of emerging insects was generally lower post-restoration (Figure 4). As with benthic samples, there were significant effects of restoration for most taxonomic groups, but none had a significant restoration x treatment interaction. Importantly, none of the effects in our models explained the variance well, with adjusted R^2 values ranging from $< 0.01 - 0.15$ (Table 1).

Like emergence samples, abundance after restoration was lower than pre-restoration levels and was significant for most groups (Figure 5). Only Ephemeroptera had a significant interaction effect, but that was driven by a reduction of Ephemeroptera post-restoration at the control site whereas the treatment sites remained nearly the same. Among all the effects examined, year and habitat explained the most variance in composition of immigrating invertebrates from the riparian zone.

We did not measure feeding habits at station AR5 (the control site); however, we found that year and site (e.g., AR4C, AR4E, AR4G, and AR4H) explained the most variance (Figure 6, Table 1). On average differences were minimal between pre- and post-restoration. However, after 2014 the number of insects per fish stomach increased above pre-restoration levels (Figure 2) as indicated by the highly significant year effect.

Because changes in community composition along spatial or temporal gradients can often occur in the absence of changes in species richness (Dornelas et al. 2014; Mori et al. 2018), we visually assessed the effects of restoration on the entire community using non-metric multidimensional scaling (NMDS). The NMDS plots show that overall community composition is not different between control and treatment sites pre- and post-restoration, and the best separation of groups was by season (Figure 7). Results of PERMANOVA confirmed our NMDS results and showed that restoration x treatment was statistically significant (p value = 0.001) but did not explain nearly as much variance as season (F value of 5.38 vs. 166.61, respectively).

Discussion

The relationship between biological diversity and habitat complexity is well established in some ecosystems and a central paradigm in ecology (e.g., MacArthur and MacArthur, 1961; Menge et al. 1985 Palmer et al., 1997), but this relationship has not been critically tested in stream benthic communities (Lepori et al. 2005; Roni et al. 2006). Specifically, there has been limited assessment of how restoration efforts have effected stream macroinvertebrates. Macroinvertebrates are essential primary consumers in stream ecosystems, linking primary producers to brown trout populations. Improving brown trout population density was the motivation for the restoration efforts. We found that habitat restoration had very little influence on macroinvertebrate abundance, richness, and community composition three years after restoration was completed on the upper Arkansas River. This result is consistent with the limited number of studies that have evaluated restoration efforts on macroinvertebrate communities (e.g., Miller et al. 2010; Luohi et al. 2011; Smith et al. 2019), _ Although we observed considerable variability in benthic samples, emergence nets and pan traps over the 8-year study, there was

little evidence that aquatic or riparian communities responded to habitat restoration. Our results also demonstrate the strength of BACI designs and the importance of continuous monitoring at reference and treatment sites throughout the period of restoration. If we simply compared abundances during the last year of monitoring (2017) to baseline values, we would have concluded that restoration increased macroinvertebrate abundances by 22%; however, control site abundances also increased by 18% during this same period. Therefore, the effect of restoration is not likely very influential compared to natural temporal variation.

A major challenge in evaluating the effects of restoration on macroinvertebrates is that nearly all stream restoration projects are designed to benefit fishes. As a result, the influence of restoration on macroinvertebrates reported in the literature has been mixed. Like our study, Luohi et al. (2011) found that macroinvertebrate density decreased following in-stream restoration designed to benefit brown trout in Finland. Importantly, Luohi et al. (2011) performed their study over a similar timeframe as our study (three years monitoring before and after the restoration). Further, their long-term (15- to 17-year restoration recovery period) space-for-time comparative study detected very limited benefits to restoration (Luohi et al. 2011). Miller et al. (2010) performed a meta-analysis assessing effects of 24 restoration projects on macroinvertebrates and found that richness was generally improved but density of macroinvertebrate populations effects were negligible. Smith et al. 2019 also reported no significant effect of restoration on stream macroinvertebrates using a BACI-designed study at Big Spring Run, Pennsylvania, USA. However, a few previous studies reported increases of abundance or richness of macroinvertebrate communities following restoration (e.g., Selego et al. 2012; Nuttle et al. 2017; Li et al. 2018). In a meta-analysis of 91 restoration projects Kail et al. (2015) effects of habitat improvements on macroinvertebrates, fishes, and macrophytes were

strongly influenced by land use, river width, and project age (Kail et al. 2015). Interestingly, they also found that effects of restoration were generally weakened with increasing project age, and response ratios of macrophyte abundance turned from positive to negative after approximately 10 years (Kail et al. 2015).

A potential limitation of this study is that our emergence trap, pan trap, and fish diets were sampled only once per year. Because emergence and terrestrial input vary seasonally and among taxa (Raitif et al. 2018; Minshall 2011; Füreder et al. 2005; Ivković et al. 2013), we may have observed different responses with more intensive sampling. However, benthic communities were sampled three times per year (Spring, Summer, and Fall), and these data also showed little response to restoration treatments. In addition, our results are based on abundance and taxonomic richness and do not reflect changes in biomass, which may be more sensitive to restoration treatments. However, since community composition and abundance did not respond to restoration, it is unlikely that biomass would be significantly affected by these treatments. Since that the restoration treatments implemented in this stream were designed to improve habitat for brown trout populations, it is possible that treatments designed specifically to improve macroinvertebrate abundance and richness (e.g., greater substrate complexity, reduced embeddedness, and large woody debris) may have had a greater effect on benthic communities.

Since invertebrate abundance and richness remained relatively stable between control and treatment sites pre- and post-restoration also evaluated the effects of restoration on community composition. Overall, community composition changed very little in our study area, except for subtle changes in benthic communities related to seasonality (Figure 7). The reason for the lack of recovery of metal-sensitive taxa in the downstream reach is still unknown, particularly when other investigators have reported improved benthic communities following reductions in metal

concentrations (Herbst et al. 2018; Hornberger et al. 2009). We hypothesized that if enough of the habitat was disturbed from heavy equipment operation during the restoration period, metal-sensitive taxa could recolonize available habitat. However, in-stream construction only denuded habitat in patches in each treatment reach, which were rapidly recolonized by metal-tolerant taxa. Therefore, the continued lower abundance of metal-sensitive taxa could in part be due to the presence of larger, metal-tolerant taxa (e.g., *Brachycentrus* and *Arctopsyche* caddisflies) that have dominated the downstream habitat even after significant reductions in metal concentrations for over 20 years (Wolff et al. 2019). Thus, the competitive advantages of caddisflies deter colonizing of metal-sensitive mayflies, particularly since investigators have shown strong competitive interactions between these two groups (Dudley et al. 1990; Kohler 1992).

Another potential explanation for continued lower abundance of metal-sensitive taxa is that metal loading from California Gulch which reaches downstream sites may still be driving macroinvertebrate communities more than explained by physical habitat. However, water column metal concentrations downstream of California Gulch are typically below EPA chronic aquatic life criteria that are considered conservative thresholds that should protect most species (US EPA 2002). Additionally, mesocosm experiments using macroinvertebrate communities from AR5 (control site in this study) exposed to zinc concentrations near the EPA hardness-adjusted criteria demonstrated a lack of mortality to most (> 95%) of taxa, including metal-sensitive groups such as grazing mayflies (Clements 2004). However, results from four mesocosm experiments performed from 2009-2017 showed that communities downstream of California Gulch consistently responded differently to experimental metal exposure compared to communities upstream of California Gulch (Wolff et al. 2019). Metal loading may not be sufficiently elevated to provide direct toxicity to aquatic macroinvertebrates, but high enough to

bioaccumulate stream benthic biofilms. Indeed, the microbiome community membership has been shown to be consistently different between upstream and downstream communities (Wolff et al. 2020; *in preparation*). Changes in microbiome communities may influence macroinvertebrate taxa, particularly those that burrow in interstitial sediments or mayfly scrapers and grazers through changes in diet composition or palatability. Metals may also bioaccumulate in the benthic biofilms that can produce toxic effects when consumed (Xie et al. 2009).

Improvements in improve water quality had dramatic effects on benthic communities in the Arkansas River (Clements 1994; Clements et al. 2010; Wolff et al. 2019). However, the general lack of response of macroinvertebrates to physical restoration suggests that restoration success may be largely dependent on watershed-scale effects and does not necessarily occur at the reach-scale (Roni et al. 2008). We have previously shown that historical metal contamination resulted in a potential regime shift of benthic communities in the upper Arkansas River (Wolff et al. 2019) and that these novel ecosystems (Hobbs et al. 2009) may not return to pre-disturbance conditions despite significant improvements in water quality. We speculate that the same ecological characteristics that maintain these novel benthic communities in the Arkansas River also make them less responsive to habitat restoration.

In spite of a lack of effect to macroinvertebrates it is plausible that the lack of an effect of habitat restoration on macroinvertebrate abundances was masked by an increase in brown trout populations. Colorado Parks and Wildlife recently reported that brown trout density and biomass after habitat restoration increased above baseline conditions by 10% and 18%, respectively, across all sites (Richer 2019). Fish density was not different between control and treatment sites, but fish biomass was 26% higher in treatment sites (Richer 2019). It is possible that restoration activities created greater habitat heterogeneity to support greater numbers of brown trout that

consumed “excess” aquatic macroinvertebrates that may have also increased abundance following restoration (Richer et al. 2019). Therefore, restoration likely has a positive effect on fish populations and therefore we would not conclude that restoration is a failure. Rather the effects of restoration across multiple trophic levels may need to be further researched to determine why fish populations (and stomach contents of individuals) seems to increase but the overall macroinvertebrate community seemed relatively unimpacted.

Tables and Figures

Table 4. Statistical results for each group model within sampling type. Significant p-values (< 0.05) are in bold text. Abbreviations: Eph = Ephemeroptera, Plec = Plecoptera, Tric = Tricoptera, Dipt = Aquatic Diptera, Other = Aquatic other, Terr. Dipt = Terrestrial Diptera, and Terr. Other = Terrestrial Other.

Group	Adj. R ²	Restoration		Treatment		Site		Year		Season		Habitat		Restoration x Treatment	
		F value	p value	F value	p value	F value	p value	F value	p value	F value	p value	F value	p value	F value	p value
Benthic Abundance															
Total	0.26	6.68	0.01	0.08	0.78	9.93	< 0.01	15.05	< 0.01	50.05	< 0.01	N/A	N/A	0.64	0.42
Eph	0.36	0.15	0.70	< 0.01	0.98	14.18	< 0.01	17.33	< 0.01	104.19	< 0.01	N/A	N/A	2.39	0.12
Plec	0.14	9.27	< 0.01	0.12	0.73	1.07	0.36	11.67	< 0.01	11.27	< 0.01	N/A	N/A	0.01	0.91
Tric	0.33	28.75	< 0.01	16.30	< 0.01	1.21	0.31	14.99	< 0.01	83.61	< 0.01	N/A	N/A	0.16	0.69
Dipt	0.38	9.50	< 0.01	1.03	0.31	5.88	< 0.01	25.30	< 0.01	97.70	< 0.01	N/A	N/A	0.40	0.53
Other	0.33	0.04	0.84	13.58	< 0.01	69.02	< 0.01	10.71	< 0.01	6.95	< 0.01	N/A	N/A	0.03	0.87
Benthic Richness															
Total	0.33	11.21	< 0.01	20.05	< 0.01	4.10	< 0.01	6.23	< 0.01	111.17	< 0.01	N/A	N/A	0.61	0.43
Eph	0.22	14.10	< 0.01	11.29	< 0.01	10.30	< 0.01	10.24	< 0.01	32.66	< 0.01	N/A	N/A	0.43	0.51
Plec	0.21	32.84	< 0.01	0.30	0.58	0.82	0.48	8.02	< 0.01	43.30	< 0.01	N/A	N/A	6.04	0.01
Tric	0.36	4.49	0.03	26.17	< 0.01	1.56	0.20	13.51	< 0.01	119.70	< 0.01	N/A	N/A	3.55	0.06
Dipt	0.15	14.20	< 0.01	46.67	< 0.01	4.80	< 0.01	6.05	< 0.01	4.50	0.01	N/A	N/A	0.22	0.64
Other	0.11	0.12	0.73	0.06	0.81	3.12	0.03	10.63	< 0.01	2.46	0.09	N/A	N/A	3.36	0.07
Emergence Abundance															
Total	0.03	6.49	0.01	6.30	0.01	0.19	0.90	2.42	0.04	N/A	N/A	N/A	N/A	0.17	0.68
Eph	0.15	17.81	< 0.01	3.58	0.06	1.16	0.32	7.43	< 0.01	N/A	N/A	N/A	N/A	0.36	0.55
Plec	< 0.01	3.37	0.08	0.08	0.78	0.45	0.72	0.67	0.65	N/A	N/A	N/A	N/A	2.63	0.11
Tric	0.07	8.30	< 0.01	11.75	< 0.01	0.33	0.80	3.28	< 0.01	N/A	N/A	N/A	N/A	0.88	0.35
Dipt	0.04	9.68	< 0.01	1.33	0.25	0.20	0.89	3.12	< 0.01	N/A	N/A	N/A	N/A	0.10	0.75
Other	< 0.01	0.03	0.86	1.74	0.20	0.49	0.69	1.20	0.34	N/A	N/A	N/A	N/A	0.08	0.78
Pan Trap Abundance															
Total	0.39	8.15	< 0.01	0.62	0.43	1.18	0.32	29.62	< 0.01	N/A	N/A	1.13	0.32	2.05	0.15
Eph	0.37	69.23	< 0.01	4.36	0.04	1.16	0.33	10.02	< 0.01	N/A	N/A	0.39	0.68	4.15	0.04
Plec	0.25	9.79	< 0.01	2.85	0.09	5.17	< 0.01	2.12	0.07	N/A	N/A	13.72	< 0.01	1.33	0.25
Tric	0.47	60.27	< 0.01	1.31	0.25	2.08	0.10	11.65	< 0.01	N/A	N/A	46.86	< 0.01	1.28	0.26
Dipt	0.41	2.64	0.11	0.40	0.53	1.00	0.39	33.79	< 0.01	N/A	N/A	0.17	0.85	2.09	0.15
Other	0.11	9.89	< 0.01	7.06	< 0.01	1.14	0.33	1.62	0.16	N/A	N/A	2.59	0.08	0.43	0.51
Terr. Dipt	0.48	10.73	< 0.01	7.03	< 0.01	3.51	0.02	19.48	< 0.01	N/A	N/A	52.70	< 0.01	< 0.01	0.93
Terr. Other	0.42	0.23	0.63	3.75	0.05	0.75	0.52	20.41	< 0.01	N/A	N/A	32.88	< 0.01	0.13	0.72
Fish Diet Abundance															
Total	0.09	2.32	0.13	N/A	N/A	0.48	0.70	13.08	< 0.01	N/A	N/A	N/A	N/A	N/A	N/A
Eph	0.20	1.77	0.18	N/A	N/A	17.30	< 0.01	9.24	< 0.01	N/A	N/A	N/A	N/A	N/A	N/A
Plec	0.11	1.01	0.32	N/A	N/A	3.86	0.01	4.32	< 0.01	N/A	N/A	N/A	N/A	N/A	N/A
Tric	0.08	0.65	0.42	N/A	N/A	1.49	0.22	10.23	< 0.01	N/A	N/A	N/A	N/A	N/A	N/A
Dipt	0.16	6.25	0.01	N/A	N/A	3.42	0.02	16.77	< 0.01	N/A	N/A	N/A	N/A	N/A	N/A
Other	0.07	3.99	0.05	N/A	N/A	1.53	0.21	2.30	0.08	N/A	N/A	N/A	N/A	N/A	N/A
Terr. Dipt	0.10	2.19	0.14	N/A	N/A	1.42	0.24	3.29	0.03	N/A	N/A	N/A	N/A	N/A	N/A
Terr. Other	0.09	6.46	0.01	N/A	N/A	3.58	0.02	0.82	0.49	N/A	N/A	N/A	N/A	N/A	N/A

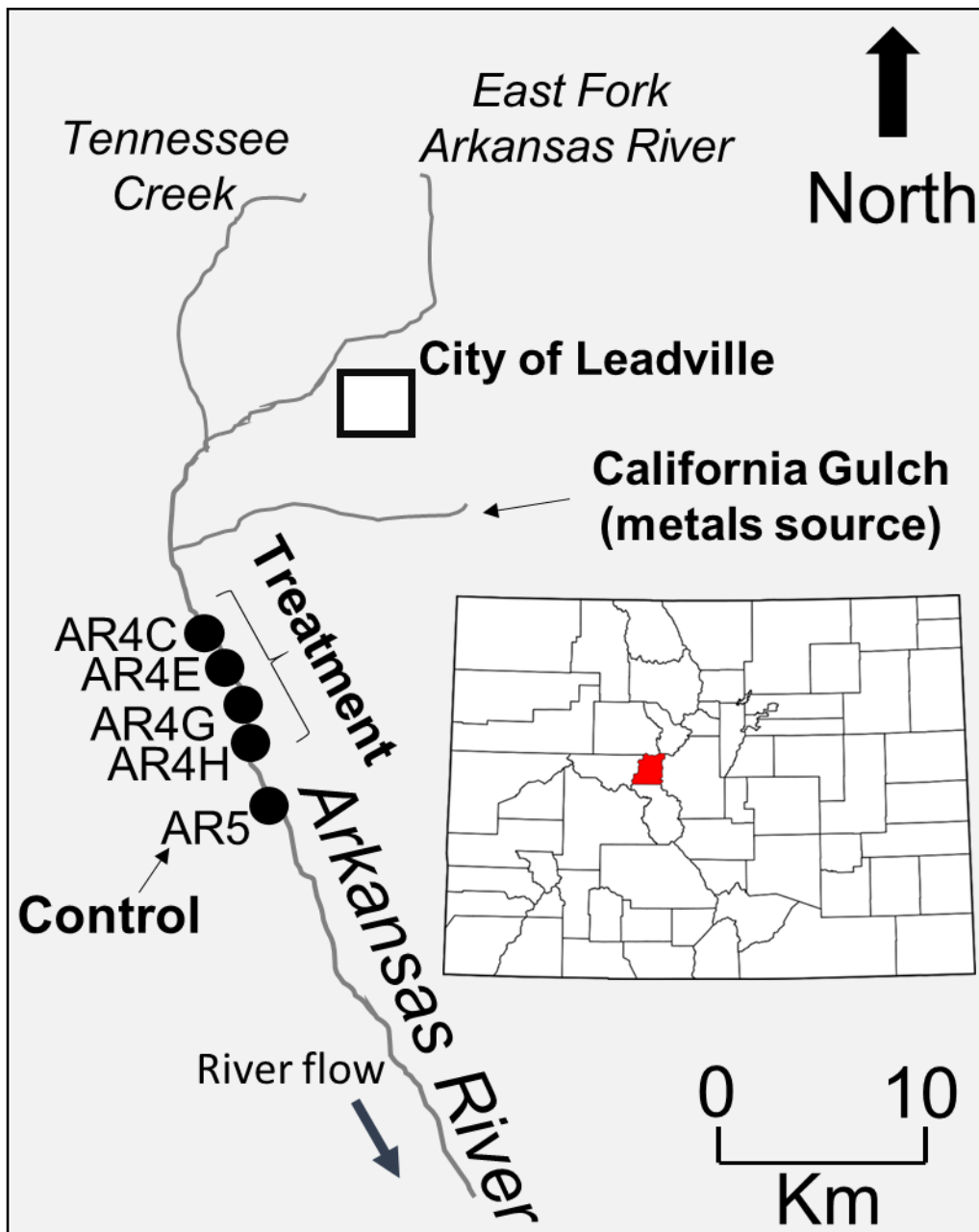


Figure 13. Map of the study area in the upper Arkansas River. The inset is a map of Colorado, USA.

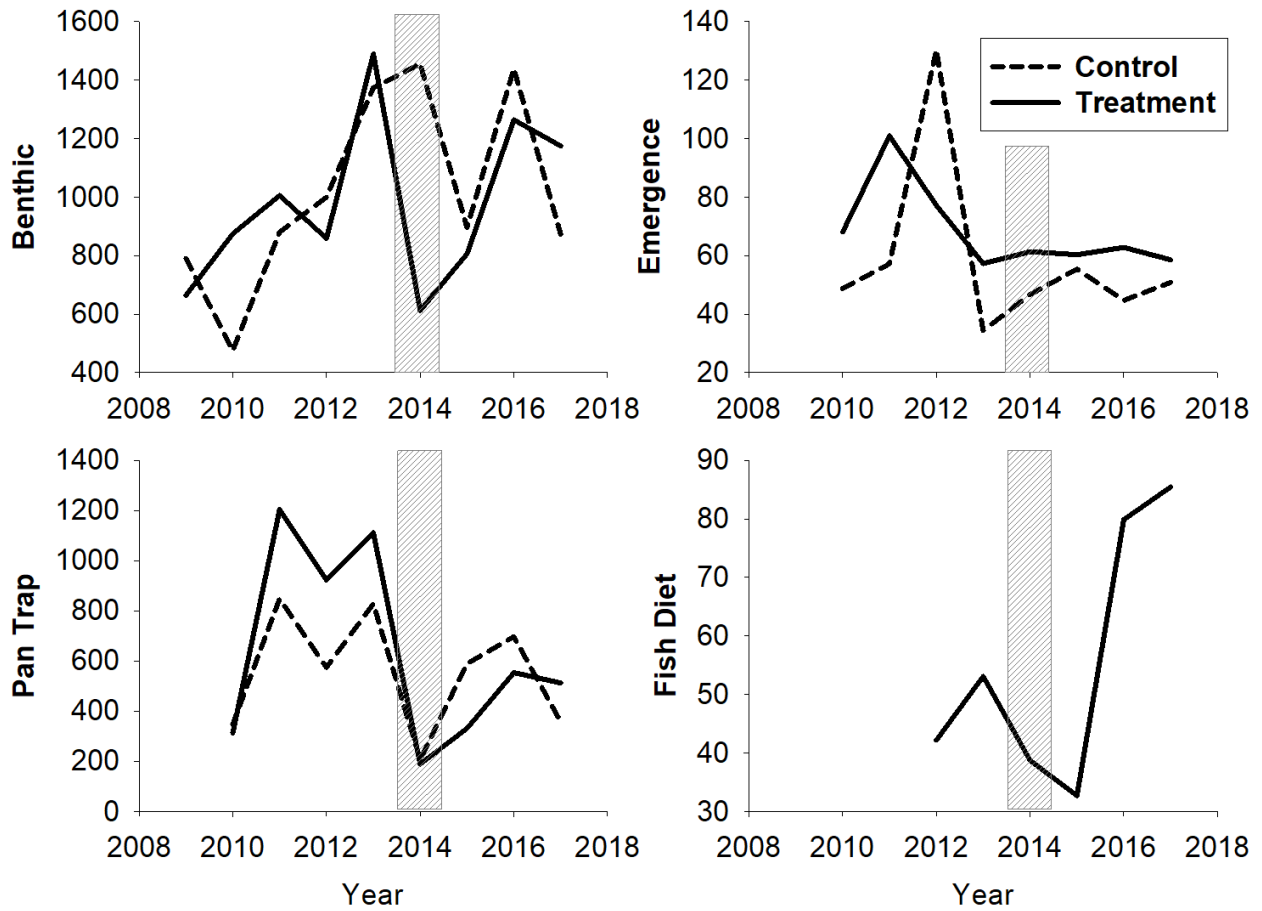


Figure 14. Mean total abundance at the control site (AR5; dashed line) and the NRDA treatment sites (AR4C, AR4E, AR4G, and AR4H combined; solid line) from each sampling method, 2009 – 2017. The shaded box represents the period of physical habitat restoration at the treatment sites.

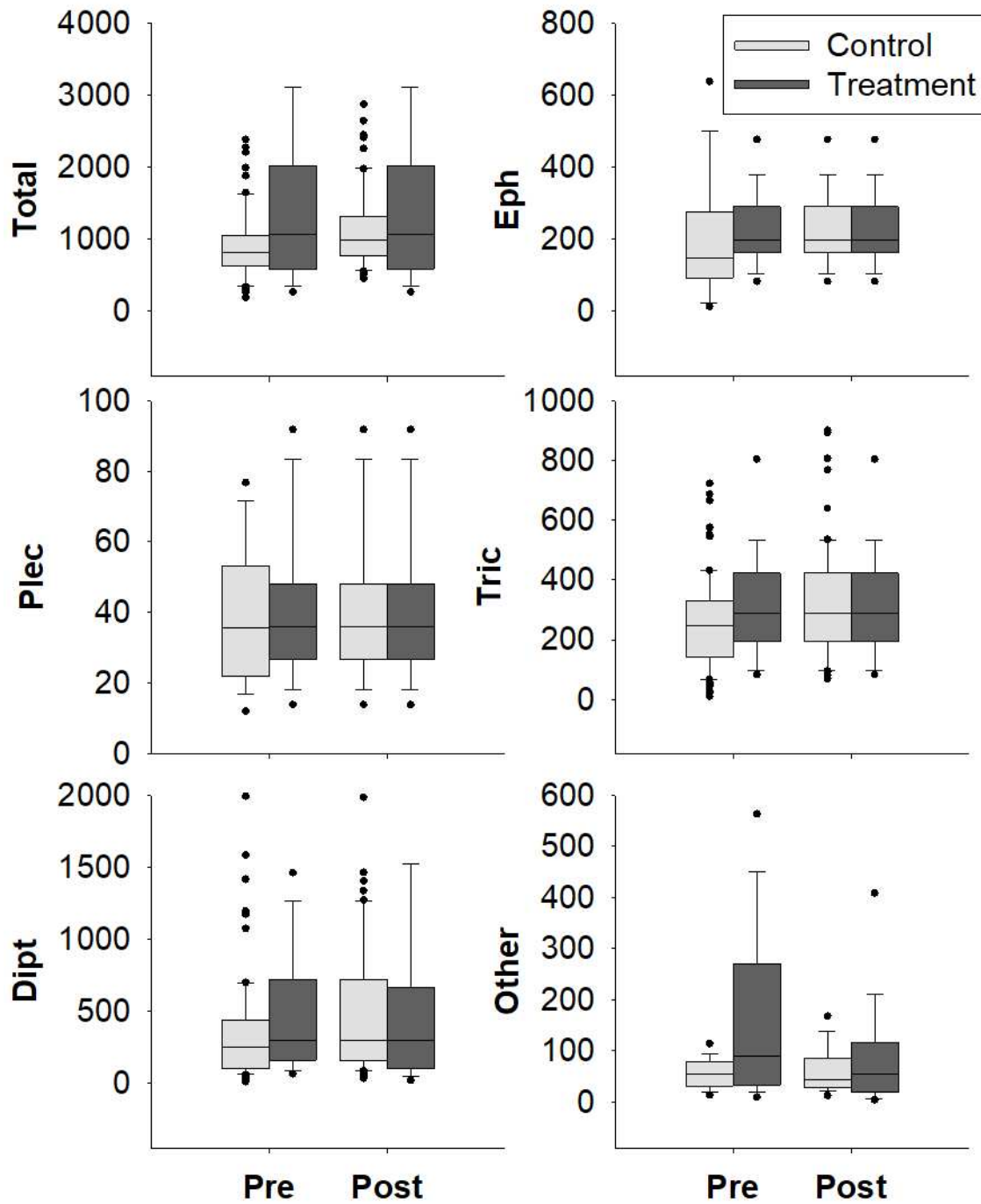


Figure 15. Benthic abundance (No. per Hess) results pre- and post-restoration at the control site (AR5; light gray) and the NRDA treatment sites (AR4C, AR4E, AR4G, and AR4H combined; dark gray). See Table 1 for abbreviations.

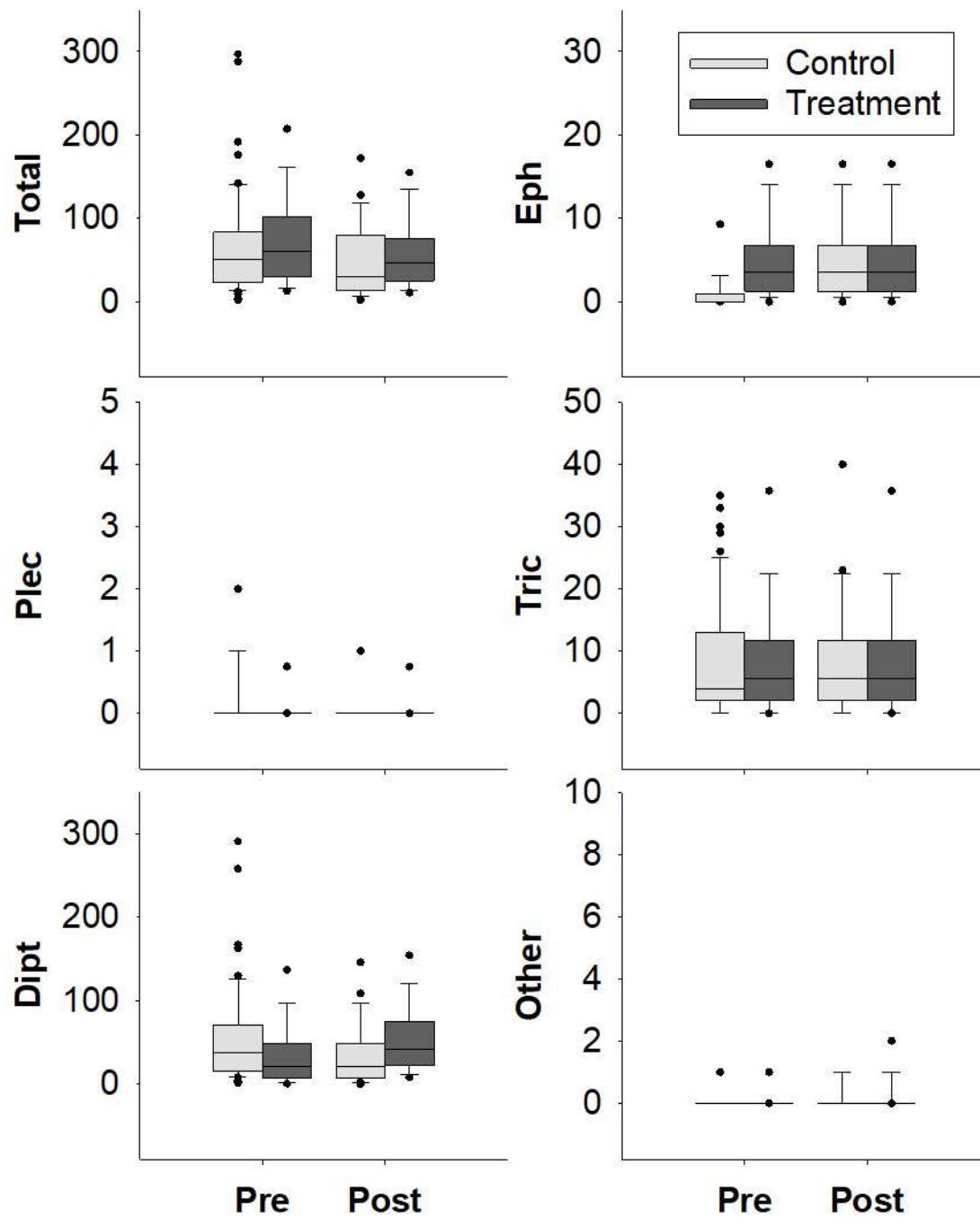


Figure 16. Adult export abundance (No. per emergence net) results pre- and post-restoration at the control site (AR5; light gray) and the NRDA treatment sites (AR4C, AR4E, AR4G, and AR4H combined; dark gray). See Table 1 for abbreviations.

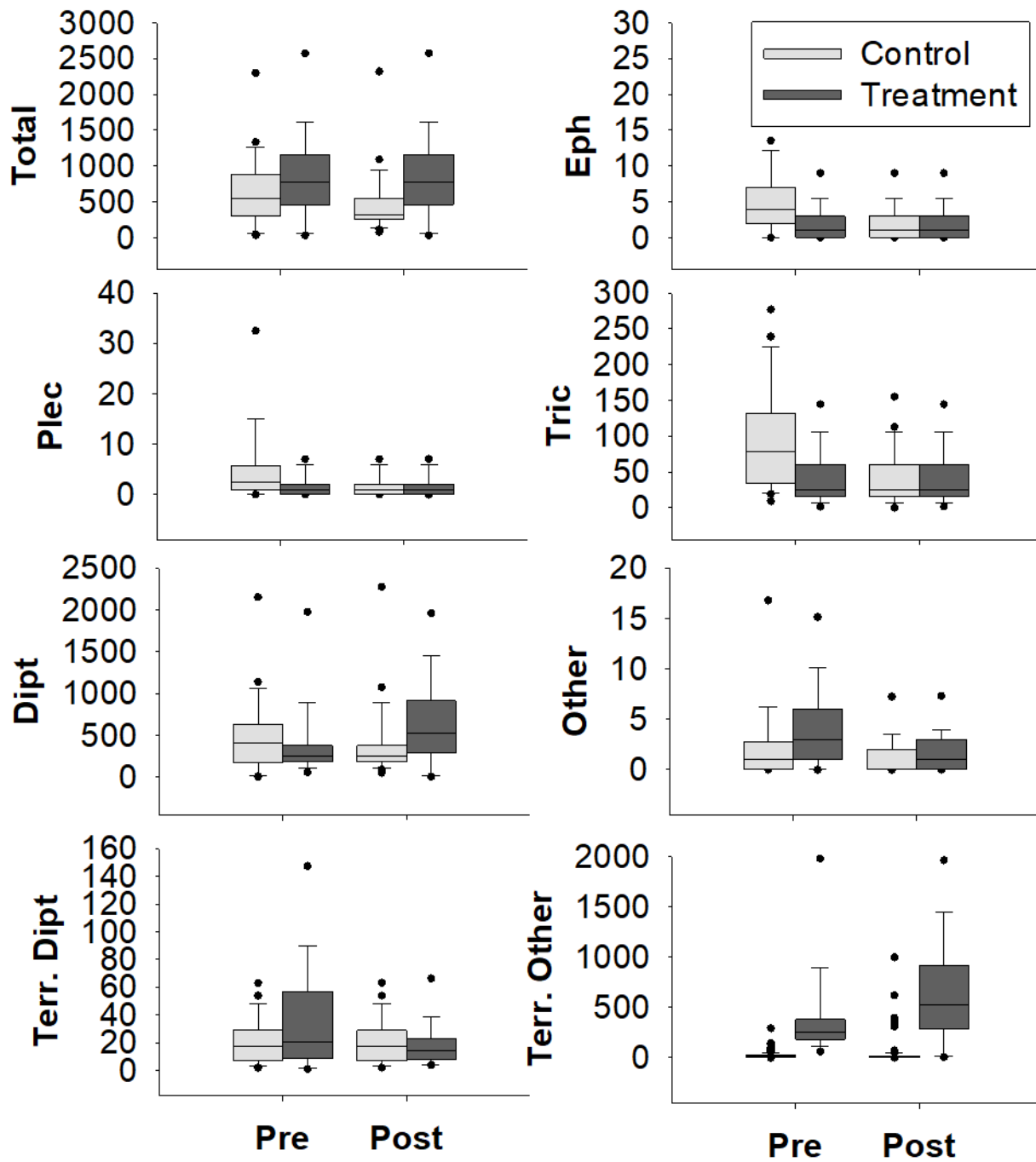


Figure 17. Adult import abundance (No. per pan trap) results pre- and post-restoration at the control site (AR5; light gray) and the NRDA treatment sites (AR4C, AR4E, AR4G, and AR4H combined; dark gray). See Table 1 for abbreviations.

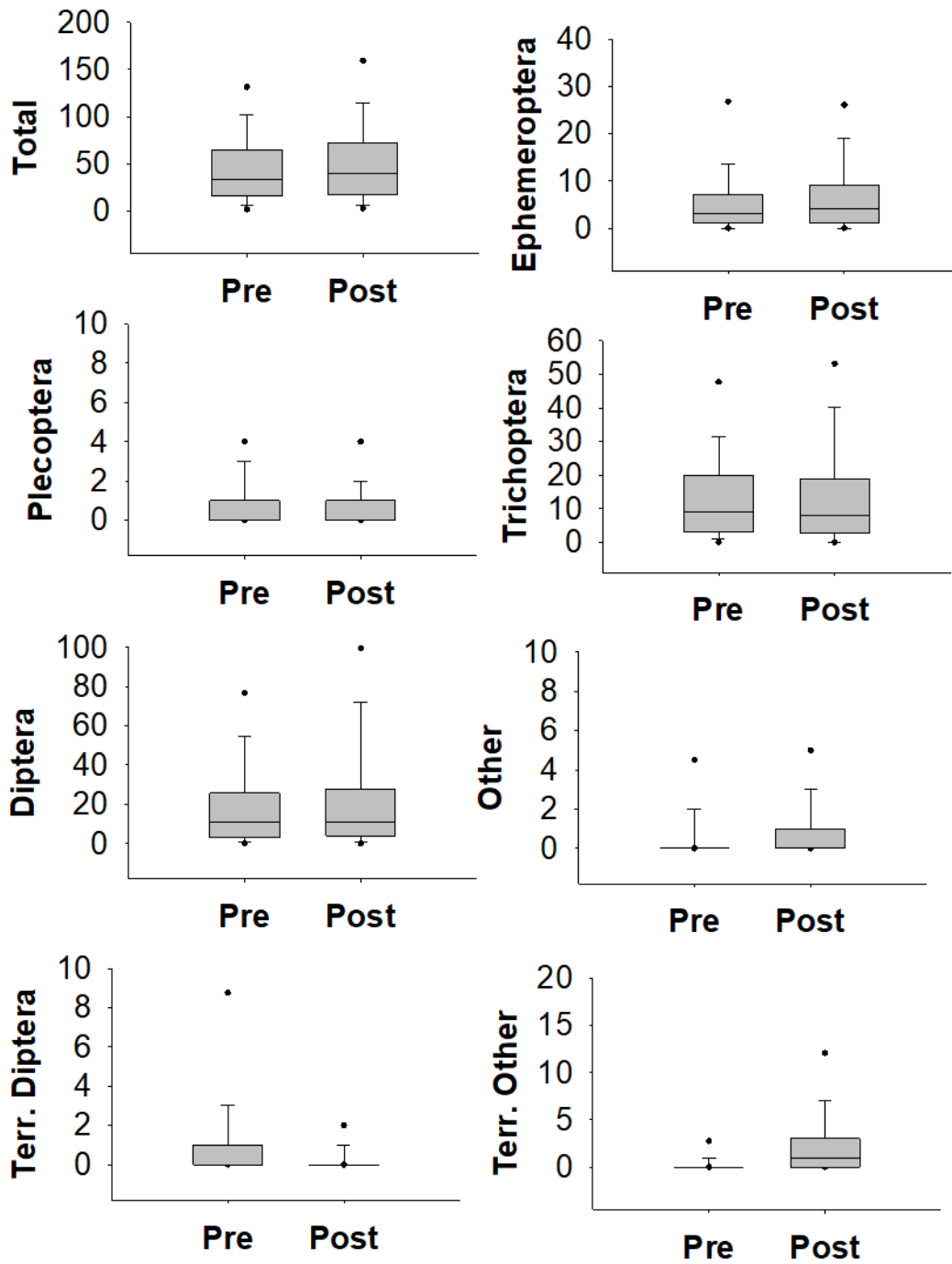


Figure 18. Fish diet abundance (No. per fish stomach) results pre- and post-restoration at the control site (AR5; light gray) and the NRDA treatment sites (AR4C, AR4E, AR4G, and AR4H combined; dark gray). See Table 1 for abbreviations.

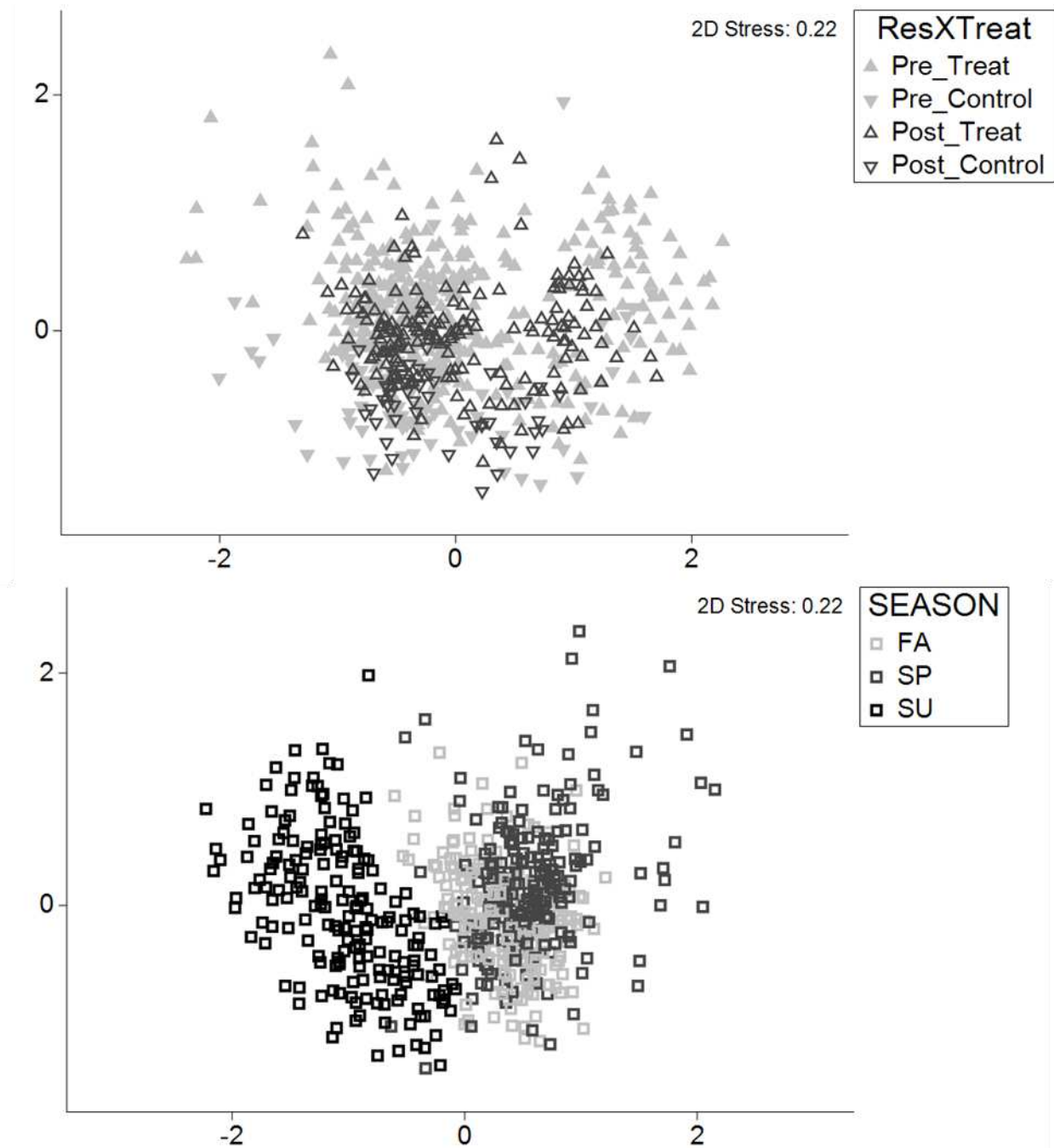


Figure 19. Non-metric Multi-Dimensional Scaling (NMDS) plot of the benthic macroinvertebrate community based on Bray-Curtis distance. The upper panel shows the grouping based on control vs treatment sites pre- and post-restoration. The lower panel shows the groupings based on season.

Literature Cited

- Bernhardt, ES, Palmer MA, Allan JD, Alexander G, et al. 2005. Synthesizing US river restoration efforts. *Science* 308:636-637.
- Burnham KP, and Anderson DR. 2002. Model selection and multimodel inference: a practical information-theoretic approach. 2nd ed. Springer, New York.
- Cadmus P, Pomeranz JPF, and Kraus JM. 2016. Low-cost floating emergence net and bottle trap: comparison of two designs. *Journal of Freshwater Ecology* 31: 1-6.
- Cadotte MW, Carscadden K, and Mirotchnick N. 2011. Beyond species: Functional diversity and the maintenance of ecological processes and services. *Journal of Applied Ecology* 48: 1079-1087.
- Clarke KR, and Gorley RN. 2006. PRIMER v6: User Manual/Tutorial (Plymouth Routines in Multivariate Ecological Research). PRIMER-E, Plymouth.
- Clements WH. 1994. Benthic community responses to heavy metals in the Upper Arkansas River Basin, Colorado. *J North Amer Benthol Soc* 13:30-44.
- Clements WH. 2004. Small-scale experiments support causal relationships between metal contamination and macroinvertebrate community responses. *Ecological Applications* 14: 954-967.
- Clements WH, and Rees DE. 1997. Effects of heavy metals on prey abundance, feeding habits, and metal uptake of brown trout (*Salmo trutta*) in the Arkansas River, Colorado. *Trans Amer Fish Soc* 126:774-785.
- Clements WH, Vieira NKM, and Church SE. 2010. Quantifying restoration success and recovery in a metal-polluted stream: a 17-year assessment of physicochemical and biological responses. *J Appl Ecol* 47:899-910.

- De Laender F, Verschuren D, Bindler R, Thas O, and Janssen CR. 2012. Biodiversity of freshwater diatom communities during 1000 years of metal mining, land use, and climate change in Central Sweden. *Environmental Science and Technology* 46: 9097-9105.
- Dornelas M, Gotelli NJ, McGill B, Shimadzu H, Moyes F, Sievers C, and Magurran AE. 2014. Assemblage time series reveal biodiversity change but not systematic loss. *Science* 344: 296-299.
- England J, and Wilkes MW. 2018. Does river restoration work? Taxonomic and functional trajectories at two restoration schemes. *Science of the Total Environment* 618: 961–970.
- Füreder L, Wallinger M, Burger R. 2005. Longitudinal and seasonal pattern of insect emergence in alpine streams. *Aquatic Ecology* 39: 67-78.
- Herbst DB., Medhurst RB, and Black NJ. 2018. Long-term effects and recovery of streams from acid mine drainage and evaluation of toxic metal threshold ranges for macroinvertebrate community reassembly. *Environmental Toxicology and Chemistry* 37: 2575-2592.
- Hobbs RJ, Higgs E, and Harris JA. 2009. Novel ecosystems: implications for conservation and restoration. *Trends in Ecology and Evolution* 24: 599-605.
- Hornberger MI, Luoma SN, Johnson ML, and Holyoak M. 2009. Influence of remediation in a mine-impacted river: Metal trends over large spatial and temporal scales. *Ecological Applications* 19: 1522-1535.
- Ivković M, Miliša M, Previšić A, Popijač A, and Mihaljević Z. 2013. Environmental control of emergence patterns: case study of changes in hourly and daily emergence of aquatic insects at constant and variable temperatures. *International Review of Hydrobiology* 98: 104-115.

- Kail J, Barbec K, Poppe M, and Januschke K. 2015. The effect of river restoration on fish, macroinvertebrates and aquatic macrophytes: a meta-analysis. *Ecological Indicators* 58: 311-321.
- Lepori F, Palm D, and Malmqvist B. 2005. Does restoration of structural heterogeneity in streams enhance fish and macroinvertebrate diversity? *Ecological Applications* 15: 2060–2071.
- Li K, Zhang Z, Yang H, Bian H, Jiang H, Sheng L, and He C. 2018. Effects of instream restoration measures on the physical habitats and benthic macroinvertebrates in an agricultural headwater stream. *Ecological Engineering* 122: 252-262.
- Luohi P, Mykra H, Paavola R, Huusko A, Vehanen T, Mäki-Petäys A, and Muotkai T. 2011. Twenty years of stream restoration in Finland: little response by benthic macroinvertebrate communities. *Ecological Applications* 21: 1950-1961.
- MacArthur RH., and MacArthur JW. 1961. On bird species diversity. *Ecology* 42:594-598.
- Menge, BA., Lubchenko J, and Ashkenas LR. 1985. Diversity, heterogeneity and consumer pressure in a tropical rocky intertidal community. *Oecologia* 65: 394-405.
- Miller SW., Budy P, and Schmidt JC. 2010. Quantifying Macroinvertebrate Responses to In-Stream Habitat Restoration: Applications of Meta-Analysis to River Restoration. *Restoration Ecology* 18: 8-19.
- Minshall GW. 2011. Structure and temporal variations of the benthic macroinvertebrate community inhabiting Mink Creek, Idaho, U.S.A., a 3rd Order Rocky Mountain stream. *Journal of Freshwater Ecology* 1: 13-26.
- Mori AS, Isbell F, and Seidl R. 2018. Beta-diversity, community assembly, and ecosystem functioning. *Trends in Ecology & Evolution*, 33, 549–564.

- Nuttle T, Logan M, Parise DJ, Foltz DA, Silvis JM, and Haibach MR. 2017. Restoration of macroinvertebrates, fish, and habitats in streams following mining subsidence: replicated analysis across 18 mitigation sites. *Restoration Ecology* 25: 820-831.
- Palmer MA, Ambrose RF, and Poff NL. 1997. Ecological theory and community restoration ecology. *Restoration Ecology* 5:291–300.
- R Development Core Team. 2010. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Raitif J, Plantegenest M, Agator O, Piscart C, and Roussel J-M. 2018. Seasonal and spatial variations of stream insect emergence in an intensive agricultural landscape. *Science of the Total Environment* 644: 594-601.
- Richer EE. 2019. Upper Arkansas River Habitat Restoration Project Annual Site Assessment. March 26, 2019. Aquatic Research Section Colorado Parks and Wildlife.
- Richer EE, Gates EA, Kondratieff MC, and Herdrich AT. 2019. Modelling changes in trout habitat following stream restoration. *River Research and Applications* 35: 680-691.
- Roni P, Bennett T, Morley S, Pess GR, Hanson K, Van Slyke D, and Olmstead P. 2006. Rehabilitation of bedrock stream channels: the effects of boulder placement on aquatic habitat and biota. *River Research and Applications* 22:967–980.
- Roni P, Hanson K, and Beechie T. 2008. Global review of the physical and biological effectiveness of stream habitat rehabilitation techniques. *North American Journal of Fisheries Management* 28: 856–890.
- Saunders WC, Fausch KD. 2007. Improved grazing management increases terrestrial invertebrate inputs that feed trout in Wyoming rangeland streams. *T Am Fish Soc* 136:1216-1230.

- Saunders WC, Fausch KD. 2012. Grazing management influences the subsidy of terrestrial prey to trout in central Rocky Mountain streams (USA). *Freshwater Biol* 57:1512-1529.
- Selego, SM, Rose CL, Merovich GT, Welsh SA, and Anderson JT. 2012. Community-level response of fishes and aquatic macroinvertebrates to stream restoration in a third-order tributary of the Potomac River, USA. *International Journal of Ecology* Article ID 753634.
- Smith RF, Neideigh EC, Rittle AM, Wallace JR. 2019. Assessing macroinvertebrate community response to restoration of Big Spring Run: expanded analysis of before-after-control-impact sampling designs. *River Research and Applications*.
<https://doi.org/10.1002/rra.3556>.
- Stewart-Oaten A, Bence JR, and Osenberg CW. 1992. Assessing effects of unreplicated perturbations: no simple solutions. *Ecology* 73:1396-1404.
- Stewart-Oaten A, Murdoch WW, and Parker KR. 1986. Environmental impact assessment: "pseudoreplication" in time? *Ecology* 67:929-940.
- US EPA (U.S. Environmental Protection Agency). 2002. National recommended water quality criteria, U.S. Environmental Protection Agency, Office of Water. EPA-822-R-02-047, Washington, DC.
- Vannote RL, Minshall GW, Cummins KW, Sedell JR, and Cushing CE. 1980. The River Continuum Concept. *Can J Fish Aquat Sci* 37:130–137.
- Waters DS, Kwak TJ, Arnott JB, Pine WE. 2004. Evaluation of stomach tubes and gastric lavage for sampling diets from blue catfish and flathead catfish. *N Am J Fish Manage* 24:258-261.

- Wolff BA, Clements WH, and Duggan SB. 2019. Resilience and regime shifts: do novel benthic communities impede ecological recovery in a historically metal-contaminated stream? *Journal of Applied Ecology*. DOI: 10.1111/1365-2664.13503.
- Wolff BA, Clements WH, and Hall EK. 2020. Metals alter membership but not diversity of a headwater stream microbiome. *In preparation*.
- Xie L, Funk DH, and Buchwalter DB. 2009. Trophic transfer of Cd from natural periphyton to the grazing mayfly *Centroptilum triangulifer* in a life cycle test. *Environmental Pollution* 158: 272-277.

APPENDIX

Chapter 1 Supplemental Tables and Figures

Table S 1. Routine physicochemical characteristics measured in stream mesocosms during the 4 metals experiments. Values in parentheses are standard deviation.

Site	Treatment	pH	Conductivity ($\mu\text{S}/\text{cm}$)	Temperature ($^{\circ}\text{C}$)	Dissolved oxygen (mg/L)
Upstream	Control	7.27 (0.27)	75.0 (2.9)	11.43 (1.06)	8.17 (1.02)
	Metals	7.12 (0.26)	77.4 (2.6)	11.41 (1.04)	7.96 (0.96)
Downstream	Control	7.26 (0.28)	74.5 (3.3)	11.42 (1.04)	8.19 (1.03)
	Metals	7.12 (0.26)	77.3 (2.5)	11.44 (1.07)	8.02 (0.99)

Table S 2. Mean (+ s.d.) concentrations of Cu and Zn measured in treated stream mesocosms during the 4 metals experiments. CCU = cumulative criterion unit, defined as the ratio of the measured metal concentration to the U.S. EPA hardness-adjusted chronic criterion value and summed for each metal (Clements et al. 2000). Values in parentheses are standard deviation.

Site	Cu ($\mu\text{g/L}$)	Zn ($\mu\text{g/L}$)	CCU
Upstream (n = 60)	20.3 (5.6)	574.8 (104.7)	18.0 (3.3)
Downstream (n = 59)	19.8 (5.3)	566.5 (114.6)	17.7 (3.6)

Table S 3. Results of 2-way ANOVA showing differences in survival between sites (upstream versus downstream), diesel treatments (control versus diesel exposure) and the site x treatment interaction term in stream mesocosm experiments.

Order	Source	F-value	p-value
Ephemeroptera	Site	58.9	< 0.0001
	Treatment	320.9	< 0.0001
	Interaction	8.1	0.0149
Plecoptera	Site	53.8	< 0.0001
	Treatment	63.9	< 0.0001
	Interaction	1.7	0.223
Trichoptera	Site	119.1	< 0.0001
	Treatment	165.2	< 0.0001
	Interaction	3.1	0.1048
Diptera	Site	13.9	0.0029
	Treatment	136.5	< 0.0001
	Interaction	1.1	0.3075

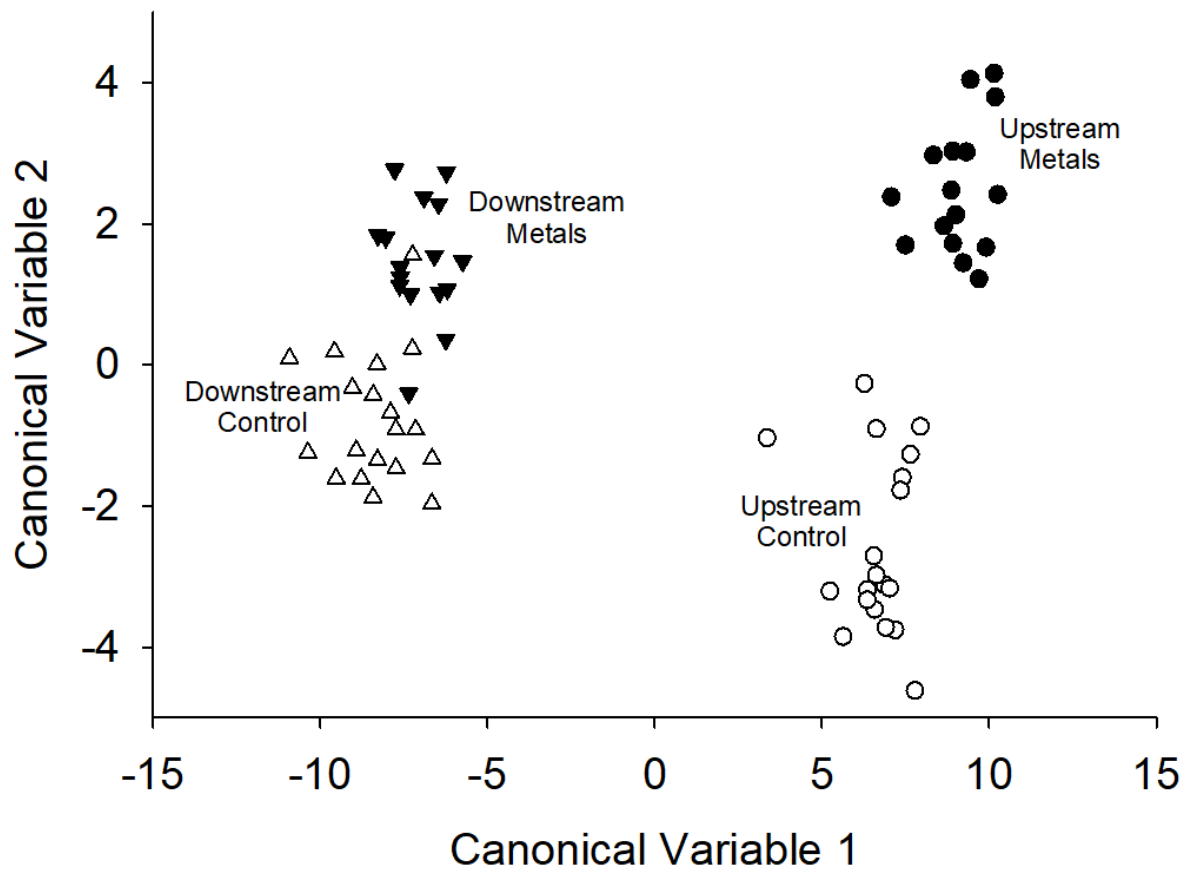


Figure S 1. Canonical Discriminant Function Analysis of community composition from all mesocosm samples.