Biogeographic classification of streams using fish community– and trait–environment relationships

Matthew J. Troia1 | Ryan A. McManamay2

1Department of Environmental Science and Ecology, The University of Texas at San Antonio, San Antonio, Texas
2Department of Environmental Science, Baylor University, Waco, Texas

Correspondence
Ryan A. McManamay, Department of Environmental Science, Baylor University, Baylor Sciences Building, BSB A. 409, One Bear Place #97266, Waco, Texas 76798-7266. Email: Ryan_McManamay@baylor.edu

Funding information
U.S. Department of Energy, Grant/Award Number: DE-AC05-00OR22725; Office of Energy Efficiency and Renewable Energy; Water Power Technologies Office

Editor: Hughes Alice

Abstract
Aim: Present a hybrid biogeographic and environmentally constrained clustering approach to classify ~853,000 stream reaches in the eastern United States. Examine the frequency of stream typologies in the landscape relative to anthropogenic stressors to identify potential stream conservation needs.

Location: Eastern United States.

Methods: Fish communities at 956 least-disturbed sampling reaches were characterized using taxonomic and functional composition of native species. Environmental variables summarized within stream reaches included stream discharge, channel gradient, hydrologic regime, summer water temperature and boundaries of freshwater ecoregions. Multivariate regression trees were used to relate environmental variables to fish communities while simultaneously developing taxa- and trait-specific biogeographic classifications. We then overlaid stream classifications with indices of anthropogenic disturbances to evaluate rarity and risk of complete loss of natural representation of stream types.

Results: Taxonomically based classes represented a combination of ecoregional boundaries and environmental gradients, whereas temperature gradients were most important in differentiating functional composition. An optimal taxonomy-based classification contained 13 classes and explained 26.4% of community variation, and the optimal trait-based classification contained 6 classes, explaining 25.2% of community variation. Overlaying class maps with an anthropogenic disturbance index revealed substantial variation in the severity of degradation among classes relative to their abundance in the landscape, especially rare classes (i.e., large rivers and cold headwater systems).

Main conclusions: We demonstrate the value of integrating local, regional and historical biogeographic factors with taxonomic and functional characteristics of freshwater biota to classify ecosystems based on mechanistic biota–environment relationships. This approach is important to addressing the challenge of developing stream classifications that reconcile the environmental drivers of community trait composition with the constraints of biogeographic regionality at large spatial scales. Furthermore, the spatially explicit and high-resolution maps disseminated here can

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited. © 2019 The Authors. Diversity and Distributions published by John Wiley & Sons Ltd
1 | INTRODUCTION

Streams and rivers display extraordinary complexity in their physicochemical characteristics. This complexity varies from reach to reach and arises from diverse geological, topographic, and land cover characteristics that emerge locally within the reach or remotely from the upstream catchment (Benda et al., 2004; Ward, Tockner, Arscott, & Clare, 2002). Such physicochemical diversity translates to equally diverse lotic biotas occupying stream ecosystems. Despite freshwater ecosystems being among the most biodiverse ecosystems in the world, they are also the most imperiled, consequently from societal water demands and other adverse anthropogenic activities (Dudgeon et al., 2006). As such, water resource managers and biodiversity conservationists are challenged with the need to adequately define and map spatial variation in the physicochemical and biological features of stream ecosystems. Such efforts are specifically necessary for identifying stream types that will respond to anthropogenic stressors, stratifying sample locations for monitoring programmes and prioritizing conservation or habitat preservation projects (Hawkins et al., 2000; Hayes et al., 2003; Pressey et al., 2000; Roux et al., 2008).

Stream classification schemes offer a means to simplify and generalize the ecological complexity of streams for the purpose of water resources management and biodiversity conservation. Approaches to stream classification are diverse and include schemes seeking to partition streams into homogenous groups (i.e., classes) based on geographic proximity (Omernik, 1995) or similarity in abiotic (Hawkins et al., 1993) or biotic characteristics (Barbour et al., 1996; Herlihy, Hughes, & Sifneos, 2006). For example, ecoregion approaches aggregate adjacent drainages based on faunal or floral differences stemming from region-specific habitat differences or biogeographic histories (Abell et al., 2008). Alternatively, abiotic habitat approaches seek to characterize spatial variation in physicochemical conditions, particularly channel geomorphology (Rosgen, 1994) and hydrology (Poff & Ward, 1989). This habitat template approach implicitly incorporates species-environment relationships by using abiotic variables known (or hypothesized) to affect the distribution of organisms or the rates of ecological processes (Townsend & Hildrew, 1994).

If the goal of stream classification is to capture abiotic factors that are causally linked to biota and ecological processes, schemes should explicitly incorporate species-environment relationships. Several recently developed statistical techniques have facilitated the application of such species-environment classification schemes (De’Ath & Fabricius, 2000; De’Ath 2002, Ferrier, Drielsma, Manion, & Watson, 2002). For example, Brenden, Wang, and Seelbach (2008) used multivariate regression trees to quantify the relationship between fish community composition and reach-scale environmental conditions, specifically summer temperature, in streams of Michigan, USA. Similarly, Leathwick et al., (2011) used generalized dissimilarity modelling to classify streams based on the response of fish and macroinvertebrate communities to reach-scale physicochemical conditions in New Zealand streams. Both of these classifications were developed at modest spatial extents (i.e., 250,000 and 268,000 km², respectively), which tends to constrain habitat and taxonomic diversity. For broader-scale classification (i.e., >1,000,000 km²) encompassing numerous drainage basins, it becomes necessary to consider not only reach-scale physicochemical factors, but also regional variation in climate and historical biogeographic boundaries (Abell et al., 2008; Olden & Kennard, 2010; Tonn, Magnuson, Rask, & Toivonen, 1990). Another advantage of these species-environment approaches to stream classification is that models can be developed using discrete sampling sites and, assuming environmental predictors are GIS-derived (e.g., National Hydrography Data set; USEPA, 2005), extrapolated to all confluence-to-confluence stream reaches in the study area (Brenden et al., 2008; Leathwick et al., 2011). This provides spatially contiguous maps which can be used to estimate the prevalence of stream classes and identify anthropogenic land use threats.

Using multivariate species-environmental relationships provides a novel approach to simultaneously understand the importance of environmental variables in explaining biogeographic patterns while also directly harnessing the variation of these relationships in clustering. The application of trait-based approaches can further improve the mechanistic underpinnings of species-environment relationships by offering a means to elucidate causal links between abiotic conditions and organismal vital rates (McGill et al., 2007; Frimpong & Angermeier, 2010). Moreover, comparing trait-environment relationships with species-environment relationships can shed light on the relative roles of environmental filtering and historical biogeography structuring biological communities (Hoeinghaus, Winemiller, & Birnbaum, 2007). Although these advantages of trait-based approaches have seen increased application in basic ecology and freshwater ecology in recent years, the practice of incorporating trait-based ecology in stream classifications has seen limited attention (but see Tolonen, Leinonen, Marttila, Erkinaro, & Heino, 2017). The overall aim of this study was to explore stream classification schemes for the eastern United States. Such tools that facilitate uniquely inform management decisions regarding anthropogenic threats to the biodiverse freshwaters of the eastern United States.

KEYWORDS

conservation prioritization, fish communities, GIS mapping, multivariate regression trees, stream classification, trait-based ecology
the understanding and management of freshwater biodiversity are greatly needed for this region as it harbours the richest temperate freshwater fauna in the world, particularly fishes, mussels and crayfishes (Abell et al., 2008; Haag, 2012; Taylor et al., 1996). This conservation concern is further exacerbated by the rapid increase in the human population in the eastern United States (Radeloff et al., 2012) and its concomitant stress on stream habitat and water resources (Richter, Mathews, Harrison, & Wigington, 2003). A specific objective was to identify causal environmental and biogeographic drivers by comparing classification schemes based on taxonomic versus functional composition of fish communities. We hypothesized that taxonomic composition would correlate with historical biogeographic boundaries that separate disparate evolutionary lineages (sensu Abell et al., 2008), whereas functional composition would correlate with environmental filters operating convergently across drainage divides (sensu Lamouroux, Poff, & Angermeier, 2002). A
second objective was to assign classes to the ~ 853,000 confluence-to-confluence stream reaches in the eastern United States, estimate the prevalence of classes, and evaluate the degree to which each class has been degraded by anthropogenic disturbance. It is our goal that public dissemination of high-resolution (i.e., reach scale) maps will inform decision-making in water resource policy and biodiversity conservation planning (sensu McManamay et al., 2018).

2 | METHODS

2.1 | Fish community data

We used previously compiled fish community data from federally administered stream fish surveys (Troia & McManamay, 2017). Similar sources of fish distributional data have been collated previously and used as a comprehensive community data set in national-scale studies of stream fish biogeography, ecology and conservation (Giam & Olden, 2016; Herlihy et al., 2006; Mitchell & Knouft, 2009). For our study, we compiled records within National Hydrology Dataset (NHD) hydrologic regions 01 through 06 (Figure 1a). We used ArcGIS (version 10.3, ESRI Inc.) to join the Cartesian coordinates of each site to NHD version 1 flowlines (i.e., confluence-to-confluence stream reaches; hereafter ‘NHD reaches’) and their associated unique identifier (i.e., COMID) to link GIS-derived environmental data layers to fish community data.

Because we were interested in describing classes based on native fish communities responding to natural environmental conditions, we retained only survey sites that were minimally impacted by anthropogenic activities. To accomplish this, we removed sites with ‘high’ or ‘very high’ cumulative disturbance indices according to the National Fish Habitat Action Plan (NFHAP; Esselman et al., 2011). We also removed species from each site that were not historically native to the NHD sub-basin (i.e., HUC8) in which that site was located. Lists of historically native species were acquired from the NatureServe Explorer data set (NatureServe, 2010). This procedure removed species that are not native to the study region (e.g., species of Eurasian origin) as well as native invaders (sensu Scott & Helfman, 2001) that have increased in prevalence due to habitat alteration and sportfish stocking.

2.2 | Trait data

We acquired existing or developed new trait data describing reproductive life-history, body morphology, trophic guild membership and temperature guild membership of native species represented in the community surveys (Table 1). These traits exemplify important axes of functional variation among North American freshwater fish species and correlate with environmental gradients associated with broad-scale climate and physiography as well as reach-scale environmental conditions associated with stream network position and catchment characteristics (Frimpong & Angermeier, 2010).

Life-history and trophic guild data were acquired from the FishTraits database (Frimpong & Angermeier, 2009). For species lacking trait values, we used NatureServe Explorer (NatureServe, 2010), FishBase (Froese & Pauly, 2013), and traditional literature searches to acquire missing trait values or identify the closest phylogenetic relative for surrogate trait values. The closest phylogenetic relative was determined as the closest clade or parental clade (subgenus), species of potential hybridization or species commonly misidentified as the species of interest (in that order of preference). For trophic traits, we coded species as herbivorous, omnivorous, invertivorous or carnivorous using a separate

| Trait category | Source | Trait | Unit |
|----------------|--------|-------|------|
| Reproductive life history | FishTraits database | Opportunistic strategist | Soft classification score |
| | | Periodic strategist | Soft classification score |
| | | Equilibrium strategist | Soft classification score |
| Body morphology | Lateral-view photographs | Geometric morphometric axis 1 | Axis score |
| | | Geometric morphometric axis 2 | Axis score |
| | | Geometric morphometric axis 3 | Axis score |
| Trophic ecology | Literature review | Herbivore | Binary dummy variable |
| | | Omnivore | Binary dummy variable |
| | | Invertivore | Binary dummy variable |
| | | Carnivore | Binary dummy variable |
| Temperature guild | Literature review | Warm-water | Binary dummy variable |
| | | Cool-water | Binary dummy variable |
| | | Cold-water | Binary dummy variable |

1 Frimpong and Angermeier (2009).  
2 Table S1.  
3 McManamay (unpublished literature review).
dummy (i.e., binary) variable for each of the four trophic guilds. For life-history traits, we used the trilateral life-history framework of Winemiller and Rose (1992). This framework describes trade-offs in age at maturity, fecundity and parental care resulting in three strategic endpoints: opportunistic, periodic and equilibrium. We used a soft classification scheme whereby species were plotted in trivariate life-history space and Euclidean distances between the species and each of the three strategic endpoints were calculated (Mims & Olden, 2012). This method produced three continuous variables representing a species’ affinity for each of the three life-history strategies. Temperature guild data were acquired from Lyons et al., (2009). We classified species as warm-water, cool-water or cold-water using a separate dummy (i.e., binary) variable for each of the three temperature guilds. Morphological traits were developed from lateral-view photographs of fish specimens that we digitized from field guides (Table S1). We used geometric morphometric analysis to characterize variation in body shape and fin origins (Franssen, Goodchild, & Shepard, 2015). We placed sixteen homologous landmarks (Figure 2a) on each specimen using tpsDig software (Rohlf, 2004a). Next, we performed a generalized Procrustes analysis to resize, translate and rotate landmarks followed by a relative warp analysis using tpsRelw software (Rohlf, 2004b). This method produced relative warp (i.e., ordination axis) scores for each species, the first three of which were retained as
TABLE 2  Data sources for seven biogeographic or environmental variables used as environmental constraints in taxonomy- and function-based stream classifications

| Predictor | Variable type | Categories or variable range | Data source |
|-----------|---------------|-----------------------------|-------------|
| Freshwater ecoregion | Categorical | Laurentian Great Lakes; St. Lawrence; Teays–Old Ohio; Cumberland; Tennessee; Mobile Bay; West Florida Gulf; Apalachicola; Florida Peninsula; Appalachian Piedmont; Chesapeake Bay | Abell et al. (2008) |
| Mean annual flow | Continuous | Cubic feet per second | NHD Plus version 1 |
| NHD reach gradient | Continuous | 0–35 m/stream km | NHD Plus version 1 |
| Summer temperature | Continuous | 12–32°C | McManamay et al. (2018) |
| Hydrology PC1 | Continuous | −1.8 to +1.8 (axis score) | McManamay et al. (2014), this study |
| Hydrology PC2 | Continuous | −2.0 to +0.6 (axis score) | McManamay et al. (2014), this study |
| Hydrology PC3 | Continuous | −1.5 to +1.5 (axis score) | McManamay et al. (2014), this study |

morphological variables describing interspecific variation in functional morphology (Figure 2b).

Analyses of trait–environmental relationships can violate the statistical assumption of independence because trait values among species are not independent of one another owing to shared evolutionary history (Felsenstein, 1985). Controlling for phylogeny ensures spatial variation in trait composition is attributable to environmental variation and not confounded by common lineage of species pools. To control for phylogeny, we used generalized linear models to predict each trait from family number based on Gaussian distributions (i.e., for continuous traits) or Poisson distributions (i.e., for binary traits) following McManamay and Frimpong (2015). For each trait model, deviance residuals were extracted for each species and used in subsequent trait analyses.

2.3 Biogeographic and environmental data

We acquired existing or developed novel GIS data layers describing biogeographic realms, stream discharge, channel gradient, summer temperature and hydrology of the ∼853,000 NHD reaches located in the eastern United States (Table 2). We selected these environmental factors because they influence taxonomic and functional composition of stream fishes (Lamouroux et al., 2002; Lyons et al., 2009; Olden & Kennard, 2010; Xenopoulos & Lodge, 2006). Moreover, each variable is available at the segment resolution for the eastern United States either directly from the NHD or can be accurately modelled using natural landscape characteristics available from the NHD. Although many anthropogenic factors (e.g., catchment land use and hydrologic alteration) influence stream fish communities, we selected only natural environmental factors to correspond with the native community data with the goal of characterizing natural community–environment relationships and developing natural classes.

Biogeographic realms for obligate freshwater organisms such as fish are defined by historical drainage and glaciation patterns that limited dispersal across drainage divides. We used the freshwater ecoregional map produced by Abell et al. (2008) as a categorical predictor variable with 12 discrete biogeographic realms spanning the eastern US study area (Table 2, Figure 1c). Stream discharge is a measure of habitat size (Xenopoulos & Lodge, 2006) and correlates with the origin and diversity of food resources available to fishes (Vannote, Minshall, Cummins, Sedell, & Cushing, 1980). We used mean annual discharge as a measure of stream network position, which was modelled previously using the Unit Runoff Method and is available for all NHD reaches (NHDPlus V1, 2010) (Figure 1d). Channel gradient correlates with channel hydraulics, bedform and substrate size (Montgomery & Buffington, 1997) and is known to affect the community composition of fishes (Lamouroux et al., 2002). We acquired channel gradient for all NHD reaches, which was an existing attribute of the NHD (Figure 1e). Mean summer temperature is a key physicochemical variable directly linked to fish community composition along gradients of stream size, elevation and latitude (Lyons et al., 2009; Rahel & Hubert, 1991). Mean summer (July–August) stream temperature from >800 monitoring sites were previously compiled and used to develop a predictive model to extrapolate summer temperatures to all NHD reaches (Figure 1e; McManamay et al., 2018). Hydrology is a putative determinant of fish community composition (Poff et al., 1997). McManamay, Bevelhimer, and Kao (2014) previously compiled records of mean daily flow from 643 reference-quality USGS flow gauges and computed 110 metrics describing the magnitude, timing, frequency, duration and rate of change in flows. Using this information, we conducted a PCA to reduce these metrics to three principal component axes and extracted site (i.e., USGS gauge) scores for each component (Table S2). We used Random Forests to model principal components using 166
natural landscape predictors available for all NHD reaches (Table S3). This algorithm is useful for characterizing non-linear relationships between predictor variables and the response variable as well as modelling high order and complex interactions among predictor variables (Cutler et al., 2007). Because predictor variables were GIS-derived and available for each of the ~853,000 NHD reaches in the eastern US, the Random Forests model was used to extrapolate component scores to all NHD reaches (Figure 1g–i).

2.4 | Stream classification analyses

We used multivariate regression trees (MRT) to develop discrete classes based on community–environment relationships (De’Ath, 2002). MRTs use a recursive partitioning procedure to sequentially split observations (e.g., fish communities) into dichotomous groups with the goal of minimizing within-group variation while simultaneously identifying environmental variables that distinguish each group from one another. Separate MRTs were developed based on taxonomic composition and functional composition of stream fishes. For the taxonomic MRT, we retained only those species occurring in ≥5% of the sites to maximize detection of predominant community–environment relationships (Poos & Jackson, 2012). As a secondary analysis, we included all species to ensure models accounted for habitats supporting highly rare or endemic species. Site-by-species matrices were Hellinger-transformed prior to implementation of the MRT analysis (Borcard, Gillet, & Legendre, 2011). For the functional MRT, we calculated abundance-weighted means for each of the 13 traits. This calculation produced weighted means for continuous traits (reproductive life history and morphology) and the proportion of individuals within a community belonging to each binary trait (trophic and temperature guild). This site-by-trait matrix was standardized from 0 to 1 prior to implementation of MRT analysis so traits with different value ranges were equally weighted. Because temperature guild membership is based on field observations of thermal habitat use (Lyons et al., 2009) and therefore could inflate model performance and explanatory capability of temperature, we also fit function-based MRTs without temperature guild membership.

We developed MRTs with two objectives in mind. First, we developed unpruned trees using all seven environmental predictors and each predictor variable on its own. Comparing explanatory capability (e.g., $R^2$) among these eight models based on taxonomic and functional community matrices allowed us to evaluate the relative community-structuring roles played by each of the environmental predictors. Second, we grew taxonomy-based and function-based trees using all seven environmental predictors and then pruned them back using the minimum cross-validation relative error rule (Borcard et al., 2011). This set of trees provided an optimal number of terminal nodes (i.e., stream classes) that maximized explanatory capability while maintaining the simplicity and interpretability of the classification scheme.

Following the development of pruned MRTs, we characterized taxonomic and functional variation within each class. For the taxonomic MRT, we used Indicator Species Analysis to quantify the affinity of each species to each of the taxonomy-based classes (Borcard et al., 2011). This technique provided an indicator value ranging from 0 to 1, where a perfect indicator species (i.e., indicator value of 1) occurs at all sites within a given class and is absent from all other classes. A Monte Carlo randomization test was used to determine the statistical significance of indicator values. For the function-based MRT, we calculated community trait means for each of the 13 traits in each class and plotted these values to visualize trait variation across classes.

We mapped both taxonomic-based and function-based classes to the ~853,000 NHD reaches to visualize the distribution of classes throughout the eastern United States. The total number of stream kilometres was summarized for each class to characterize prevalence. The National Fish Habitat Action Plan cumulative disturbance index (Esselman et al., 2011; Figure 1b) was used to estimate the proportion of stream kilometres within each class considered to be disturbed by anthropogenic land use in the upstream catchment.

3 | RESULTS

3.1 | Environmental layers and fish data

The seven environmental predictor variables acquired or modelled for the eastern US study area included a total of 852,543 NHD reaches. Random Forest models produced accurate predictions of scores for the three hydrology PC axes. Mapped hydrology axes were consistent with other assessments of spatial variability in hydrology in the eastern United States (McManamay et al., 2014; Olivero & Anderson, 2008). Hydrology PC1 explained 40.8% of the variation in the 110 hydrologic metrics across the 643 gauged stream reaches. This first axis represented a gradient of flow variability, with negative values representing streams with stable regimes and positive values representing streams with variable regimes (Figure 1g, j). For Hydrology PC2 (14.3% of the variation), negative values represented metrics of high flow duration and positive values represented metrics of high flow frequency (Figure 1h,k). For Hydrology PC3 (9.2%), negative values represented metrics of springtime mean flows and positive values represented metrics of autumn maximum flows (Figure 1i,l). Pairwise correlation analysis indicated weak concordance among the seven environmental predictor variables (Figure 3). A total of 371 native fish species were documented at 111 sites. Of these species, 63 occurred at <5% (55) of the sites and were considered rare species. Taxonomic MRTs were constructed that excluded and included these rare species.

3.2 | Unpruned MRTs and explanatory capability of environmental variables

The variation explained by each MRT increased with the size of the unpruned tree. All seven environment predictors together explained more variation in both taxonomic (Figure 4a, black line) and functional (Figure 4b, black line) community composition than did any single predictor. Ecoregion explained more variation in taxonomic
composition than any other individual predictor (Figure 4a, grey line), whereas summer temperature explained the most variation in functional composition (Figure 4b, red line). In other words, the explanatory capability of ecoregion was similar for taxonomic and functional descriptors of the community (Figure 4c), whereas the explanatory capability of the six strictly environmental predictors was approximately 10% higher for functional composition compared to taxonomic composition (Figure 4d). Including all species in the taxonomic MRTs noticeably reduced the explanatory power of all models (Figure S1). In the functional MRT, however, removing temperature guild membership did not change the relative explanatory capability of environmental or ecoregional predictors (Figure S2).

3.3 Pruned trees and stream class mapping

The optimal taxonomic MRT for non-rare species explained 26.4% of the variation in community composition and retained 13 terminal nodes (i.e., leaves), which we interpret as ecologically informative classes hereafter referred to as T1 through T13 (Figure 5a). When considering all species, variance explained by the optimal taxonomic
MRT reduced to 22.9%, although tree structure (including important variables) was similar to that for non-rare species (Figure S3). From here on, we describe only the results of the non-rare taxonomic MRT. Of the twelve internal nodes representing environmental/biogeographic divisions, three involved ecoregional divides, three involved summer temperature thresholds, three involved discharge thresholds, two involved gradient thresholds and one involved a hydrology threshold. The first division involved an ecoregional divide separating Atlantic coast drainages (T1 through T6) from the Mississippi River and Great Lakes drainages (T7 through T13).

Within the Atlantic Coast classes, summer temperature thresholds distinguished cold- and cool-water streams (T1 through T3) from warm-water streams (T4 through T6). Cold- and cool-water streams were further divided into high-gradient cold-water streams (T1), high-gradient cool-water streams (T2) or low-gradient cold-water or cool-water streams (T3). Warm-water streams on the Atlantic Coast were separated into very warm streams on the southeastern coastal plain (T4), and moderately warm streams on the southeastern Piedmont (T5), and a geographically disjunct group of moderately warm streams in the northeast and northern gulf coast (T6). Within the Mississippi–Great Lakes classes, low Hydrology PC1 separated streams with variable flow regimes (T7 through T10) from streams with stable flow regimes (T11 through T13). The former group was further separated by an ecoregional divide separating small streams (T7) and large rivers (T8) in the Great Lakes basin from small streams (T9) and large rivers (T10) in the Mississippi Basin. The latter group (T11 though T13) included small high-gradient streams (T11), small low-gradient streams (T12) and large rivers (T13).

Statistically significant indicator species were identified for all taxonomy-based classes except T9 (Table 3). These indicator species generally aligned with putative temperature and physiographic associations. For example, brook trout (Salvelinus fontinalis) was most representative (i.e., highest indicator value) of the T1 class, which represents cold-water streams of New England and central and southern Appalachia. Similarly, blacknose dace (Rhinichthys atratulus) was most representative of the T2 class, which includes streams slightly warmer streams in the same geographic areas. By contrast, the best indicators of the T4 class included two centrarchids (Lepomis spp.) and the pirate perch (Aphredoderus sayanus), all of which are common in warm, low-gradient streams of the southeastern United States. In the

**FIGURE 4** Explanatory capability as a function of tree size for (a) taxonomy-based and (b) function-based multivariate regression trees. Explanatory capability for the full model (all 7 predictor variables) and models using each predictor variable, in turn, are shown. Explanatory capability for (c) only ecoregional (i.e., biogeographic) boundaries as predictors and (d) only the six environmental (i.e., non-biogeographic) variables are shown for taxonomy-based versus function-based communities.
Great Lakes region, small stream species (e.g., mudminnow [Umbra limi], johnny darter [Etheostoma nigrum] and common shiner [Luxilus cornutus]) appropriately represented the T7 class, whereas species commonly found in large rivers in this region (e.g., smallmouth bass [Micropterus dolomieu], rock bass [Ambloplites rupestris] and blackside darter [Percina maculata]) best represented the T8 class.

The optimal functional MRT explained 25.2% of variation in community composition and retained six terminal nodes hereafter referred to as F1 through F6 (Figure 5b). The optimal functional MRT without temperature guild membership included in the trait matrix explained 26.1% of the variation and also retained six terminal nodes (Figure S4). Of the five internal nodes representing environmental/biogeographic divisions, three involved ecoregional divides and two involved summer temperature thresholds. The first division separated warm-water streams (F1 through F3) from cool- and cold-water streams (F4 through F6). Among the warm-water streams, an ecoregional divide separated mostly northern ecoregions and the Mobile basin (F1 and F2) from the Tennessee drainage and Atlantic and Gulf coastal drainages of the Florida Panhandle (F3). The former group was further separated into a group of streams draining the Atlantic coast and northern Gulf coast (F1) and a group of streams in the Mobile drainage and south-flowing streams of the Ohio River basin (F2). The cool- and cold-water streams (F4 through F6) were further divided into a cool-water group (F4 and F5) and cold-water group (F6). Cool-water streams were further divided by ecoregion, with the F4 class representing streams in the Great Lakes basin, Northeastern Atlantic drainages and southern Appalachia and the F5 class representing streams in the Ohio and Tennessee basins.

Temperature guilds and trophic guilds exhibited the greatest variation among function-based classes (Figure 6). Cold- and cool-water species were most abundant in the F4 class, followed by the F4 and F5 classes. Omnivores made up approximately half of the community for all function-based classes except the F3 class which was represented by only about 15% omnivores. Invertivores were particularly abundant in the F1 and F3 classes, comprising the majority of individuals, on average, in those communities. Life history and body morphology varied more modestly among function-based stream classes. Equilibrium
Table 3: Top three statistically significant (p < .05) indicator species for 13 taxonomy-based stream classes

| Class | Species           | Indicator value |
|-------|-------------------|-----------------|
| T1    | Salvelinus fontinalis | 0.42            |
| T2    | Rhinichthys atratulus | 0.35            |
| T3    | Semotilus corporalis | 0.34            |
|       | Etheostoma olmstedii | 0.17            |
|       | Catostomus commersonii | 0.17           |
| T4    | Lepomis gulosus     | 0.28            |
|       | Aphredoderus sayanus | 0.28            |
|       | Lepomis auritus     | 0.27            |
| T5    | Noemis leptocephalus | 0.54            |
|       | Noturus insignis    | 0.21            |
|       | Clinostomus funduloides | 0.17           |
| T6    | Anguilla rostrata   | 0.24            |
|       | Notrops hudsonius   | 0.11            |
|       | Notrops rubellus    | 0.07            |
| T7    | Umbra limi         | 0.45            |
|       | Etheostoma nigrum  | 0.17            |
|       | Luxilus cornutus   | 0.15            |
| T8    | Micropterus dolomieu | 0.48            |
|       | Ambloplites rupestris | 0.27           |
|       | Percina maculata   | 0.18            |
| T9    | No significant indicator species |
| T10   | Hypentelium nigricans | 0.25            |
|       | Moxostoma duquesnei | 0.19            |
|       | Etheostoma zonale  | 0.15            |
| T11   | Semotilus atromaculatus | 0.23           |
| T12   | Lepomis cyaneus     | 0.21            |
|       | Notrops buccatus    | 0.17            |
|       | Etheostoma spectabile | 0.16            |
| T13   | Etheostoma blemioides | 0.26            |
|       | Moxostoma erythrum | 0.23            |
|       | Notrops stramineus  | 0.23            |

species were most indicative of the F1, F2 and F3 classes and least indicative of the F6 class. Morphology axis 1 scores were lowest for the F3 class, which corresponds to a community composed of species with greater distances between anterior and posterior dorsal fin origins (indicative of sunfishes and darters). Morphology axis 3 scores were lowest for the F3 class, which corresponds to a community composed of species with greater distances between anterior and posterior dorsal fin origins (indicative of sunfishes and darters). Morphology axis 3 scores were low for the F1, F2 and F3 classes and corresponded to deeper-bodied species with derived (i.e., anterior oriented) pelvic fin origins.

3.4 Prevalence and anthropogenic degradation of stream classes

The 13 taxonomy-based classes varied substantially in prevalence, which we characterized as the number of stream kilometres per class distributed among the 852,543 NHD reaches (Figure 7l). The T12 class was most prevalent (269,326 stream kilometres), followed by the T4 (158,427 km) and T5 (152,614 km) classes. Classes representing larger streams (i.e., mean annual discharges greater than 37 cubic feet per second) were the least prevalent and included the T8 (5,743 km), T10 (29,613 km) and T13 (40,966 km) classes. It is important to consider that although these classes are represented by relatively few stream kilometres, their large discharge provides substantially more habitat area per stream kilometre than do the smaller streams represented by more prevalent classes (e.g., T11 and T12). Stream kilometres was more evenly distributed among function-based classes compared to the taxonomy-based classes (Figure 8f). The F6 (282,775 km) and F1 (269,498 km) classes were most prevalent, whereas the F2 (187,488 km) and F3 (187,247 km) classes were least prevalent.

The degree to which stream classes have been anthropogenically degraded varied substantially among classes (Table 4). In general, those classes in regions with high agricultural land use (i.e., Ohio, Indiana, Mississippi delta) or high human population densities (i.e., northeast) were the most degraded. Among taxonomy-based classes, the T12 class exhibited the greatest degradation; 45% of stream kilometres were classified by NFHAP as high or very high disturbance. At least 30% of the T2, T5 and T13 classes were classified as high or very high disturbance. Among the least disturbed taxonomy-based classes were the T1 and T8 classes. Function-based classes exhibited similar and moderate anthropogenic degradation, with all six classes ranging from 21% to 36% of stream kilometres classified by NFHAP as high or very high disturbance. It is important to note that reaches of long rivers inundated by reservoirs (e.g., the entire Tennessee River) are not scored by NFHAP. Given that these reaches are impounded, it is most appropriate to view them as degraded. As such, the degradation reported for the T3, T4 and T12 classes in the ‘High’ and ‘Very high’ columns of Table 4 likely underestimate the degree of true degradation owing to large percentages in the ‘Unscored’ column.

4 DISCUSSION

Our overall aim was to use species- and trait-environment relationships of fish communities to classify streams across the physiographically and biologically diverse eastern United States. Developing classifications for large regions can be challenging, given that community composition is a product of historical biogeography and numerous contemporary environmental conditions that vary at multiple spatial scales (Diamond, 1975; Tonn et al., 1990; Weeks, Claramunt, & Cracraft, 2016). By comparing environmental and biogeographic associations of taxonomic versus functional composition, we were able to estimate the relative importance of these environmental factors in structuring different aspects of fish communities. Moreover, classification schemes mapped in a spatially contiguous manner (e.g., all confluence-to-confluence stream reaches) can inform water resources management and biodiversity conservation using GAP analyses (Turak & Linke, 2011). By integrating fish communities at discrete sites with the spatially contiguous
environmental layers originating from the frequently used National Hydrography Dataset, we were able to quantify the spatial distribution and prevalence of fish-based stream classes as well as estimate anthropogenic degradation of these classes.

4.1 | Natural drivers of assemblage composition

Species- and trait-environment relationships revealed by our environmentally constrained classification schemes were generally congruent with the findings of previous studies on stream fishes in North America and on other continents. Our findings are congruent with temperature thresholds identified at narrower spatial extents within the eastern United States (Brenden et al., 2008; Lyons et al., 2009). For example, eastern North America’s only native inland salmonid, the brook trout, was the top indicator species for the coldest (<17.8°C) stream class. Similarly, cool-water minnow species (e.g., blacknose dace) were top indicators for stream classes of intermediate temperature (17.8–21.2°C). From a functional perspective, species belonging to cold- and cool-water guilds were numerically dominant in streams less than 19.7°C. Stream size and gradient were also important drivers of taxonomic composition, which is congruent with putative patterns of community nestedness and turnover from upstream-to-downstream gradients in North America and in other temperate regions of the world.

Ultimately, the variation explained by MRTs for both taxonomic and functional community composition was relatively low, which suggests that other important environmental variables or historical dispersal mechanisms could have been unaccounted for in our models. Our analysis was conducted at the stream-reach scale whereas fish communities were assessed within smaller sub-reaches. Hence, our analysis might have failed to incorporate finer-resolution variables, such as substrate or channel width, important to distinguishing communities (Marsh-Matthews & Matthews, 2000). Although ecoregions have been shown to explain considerable variation in fish community composition (McManamay et al., 2014) and served to constrain the geographical extent of fish distributions in our models, incorporation of spatial dependency structure might have increased model explanatory power through higher spatial fidelity. For instance, spatial eigenvectors explaining autocorrelative patterns in community composition have proven to be a powerful technique for incorporating spatial structure in models (Borcard & Legendre, 2002).
Components of the natural flow (hydrologic) regime were of minimal importance in distinguishing communities based on taxonomic and functional composition. This is surprising given that previous research in North America has shown strong relationships between hydrologic variability and fish communities, particularly life-history variation (McManamay & Frimpong, 2015; Mims & Olden, 2012). Several factors could have contributed to this limited effect of hydrology. First, response of life-history strategies to hydrologic regimes may not be detectable via the analysis of community trait means. For example, Troia & Gido (2015) found that river mainstems

**FIGURE 7** (a–k) Geographic distribution and (l) relative frequency of 13 taxonomy-based stream classes in the eastern United States
have a greater diversity of life-history strategists compared to tributary streams in the central United States. Likewise, Mims & Olden (2012) were only able to detect relationships between hydrology and life-history strategies via quantile regression, which also suggests community mean values may not respond predictably to hydrology. Trait-based studies of more generalized applicability also highlight the importance of whole-community metrics characterizing not only trait means but also trait variances (Cornwell, Schwilk, & Ackerly, 2006). A second factor that may have contributed to the limited explanatory capability of hydrology is the relative strength of other community-structuring environmental factors. The eastern US study area spans a broad latitudinal gradient which corresponded to a strong community-structuring gradient of summer temperature. By contrast, natural hydrologic regimes depend largely on regional precipitation regimes, which are not particularly variable within the eastern United States. Indeed, those studies in North America that have identified hydrology as an important driver of fish life histories have spanned the complete east-coast-to-west-coast precipitation gradient (McManamay & Frimpong, 2015; Mims & Olden, 2012), which corresponds to strong hydrologic gradients (McManamay et al., 2014; Poff & Ward, 1989).

Studies that compare community-environment relationships based on taxonomic and functional composition can shed light on the relative roles of contemporary environmental filtering versus historical dispersal and speciation (Hoeinghaus et al., 2007; but see Heino, 2008). Our taxonomy- and trait-based classification of fish communities in the eastern United States supported this assertion. That environmental factors were more effective predictors of functional composition than taxonomic composition suggests that aspects of taxonomic variation among communities is not a consequence of contemporary environmental filters and therefore is related to non-environmental factors (e.g., historical dispersal and speciation). The importance of ecoregional boundaries in explaining taxonomic composition of fishes is not surprising, given the complex biogeographic history of the fish fauna in eastern North America (Mayden, 1992). Including rare species in the taxonomy-based classification scheme decreased the explanatory power of MRTs. On one hand, the inclusion of rare species could be hypothesized to increase the explanatory power of ecoregional boundaries because most rare species in our data set are narrow-ranging endemics that are functionally similar to broad-ranging congeners (Etnier & Starnes, 1993). However, including rare species also induces zero inflation in models, potentially impacting model stability, and induces model bias, as MRTs might select for environmental variables serving as mere artefacts of systems where rare species are found, such as remnant patches (e.g., small streams) of former distributions, as opposed to true environmental gradients associated with species occurrences.

To summarize, these findings demonstrate the need for stream classification systems based on species-environment relationships to include contemporary environmental and historical biogeographic factors, particularly for those schemes developed at broad spatial scales and/or in regions with complex biogeographic histories.
4.2 | Conservation and management implications

Stream classifications are essential for decision-making in water resources management and biodiversity conservation. Some have suggested that conserving habitat typologies, that is surrogates for ecological process supporting biodiversity, are a superior approach to climate change adaption, as creating reserves representing the diversity of these types will simultaneously preserve processes that maintain species diversity (Beier & Brost, 2010). Our classifications effectively identified several stream habitats that are currently limited in prevalence, have been disturbed by anthropogenic activities or are likely to be affected by future anthropogenic environmental change. For example, cold-water streams were not very prevalent, given their restriction to northern latitudes and high elevations. Ongoing and future climate change threatens cold-water fishes generally (Comte, Buisson, Daufresne, & Grenouillet, 2013) and eastern cold-water species specifically (Meisner, 1990). The immediacy of this threat depends on the velocity of climate change which has been evaluated in the western United States (Isaak et al., 2016), but not yet in the eastern United States. Another conservation concern is large river habitats, which are least prevalent in terms of cumulative length and are also highly degraded owing to flow regulation for flood control, commercial navigation and hydroelectric generation (Vörösmarty et al., 2010). Great rivers of North America (e.g., Mississippi, Missouri, and Ohio Rivers) and their large tributaries are important reservoirs of fish diversity and require conservation action (Pracheil, McIntyre, & Lyons, 2013).

Management and conservation decisions should not only depend on fishes, but on all stream-dwelling taxa including macroinvertebrates, crayfishes and mussels. This is particularly important in the eastern United States where these invertebrate taxa exhibit high regional richness and local endemism (Stein, Kutner, & Adams, 2000). It is unclear whether our fish-based classifications would accurately represent spatial variation in other stream-dwelling fauna and flora. Some previous studies suggest that fishes and invertebrate taxa respond similarly to natural environmental conditions and/or anthropogenic disturbances (Johnson & Hering, 2010), whereas others indicate differing responses (Piliere et al., 2014; Troia, Williams, Williams, & Ford, 2015). Presently, compilation of occurrence data for non-fish taxa lags behind that of fishes in the eastern United States (Troia & McManamay, 2016); however, the growth of open-access biodiversity data will soon allow the completion of classification schemes for the eastern United States that incorporate these other important taxa. A final caveat is that many freshwater taxa occupy lentic habitats whereas our classification schemes only correspond to lotic habitats. Although this is a common limitation of many freshwater classification schemes (Melles, Jones, & Schmidt, 2012), it is important to acknowledge that numerous species and functional components of the eastern North American freshwater fish fauna may not be adequately represented. This is particularly important in the southeastern United States where many unique fishes occupy wetland habitats (e.g., species of the family Elassomatidae) and the upper Midwest where glacial lakes support unique species (e.g., lake trout [Salvelinus namaycush] and species of the family Coregonidae).

4.3 | Broader implications and conclusions

Our study highlights at least three important considerations when developing habitat classifications for streams or, more generally, for other aquatic or terrestrial habitats. First, we have provided a straightforward, yet powerful, technique to directly (as opposed to indirectly) incorporate the relative importance of environmental variables in organizing ecological communities into stream classification. Secondly, our work emphasizes that spatial variation in biodiversity is a product of environmental conditions and historical factors that vary at a range of spatial scales (Diamond, 1975; Weeks et al., 2016). For example, our classification revealed the importance of reach-scale channel gradient but also region-wide biogeographic boundaries in distinguishing biological communities. Anthropogenic alterations to ecosystems increasingly present continental- and even global-scale challenges (Hulme, 2005; Vörösmarty et al., 2010). As such, classification schemes should continue to include appropriately scaled environmental predictors as we have shown with eastern US stream fish assemblages. Finally, trait-based approaches can improve confidence in the causative linkages between biotas and environmental conditions, be those natural or anthropogenic.
(Frimpong & Angermeier, 2010; McGill, Enquist, Weiher, & Westoby, 2006). These considerations will continue to benefit from the ever-increasing availability of open-access data describing environmental conditions (e.g., NHD), species distributions (e.g., GBIF) and species' traits (e.g., FishTraits).

ACKNOWLEDGEMENTS

The authors have no conflict of interest in the publication of this research. This research was sponsored by the United States Department of Energy’s Office of Energy Efficiency and Renewable Energy, Water Power Technologies Office. We thank Julian Olden, Tim Bonner and Cody Craig for thoughtful comments on data analysis and interpretations.

DATA AVAILABILITY STATEMENT

Data are posted on figshare at: https://doi.org/10.6084/m9.figshare.9917081.v1

ORCID

Ryan A. McManamay https://orcid.org/0000-0002-5551-3140

REFERENCES

Abell, R., Thieme, M. L., Revenga, C., Bryer, M., Kottelat, M., Bogutskaya, N., ... Petry, P. (2008). Freshwater ecoregions of the world: A new map of biogeographic units for freshwater biodiversity conservation. *BioScience*, 58, 40–414. https://doi.org/10.1641/B580507

Barbour, M. T., Gerritsen, J., Griffith, G. E., Frydenborg, R., McCarron, E., White, J. S., & Bastian, M. L. (1999). A framework for biological criteria for Florida streams using benthic macroinvertebrates. *Journal of the North American Benthological Society*, 15, 185–211. https://doi.org/10.2307/1467948

Beier, P., & Brost, B. (2010). Use of land facets to plan for climate change: Conserving the arenas, not the actors. *Conservation Biology*, 24, 701–710. https://doi.org/10.1111/j.1523-1739.2009.01422.x

Benda, L., Poff, N. L., Miller, D., Dunne, T., Reeves, G., Pess, G., & Pollock, M. (2004). The network dynamics hypothesis: How channel networks structure riverine habitats. *Ecology*, 85, 413–427. https://doi.org/10.1641/0003-0563(2004)054[0413:TNDHH]2.0.CO;2

Borcard, D., Gillet, F., & Legendre, P. (2011). *Numerical ecology with R*. New York, NY: Springer.

Borcard, D., & Legendre, P. (2002). All-scale spatial analysis of ecological data by means of principal coordinates of neighbor matrices. *Ecological Modelling*, 153, 51–68.

Brenden, T. O., Wang, L., & Seelbach, P. W. (2008). A river valley segment classification of Michigan streams based on fish and physical attributes. *Transactions of the American Fisheries Society*, 137, 1621–1636. https://doi.org/10.1577/T07-166.1

Comte, L., Buissin, L., Daufresne, M., & Grenouillet, G. (2013). Climate-induced changes in the distribution of freshwater fish: Observed and predicted trends. *Freshwater Biology*, 58, 625–639. https://doi.org/10.1111/fw.12081

Cornwell, W. K., Schwilk, D. W., & Ackerly, D. D. (2006). A trait-based test for habitat filtering: Convex hull volume. *Ecology*, 87, 1465–1471. https://doi.org/10.1890/0012-9658(2006)87[1465:ATTFH]2.0.CO;2

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, 88, 2783–2792. https://doi.org/10.1890/07-0539.1

De’Ath, G. (2002). Multivariate regression trees: A new technique for modeling species–environment relationships. *Ecology*, 83, 1105–1117. De’Ath, G., & Fabricius, K. E. (2000). Classification and regression trees: A powerful yet simple technique for ecological data analysis. *Ecology*, 81, 3178–3192. https://doi.org/10.1890/0012-9658(2000)081[3178:CARP]2.0.CO;2

Diamond, J. (1975) Assembly of species communities. In M. L. Cody, & J. Diamond (Eds.), *Ecology and evolution of communities* (pp. 342–444). Cambridge, MA: Harvard University Press.

Dudgeon, D., Arthington, A. H., Gessner, M. O., Kawabata, Z.-I., Knowler, D. J., Lévêque, C., ... Sullivan, C. A. (2006). Freshwater biodiversity: Importance, threats, status and conservation challenges. *Biological Reviews*, 81, 163–182. https://doi.org/10.1017/S1464793105006950

Esselman, P. C., Infante, D. M., Wang, L., Wu, D., Cooper, A. R., & Taylor, W. W. (2011). An index of cumulative disturbance to river fish habitats of the conterminous United States from landscape anthropogenic activities. *Ecological Restoration*, 29, 133–151. https://doi.org/10.3368/er.29.1.2.133

Etner, D. A., & Starnes, W. C. (1993). *The Fishes of Tennessee* (p. 689p). Knoxville, TN: The University of Tennessee Press.

Felsenstein, J. (1985). Phylogenies and the comparative method. *The American Naturalist*, 125, 1–15. https://doi.org/10.1086/284325

Ferrier, S., Drielsma, M., Manion, G., & Watson, G. (2002). Extended statistical approaches to modelling spatial pattern in biodiversity in north-east New South Wales: II. Community-level Modelling. *Biodiversity and Conservation*, 1, 2309–2338.

Franssen, N. R., Goodchild, C. G., & Shepard, D. B. (2015). Morphology predicting ecology: Incorporating new methodological and analytical approaches. *Environmental Biology of Fishes*, 98, 713–724. https://doi.org/10.1007/s10641-014-0306-z

Frimpong, E. A., & Angermeier, P. L. (2009). Fish traits: A database of ecological and life-history traits of freshwater fishes of the United States. *Fishes*, 34, 487–495. https://doi.org/10.1577/1548-8446-34.10.487

Frimpong, E. A., & Angermeier, P. L. (2010). Trait-based approaches in the analysis of stream fish communities. *American Fisheries Society Symposium*, 73, 109–136.

Giam, X., & Olden, J. D. (2016). Environment and predation govern fish community assembly in temperate streams. *Global Ecology and Biogeography*, 25, 1194–1205. https://doi.org/10.1111/geb.12475

Haag, W. R. (2012). *North American freshwater mussels: Natural history, ecology, and conservation*. Cambridge, MA: Cambridge University Press.

Hawkins, C. P., Kershner, J. L., Bisson, P. A., Bryant, M. D., Deckler, L. M., Gregory, S. V., ... Young, M. K. (1993). A hierarchical approach to classifying stream habitat features. *Fishes*, 18, 3–12. https://doi.org/10.1577/1548-8446(1993)018<0003:AHATC>2.0.CO;2

Hawkins, C. P., Norris, R. H., Gerritsen, J., Hughes, R. M., Jackson, S. K., Johnson, R. K., & Stevenson, R. J. (2000). Evaluation of the use of landscape classifications for the prediction of freshwater biota: Synthesis and recommendations. *Journal of the North American Benthological Society*, 19, 541–556. https://doi.org/10.2307/1468113

Hayes, D., Baker, E., Bednarz, R., Borgeson, D., Braunscheidel, J., Harrington, A., ... Zorn, T. (2003). Developing a standardized sampling program: The Michigan experience. *Fishes*, 28, 18–25. https://doi.org/10.1577/1548-8446(2003)28[18:DASSP]2.0.CO;2

Heino, J. (2008). Patterns of functional biodiversity and function-environment relationships in lake littoral macroinvertebrates. *Limnology and Oceanography*, 53, 1446–1455.
Rosgen, D. L. (1994). A classification of natural rivers. *Catena*, 22, 169–199. https://doi.org/10.1016/0341-8162(94)90001-9

Roux, D. J., Nel, J. L., Ashton, P. J., Deacon, A. R., de Moor, F. C., Hardwick, D., ... Scholes, R. J. (2008). Designing protected areas to conserve riverine biodiversity: Lessons from a hypothetical redesign of the Kruger National Park. *Biological Conservation*, 141, 100–117. https://doi.org/10.1016/j.biocon.2007.09.002

Scott, M. C., & Helfman, G. S. (2001). Native invasions, homogenization, and the mismeasure of integrity of fish assemblages. *Fisheries*, 26, 6–15. https://doi.org/10.1577/1548-8446(2001)026<0006:NIIHMA>2.0.CO;2

Stein, B. A., Kutner, L. S., & Adams, J. S. (2000). *Precious heritage: The status of biodiversity in the United States*. New York, NY: Oxford University Press.

Taylor, C. A., Warren, M. L., Fitzpatrick, J. F., Hobbs, H. H., Jezerinac, R. F., Pfieger, W. L., & Robison, H. W. (1996). Conservation status of crayfishes of the United States and Canada. *Fisheries*, 21, 25–38. https://doi.org/10.1577/1548-8446(1996)021<0025:CSOCT>2.0.CO;2

Tolonen, K. E., Leinonen, K., Marttila, H., Erkinaro, J., & Heino, J. (2017). Environmental predictability of taxonomic and functional community composition in high-latitude streams. *Freshwater Biology*, 62, 1-16. https://doi.org/10.1111/fwb.12832

Tonn, W. M., Magnuson, J. J., Rask, M., & Toivonen, J. (1990). Intercontinental comparison of small-lake fish assemblages: The balance between local and regional processes. *The American Naturalist*, 136, 345–375. https://doi.org/10.1086/285102

Townsend, C. R., & Hildrew, A. G. (1994). Species traits in relation to a habitat template for river systems. *Freshwater Biology*, 31, 265–275. https://doi.org/10.1111/j.1365-2427.1994.tb01740.x

Troia, M. J., & McManamay, R. A. (2016). Filling in the GAPS: Evaluating completeness and coverage of open-access biodiversity databases in the United States. *Ecology and Evolution*, 6, 4654–4669. https://doi.org/10.1002/ece3.2225

Troia, M. J., & McManamay, R. A. (2017). Completeness and coverage of open-access freshwater fish distribution data in the United States. *Diversity and Distributions*, https://doi.org/10.1111/ddi.12637

Troia, M. J., Williams, L. R., Williams, M. G., & Ford, N. B. (2015). The process domains concept as a framework for fish and mussel habitat in a coastal plain river of southeastern North America. *Ecological Engineering*, 75, 484–496. https://doi.org/10.1016/j.ecoleng.2014.12.016

Turak, E., & Linke, S. (2011). Freshwater conservation planning: An introduction. *Freshwater Biology*, 56, 1–5. https://doi.org/10.1111/j.1365-2427.2010.02515.x

(U.S. EPA) United States Environmental Protection Agency. (2005). *National Hydrography Dataset Plus—NHDP Plus Edition 1.0*. Washington, DC: U.S. Environmental Protection Agency (USEPA).

Vannote, R. L., Minshall, G. W., Cummins, K. W., Sedell, J. R., & Cushing, C. E. (1980). The river continuum concept. *Canadian Journal of Fisheries and Aquatic Sciences*, 37, 130–137. https://doi.org/10.1139/F80-017

Vörösmarty, C. J., McIntyre, P. B., Gessner, M. O., Dudgeon, D., Prusevich, A., Green, P., ... Davies, P. M. (2010). Global threats to human water security and river biodiversity. *Nature*, 467, 555–561. https://doi.org/10.1038/nature09440

Ward, J. V., Tockner, K., Arscott, D. B., & Claret, C. (2002). Riverine landscape diversity. *Freshwater Biology*, 47, 517–539. https://doi.org/10.1111/j.1365-2427.2002.00893.x

Weeks, B. C., Claramunt, S., & Cracraft, J. (2016). Integrating systematics and biogeography to disentangle the roles of history and ecology in biotic assembly. *Journal of Biogeography*, 43, 1546–1559. https://doi.org/10.1111/jbi.12747

Winemiller, K. O., & Rose, K. A. (1992). Patterns of life-history diversification in North American fishes: Implications for population regulation. *Canadian Journal of Fisheries and Aquatic Sciences*, 49, 2196–2218. https://doi.org/10.1139/f92-242

Xenopoulos, M. A., & Lodge, D. M. (2006). Going with the flow: Using species-discharge relationships to forecast losses in fish biodiversity. *Ecology*, 87, 1907-1914. https://doi.org/10.1890/0012-9658(2006)087[1907:GWTFU]2.0.CO;2

**BIOSKETCHES**

**Dr. Matthew Troia** earned a PhD in Biology from Kansas State University in 2014, completed postdoctoral research at Oak Ridge National Laboratory and the University of Tennessee. He is currently an Assistant Professor in the Department of Environmental Science and Ecology at the University of Texas at San Antonio. His research interests include biodiversity science, freshwater fish ecology and conservation, thermal ecology and climate change biology. MT co-conceived the idea, acquired data, conducted the analysis, and drafted the manuscript.

**Dr. Ryan McManamay** is an Assistant Professor of Environmental Science at Baylor University in Waco Texas. He is a spatial ecologist that studies human–environmental systems, particularly balancing ecosystem needs with societal demands. His interests include hydrography, characterizing biodiversity impacts of energy development and urban land–energy–water dynamics. Prior to his current appointment, he was a research scientist at Oak Ridge National Laboratory for 6 years. RM conceived the idea, acquired data, edited the manuscript, and acquired funding.

**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article**: Troia MJ, McManamay RA. Biogeographic classification of streams using fish community- and trait-environment relationships. *Divers Distrib.* 2020;26:108–125. https://doi.org/10.1111/ddi.13001