Environmental Toxicology and Chemistry—Volume 41, Number 5—pp. 1260–1275, 2022
Received: 9 September 2021 | Revised: 28 October 2021 | Accepted: 23 January 2022

Environmental Toxicology

Interlaboratory Comparison of Three Sediment Bioaccumulation Tests

Guilherme R. Lotufo,[a,*] James M. Biedenbach,[a] J. Daniel Farrar,[a] Michael K. Chanov,[b] Brian W. Hester,[c] C. Ryan Warbritton,[d] Jeffery A. Steevens,[d] Jenifer M. Netchaev,[a] Anthony J. Bednar,[a] and David W. Moore[a]

[a]US Army Engineer Research and Development Center, Vicksburg, Mississippi, USA
[b]EA Engineering, Science, and Technology, Sparks, Maryland, USA
[c]EcoAnalysts, Port Gamble, Washington, USA
[d]US Geological Survey, Columbia Environmental Research Center, Columbia, Missouri, USA

Abstract: Standard bioaccumulation tests are commonly conducted using *Macoma nasuta* (clam), and *Alitta virens* (polychaete) for marine tests, and *Lumbriculus variegatus* (an oligochaete) for freshwater tests. Because the interlaboratory variability associated with these tests is unknown, four experienced laboratories conducted standard 28-day bioaccumulation tests with the above species using sediments contaminated with polychlorinated biphenyls (PCBs) and polycyclic aromatic hydrocarbons (PAHs). Chemical analysis of tissue samples was performed by a single laboratory. The intralaboratory variance among replicates was relatively low for PCB tissue concentrations, with coefficients of variation (CVs) ranging from 9% to 28% for all laboratories and species, with the exception of one laboratory reporting higher variability for *L. variegatus* (CV = 51%). Intralaboratory variance for PCB tissue concentrations was higher than interlaboratory variance for *A. virens* and *L. variegatus*, and the magnitude of difference (MOD) for laboratory means ranged from 1.4 to 2.0 across species. Intralaboratory variability was also low for lipid content, and lipid normalization of PCB and PAH body residues generally had little impact on variability. In addition to variability across bioassay laboratories, analytical variability was evaluated by different laboratories measuring the concentration of PCBs and total lipids in a subsample of tissue homogenate of sediment-exposed test organisms. Variability associated with tissue analysis was higher than bioassay laboratory variability only in tests with *L. variegatus*. Statistical differences between samples may be observed due to the low intralaboratory variability; however, the biological significance of these differences may be limited because the MOD is low. Considering the MOD when comparing bioaccumulation across treatments accounts for uncertainty related to inherent variability of the test in the interpretation of statistically significant results. *Environ Toxicol Chem* 2022;41:1260–1275. © 2022 The Authors. *Environmental Toxicology and Chemistry* published by Wiley Periodicals LLC on behalf of SETAC. This article has been contributed to by US Government employees and their work is in the public domain in the USA.

Keywords: Bioaccumulation; Benthic macroinvertebrates; Sediment assessment; Polychlorinated biphenyls; *Macoma nasuta*; *Alitta virens*; *Lumbriculus variegatus*

INTRODUCTION

Bioaccumulation is defined as the net accumulation of contaminants from water, diet, and sediment into the tissues of exposed organisms (Weisbrod et al., 2009). Bioaccumulation tests are intended to directly measure the uptake of contaminants into the tissues of the test organism and therefore provide quantitative information concerning chemical exposure at a contaminated site (Van Geest et al., 2010). The primary objective of evaluating bioaccumulation resulting from exposure to contaminated sediments is to obtain a measurement of exposure (USEPA, 2000). Bioaccumulation data can be used to estimate effects in invertebrates directly or support exposure estimates for fish and wildlife that feed on invertebrates (McElroy et al., 2011). In addition, health hazards to humans may ensue from the consumption of shellfish or pelagic fish that may have directly or indirectly bioaccumulated sediment-associated contaminants. Consequently, assessment of

This article includes online-only Supporting Information.

* Address correspondence to Guilherme.Lotufo@usace.army.mil
Published online 25 January 2022 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/etc.5300

bioaccumulation is commonly included in sediment quality evaluations (USEPA, 2016).

The US Army Corps of Engineers (USACE) and the US Environmental Protection Agency (USEPA) share federal responsibility for regulating the management of dredged material within waters of the United States under section 404 of the Clean Water Act and in ocean waters under section 103 of the Marine Protection, Research, and Sanctuaries Act. The USACE and USEPA have jointly developed testing and interpretative guidance for the evaluation of dredged materials proposed for aquatic placement (USEPA & USACE, 1991, 1998). Regional guidance from USEPA and USACE offices provide region-specific and often more detailed guidance relative to national guidance (e.g., USEPA & USACE, 2008; Northwest Regional Sediment Evaluation Team, 2018). The evaluation of dredged material proposed for open-water placement may require an assessment of benthic bioaccumulation potential using laboratory sediment bioaccumulation tests in which invertebrate test species are exposed to material proposed for dredging and sediment collected from one or more designated reference areas. Infaunal invertebrates are used as test species and are exposed within environmentally controlled conditions to sediment for 28 days. At test termination, organisms are collected from each of the replicate chambers and placed in water for purging of gut content; then the organisms' tissues are homogenized for analytical determination of contaminants of concern. Concentrations measured in the tissues of organisms exposed to dredged sediments are statistically compared with concentrations measured in tissues of reference-sediment exposed organisms, typically via analysis of variance (ANOVA) or nonparametric pair-wise comparison to determine which analytes are statistically significantly elevated at the 95% confidence level. Any significant differences are then further evaluated to determine adverse outcomes in higher trophic level receptors. Appropriate interpretation of bioaccumulation test results is paramount to ensuring accurate conclusions so that risks to fish, wildlife, or humans are appropriately managed (Lotufo et al., 2014; McQueen et al., 2020).

Standard test methods for conducting sediment bioaccumulation tests are still under development, and current guidance is provided by ASTM International (2019) and the Organization for Economic Co-operation and Development (OECD, 2008). The use of species with a high potential to bioaccumulate contaminants is desirable, because wide differences have been observed among different species (ASTM International, 2019). Relatively few freshwater, marine, and estuarine species are routinely used in sediment bioaccumulation tests. The most commonly used test species include the freshwater oligochaete *Lumbriculus variegatus* (California blackworm) for evaluation of freshwater sediments and the polychaete *Alitta virens* (sandworms; formerly *Nereis virens*) and the marine bivalve *Macoma nasuta* (bent-nosed clam) for marine and estuarine evaluations. Testing methods for these species serve as the primary basis of the generic ASTM International (2019) guidance for sediment bioaccumulation testing.

Bioaccumulation of contaminants in benthic invertebrates is a complex phenomenon because it occurs through multiple exposure routes including dietary assimilation, transport across respiratory surfaces, and dermal absorption, and it is influenced by a variety of factors that control contaminant bioavailability (Selck et al., 2012). Due to that complexity, predicting bioaccumulation in organisms inhabiting field sites using laboratory exposures of a model organism in sediment collected from those sites is inherently uncertain (Beckingham & Ghosh, 2010; Burkhard et al., 2012; Selck et al., 2012). Even under standardized and controlled conditions, a multitude of factors contribute to variability when one is measuring bioaccumulation using sediment bioaccumulation tests. Those factors include the homogeneity of sediment distributed across experimental replicates, the selection of contaminants of concern (e.g., loss of water-soluble contaminants of concern occurs during renewal of overlying water), and the variability and complexity associated with analytical detection of contaminants of concern in tissue (Ingersoll et al., 1995). Factors associated with the test organism such as their initial health, lipid content and baseline body burden of contaminants of concern, feeding (e.g., sediment-ingesting organisms may not actively ingest sediment during laboratory exposures), and sediment avoidance (e.g., infaunal organisms may avoid burrowing due to high NH_4 concentration) also contribute to inherent variability (McElroy et al., 2011; Burkhard et al., 2015). Moreover, guidance documents (see ASTM International, 2019) provide only recommendations for many procedural aspects, potentially leading to methodological differences among laboratories following the same guidance document; for example, the overlying water in the exposure vessel may be exchanged on a periodic basis or continuously renewed using flow-through systems. Despite those sources of variability, laboratory-testing variability of the outcome of bioaccumulation assessment is expected to be considerably lower than the variability associated with natural environments (Burkhard et al., 2012; Hoke et al., 2016). Even though laboratory sediment bioaccumulation tests have yielded relatively precise and repeatable results (Egeler et al., 2006; McQueen et al., 2020; Van Geest et al., 2011a), method performance and analytical variability are largely unknown for sediment bioaccumulation tests.

Our study presents the results of an interlaboratory comparison conducted to quantify variability associated with bioaccumulation sediment tests using *L. variegatus*, *A. virens*, and *M. nasuta*. Although these are the most used species in sediment bioaccumulation testing, interlaboratory comparisons of standard tests using these species have not been conducted to date. Therefore, the primary objective of our study was to examine interlaboratory bioaccumulation variability using a homogenous batch of sediment and organisms supplied out of a single pool. Sediment and tissue samples generated from each bioassay laboratory were analyzed for polychlorinated biphenyl (PCB) congeners and polycyclic aromatic hydrocarbon (PAH) compounds by a single analytical laboratory.

A secondary focus of the study was the investigation of the agreement among analytical laboratories using the same methodology on whole-body tissue homogenate produced for each species. Testing of procedural and instrument variability aims to quantify uncertainty during sediment bioaccumulation

testing and hence improve our ability to characterize exposure risk associated with dredging material. Specifically, when comparing a site sediment to a reference, the magnitudes of difference (MODs) may be de minimis in the context of ecological and human health risk.

MATERIALS AND METHODS

Overview of interlaboratory comparison of sediment bioaccumulation tests

Sediment from a heavily contaminated site (New Bedford Harbor, MA, USA) containing elevated concentrations of PCBs was mixed with either relatively uncontaminated freshwater or marine sediment to achieve a concentration range where test organisms are likely to accumulate organic chemicals but not result in significant mortality (Gidley et al., 2019). The two resulting sediments (hereinafter termed "test sediments," one for marine–estuarine testing and one for freshwater testing) were apportioned and shipped to four participating laboratories for the interlaboratory sediment bioaccumulation testing. Participating laboratories included the US Army Engineer Research and Development Center (ERDC; Vicksburg, MS), the US Geological Survey Columbia Environmental Research Center (CERC; Columbia, MO), EA Engineering, Science, and Technology (EA; Hunt Valley, MD). and EcoAnalysts (ECO; Port Gamble, WA), all of which had a proven track record of successfully performing sediment bioaccumulation tests for research or site assessments. The ERDC provided sediment and test organisms to all participating laboratories. All laboratories conducted the *M. nasuta, A. virens*, and *L. variegatus* 28-day sediment bioaccumulation tests following basic procedures provided by the ERDC, which were based on national guidance as detailed in the *Alitta virens and M. nasuta bioaccumulation tests* and the *Lumbriculus variegatus bioaccumulation tests* sections. All laboratories followed the same procedure for aquaria size and setup, sediment volume, test organism loading, temperature, and organism retrieval, handling, and shipping at test termination (ASTM International, 2019). Test maintenance including water exchanges and water quality parameter measurements followed laboratory-specific testing protocols and met minimum test requirements to meet test acceptability (ASTM International, 2019). Per the test protocols, supplemental feeding was not provided. Each test sediment was tested in five replicate aquaria, with each replicate generating one sample containing all surviving organisms for subsequent chemical analyses. Test sediment and tissue samples of organisms exposed to test sediments were analyzed by one analytical laboratory for total lipids, select PCB congeners, and the 16 USEPA Priority PAHs (Keith, 2015). The 17 PCB congeners (IUPAC numbers 8, 18, 28, 44, 52, 66, 101, 105, 118, 128, 138, 153, 170, 180, 187, 195, and 206) selected for analysis, with the addition of the 209 are used by the US National Oceanic and Atmospheric Administration (NOAA) National Status and Trends Program in the United States (Lauenstein & Cantillo, 1993) and will be termed NOAA PCBs. These congeners are significant components of the most common commercial PCB mixtures (i.e., Aroclor 1016, 1248, 1254, and 1260). This congener mix also contains six congeners from the International Council for Exploration of the Sea list (i.e., 52, 101, 118, 138, 153, and 180) proposed as pollution indicators (Duinker et al., 1988).

Sediments

The New Bedford Harbor sediment used to prepare test sediments was collected from an estuarine location at the New Bedford Harbor Superfund site (Nelson & Bergen, 2012) in September 2014. The sediment was transported to the ERDC and stored in steel drums at 2.8–4.0 °C. The sediment used in the present study was considered "highly contaminated," because the total concentration of PCB (∑PCBs) congeners was 45 mg/kg dry weight, and the total concentration of PAHs (∑PAHs) was 10 mg/kg dry weight; see Schmidt et al., 2017 for additional information). The sum concentration of NOAA PCBs (∑NOAA-PCBs) corresponded to 36% of the total PCB concentration in the New Bedford Harbor sediment. Sediments with no detectable concentration of PCBs were obtained from an estuarine site in Sequim Bay (Sequim, WA, USA) for use in the marine bioaccumulation tests; freshwater bioaccumulation tests utilized no detectable PCB sediments from Horseshoe Lake, an oxbow lake adjacent to the Mississippi River in Warren County (MS, USA) (Knight et al., 2009; Kennedy et al., 2010). Each of these sediments was mixed with New Bedford Harbor sediment until thoroughly homogenized (to uniform color and consistency, for ~20 min.) using an impeller mixer. The amount of New Bedford Harbor sediment added was targeted to achieve a ∑PCBs of approximately 1 and 2 mg/kg dry weight for the marine and freshwater test sediments, respectively. A higher target concentration was used for freshwater testing because of smaller target tissue mass/replicate. Those target concentrations are higher than total PCB concentrations reported for navigation channels and harbors in the United States (see Fredette et al., 2007; McQueen et al., 2020; Steevens et al., 2008) but were selected to ensure detection of NOAA PCBs in tissue samples. Benthic toxicity was not expected at that range of concentrations, based on Swartz (1999) and Finkelstein et al. (2021). Following vigorous mixing for approximately 20 min, three 0.5-L aliquots of the sediments for marine and freshwater testing were assessed for homogeneity by analyzing Aroclors concentrations in three representative aliquots and were deemed sufficiently homogeneous for use in the interlaboratory comparison based on the variability of the concentrations of ∑PCBs among sediment aliquots (Supporting Information, Table S1).

Experimental organisms

Alitta virens constructs deep (8–10 cm or deeper), vertical, well-irrigated, semipermanent burrows lined with mucus, and has been reported to be an omnivorous feeder that also nonselectively deposit-feeds (Ciarelli et al., 2000). *Macoma nasuta* is a free-burrowing clam that deposit-feeds by siphoning the top few millimeters of the sediment surface; suspension

feeding is considered a supplementary feeding mode for this organism (Hylleberg & Gallucci, 1975). *Lumbriculus variegatus* feeds on decaying vegetation, microorganisms, and sediment (Brinkhurst & Gelder, 1991; Penttinen et al., 1996), burrowing into the sediment head first and keeping its tail in the water, where respiration and photoreception occurs (Penttinen et al., 1996). *Alitta virens* and *M. nasuta* were purchased from Aquatic Research Organisms in Hampton (NH, USA) and shipped overnight by the vendor directly to the participating laboratories. *Alitta virens* were collected from the Damariscotta River (Boothbay Harbor, ME, USA), and *M. nasuta* were collected from Tomales Bay (Dillon Beach, CA, USA). *Lumbriculus variegatus* was purchased from California Blackworm Company (Fresno, CA, USA) and was maintained at the ERDC laboratory for approximately 10 weeks prior to being shipped overnight to the participating laboratories. All participating laboratories reported receiving the organisms in adequate health status based on appearance and behavior, except for *L. variegatus* received at CERC. Because of insufficient mass at the ERDC for shipping a second batch, CERC used their in-house cultured organisms. On arrival, organisms were acclimated for least 24 h to the test temperature in the test water under aeration.

Alitta virens and M. nasuta bioaccumulation tests

Marine sediment 28-day bioaccumulation tests were conducted in accordance with national testing guidance for sediment (ASTM International, 2019) and dredged material evaluations (USEPA & USACE, 1998). Five replicates of the test sediment and control sediment were setup by adding approximately 6 kg of sediment and the appropriate seawater (30 ppt) to 20-L aquaria. The EA and ERDC laboratories used activated carbon–treated tap water or de-ionized well water (CERC) mixed with commercially available synthetic sea salts; ECO used filtered natural seawater. Aquaria were maintained at 20 ± 1 °C (*A. virens*) or 15 ± 1 °C (*M. nasuta*) under a 16:8-h light:dark cycle using ambient light intensity with aeration. Overlying water quality (temperature, pH, dissolved oxygen, salinity, and ammonia) was recorded for each replicate aquarium at bioassay initiation. A variable number of organisms were added to each aquarium at test initiation. For *A. virens*, the CERC, ERDC, and EA laboratories added 5–8 individuals, targeting 30–35 g of total biomass, and the ECO laboratory added 10–13 individuals. For *M. nasuta* 10–14 clams weighing 167–184 g, including the shells, were added across all laboratories. Two laboratories (ECO and EA) conducted overlying water renewals of 2 volume additions/day, and CERD and ERDC exchanged 50%–60% of the water volume three times weekly. Temperature, dissolved oxygen, pH, and salinity were measured during the test, at a minimum for one replicate/treatment weekly and for all replicates on day 28.

After 28 days, each replicate aquarium was terminated by gently separating surviving organisms from the sediment by hand to avoid physical damage. Surviving *A. virens* or *M. nasuta* from each replicate were enumerated and placed in clean water at the test temperature for 24 h for gut purging. Following gut purging, the soft tissue of *M. nasuta* was separated from the shells, tissues from each replicate were placed in a preweighed jar. and total tissue mass (wet wt) was determined. Tissue samples were frozen for at least 24 h and subsequently shipped overnight on ice to the ERDC, where they were grouped with samples from the other laboratories and sent to the designated analytical laboratory for analysis. Tissue and sediment samples were labeled without including the identity of the participating laboratory so as to avoid bias in the results.

Lumbriculus variegatus bioaccumulation test

Freshwater sediment 28-day bioaccumulation experiments were conducted in accordance with national testing guidance for sediment (ASTM International, 2019; USEPA, 2000) and dredged material evaluations (USEPA & USACE, 1998). Five replicates of the test sediment and control sediment were set up by adding approximately 2.8 kg of sediment and non-chlorinated water (i.e., chlorine-free or dechlorinated) to 8-L aquaria. Aquaria were maintained at 23 ± 1 °C with aeration under a 16:8-h light:dark cycle using ambient light intensity. Overlying water quality (temperature, pH, dissolved oxygen, conductivity, and ammonia) were recorded for each replicate aquarium at bioassay initiation. Approximately 10 g of *L. variegatus* were added to each aquarium at test initiation. Three of four laboratories (i.e., CERC, ECO, and EA) conducted overlying water renwals of 2 volume additions/day, and ERDC exchanged 50-60% of the water volume three times weekly. Temperature, dissolved oxygen, pH, and conductivity were measured during the test, at a minimum for one replicate/treatment weekly and for all replicates on day 28. After 28 days, each replicate aquarium was terminated by passing the sediment through a 250-µm screen. Surviving organisms from each replicate were recovered and placed for 24 h in chambers specially designed to facilitate separation of *L. variegatus* from materials (primarily plant detritus) retained on the sieve during recovery of test organisms. Details of this separation method are described in Lotufo et al. (2021). Following separation from detritus and gut purging, organisms from each replicate were placed in a preweighed jar, and total tissue mass (wet wt) was determined. Tissue samples were frozen for at least 24 h and then shipped overnight on ice to the ERDC, where they were grouped with samples from the other laboratories and, along with sediment samples, sent to the designated analytical laboratory for analysis.

Interlaboratory comparison of chemical analysis of sediment bioaccumulation test organisms

Concurrent with the bioaccumulation tests just described, five additional replicate aquaria for each test species evaluated (*A. virens*, *M. nasuta*, and *L. variegatus*) were set up at the ERDC laboratory to generate tissue for an interlaboratory comparison of chemical analysis of tissue samples. The replicate aquaria were set up using the same batch of test sediment and organisms used in the tests just described. Exposed organisms of each species were recovered and handled as just

described and composited across all replicates. Thoroughly homogenous pooled samples of *A. virens*, *M. nasuta*, and *L. variegatus* were each equally split into four aliquots, each analyzed by a different participating analytical laboratory. Participating analytical laboratories included the ERDC, CERC, and two unnamed commercial laboratories, all of which had experience analyzing aquatic invertebrate tissue samples. Each participating laboratory analyzed a single tissue sample of each test species for NOAA PCBs and total lipids.

Chemical analysis of sediment and tissue samples for the bioaccumulation test interlaboratory comparison

Sediment samples collected at test initiation (one sample/replicate/laboratory for the *M. nasuta* test and one single sample/laboratory for the *L. variegatus* test) were analyzed for PAHs and PCBs. Five replicate tissue samples were taken at the ERDC laboratory at test initiation for each species from the pool of test organisms for analysis of "baseline" pre-exposure concentrations of contaminants of concern. Five additional samples were taken at CERC of *L. variegatus* pre-exposure organisms because they represented a different test population. Organisms recovered from control aquaria were not analyzed. Sediment samples were extracted by pressurized fluid extraction using hexane/acetone (USEPA method 3545A; 2007a) and with the extract treated with silica gel (modified USEPA method 3630C; 1996a) or sulfuric acid (modified USEPA method 3665A; 1996b) for PAHs and PCBs, respectively, to reduce matrix interferences. The total organic carbon content (TOC) of the test sediments was determined following the Lloyd Kahn method (USEPA, 1988). Grain size was analyzed according to ASTM International (2007).

Tissue samples were extracted with hexane (95%; Fisher Scientific) by sonication, following a modification of USEPA method 3550C (2007b). The PCB congeners were extracted either in a sonic bath overnight with hexane for larger tissue amounts, or in a single extraction with hexane using a sonic probe for smaller amounts of tissue. Tetrachloro-m-xylene and PCB congener 195 were used as extraction surrogates for all PCB analyses. Lipids and other interfering compounds were removed from sample extracts by treatment with sulfuric acid (modified USEPA method 3665A; 1996b). Dual column gas chromatography with electron capture detection (ECD) according to USEPA method 8082 (2007c) was used for analysis of PCB congeners as previously described by Schmidt et al. (2017). The PAHs were analyzed using gas chromatography–mass spectrometry (GC–MS) in selected ion monitoring mode following modifications of USEPA method 8270E (2018). This analysis used deuterated PAH compounds for internal standards and 2-fluorobiphenyl and terphenyl for extraction surrogate compounds. For tissues or sediment, a single value was used for all individual PCB congeners as a reporting limit, with a detection limit at one-third this concentration. Detection limits were 0.4, 0.6, and 0.3 μg/kg for sediment, *A. virens* and *M. nasuta*, and *L. variegatus* samples, respectively. The reporting limit was used as the lowest calibration standard and is supported by method detection limit studies performed per 40 CFR 136 Appendix B guidelines (USEPA, 2011). No co-elution was reported. For PAHs, detection limits were 4, 5, and 20 μg/kg for sediment, *A. virens* and *M. nasuta*, and *L. variegatus* samples, respectively. Total lipid content of tissues was measured using the standard Bligh–Dyer macrogravimetric determination (Bligh and Dyer, 1959).

Chemical analysis of tissue samples for the analytical chemistry interlaboratory comparison

Tissue samples for the analytical comparison study were analyzed for PCBs following the basic guidance of USEPA method 8082 (2007c) at each of the participant laboratories: ERDC, CERC, and two commercial facilities (termed Lab A and Lab B) separate from the bioassay laboratories. Tissue analysis performed at ERDC followed the sonication extraction method (USEPA method 3550C; 2007b), with sulfuric acid treatment, followed by GC–ECD analysis (USEPA method 8082; 2007c) as just described.

The CERC laboratory extracted the tissue samples following the procedure described in Gale et al. (2009), using PCBs 29, 155, and 204 as extraction surrogates. Analysis of the extracts was accomplished by GC–ECD after interference reduction using size exclusion chromatography and silica gel columns following the procedures outlined in Gale et al. (2009).

The two commercial laboratories followed USEPA method 8082 (2007c), with both using tetrachloro-m-xylene as one extraction surrogate, athough the second surrogate, the method of extraction, and the method of extraction cleanup differed between the two facilities. Laboratory A used sonication for extraction (USEPA method 3550C; 2007b) and PCB congeners 103 and 192 as additional extraction surrogates, with interference reduction achieved using gel permeation chromatography (USEPA method 3640A; 1994a). Laboratory B used automated Soxhlet extraction (USEPA method 3541; 1994b) and PCB 205 as a second surrogate. Laboratory B also treated the extracts for sulfur (USEPA method 3660B; 1996c) and with acid (USEPA method 3665A; 1996b) for interference reduction prior to analysis.

Data analysis

Lipid normalized concentrations were determined by dividing bulk tissue concentration by the organism sample-specific fraction lipid content (data are expressed as mg PCB/kg lipid). Biota-to-sediment accumulation factors (BSAFs) were calculated as described in ASTM International (2019). To quantify intralaboratory variability, coefficients of variation (CV; expressed as a percentage) were computed for analytical chemistry data derived from each individual laboratory.

Statistical comparisons and determinations of data normality (Kolmogorov–Smirnov test) and homogeneity (Levene's test) were performed using SigmaStat Ver 3.5 software (SSPS). One-way ANOVAs was performed to determine statistically significant differences ($\alpha = 0.05$) across participating

laboratories. When the data did not meet the assumption of normality, the Kruskal–Wallis one-way ANOVA on ranks was applied. The Holm–Sidak method was employed as an all-pairwise multiple comparison procedure to determine statistical significance in bioaccumulation between different participating laboratories. The MOD was determined using the ratio between the highest and the lowest means among laboratories. Variability was also quantified for each bioaccumulation test (i.e., *A. virens*, *M. nasuta*, and *L. variegatus* tests) using replicate bioaccumulation measurements across all participating laboratories and subsequently calculating a variation factor (VF) for PCB congeners, ∑NOAA-PCBs, and lipid content as described in Jonker et al. (2018). This factor was calculated by taking the ratio of the 95th percentile (PCTL) value of the averaged body burden/target chemical to the 5th percentile value:

$$\text{VF} = \frac{\text{95th PCTL}}{\text{5th PCTL}}$$

According to Jonker et al. (2018), the use of this index of variability assumes that experimentally determined concentrations exhibited normal distributions. Therefore, normality was assessed using the Kolmogorov–Smirnov test.

RESULTS

Marine and freshwater test sediments

One sample for each batch of test sediments (i.e., marine and freshwater) was used for TOC and grain size analysis. The TOC was 3.0% for the marine test sediment and 3.5% for the freshwater sediment. Grain size distribution for the marine (coarse sand 0.5%, medium sand 2.8%, fine sand 12.8%, silt 47.7%, clay 38.2%) and freshwater (coarse sand 0.6%, medium sand 2.4%, fine sand 4.2%, silt 37.1%, and clay 55.7%) sediments indicated a predominance of fine particles.

While sediment was added to replicate aquaria, participating laboratories obtained five samples of the marine test sediment used in the interlaboratory comparison, one from each replicate, but obtained only one sample of the freshwater test sediment. For ∑NOAA-PCBs, concentrations were similar across individual laboratory samples and across means for all laboratories (Figure 1), indicating that the single batch sediment was well homogenized. Variability was also low for ∑NOAA-PCBs concentrations for the single sample of the freshwater test sediment (Figure 1). Congener-specific concentrations in test sediments are presented in the Supporting Information, Tables S2 and S3. Using the mean ∑NOAA-PCB concentration for replicate samples from all laboratories and the concentration measured in the concentration of ∑NOAA-PCBs in the New Bedford Harbor sediment, the dilution factor was 49 for the marine and 21 for the freshwater test sediment. For ∑PAHs, concentrations varied more widely across individual laboratory samples and across means for all laboratories (Supporting Information, Figure S1). Individual PAH-specific concentrations in test sediments are presented in the Supporting Information, Tables S4 and S5.

Organismal recovery and lipid content

For *A. virens*, survival was 100% for all replicates for all laboratories, with replicate final biomass ranging from 26 to 32 g for CERC and ERDC and from 45 to 66 for EA and ECO. For *M. nasuta*, survival ranged from 83% to 100% for all replicates for all laboratories, with the mean for each laboratory ranging from 92% to 100%. Replicate final biomass of soft tissue ranged from 35 to 45 g for CERC, ERDC, and ECO and from 56 to 80 g for EA. For *L. variegatus*, replicate final biomass of soft tissue ranged from 4.9 to 8.2 g, corresponding to a percentage recovery of added mass ranging from 49% to 75.2%. Note that exhaustive recovery of organisms was not attempted in the present study.

Mean baseline (i.e., at test initiation) lipid content was highest for *L. variegatus*, followed by *A. virens* and *M. nasuta* (Figure 2). For *A. virens* and *M. nasuta*, mean lipid content at experiment termination was within ±23% and 39%, respectively, the mean baseline lipid content. For *L. variegatus*, mean baseline lipid content for the organism batch used by EA, ECO, and ERDC was virtually identical to that for the organism batch used by CERC (Figure 2). The mean lipid content at

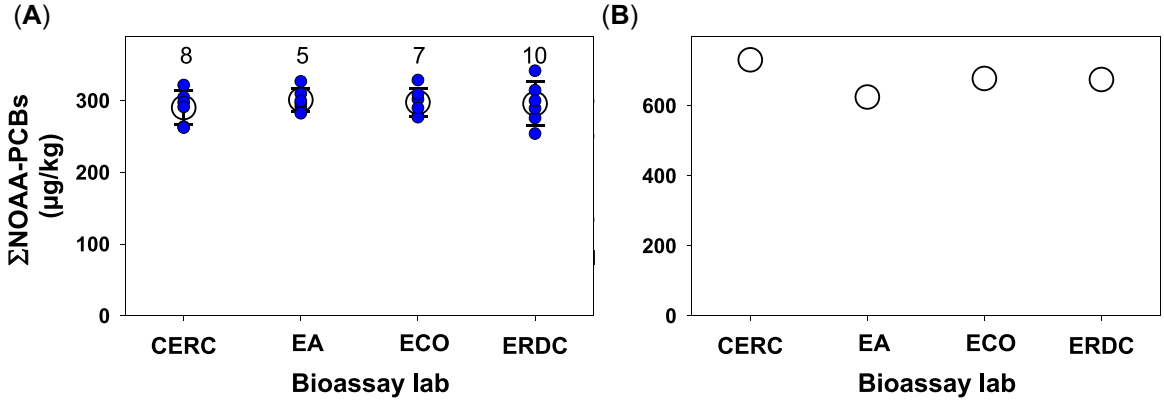

FIGURE 1: Replicate (small circles) and mean concentrations (large circles) of ∑NOAA-PCBs in marine sediment (**A**) and concentration of ∑NOAA-PCBs in freshwater sediment (**B**) sampled from each bioassay laboratory. Only one sample generated in each bioassay laboratory was analyzed for the freshwater sediment. Error bars are ±1 standard deviation ($n = 5$). The number over the error bar is the coefficient of variation as a percentage. For abbreviations, see footnote to Table 1.

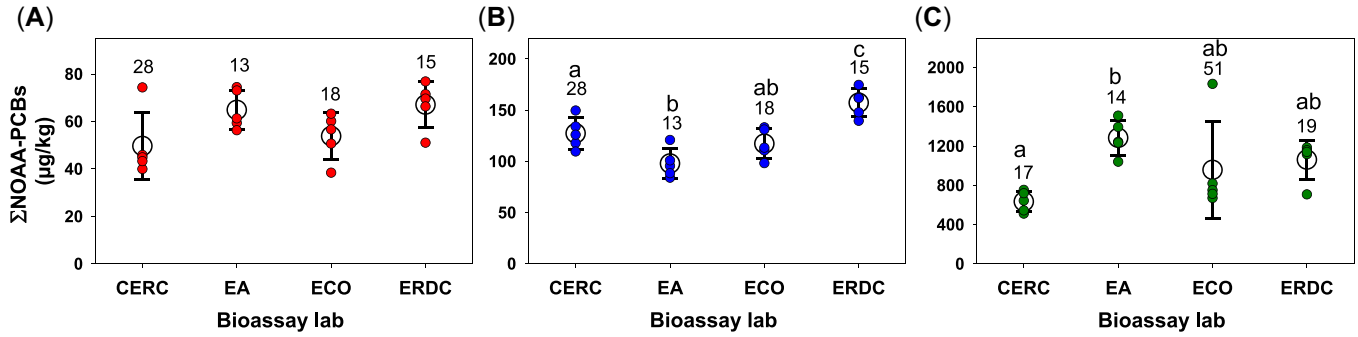

FIGURE 2: Replicate (small circles) and mean (large circles) lipid content as percentage of the wet weight for *Alitta virens* (**A**), *Macama nasuta* (**B**), and *Lumbriculus variegatus* (**C**), both unexposed ("baseline") and exposed to sediment for 28 days at each bioassay laboratory. Error bars are ±1 standard deviation ($n = 5$). The number over the error bar is the coefficient of variation as a percentage. For **A**, different letters indicate significant differences from pairwise comparisons following one-way analysis of variance. For *L. variegatus*, "Baseline-C" represents the batch used for the US Geological Survey Columbia Environmental Research Center laboratory, and "Baseline" represents the batch used by the other bioassay laboratories. For abbreviations, see footnote to Table 1.

experiment termination was lower than the mean baseline by 37% or less. Overall, mean lipid content varied relatively little across replicates within treatments (i.e., baseline and bioassay laboratories), and across species, with CVs ranging from 9% to 25% for all treatments except for *M. nasuta* for the baseline and EA treatments, for which the CV was 54% and 60%, respectively.

Alitta virens bioaccumulation of PCBs

The concentrations of all NOAA PCBs were below detection limits for all baseline samples. Intralaboratory variability was low: the mean concentrations of ∑NOAA-PCBs ranged from 50 to 67 µg/kg, with a MOD of 1.4, and there was no significant difference between participating laboratories (Table 1 and Figure 3). Intralaboratory variability was also relatively low, with CVs ranging from 13% to 28%. When lipid normalized, the interlaboratory variability increased, with a MOD of 1.8 and statistical differences between laboratories (Table 1), but intralaboratory variability remained similar (Figure 4). The ∑NOAA-PCB and congener-specific bioaccumulation data are presented in the Supporting Information, Tables S6 and S7. Across the 20 replicate samples analyzed by all laboratories, variability was low (MOD = 2.0; CV = 21%) for ∑NOAA-PCBs concentrations but was higher (MOD = 3.2; CV = 31%) for lipid-normalized ∑NOAA-PCBs.

Macoma nasuta bioaccumulation of PCBs

The concentrations of all NOAA PCBs were below detection limits for all baseline samples. Intralaboratory variability was low: the mean concentrations of ∑NOAA-PCBs ranged from 98 to 157 µg/kg, with a MOD of 1.6, but statistical differences between laboratories were detected (Table 1 and Figure 3). Intralaboratory variability was low, with CVs ranging from 9 to 15%. When lipid normalized, the interlaboratory variability decreased (MOD = 1.3) and there were no statistical differences between laboratories (Table 1), but intralaboratory variability increased (Figure 4). The ∑NOAA-PCB and congener-specific bioaccumulation data are presented in the Supporting

TABLE 1: Summary of analysis of variance for total lipids, concentration of ∑NOAA-PCBs (∑PCBs), and lipid-normalized ∑NOAA-PCBs ($\sum PCBs_{lipid}$) concentration data from 28-day bioaccumulation tests conducted by four bioassay laboratories

			Source of variation (percent of total)		Ratio of means			
Analysis	Bioaccumulation test	*p* value	Inter-laboratory	Intra-laboratory	CERC	EA	ECO	ERDC
Total lipids	*Alitta virens*	<0.01	66%	34%	1.3	1.0	1.0	**1.5**
	Macoma nasuta	0.41	ND[a]	ND	1.00	**1.05**	1.04	1.04
	Lumbriculus variegatus	0.44	ND	ND	1.00	1.03	1.02	**1.04**
∑PCBs	*Alitta virens*	0.055	37%	63%	1.0	1.3	1.1	**1.4**
	Macoma nasuta	<0.01	73%	27%	1.3	1.0	1.2	**1.6**
	Lumbriculus variegatus	0.02	46%	54%	1.0	**2.0**	1.5	1.7
$\sum PCBs_{lipid}$	*Alitta virens*	<0.01	56%	44%	1.0	**1.8**	1.4	1.2
	Macoma nasuta	0.39	17%	83%	1.2	1.0	1.0	**1.3**
	Lumbriculus variegatus	0.72	16%	84%	1.0	1.0	**1.3**	1.0

The ratio of means for a given laboratory represents the ratio between the average for that laboratory and the lowest average among laboratories. The magnitude of difference (MOD; ratio of highest and lowest mean) for each analysis is in bold type, and *p* values indicative of significant differences are in red type.
[a]ND = not determined because test of data normality failed, and thus nonparametric statistical tests were used.
∑NOAA-PCBs = sum National Oceanic and Atmospheric Administration-polychlorinated biphenyls; CERC = US Geological Survey Columbia Environmental Research Center; EA = EA Engineering, Science, and Technology; ECO = EcoAnalysts; ERDC = US Army Engineer Research and Development Center.

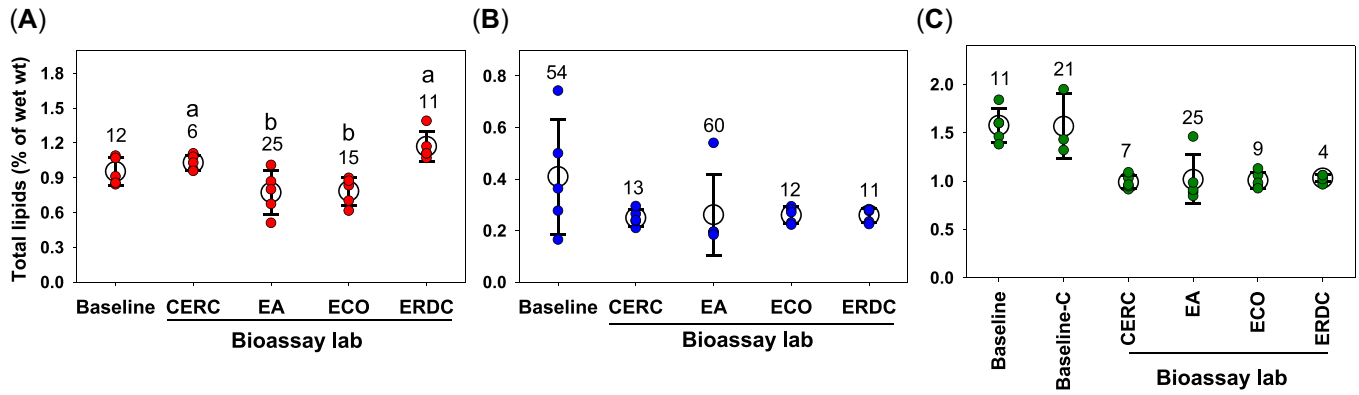

FIGURE 3: Replicate (small circles) and mean concentrations of ∑NOAA-PCBs in *Alitta virens* (**A**), *Macoma nasuta* (**B**), and *Lumbriculus variegatus* (**C**) exposed to sediment for 28 days at each bioassay laboratory. Error bars are ±1 standard deviation ($n = 5$). The number over the error bar is the coefficient of variation as a percentage. For **B** and **C**, different letters indicate significant differences from pairwise comparisons following one-way analysis of variance. For abbreviations, see footnote to Table 1.

Information, Tables S8 and S9. Across the 20 replicate samples analyzed by all laboratories, variability was low (MOD = 2.1; CV = 21%) for ∑NOAA-PCBs concentrations but was higher (MOD = 3.8; CV = 24%) for lipid-normalized ∑NOAA-PCBs.

Lumbriculus variegatus bioaccumulation of PCBs

The concentrations of all NOAA PCBs were below detection limits for all baseline samples. Intralaboratory variability was low: the mean concentrations of ∑NOAA-PCBs ranged from 634 to 1283 µg/kg, with a MOD of 2.0, but statistical differences between laboratories were detected (Table 1 and Figure 3). Intralaboratory variability was typically low, with CVs ranging from 14% to 19% for three laboratories, but was 51% for one laboratory. When lipid normalized, the interlaboratory variability decreased (MOD = 1.3) and there were no statistical difference between laboratories (Table 1), but intralaboratory variability increased (Figure 4). The ∑NOAA-PCB and congener-specific bioaccumulation data are presented in the Supporting Information, Tables S10 and S11. Across the 20 replicate samples analyzed by all laboratories, variability was high (MOD = 3.6; CV = 36%) relative to that for *A. virens* and *M. nasuta* for ∑NOAA-PCB concentrations and increased slightly (MOD = 3.9; CV = 39%) for lipid-normalized ∑NOAA-PCBs.

Sediment and tissue congener profiles for PCBs

The relative concentration of the individual NOAA PCBs for *M. nasuta* and *A. virens* was similar between the two species and for the marine test sediment (Figure 5). The most notable differences were for PCB 28, which accounted on average for 29% of the ∑NOAA-PCBs for sediment, 35% for *M. nasuta*, and 22% for *A. virens*, for PCB 52, which accounted for approximately 16%–17% of the sum for sediment and *M. nasuta* but 29% for *A. virens*, and for PCB 44, which accounted for approximately 9% of the sum for sediment but only approximately 2% for *A. virens* and *M. nasuta*. The relative concentration of the individual NOAA PCBs for *L. variegatus* was similar to that for freshwater sediment (Figure 5). The most notable differences were for PCB 28, which accounted on average for 27% of the ∑NOAA-PCBs for sediment, but only 16% for tissue, and for PCBs 101, 138. and 153, which were 50%–81% higher in tissue than in sediment.

BSAFs for PCBs

The BSAF values were calculated using replicate lipid-normalized PCB or PAH tissue concentrations divided by the

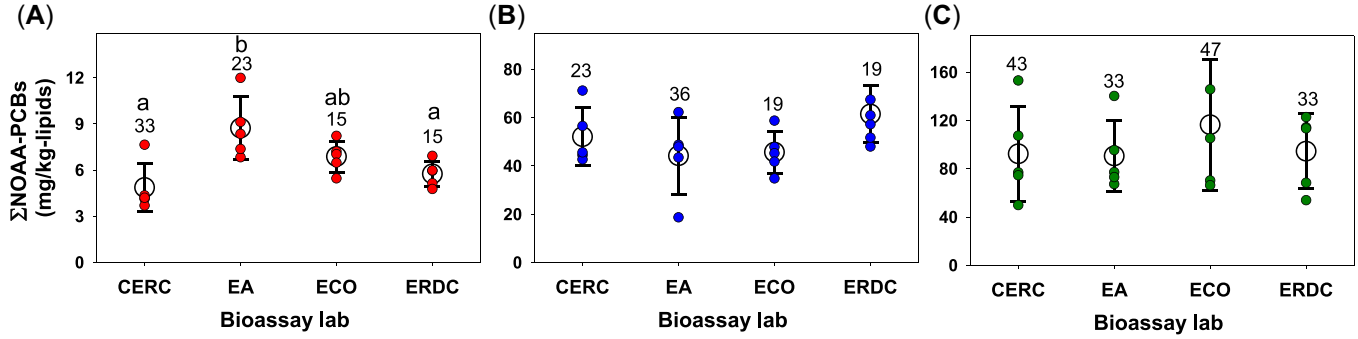

FIGURE 4: Mean concentrations of ∑NOAA-PCBs normalized by the lipid content in *Alitta virens* (**A**), *Macoma nasuta* (**B**), and *Lumbriculus variegatus* (**C**) exposed to sediment for 28 days at each bioassay laboratory. Error bars are ±1 standard deviation ($n = 5$). The number over the error bar is the coefficient of variation as a percentage. For **A**, different letters indicate significant differences from pairwise comparisons following one-way analysis of variance. For abbreviations, see footnote to Table 1.

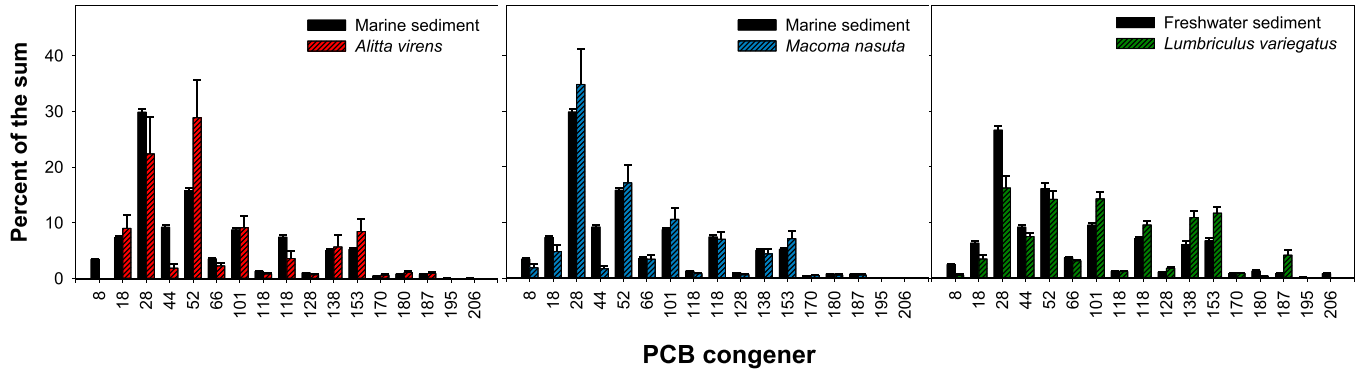

FIGURE 5: Polychlorinated biphenyl (PCB) congener profile for the bioassay sediment and for *Alitta virens* (**A**), *Macoma nasuta* (**B**), and *Lumbriculus variegatus* (**C**) exposed to sediment for 28 days. Each bar represents the mean percentage of the ∑NOAA-PCBs concentration contributed by each congener calculated using replicates from all bioassay laboratories. Error bars are ±1 standard deviation ($n = 20$).

mean organic carbon–normalized concentration for all sediment samples (i.e., a single value). Therefore, the variability just described for lipid-normalized PCB tissue concentrations applies to BSAF values. The BSAF results were compared across species. For *A. virens*, ∑NOAA-PCB BSAF values ranged from 0.2 to 1.2 (mean = 0.7, SD = 0.2) across 20 replicates. For *M. nasuta*, ∑NOAA-PCB BSAF values ranged from 1.9 to 7.2 (mean = 5.1, SD = 1.2) across 20 replicates. For *L. variegatus*, ∑NOAA-PCB BSAF values ranged from 2.6 to 10.1 (mean = 5.1, SD = 2.0) across 20 replicates. The mean ∑NOAA-PCB BSAFs for *M. nasuta* and *L. variegatus* were virtually identical and were approximately eight times higher than that for *A. virens* (Figure 6 and Supporting Information, Table S12). Congener-specific mean BSAFs varied widely for all species (Figure 6 and Supporting Information, Table S12). For *A. virens*, the lowest value (0.4) was for PCB 44 and the highest (1.3) for PCB 52. For *M. nasuta*, the lowest value (1.0) was for PCB 44 and the highest (7.2) for PCB 153. For *L. variegatus*, the lowest value (0.8) was for PCB 206 and the highest (23.4) for PCB 187, with the second highest value (9.0) occurring for PCB 138.

Summary of intralaboratory variability for bioaccumulation of PCBs

The CVs for all PCB congeners ranged widely but were relatively low for *A. virens*, *M. nasuta*, and *L. variegatus* for wet weight (mean = 24%, 15%, and 29%, respectively, $n = 68$) and lipid-normalized (mean = 26%, 25%, and 41%, respectively, $n = 68$) concentrations (Supporting Information, Tables S13 and S14). For ∑NOAA-PCBs, CVs ranged from 9% to 51% for all species for wet weight concentrations and from 15% to 47% for lipid-normalized concentrations (Supporting Information, Tables S13 and S14), indicating low overall variability across replicates for a given laboratory. Intralaboratory variability was also low overall for percentage of lipids (Supporting Information, Table S13).

Summary of interlaboratory variability for bioaccumulation of PCBs

As revealed by ANOVA, the within-treatment fraction of the total variability (i.e., the intralaboratory variability) for *A. virens*,

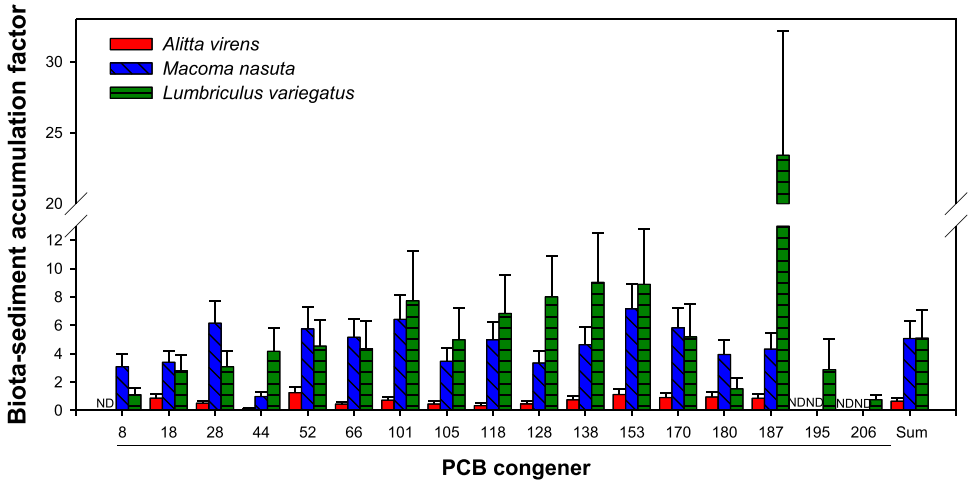

FIGURE 6: Mean biota sediment accumulation factors (BASFs) for PCB congeners and for ∑NOAA-PCBs for *Alitta virens* (**A**), *Macoma nasuta* (**B**), and *Lumbriculus variegatus* (**C**) exposed to sediment for 28 days calculated using replicates from all bioassay laboratories. Error bars are ±1 standard deviation ($n = 20$). For abbreviations, see footnote to Table 1.

M. nasuta, and *L. variegatus* was 63%, 27%, and 54%, respectively, for ∑NOAA-PCB wet weight concentrations and 44%, 83%, and 84%, respectively, for lipid-normalized concentrations (Table 1), therefore showing higher within-treatment variance compared with between-treatment (i.e., interlaboratory variability) in most cases. Differences were statistically significant for ∑NOAA-PCB wet weight concentrations for *M. nasuta* and *L. variegatus* and for *A. virens* only for lipid-normalized concentrations. The MOD ranged from 1.4 and 1.7 across species for ∑NOAA-PCB wet weight concentrations and from 1.3 to 1.8 for lipid-normalized concentrations (Table 1). Mean lipid content was remarkably similar and not statistically different for *M. nasuta*, and *L. variegatus* but was significantly different across laboratories, with 1.5 being the MOD for *A. virens* (Table 1).

Variability comparison among bioaccumulation tests

Variability in each bioaccumulation test (i.e., tests with each species) was compared using all replicate data across laboratories for lipid content and PCB congeners, and ∑NOAA-PCBs using VF. For individual congeners, VF values ranged from 1.6 to 3.9 across all species (Supporting Information, Table S15), and the mean VF was lowest for *M. nasuta* and highest for *L. variegatus* (Table 2). For lipid-normalized bioaccumulation, VF values for individual congeners ranged from 1.8 to 4.3 across all species (Supporting Information, Table S15), and the mean VF was lowest for *M. nasuta* and highest for *L. variegatus* (Table 2), showing relatively little differences in variability across species and between wet weight and lipid-normalized bioaccumulation, which is explained by the low variability for lipid content. The VF values for lipid content were lowest for *L. variegatus* and highest for *A. virens* (Table 2).

Analytical laboratory variability

The investigation of the variability associated with analyzing PCB congener concentrations for each tissue type associated with our study (i.e., whole-body homogenates of exposed *A. virens*, *M. nasuta*, and *L. variegatus*) revealed that the tissue concentrations of PCB congeners and ∑NOAA-PCBs were similar across laboratories for *A. virens* and for *M. nasuta* but were variable for *L. variegatus* tissue (Figure 7 and Supporting Information, Table S16), as shown by their respective MODs (e.g., MOD = 1.4, 1.3, and 2.5 for *A. virens*, *M. nasuta*, and *L. variegatus*, respectively; Table S16). The laboratories reported similar total lipid contents for *L. variegatus*, but variable values for *A. virens* and *M. nasuta* (Figure 7 and Supporting Information, Table S16). As a result of high variability of lipid content for tissue homogenates from all species, the concentration of lipid-normalized ∑NOAA-PCBs varied widely, as reflected in the high MOD (i.e., by 6.7, 11.0, and 4.3-fold for *A. virens*, *M. nasuta*, and *L. variegatus*, respectively; Supporting Information, Table S16).

TABLE 2: Variation factors (VFs) for percentage lipids and the average VF (VF_{avg}) for PCB congeners for wet weight and lipid-normalized concentrations derived for *Alitta virens*, *Macoma nasuta*, and *Lumbriculus variegatus* exposed to sediment for 28 days calculated using replicate data across four bioassay laboratories

		Congener average VF[a]		∑NOAA-PCBs average VF	
Bioaccumulation test	% Lipids VF	Wet wt	Lipid-norm	Wet wt	Lipid-norm
Alitta virens	2.2	2.4 ± 0.6	3.0 ± 0.9	1.9	2.3
Macoma nasuta	1.6	2.0 ± 0.3	2.3 ± 0.7	1.9	2.0
Lumbriculus variegatus	1.3	3.0 ± 0.5	4.2 ± 1.3	2.8	2.9

[a]Average ± 1 standard deviation.
∑NOAA-PCBs = sum National Oceanic and Atmospheric Administration-polychlorinated biphenyls.

Bioaccumulation of PAHs

All 16 target analyte PAHs were detected in the sediments used for our study. However, most PAHs were below the limits of detection in tissue samples (Supporting Information). Interlaboratory comparison for the bioaccumulation of PAHs was conducted for *M. nasuta* using the sum of a subset of PAHs that were detected in all replicates (Supporting Information). Intra- and interlaboratory variability were relatively low and were similar to that observed for ∑NOAA-PCBs for that species.

DISCUSSION

In the present study, the intralaboratory variability associated with whole-body residue measurements for benthic organisms exposed to the same sediment treatment under identical conditions was relatively low for ∑NOAA-PCBs, with CV ranging from 9% to 28% for all laboratories and species, with the exception of one laboratory reporting higher variability for *L. variegatus* (CV = 51%). Variance among replicates was also relatively low in previous sediment bioaccumulation studies using *M. nasuta* and *N. virens* (Boese et al., 1995; Kennedy et al., 2010; McQueen et al., 2020; Rubinstein et al., 1983). For example, after species exposure to sediment collected from the Arthur Kill (New York, NY, USA), CVs of tissue ∑PCB concentrations were 13% for *A. virens* and 19% for *M. nasuta* (Kennedy et al., 2010), and CVs of tissue ∑-coplanar-PCBs ranged from 5% to 17% for exposures to six different sediments collected from the New York–New Jersey Bight (USA; McQueen et al., 2020). Relatively low variability for PCB body residues across replicates has also been reported for *L. variegatus* for laboratory sediment bioaccumulation tests (Ankley et al., 1992; Beckingham & Ghosh, 2010; McQueen et al., 2020; Van Geest et al., 2011a, 2011b). For example, for sediments collected from 12 different locations in the Great Lakes area (USA), CVs for ∑PCB tissue concentrations ranged from 3% to 35% and were less than 25% for 92% of the sediments evaluated (McQueen et al., 2020).

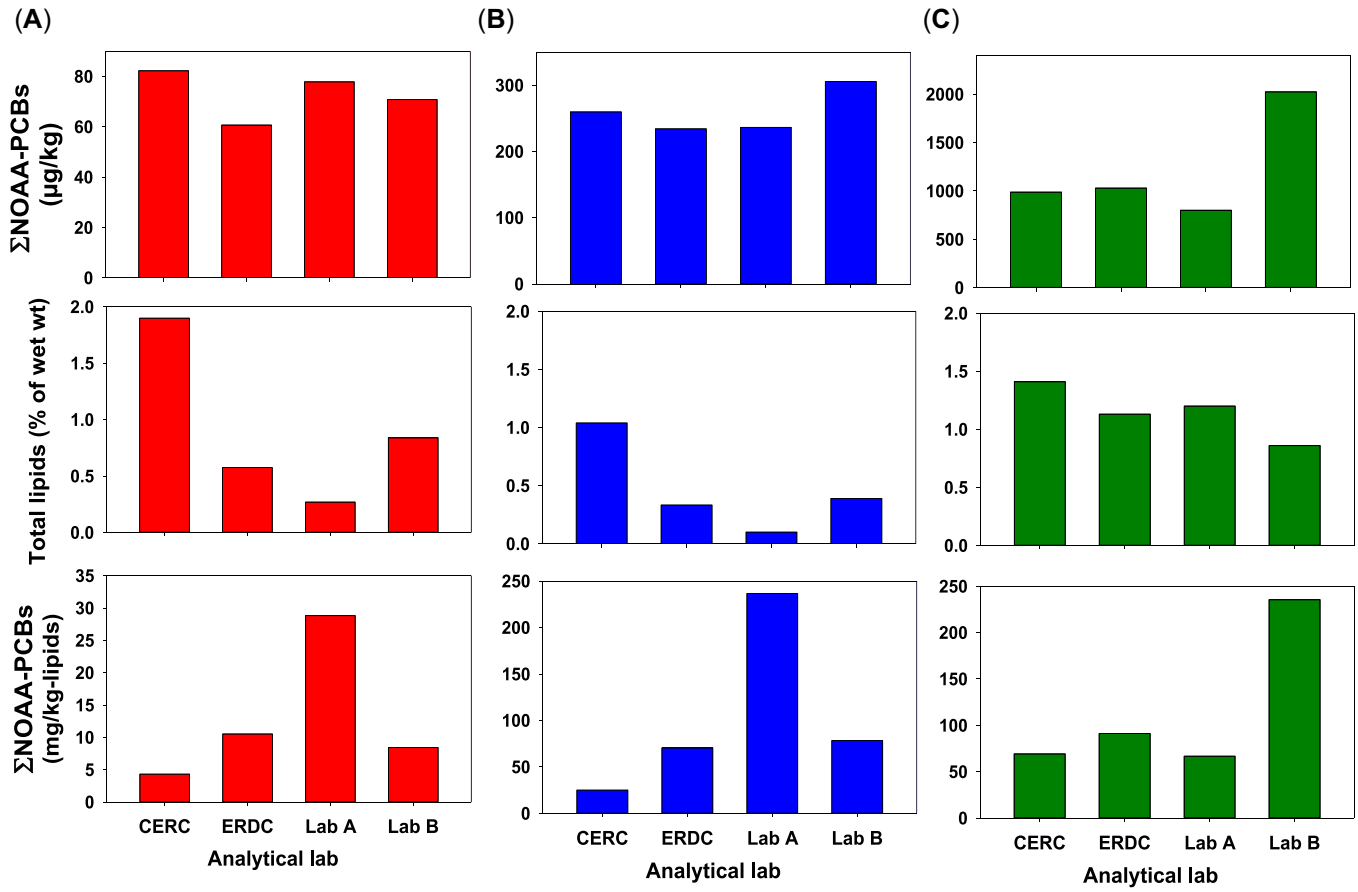

FIGURE 7: Concentrations of ∑NOAA-PCBs (top row), total lipids (as percentage of the wet wt; middle row), and lipid-normalized concentrations of ∑NOAA-PCBs (bottom row) reported by each analytical laboratory for aliquots of homogenous tissue mass for *Alitta virens* (**A**), *Macoma nasuta* (**B**), and *Lumbriculus variegatus* (**C**) exposed to sediment for 28 days. For abbreviations, see footnote to Table 1.

It has been previously unknown whether the low variability in PCB bioaccumulation test organism body residues typically obtained across replicates by a given laboratory also applies to interlaboratory variability. Our study is the first to show that sediment bioaccumulation tests commonly performed in the United States can be conducted with a low degree of variability by different laboratories. No evidence was found of a consistent bias among laboratories. Interlaboratory variability was also remarkably low. For ∑NOAA-PCBs wet weight concentrations, the MOD for laboratory means ranged from 1.4 to 2.0 across species. The MOD was low for *A. virens* and *M. nasuta* when all replicates ($n = 20$) across all laboratories instead of means were considered (2.0 and 2.1), but it was higher (3.6) for *L. variegatus*. Moreover, the intralaboratory variance was higher than the interlaboratory variance for *A. virens* and *L. variegatus*. These results are not surprising considering that all laboratories were experienced in conducting bioaccumulation tests. In addition, all laboratories followed the same basic procedures that were based on national guidance for use in the United States. The major methodological difference across laboratories was the frequency and volume of overlying water renewals. Two laboratories (ECO and EA) conducted overlying water renewals of 2 volume additions/day, whereas CERC and ERDC exchanged 50%–60% of the water volume three times weekly (i.e., the total volume of water exchanged during the test was approximately eight times higher for ECO and EA). Even though renewing the overlying water is expected to effectively remove soluble contaminants from sediments (Kennedy et al., 2010), no trend of greater bioaccumulation for laboratories that exchanged a smaller volume of water was detected, even for low-molecular-mass PCB congeners and naphthalene (Supporting Information, Tables S6, S8, S10, S17, and S18). In our study, wet samples of sediment-exposed benthic organisms were analyzed for PCB and PAH body residues. According to Egeler et al. (2006), the water adhering to the wet *L. variegatus* samples is thought to be a major source of variability in tissue residues and could have contributed to the higher variability associated with body residues for that species compared with the larger marine invertebrates used in the present study.

Although all 16 PAHs were detected in the sediments used in 28-day exposures in our study, only a few PAHs were detected in tissue samples. Adequate interlaboratory comparisons for the bioaccumulation of PAHs was only possible for *M. nasuta*. Intra- and interlaboratory variability were relatively low and similar to that observed for ∑NOAA-PCBs for that species. Variability was also relatively low (CVs of 24% or less) for total PAHs in *M. nasuta* exposed to sediments from San Diego Bay (CA, USA) using standard 28-day bioaccumulation tests (Exponent, 2003).

Despite changes in lipid content during the experiment, mostly decreasing, final lipid content varied relatively little across replicates for all species and laboratories, except for *M. nasuta* lipids for one laboratory (Figure 2). Lipid contents reported in our study are similar to those reported previously for *A. virens* (Kennedy et al., 2010; McLeese et al., 1980) and *L. variegatus* (Burkhard et al., 2015; Egeler et al., 2006; Van Geest et al., 2011a) but are lower than the range (1.0%–2.8%) previously reported for *M. nasuta* exposed to sediment for 28 days (Kennedy et al., 2010; McFarland et al., 1994). Decreases in lipid content during bioaccumulation tests have been reported previously for *M. nasuta* (Boese et al., 1995) and *L. variegatus* (Burkhard et al., 2013).

Lipids are the major compartment for the partitioning of neutral organic chemicals in animal tissues, and lipid normalization is used in calculating BSAFs, which provides valuable insights into the behavior and risks associated with toxic organic contaminants (Burkhard et al., 2012). Besides the use of lipid-normalized bioaccumulation to calculate BSAFs, it is common practice to normalize tissue concentrations for lipid content to decrease variability, because lipid content of the exposed organisms influences estimates of tissue concentrations (Burkhard et al., 2003; Di Toro et al., 1991). Intralaboratory variability (as CV) and interlaboratory variability (as MOD) were typically higher for lipid-normalized body residues compared with body residues expressed on a wet weight basis (Figures 2 and 3). After lipid normalization, the MOD for all replicates ($n = 20$) across all laboratories increased from 2.0 to 3.2 for *A. virens*, from 2.1 to 3.8 for *M. nasuta*, and from 3.6 to 3.9 for *L. variegatus*. Increase in variability by lipid normalization was also reported by Van Geest et al. (2011a) and may be at least partly explained by the expected effect of compounding the uncertainties associated with the variability that is connected to analytical measurements of two independently measured parameters. For the bioaccumulation test species investigated in the present study, an interlaboratory study was performed only for *L. variegatus* exposed to sediment amended with radiolabeled hexachlorobenzene (Egeler et al., 2006). The MOD for the mean tissue residue across five laboratories was 2.7, which was higher than the MOD (2.0) across-laboratory mean lipid-normalized ∑NOAA-PCBs for the present study.

In addition to CV and MOD values, ours study also used VF to quantify variability. According to Jonker et al. (2018), the range in body burden and lipid content can be quantified and expressed intuitively as a factor, while excluding outliers, which are not excluded using MOD. Variability was compared for lipid content, PCB congeners, and ∑NOAA-PCBs using VF calculated for all replicate data across laboratories. As expected, the VFs (Table 2 and Supporting Information, Table S18) were lower than the MODs (Table 1), and the VFs were similar for wet weight and lipid-normalized body residues for *M. nasuta* and *L. variegatus*, but higher for lipid-normalized body residues for *A. virens*. Comparison of VFs also indicated that variability was higher for *L. variegatus* than for the other two species for wet weight body residues. However, for lipid-normalized body residues, VFs were similar for *A. virens* and *L. variegatus* and lower for *M. nasuta*. Interlaboratory variability expressed by mean VFs was similar for PCB body residues derived from sediment bioaccumulation tests in our study (2.0–3.0; Supporting Information, Table S18) and for freely dissolved concentrations derived through passive sampling (2.4 and 2.6; Jonker et al., 2018). The similarity in variability for benthic organism exposure and polymer sampling is remarkable considering that uptake of PCBs in polymer samplers is a result of diffusion whereas bioaccumulation in benthic biota is a result of partitioning into multiple phases over multiple routes of exposure via sediment, diet, and overlying water compartments. In addition, other factors such as biotransformation, feeding rate, size, age, health, and others have been reported to influence bioaccumulation of organic contaminants (Diepens et al., 2015; Rust et al., 2004).

In the present study, statistically different bioaccumulation of ∑NOAA-PCBs was detected between laboratories for *M. nasuta* and *L. variegatus* (wet wt body burden) and for *A. virens* (lipid normalized) in a total of seven statistically different pairwise comparisons. The MOD for the statistically different means were 1.2, 1.3, 1.3, 1.5, 1.6, 1.8, and 2.0. Because of the low variability among replicates, statistical differences (using an α value of 0.05, selected to prevent Type II error) between laboratories occurred when the MOD was low (i.e., between 1.2 and 2.0). Small differences in contaminant body residue (i.e., MOD less than 2) have also resulted in statistically significant differences between treatments in previous sediment bioaccumulation studies using *A. virens, M. nasuta*, and *L. variegatus* and were attributed to low variability among replicates, which has been frequently observed for benthic sediment bioaccumulation testing with those species (McQueen et al., 2020). Detection of statistically significant differences between similar averages (e.g., MOD less than 2) could result in a Type I error of assuming increases in bioaccumulation when there is none. Joint USEPA and USACE guidance recommends using specific factors (e.g., MOD, toxicological importance of constituent, and propensity to biomagnify) to evaluate statistically elevated contaminants of concern from sediment bioaccumulation tests and make decisions concerning suitability for aquatic placement (USEPA & USACE, 1991, 1998). Consideration of the MOD, defined as the "magnitude by which bioaccumulation from the dredged material exceeds bioaccumulation from the reference material" (USEPA & USACE, 1991) could be a relevant element of the evaluation (McQueen et al., 2020). The MOD is advantageous because it allows the practitioner to account for uncertainty related to inherent variability of the test in the interpretation of statistically significant bioaccumulation results. Establishing a minimum MOD above which a meaningful level of bioaccumulation relative to reference sediment has been recently proposed (McQueen et al., 2020). Data sets with MODs less than 2 were considered by McQueen et al. (2020) to be within the inherent variability range of the methods associated with laboratory sediment bioaccumulation testing. The results we present for an interlaboratory evaluation of variability in sediment bioaccumulation tests corroborate the use of a MOD of 2 as a minimum relative threshold when one is applying MOD as a line of evidence in delineating an appropriate level of

bioaccumulation relative to reference sediment warranting further evaluation.

Equilibrium partitioning theory hypothesizes that the chemical activity between sediment and organisms should be equal, and that the lipid and organic carbon normalization should result in BSAFs that are similar across species (Di Toro et al., 1991). However, net bioaccumulation of contaminants from sediment is influenced by species traits such as body size, surface-area-to-volume, respiratory strategies, diet, and dietary assimilation (Gaskell et al., 2007; Rubach et al., 2011); therefore, species traits commonly cause differences in lipid-normalized concentrations in different species exposed to the same sediment. In the preset study, the mean ∑NOAA-PCB 28-day BSAF for *M. nasuta* was approximately eight times higher than that for *A. virens*. Few published studies comparing the lipid-normalized bioaccumulation of PCBs in *M. nasuta* and *A. virens* exposed to the same sediment were found. Lipid-normalized bioaccumulation of ∑PCBs in *M. nasuta* were 6 to 10 times higher than in *A. virens* exposed to sediment collected from Port Chester (New York, NY, USA; Barrows et al., 1996), but BSAFs for *M. nasuta* and *A. virens* exposed to sediment from New York–New Jersey Harbor (USA) were within a factor of 2 (Kennedy et al., 2010).

Although the focus of the present study was to investigate the interlaboratory variability associated with conducting standard sediment bioaccumulation tests, analytical variability was also evaluated by assigning different analytical laboratories a sample split of tissue homogenate for *A. virens*, *M. nasuta*, and *L. variegatus* exposed to contaminated sediment. For ∑NOAA-PCB wet weight concentrations, the variability, measured using MOD, associated with different analytical laboratories measuring the concentration associated with one sample of tissue ranged from 1.3 to 2.5 and was therefore similar to the variability of mean body residue across different bioassay laboratories. However, the variability across analytical laboratories for lipid content was high for *A. virens* and *M. nasuta*, but not for *L. variegatus*. Whereas the variability for wet weight concentration and lipid-normalized concentrations were relatively similar for *L. variegatus*, it was substantially higher for lipid-normalized concentrations for *A. virens* and *M. nasuta*. Because considerable effort was expended to ensure a homogenous tissue sample for the analytical interlaboratory comparison, the variability in lipid concentrations was likely caused by analytical variability and not differences in lipid content across subsamples. High variability in lipid content for two species but low variability for one species may be related to higher extraction efficiency associated with the softer nature of *L. variegatus* tissues. Interlaboratory comparisons for the analysis of PCBs or lipids in benthic invertebrate tissues were not found in the available literature. Although the limited evaluation in the present study suggests a relatively low variability of results experienced across laboratories, an expanded round-robin effort including a larger number of laboratories and precision measurement by each laboratory would be useful, considering that variability in analytical detection has been identified as an important potential confounding factor in bioaccumulation studies (Weisbrod et al., 2009).

The interlaboratory variability measured for PCBs in the present study may not be indicative of the variability associated with the bioaccumulation for other relevant pollutants, such as organometals (e.g., methylmercury) because the geochemical complexities influencing uptake (see Lawrence & Mason, 2001) may provide additional sources of variability. Therefore, studies of the interlaboratory variability associated with measuring the bioaccumulation of organometals using sediment bioaccumulation testing are warranted.

CONCLUSIONS

Variance among replicates was relatively low for PCBs for all test species across all laboratories. The MOD for laboratory means was 2 or less across species, indicating low interlaboratory variability. The results of our study indicate that the intralaboratory variability in measuring PCB bioaccumulation was greater than interlaboratory variability for *M. nasuta* and *A. virens*. However, the data cannot be used to determine whether this variability is introduced by multiple tests or by multiple laboratories, because multiple tests within one laboratory may produce a similar level of variability as that we measured across laboratories. A measurement of intralaboratory variability, obtained through the repeated measurement of bioaccumulation for the same sediment would provide a comparison with the interlaboratory variability measured in our study. Such a study is difficult to conduct with sediments because bioavailability may be altered by long-term sediment storage as well as the variability associated with field-collected test organisms. Analytical variability was similar to bioassay laboratory variability for the analysis of PCBs in *M. nasuta* and *A. virens* but was high for the analysis of lipid contents for those species, which warrants additional investigation. Limited results for PAHs also showed low variance among replicates. The present interlaboratory evaluation of variability for sediment bioaccumulation testing, although relatively limited in scope, supports the usefulness of MOD when one is interpreting the biological significance of statistically significant results (McQueen et al., 2020). Statistical significances associated with a low MOD similar to those reported in the present study and elsewhere may not constitute meaningful differences and may represent Type I error (McQueen et al., 2020).

Supporting Information—The Supporting information is available on the Wiley Online Library at https://doi.org/10.1002/etc.5300.

Acknowledgments—The Dredging Operations Environmental Research Program of the US Army Corps of Engineers (T. Bridges, director; D. Moore, focus area manager) funded the present study (Project 1707). The authors thank G. Saari for reviewing this manuscript.

Disclaimer—The use of trade, product, or firm names in this report is for descriptive purposes only and does not imply endorsement by the US Government. Permission was granted to publish this information by the Chief of Engineers of the US

Army Corps of Engineers. The findings of this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.The views and opinions expressed in the present study are those of the individual authors and not those of the US Army or other sponsor organizations.

Data Availability Statement—Data, associated metadata, and calculation tools are available from the corresponding author (Guilherme.Lotufo@usace.army.mil).

REFERENCES

Ankley, G. T., Cook, P. M., Carlson, A. R., Call, D. J., Swenson, J. A., Corcoran, H. F., & Hoke, R. A. (1992). Bioaccumulation of PCBs from sediments by oligochaetes and fishes—Comparison of laboratory and field studies. *Canadian Journal of Fisheries and Aquatic Sciences*, *49*, 2080–2085. https://doi.org/10.1139/f92-231

ASTM International. (2007). *Standard test method for particle-size analysis of soils*. D 422-63.

ASTM International. (2019). *Standard test method for measuring the toxicity of sediment-associated contaminants with freshwater invertebrates*. E 1706-05 (Reapproved 2019).

Barrows, E., Mayhew, H., Word, J., & Tokos, J. J. (1996). *Evaluation of dredged material proposed for ocean disposal from Port Chester, New York*. Pacific Northwest Lab. https://doi.org/10.2172/418251

Beckingham, B., & Ghosh, U. (2010). Comparison of field and laboratory exposures of *Lumbriculus variegatus* to polychlorinated biphenyl-impacted river sediments. *Environmental Toxicology and Chemistry*, *29*(12), 2851–2858. https://doi.org/10.1002/etc.353

Bligh, E. G., & Dyer, W. J. (1959). A rapid method of total lipid extraction and purification. *Canadian Journal of Biochemistry and Physiology*, *37*, 911–917. https://doi.org/10.1139/o59-099

Boese, B. L., Winsor, M., Lee, H., Echols, S., Pelletier, J., & Randall, R. (1995). PCB congeners and hexachlorobenzene biota sediment accumulation factors for *Macoma nasuta* exposed to sediments with different total organic carbon contents. *Environmental Toxicology and Chemistry*, *14*(2), 303–310. https://doi.org/10.1002/etc.5620140216

Brinkhurst, R., & Gelder, S. (1991). Annelida: Oligochaeta and Branchiobdellida. In V. H. Thorp & A. P. Covith (Eds.), *Ecology and classification of North American freshwater invertebrates* (pp. 431–463). Academic Press.

Burkhard, L. P., Arnot, J. A., Embry, M. R., Farley, K. J., Hoke, R. A., Kitano, M., Leslie, H. A., Lotufo, G. R., Parkerton, T. F., Sappington, K. G., Tomy, G. T., & Woodburn, K. B. (2012). Comparing laboratory- and field-measured biota-sediment accumulation factors. *Integrated Environmental Assessment and Management*, *8*(1), 32–41. https://doi.org/10.1002/ieam.218

Burkhard, L. P., Endicott, D. D., Cook, P. M., Sappington, K. G., & Winchester, E. L. (2003). Evaluation of two methods for prediction of bioaccumulation factors. *Environmental Science and Technology*, *37*(20), 4626–4634. https://doi.org/10.1021/es0303889

Burkhard, L. P., Hubin-Barrows, D., Billa, N., Highland, T. L., Hockett, J. R., Mount, D. R., Norberg-King, T. J., Hawthorne, S., Miller, D. J., & Grabanski, C. B. (2015). Sediment bioaccumulation test with *Lumbriculus variegatus*: Effects of feeding. *Archives of Environmental Contamination and Toxicology*, *68*(4), 696–706. https://doi.org/10.1007/s00244-015-0148-5

Burkhard, L. P., Mount, D. R., Highland, T. L., Hockett, J. R., Norberg-King, T., Billa, N., Hawthorne, S. B., Miller, D. J., & Grabanski, C. B. (2013). Evaluation of PCB bioaccumulation by *Lumbriculus variegatus* in field-collected sediments. *Environmental Toxicology and Chemistry*, *32*(7), 1495–1503. https://doi.org/10.1002/etc.2207

Ciarelli, S., Kater, B. J., & Van Straalen, N. M. (2000). Influence of bioturbation by the amphipod *Corophium volutator* on fluoranthene uptake in the marine polychaete *Nereis virens*. *Environmental Toxicology and Chemistry*, *19*(6), 1575–1581. https://doi.org/10.1002/etc.5620190614

Di Toro, D. M., Zarba, C. S., Hansen, D. J., Berry, W. J., Cowan, C. E., Pavlou, S. P., Allen, H. E., Thomas, N. A., & Paquin, P. R. (1991). Technical basis for establishing sediment quality criteria for nonionic organic chemicals using equilibrium partitioning. *Environmental Toxicology and Chemistry*, *10*, 1541–1583. https://doi.org/10.1002/etc.5620101203

Diepens, N. J., Van den Heuvel-Greve, M. J., & Koelmans, A. A. (2015). Modeling of bioaccumulation in marine benthic invertebrates using a multispecies experimental approach. *Environmental Science & Technology*, *49*(22), 13575–13585. https://doi.org/10.1021/acs.est.5b02500

Duinker, J. C., Schultz, D. E., & Petrick, G. (1988). Selection of chlorinated biphenyl congeners for analysis in environmental samples. *Marine Pollution Bulletin*, *19*(1), 19–25. https://doi.org/10.1016/0025-326X(88)90748-5

Egeler, P., Meller, M., Schallnaß, H. J., & Gilberg, D. (2006).*Validation of a sediment bioaccumulation test with endobenthic aquatic oligochaetes by an international ring test* (Technical Report FKZ 202 67 437 prepared for the German Federal Environmental Agency). German Federal Environmental Agency (Umweltbundesamt). https://ect.de/validation-of-a-sediment-bioaccumulation-test-with-endobenthic-aquatic-oligochaetes-by-an-international-ring-test/

Exponent. (2003). *Report for National Steel and Shipbuilding Company (NASSCO) and Southwest Marine. NASSCO and Southwest Marine detailed sediment investigation. Vols I, II, and III*. Prepared by Exponent NASSCO and Southwest Marine.

Finkelstein, K., Wirth, E. F., Chung, K. W., Shaddrix, B. S., Pisarski, E. C., & Rios, C. (2021). Acute polychlorinated biphenyl benthic invertebrate toxicity testing to support the 2017 chronic dose–response sediment injury model. *Environmental Toxicology and Chemistry*, *40*(4), 1188–1193. https://doi.org/10.1002/etc.4977

Fredette, T. J., Guza, O., Farris, C. N., Lefkovitz, L. F., & Naber, S. J. (2007). Exploring relationships in a sediment-tissue bioaccumulation database. *Proceedings of the Western Dredging Association Twenty-Sixth Technical Conference and Thirty-Eighth Texas A&M Dredging Seminar, 26–28 June 2006, San Diego, CA* (pp. 349–361). Center for Dredging Studies, Ocean Engineering Program, Civil Engineering Department, Texas A&M University. https://www.westerndredging.org/index.php/information/category/43-26th-annual-weda-conference?download=393:11-fredette-exploring-relationships-in-a-sediment-tissue-bioaccumulation-database

Gale, R. W., Orazio, C. E., & McKee, M. J. (2009). *Determination of polychlorinated biphenyls, selected persistent organochlorine pesticides, and polybrominated flame retardants in fillets of fishes from the 2007 Missouri Department of Conservation Monitoring Programs* (US Geological Survey Open-File Report, 1053). p. 20. https://pubs.usgs.gov/of/2008/1028/index.html

Gaskell, P. N., Brooks, A. C., & Maltby, L. (2007). Variation in the bioaccumulation of a sediment-sorbed hydrophobic compound by benthic macroinvertebrates: Patterns and mechanisms. *Environmental Science and Technology*, *41*(5), 1783–1789. https://doi.org/10.1021/es061934b

Gidley, P. T., Kennedy, A. J., Lotufo, G. R., Wooley, A. H., Melby, N. L., Ghosh, U., Burgess, R. M., Mayer, P., Fernandez, L. A., Schmidt, S. N., Wang, A. P., Bridges, T. S., & Ruiz, C. E. (2019). Bioaccumulation in Functionally Different Species: Ongoing Input of PCBs with Sediment Deposition to Activated Carbon Remediated Bed Sediments. *Environmental Toxicology and Chemistry*, *38*, 2326–2336. https://doi.org/10.1002/etc.4526

Hoke, R., Huggett, D., Brasfield, S., Brown, B., Embry, M., Fairbrother, A., Kivi, M., Paumen, M. L., Prosser, R., Salvito, D., & Scroggins, R. (2016). Review of laboratory-based terrestrial bioaccumulation assessment approaches for organic chemicals: Current status and future possibilities. *Integrated Environmental Assessment and Management*, *12*(1), 109–122. https://doi.org/10.1002/ieam.1692

Hylleberg, J., & Gallucci, V. V. (1975). Selectivity in feeding by the deposit-feeding bivalve *Macoma nasuta*. *Marine Biology*, *32*, 167–178. https://doi.org/10.1007/BF00388509

Ingersoll, C. G., Brunson, E. L., Dwyer, F. J., Ankley, G. T., Benoit, D. A., Norberg-King, T. J., Burton, G. A., Hoke, R. A., Landrum, P. F., & Winger, P. V. (1995). Toxicity and bioaccumulation of sediment-associated contaminants using freshwater invertebrates: A review of methods and applications. *Environmental Toxicology and Chemistry*, *14*, 1885–1894. https://doi.org/10.1002/etc.5620141110

Jonker, M. T. O., van der Heijden, S. A., Adelman, D., Apell, J. N., Burgess, R. M., Choi, Y., Fernandez, L. A., Flavetta, G. M., Ghosh, U., Gschwend, P. M., Hale, S. E., Jalalizadeh, M., Khairy, M., Lampi, M. A., Lao, W., Lohmann, R., Lydy, M. J., Maruya, K. A., Nutile, S. A., … Wu, Y. (2018). Advancing the use of passive sampling in risk assessment and management of sediments contaminated with hydrophobic organic

chemicals: Results of an international ex situ passive sampling interlaboratory comparison. *Environmental Science and Technology*, *52*(6), 3574. https://doi.org/10.1021/acs.est.7b05752

Keith, L. H. (2015). The source of U.S. EPA's sixteen PAH priority pollutants. *Polycyclic Aromatic Compounds*, *35*(2–4), 147–160. https://doi.org/10.1080/10406638.2014.892886

Kennedy, A. J., Lotufo, G. R., Steevens, J. A., & Bridges, T. S. (2010). *Determining steady-state tissue residues for invertebrates in contaminated sediment. Dredging Operations and Environmental Research Program* (ERDC/EL TR-10-2). U.S. Army Engineer Research and Development Center. https://apps.dtic.mil/sti/citations/ADA529313

Knight, S. S., Lizotte, R. E., & Shields, F. D. (2009). *Hyalella azteca* (Saussure) Responses to Coldwater River Backwater Sediments in Mississippi, USA. *Bulletin of Environmental Contamination and Toxicology*, *83*, 493–496. https://doi.org/10.1007/s00128-009-9804-7

Lauenstein, G. G., & Cantillo, A. Y. (1993). Sampling and analytical methods of the National Status and Trends Program. National Benthic Surveillance and Mussel Watch Projects, 1984–1992. NOAA Technical Memorandum NOS ORCA 71. National Oceanic and Atmospheric Administration. https://repository.library.noaa.gov/view/noaa/2893/noaa_2893_DS1.pdf?

Lawrence, A. L., & Mason, R. P. (2001). Factors controlling the bioaccumulation of mercury and methylmercury by the estuarine amphipod *Leptocheirus plumulosus*. *Environmental Pollution*, *111*(2), 217–231. https://doi.org/10.1016/S0269-7491(00)00072-5

Lotufo, G. R., Burton, G. A., Rosen, G., Fleeger, & J. W. (2014). Assessing biological effects, processes, assessment and remediation of contaminated sediments assessment and remediation of contaminated sediments (pp. 131–176). Springer Science and Business Media. https://doi.org/10.1007/978-1-4614-6726-7

Lotufo, G. R., Melby, N. L., Biedenbach, J. M., Moore, D. W., & Farrar, J. D. (2021). Streamlining freshwater bioaccumulation bioassays: Letting the worms do the hard work. *Environmental Toxicology and Chemistry*, *40*(6), 1673–1677. https://doi.org/10.1002/etc.5019

McElroy, A. E., Barron, M. G., Beckvar, N., Driscoll, S. B. K., Meador, J. P., Parkerton, T. F., Preuss, T. G., & Steevens, J. A. (2011). A review of the tissue residue approach for organic and organometallic compounds in aquatic organisms. *Integrated Environmental Assessment and Management*, *7*(1), 50–74. https://doi.org/10.1002/ieam.132

McFarland, V. A., Clarke, J. U., Lutz, C. H., Jarvis, A. S., Mulhearn, B., Reilly Jr., F. J. (1994). *Bioaccumulation potential of contaminants from bedded and suspended Oakland Harbor Deepening Project sediments to San Francisco Bay flatfish and bivalve mollusks* (Miscellaneous Paper EL-94-7). U.S. Army Engineer Waterways Experiment Station. https://apps.dtic.mil/sti/citations/ADA284095

McLeese, D. W., Metcalfe, C. D., & Pezzack, D. S. (1980). Uptake of PCBs from sediment by Nereis virens and Crangon septemspinosa. *Archives of environmental contamination and toxicology*, *9*(5), 507–518. https://doi.org/10.1007/BF01056931

McQueen, A. D., Lotufo, G. R., Pickard, S. W., Lenox, A. M., Moore, D. W., von Stackelberg, K., & Suedel, B. C. (2020). Evaluation of dredged sediment for aquatic placement: Interpreting contaminant bioaccumulation. *Environmental Monitoring and Assessment*, *192*(5), 277. https://doi.org/10.1007/s10661-020-8236-z

Nelson, W. G., & Bergen, B. J. (2012). The New Bedford Harbor Superfund site long-term monitoring program (1993–2009). *Environmental Monitoring and Assessment*, *184*(12), 7531–7550. https://doi.org/10.1007/s10661-012-2517-0

Northwest Regional Sediment Evaluation Team. (2018). *Sediment evaluation framework for the Pacific Northwest*. Prepared by the RSET Agencies. https://usace.contentdm.oclc.org/digital/collection/p16021coll11/id/684/

Organization for Economic Co-operation and Development (OECD). (2008). *Test No. 315: Bioaccumulation in sediment-dwelling benthic oligochaetes OECD Guidelines for the Testing of Chemicals*. https://doi.org/10.1787/9789264067516-en.

Penttinen, O. P., Kukkonen, J., & Pellinen, J. (1996). Preliminary study to compare body residues and sublethal energetic responses in benthc invertebrates exposed to sediment bound 2,4,5-trichlorophenol. *Environmental Toxicology and Chemistry*, *15*(2), 160–166. https://doi.org/10.1002/etc.5620150214

Rubach, M. N., Ashauer, R., Buchwalter, D. B., De Lange, H. J., Hamer, M., Preuss, T. G., Töpke, K., & Maund, S. J. (2011). Framework for traits-based assessment in ecotoxicology. *Integrated Environmental Assessment and Management*, *7*(2), 172–186. https://doi.org/10.1002/ieam.105

Rubinstein, N. I., Lores, E., & Gregory, N. R. (1983). Accumulation of PCBs, mercury and cadmium by *Nereis virens, Mercenaria mercenaria* and *Palaemonetes pugio* from contaminated harbor sediments. *Aquatic Toxicology*, *3*(3), 249–260. https://doi.org/10.1016/0166-445X(83)90045-0

Rust, A. J., Burgess, R. M., Brownawell, B. J., & McElroy, A. E. (2004). Relationship between metabolism and bioaccumulation of benzo [α] pyrene in benthic invertebrates. *Environmental Toxicology and Chemistry*, *23*(11), 2587–2593. https://doi.org/10.1897/03-354

Schmidt, S. N., Wang, A. P., Gidley, P. T., Wooley, A. H., Lotufo, G. R., Burgess, R. M., Ghosh, U., Fernandez, L. A., & Mayer, P. (2017). Cross validation of two partitioning-based sampling approaches in mesocosms containing PCB contaminated field sediment, biota, and activated carbon amendment. *Environmental Science and Technology*, *51*(17), 9996–10004. https://doi.org/10.1021/acs.est.7b01909

Selck, H., Drouillard, K., Eisenreich, K., Koelmans, A. A., Palmqvist, A., Ruus, A., Salvito, D., Schultz, I., Stewart, R., Weisbrod, A., van den Brink, N. W., & van den Heuvel-Greve, M. (2012). Explaining differences between bioaccumulation measurements in laboratory and field data through use of a probabilistic modeling approach. *Integrated Environmental Assessment and Management*, *8*(1), 42–63. https://doi.org/10.1002/ieam.217

Steevens, J. A., Kennedy, A. J., Farrar, J. D., Bridges, T. S., Reiss, M. R., kropp, R. K., & Doi, J. (2008). *Dredged material analysis tools: Performance of acute and chronic sediment toxicity methods* (U.S. Army Engineer Research and Development Center Report ERDC/EL TR- 08-16). https://apps.dtic.mil/sti/citations/ADA485346

Swartz, R. C. (1999). Consensus sediment quality guidelines for polycyclic aromatic hydrocarbon mixtures. *Environmental Toxicology and Chemistry*, *18*(4), 780–787. https://doi.org/10.1002/etc.5620180426

U.S. Environmental Protection Agency. (1988). Determination of total organic carbon in sediment (Lloyd Khan Method). Environmental Service Division.

U.S. Environmental Protection Agency. (1994a). Method 3640A: Gel permeation cleanup.

U.S. Environmental Protection Agency. (1996b). Method 3541 (SW-846): Automated Soxhlet extraction.

U.S. Environmental Protection Agency. (1996a). Method 3630C (SW-846): Silica gel cleanup.

U.S. Environmental Protection Agency. (1996b). Method 3665A (SW-846): Sulfuric acid/permanganate cleanup.

U.S. Environmental Protection Agency. (1996c). Method 3660B: Sulfur cleanup (revision 2).

U.S. Environmental Protection Agency. (2000*). Methods for measuring the toxicity and bioaccumulation of sediment-associated contaminants with freshwater invertebrates* (2nd ed., EPA 600/R-99/064).

U.S. Environmental Protection Agency. (2007a). Method 3545A (SW-846): Pressurized fluid extraction (PFE) (revision 1).

U.S. Environmental Protection Agency. (2007b). Method 3550C (SW-846): Ultrasonic extraction.

U.S. Environmental Protection Agency. (2007c). Method 8082A (SW-846): Polychlorinated biphenyls (PCBs) by gas chromatography (revision 1).

U.S. Environmental Protection Agency. (2016). Weight of evidence in ecological assessment, EPA/100/R-16/001. Washington, DC, USA.

U.S. Environmental Protection Agency. (2018). Method 8270E (SW-846): Semivolatile organic compounds by gas chromatography/mass spectrometry (GC/MS) (revision 6).

U.S. Environmental Protection Agency. (2011). 40 CFR Appendix B to Part 136. Guidelines for establishing test procedures for the analysis of pollutants. Definition and procedure for the determination of the method detection limit. Revision 1.11. Subchapter D. Water Programs.

U.S. Environmental Protection Agency & US Army Corps of Engineers. (1991). *Evaluation of dredged material proposed for ocean disposal (testing manual)* (EPA 503/8-91/001). https://www.epa.gov/sites/default/files/2015-10/documents/green_book.pdf

U.S. Environmental Protection Agency & US Army Corps of Engineers. (1998). *Evaluation of dredged material proposed for disposal in waters of the U.S.—Testing manual (inland testing manual)* (EPA-823-B-98-O04). Office of Water, Office of Science and Technology, and Operation, Construction, and Readiness Division, USEPA and Department of U.S. Army Corps of Engineers. https://www.epa.gov/sites/default/files/2015-08/documents/inland_testing_manual_0.pdf

U.S. Environmental Protection Agency & US Army Corps of Engineers. (2008). *Southeast regional implementation manual (SERIM) for*

requirements and procedures for evaluation of the ocean disposal of dredged material in southeastern U.S. Atlantic and Gulf Coast waters (EPA 904-B-08-001). U.S. Environmental Protection Agency Region 4 and U.S. Army Corps of Engineers, South Atlantic Division.

Van Geest, J. L., Poirier, D. G., Sibley, P. K., & Solomon, K. R. (2010). Measuring bioaccumulation of contaminants from field-collected sediment in freshwater organisms: A critical review of laboratory methods. *Environmental Toxicology and Chemistry*, *29*(11), 2391–2401.

Van Geest, J., Poirier, D., Solomon, K., & Sibley, P. (2011a). The effect of organism density on bioaccumulation of contaminants from sediment in three aquatic test species: A case for standardizing to sediment organic carbon. *Archives of Environmental Contamination and Toxicology*, *60*(4), 626–635. https://doi.org/10.1007/s00244-010-9573-7

Van Geest, J. L., Poirier, D. G., Solomon, K. R., & Sibley, P. K. (2011b). A comparison of the bioaccumulation potential of three freshwater organisms exposed to sediment-associated contaminants under laboratory conditions. *Environmental Toxicology and Chemistry*, *30*(4), 939–949. https://doi.org/10.1002/etc.456

Weisbrod, A. V., Woodburn, K. B., Koelmans, A. A., Parkerton, T. F., McElroy, A. E., & Borga, K. (2009). Evaluation of bioaccumulation using in vivo laboratory and field studies. *Integrated Environmental Assessment and Management*, *5*(4), 598–623. https://doi.org/10.1897/IEAM_2009-004.1

