

David Costello ORCID iD: 0000-0002-1532-5399

Anna Harrison ORCID iD: 0000-0002-4752-5635

Running head: Homogenization resets toxicity of metal-amended sediment

## Hitting reset on sediment toxicity: Sediment homogenization alters the toxicity of metal-amended sediments

David M. Costello<sup>a,\*</sup>, Anna M. Harrison<sup>b</sup>, Chad R. Hammerschmidt<sup>c</sup>, Raissa M. Mendonca<sup>a</sup>, and G. Allen Burton, Jr.<sup>b</sup>

<sup>a</sup> Department of Biological Sciences, Kent State University, Kent, OH, USA

<sup>b</sup> School for Environment and Sustainability, University of Michigan, Ann Arbor, MI, USA

<sup>c</sup> Earth and Environmental Sciences, Wright State University, Dayton, OH, USA

\* Address correspondence to David Costello: [dcostel3@kent.edu](mailto:dcostel3@kent.edu)

### Acknowledgements

Jennifer Daley, Kyle Fetters, Maggie Grundler, Sara Nedrich, Olivia Rath, Lauren Estes, and Larissa Sano assisted with experimental setup and sampling. Alison Agather, Katelynn Alcorn, Deepthi Nalluri, Nick Johnson, Daniel Marsh, and Brendan Shields

This is the author manuscript accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1002/etc.4512](https://doi.org/10.1002/etc.4512).

This article is protected by copyright. All rights reserved.

assisted with sample analysis. Chris Schlekat and four anonymous reviewers provided us with comments that improved the manuscript. Funding for this project was provided by Rio Tinto, Nickel Producers Environmental Research Organization, Copper Alliance, International Lead Zinc Research Organization, Vanitec, and Cobalt Development Institute.

### **Data accessibility**

All data and code for analyses and figures is available on GitHub at <https://doi.org/10.5281/zenodo.3236346>

### **Abstract**

Laboratory testing of sediments frequently involves manipulation by amendment with contaminants and homogenization, which changes sediment physicochemical structure. These changes can influence the bioavailability of divalent metals, and field and mesocosm experiments have shown that laboratory-derived thresholds are often overly conservative. We assess the mechanisms that lead to divergence between laboratory- and field-derived thresholds; specifically, we assessed the importance of slow equilibration to solid-phase ligands and vertical stratification. To mimic natural physicochemical conditions, we uniquely aged sediment with a flow-through exposure system. These sediments were then homogenized and compared, toxicologically, to freshly metal-amended sediments in a 28-day chronic toxicity bioassay with the amphipod *Hyaella azteca*. We assessed concentration–response relationships for three metals (Cu, Ni, Zn) and five geochemically distinct sediments. We observed minimal differences in EC10 values and survival of *H. azteca* between aged and freshly spiked sediments across all

This article is protected by copyright. All rights reserved.

sediments and metals. These trends suggest that a loss of toxicity observed during long-term sediment aging is reversed upon sediment homogenization. By comparison to mesocosm experiments, we demonstrate that homogenizing sediment immediately before toxicity assays may produce artificially high toxicity thresholds. We suggest that toxicity assays with sediments that maintain vertical redox gradients are needed to generate field-relevant sediment metal toxicity thresholds.

## **Introduction**

Risk assessment of chemical contaminants in sediments requires a multi-tiered approach of field observations and laboratory tests (Burton et al. 2002, Adams 2005, Dale et al. 2008). Laboratory exposures allow for control of experimental conditions and facilitate development of causal links between chemical concentration and toxicological endpoints. However, laboratory testing often requires manipulation of sediments that is often not representative of conditions in natural ecosystems (Burton et al. 2002). This is of particular concern for divalent metals (e.g., copper (Cu), nickel (Ni), and zinc (Zn)), because their fate and availability is modified by environmental conditions (Campbell and Tessier 1996, Chapman and Wang 1998). Different toxicological results between laboratory and field tests of metals in sediments are of particular concern because (1) adding metals to sediment can alter physicochemical conditions, (2) natural sediments are not homogenous and typically are characterized by sharp vertical physicochemical gradients, and (3) reestablishment of physicochemical conditions (e.g., pH, redox, overlying and porewater chemistry) is slow and can often take longer than recommended exposure periods of toxicity test organisms (Hutchins et al. 2009).

Recent advances in methods for metal addition to sediment (Simpson et al. 2004, Hutchins et al. 2009, Brumbaugh et al. 2013) have maintained natural physicochemical conditions, particularly anoxia. Estimates for porewater equilibration times for various metals range from 10 to 70 days (Simpson et al. 2004) if amendments are done under proper pH buffered conditions (Hutchins et al. 2008). Even with properly buffered and equilibrated sediments, overlying water must be exchanged frequently during toxicity tests to avoid concentrations of metal in overlying water that exceed toxicity thresholds (U.S. EPA 2000, ASTM 2010, Brumbaugh et al. 2013). However, even with best-practices for sediment preparation (i.e., long equilibration and pH buffering) and sufficient water exchanges, mesocosm studies have demonstrated that the geochemistry and toxicity of sediments can continue to change with time (De Jonge et al. 2012, Costello et al. 2015, 2016). This suggests that even state-of-the-art laboratory sediment toxicity tests are still falling short of simulating natural conditions.

Field and mesocosm experiments have highlighted the limitations of laboratory assays for translating impairment thresholds for metals into effects thresholds observed under field conditions (Burton et al. 2005, Costello et al. 2011, 2015, 2016, Mendonca et al. 2017). In general, laboratory-derived thresholds for metals are overly conservative when compared to concentrations at which effects are observed in either mesocosm or field studies (Costello et al. 2011, 2015, 2016). The disparity of results laboratory and field studies have been hypothesized to result from physicochemical changes to sediments that occur during aging under field conditions, which includes advection of porewater metal, oxidation of sediment surfaces, and shifting metal fractionation among solid-phase ligands (Besser et al. 2008, Costello et al. 2011, 2015, Hong et al. 2011, De

Jonge et al. 2012). Advection of porewater metal can be minimized in laboratory tests with pH buffering and extended equilibrium times (Hutchins et al. 2009, Brumbaugh et al. 2013), but more effort is needed to understand the role of surface sediment oxidation and slow changes in solid-phase metal fractionation (hereafter referred to as “slow equilibration”) in modifying the toxicity of sediment. Furthermore, if differences in toxicity are related to vertical stratification then this can have implications for risk assessment of frequently disturbed sediment. For example, suspended and redeposited sediment particles can be more toxic than undisturbed bedded sediment (Fetters et al. 2016) and reworking of sediments by bioturbating organisms can alter exposure and toxicity of metals (Remaili et al. 2016, 2017).

In prior studies, we demonstrated that aging metal-amended sediment in a flow-through mesocosm altered metal partitioning and reduced toxicity to the epibenthic amphipod *Hyalella azteca* (Costello et al. 2015, 2016). We argued that the toxicological and geochemical characteristics of these sediments at the end of the mesocosm incubation more closely resembled field-contaminated sediments. Here we collected and gently homogenized the aged sediments (hereafter “aged treatment”) and prepared fresh metal amended sediment (hereafter “freshly spiked treatment”) to test whether the change in toxicity through time observed in the mesocosm tests was due to either slow equilibration of metal or development of an oxidized surface layer. We predicted that if slow equilibration were the mechanism by which the sediments in the mesocosm were less toxic through time, then the aged sediment will exhibit lower toxicity than the freshly spiked sediment. We also hypothesized that if the oxidized surface layer were critical to lowered toxicity, then the aged sediment would have similar toxicity as freshly prepared

sediment because the oxidized layer will not be present in either sediment. Through this study, we sought to gain mechanistic understanding of differences in apparent toxicity between field and laboratory sediment and better understand how spatial heterogeneity and disturbance of sediment may influence bioavailability and toxicity of metals in sediment.

### **Materials and methods**

Study design. Three separate experiments were completed for each of three metals: Cu, Ni, and Zn. For each experiment, sediment was amended with metal and either immediately used for standard toxicity assays (U.S. EPA 2000, ASTM 2010) or aged in a flow-through mesocosm for >100 days prior to use in toxicity tests. Although sediment preparation times differed for freshly spiked and aged sediments, preparation was staggered so all toxicity assays for a single metal started within 3 d of each other. Sediment preparation techniques and daily experimental maintenance differed among metals and were modeled after the most robust methods (i.e., paired geochemistry and toxicity) in the literature at the time of the experiments (2013–2014; Table 1). Using this approach required Cu tests with less sophisticated metal amendment methods (i.e., no pH buffering or equilibration under N<sub>2</sub> headspace) than tests for either Ni or Zn. All toxicity tests used the amphipod *H. azteca* in chronic tests (28-d) with assessment of both mortality and growth endpoints. Sediment, porewater, and overlying water were sampled at the beginning and end of experimental treatments to characterize differences in metal partitioning and bioavailability between sediment preparation techniques.

Sediment preparation. Surficial sediments (0–10 cm depth) were sampled from uncontaminated stream ecosystems. Dow Creek and River Raisin are low gradient streams in Michigan with no history of metal contamination. The Ocoee and Burtwood Rivers have a history of Cu and Ni mining, respectively, but upstream tributaries were sampled for this study. For Cu and Ni tests, two sediments that differed in key metal-binding ligand concentrations – notably, acid volatile sulfide (AVS), organic matter, iron (Fe) oxides – were used for each test; for the Zn experiment, a single sediment with a high concentration of AVS and organic matter was used for the fresh versus aged comparison (Table 2). Sediments were stored at room temperature under N<sub>2</sub> headspace until amendment with metal (i.e., spiking). Sediments for the aged treatment were amended with metal following previously described approaches using indirect spiking and pH buffering (Hutchins et al. 2008, 2009, Costello et al. 2011, 2015, 2016, Brumbaugh et al. 2013). Five Cu and Ni treatment concentrations were prepared to span an expected non-toxic to toxic range (four amended and one reference; Costello et al. 2015, 2016), and six treatments were made for the Zn experiment (five amended and one reference). Importantly, aging of sediments in the mesocosm did not change total metal concentrations but altered the vertical redox gradients (i.e., oxidized the top few mm), which influenced the partitioning of metal to porewater (Costello et al. 2015, 2016). The aged sediments were prepared for the beaker tests by collecting sediments from the mesocosm for each metal treatment and homogenizing them (placed on a bottle roller at 30 rpm for 1 h) prior to dividing homogenized sediment into six replicate 300-mL beakers (sediment:water ratio in Table 1).

Concurrent with the collection of aged sediment from the mesocosms, sediments from the same field site were freshly amended with the same nominal total metal concentrations. The amendment (i.e., “spiking”) methods differed for each metal but were selected to best replicate the conditions used in previous studies that generated high-quality chronic *H. azteca* concentration–response thresholds: Cu additions replicated Roman and colleagues (2007) who used laboratory-formulated sediment and brief metal equilibration times, Ni amendments followed Besser and colleagues (2013) who used indirect spiking and long equilibration times, and Zn additions followed an indirect method similar to Nguyen and colleagues (2005). Metal additions for Ni and Zn beaker tests were comparable to sediment preparation prior to aging, but the Cu methods differed substantially because equilibration was done within the beakers (Roman et al. 2007). Dissolved Cu (<0.2  $\mu\text{m}$ ) was measured in overlying water to confirm that this equilibration approach did not lead to waterborne Cu toxicity. Six replicates of each treatment of freshly spiked sediment were added to 300-mL beakers in the same sediment:water ratio as aged sediments (Table 1). Prior to addition of organisms, sediments in exposure beakers were further equilibrated within the water renewal system for 1–10 d to flush any excess dissolved metal (Table 1).

*Hyalella azteca* assays. Toxicity of sediments was determined with 28-d chronic *H. azteca* assays. Amphipods were 7–10 d old at the start of each experiment and 10 individuals were added to each of the five replicates of each treatment. The dry mass of *H. azteca* ( $\pm 0.001$  mg) at the start of each experiment was determined by weighing 5–6 groups of 10 randomly sampled individuals. During the 28-d exposure, overlying water was exchanged at a rate that closely replicated the previously mentioned experiments



(i.e., Cu and Ni, Roman et al. 2007, Besser et al. 2013) or followed U.S. EPA guidance (i.e., Zn, U.S. EPA 2000). Water exchange rates differed among experiments; in the Cu test, water was exchanged 3 times a week, whereas in the Ni and Zn experiments an automated system exchanged the water every 3 or 12 hours, respectively (Table 1). Amphipods were fed daily (Table 1), and in tests with frequent water exchange, feeding was done immediately after a water change. Dissolved oxygen, pH, and temperature were monitored weekly in a representative beaker of each treatment, while conductivity, hardness, and ammonia were measured at the beginning and end of the experiment. All water quality parameters passed acceptability thresholds (U.S. EPA 2000). At the end of the 28-d exposure, surviving *H. azteca* were removed from the beakers and placed into 50  $\mu\text{M}$  EDTA for 24 h to void their gut contents and remove adsorbed metal (Neumann et al. 1999). After gut evacuation, *H. azteca* were dried and weighed to determine mass-specific relative growth rates (RGR) (Welton and Clarke 1980). Dried *H. azteca* from Ni and Zn tests were digested with  $\text{HNO}_3$  and  $\text{H}_2\text{O}_2$  and analyzed by inductively coupled plasma mass spectrometry (ICP-MS) for determination of tissue metal concentrations (Norwood et al. 2006).

Sediment and water physicochemistry. Overlying water was sampled and filtered (0.2  $\mu\text{m}$  PES filter) for determination of filtered metal concentrations during the equilibration and exposure periods. For the Cu test, which utilized the shortest equilibration period, we measured dissolved Cu (by ICP-MS) during the 10-d equilibration period (prior to all water changes) to examine Cu flux into overlying water. A single beaker from each sediment treatment was destructively sampled at the beginning and end of the 28-d exposure period for determination of sediment chemistry. Sediments

in the beaker were divided into surface (0–1 cm) and deep (1–3 cm) strata, promptly frozen, and analyzed for AVS, simultaneously extracted metal (SEM), total metal, organic matter (OM), amorphous Fe oxide and associated metal, crystalline Fe oxide and associated metal, and Mn oxide concentrations. A single peeper (Brumbaugh et al. 2007) was placed within the sediment at the bottom of the beakers reserved for sediment geochemistry. Recovered peepers were rinsed of particles, and peeper fluid was removed and acidified with HNO<sub>3</sub> (to 1%) prior to analysis of dissolved organic carbon (DOC) (Cu and Ni only) with an infrared combustion analyzer and dissolved metals by ICP-MS. No peepers were collected for the initial Zn sediment chemistry because of insufficient time for peepers to equilibrate with pore fluids (<7 d).

Sediments were thawed prior to analysis of redox-sensitive constituents; AVS, SEM, and Fe/Mn oxides were analyzed according to standard procedures (Allen et al. 1993, Kostka and Luther 1994, U.S. EPA 2005). Due to very low concentrations of AVS in Burntwood sediments (<0.1 μmol g<sup>-1</sup>), Ni<sub>SEM</sub> was determined with a modified AVS technique where the cold acid extraction was completed but sulfide was not measured (Costello et al. 2016). Oxidized Fe and associated Cu, Ni, or Zn were analyzed with either extractions selective for amorphous Fe (buffered ascorbate) or total reducible Fe and Mn (buffered sodium hydrosulfite; Kostka and Luther 1994). Crystalline Fe oxides and associated metals were calculated as the difference between the total Fe oxide pool and amorphous Fe oxides. Total extractable metals were measured by digesting dried sediment with concentrated HNO<sub>3</sub> at reflux temperatures in a hot block (Environmental Express). Aqueous solutions from SEM, oxide-associated, and total metals extractions were analyzed by inductively couple plasma optical emission spectrometry (ICP-OES).

Organic matter was measured as the mass loss of dried sediment after burning at 400 °C for 4 h. Quality control of analytical chemistry is reported in Supporting Information (Table S1).

Data analysis. Sediment and porewater chemistry were analyzed using three-way analysis of variance (ANOVA) with sampling time (initial or final), sediment depth (surface or deep), and sediment preparation (freshly spiked or aged) as main effects. A statistical covariate accounting for the total metal concentration was included in all ANOVAs with the exception of those physicochemical variables that were not expected to be affected by metal amendments (i.e., all measures of Fe and Mn). A significant time × preparation effect would indicate that the temporal dynamics of metal partitioning differed between freshly spiked and aged sediments. Overlying water Cu concentrations during the equilibration period were analyzed with analysis of covariance (ANCOVA) to test for a difference in Cu flux between sediment preparation treatments. Survival and growth of *H. azteca* were analyzed separately for each sediment treatment using a generalized linear model (survival) or nonlinear least-squares regression (growth). Estimates of LC50 for survival and EC10 for growth were calculated from model fits with corresponding 95% confidence intervals. EC50 estimates, which were calculated to facilitate direct comparison to LC50, can be found in Supporting Information (Table S2). Tissue metal concentrations for Ni and Zn experiments were analyzed with ANCOVA and two-way ANOVA, respectively, to determine if accumulation of metal was related to sediment metal and differed among sediment preparations (Ni reference sediments omitted due to non-linearity). All statistical tests were performed in R version 3.4.3 (R Core Team 2017). The assumptions of linear models were assessed with Shapiro-Wilks

test for normality and diagnostic plots of residuals. Statistical models that did not meet assumptions were normalized with log transformations of response variables and total metal covariate.

## Results

Sediment and water physicochemistry. In general, physical and chemical conditions during the experiments were similar between aged and freshly spiked treatments and between initial and final samples. Although some physicochemical parameters differed significantly between treatments and time points, the magnitude of those differences were relatively small (effect size <15%). However, the two distinct sediments used in the Cu and Ni experiments maintained their geochemical differences (Table 2), which led to distinct metal speciation and bioavailability between sediment types for the same metal. Here we provide a summary of the sediment and overlying water physicochemistry for each experiment, and more detailed information is provided in the Supporting Information (Appendix S1).

For the Cu experiment, a substantial amount of Cu was released into the water column during the 10-d equilibration period, and a greater amount of Cu was released from the freshly spiked sediment than aged sediment ( $p < 0.001$ ). The flux of Cu declined exponentially in both aged and freshly spiked treatments, which indicated rapid equilibration of Cu and much lower flux rates prior to *H. azteca* being added to the beakers (Supporting Information, Figure S1). Overlying water pH was lower in freshly spiked treatments with high total Cu concentrations ( $\text{Cu}_{\text{TOT}}$ ) immediately after metal addition, but overlying water pH had stabilized to  $7.8 \pm 0.5$  (mean  $\pm$  1 standard deviation)

at the end of equilibration period. Sediment pH showed similar patterns to overlying water with lower pH in freshly spiked sediments with high Cu concentrations; the differences in sediment pH were still evident when *H. azteca* were added to the beakers, but at the end of the experiment freshly spiked and aged sediments had similar pH (Supporting Information, Figure S2). Ocoee sediment AVS remained higher than Dow sediment AVS throughout the experimental period, and as expected, AVS was lower in Ocoee sediment with greater  $Cu_{TOT}$  (U.S. EPA 2005). Sediment Fe and Mn concentrations (total and oxidized) were relatively stable through time and between sediment preparation treatments (<15% difference). The two preparation treatments produced sediments that had statistically similar total Cu concentrations ( $p > 0.05$ ) but the partitioning of that Cu into porewater and particulate fractions differed between the aged and freshly spiked treatments. Bulk partitioning coefficients (i.e.,  $K_D$ ) indicated that Dow sediment had a greater proportion of Cu in the pore waters relative to solid-phases when compared to Ocoee (Supporting Information, Figure S3). For both sediments, freshly spiked treatments had lower  $K_D$  (i.e., greater proportion of Cu in pore waters) than aged treatments, but differences were small enough to be non-significant (all  $p > 0.07$ ). For partitioning to specific solid-phase fractions, we found a relatively small pool of Cu in the oxidized metal fraction (0.1–11%), but the Cu in that fraction showed the largest differences between freshly spiked and aged treatments. Overall, freshly spiked sediments had greater concentrations of Cu in acid extractable and reducible solid-phase fractions than aged sediments.

For the Ni experiments, we measured Ni flux into overlying water but there were no differences in flux rates between aged and freshly spiked treatments (Supporting

Information, Figure S4). Also, the frequent exchange of overlying water (every three hours) maintained filtered Ni concentrations below the toxic threshold for *H. azteca* (Besser et al. 2013). Sediment pH was circumneutral (6.8–7.4) for all treatments throughout the exposure period. We did detect some significant differences in sediment pH between preparation treatments, but the magnitude of differences was small such that it would likely not affect Ni partitioning or toxicity (Supporting Information, Figure S5). Similar to our observations of the Cu experiment, differences in AVS between Burntwood and Raisin sediments were maintained over the experiment (Table 2), but AVS was lower with increasing sediment total Ni ( $Ni_{TOT}$ ) in the naturally high-AVS Raisin sediment (U.S. EPA 2005). We observed minor differences in Fe and Mn concentrations through time and between depths and sediment preparations (i.e., freshly spiked vs. aged), but there were large differences between Raisin and Burntwood sediments (Table 2). Sediment  $Ni_{TOT}$  was stable through time and consistent across depths, but Burntwood freshly spiked treatments had on average 10% greater  $Ni_{TOT}$  than the aged sediment for the same nominal treatments. Bulk partitioning of Ni to porewater was greater in Burntwood sediment than Raisin sediment, and porewater Ni in Burntwood changed through time (Supporting Information, Figure S6). Interestingly, for both Cu and Ni experiments, the sediment characterized by low AVS concentrations (i.e., Burntwood and Dow) had more unstable  $K_D$  through time when compared to high AVS sediments (i.e., Ocoee and Raisin). In the solid phase, a majority of Ni was associated with Fe oxide minerals for both sediments. In Raisin sediment, Ni was primarily associated with amorphous minerals (34–74%), whereas in Burntwood sediment, Ni was

more evenly distributed between amorphous (24–54%) and crystalline (32–38%) Fe oxide minerals.

For the Zn experiments, we collected sediment from the same location used for the Ni experiments (River Raisin), but the sediment chemistry differed slightly between the two experiments, most notably in the speciation of Fe (Table 2). A positive relationship between AVS and sediment metals was observed for the Zn experiments, whereas the opposite relationship was observed for Cu and Ni sediments; this emphasizes that the metal amendment methods did not oxidize AVS and that ZnS has greater solubility than either CuS or NiS (U.S. EPA 2005). Sediment pH was circumneutral (7.3–7.9) in all treatments and did not change during the experiment. Similar to Cu and Ni experiments, there were occasional differences in Fe and Mn concentrations both through time and between aged and freshly spiked treatments, but the magnitude of differences was small (i.e., <10%). Sediment total Zn ( $Zn_{TOT}$ ) in the freshly spiked treatments was on average 10% greater than in the aged treatment for the same nominal concentrations, but the Zn concentrations were stable through time. Water overlying the freshly-spiked sediment had greater Zn concentrations than aged treatments, but partitioning to porewater Zn did not differ between sediment preparations ( $\log K_D = 4.7 \pm 0.7 \text{ L kg}^{-1}$ ). Among solid-phase fractions, we saw minimal differences between aged and freshly spiked sediment.  $Zn_{SEM}$  exceeded AVS (molar) at the two highest treatment concentrations of both sediment preparations (i.e.,  $Zn_{TOT} > 1400 \text{ mg kg}^{-1}$ ). Additionally, those highest  $Zn_{TOT}$  treatments were the only sediments with detectable Zn found associated with amorphous Fe oxides. This suggests that fractionation of Zn to

amorphous Fe oxides may account for a significant amount of  $Zn_{TOT}$  only when binding capacity of AVS has been titrated.

Sediment Cu toxicity. The *Hyalella azteca* survival and growth data indicated large differences in concentration–response between Ocoee and Dow sediments and small differences due to sediment preparation treatments. Both Dow and Ocoee freshly spiked sediments caused complete mortality at the highest  $Cu_{TOT}$  concentrations (780 and 2400  $mg\ kg^{-1}$ , respectively) (Figure 1). Conversely, aged Dow and Ocoee sediments at similar concentrations (740 and 2300  $mg\ kg^{-1}$ , respectively) reduced survival (<50%), but complete mortality was only observed in a single replicate (Dow 740  $mg\ kg^{-1}$ ). The increased mortality in freshly spiked sediments is reflected in slightly lower LC50 values when compared to aged sediments, but LC50s for freshly spiked and aged sediments were not significantly different as evidenced by overlapping 95% confidence intervals (CI) (Table 3).

Growth rates of *H. azteca* were also reduced in high-Cu sediments, but again differences between sediment preparation treatments were minimal (Figure 1). For Dow sediment, a 10% reduction in RGR relative to reference sediments occurred at an approximately 2-fold higher  $Cu_{TOT}$  concentration in aged sediments when compared to freshly spiked sediments (Table 3). Although the magnitude of difference in RGR between aged and freshly spiked Dow sediments was relatively large, the 95% confidence intervals around these estimates also were fairly large, and the EC10 values did not differ significantly between sediment preparations (Table 3). For Ocoee sediments, *H. azteca* growth rates were reduced at higher  $Cu_{TOT}$  concentrations but concentration–response curves were very similar between freshly spiked and aged



treatments (Figure 1). Mortality and growth thresholds for Ocoee sediments were higher than those calculated for Dow sediments, likely reflecting the lower porewater Cu concentrations in Ocoee sediments.

Threshold EC10 values calculated for each sediment  $\times$  preparation combination were positively correlated with bulk Cu partitioning between aqueous and solid phases (Figure 2A). Relating *H. azteca* growth rates to bioavailable Cu, which was measured as  $Cu_{SEM}$  in excess of AVS normalized to the organic carbon content (i.e.,  $(Cu_{SEM} - AVS)/f_{OC}$ ), minimized the concentration–response differences between Ocoee and Dow sediments and freshly spiked and aged preparations (Figure 3A). Although concentration–response relationships for Dow and Ocoee sediments were more similar when the Cu was expressed as the bioavailable fraction when compared to  $Cu_{TOT}$ , point estimates from the concentration–response curves remained significantly different from each other. When both aged and freshly spiked sediments were combined, Dow sediments had an EC10 of  $42 \mu\text{mol g}^{-1}$  (95% CI = 26–61  $\mu\text{mol g}^{-1}$ ) whereas Ocoee sediments had an EC10 of  $13 \mu\text{mol g}^{-1}$  (95% CI = 8–22  $\mu\text{mol g}^{-1}$ ). Interestingly, Ocoee sediments were toxic at lower concentrations than Dow sediments when Cu was expressed as the bioavailable fraction; this result is likely due to the greater proportion of Cu adsorbed to Fe oxides in Dow (3–12% of  $Cu_{TOT}$ ) than Ocoee sediments (2–3% of  $Cu_{TOT}$ ). Fe oxides are a solid-phase fraction that is not considered bioavailable, yet is not accounted for in standard bioavailability models. These results agree with the prevailing ecotoxicological models that assume porewater metals are the primary route of toxicity.

Sediment Ni toxicity. The *H. azteca* toxicity endpoints (survival, growth, and tissue concentrations) indicated differences in concentration–response between

Burntwood and Raisin sediments but minimal differences between sediment preparation techniques. Complete mortality was observed in a few replicates of the highest Ni<sub>TOT</sub> treatment of both aged and freshly spiked Raisin sediment (i.e., 3000 mg kg<sup>-1</sup>) but no complete mortality was observed in Burntwood sediment (Figure 4). There was minimal mortality in the aged Burntwood sediments, and thus concentration–response thresholds (i.e., LC50) could not be calculated (Table 3). Raisin aged and freshly spiked sediments with high Ni<sub>TOT</sub> caused mortality (Figure 4), but LC50s were not statistically different between sediment preparations (Table 3). Growth rates of *H. azteca* were lower on sediments with greater Ni<sub>TOT</sub>, and differences were observed between Raisin and Burntwood sediments (Figure 4). EC10 values were similar between aged and freshly spiked sediments for both Burntwood (951 and 1010 mg kg<sup>-1</sup>, respectively) and Raisin sediments (242 and 499 mg kg<sup>-1</sup>, respectively). Although concentration–response thresholds for Burntwood were 2–4-fold higher than thresholds for Raisin sediments, the 95% confidence intervals around these estimates were large and often overlapped (Table 3). Exposure of *H. azteca* to both Burntwood and Raisin sediments resulted in Ni accumulation in their tissues, and there was a positive relationship (log–log) between sediment and tissue Ni concentrations (Supporting Information, Figure S7). There were no differences in tissue Ni concentrations between *H. azteca* exposed to either aged or freshly spiked preparations of either sediment ( $p > 0.70$ ), but *H. azteca* exposed to Raisin sediments had a slightly greater accumulation of Ni relative to Ni<sub>TOT</sub> (i.e., greater slope) than those exposed to Burntwood sediments (Supporting Information, Figure S7).

In contrast to Cu exposures, threshold EC10 values calculated for each sediment × preparation combination were not correlated to the bulk partitioning of Ni between pore

water and sediment (Figure 2B). The lack of relationship between bulk partitioning and Ni EC10 is likely due to the spatial disconnect between organism exposure at the surface and pore water measurement in anoxic sediment. Bulk partitioning in surface sediments where sorption to Fe oxides would dominate over coprecipitation with AVS may yield better correlation to Ni EC10. Within each sediment type, bioavailable Ni (i.e.,  $(\text{Ni}_{\text{SEM-AVS}})/f_{\text{OC}}$ ) was related to the reduction in *H. azteca* growth rates and we again observed no differences between sediment preparations (Figure 3B). Combining aged and freshly spiked sediments, we estimated EC10s for RGR of  $160 \mu\text{mol g}^{-1}$  (95% CI = 111–423  $\mu\text{mol g}^{-1}$ ) for Burntwood and  $3 \mu\text{mol g}^{-1}$  (95% CI = 1–7  $\mu\text{mol g}^{-1}$ ) for Raisin sediments. Expressing Ni exposure as only the bioavailable fraction is expected to reduce the variability in toxicological thresholds (U.S. EPA 2005), but we observed greater differences (50-fold) when expressing dose as bioavailable Ni when compared to  $\text{Ni}_{\text{TOT}}$ . Furthermore, Raisin sediment (relatively high AVS and low Fe) was toxic at lower bioavailable Ni concentrations than Burntwood sediment (relatively low AVS and high Fe), which emphasizes that Ni adsorption to Fe oxide minerals likely reduces Ni exposure but is not accounted for in current bioavailability models (Costello et al. 2011, 2016, Mendonca et al. 2017).

Sediment Zn toxicity. We observed minimal mortality of *H. azteca* and only minor responses in growth and tissue Zn as a result of exposure to aged or freshly spiked sediment. Limited mortality of *H. azteca*, even when exposed to the greatest  $\text{Zn}_{\text{TOT}}$  (i.e., 2600 and 3300  $\text{mg kg}^{-1}$ ), did not allow for calculation of LC50s (Table 3). Growth of *H. azteca* was reduced ~5% in the greatest  $\text{Zn}_{\text{TOT}}$  treatments relative to controls for both the aged and freshly spiked treatments (Figure 5). However, a concentration–response

threshold could be calculated for aged sediments only, but the estimate was above the highest  $Zn_{TOT}$  treatment ( $3300 \text{ mg kg}^{-1}$ ) with an undefined positive confidence limit (Table 3) and we urge caution in interpreting this threshold. Tissue Zn concentrations exhibited similar patterns to growth, as the *H. azteca* exposed to the highest  $Zn_{TOT}$  treatments were the only amphipods with tissue Zn above controls and there were no differences among sediment preparation methods (Supporting Information, Figure S8). The *H. azteca* exposed to the  $>2500 \text{ mg kg}^{-1} Zn_{TOT}$  had Zn tissue concentrations 2-fold greater than other treatments. Expressing the dose of sediment Zn as bioavailable Zn (i.e.,  $(Zn_{SEM}-AVS)/f_{OC}$ ) showed similarities between aged and freshly spiked sediments, but no concentration–response thresholds could be calculated. The relatively high AVS concentrations in both aged and freshly spiked Raisin sediments resulted in AVS exceeding  $Zn_{SEM}$  in 4 of the 6 metal amended treatments and  $(Zn_{SEM}-AVS)/f_{OC} < 90 \mu\text{mol g}^{-1}$  in all treatments, which is below the threshold at which toxicity may occur (i.e.,  $120 \mu\text{mol g}^{-1}$ , Burton et al. 2005, Di Toro et al. 2005).

Comparison to intact sediment. The Ni and Cu experiments can be compared to mesocosm experiments that used the same sediments but tested *H. azteca* growth response to intact sediment (Costello et al. 2015, 2016). Caution must be used for these comparisons because of the differences in experimental conditions (e.g., duration of toxicity assay, *H. azteca* access to sediment, overlying water exchange rate), but comparisons between sediments within experiments do reveal some important patterns. First, sediments retained within the mesocosm and tested without manipulation (i.e., intact un-homogenized sediment) showed increases in EC10 during aging, which indicates that these sediments became less toxic as they aged (Table 4). However, the

magnitude of change in EC10s differed among sediments, as Dow (Cu tests) and Raisin (Ni tests) EC10s changed more than for Ocoee and Burntwood, respectively. This difference in the magnitude of change in EC10s during aging established a switch in relative toxicity where the sediments with lower concentration–response thresholds early in the aging process (EC10s: Dow < Ocoee and Raisin < Burntwood) had greater concentration–response thresholds late in the aging period (EC10s: Dow > Ocoee and Raisin > Burntwood).

As mentioned above, we observed only minor differences between freshly spiked and aged sediment preparations for the homogenized sediment. Comparing between homogenized sediment tests for Cu and Ni, we found, regardless of sediment preparation, EC10s were lower for Dow than Ocoee and lower for Raisin than Burntwood. The relative toxicity of these homogenized sediments was more similar to intact sediments tested in the mesocosm early during the aging process rather than late (Table 4).

Although the aged treatments for the experiments described here were taken immediately after the late sampling period within the mesocosm experiment, the toxicity results more closely resemble those observed early in the mesocosm experiment, thus suggesting a ‘reset’ of sediment toxicity following homogenization. This comparison provides additional support for our hypothesis that declines in toxicity observed during sediment aging are not due to slow equilibration of metals, but rather are related to changes in vertical heterogeneity of sediment due to oxidation of surface sediment.

## Discussion

Our data demonstrate that slow equilibration is not the mechanism by which aging of sediment can alter toxicity, but rather oxidation of surface sediment, and the persistence of that oxidized surface layer (and associated metal oxide ligands), is the likely explanation for altered toxicity during sediment aging. The most striking differences in physicochemistry between sediment preparations were in the Cu experiments, where freshly spiked sediments had greater pore and overlying water Cu and a larger fraction of Cu partitioned to Fe oxides than aged sediments. Despite those large physicochemical differences, the toxicological response to those sediment treatments were very similar. For the Ni and Zn experiments, there were minimal differences between freshly spiked and aged sediments for porewater metals (with the exception of Ni in Burntwood sediment), metal flux into overlying water, and metal partitioning to solid-phase ligands. These physicochemical trends were reflected in similar concentration–response thresholds. Differences between sediment preparations observed in the Cu experiments can be linked to instability of sediment physicochemistry in the freshly spiked treatments. Following metal amendments, sediment pH was lower as a result of  $\text{CuCl}_2$  hydrolysis, and pH declines were greatest in high Cu sediments. Although sediment pH eventually returned to reference conditions by the end of the exposure period, AVS concentrations in freshly spiked treatments were lower than aged sediments; this likely resulted from dissolution of reduced sulfur compounds under low pH conditions, rapid titration by added Cu, and oxidation. For Ni and Zn tests, the buffering with NaOH (which was not used in the Cu test) titrated the hydrolysis products that are generated during metal addition, and smaller differences in sediment pH were

observed between our sediment preparation treatments (no sediment with pH <6). Consequently, AVS was either stable (Zn experiments) or only slightly lower in freshly spiked than aged treatments (Raisin sediment in Ni experiments). These data add to the evidence that regulation of pH during metal amendment is of critical importance for producing sediments with stable geochemistry that mirror field-contaminated ecosystems (Hutchins et al. 2009, Brumbaugh et al. 2013).

Our sediment preparation methods for freshly spiked treatments were selected to best match the methods used for studies that generated high-quality *H. azteca* toxicity data (Nguyen et al. 2005, Roman et al. 2007, Besser et al. 2013). However, that approach resulted in important differences in the handling of our freshly spiked treatments among experiments for different metals. Most importantly, the Cu test did not use the state-of-the-art, metal amendment approaches (i.e., direct metal addition, no pH correction, short equilibration time), whereas Ni and Zn tests used more sophisticated metal amendment methods with stabilized pH, and those approaches matched sediment treatments for the mesocosm experiments (Costello et al. 2015, 2016). Consequently, we are confident our sediment preparation techniques used in the Ni and Zn experiments allowed us to isolate the effects of slow equilibration processes, but the Cu experiment compared two manipulations simultaneously: slow equilibration and pH buffering during spiking. In both the Ni and Zn beaker tests, sediment geochemistry differed minimally and there were no differences in *H. azteca* concentration–response thresholds between aged and freshly spiked sediments. This stands in contrast to results from the complementary mesocosm experiments, which showed large differences in sediment geochemistry and concentration–response thresholds during aging (Costello et al. 2016). This suggests that

surface sediment oxidation in the mesocosm changed metal bioavailability and toxicity, but these effects did not carry over into beaker experiments, likely due to homogenization of sediment prior to these tests. Equilibrating porewater pH during amendment is critical for environmentally realistic metal partitioning (Hutchins et al. 2009, Brumbaugh et al. 2013), but the Cu test, which did not include pH equilibration, still demonstrated minimal differences in toxicological response between aged and freshly spiked sediments. Therefore, slow equilibration of sediments (for at least 100 days) prior to standard 28-d sediment toxicity tests elicits minimal effects on sediment geochemistry and concentration–response relationships.

In the mesocosm experiments, toxicity and geochemistry were monitored throughout the aging process (Costello et al. 2015, 2016), and 3 of 4 study sediments had shifts in metal speciation and declines in toxicity. By the end of the aging period, one sediment from each of the Cu and Ni tests (Dow and Raisin, respectively) were non-toxic at the concentrations tested. When these aged sediments were homogenized and transitioned to beakers, there were significant increases in toxicity. The mesocosm-to-beaker comparison demonstrated a return of toxicity at low total metal concentrations for Dow (Cu test) and Raisin (Ni test) sediments, which were absent at the end of the mesocosm experiments. The Cu and Ni experiments indicated that the sediment toxicity reverted to what we observed early in the mesocosm aging experiments: toxicity at lower Cu concentrations in Dow when compared to Ocoee sediments, and Raisin sediments that were toxic at lower Ni concentrations than Burntwood sediment. This suggests that homogenization of sediment is critically important and may be responsible for the minimal effects of our sediment preparation methods on metal partitioning and toxicity.



In the mesocosm experiments, AVS declined in surface sediments due to diffusion and advection of O<sub>2</sub>, and Cu and Ni partitioning shifted from being dominated by reduced sulfur to Fe oxides (Costello et al. 2015, 2016). While these high AVS sediments were equilibrating to the flow-through conditions, they were toxic to *H. azteca* at lower total Ni concentrations. The sediment chemistry in the beaker experiments described here was frequently similar across depths, which indicates that homogenization eliminated any vertical gradients. With the exception of Ocoee sediment in the Cu test (i.e., lower AVS in surface), the 28-d experimental period was not sufficiently long to reestablish vertical gradients in physicochemistry. The transition between oxic and anoxic sediment, and the associated changes in redox-sensitive ligand concentrations, occurs over short distances (<8 mm; Costello et al. 2015, 2016), and homogenization of the 7 cm deep sediment column diluted any oxidized ligands and associated metals, and redistributed reduced ligands to the surface. This result adds to the evidence that physical disturbance of sediment can alter metal speciation and toxicity (Simpson et al. 1998, Fetters et al. 2016), and in this experiment it reversed the changes observed during sediment aging. Thus, these data indicate that the addition of an aging step prior to standard sediment toxicity assays (i.e., includes homogenization) would not create experimental conditions that more closely represent field conditions. However, intact core experiments or aging without homogenization, which are designed to maintain vertical gradients in redox and metal partitioning, would likely be needed to best represent contaminated sediments encountered by organisms in the field.

Although there were minimal differences between aged and freshly spiked treatments related to aging, we did observe variation in concentration–response and metal

partitioning when comparing among our different sediment types. This variation in concentration–response across sediment types (i.e., EC10s) and metals can prove to be an important test of different sediment quality guidelines that are currently in use. Observed EC10 concentrations were all above the consensus-based threshold effect concentrations (TECs, MacDonald et al. 2000) for all but Dow freshly spiked sediment, which is expected because TECs are conservative estimates of toxicity based on total metal concentrations. Although EC10s almost always exceeded TECs, the magnitude of difference among metals differed; EC10:TEC ratios for Cu were 0.9–7.5 (Cu TEC = 31.6 mg kg<sup>-1</sup>), but Ni and Zn EC10:TEC were always >10 (22.7 and 121 mg kg<sup>-1</sup> for Ni and Zn, respectively). The low EC10:TEC for Cu experiments are likely due to the short equilibration times and high bioavailability of Cu in these experiments. Accounting for bioavailability with SEM/AVS approaches (U.S. EPA 2005) reduced variability in concentration–response thresholds for Cu only. Differences in concentration–response between Raisin and Burntwood sediments amended with Ni were not minimized by accounting for AVS and OC or bulk partitioning to pore water in anoxic sediment, and this supports the observation from other studies that Fe oxides in surface sediment are a more important ligand for Ni binding (Costello et al. 2011, 2016, Besser et al. 2013, Mendonca et al. 2017). Regardless of sediment type, preparation technique, or metal, no laboratory-prepared sediment with AVS in excess of SEM caused toxicity, which has been demonstrated repeatedly in laboratory and field studies (Di Toro et al. 1992, Burton et al. 2005, U.S. EPA 2005, Costello et al. 2011).

Although traditional standardized conditions used for sediment preparation and toxicity testing (e.g., Hutchins et al. 2009, Brumbaugh et al. 2013) are crucial for accurate

concentration–response thresholds, we argue that laboratory-prepared sediments that are homogenized do not account for surface sediment oxidation and may still be conservative with respect to intact aged sediments (Costello et al. 2015, 2016). Furthermore, field-collected contaminated sediments may also have vertical gradients (e.g., recently settled surface particles) that are disturbed during homogenization prior to laboratory testing. Collection, storage, and homogenization of sediment may overlook or minimize the role of bioturbating organisms in modifying gradients in redox (Remaili et al. 2016). We advocate for additional research comparing metal partitioning and toxicity between intact cores and homogenized sediments to improve our understanding of the relationships between slow oxidation of surface sediment, ambient physicochemical gradients in sediment, and resulting toxicity to epibenthic organisms. Overall, our findings suggest that laboratory-derived, sediment quality guidelines currently in use are useful as conservative benchmarks for initial screening of sediments. Nevertheless, without an understanding of the role of sediment oxidation and the stability of metal-binding ligands under oxidizing conditions influence metal partitioning and toxicity, traditional sediment toxicity methods are not well suited to predict *in situ* toxicity. This study adds to the growing body of literature demonstrating the importance of oxidized ligands, sediment disturbance, and non-equilibrium dynamics in surface sediments (Simpson et al. 1998, 2012, Atkinson et al. 2007, Costello et al. 2011, 2015, Fetters et al. 2016, Mendonca et al. 2017), and emphasizes the need for well-designed experiments for improved accuracy in ecological risk assessment.

## References

- Adams, S. M. 2005. Assessing cause and effect of multiple stressors on marine systems. *Marine Pollution Bulletin* 51:649–657.
- Allen, H. E., G. Fu, and B. Deng. 1993. Analysis of acid-volatile sulfide (AVS) and simultaneously extracted metals (SEM) for the estimation of potential toxicity in aquatic sediments. *Environmental Toxicology and Chemistry* 12:1441–1453.
- ASTM. 2010. Standard Test Method for Measuring the Toxicity of Sediment-Associated Contaminants with Freshwater Invertebrates. American Society for Testing and Materials, Philadelphia, PA.
- Atkinson, C. A., D. F. Jolley, and S. L. Simpson. 2007. Effect of overlying water pH, dissolved oxygen, salinity and sediment disturbances on metal release and sequestration from metal contaminated marine sediments. *Chemosphere* 69:1428–1437.
- Besser, J. M., W. G. Brumbaugh, C. G. Ingersoll, C. D. Ivey, J. L. Kunz, N. E. Kemble, C. E. Schlekat, and E. R. Garman. 2013. Chronic toxicity of nickel-spiked freshwater sediments: Variation in toxicity among eight invertebrate taxa and eight sediments. *Environmental Toxicology and Chemistry* 32:2495–2506.
- Besser, J. M., W. G. Brumbaugh, C. D. Ivey, C. G. Ingersoll, and P. W. Moran. 2008. Biological and chemical characterization of metal bioavailability in sediments from Lake Roosevelt, Columbia River, Washington, USA. *Archives of Environmental Contamination and Toxicology* 54:557–570.

- Brumbaugh, W. G., J. M. Besser, C. G. Ingersoll, T. W. May, C. D. Ivey, C. E. Schlekat, and E. Rogevich Garman. 2013. Preparation and characterization of nickel-spiked freshwater sediments for toxicity tests: Toward more environmentally realistic nickel partitioning. *Environmental Toxicology and Chemistry* 32:2482–2494.
- Brumbaugh, W. G., T. May, J. M. Besser, A. Allert, and C. Schmitt. 2007. Assessment of elemental concentrations in streams of the New Lead Belt in southeastern Missouri 2002-05. Pages 1–57 U.S. Geological Survey Scientific Investigations Report 2007–5057. Reston, VA.
- Burton, G. A., P. M. Chapman, and E. P. Smith. 2002. Weight-of-evidence approaches for assessing ecosystem impairment. *Human and Ecological Risk Assessment* 8:1657–1673.
- Burton, G. A., L. T. H. Nguyen, C. Janssen, R. Baudo, R. A. McWilliam, B. Bossuyt, M. Beltrami, and A. Green. 2005. Field validation of sediment zinc toxicity. *Environmental Toxicology and Chemistry* 24:541–553.
- Campbell, P. G. C., and A. Tessier. 1996. Ecotoxicology of metals in the aquatic environment: Geochemical aspects. Pages 11–58 *in* M. C. Newman and C. H. Jagoe, editors. *Ecotoxicology: A hierarchical treatment*. CRC Press, New York, NY, USA.
- Chapman, P., and F. Wang. 1998. Ecotoxicology of metals in aquatic sediments: binding and release, bioavailability, risk assessment, and remediation. *Canadian Journal of Fisheries and Aquatic Sciences* 55:2221–2243.
- Costello, D. M., G. A. Burton, C. R. Hammerschmidt, E. C. Rogevich, and C. E.

- Schlekat. 2011. Nickel phase partitioning and toxicity in field-deployed sediments. *Environmental Science & Technology* 45:5798–5805.
- Costello, D. M., C. R. Hammerschmidt, and G. A. Burton. 2015. Copper sediment toxicity and partitioning during oxidation in a flow-through flume. *Environmental Science & Technology* 49:6926–6933.
- Costello, D. M., C. R. Hammerschmidt, and G. A. Burton. 2016. Nickel partitioning and toxicity in sediment during aging: Variation in toxicity related to stability of metal partitioning. *Environmental Science & Technology* 50:11337–11345.
- Dale, V. H., G. R. Biddinger, M. C. Newman, J. T. Oris, G. W. Suter, T. Thompson, T. M. Armitage, J. L. Meyer, R. M. Allen-King, G. A. Burton Jr., P. M. Chapman, L. L. Conquest, I. J. Fernandez, W. G. Landis, L. L. Master, W. J. Mitsch, T. C. Mueller, C. F. Rabeni, A. D. Rodewald, J. G. Sanders, and I. L. van Heerden. 2008. Enhancing the ecological risk assessment process. *Integrated Environmental Assessment and Management* 4:306–313.
- Fetters, K. J., D. M. Costello, C. R. Hammerschmidt, and G. A. Burton. 2016. Toxicological effects of short-term resuspension of metal-contaminated freshwater and marine sediments. *Environmental Toxicology and Chemistry* 35:676–686.
- Hong, Y. S., K. A. Kinney, and D. D. Reible. 2011. Effects of cyclic changes in pH and salinity on metals release from sediments. *Environmental Toxicology and Chemistry* 30:1775–1784.
- Hutchins, C. M., P. R. Teasdale, S. Y. Lee, and S. L. Simpson. 2008. Cu and Zn

concentration gradients created by dilution of pH neutral metal-spiked marine sediment: A comparison of sediment geochemistry with direct methods of metal addition. *Environmental Science & Technology* 42:2912–2918.

Hutchins, C., P. Teasdale, S. Lee, and S. Simpson. 2009. The effect of sediment type and pH-adjustment on the porewater chemistry of copper- and zinc-spiked sediments. *Soil & Sediment Contamination* 18:55–73.

De Jonge, M., J. Teuchies, P. Meire, R. Blust, and L. Bervoets. 2012. The impact of increased oxygen conditions on metal-contaminated sediments part I: effects on redox status, sediment geochemistry and metal bioavailability. *Water Research* 46:2205–2214.

Kostka, J., and G. Luther. 1994. Partitioning and speciation of solid phase iron in saltmarsh sediments. *Geochimica et Cosmochimica Acta* 58:1701–1710.

MacDonald, D. D., C. G. Ingersoll, and T. A. Berger. 2000. Development and evaluation of consensus-based sediment quality guidelines for freshwater ecosystems. *Archives of Environmental Contamination and Toxicology* 39:20–31.

Mendonca, R. M., J. M. Daley, M. L. Hudson, C. E. Schlekot, G. A. Burton, and D. M. Costello. 2017. Metal oxides in surface sediment control nickel bioavailability to benthic macroinvertebrates. *Environmental Science and Technology* 51:13407–13416.

Neumann, P. T. M., U. Borgmann, and W. Norwood. 1999. Effect of gut clearance on metal body concentrations in *Hyalella azteca*. *Environmental Toxicology and*

Chemistry 18:976–984.

Nguyen, L. T. H., Y. E. Roman, M. B. Vandegehuchte, and C. R. Janssen. 2005.

Ecotoxicity of zinc to the amphipod *Hyalella azteca* tested in natural freshwater sediment: Final report submitted to International Lead Zinc Research Organisation. Ghent, Belgium.

Norwood, W. P., U. Borgmann, and D. G. Dixon. 2006. Saturation models of arsenic, cobalt, chromium and manganese bioaccumulation in *Hyalella azteca*.

*Environmental Pollution* 143:519–528.

R Core Team. 2017. R: A language and environment for statistical computing. R

Foundation for Statistical Computing, Vienna, Austria.

Remali, T. M., S. L. Simpson, E. D. Amato, D. A. Spadaro, C. V. Jarolimek, and D. F.

Jolley. 2016. The impact of sediment bioturbation by secondary organisms on metal bioavailability, bioaccumulation and toxicity to target organisms in benthic bioassays: Implications for sediment quality assessment. *Environmental Pollution* 208:590–599.

Remali, T. M., S. L. Simpson, and D. F. Jolley. 2017. Effects of enhanced bioturbation intensities on the toxicity assessment of legacy-contaminated sediments.

*Environmental Pollution* 226:335–345.

Roman, Y. E., K. A. C. De Schamphelaere, L. T. H. Nguyen, and C. R. Janssen. 2007.

Chronic toxicity of copper to five benthic invertebrates in laboratory-formulated sediment: Sensitivity comparison and preliminary risk assessment. *Science of the*



Total Environment 387:128–140.

Simpson, S. L., B. M. Angel, and D. F. Jolley. 2004. Metal equilibration in laboratory-contaminated (spiked) sediments used for the development of whole-sediment toxicity tests. *Chemosphere* 54:597–609.

Simpson, S. L., S. C. Apte, and G. E. Batley. 1998. Effect of short-term resuspension events on trace metal speciation in polluted anoxic sediments. *Environmental Science & Technology* 32:620–625.

Simpson, S. L., D. Ward, D. Strom, and D. F. Jolley. 2012. Oxidation of acid-volatile sulfide in surface sediments increases the release and toxicity of copper to the benthic amphipod *Melita plumulosa*. *Chemosphere* 88:953–961.

Di Toro, D. M., J. D. Mahony, D. J. Hansen, K. J. Scott, A. R. Carlson, and G. T. Ankley. 1992. Acid volatile sulfide predicts the acute toxicity of cadmium and nickel in sediments. *Environmental Science & Technology* 26:96–101.

Di Toro, D. M., J. A. McGrath, D. J. Hansen, W. J. Berry, P. R. Paquin, R. Mathew, K. B. Wu, and R. C. Santore. 2005. Predicting sediment metal toxicity using a sediment biotic ligand model: Methodology and initial application. *Environmental Toxicology and Chemistry* 24:2410–2427.

U.S. EPA. 2000. Methods for measuring the toxicity and bioaccumulation of sediment-associated contaminants with freshwater invertebrates. Washington DC, USA.

U.S. EPA. 2005. Procedures for the derivation of equilibrium partitioning sediment

benchmarks (ESBs) for the protection of benthic organisms: metal mixtures (cadmium, copper, lead, nickel, silver and zinc). US Environmental Protection Agency, Washington, DC.

Welton, J. S., and R. T. Clarke. 1980. Laboratory studies on the reproduction and growth of the amphipod, *Gammarus pulex* (L.). *Journal Of Animal Ecology* 49:581–592.

**FIGURE CAPTIONS**

Figure 1. Survival and relative growth rates of *Hyaletta azteca* that were exposed to Cu contaminated sediments that were either freshly spiked or aged. Mortality of all organisms was observed for the highest sediment Cu treatments of the freshly spiked treatments, and growth rates could not be calculated for these treatments. Lines indicate best-fits from GLM (survival) or non-linear least-squares regression using a logistic function (growth). Solid symbols and horizontal lines indicate the EC10 and 95% confidence interval for each exposure scenario. Note the log scale on the x-axis and the different scales, as Dow sediment had lower Cu concentrations than Ocoee sediment.

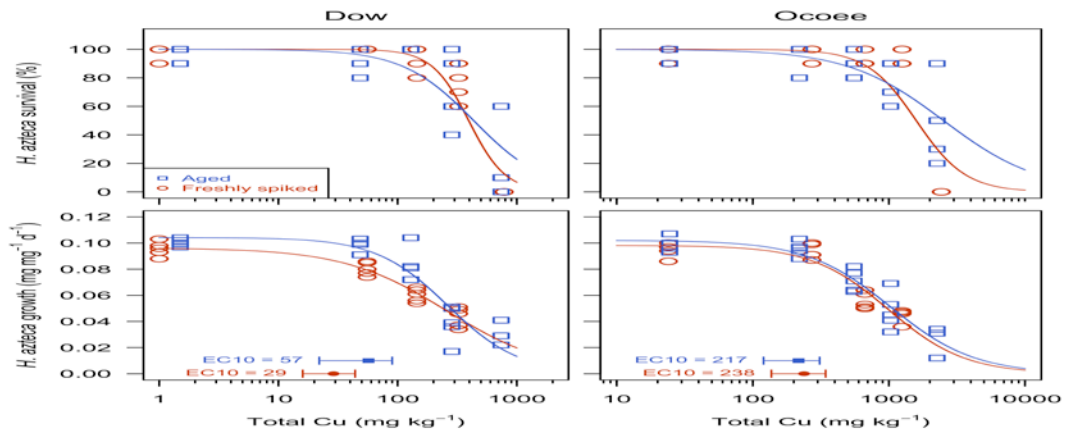


Figure 2. The partitioning of Cu (A) and Ni (B) between solid and aqueous phases (i.e.,  $K_D$ ) and toxicological thresholds for sediments amended with metals with two different techniques. Sediments with more Cu in pore waters relative to solid-phase Cu concentrations (i.e., low  $\log K_D$ ) caused reduced *H. azteca* growth at lower  $Cu_{TOT}$  concentrations ( $p = 0.007$ ). There was no relationship between Ni partitioning to pore waters and concentration–response thresholds ( $p = 0.16$ ). Vertical lines are 95% confidence intervals and the dashed line is a best-fit line from linear regression (equation given on plot). Note the different scales on the x- and y-axes.

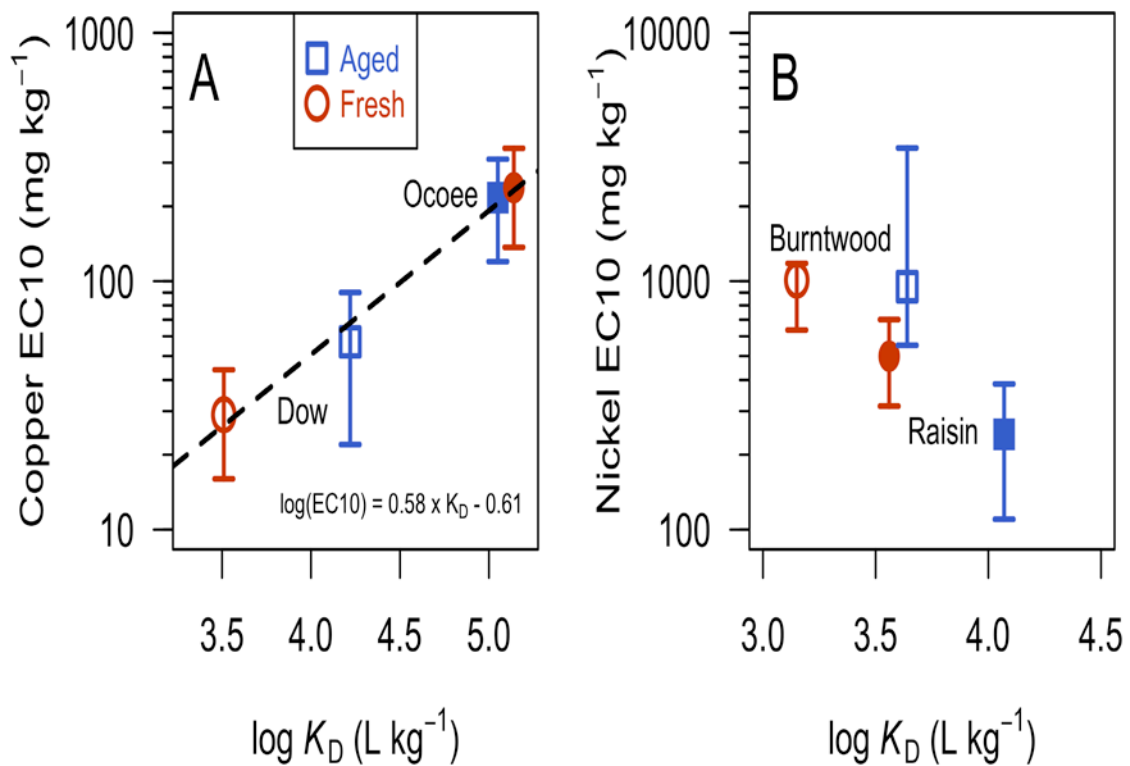


Figure 3. *Hyalella azteca* growth response relative to the concentration of Cu (A) or Ni (B) in excess of AVS, normalized by sediment organic carbon concentration. Response in RGR did not differ between sediment preparation (i.e., freshly spiked or aged) of the same sediment type and best-fit lines show concentration–response relationships for different sediment types only. Because the x-axis is on a log scale, negative values of  $(SEM-AVS)/f_{OC}$  (i.e., AVS in excess of SEM) are placed at 1.

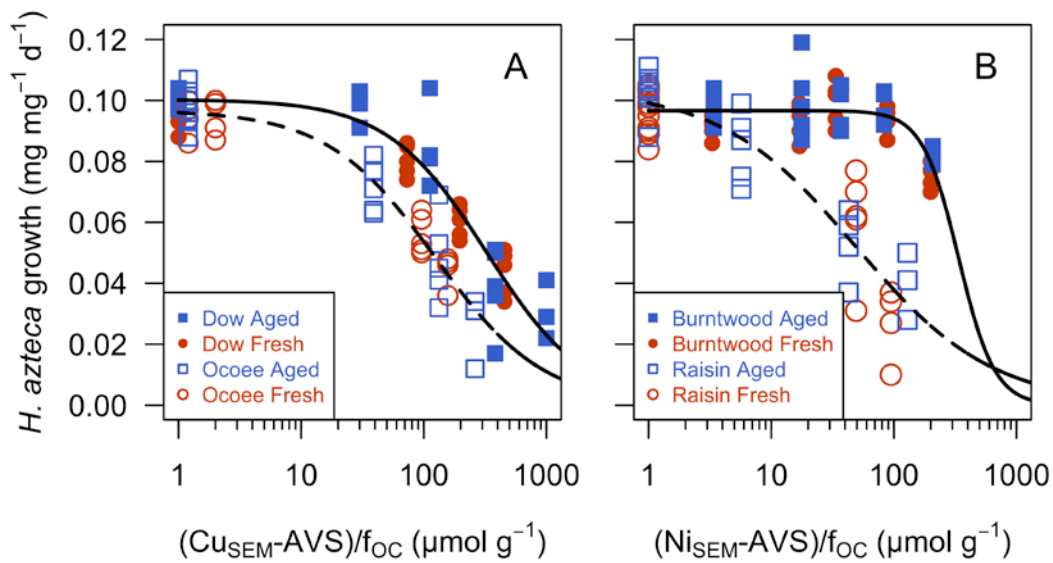


Figure 4. Survival and relative growth rates of *Hyalella azteca* that were exposed to Ni contaminated sediments that were either freshly spiked or aged. Lines indicate best-fits from GLM (survival) or non-linear least-squares regression using a logistic function (growth). Solid symbols and horizontal lines indicate the EC10 and 95% confidence interval for each exposure scenario. Note the log scale on the x-axis and the different scales, as Burntwood sediment had lower Ni concentrations than Raisin sediment.

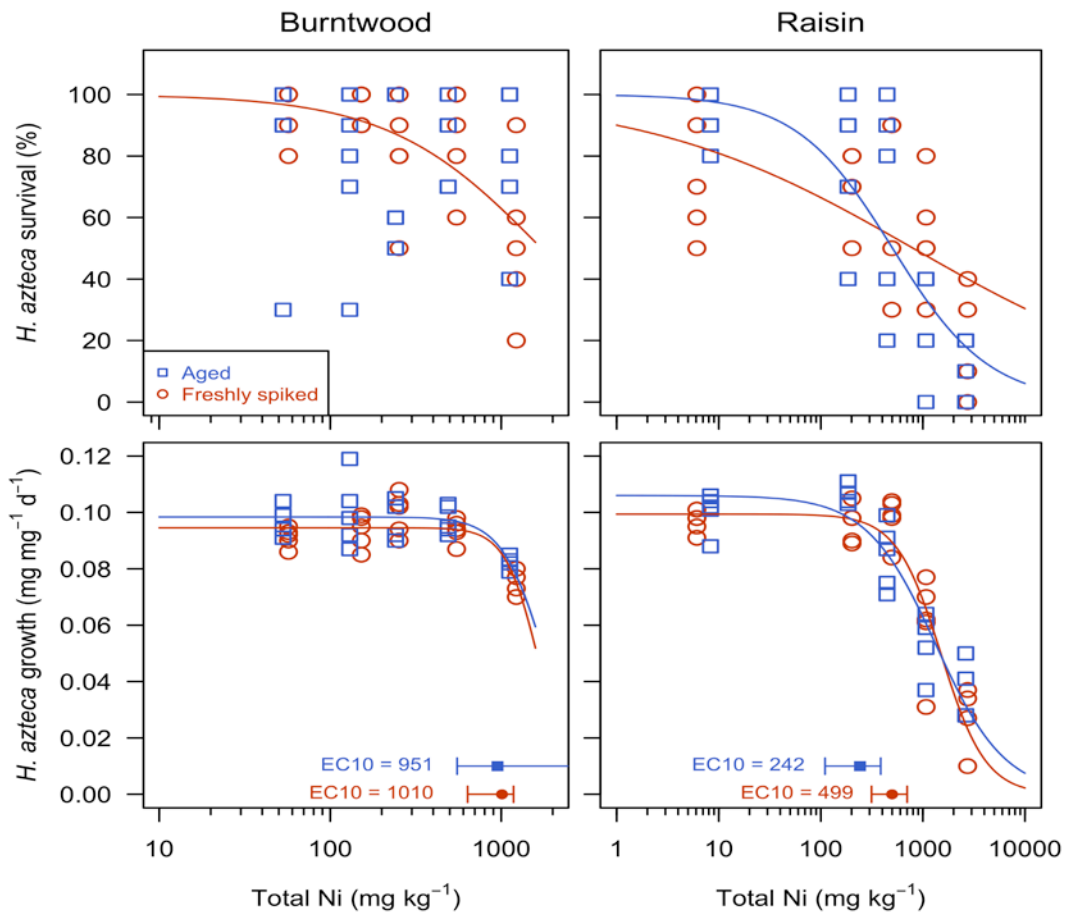


Figure 5. Survival and relative growth rates of *Hyalella azteca* exposed to Zn amended Raisin sediments that were either freshly spiked or aged. The line indicates best-fit from non-linear least-squares regression using a logistic function. All survival data and the RGR response on aged sediment produced non-significant concentration–response relationships. The solid symbol and horizontal line indicate the EC10 and 95% confidence interval for the freshly spiked sediments. Note the log scale on the x-axes.

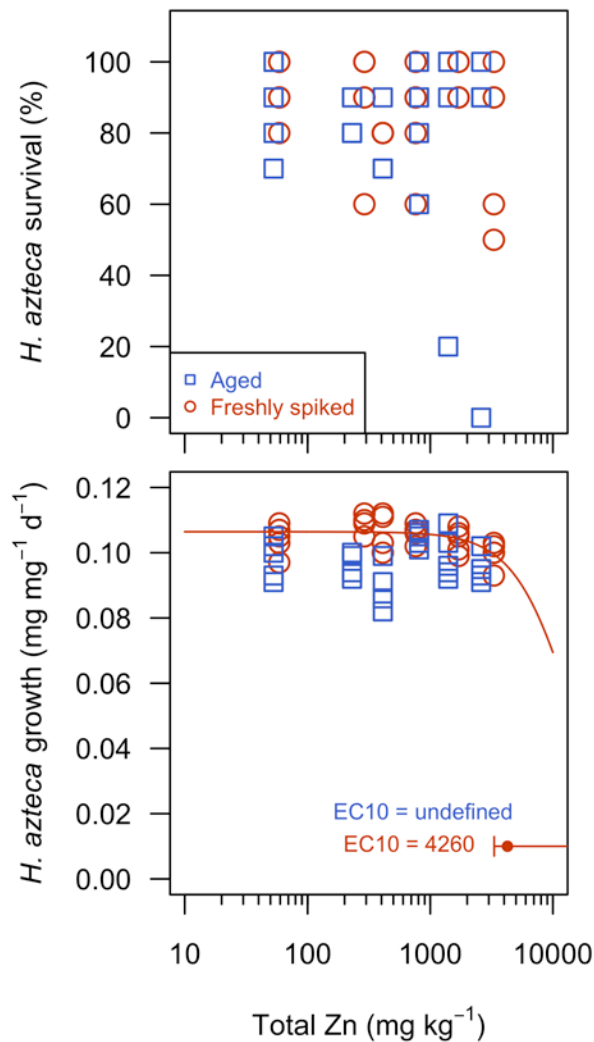


Table 1. Experimental conditions for *Hyalella azteca* toxicity assays. Experimental conditions varied between the experiments for copper, nickel, and zinc, but reflected the most robust recent 28-d *H. azteca* bioassays with growth as the main endpoint of toxicity.

	<b>Cu</b>	<b>Ni</b>	<b>Zn</b>
Spike equilibration	1 h	8 weeks	40 d
Metal addition	Direct	Indirect	Indirect
pH correction?	N	Y	Y
Aging time (d)	221	108	106
Test chamber	10	7	1
equilibration (d)			
Sed:H <sub>2</sub> O (mL:mL)	75:175	100:150	100:150
Water renewal	3x weekly	8x daily	2x daily
Feeding	1 mL YCT:algae	0.6 mL YCT	1 mL YCT
(mL d <sup>-1</sup> beaker <sup>-1</sup> )	(3:2)		
Nominal metal	Dow: 60–800	Burntwood: 100–1300	Raisin: 150–2100
conc. (mg kg <sup>-1</sup> ) <sup>a</sup>	Ocoee: 160–2100	Raisin: 200–2800	
Source	Roman et al. 2007	Besser et al. 2013	Nguyen et al. 2005

<sup>a</sup> Reference sediment with no added metal were included in all bioassays but concentrations are not reported here; YCT = Yeast, cerophyl, trout chow food

Table 2. Key physicochemical characteristics of sediments amended with metal and used in 28-d *Hyalella azteca* bioassays.

Sediment	Meta l	AVS ( $\mu\text{mol g}^{-1}$ )	O M (%)	Total carbo n (%)	Total Fe ( $\text{mg kg}^{-1}$ )	Fe <sub>HF</sub> ( $\text{mg kg}^{-1}$ )	Fe <sub>CF</sub> ( $\text{mg kg}^{-1}$ )	Total Cu ( $\text{mg kg}^{-1}$ )	Total Ni ( $\text{mg kg}^{-1}$ )	Total Zn ( $\text{mg kg}^{-1}$ )
Dow	Cu	0.54	2.1	0.68	6800	840	1700	6	na	na
					4600		1700			
Ocoee	Cu	6.3	8.2	3.6	0	1700	0	27	na	160
Burntwood					4200		1300			
d	Ni	0.01	10	4.2	0	3200	0	35	55	65
					1300					
Raisin	Ni	9.1	15	8.2	0	5300	2900	na	5	na
Raisin	Zn	9.2	19	na	1500	1000	7100	na	na	56



AVS = acid volatile sulfide; OM = organic matter; Fe<sub>HFO</sub> = amorphous Fe oxides; Fe<sub>CFO</sub> = crystalline Fe oxides; na = data not available

Table 3. Total metal (mg kg<sup>-1</sup>) toxicological thresholds ( $\pm$  95% confidence intervals) from *Hyaella azteca* survival and growth concentration–response curves. Standard 28-d bioassays were used to compare metal amended sediments that were either freshly spiked or aged in a mesocosm for >100 d.

<b>Metal</b>	<b>Sediment</b>	<b>Preparation</b>	<b>LC50</b>	<b>EC10 (growth)</b>
Cu	Dow	Aged	435 (341–590)	57 (22–90)
	Dow	Freshly spiked	395 (339–465)	29 (16–44)
	Ocoee	Aged	2520 (1800–4410)	217 (120–310)
	Ocoee	Freshly spiked	1590 (1360–1900)	238 (137–343)
Ni	Burntwood	Aged	>1120 <sup>a</sup>	951 (552–3440)
	Burntwood	Freshly spiked	1720 (1090–3800)	1010 (636–1180)

	Raisin	Aged	504 (355–702)	242 (110–386)
	Raisin	Freshly spiked	803 (364–2390)	499 (314–701)
Zn	Raisin	Aged	>2600 <sup>a</sup>	>2600 <sup>a</sup>
	Raisin	Freshly spiked	>3300 <sup>a</sup>	4260 (3320–inf. <sup>b</sup> )

<sup>a</sup> no significant concentration–response curve could be calculated and threshold is reported as greater than the maximum concentration tested; <sup>b</sup> upper confidence interval was unbounded (i.e., positive infinity).

**Table 4. Comparison of EC10s between experiments using homogenized and intact sediments. EC10s are from concentration–response relationships between total metal concentrations (mg kg<sup>-1</sup>) and *Hyalella azteca* relative growth rates. Test conditions differed between experiments with homogenized and intact sediment, most notably in the exposure chamber, test duration, and overlying water conditions. Homogenized sediment: beakers giving *H. azteca* free access to sediment, 28-d toxicity assay, and static renewal of overlying water; Intact (mesocosm) sediment: *H. azteca* caged at surface of sediment, 7-d toxicity assay, and flow-through water column.**

Metal	Sediment	Homogenized		Intact <sup>a</sup>	
		Freshly spiked	Aged	Early <sup>b</sup>	Late <sup>c</sup>
Cu	Dow	29	57	18	>800 <sup>d</sup>
Cu	Ocoee	238	217	43	92

Ni	Burntwood	1010	951	239	275
Ni	Raisin	499	242	20	>3000 <sup>d</sup>

---

<sup>a</sup> Data for 7-d *H. azteca* assays completed within the mesocosm (details in Costello et al. 2015, 2016); <sup>b</sup> early EC10 are from assays that were completed 2–7 days after the start of aging; <sup>b</sup> late EC10 are from assays that were completed 213–220 (Cu test) and 100–107 days (Ni test) after the start of aging; <sup>d</sup> calculated concentration–response threshold was beyond the range of total metal concentrations used in the test, and threshold is reported as greater than the maximum concentration tested.