Environmental Monitoring and Assessment

Comparison of Benthic Macroinvertebrate Biomonitoring Approaches for Evaluating Community Response Across a Stressor Gradient in Headwater Streams

--Manuscript Draft--

Manuscript Number:	
Full Title:	Comparison of Benthic Macroinvertebrate Biomonitoring Approaches for Evaluating Community Response Across a Stressor Gradient in Headwater Streams
Article Type:	Original Research
Keywords:	salinization; rapid bioassessment; full enumeration; coal mining
Corresponding Author:	Thomas Cianciolo, M.S. Trout Unlimited Klamath Falls, OR UNITED STATES
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Trout Unlimited
Corresponding Author's Secondary Institution:	
First Author:	Rachel A. Pence, M.S.
First Author Secondary Information:	
Order of Authors:	Rachel A. Pence, M.S.
	Thomas Cianciolo, M.S.
	Damion R. Drover
	Daniel L. McLaughlin
	David J. Soucek
	Anthony J. Timpano
	Carl E. Zipper
	Stephen H. Schoenholtz
Order of Authors Secondary Information:	
Funding Information:	Powell River Project — Dr. Stephen H. Schoenholtz
Abstract:	Benthic macroinvertebrate community assessments are commonly used to characterize aquatic systems and increasingly for identifying their impairment caused by myriad stressors. Yet, sampling and enumeration methods vary, and research is needed to compare their ability to detect macroinvertebrate community response to specific water quality variables. A common assessment approach, rapid bioassessment uses sub-sampling procedures to identify a fixed number of individuals regardless of total sample abundance. In contrast, full-enumeration assessments typically allow for expanded community characterization resulting from higher numbers of identified organisms. Here, we compared these two sampling and enumeration methods and their ability to detect benthic macroinvertebrate response to freshwater salinization, a common stressor of streams worldwide. We applied both methods in headwater streams along a salinity gradient within the coal-mining region of central Appalachia USA. Full-enumeration samples had approximately seven times higher abundance and estimates of taxonomic richness were 26-92% higher than rapid-bioassessment samples. Metrics of community composition, trophic function, dominance, tolerance, and habit also differed between the methods. Full-enumeration was generally a stronger predictor of benthic community responses to salinization and was more sensitive than rapid-bioassessment for detecting community changes in salinized streams. However, the more cost-effective and widely-employed rapid-

	bioassessment method was capable of characterizing severe community alterations associated with the mining-induced salinity gradient. These findings can help inform decisions regarding such tradeoffs for assessments of freshwater salinization in temperate forested headwater streams and highlight the need for similar research of sampling and enumeration methodology in other aquatic systems and for other stressors.
Suggested Reviewers:	David Arscott Stroud Water Research Center darscott@stroudcenter.org
	Margaret Passmore EPA: Environmental Protection Agency passmore.margaret@epa.gov
	Hanneke Keizer-Vlek Wageningen University Research Centre: Wageningen University & Research hanneke.keizer-vlek@wur.nl
	Nick Everall Aquascience Consultancy Ltd maquaconsult@aol.com

Introduction

Characterization of benthic macroinvertebrate communities is a fundamental feature in many bioassessments of aquatic systems. In freshwater streams impacted by stressors such as salinity, nutrients, sediments, or toxic elements, macroinvertebrate community structure may be altered in comparison to a relatively undisturbed reference condition; biomonitoring and bioassessment procedures are commonly employed to detect such alterations. Indeed, macroinvertebrate biomonitoring and bioassessments are used by most states in the USA to aid enforcement of the federal Clean Water Act and its mandate to restore and maintain "chemical, physical and biological integrity of the Nation's waters" (Governor et al., 2017). Macroinvertebrate biomonitoring procedures and data are also used in other nations and world regions to aid water management (Morse et al., 2007; Birk et al., 2012; Nichols et al., 2017). However, a wide variety of biomonitoring methods are available for use (Davies, 2001). Among major differences, in addition to sampling devices, are numbers of individuals collected and identified, and whether complete samples are characterized or if subsampling is employed to reduce sample size and, hence, biomonitoring costs.

A common method for biomonitoring in the USA is the rapid bioassessment protocol (RBP), which employs a sampling method with fixed-count subsampling (Barbour et al., 1999). Such sampling commonly occurs over an approximated area, such as riffle-run habitats in high-gradient streams using a kick-net or D-frame net. Multiple samples are then composited and subsampled to obtain a fixed count of organisms (e.g., 200 ± 10%) for identification regardless of actual macroinvertebrate sample size. Macroinvertebrate community metrics such as taxonomic

richness, relative abundance, and diversity are derived from the fixed-count subsample, potentially overlooking presence of rare taxa not included in the subsample (Courtemanch, 1996). Consequently, the fixed number of organisms in the subsample may limit the utility of such sampling for accurate and complete characterization of macroinvertebrate community structure.

In contrast, full-enumeration sampling methods typically collect all organisms from a defined area of the stream bottom (e.g., using a Surber sampler), typically in greater numbers than the fixed-count subsamples employed by conventional RBP methods, thereby enumerating and identifying all organisms collected. Such full-enumeration samples are more time-consuming and expensive to process; however, they allow for more accurate estimates of taxonomic richness, abundance, and diversity, and potentially may better detect responses of macroinvertebrate assemblages to disturbances relative to RBP methods (Courtemanch, 1996; Vinson & Hawkins, 1996). Although some research has compared biomonitoring methods for differences in community characterization (e.g., Buss & Borges, 2008; Tubić et al., 2017), there has been limited study on how methods may differ in detecting community change along specific stressor gradients.

Anthropogenic salinization of freshwater ecosystems is a growing and global concern and one often assessed through benthic macroinvertebrate-based bioassessments. Salinization results from many land uses, including agriculture, municipal and industrial discharges, and road de-icing as well as mining and can affect aquatic ecosystem structure and function (Schäfer et al., 2012; Canedo-Arguelles, 2013; Kaushal et al., 2018). Here we focus on surface coal mining in

central Appalachia USA, where accelerated weathering of deposited waste rock material increases concentrations of dissolved ions and associated salinity in receiving headwater streams (Bryant et al., 2002; Hartman et al., 2005; Pond et al., 2008; Lindberg et al., 2011). Elevated mining-origin salinity in this region is associated with alterations of aquatic life, particularly shifts in benthic macroinvertebrate communities to more salt-tolerant taxa (e.g., Pond et al., 2008; Boehme et al., 2016; Timpano et al., 2018a). Macroinvertebrate taxa respond differentially to mining-origin water influences, including increased salinity, with individuals in the Ephemeroptera order and scraper functional feeding group particularly sensitive, whereas others such as the ubiquitous headwater stream stonefly genera *Amphinemura* and *Leuctra*, are often found in higher relative abundances in salinized streams (Pond, 2010; Pond et al., 2014; Boehme et al., 2016; Timpano et al., 2018a; Drover et al., 2020).

Studies characterizing these salinity-induced responses in central Appalachia have relied heavily on RBP methodology with fixed-count subsampling. Critically, the same approach was also used by USEPA (2011) to establish an aquatic life benchmark for salinity (measured as specific conductance; SC: electrical conductivity at 25˚C) in Appalachian streams based on extirpation of sensitive benthic macroinvertebrate taxa. In contrast, full-enumeration sampling has rarely been used to characterize macroinvertebrate communities and their response to elevated salinity in central Appalachian headwater streams (Hartman et al., 2005; Drover et al., 2020). Indeed, full-enumeration sampling has been recommended for obtaining samples representative of riffle-dwelling macroinvertebrate communities, which are common in these headwater systems (Davies, 2001).

Given that sampling methodology has potential to influence assessment of macroinvertebrate metrics and thus bioassessment outcome, a comparison of fixed-count RBP with full-enumeration sampling techniques in salinized central Appalachian streams is warranted yet, to our knowledge, has not been performed. To fill this research gap, we applied and compared these methods across 15 headwater streams representing a salinity gradient from reference to highly salinized, mining-influenced streams. Our objectives were to compare measures of benthic macroinvertebrate community structure and their response to elevated salinity based on metrics generated by the two sampling methods. We predicted that the full-enumeration sampling method would capture greater taxonomic richness than the fixed-count RBP method and, in doing so, would provide structural metric values that are more sensitive to elevated salinity.

Methods

Site selection

We selected 15 1st-order streams in central Appalachia to assess effects of salinization from surface coal mining on benthic macroinvertebrate community structure characteristic of riffle habitat. Four of these study streams were representative of reference conditions and were located in minimally disturbed, forested catchments not influenced by surface coal mining (Table 1). The remaining 11 streams were influenced by historical (i.e., before 2011) surface coal mining with elevated concentrations of dissolved major ions relative to reference streams (Online Resource Table S1). Surface coal mining releases elevated concentrations of sulfate (SO_4^{2-}), calcium (Ca^{2+}), magnesium (Mg^{2+}), and bicarbonate (HCO_3^-), and trace elements such as manganese

(Mn) and selenium (Se) to drainage waters (Pond et al. 2014; Timpano et al. 2018b). Non-mined areas of all mining-influenced watersheds were predominantly forested with intact forest in the riparian zone. All study streams were located within the central Appalachian coalfield of Virginia and West Virginia (Figure 1; Table 1) and within US EPA Ecoregion 69d (Level IV, Omernik, 1987).

Study streams were chosen using a rigorous selection process (see methods in Timpano et al., 2015) endeavoring to isolate salinity as the stressor to aquatic life whereby non-salinity stressors (e.g., altered riparian buffer, extreme pH, excessive sedimentation) were minimized. At each study stream, we collected both full-enumeration and fixed-count RBP benthic macroinvertebrate community samples during Spring 2014 and Fall 2017. However, one study stream was removed for Fall 2017 sampling because of its recent streamside development for natural gas extraction, which removed riparian vegetation and increased sedimentation.

Table 1: Stream attributes for selected study streams in the central Appalachian coalfield.

Stream	Stream ID	Stream Type	County	State	Mined Area in Watershed[a] (%)	Annual Mean SC[b] Spring 2014 (µS cm^{-1})	Annual Mean SC[c] Fall 2017 (µS cm^{-1})
Copperhead Branch	COP	Reference	Buchanan	VA	0	128	115
Crooked Branch	CRO	Reference	Dickenson	VA	0	65	57
Eastland Creek	EAS	Reference	Wise	VA	0	26	25
Hurricane Fork	HCN	Reference	McDowell	WV	0	68	86
Fryingpan Creek	FRY	Mining-influenced	Dickenson	VA	6.0	361	398
Fryingpan Creek Right Fork[d]	RFF	Mining-influenced	Dickenson	VA	20.8	430	-
Grape Branch	GRA	Mining-influenced	Buchanan	VA	3.2	239	238
Kelly Branch	KEL	Mining-influenced	Wise	VA	60.0	759	737
Kelly Branch Unnamed Tributary	KUT	Mining-influenced	Wise	VA	38.8	1,064	1,013
Left Fork/Laurel Fork/Coal Fork	LLC	Mining-influenced	Kanawha	WV	28.2	1,212	1,132
Mill Branch West Fork	MIL	Mining-influenced	Wise	VA	51.2	599	540
Powell River	POW	Mining-influenced	Wise	VA	58.6	773	814
Rickey Branch	RIC	Mining-influenced	Wise	VA	32.1	1,484	1,404
Rockhouse Creek	ROC	Mining-influenced	Raleigh	WV	34.2	694	667
Spruce Pine Creek	SPC	Mining-influenced	Buchanan	VA	2.2	352	371

[a]Mined area of watersheds, 1985–2015, calculated using yearly surface mining layers produced by Pericak et al. (2018).

[b]Annual mean specific conductance (SC) calculated using data collected at 30-minute intervals for 12 months prior to Spring 2014 macroinvertebrate sampling.

[c]Annual mean specific conductance (SC) calculated using data collected at 30-minute intervals for 12 months prior to Fall 2017 macroinvertebrate sampling.

[d]Excluded from fall 2017 analyses.

Fig. 1 Location of headwater streams selected for study in the central Appalachian coalfield

Field collection and laboratory sample processing

Specific conductance

At each study stream, freshwater conductivity loggers (HOBO, Onset Computer Corp., Bourne, MA, USA) measured and recorded SC and temperature at 30-minute intervals from Fall 2011through Fall 2017 in each stream. At the time of benthic macroinvertebrate sampling, water temperature, dissolved oxygen, SC, and pH were measured with a handheld multi-probe meter (YSI Professional Plus; YSI Inc., Yellow Springs, OH, USA) to provide synoptic characterization of water-quality parameters. Synoptic measurements of SC were used to calibrate continuous conductivity loggers and adjust for SC data drift.

Benthic macroinvertebrates

Fixed-count RBP benthic macroinvertebrate samples were collected following US EPA RBP for the single-habitat riffle-run method in high-gradient streams (Barbour et al., 1999) from 11-13 April 2014 (hereafter "Spring") and from 4-31 October 2017 (hereafter "Fall"). A 30-cm D-frame kicknet with 500-µm mesh was used to collect six 1 x 0.3-m samples from randomly selected riffle habitats within each stream (n = 15 streams in Spring; n = 14 streams in Fall). Sampling of benthic macroinvertebrates at each stream occurred along a 100-m reach approximately centered on the location of the continuous SC logger. The six kick-net samples were combined into a single composite sample (approximately 2 m^2) at each stream. Samples

were preserved in 95% ethanol immediately upon collection in the field and returned to the laboratory for sorting, identification, and enumeration.

Samples collected using the RBP method were sub-sampled randomly in the laboratory using a standardized gridded tray to obtain a 200 (±10%) organism count following Virginia Department of Environmental Quality methods (VDEQ 2008) adapted from US EPA RBP (Barbour et al., 1999). Individuals in each sample were counted and identified to genus-level (except for individuals in family Chironomidae and sub-class Oligochaeta, which were identified at those levels) using standard keys (Wiggins, 1996; Merritt et al., 2008).

Full-enumeration benthic macroinvertebrates samples were collected using a 0.0929-m^2 Surber stream-bottom sampler (Wildco; Yulee, FL, USA) with a 500-µm mesh, from 25-29 April 2014 and 4-31 October 2017. During Fall 2017, RBP and full-enumeration samples were collected on the same date for each stream. During Spring 2014, sampling using RBP and full-enumeration methods were not always collected on the same date. At each stream, three separate samples for full-enumeration were collected along the same 100-m reach from which RBP samples were collected. Care was taken to avoid collection of samples in any area previously disturbed by RBP sampling. Each full-enumeration sample was stored separately in 95% ethanol upon collection and returned to the laboratory for sorting, identification, and enumeration.

Laboratory analysis involved sorting and identification of all individuals in each full-enumeration sample in an identical manner to that applied to RBP samples, but without subsampling. For each stream, the three full-enumeration samples collected during a particular

sampling event were initially sorted, identified, and counted separately, but the data were combined prior to metric calculation to produce a single composite sample for comparison to the composited RBP sample collected at the same stream.

Data analysis

Specific conductance

Continuous SC data were summarized to obtain an annual SC mean value for the year preceding each Spring and Fall sample to provide a single variable for use in correlation and regression analyses that was more representative of stream salinity than synoptic samples taken at one time. However, data gaps were occasionally present in the continuous SC time series of some study streams as a result of datalogger malfunction and burial following storm events. For data gaps of more than 15 consecutive days, substitutions were made using historical data from the same time period in the previous year. This procedure was employed because central Appalachian headwater streams exhibit a strong seasonal SC pattern (Timpano et al., 2018b). Maximum extents of continuous-SC data gaps ranged from 0 to 46 days among the 15 streams.

Benthic macroinvertebrates

Thirty-seven metrics of benthic macroinvertebrate community structure were selected for analysis. These metrics are indicators of taxonomic richness, composition, functional feeding groups, dominance, diversity, tolerance, and habit of the benthic macroinvertebrate community

(Table 2). Metrics were selected to include those used for quantification of benthic community structure by Barbour et al. (1999) and by previous studies conducted in headwater streams of the central Appalachian coalfield (Boehme et al., 2016; Timpano et al., 2018a; Cianciolo et al., 2020; Drover et al., 2020). Pielou's evenness (*J'*) was added as a measure of the degree of evenness in distribution of individuals among taxa of the benthic community (Pielou, 1966). Functional feeding groups, tolerance values, and habits were ascribed to each genus using references by Merritt et al. (2008) and Barbour et al. (1999); taxa with pollution-tolerance values > 6 were designated as tolerant and taxa with pollution-tolerant values <4 were considered as intolerant (after Pond et al., 2013). All 37 metrics were calculated for both fixed-count RBP and full-enumeration composite samples using R software (RStudio Team 2018).

196 Table 2: Benthic macroinvertebrate metrics by category, with abbreviations and expected

197 response to increasing environmental disturbance.

Category	Metric	*Abbrev.*	Expected Response to Perturbation[a]
Taxonomic Richness	Total Taxa	*rTotal*	Decrease
	Ephemeroptera, Plecoptera, and Trichoptera (EPT) Taxa	*rEPT*	Decrease
	EPT Taxa less Hydropsychidae	*rEPT-H*	Decrease
	Ephemeroptera Taxa	*rE*	Decrease
	Ephemeroptera Taxa less Baetidae	*rE-B*	Decrease
	Plecoptera Taxa	*rP*	Decrease
	Trichoptera Taxa	*rT*	Decrease
	Trichoptera Taxa less Hydropsychidae	*rT-H*	Decrease
	Diptera Taxa	*rDip*	Decrease
Composition	% EPT	*pEPT*	Decrease
	% EPT less Hydropsychidae	*pEPT-H*	Decrease
	% Ephemeroptera	*pE*	Decrease
	% Ephemeroptera less Baetidae	*pE-B*	Decrease
	% Plecoptera	*pP*	Decrease
	% Trichoptera	*pT*	Decrease
	% Trichoptera less Hydropsychidae	*pT-H*	Decrease
	% Plecoptera plus Trichoptera Taxa less Hydropsychidae	*pPT-H*	Decrease
	% Diptera	*pDip*	Increase
	% Chironomidae	*pChi*	Increase
	% Oligochaeta	*pOligo*	Increase
Functional Feeding Groups	% Collector-Gatherers	*pCG*	Decrease
	% Collector-Filterers	*pCF*	Variable
	% Predators	*pPR*	Variable
	% Scrapers	*pSC*	Decrease
	% Shredders	*pSH*	Decrease
	Scraper Taxa Richness	*rSC*	Decrease
Dominance	% 1 Dominant Taxon	*p1dom*	Increase
	% 2 Dominant Taxa	*p2dom*	Increase
	% 5 Dominant Taxa	*p5dom*	Increase
	Pielou's Evenness	*even*	Decrease
Diversity	Shannon Diversity	*shan*	Decrease
	Simpson Diversity	*simp*	Decrease
Tolerance	Intolerant Taxa Richness	*rINT*	Decrease
	% Tolerant	*pTOL*	Increase
	Hilsenhoff Biotic Index	*HBI*	Increase
Habit	% Clingers	*pCling*	Variable
	Clinger Taxa Richness	*rCling*	Variable

198 [a]Source: Barbour et al. (1999).

199

200 Differences in community metrics between the two sampling methods for each stream were

201 assessed with a paired-sample Wilcoxon signed-rank tests using the *exactRankTests* package

(Hothorn & Hornik, 2018; R Core Team, 2018) in R. In each stream for each sampling period (i.e., Spring and Fall), mean percent differences for all 37 metrics were calculated for each RBP vs. full-enumeration comparison (Equation 1).

Equation 1

$$Percent\ Difference = \left[\frac{Metric\ Value_{quantitative full-enumeration} - Metric\ Value_{semi-quantitative RBP}}{(Metric\ Value_{quantitative full-enumeration} + Metric\ Value_{semi-quantitative RBP})/2}\right] * 100$$

Spearman rank correlations were used to test for association between annual mean SC and 1) RBP sample metrics, 2) full-enumeration sample metrics, and 3) differences in sample metrics (full-enumeration minus RBP). Correlation results were evaluated to determine if the two methods similarly characterize both the direction and magnitude of community response to elevated SC. Mean values of preceding-year SC for each stream were used for these calculations. In order to compare the two sampling methods for responsiveness to SC directly, the seven metrics with differences showing consistent and significant relationships with SC during both seasons were analyzed using linear regression vs. SC.

Results

For the 15 streams sampled in Spring, a total of 83 taxa from full-enumeration samples (26,208 individuals) and 64 taxa from RBP samples (3,134 individuals) were identified; 30 taxa were unique to full-enumeration samples, whereas 11 taxa were unique to RBP samples (Online

225 Resource Table S2). For the 14 streams sampled in Fall, 87 taxa from full-enumeration samples
226 (14,796 individuals) and 71 taxa from RBP samples (2,836 individuals) were identified; 23 taxa
227 were unique to full-enumeration samples, whereas 9 taxa were unique to RBP samples (Online
228 Resource Table S2). Mean (+/- SE) organism abundances in full-enumeration samples were
229 1,747 ± 217 and 1,057 ± 160 individuals in Spring and Fall, respectively, whereas, by design,
230 RBP samples were sub-sampled by collecting 200 (± 10%) organisms (actual means of 209 ± 2
231 and 203 ± 5 individuals in Spring and Fall, respectively).
232
233 Mean of 12-month antecedent SC values for reference streams ranged from 26 to 128 µS/cm for
234 Spring, whereas test-stream mean 12-month antecedent SC ranged from 239 to 1,848 µS/cm
235 (Online Resource Table S1). For Fall, mean 12-month antecedent SC ranged from 25 to 115
236 µS/cm for reference streams and from 238 to 1,404 µS/cm for test streams (Online Resource
237 Table S1).
238
239 Comparison of sampling methods for community characterization
240
241 Taxonomic Richness
242
243 Full-enumeration sampling produced higher estimates than RBP sampling for all nine taxonomic
244 richness metrics for both Spring and Fall. All nine measures of taxonomic richness derived from
245 full-enumeration samples differed significantly from those derived from RBP samples in Spring
246 and all metrics except Ephemeroptera richness less Baetidae (*rE-B*) differed significantly in Fall
247 (Figure 2; Tables 3 and 4). When averaged across streams, differences in taxonomic richness

248 metrics ranged from 35.7 to 92.1% higher in Spring full-enumeration samples and from 25.9 to
249 64.6% higher in Fall full-enumeration samples (Figure 2).
250
251 Community Composition
252
253 Significant differences were detected between the two sampling techniques for several
254 community-composition metrics. Percent EPT (*pEPT*), percent EPT less Hydropsychidae (pEPT-
255 *H*), percent Plecoptera (*pP*), percent Plecoptera plus Trichoptera taxa less Hydropsychidae (*pPT-*
256 *H*), and percent Oligochaeta (*pOligo*) were significantly higher in full-enumeration samples
257 collected in Spring (Table 3). Significant differences between sample methods were also
258 detected for percent Diptera (*pDip*) for Spring and percent Chironomidae (*pChi*) for Fall, where
259 RBP metric values exceeded full-enumeration values by 23.5% (±14.1) and 26.2% (±10.0),
260 respectively (Figure 2; Tables 3 and 4).
261
262 Functional Feeding Groups, Dominance, Diversity, Tolerance, and Habit
263
264 All functional feeding group metrics, except for percent predators (*pPR*), differed significantly
265 between sample methods for at least one of the sample periods (Figure 2; Tables 3 and 4). Rapid
266 bioassessment samples produced significantly higher measures for percent collector-filterers
267 (*pCF*) (average difference 46.6% ±14.8) and percent scraper (*pSC)* (average difference 55.7%
268 ±18.6) in Spring, and for percent collector-gatherers (*pCG)* in Fall (average difference 30.0%
269 ±12.4). Percent shredders (*pSH*) was significantly higher in full-enumeration samples collected
270 in Spring, and the average difference between sample methods was 22.1% (±7.3). For both

sampling periods, scraper taxa richness (*rSC*) was significantly higher in full-enumeration samples than corresponding RBP samples (average difference of 32.9% ±11.2 in Spring and 29.2% ±10.1 in Fall) (Figure 2; Tables 3 and 4).

Pileou's evenness (*even*), intolerant taxa richness (*rINT*), and clinger taxa richness (*rCling*) were the only metrics from the Dominance, Diversity, Tolerance, and Habit categories to display significant differences between the two sampling techniques (Figure 2; Tables 3 and 4). For Pielou's evenness (*even*), a small but statistically significant difference between sampling methods was detected for Spring, where RBP measures exceeded corresponding full-enumeration measures by 19.4% (±4.8) (Figure 2; Table 3). Intolerant taxa richness (*rINT*) for full-enumeration samples was 34.9% (±4.5) and 43.9% (±6.2) higher than for RBP samples collected in Spring and Fall, respectively. Clinger taxa richness (*rCling*) was also significantly higher in full-enumeration samples collected during both sampling periods (Figure 2; Tables 3 and 4). A total of 73 intolerant taxa were observed throughout the course of this study, and more than half of these taxa were also classified as clinger taxa (data not shown).

Fig. 2 Mean differences (± SE) of macroinvertebrate community metrics between sample methodologies (Equation 1). Differences are expressed as percentages (%) and grouped by metric category for Spring and Fall sampling periods ("Dom" = Dominance; "Tol" = Tolerance; "Hab" = Habit). Positive differences reflect higher values in full-enumeration samples. Metric abbreviations are defined in Tables 3 and 4. Only metrics with significant differences (paired-sample Wilcoxon signed-rank test <0.05) between sample types are displayed

Comparison of methods for detecting community response to SC

Many benthic community metrics were correlated with SC gradient, and differences in community metrics between sample methods were also correlated with SC (Tables 3 and 4). The largest differences between full-enumeration and RBP richness metrics were typically observed at reference streams because of higher levels of taxonomic richness present at those streams. Differences between full-enumeration and RBP richness metrics were reduced in high-SC streams, a function of taxa extirpation at elevated SC (Figure 3; Online Resource Table S3).

Fig. 3 Regressions of three taxa richness measures (total taxa, EPT taxa, and Ephemeroptera taxa less Baetidae) in response to 12-month antecedent specific conductance (SC) as derived from Rapid Bioassessment Protocol (RBP) and full-enumeration sampling during Spring (left) and Fall (right). All regressions are significant at $p<0.05$. See Online Resource, Table S3, for further detail

Taxonomic richness

Most taxonomic richness metrics (*rTotal*, *rEPT*, *rEPT-H*, *rE*, *rE-B*, and *rP*) were negatively correlated with SC for both Spring and Fall for both sampling methods (Tables 3 and 4). Among the pollution sensitive EPT orders, Ephemeroptera are the most sensitive to elevated SC in central Appalachian coalfield streams, but members of the family Baetidae appear to be more tolerant to this type of pollution (Pond 2010; Timpano et al. 2018a). Consequently, exclusion of salt tolerant Baetidae (Ephemeroptera taxa richness less Baetidae, *rE-B*) generally resulted in

317 stronger responses than the aggregate Ephemeroptera taxa richness metric (*rE*) for both sampling
318 periods. Plecoptera are also recognized as a pollution-sensitive order (Barbour et al. 1999), but
319 relationships with SC were not as strong compared to relationships with *rE* and *rE-B* (Tables 3
320 and 4). Negative relationships between SC and Trichoptera taxa richness (*rT*), as well as between
321 SC and Trichoptera taxa less Hydropsychidae (*rT-H*), were also observed but only for full-
322 enumeration samples collected during Spring.
323
324 Differences in taxonomic richness metrics (*rTotal*, *rEPT*, *rEPT-H*, *rE*, *rE-B*, and *rT*) between the
325 two sample methods were negatively correlated with SC for both Spring and Fall, and
326 relationships were generally stronger in Spring (Tables 3 and 4). Differences in EPT taxa
327 richness (*rEPT*) between Spring samples had the strongest relationship with SC, followed by
328 Ephemeroptera taxa richness less Baetidae (*rE-B*). Differences in *rE* and *rE-B* between sample
329 methodologies quickly diminished with increasing SC because of overall Ephemeroptera taxa
330 declines across the SC gradient with extirpations of sensitive non-Baetidae taxa occurring at
331 high-SC streams.
332
333 Community composition
334
335 Percent Ephemeroptera (*pE*) and percent Ephemeroptera less Baetidae (*pE-B*) were significantly
336 and negatively correlated with SC for full-enumeration and RBP samples during both sampling
337 periods. In Spring, the metrics percent Plecoptera (*pP)* and percent Plecoptera plus Trichoptera
338 less Hydropsychidae (*PT-H*) were significantly and positively correlated with SC for both
339 sampling methods. Additionally, in Fall, *pEPT* was positively correlated with SC for both
340 sampling methods. However, during this sampling period, *pP* and *pT-H* were only correlated

(negatively) with SC for full-enumeration samples, whereas *pDip* and *pChi* were only correlated (negatively) for RBP samples. No significant correlations of compositional sampling differences (full-enumeration minus RBP) with SC were noted for either sampling period (Tables 3 and 4).

Functional feeding groups, dominance, diversity, tolerance, and habit

The metrics percent collector-gatherers (*pCG*), percent predators (*pPR*), percent scrapers (*pSC*), and scraper taxa richness (*rSC*) were correlated negatively, and percent Shredders (*pSH*) was correlated positively with SC for both sampling methods in Spring; whereas all of these metrics except *pPR* were correlated similarly for either one or both sampling methods in Fall (Tables 3 and 4). No significant correlations of sampling differences (full-enumeration minus RBP) with SC were noted for macroinvertebrate functional feeding groups.

All the Dominance and Diversity metrics were correlated either positively (*p1dom*, *p2dom*, *p5dom*) or negatively (*even*, *shan*, *simp*) with SC for both sampling methods in Spring and correlations were stronger for full-enumeration samples. In Fall, these metrics were similarly correlated with SC for one or both sampling methods. However, no significant correlations between sampling difference and SC were noted.

Intolerant taxa richness (*rINT*) was negatively correlated with SC in both Spring and Fall and the strongest relationship was detected for full-enumeration samples collected in Spring. In Spring, sample differences in *rINT* exhibited a moderate negative association with SC (Table 3), with

363 larger differences occurring at reference streams and generally diminishing across the salinity
364 gradient.
365
366 Spearman correlations detected a negative relationship between percent clingers (*pCling*) and SC
367 for RBP samples collected in Spring, and positive relationships between *pCling* and SC for both
368 sample types collected in Fall. Clinger taxa richness (*rCling*) was negatively correlated with SC
369 for both sample types collected in Spring and Fall (Tables 3 and 4). Differences in *pCling*
370 between sample types were positively correlated with SC in Spring and differences in *rCling*
371 between sample types exhibited negative relationships with SC for both sampling periods.
372 Although overall clinger taxa richness declines with increasing SC, certain salt-tolerant clinger
373 taxa (including members of families Leuctridae, Hydropsychidae, and Simuliidae) were present
374 at high abundances in high-SC streams.
375

Table 3: Mean values of community metrics for full-enumeration (FE) samples and Rapid Bioassessment Protocol (RBP) samples, for FE minus RBP, and coefficients of Spearman correlation (ρ) of those measures vs. specific conductance (SC) for Spring.

			FE samples		RBP samples		FE minusRBP (Difference)	
Category	Metric	Abbrev.	Mean	ρ vs. SC[a]	Mean	ρ vs. SC	Mean	ρ vs. SC
Taxonomic Richness	Total Taxa	*rTotal*	33.7	**-0.91**	20.9	**-0.66**	**12.8**	**-0.82**
	Ephemeroptera, Plecoptera, and Trichoptera (EPT) Taxa	*rEPT*	21.5	**-0.92**	13.1	**-0.72**	**8.5**	**-0.85**
	EPT Taxa less Hydropsychidae	*rEPT-H*	18.3	**-0.93**	11.0	**-0.78**	**7.3**	**-0.81**
	Ephemeroptera Taxa	*rE*	6.5	**-0.95**	4.1	**-0.69**	**2.4**	**-0.81**
	Ephemeroptera Taxa less Baetidae	*rE-B*	4.9	**-0.94**	2.8	**-0.86**	**2.1**	**-0.84**
	Plecoptera Taxa	*rP*	6.7	**-0.72**	4.7	**-0.61**	**1.9**	-0.25
	Trichoptera Taxa	*rT*	8.4	**-0.68**	4.3	0.03	**4.1**	**-0.55**
	Trichoptera Taxa less Hydropsychidae	*rT-H*	5.1	**-0.58**	2.2	-0.22	**2.9**	-0.33
	Diptera Taxa	*rDip*	7.2	0.04	4.3	-0.09	**2.9**	0.17
Composition	% EPT	*pEPT*	80.0	0.42	74.0	-0.05	**6.0**	0.36
	% EPT less Hydropsychidae	*pEPT-H*	73.8	0.35	64.2	-0.13	**9.6**	0.25
	% Ephemeroptera	*pE*	16.6	**-0.91**	19.0	**-0.85**	-2.5	0.13
	% Ephemeroptera less Baetidae	*pE-B*	8.9	**-0.94**	13.3	**-0.90**	-4.3	0.34
	% Plecoptera	*pP*	53.3	**0.83**	42.7	**0.71**	**10.5**	0.30
	% Trichoptera	*pT*	10.1	-0.06	12.2	0.40	-2.1	-0.34
	% Trichoptera less Hydropsychidae	*pT-H*	4.0	-0.40	2.4	-0.25	1.5	-0.29
	% Plecoptera plus Trichoptera Taxa less Hydropsychidae	*pPT-H*	57.2	**0.84**	45.2	**0.69**	**12.1**	0.24
	% Diptera	*pDip*	15.4	-0.14	19.0	0.26	**-3.5**	-0.40
	% Chironomidae	*pChi*	11.6	-0.04	9.9	0.24	**1.7**	-0.39
	% Oligochaeta	*pOligo*	1.0	0.00	0.2	0.19	**0.7**	-0.10
Functional Feeding Groups	% Collector-Gatherers	*pCG*	24.6	**-0.67**	25.2	**-0.57**	-0.6	-0.12
	% Collector-Filterers	*pCF*	11.0	-0.19	17.3	0.31	**-6.3**	-0.30
	% Predators	*pPR*	6.3	**-0.63**	6.5	**-0.71**	-0.3	0.06
	% Scrapers	*pSC*	6.8	**-0.72**	10.2	**-0.80**	**-3.4**	0.31
	% Shredders	*pSH*	51.3	**0.84**	40.8	**0.71**	**10.6**	0.27
	Scraper Taxa Richness	*rSC*	5.3	**-0.82**	4.0	**-0.65**	**1.3**	-0.49
Dominance	% 1 Dominant Taxon	*p1dom*	38.7	**0.80**	37.7	**0.77**	0.9	0.25
	% 2 Dominant Taxa	*p2dom*	59.1	**0.78**	52.7	**0.77**	6.4	0.41
	% 5 Dominant Taxa	*p5dom*	78.9	**0.83**	77.4	**0.78**	1.5	0.16
	Pielou's Evenness	*even*	0.6	**-0.75**	0.7	**-0.71**	**-0.1**	-0.29
Diversity	Shannon Diversity	*shan*	2.0	**-0.84**	2.1	**-0.76**	-0.1	-0.44
	Simpson Diversity	*simp*	0.8	**-0.81**	0.8	**-0.69**	0.0	-0.18
Tolerance	Intolerant Taxa Richness	*rINT*	18.5	**-0.92**	13.3	**-0.71**	**5.2**	**-0.55**
	% Tolerant	*pTOL*	0.0	0.06	0.0	NA	0.0	0.06
	Hilsenhoff Biotic Index	*HBI*	2.2	**-0.61**	2.5	0.12	-0.4	**-0.59**
Habit	% Clingers	*pCling*	53.6	0.00	48.5	**-0.55**	5.1	**0.67**
	Clinger Taxa Richness	*rCling*	21.5	**-0.87**	13.9	**-0.63**	**7.5**	**-0.74**

[a]Bold indicates a significant result at alpha of 0.05.

Table 4: Mean values of community metrics for full-enumeration (FE) samples and rapid bioassessment protocol (RBP) samples, for FE minus RBP, and coefficients of Spearman correlation (ρ) of those measures vs. specific conductance (SC) for Fall.

			FE samples		RBP samples		FE minusRBP (Difference)	
Category	Metric	Abbrev.	Mean	ρ vs. SC[a]	Mean	ρ vs. SC	Mean	ρ vs. SC
Taxonomic Richness	Total Taxa	*rTotal*	33.7	**-0.75**	22.0	**-0.84**	**11.7**	**-0.61**
	Ephemeroptera, Plecoptera, and Trichoptera (EPT) Taxa	*rEPT*	20.9	**-0.78**	13.7	**-0.81**	**7.2**	**-0.60**
	EPT Taxa less Hydropsychidae	*rEPT-H*	17.1	**-0.77**	10.8	**-0.81**	**6.4**	**-0.57**
	Ephemeroptera Taxa	*rE*	4.9	**-0.89**	3.4	**-0.87**	**1.6**	**-0.58**
	Ephemeroptera Taxa less Baetidae	*rE-B*	4.1	**-0.92**	2.8	**-0.89**	1.4	**-0.60**
	Plecoptera Taxa	*rP*	7.1	-0.51	3.8	**-0.64**	**3.4**	-0.09
	Trichoptera Taxa	*rT*	8.9	-0.42	6.6	-0.08	**2.3**	**-0.56**
	Trichoptera Taxa less Hydropsychidae	*rT-H*	5.1	-0.33	3.9	-0.18	**1.4**	-0.38
	Diptera Taxa	*rDip*	7.2	-0.50	4.4	-0.48	**2.8**	-0.37
Composition	% EPT	*pEPT*	74.9	**0.66**	69.0	**0.65**	5.9	-0.28
	% EPT less Hydropsychidae	*pEPT-H*	43.2	0.01	39.5	0.02	3.7	0.02
	% Ephemeroptera	*pE*	6.7	**-0.79**	7.1	**-0.93**	-0.5	-0.11
	% Ephemeroptera less Baetidae	*pE-B*	6.2	**-0.83**	6.3	**-0.92**	-0.1	-0.05
	% Plecoptera	*pP*	29.6	**0.64**	25.4	0.31	4.2	0.26
	% Trichoptera	*pT*	38.7	0.11	36.5	0.47	2.2	-0.48
	% Trichoptera less Hydropsychidae	*pT-H*	7.0	-0.**70**	7.0	-0.28	-0.1	-0.50
	% Plecoptera plus Trichoptera Taxa less Hydropsychidae	*pPT-H*	36.6	0.45	32.4	0.37	4.2	-0.04
	% Diptera	*pDip*	16.5	-0.38	19.9	**-0.56**	-3.4	0.20
	% Chironomidae	*pChi*	11.1	-0.31	15.1	**-0.54**	-4.0	0.27
	% Oligochaeta	*pOligo*	1.6	-0.15	3.1	-0.44	**-1.5**	0.27
Functional Feeding Groups	% Collector-Gatherers	*pCG*	16.7	-0.46	22.5	**-0.73**	**-5.8**	0.41
	% Collector-Filterers	*pCF*	36.0	0.20	31.5	0.45	4.5	-0.48
	% Predators	*pPR*	7.6	-0.50	7.6	-0.19	0.0	-0.49
	% Scrapers	*pSC*	8.5	**-0.79**	10.6	**-0.71**	-2.2	-0.08
	% Shredders	*pSH*	31.2	**0.78**	27.8	0.33	3.4	0.30
	Scraper Taxa Richness	*rSC*	5.9	**-0.84**	4.2	**-0.73**	**1.7**	-0.41
Dominance	% 1 Dominant Taxon	*p1dom*	32.6	**0.65**	35.6	0.35	-3.0	0.19
	% 2 Dominant Taxa	*p2dom*	48.1	**0.76**	51.4	0.50	-3.3	0.39
	% 5 Dominant Taxa	*p5dom*	71.3	**0.83**	76.2	**0.84**	-4.9	0.34
	Pielou's Evenness	*even*	0.7	-0.48	0.7	**-0.55**	0.0	-0.10
Diversity	Shannon Diversity	*shan*	2.3	**-0.79**	2.2	**-0.61**	0.2	-0.29
	Simpson Diversity	*simp*	0.8	**-0.75**	0.8	-0.50	0.0	-0.20
Tolerance	Intolerant Taxa Richness	*rINT*	19.1	**-0.74**	12.3	**-0.74**	**6.8**	-0.40
	% Tolerant	*pTOL*	0.1	-0.10	0.0	0.45	0.1	-0.30
	Hilsenhoff Biotic Index	*HBI*	2.4	-0.45	2.7	-0.28	-0.3	0.07
Habit	% Clingers	*pCling*	73.9	**0.60**	70.7	**0.74**	3.1	-0.34
	Clinger Taxa Richmess	*rCling*	19.9	**-0.81**	13.9	**-0.69**	**6.0**	**-0.66**

[a]Bold indicates a significant result at alpha of 0.05.

Discussion

Two sampling methods were used to produce metrics describing various aspects of benthic macroinvertebrate community structure and to evaluate their capacity to detect community changes induced by mining influence as indicated by elevated SC. Our data demonstrate that

392 full-enumeration samples consistently produced higher estimates of taxonomic richness than
393 corresponding RBP samples (i.e., subsampled to 200±10% organisms) for both Spring and Fall
394 sampling periods. Most metrics exhibiting significant response to SC for full-enumeration
395 samples also produced such results for RBP samples, but full-enumeration samples yielded more
396 sensitive relationships with SC as indicated by larger correlation coefficients and/or steeper
397 regression slopes. In addition, differences between richness metrics obtained from the two
398 sampling methods declined as SC increased among streams, especially for those metrics that
399 were most sensitive to elevated SC. Below, we discuss these findings and their implications for
400 bioassessments in our study systems and more broadly.
401
402 Comparison of sampling methods for community characterization
403
404 Full-enumeration sampling consistently produced higher estimates than RBP sampling for
405 richness metrics in both seasons, excepting only Ephemeroptera less Baetidae (*rE-B*) in the fall
406 season. The *rE-B* metric exhibited large differences between sampling methods at reference
407 streams, but the non-parametric comparison did not reveal significant differences as the highest-
408 SC streams yielded zero specimens for both sampling methods whereas RBP-based *rE-B* was
409 slightly greater than the full-enumeration measure at other high-SC streams with low relative
410 abundances compared to reference streams (Figure 3).
411
412 Our finding of generally greater taxa richness in the full-enumeration samples, relative to the
413 RBP samples with lower abundance, is consistent with ecological principles, given the well-
414 known positive relationship between taxonomic richness and sampling effort (Fisher et al., 1943)

as indicated in this case by numbers of individuals captured. Other researchers have reported similar results. Bioassessments conducted within streams of Malaysia (Ghani et al., 2016), central Brazil (Silva et al., 2005), southwestern Australia (Storey et al., 1991), and ephemeral lakes of Ireland (O'Connor et al., 2004) reported richer macroinvertebrate assemblages using full-enumeration methods relative to fixed-count samples with fewer organisms. Further, fixed-count samples collected from a river in southeastern Australia contained only 55 to 63% of the total taxa collected by extensive full-enumeration sampling at the same site (Gillies et al., 2009). Everall et al. (2017) collected full-enumeration samples from rivers in the U.K. and found that they contained significantly higher total taxa and EPT taxa richness than fixed-count samples. Similarly, in our study, full-enumeration measures of total taxa richness (*rTotal*) were 48.0% (±4.1) higher than those from RBP fixed-count samples collected in Spring and 40.2% (±4.8) higher in Fall (Figure 3). Previous studies (Armitage et al., 1983) have suggested that collection of at least 75% of total macroinvertebrate taxa is needed to provide a reliable assessment of water quality in temperate lotic systems. Our findings show RBP methodology can miss approximately 50% of the total taxa collected using full-enumeration methods. Consequently, previous studies of central Appalachian headwater streams using RBP methods have likely underestimated and underrepresented benthic macroinvertebrate taxonomic richness, an important descriptor of community structure. However, we note that methodological differences in richness metrics did not translate broadly across all our metrics used to characterize community composition, functional feeding group, dominance, diversity, tolerance, and habit.

A principal difference between sampling methods is that full-enumeration sampling quantifies all organisms collected, whereas the common RBP method identifies a fixed-count subsample.

438 Laboratory processing of RBP samples often requires subsampling using a gridded pan method
439 to obtain a standardized, random sample (Plafkin et al., 1989) potentially introducing sources of
440 error such as uneven distribution of organisms across the grid. Subsampling procedures tend to
441 be biased in selection of larger organisms (Gillies et al., 2009; Vinson & Hawkins, 1996)
442 because they are easier to see and spread consistently and evenly across the subsampling grid.
443 Thus, smaller organisms (such as individuals of the genus *Leuctra*) may be overlooked during
444 sample processing and subsequently underrepresented in community composition metrics.
445 Indeed, we found that percent (%) *Leuctra* differed significantly between full-enumeration and
446 RBP samples collected in Spring, and full-enumeration measures of % *Leuctra* exceeded RBP
447 measures by 106.2% (±25.3). These findings indicate that RBP-derived community
448 compositional metrics based on relative abundance data are associated with a degree of sampling
449 imprecision, potentially rendering RBP-derived data as inconsistent and potentially unreliable for
450 comparison of community-compositional metrics estimated from samples collected during a
451 particular season or over multiple seasons. Others have noted that subsampling procedures
452 introduce imprecision into biomonitoring results, rendering results as less reliable indicators of
453 stream condition and reducing their discriminatory capability relative to full enumeration of the
454 collected samples (Doberstein et al., 2000).
455
456 Comparison of methods for determining benthic community structure response to SC
457
458 Numerous benthic macroinvertebrate community metrics responded to the SC gradient in both
459 Spring and Fall sample periods. Most of the responses documented in this study were similar to
460 those found by previous studies in central Appalachia (Pond et al., 2008, 2014; Timpano et al.,

2015, 2018a; Boehme et al., 2016; Cianciolo et al., 2020) including Ephemeroptera taxa and the scraper functional feeding group being especially sensitive and prone to decline with increasing SC. Our data show that correlations of multiple metrics with SC were significant for both full-enumeration and RBP sampling data in most cases. This finding indicates that both sampling methods can detect the presence of an SC-induced gradient of community structure, particularly differences between metrics in high-salinity streams and reference streams. However, correlations of community metrics with SC were generally stronger (higher-magnitude correlation coefficients, lower p-values) for full-enumeration sampling data than for RBP sampling data, particularly for richness metrics during the spring season, indicating greater capability to reveal statistical significance, especially when magnitude differences are small. Further, multiple richness metrics showed greater rates of change (i.e., steeper slopes) along the SC gradient in full-enumeration versus RBP sampling data. Consequently, differences between full-enumeration and RBP samples also declined in response to SC for several metrics, including *rTotal*, *rEPT*, *rEPT-H*, *rE*, *rE-B*, *rT*, and *rCling.* The difference between sample types was especially pronounced for these richness metrics at low-SC streams, where RBP sampling failed to capture multiple taxa that were included in full-enumeration samples. In contrast, richness values from RBP samples more closely approximated full-enumeration richness values at high-SC streams due to reduced occurrence of unique taxa in the full-enumeration samples. It follows that RBP sampling data are less capable of revealing differences between moderately salinized streams versus reference conditions, particularly for richness metrics, and less capable of doing so with statistical certainty. Analogous to our findings, Cao et al. (1998) found that among streams spanning a stressor gradient, taxonomic richness differences among streams increased with increasing sample size.

Implications for Bioassessment of Freshwater Streams and Future Work

Freshwater stream bioassessments are typically conducted by comparing biomonitoring metrics to a reference level that defines what is acceptable and is often based on conditions in reference streams such as those included in this study. In the Appalachian coalfield, bioassessments are commonly conducted using data derived from fixed-count RBP samples with subsampling to a fixed organism count.

Other researchers have explored issues concerning sample-size effects on utility of biomonitoring data, noting that smaller samples often fail to capture taxa that are rare within a given environment and that the utility of relatively small samples often depends on the purpose for the bioassessments. When purposes include detection of stressor effects along a strong environmental gradient or gross separations of highly degraded from relatively unimpaired streams, smaller sample sizes are often quite adequate (Cao et al., 1998; Arscott et al., 2006). Such has been the purposes of many studies of mining-origin salinity conducted to date in the Appalachian coalfield that have utilized RBP methods with fixed-count subsampling (e.g. Pond et al., 2008, 2014). Many of the SC-gradient effects noted by such studies (e.g. reduced richness of total taxa and of sensitive taxa such as Ephemeroptera; reduced relative abundances of scrapers and or Ephemeroptera at high -SC streams) are replicated here in our study in data generated by both sampling methods. For other purposes, however, larger samples, which capture larger numbers of taxa play a more important role. Sample accuracy as an estimator of the sampled community condition increases with increasing sample size (Vlek et al., 2006).

Hence, biomonitoring and assessment procedures such as defining conditions in relatively undisturbed habitats as a baseline for evaluating future impacts, discriminating moderately impaired from relatively unimpaired and/or highly impaired streams or otherwise separating stream sites into multiple categories are best accomplished using highly enumerated samples (Doberstein et al., 2000; Arscott et al., 2006).

Conclusions

This study demonstrates that full-enumeration bioassessments offer advantages over RBP fixed-count bioassessments of benthic macroinvertebrate communities in mining-influenced central Appalachian headwater streams. Full-enumeration sampling provides more complete estimates of taxonomic richness, and hence estimates of diversity that are more accurate, compared to the RBP sampling methods with subsampling to fixed and often far lower organism counts as is commonly applied for assessment of mining impacts to stream communities in the Appalachian coalfield. In addition, full-enumeration sampling enhances the ability to interpret biological condition in headwater streams by characterizing community-metric responses to a stressor gradient and can be particularly useful when comparing less-degraded systems to those in a reference condition. Rapid bioassessment samples, with comparatively small numbers of individuals, are often able to discriminate highly degraded systems from those in a reference condition but may not detect more subtle changes.

Both RBP and full-enumeration sampling revealed differences in community condition across a mining-induced salinity gradient. Although RBP sampling showed stream-community sensitivity

530 to the SC stressor evaluated in this study, the extra time and effort required for full-enumeration
531 sampling can provide additional insight into benthic macroinvertebrate community structure and
532 stressor response.
533
534 **References**
535
536 Armitage, P. D., Moss, D., Wright, J.F., & Furse, M.T. (1983). The performance of a new
537 biological water quality score system based on macroinvertebrates over a wide range of
538 unpolluted running water sites. *Water Research*, 17(3): 333-347. https://doi.org/10.1016/0043-
539 1354(83)90188-4
540
541 Arscott, D. B., Jackson J. K., & Kratzer E. B. (2006). Role of rarity and taxonomic resolution in
542 a regional and spatial analysis of stream macroinvertebrates. *Journal of the North American*
543 *Benthological Society*, 25, 977-997. https://doi.org/10.1899/0887-
544 3593(2006)025[0977:RORATR]2.0.CO;2
545
546 Barbour, M. T., Gerrtisen, J., Snyder, B. D., & Stribling, J. B. (1999). Rapid bioassessment
547 protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates, and
548 fish, Second Edition. USEPA, Washington, DC.
549
550 Birk, S., Bonne W., Borja A., Brucet S., Courrat A., et al. 2012. Three hundred ways to assess
551 Europe's surface waters: An almost complete overview of biological methods to implement the

Water Framework Directive. *Ecological Indicators*, 18, 31-41. https://doi.org/10.1016/j.ecolind.2011.10.009

Boehme, E. A., Zipper, C. E., Schoenholtz, S. H., Soucek, D. J., & Timpano, A. J. (2016). Temporal dynamics of benthic macroinvertebrate communities and their response to elevated specific conductance in Appalachian coalfield headwater streams. *Ecological Indicators*, 64, 171-180. https://doi.org/10.1016/j.ecolind.2015.12.020

Bryant, G., McPhilliamy, S., & Childers, H. (2002). A survey of the water quality of streams in the primary region of mountaintop/valley fill coal mining, October 1999 to January 2001. Mountaintop mining/valley fill programmatic environmental impact statement. Region 3, US Environmental Protection Agency, Philadelphia, Pennsylvania.

Buss, D. F., & Borges E. L. (2008) Application of rapid bioassessment protocols (RBP) for benthic macroinvertebrates in Brazil: Comparison between sampling techniques and mesh sizes. *Neotropical Ecology*, 37: 288-295. https://doi.org/10.1590/S1519-566X2008000300007

Cañedo-Argüelles M., Kefford B. J., Piscart C., Prat N., Schäfer R. B., & Schulz C.-J., Salinisation of rivers: An urgent ecological issue. *Environmental Pollution*. 173, 157-167. https://doi.org/10.1016/j.envpol.2012.10.011

Cao, Y., Williams D. D., Williams N. E. (1998). How important are rare species in aquatic community ecology and bioassessment? *Limnology and Oceanography*, 43, 1403-1409. https://doi.org/10.4319/lo.1998.43.7.1403

Cianciolo, T. R., McLaughlin, D. L., Zipper, C. E., Timpano, A. J., Soucek, D. J., & Schoenholtz, S.H. (2020). Impacts to water quality and biota persist in mining-influenced Appalachian streams. *Science of The Total Environment*, 717, p.137216. https://doi.org/10.1016/j.scitotenv.2020.137216

Courtemanch, D. L. (1996). Commentary on the subsampling procedures used for rapid bioassessments. *Journal of the North American Benthological Society*, 15, 381-385. https://doi.org/10.2307/1467284

Davies, A. (2001). The use and limits of various methods of sampling and interpretation of benthic macro-invertebrates. *Journal of Limnology*, 60(1s), 1-6.

Doberstein, C. P., Karr J. R., & Conquest L.L. (2000). The effect of fixed-count subsampling on macroinvertebrate biomonitoring in small streams. *Freshwater Biology*, 44, 355-371. https://doi.org/10.1046/j.1365-2427.2000.00575.x

Drover, D. R., Schoenholtz, S. H., Soucek, D. J., & Zipper, C. E. (2020). Multiple stressors influence benthic macroinvertebrate communities in central Appalachian coalfield streams. *Hydrobiologia*, 847: 191–205, https://doi.org/10.1007/s10750-019-04081-4.

Everall, N. C., Johnson, M. F., Wood, P., Farmer, A., Wilby, R. L., & Measham, N. (2017). Comparability of macroinvertebrate biomonitoring indices of river health derived from semi-quantitative methodologies. *Ecological Indicators*, 78, 437-448. https://doi.org/10.1016/j.ecolind.2017.03.040

Fisher, R. A., Corbet A. S., & Williams C. B. (1943). The relation between the number of species and the number of individuals in a random sample of an animal population. *Journal of Animal Ecology*, 12, No. 1, pp. 42-58. https://doi.org/10.2307/1411

Ghani, W. M., Rawi, C. S., Hamid, S. A., & Al-Shami, S. A. (2016). Efficiency of different sampling tools for aquatic macroinvertebrate collections in Malaysian streams. *Tropical Life Sciences Research*, 27(1), 115-133.

Gillies, C. L, Hose, G. C., & Turak, E. (2009). What do qualitative rapid bioassessment collections of macroinvertebrates represent? A comparison with extensive quantitative sampling. *Environmental Monitoring and Assessment*, 149, 99-112. https://doi.org/10.1007/s10661-008-0186-9

Governor, H., Krometis, L. A. H., & Hession, W. C. (2017). Invertebrate-based water quality impairments and associated stressors identified through the US Clean Water Act. *Environmental Management*, 60, 598–614. https://doi.org/10.1007/s00267-017-0907-3

Hartman, K. J., Kaller, M. D., Howell, J. W., & Sweka, J. A. (2005). How much do valley fills influence headwater streams? *Hydrobiologia*, 532, 91-102. https://doi.org/10.1007/s10750-004-9019-1

Hothorn & Hornik. (2018). exactRankTests: Exact Distributions for Rank and Permutation Tests. R package version 0.8-31. https://CRAN.R-project.org/package=exactRankTests

Kaushal, S.S., Likens G.E., Pace M.L., Utz R.M., Haq S., Gorman J. & Grese M. (2018) Freshwater salinization syndrome on a continental scale. *Proceedings of the National Academy of Sciences*, 115, E574-E583. https://doi.org/10.1073/pnas.1711234115

Lindberg, T. T., Bernhardt, E. S., Vier, R., Helton, A. M., Merola, R. B., Vengosh, A., & Di Giulio, R. T. (2011). Cumulative impacts of mountaintop mining on an Appalachian watershed. *Proceedings of the National Academy of Sciences*, 108(52), 20929-20934. https://doi.org/10.1073/pnas.1112381108

Merritt, R. W., Cummins, K. W., & Berg, M. B. (2008). An Introduction to the Aquatic Insects of North America. Dubuque: Kendall/Hunt.

Morse, J. C., Bae Y. J., Munkhjargal G., Sangpradub N., Tanida K., et al. (2007). Freshwater biomonitoring with macroinvertebrates in East Asia. *Frontiers in Ecology and Environment*, 5, 33–42. https://doi.org/10.1890/1540-9295(2007)5[33:FBWMIE]2.0.CO;2

Nichols, S. J, Barmuta L. A., Chessman B. C., Davies P. E., Dyer F. J. (2017). The imperative need for nationally coordinated bioassessment of rivers and streams. *Marine and Freshwater Research*, 68, 599–613. https://doi.org/10.1071/MF15329

O Connor, Á., Bradish, S., Reed, T. et al. (2004). A Comparison of the Efficacy of Pond-Net and Box Sampling Methods in Turloughs - Irish Ephemeral Aquatic Systems. *Hydrobiologia*, 524, 133–144. https://doi.org/10.1023/B:HYDR.0000036128.83998.44

Omernik, J. M. (1987). Ecoregions of the conterminous United States. *Annals of the Association of American Geographers*, 77, 118-125. https://doi.org/10.1111/j.1467-8306.1987.tb00149.x

Pericak, A. A., Thomas, C. J., Kroodsma, D. A., Wasson, M. F., Ross, M. R., Clinton, N. E., Campagna, D. J., Franklin, Y., Bernhardt, E. S., & Amos, J. F. (2018). Mapping the yearly extent of surface coal mining in Central Appalachia using Landsat and Google Earth Engine. *PloS one*, 13(7). https://doi.org/10.1371/journal.pone.0197758.

Pielou, E. C. (1966). The measurement of diversity in different types of biological collections. *Journal of Theoretical Biology*, 13, 131-144. https://doi.org/10.1016/0022-5193(66)90013-0

Plafkin, J. L., Barbour, M. T., Porter, K. D., Gross, S. K., & Hughes, R. M. (1989). Rapid bioassessment protocols for use in streams and rivers: benthic macroinvertebrates and fish. USEPA, Washington, D.C.

Pond, G. J., Passmore, M. E., Borsuk, F. A., Reynolds, L., & Rose, C. J. (2008). Downstream effects of mountaintop coal mining: comparing biological conditions using family- and genus-level macroinvertebrate bioassessment tools. *Journal of the North American Benthological Society*, 27(3), 717-737. https://doi.org/10.1899/08-015.1

Pond, G. J. (2010). Patterns of Ephemeroptera taxa loss in Appalachian headwater streams (Kentucky, USA). *Hydrobiologia*, 641, 185-201. https://doi.org/10.1007/s10750-009-0081-6

Pond, G. J., Bailey, J. E., Lowman, B. M., & Whitman, M. J. (2013). Calibration and validation of a regionally and seasonally stratified macroinvertebrate index for West Virginia wadeable streams. *Environmental Monitoring and Assessment*, 185, 1515-1540. https://doi.org/10.1007/s10661-012-2648-3

Pond, G. J., Passmore, M. E., Pointon, N. D., Felbinger, J. K., Walker, C. A., Krock, K. J. G., Fulton, J. B., & Nash, W. L. (2014). Long-term impacts on macroinvertebrates downstream of reclaimed mountaintop mining valley fills in central Appalachia. *Environmental Management*, 54, 919-933. https://doi.org/10.1007/s00267-014-0319-6

RStudio Team. (2018). RStudio: Integrated development environment for R. RStudio, Inc., Boston, Massachusetts USA.

Schäfer, R. B., Bundschuh, M., Rouch, D. A., Szöcs, E., Von Der Ohe, P. C., Pettigrove, V., Schulz, R., Nugegoda, D., & Kefford, B. J. (2012). Effects of pesticide toxicity, salinity and

other environmental variables on selected ecosystem functions in streams and the relevance for ecosystem services. *Science of the Total Environment*, 415, 69-78. https://doi.org/10.1016/j.scitotenv.2011.05.063

Silva, L. C. F., Vieira, L. C. G., Costa, D. A., Lima Filho, G. F., Vital, M. C. C, Carvalho, R. A., Silveira, A. V. T., & Oliveira, L. C. (2005). Qualitative and quantitative benthic macroinvertebrate samplers in Cerrado streams: a comparative approach. *Acta Limnologica Brasiliensia*, 17, 123-128.

Storey, A. W., Edward, D. H. D., & Gazey, P. (1991). Surber and kick sampling: a comparison for the assessment of macroinvertebrate community structure in streams of south-western Australia. *Hydrobiologia*, 211, 111-121. https://doi.org/10.1007/BF00037367

Timpano, A. J., Schoenholtz, S. H., Soucek, D. J., & Zipper, C. E. (2015). Salinity as a limiting factor for biological condition in mining-influenced Central Appalachian headwater streams. *Journal of the American Water Resources Association*, 51(1), 240-250. https://doi.org/10.1111/jawr.12247

Timpano, A. J., Schoenholtz, S. H., Soucek, D. J., & Zipper, C. E. (2018a). Benthic macroinvertebrate community response to salinization in headwater streams in Appalachia USA over multiple years. *Ecological Indicators*, 91, 645-656. https://doi.org/10.1016/j.ecolind.2018.04.031

Timpano, A. J., Zipper, C. E., Soucek, D. J., & Schoenholtz, S. H. (2018b). Seasonal pattern of anthropogenic salinization in temperate forested headwater streams. *Water Research*, 133, 8-18. https://doi.org/10.1016/j.watres.2018.01.012

Tubić B. P., Popović N. Z., Raković M. J., Petrović A. S., Simić V. M. & Paunović M. M. (2017). Comparison of the effectiveness of kick and sweep hand net and Surber net sampling techniques used for collecting aquatic macroinvertebrate samples. *Archives of Biological Sciences*, 69, 233-238. https://doi.org/10.2298/ABS160622087T

USEPA. (2011). A Field-Based Aquatic Life Benchmark for Conductivity in Central Appalachian Streams. Office of Research and Development, National Center for Environmental Assessment, Washington, D.C.

VDEQ. (2008). Biological monitoring program quality assurance project plan for wadeable streams and rivers. Virginia Department of Environmental Quality, Water Quality Monitoring and Assessment Programs, Richmond, Virginia.

Vinson, M. R., & Hawkins, C. P. (1996). Effects of sampling area and subsampling procedure on comparisons of taxa richness among streams. *Journal of the North American Benthological Society*, 15(3), 392-399. https://doi.org/10.2307/1467286

Vlek, H. E., Sporka, F., & Krno, I. (2006). Influence of macroinvertebrate sample size on bioassessment of streams. *Hydrobiologia*, 566, 523–542. https://doi.org/10.1007/978-1-4020-5493-8_35

Wiggins, G.B. (1996). Larvae of the North American caddisfly genera (Trichoptera). University of Toronto Press, Toronto.

Title

Comparison of Benthic Macroinvertebrate Biomonitoring Approaches for Evaluating Community Response Across a Stressor Gradient in Headwater Streams

Author Information

Rachel A. Pence

Rivanna Conservation Alliance, Charlottesville, VA, USA

Thomas R. Cianciolo

Trout Unlimited, Klamath Falls, OR, USA

Damion R. Drover

Parker Design Group, Roanoke, VA, USA

Daniel L. McLaughlin

Department of Forest Resources and Environmental Conservation, Virginia Tech, Blacksburg, VA, USA

David J. Soucek

Illinois Natural History Survey, Champaign, IL, USA

Anthony J. Timpano

Department of Fish and Conservation, Virginia Tech, Blacksburg, VA, USA

Carl E. Zipper

School of Plant and Environmental Sciences, Virginia Tech, Blacksburg, VA, USA

Stephen H. Schoenholtz (corresponding author; schoenhs@vt.edu)

Virginia Water Resources Research Center, Virginia Tech, Blacksburg, VA, USA

Abstract

Benthic macroinvertebrate community assessments are commonly used to characterize aquatic systems and increasingly for identifying their impairment caused by myriad stressors. Yet, sampling and enumeration methods vary, and research is needed to compare their ability to detect macroinvertebrate community response to specific water quality variables. A common assessment approach, rapid bioassessment uses sub-sampling procedures to identify a fixed number of individuals regardless of total sample abundance. In contrast, full-enumeration assessments typically allow for expanded community characterization resulting from higher numbers of identified organisms. Here, we compared these two sampling and enumeration

methods and their ability to detect benthic macroinvertebrate response to freshwater salinization, a common stressor of streams worldwide. We applied both methods in headwater streams along a salinity gradient within the coal-mining region of central Appalachia USA. Full-enumeration samples had approximately seven times higher abundance and estimates of taxonomic richness were 26-92% higher than rapid-bioassessment samples. Metrics of community composition, trophic function, dominance, tolerance, and habit also differed between the methods. Full-enumeration was generally a stronger predictor of benthic community responses to salinization and was more sensitive than rapid-bioassessment for detecting community changes in salinized streams. However, the more cost-effective and widely-employed rapid-bioassessment method was capable of characterizing severe community alterations associated with the mining-induced salinity gradient. These findings can help inform decisions regarding such tradeoffs for assessments of freshwater salinization in temperate forested headwater streams and highlight the need for similar research of sampling and enumeration methodology in other aquatic systems and for other stressors.

Keywords

salinization, rapid bioassessment, full enumeration, coal mining

Declarations

Funding: Support for this research was provided by the Powell River Project at Virginia Tech.

Conflicts of interest: The authors declare no conflicts of interest.

Availability of data and material: The benthic macroinvertebrate monitoring data for this study are available as Tables A-10 through A-13 in: Pence, R.A. 2019. Comparison of Quantitative and Semi-Quantitative Assessments of Benthic Macroinvertebrate Community Response to Elevated Salinity in Central Appalachian Coalfield Streams. M.S. Thesis. Virginia Tech, Blacksburg. https://vtechworks.lib.vt.edu/handle/10919/86787

Code availability: Custom R code was developed for this research project and is available upon reasonable request.

Figure 1

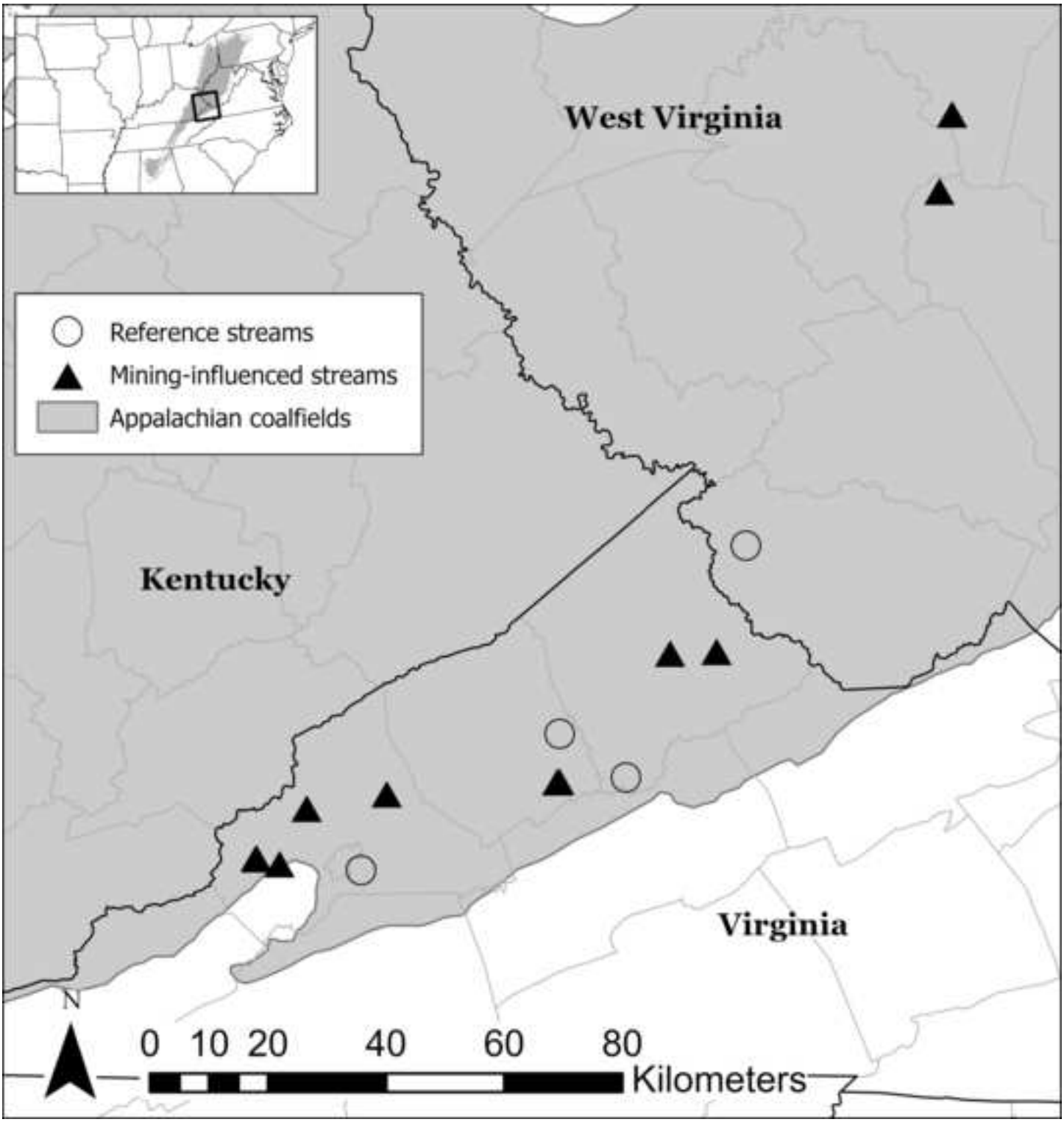

Figure 2

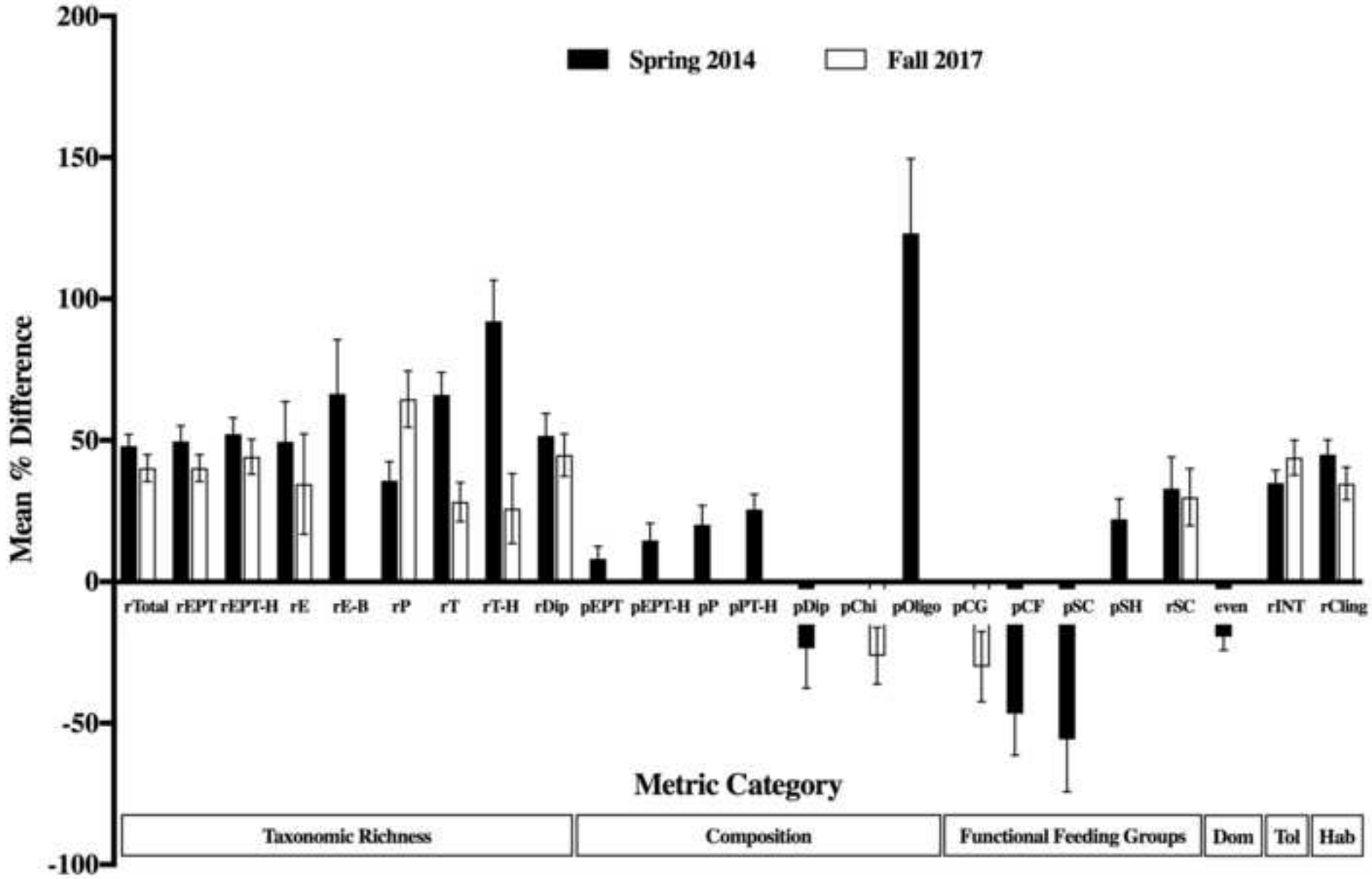

Figure 3