

# Regional species richness, hydrological characteristics and the local species richness of assemblages of North American stream fishes

SOPHIA QIAN NIU\*, MICHAEL P. FRAN CZYK\*\*<sup>1</sup> AND JASON H. KNOUFT\*\*<sup>†</sup>

\*Department of Biology, Saint Louis University, 3507 Laclede Ave., St. Louis, MO 63103, U.S.A.

<sup>†</sup>Center for Environmental Sciences, Saint Louis University, 3507 Laclede Ave., St. Louis, MO 63103, U.S.A.

## SUMMARY

1. Local assemblage structure, from a deterministic perspective, is presumably dictated by the regional species pool as well as regulated by local factors. We examined the relationships of the regional species pool and local hydrological characteristics to local species richness of North American freshwater fishes using data sets collected during the National Water Quality Assessment program conducted by the United States Geological Survey.

2. We predicted that local species richness is ultimately constrained by the composition of the regional species pool and further associated with local hydrological factors. Moreover, we predicted that variation in local species richness within major families can be explained by different combinations of hydrological characteristics that represent lineage-specific responses to the environment.

3. Daily discharge and regional and local species richness data were assembled from 41 stream localities across the United States. Multiple stepwise regressions were conducted to predict local species richness, based on regional species richness, mean discharge and hydrological characteristics quantified by nine variables characterising flow variability. Species richness at each site was calculated for the entire assemblage as well as within the four most species-rich families in the data set (Catostomidae, Centrarchidae, Cyprinidae and Percidae).

4. Local species richness was best predicted by a combination of regional species richness and discharge magnitude when all species were considered. Regional species richness was a significant explanatory variable of local species richness for three of four families (Catostomidae, Centrarchidae, Cyprinidae), but not for Percidae. Local richness in Centrarchidae and Cyprinidae was positively correlated with temporal flow variability as well as high and low flow duration, respectively, while richness in Catostomidae and Percidae tended to be associated with discharge volume. In addition, local species richness for three of the four major families was positively correlated with species richness of the other families in the assemblage, potentially suggesting the influence of local habitat quality and heterogeneity.

5. Results suggest the importance of the combined influences of the regional species pool and local hydrological characteristics on local richness in freshwater fishes, with variation in richness within each family predicted by different characteristics of flow regimes.

*Keywords:* discharge, evolutionary constraints, flow variability, hydrology, national water quality assessment

## Introduction

Deterministic and stochastic processes are both responsible for species composition in local assemblages (Chase &

Myers, 2011). Deterministic processes are generally regulated by non-random, niche-based factors, while stochastic processes include events such as chance colonisation, dispersal and extinction, and random changes in

Correspondence: Sophia Q. Niu, 3507 Laclede Ave., Macelwane Hall, St. Louis, MO 63103, U.S.A.

E-mail: qniu1@slu.edu

<sup>1</sup>Present address: Department of Zoology, Southern Illinois University Carbondale, Carbondale, IL 62901, U.S.A.

population density. There is increasing evidence indicating that deterministic and stochastic forces each explain a proportion of variation in local species composition among aquatic assemblages (e.g. Arrington & Winemiller, 2006; Thompson & Townsend, 2006; Brown & Swan, 2010).

From a deterministic perspective, understanding the relative influence of regional and local factors on local species richness is fundamental to predicting variation in spatial patterns of biodiversity. The overall species pool within a particular geographic area is formed through regional and historical processes, such as dispersal and the influence of climatic and geographic templates on evolutionary history, while local processes such as biotic interactions and microhabitat availability serve as a hierarchy of ecological filters that influences the probability of persistence of local populations with particular biological traits (Ricklefs, 1987; Poff, 1997). Consequently, contemporary diversity patterns originate from the history of each region, with the interaction between biogeographic and ecological factors ultimately regulating the relationship between regional and local species richness as well as local assemblage structure (Cornell & Lawton, 1992).

North America contains the greatest number of temperate species of freshwater fishes on the Earth (Page & Burr, 2011). While evolutionary history (Knouft & Page, 2003), long-term dispersal (Knouft, 2004; Griffiths, 2012) and regional climate and landscape characteristics (Knouft & Page, 2011) regulate spatial variation in freshwater fish species richness in this region, flow regimes are recognised as important agents regulating aquatic assemblages at a local scale (Bunn & Arthington, 2002; Lytle & Poff, 2004). Natural flow regimes can be characterised according to discharge magnitude as well as the frequency, timing, duration and changing rates of the periodic flow events (Richter *et al.*, 1997). Flow variability influences aquatic assemblages through several interacting processes via: (i) selection of species behavioural and life-history traits associated with predictable flow events, such as fecundity and generation time (Meffe, 1984; Lytle & Poff, 2004; Lytle, Bogan & Finn, 2008); (ii) recruitment success (survival of young-of-year fish) because of flood timing and duration (Schlosser, 1982; Warner & Chesson, 1985; Piffady *et al.*, 2010); and (iii) changes in composition of lower trophic levels, and hence trophic interactions, which in turn alter the structure of higher trophic levels (Franssen, Gido & Propst, 2007; Power, Parker & Dietrich, 2008; Cross *et al.*, 2011). Finally, discharge volume can determine available habitat area and maintain the lateral and

longitudinal connectivity of nutrients (Bunn & Arthington, 2002).

Given the profound influence of flow regimes on stream assemblages, local species richness should be correlated with local hydrological characteristics while being ultimately constrained by the composition of the regional species pool. Once the group of species available to establish in local assemblages is defined by the regional species pool, we anticipate that higher species richness in local stream fish assemblages is associated with higher discharge volume based on the positive habitat area–species richness relationship reported in other freshwater systems (Angermeier & Schlosser, 1989; Oberdorff, Huguény & Guegan, 1997; Xenopoulos & Lodge, 2006). In addition, relatively higher species richness should be also observed at sites with intermediate to high flow variability based on the predictions of the storage effect hypothesis (*sensu* Warner & Chesson, 1985) as well as the intermediate disturbance hypothesis (Connell, 1978). Variation in recruitment successes among species mediated by flow variation allows species to recover from lower densities, thus increasing the probability of species coexistence (Chesson, 2000). The latter hypothesis predicts that the highest levels of diversity are associated with intermediate disturbance regimes because of trade-offs in species functional traits (between colonisation and competition ability; Connell, 1978). The positive influence of hydrological variability on local species richness is also likely due to fine-scale habitat heterogeneity and the availability of ‘habitat shelters’ generated by spatiotemporally varying hydrological conditions interacting with the geomorphology of the stream channel (Ward, 1998). However, the unique responses of various freshwater taxa are ultimately constrained by the evolutionary history within lineages because of the phylogenetically constrained life-history traits associated with particular taxonomic groups (Cattaneo *et al.*, 2001; Olden & Kennard, 2010; Knouft & Page, 2011).

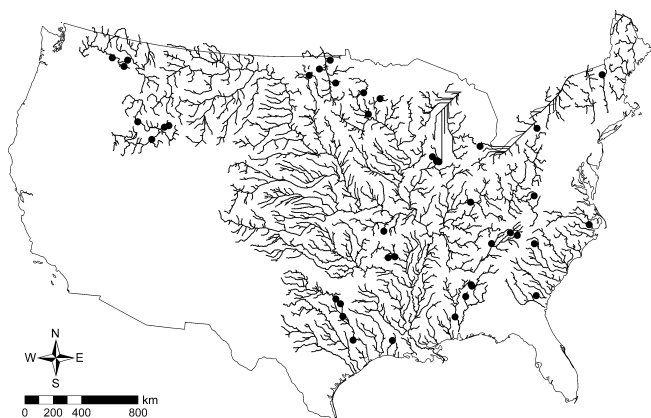
We investigated the relative importance of regional species richness and local hydrology on local species richness in North American stream fishes across a broad geographic extent in the United States. We hypothesised that within the constraints of the regional species pool, local hydrological variables act as habitat filters that further constrain local fish species richness. Moreover, because of the presumed evolutionary constraints within lineages, we expected that variation in local species richness within families will be explained by various features of hydrological regimes (e.g. flow event duration and magnitude of high and low flow events).

## Methods

### Data collection

Fish assemblage data were compiled from the National Water Quality Assessment (NAWQA) program conducted by the United States Geology Survey (USGS) (<http://water.usgs.gov/nawqa/>). The NAWQA data set, in part, includes fish species data from nearly 1000 sites representing most physiographic regions of the United States (Leahy, Rosenshein & Knopman, 1990; Walsh & Meador, 1998). NAWQA sampling generally occurred in low-flow seasons, and sampling sites were identified based on physical characteristics of the stream as well as a standardised reach length criterion. Fishes were collected using a combination of electrofishing and seining to ensure a representative sample of the fish assemblage (Walsh & Meador, 1998; Moulton *et al.*, 2002). Using fish assemblage data from 751 NAWQA sites, Mitchell & Knouft (2009) demonstrated a negative relationship between the number of introduced fish species and native fish species richness. We use the same data set presented in Mitchell & Knouft (2009), but because of the apparent influence of non-native species on native species richness, we limited potential sites for this study to the 233 localities that contained only native species. Hydrological data for this study were assembled from daily flow measurements from USGS gauging stations. Of the 233 native assemblage localities, we only used sites within 500 m of a USGS gauging station to assure reasonably accurate hydrological estimates. This criterion resulted in a final data set of 41 sites (Fig. 1).

The number of species sampled from each site during the NAWQA program represents a measure of local species richness. Because local species richness estimates can be biased by variation in sampling efforts among sites,



**Fig. 1** Map showing the locations of the 41 national water quality assessment (NAWQA) study sites.

we used a rarefaction technique to generate rarefied estimates of local species richness (Gotelli & Colwell, 2001; Gotelli & Entsminger, 2004; as in Mitchell & Knouft, 2009). The minimum number of individuals at a site (61) in the NAWQA data set was used as the sample size during rarefaction resampling at all sites. Resampling of data without replacement at each site was repeated 1000 times to generate a rarefied estimate of species richness (Ecosim, ver. 7.0; Gotelli & Entsminger, 2004). We also assembled and then rarefied local species richness within the four most species-rich families (Catostomidae, Centrarchidae, Cyprinidae and Percidae) to assess the influence of evolutionary constraints on factors regulating local species richness.

Regional species richness for each sampling site was estimated as the number of species with distributions overlapping the site and was calculated based on species distribution maps for North American freshwater fishes (Page & Burr, 2011), assembled in Knouft & Page (2011). This estimate of the regional species pool is conservative in that it makes no assumptions about a species' ability to migrate to novel regions outside of the native range, nor does it make any assumptions about the undocumented historical distribution of a species that may have been more extensive than the current distribution. In this study, if a species' range, based on documented collections (Page & Burr, 2011), is known to overlap a NAWQA site (and thus occur in the same interconnected catchment), then the species is considered to be part of the regional species pool and a potential occupier of the local assemblage.

Ten years of flow discharge data (1990–1999) were assembled for each locality, with the exception of three sites for which only 5–8 years of data were available. Olden & Poff (2003) assessed the redundancy among 171 hydrological indices for streams in the United States and presented an informative subset of those indices characterising four properties of the flow regime (magnitude of discharge, period, timing and changing rate of low/high flow events). Based on recommendations in Olden & Poff (2003), we selected nine hydrological variables to predict variation in local species richness among sites (Table 1). These variables were calculated using INDEX OF HYDROLOGIC ALTERATION v. 3.0 software (<http://conserveonline.org/workspaces/iha/>; Richter *et al.*, 1997). Although our flow estimates may not equally reflect the relevant hydrological components experienced by all fishes in the data set (e.g. different species may be subject to flow variability under finer or coarser time scale), these measures are assumed to be general estimates of hydrological conditions at each site (Chinnayakanahalli *et al.*, 2011).

**Table 1** Hydrological indices used to characterise local hydrology calculated with the INDEX OF HYDROLOGIC ALTERATION software (version 3.0)

Index	Abbreviation	Definition
Coefficient of variation in annual mean discharge	Interyr_cv	Coefficient of variation in annual mean discharge among all the years
Coefficient of variation in daily discharge	Daily_cv	Coefficient of variation in daily discharge within each year averaged across all the years
Thirty-day maximum/annual median discharge	30d_max	Mean of 30-day maximum flow divided by median annual daily discharge, averaged across all the years
Seven-day minimum/annual median discharge	7d_min	Mean of seven-day minimum flow divided by median annual daily discharge, averaged across all the years
High flow duration (days)	HF_dur	Mean duration of annual occurrences during which the discharge remains above 75th percentile of daily discharge
Low flow duration (days)	LF_dur	Mean duration of annual occurrences during which the discharge remains below 25th percentile of daily discharge
Flood-free days	FL_free	Maximum proportion of the year (measured by days), during which discharge remains below 75th percentile of daily discharge, averaged across all the years.
Flow constancy	Cons	The degrees to which flow remains stable across a year (temporal variation; Colwell, 1974)
Flow contingency	Cont	The degree to which time determines the flow state (repeatability of periodic events; Colwell, 1974)

We used mean annual discharge, temporal flow variability, low/high flow magnitude relative to the median discharge, duration of low/high flow events, as well as the predictability of hydrological regimes (constancy and contingency) to predict variation in local species richness (Table 1). To reduce the redundancy among hydrological variables and to constrain the number of independent variables used in the following multiple regressions, a principal components analysis (PCA) was conducted to generate a set of uncorrelated variables [principle component (PC) scores] representing the hydrological characteristics at each site. Because the sites used for each family ( $n = 25\text{--}38$ ) were subsets of the overall species data set ( $n = 41$ ), we conducted a PCA using sites from the total species data set as well as separate PCAs using sites for each family. A correlation matrix was used to generate PC scores, and the PCs with eigenvalues greater than 1.0 were retained as uncorrelated univariate data for further analyses. Prior to the analyses, Daily\_cv (coefficient of variation in daily discharge), HF\_dur (high flow duration), LF\_dur (low flow duration), FL\_free (flood-free days) and 30d\_max (30-day maximum flow relative to median discharge) were  $\log_{10}$ -transformed, and Cont (contingency), Cons (constancy) and 7d\_min (7-day minimum flow relative to median discharge) were arcsine-square-root transformed to normalise each data set.

#### Statistical analyses

Multiple regressions were employed to investigate the relationships between regional species richness, local

hydrological characteristics and local species richness. We followed a general strategy in model selection described in Kleinbaum *et al.* (2008). Regional species richness ( $R_{\text{reg}}$ ), mean annual discharge ( $Q$ ) and PC scores (PC1, PC2 and PC3) were used as explanatory variables, and separate analyses were conducted on original (non-rarefied) and rarefied local species richness ( $R_{\text{loc}}$ ) data sets. A forward stepwise method was used to select the optimum model based on Akaike Information Criteria (AIC) (see details in Yamashita, Yamashita & Kamimura, 2007), which considers a balance between the amount of information represented by a model and the number of predictors used (Burnham & Anderson, 2004). Addition or removal of a candidate variable was decided by whether the AIC value of the resultant model was reduced by more than 2 (selection criteria:  $\Delta\text{AIC} \geq 2$ ). We also assessed interactions between regional species richness and hydrological variables ( $Q$  and PC scores) and considered quadratic relationships in  $R_{\text{reg}}$  and  $Q$ . A cross-validation was conducted to assess the reliability of resultant models.

It is impractical to have a maximum model, from which all the competing terms of the basic predictors are to be selected (including all pair-wise interaction and polynomial terms). Consequently, we conducted several progressive steps to seek the optimal model predicting local species richness. First, we conducted a stepwise regression based on AIC scores including only linear terms of basic predictors; thus, the maximum model is as follows:

$$R_{\text{loc}} = \beta_0 + \beta_1 R_{\text{reg}} + \beta_2 Q + \beta_3 \text{PC1} + \beta_4 \text{PC2} + \beta_5 \text{PC3},$$

where  $\beta_k$  denotes the regression coefficient in an analysis involving  $k$  predictor. We then tested the interactions among selected predictors. Using non-rarefied total richness as an example, local species richness was best predicted by regional species richness and discharge; thus, the maximum model, including interactions is as follows:

$$R_{\text{loc}} = \beta_0 + \beta_1 R_{\text{reg}} + \beta_2 Q + \beta_3 R_{\text{reg}} \times Q.$$

We did not consider any higher order interactions because of a concern regarding the number of samples in the NAWQA data set. For example, Kleinbaum *et al.* (2008) suggest that it is preferable to have at least five to ten observations per variable in multiple regression analyses. If there were no significant interactions in our models, we removed the interaction term and subsequently tested whether quadratic relationships improve the explanatory power of the models including only linear relationships. We considered quadratic terms of regional species richness and mean annual discharge, but not for PC scores, because they are linear combinations of individual hydrological variables and it is difficult to interpret the squared value of these measures. The maximum model at this step was (in cases where no interaction was detected) as follows:

$$R_{\text{loc}} = \beta_0 + \beta_1 R_{\text{reg}} + \beta_2 R_{\text{reg}}^2 + \beta_3 Q + \beta_4 Q^2 + \beta_5 \text{PC1} + \beta_6 \text{PC2} + \beta_7 \text{PC3}.$$

Conserved traits associated with the diversification of particular lineages (i.e. niche conservatism: Wiens & Graham, 2005) of North American freshwater fishes likely imposes constraints on species distributions because of lineage-specific responses to environmental conditions (Knouft & Page, 2011). Therefore, the relationships between regional species richness, local hydrology and local species richness were also investigated within the four most diverse families in the data set (Catostomidae, Centrarchidae, Cyprinidae and Percidae). We anticipated that the number of species present outside of the number of species in the focal family might also explain a portion of the variation in local richness for the focal family of interest. Thus, species richness of other families [richness of others ( $R_{\text{other}}$ )] was added as an independent variable into multiple regressions for family-level analyses. For instance, in addition to regional species richness of Cyprinidae, mean annual discharge and local hydrological variables, we used non-cyprinid local species richness as one of the predictors of local Cyprinidae species richness. The same procedure used to test for interactions in the total richness analysis was applied for each family data set. The family-level data sets have smaller sample

sizes than the total species data set; therefore, instead of adding a quadratic term of  $R_{\text{other}}$  into the maximum model (which would result in nine predictors), we conducted stepwise selection without  $R_{\text{other}}^2$ , and when  $R_{\text{other}}$  was selected, we tested whether the quadratic relationship improves the model by comparing AIC values between the models with and without  $R_{\text{other}}^2$ .

We used fourfold cross-validation to test whether a resultant model could accurately predict local species richness using 75% of the observations as a 'training group' (Kleinbaum *et al.*, 2008). The regression coefficient in the resultant model was estimated based on 30 randomly selected observations (training group), and the remaining eleven observations (testing group) were fitted with this function to calculate an estimated error (residual). We assessed whether the model was valid by comparing the estimated error in the complete model and the mean of the estimated error among testing groups, as well as visually evaluating the plots of predicted-versus-observed values for testing groups distributed around the regression line generated by the training group. No apparent differences in estimated errors and minimal departure of data points in the training group from regression equations suggest statistically accurate predictions (Picard & Cook, 1984). Finally, to assess the relative importance of each variable in the resultant optimal model, the percentage of variation explained by each independent variable was calculated based on the hierarchical partitioning approach described in Chevan & Sutherland (1991). All richness measures and mean annual discharge were  $\log_{10}$ -transformed prior to statistical analyses. All statistical analyses were performed with R, version 2.15.2 (R Development Core Team, 2012).

## Results

The eigenvalues for the first three principle components from the PCA of hydrological indices were greater than 1.0 and collectively explained 84% of the variation among hydrology data sets (PC1, 48%; PC2, 24%; and PC3, 12%) for the total species sites and were used in the subsequent analyses. PC1 describes the temporal variability and predictability in flows and the magnitude of low/high flows relative to the median (Table 2). PC2 describes the duration of low/high flow events as well as inter-year variability (Table 2). PC3 is weakly correlated with both flow variability and low flow duration (Table 2). In the separate PCAs conducted for each family, the first three PCs had eigenvalues >1.0 and were associated with similar hydrological indices as the total species data set

**Table 2** Loadings of each hydrological index at the first three principle components (PCs) from the total species richness analysis. See Table 1 for abbreviation of each index

	PC1	PC2	PC3
Interyr_cv	-0.292	-0.368	-0.385
Daily_cv	-0.441	0.088	-0.252
30d_max	-0.421	0.055	-0.392
7d_min	0.449	0.011	-0.129
HF_dur	0.188	-0.517	-0.217
LF_dur	-0.046	-0.508	0.553
FL_free	-0.014	-0.571	-0.156
Cons	-0.358	-0.048	0.381
Cont	0.420	-0.005	-0.31

(see the individual PCA results for each family in Table S1).

With the original (non-rarefied) data sets, between 59 and 70% of the variation in local species richness and

within-family species richness was explained by the resultant models (Table 3a). Local species richness was positively correlated with regional species richness and discharge volume in the 'All species' data set, with 78% of the explained variation attributed to regional species richness. Local total richness exhibited quadratic association with discharge volume, indicating local richness increased and then decreased as discharge increased. Hydrological characteristics were important explanatory variables in local richness within each of the four families, the relative amount of variation explained by these variables varied among families (discharge, 0–27%; flow variability, 9–32%; Table 3a). Discharge volume explained a relatively greater amount of variation in species richness for Catostomidae and Percidae. In addition, a significant interaction between regional species richness and mean discharge (positive) in both original and rarefied data sets was detected in local species richness

**Table 3.** Selected variables in the optimum model predicting (a) original and (b) rarefied local species richness, with percentage of the explained variation in the dependent variable associated with each independent variable

	Selected variables	Coefficient (SE)	% variation explained	<i>t</i>	<i>P</i>
(a) Original local richness group					
All species (adjust $R^2 = 0.67$ , $n = 41$ )	$R_{reg}$	0.86 (0.10)	77.7	8.60	<0.001
	$Q$	0.79 (0.21)	12.4	3.71	<0.001
	$Q^2$	-0.12 (0.04)	9.8	-3.07	0.004
Catostomidae (adjust $R^2 = 0.62$ , $n = 31$ )	$R_{reg}$	0.52 (0.09)	70.0	3.99	<0.001
	$Q$	0.22 (0.04)	12.2	3.63	0.001
	$R_{reg} \times Q$	0.56 (0.14)	17.8	3.08	0.004
Centrarchidae (adjust $R^2 = 0.70$ , $n = 29$ )	$R_{reg}$	0.80 (0.31)	35.1	2.15	0.018
	$R_{other}$	0.54 (0.11)	32.7	4.67	<0.001
	PC1	-0.06 (0.01)	32.1	-4.08	<0.001
Cyprinidae (adjust $R^2 = 0.59$ , $n = 38$ )	$R_{reg}$	0.51 (0.15)	50.2	3.39	0.002
	$R_{other}$	0.39 (0.12)	40.3	3.34	0.002
	PC2	-0.06 (0.02)	9.4	-2.83	0.008
Percidae (adjust $R^2 = 0.67$ , $n = 25$ )	$R_{other}$	1.07 (0.21)	73.1	5.22	<0.001
	$Q$	0.11 (0.06)	26.9	1.98	0.060
(b) Rarefied local richness group					
All species (adjust $R^2 = 0.39$ , $n = 41$ )	$R_{reg}$	0.44 (0.09)	86.9	5.05	<0.001
	$Q$	0.08 (0.03)	13.1	2.20	0.034
Catostomidae (adjust $R^2 = 0.49$ , $n = 31$ )	$R_{reg}$	0.35 (0.27)	54.4	5.27	<0.001
	$Q$	0.14 (0.10)	11.9	4.74	<0.001
	$R_{reg} \times Q$	0.53 (0.13)	24.6	4.17	<0.001
	$R_{other}$	-0.50 (0.18)	9.1	-2.79	0.010
Centrarchidae (adjust $R^2 = 0.50$ , $n = 29$ )	PC1	-0.06 (0.02)	52.7	-2.96	0.006
	$R_{reg}$	1.23 (0.46)	47.2	2.67	0.013
Cyprinidae (adjust $R^2 = 0.42$ , $n = 38$ )	$R_{reg}$	0.59 (0.14)	90.6	3.95	<0.001
	PC2	-0.04 (0.02)	9.4	-1.76	0.086
Percidae (adjust $R^2 = 0.18$ , $n = 25$ )	$Q$	0.16 (0.06)	-	2.50	0.020

$R_{reg}$  =  $\log_{10}$ -transformed regional species richness;  $R_{other}$  =  $\log_{10}$ -transformed local richness of species other than the given family of interest;  $Q$  =  $\log_{10}$ -transformed discharge ( $\text{m}^3 \text{s}^{-1}$ ).

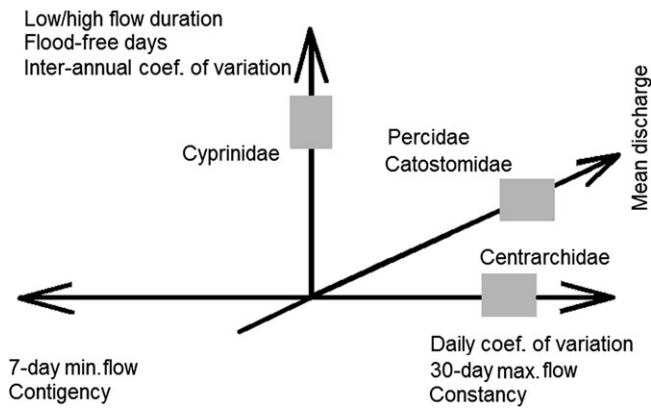


Fig. 2 A diagram representing associations of local species richness with local hydrological characteristics in four freshwater fish families, indicated by the position of each family on the axis composed by hydrological variables.

of Catostomidae, indicating that the positive relationship between local and regional species richness varies among different levels of mean discharge. Hydrological variability appeared to be more important in explaining variation in richness of Centrarchidae and Cyprinidae. In particular, Centrarchidae richness was positively correlated with high flow magnitude relative to median discharge and coefficient of variation in daily discharge (PC1); Cyprinidae richness was positively correlated with low/high flow duration, and to a lesser extent with the coefficient of variation in daily discharge (PC2) (Fig. 2). Richness of other species ( $R_{\text{other}}$ ) was positively correlated with local richness in the models for three families (Centrarchidae, Cyprinidae and Percidae) in the non-rarefied data sets (Table 3).

Rarefied local species richness in the 'All species' analysis was best predicted by regional species richness and discharge volume; the selected linear models explained 18–59% of the variation in local within-family richness (Table 3b). Results of analyses of rarefied data sets are largely in accordance with original data sets, except that rarefied data sets generally have lower explanatory power, and  $R_{\text{other}}$  was not a significant predictor of variation in local richness in three of the four families (except Catostomidae). Consistent with the original data set, hydrological variability (PC1 and PC2) was an important predictor for Centrarchidae and Cyprinidae, and discharge volume tended to be more important for Catostomidae and Percidae.

According to results from the cross-validation techniques, models based on original data sets yielded more reliable predictions than rarefied data sets, and considerable prediction accuracy was seen in the original local richness analyses for the 'All species' data set as well as in

Centrarchidae and Cyprinidae (Fig. S1). In contrast, apparent deviation among regressions generated by training groups and departure of test data from these regression equations were observed in Catostomidae and Percidae, especially in rarefied data sets, suggesting less reliable predictions in these two families.

## Discussion

We anticipated that deterministic processes associated with hydrological characteristics would predict variation in local species richness in stream fish assemblages when the constraints of the regional species pool were considered. Using multiple regressions, we found that the regional species pool does play an important role in predicting local species richness, and hydrological characteristics explain an additional proportion of variation. In addition, species richness of other families present in local fish assemblages also appeared to be important in explaining local richness of each family, possibly serving as an indicator of the overall quality and heterogeneity of the local physical habitat.

With the exception of Percidae, the regional species pool explained 50–70% (original data set) and 54–90% (rarefied data set) of the total explained variation in local family-level richness. These results support the notion that local assemblages are limited by the regional species pool, which is an outcome of biogeographic processes attributed to climate, geography and phylogenetic history (Cornell & Lawton, 1992; Ricklefs & Schluter, 1993). Similarly, regional factors tend to be a primary determinant in local richness of periphytic diatoms (Hillebrand *et al.*, 2001; Soininen, Paavola & Muotka, 2004; but see Pan *et al.*, 2000), macroinvertebrates (Heino, 2001; Grönroos & Heino, 2012) and fishes (Angermeier & Winston, 1998; Marsh-Matthews & Matthews, 2000).

High and low flow duration and hydrological variability (PC1 and PC2) accounted for 9–53% of the explained variation in local species richness of Centrarchidae and Cyprinidae. The models predicting family-level local richness suggested that local richness of each family was associated with the regional species pool and local hydrological factors to various degrees. Taxon-specific responses to hydrological characteristics were also observed by Schlosser (1982), where assemblages of suckers (Catostomidae) and darters (Percidae) were not correlated with flow conditions in two streams in Illinois, while abundance of juvenile cyprinids was negatively associated with floods. Although high flow events cause mortality of juvenile fishes (Schlosser, 1982), positive effects of flow variability on cyprinid recruitment and

richness were observed in a French river (Cattanéo *et al.*, 2001). Temporal variation in recruitment induced by flow events is likely to be responsible for the positive association of species richness with duration of flow events and flow variability. Stochastic changes in recruitment success coupled with varying flow conditions favour different groups of populations and species and thus enhance species coexistence (i.e. storage effects: Warner & Chesson, 1985; Chesson, 2000).

A framework based on life-history traits of fish species may also help to interpret taxon-specific associations with hydrological characteristics. Fish species can be categorised into three groups based on demographic parameters related to recruitment and fecundity (i.e. 'periodic-opportunistic-equilibrium trichotomy') (Winemiller, 2005; Mims *et al.*, 2010; Olden & Kennard, 2010). Opportunistic species, such as shiners (Cyprinidae), are mainly early breeders with low fecundity per spawning and low juvenile survivorship. Periodic groups such as suckers (Catostomidae) are mainly large-bodied species with late maturation, high fecundity, but low juvenile survivorship. Equilibrium groups show low fecundity per spawning event but high juvenile survivorship (i.e. provide parental care), as exemplified by bass and sunfish (Centrarchidae; Mims *et al.*, 2010). We observed positive relationships between local species richness of Centrarchidae (equilibrium) and Cyprinidae (opportunistic) and temporal flow variability as well as flow event durations. Patterns of life-history composition of river basins in the United States and Australia responded similarly along a gradient of hydrological variability, which was also positively associated with opportunistic strategies and negatively associated with periodic strategies (Olden & Kennard, 2010).

Discharge volume appeared to be an important predictor of local species richness in Catostomidae and Percidae. Fish species richness has been documented to increase with habitat volume among small streams (Angermeier & Schlosser, 1989) and increased with surface discharge (Xenopoulos & Lodge, 2006) or drainage basin area (Oberdorff *et al.*, 1997) among larger river basins. As expected, stream discharge in many regions has been reduced because of anthropogenic hydrological regulation, resulting in severe declines in freshwater biodiversity (Vörösmarty *et al.*, 2010).

Our results indicate inconsistencies between original and rarefied data sets in terms of explanatory power of the optimal models and relative importance of richness of other species ( $R_{\text{other}}$ ). When using the original data sets, family-level richness was positively correlated with  $R_{\text{other}}$ , probably due to positive effects of local habitat quality and

heterogeneity on local richness. Our statistical analyses do not permit inference of ecological processes, and this tentative explanation remains speculative. Moreover, the low explanatory power in models generated by rarefied data sets seems counterintuitive. Rarefied data sets should represent a 'core proportion' of local species (common species), while original data sets represent a mixture of common and rare species; one may expect common species to show a higher degree of affinity to environmental conditions (less subject to stochastic processes). It is noticeable that the distribution of rarefied species richness was strongly skewed to lower values (except Cyprinidae), and there may be an insufficient degree of variation in the dependent variable to reveal any pattern under the present statistical approach. Therefore, we interpret that the low explanatory power may be a reflection of weak patterns in rarefied data sets or a statistical outcome partially because of the skewed data distribution.

Chinnayakanahalli *et al.* (2011) recently conducted a similar analysis of the influence of local hydrology (as well as water temperature) on stream invertebrate species richness using 63 sites from the NAWQA data set. Although direct comparisons of results are difficult because of differences in methodologies and taxonomic groups, the relative contributions of local hydrology to variation in species richness among stream macroinvertebrates and fishes appear to be generally similar. However, the inclusion of regional species richness data in this study allowed for greater predictive ability of variation in local species richness, suggesting the importance of historical contributions to local processes (Wiens & Donoghue, 2004). The importance of the regional species pool was also identified in another study on macroinvertebrates from streams in Finland (Grönroos & Heino, 2012), where regional species richness and local habitat characteristics were used to predict local macroinvertebrate species richness. Linear local-versus-regional richness relationships were found in most of the groups, while significant amounts of variation in local species richness were attributed to local habitat characteristics.

There is still a substantial proportion of unexplained variation in the regression models, and models for Catostomidae and Percidae did not provide reliable predictions. We acknowledge that stochastic components (neutral processes, such as chance colonisation, random extinction and stochastic changes in population abundance) are important in determining local assemblage structure, but the present framework was not able to address this perspective. For instance, Thompson & Townsend (2006) revealed that macroinvertebrate assemblages in ten local streams are influenced by both environmental conditions



and spatial distance, and the influences varied among dispersal and trophic groups. Patch-dynamics hypotheses, based on metacommunity theory, again support the notion that both the spatial distribution of habitat patches and deterministic factors (species traits) contribute to the structuring of local fish assemblages (Arrington & Winemiller, 2006). The spatial configuration of stream networks affects the association of local fish assemblages with environmental factors and dispersal limitation (Brown & Swan, 2010), and the existence of dispersal barriers shapes the distribution and evolutionary history of lineages (Poissant, Knight & Ferguson, 2005). Moreover, stochastic ecological processes can influence the importance of deterministic processes in generating assemblage structure, and the relative importance of stochastic and deterministic factors varies depending on, for example, productivity and environmental perturbations on the system (Chase, 2007; Chase & Myers, 2011). Nevertheless, our results revealed family-specific relationships between local hydrology and species richness, and such information derived from broad geographic scales is helpful in generalising hydro-ecological associations in freshwater fishes and points to the value of a species trait-based framework to understand these associations.

As anticipated, regional species richness played a major role in constraining local richness in the total species analyses. However, within-family richness patterns were regulated by the regional species pool and local hydrology to various degrees, with flow variability explaining most of the remaining variation in richness of centrarchids and cyprinids, and discharge volume explaining most of the variation in catostomids and percids. The variation in response to hydrological characteristics among families may be due to differences in life-history traits associated with evolutionary constraints within lineages. Based on these findings, current understanding of hydro-ecological relationships will benefit from broad-scale analytical approaches investigating the relative importance of local and regional factors, combined with empirical studies that compare species traits as well as other ecological processes along gradients of hydrological characteristics.

### Acknowledgments

We thank M. Harris (USGS) for providing the NAWQA data. We also thank M. Anthony, M. Chu, M. Michel and two anonymous reviewers for valuable comments on previous drafts of this manuscript. Funding for this research was provided to JHK by the Environmental Protection Agency's (EPA) Science to Achieve Results (STAR) Consequences of Global Change for Water Quality

program (EPA-G2008-STAR-D2) and from the National Science Foundation (DEB-0844644).

### References

- Angermeier P.L. & Schlosser I.J. (1989) Species-area relationship for stream fishes. *Ecology*, **70**, 1450–1462.
- Angermeier P.L. & Winston M.R. (1998) Local vs. regional influences on local diversity in stream fish communities of Virginia. *Ecology*, **79**, 911–927.
- Arrington D.A. & Winemiller K.O. (2006) Habitat affinity, the seasonal flood pulse, and community assembly in the littoral zone of a Neotropical floodplain river. *Journal of the North American Benthological Society*, **25**, 126–141.
- Brown B.L. & Swan C.M. (2010) Dendritic network structure constrains metacommunity properties in riverine ecosystems. *Journal of Animal Ecology*, **79**, 571–580.
- Bunn S.E. & Arthington A.H. (2002) Basic principles and ecological consequences of altered flow regimes for aquatic biodiversity. *Environmental Management*, **30**, 492–507.
- Burnham K.P. & Anderson D.R. (2004) Multimodel inference – understanding AIC and BIC in model selection. *Sociological Methods & Research*, **33**, 261–304.
- Cattanéo F., Carrel G., Lamouroux N. & Breil P. (2001) Relationship between hydrology and cyprinid reproductive success in the Lower Rhône at Montelimar, France. *Archiv für Hydrobiologie*, **151**, 427–450.
- Chase J.M. (2007) Drought mediates the importance of stochastic community assembly. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 17430–17434.
- Chase J.M. & Myers J.A. (2011) Disentangling the importance of ecological niches from stochastic processes across scales. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **366**, 2351–2363.
- Chesson P. (2000) Mechanisms of maintenance of species diversity. *Annual Review of Ecology and Systematics*, **31**, 343–366.
- Chevan A. & Sutherland M. (1991) Hierarchical partitioning. *The American Statistician*, **45**, 90–96.
- Chinnayakanahalli K.J., Hawkins C.P., Tarboton D.G. & Hill R.A. (2011) Natural flow regime, temperature and the composition and richness of invertebrate assemblages in streams of the western United States. *Freshwater Biology*, **56**, 1248–1265.
- Colwell R.K. (1974) Predictability, constancy, and contingency of periodic phenomena. *Ecology*, **55**, 1148–1153.
- Connell J.H. (1978) Diversity in tropical rain forests and coral reefs. *Science*, **199**, 1302–1310.
- Cornell H.V. & Lawton J.H. (1992) Species interactions, local and regional processes, and limits to the richness of ecological communities: a theoretical perspective. *Journal of Animal Ecology*, **61**, 1–12.

- Cross W.F., Baxter C.V., Donner K.C., Rosi-Marshall E.J., Kennedy T.A., Hall R.O. Jr *et al.* (2011) Ecosystem ecology meets adaptive management: food web response to a controlled flood on the Colorado River, Glen Canyon. *Ecological Applications*, **21**, 2016–2033.
- Franssen N.R., Gido K.B. & Propst D.L. (2007) Flow regime affects availability of native and nonnative prey of an endangered predator. *Biological Conservation*, **138**, 330–340.
- Gotelli N.J. & Colwell R.K. (2001) Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letter*, **4**, 379–391.
- Gotelli N.J. & Entsminger G.L. (2004) *EcoSim: Null Models Software for Ecology*. Version 7. Acquired Intelligence Inc. & Kesey-Bear, Jericho, VT. Available at: <http://garyentsminger.com/ecosim/index.htm>.
- Griffiths D. (2012) Body size distributions in North American freshwater fish: large-scale factors. *Global Ecology and Biogeography*, **21**, 383–392.
- Grönroos M. & Heino J. (2012) Species richness at the guild level: effects of species pool and local environmental conditions on stream macroinvertebrate communities. *Journal of Animal Ecology*, **81**, 679–691.
- Heino J. (2001) Regional gradient analysis of freshwater biota: do similar biogeographic patterns exist among multiple taxonomic groups? *Journal of Biogeography*, **28**, 69–76.
- Hillebrand H., Watermann F., Karez R. & Berninger U.G. (2001) Differences in species richness patterns between unicellular and multicellular organisms. *Oecologia*, **126**, 114–124.
- Kleinbaum D.G., Kupper L.L., Nizam A. & Muller K.E. (2008) *Applied Regression Analysis and Other Multivariable Methods*, 4th edn. Thomson Brook/Cole, Belmont, CA.
- Knouft J.H. (2004) Latitudinal variation in the shape of the species body size distribution: an analysis using freshwater fishes. *Oecologia*, **139**, 408–417.
- Knouft J.H. & Page L.M. (2003) The evolution of body size in extant groups of North American freshwater fishes: speciation, size distributions, and Cope's rule. *The American Naturalist*, **141**, 413–421.
- Knouft J.H. & Page L.M. (2011) Assessment of the relationships of geographic variation in species richness to climate and landscape variables within and among lineages of North American freshwater fishes. *Journal of Biogeography*, **38**, 2259–2269.
- Leahy P.P., Rosenshein J.S. & Knopman D.S. (1990). Implementation plan for the National Water-Quality Assessment Program: United States geological survey, Open-File Report 90-174.
- Lytle D.A., Bogan M.T. & Finn D.S. (2008) Evolution of aquatic insect behaviours across a gradient of disturbance predictability. *Proceedings of the Royal Society B: Biological Sciences*, **275**, 453–462.
- Lytle D.A. & Poff N.L. (2004) Adaptation to natural flow regimes. *Trends in Ecology & Evolution*, **19**, 94–100.
- Marsh-Matthews E. & Matthews W.J. (2000) Geographic, terrestrial and aquatic factors: which most influence the structure of stream fish assemblages in the midwestern United States? *Ecology of Freshwater Fish*, **9**, 9–21.
- Meffe G.K. (1984) Effects of abiotic disturbance on coexistence of predator-prey fish species. *Ecology*, **65**, 1525–1534.
- Mims M.C., Olden J.D., Shattuck Z.R. & Poff N.L. (2010) Life history trait diversity of native freshwater fishes in North America. *Ecology of Freshwater Fish*, **19**, 390–400.
- Mitchell A.L. & Knouft J.H. (2009) Non-native fishes and native species diversity in freshwater fish assemblages across the United States. *Biological Invasions*, **11**, 1441–1450.
- Moulton S.R., Kennen J.G., Goldstein R.M. & Hambrook J.A. (2002) Revised protocols for sampling algal, invertebrate, and fish communities as part of the National Water-Quality Assessment Program: United States geological survey. Open-File Report 02-150.
- Oberdorff T., Huguéy B. & Guegan J.F. (1997) Is there an influence of historical events on contemporary fish species richness in rivers? Comparisons between Western Europe and North America *Journal of Biogeography*, **24**, 461–467.
- Olden J.D. & Kennard M.J. (2010) Intercontinental comparison of fish life history strategies along a gradient of hydrologic variability. In: *Community Ecology of Stream Fishes: Concepts, Approaches, and Techniques* (Eds K.B. Gido & D.A. Jackson), *American Fisheries Society Symposium* **73**, 83–107.
- Olden J.D. & Poff N.L. (2003) Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. *River Research and Applications*, **19**, 101–121.
- Page L.M. & Burr B.M. (2011) *Peterson Field Guide to Freshwater Fishes of North America North of Mexico*, 2nd edn. Houghton Mifflin Harcourt Publishing Company, New York, NY.
- Pan Y, Stevenson R.J., Hill B.H. & Herlihy A.T. (2000) Ecoregions and benthic diatom assemblages in Mid-Atlantic Highlands streams, USA. *Journal of the North American Benthological Society*, **19**, 518–540.
- Picard R.R. & Cook R.D. (1984) Cross-validation of regression models. *Journal of the American Statistical Association*, **79**, 575–583.
- Piffady J., Souchon Y., Capra H. & Parent E. (2010) Quantifying the effects of temperature and flow regime on the abundance of 0 + cyprinids in the upper River Rhone using Bayesian hierarchical modelling. *Freshwater Biology*, **55**, 2359–2374.
- Poff N.L. (1997) Landscape filters and species traits: towards mechanistic understanding and prediction in stream ecology. *Journal of the North American Benthological Society*, **16**, 391–409.
- Poissant J., Knight T.W. & Ferguson M.M. (2005) Nonequilibrium conditions following landscape rearrangement: the relative contribution of past and current hydrological landscapes on the genetic structure of a stream-dwelling fish. *Molecular Ecology*, **14**, 1321–1331.

- Power M.E., Parker M.S. & Dietrich W.E. (2008) Seasonal reassembly of a river food web: floods, droughts, and impacts of fish. *Ecological Monographs*, **78**, 263–282.
- R Development Core Team (2012) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL: <http://www.R-project.org/>.
- Richter B.D., Baumgartner J.V., Wigington R. & Braun D.P. (1997) How much water does a river need? *Freshwater Biology*, **37**, 231–249.
- Ricklefs R.E. (1987) Community diversity: relative roles of local and regional processes. *Science*, **235**, 167–171.
- Ricklefs R.E. & Schluter D. (1993) Species diversity: regional and historical influences. In: *Species Diversity in Ecological Communities: Historical and Geographical Perspectives* (Eds R.B. Ricklefs & D. Schluter), pp. 350–363. University of Chicago Press, Chicago, IL.
- Schlosser I.J. (1982) Fish community structure and function along two habitat gradients in a headwater stream. *Ecological Monographs*, **52**, 395–414.
- Soininen J., Paavola R. & Muotka T. (2004) Benthic diatom communities in boreal streams: community structure in relation to environmental and spatial gradients. *Ecography*, **27**, 330–342.
- Thompson R.M. & Townsend C.R. (2006) A truce with neutral theory: local deterministic factors, species traits and dispersal limitation together determine patterns of diversity in stream invertebrates. *Journal of Animal Ecology*, **76**, 476–484.
- Vörösmarty C.J., McIntyre P.B., Gessner M.O., Dudgeon D., Prusevich A., Green P. *et al.* (2010) Global threats to human water security and river biodiversity. *Nature* **467**, 555–561.
- Walsh S.J. & Meador M.R. (1998) Guidelines for quality assurance and quality control of fish taxonomic data collected as part of the National Water-Quality Assessment Program: United States geological survey, Water-Resources Investigations Report 98-4239.
- Ward J.V. (1998) Riverine landscapes: biodiversity patterns, disturbance regimes, and aquatic conservation. *Biological Conservation*, **83**, 269–278.
- Warner R.R. & Chesson P.L. (1985) Coexistence mediated by recruitment fluctuations: a field guide to the storage effect. *The American Naturalist*, **125**, 769–787.
- Wiens J.J. & Donoghue M.J. (2004) Historical biogeography, ecology, and species richness. *Trends in Ecology and Evolution*, **19**, 639–644.
- Wiens J.J. & Graham C.H. (2005) Niche conservatism: integrating evolution, ecology, and conservation biology. *Annual Review of Ecology, Evolution, and Systematics*, **36**, 519–539.
- Winemiller K.O. (2005) Life history strategies, population regulation, and implications for fisheries management. *Canadian Journal of Fisheries and Aquatic Sciences*, **62**, 872–885.
- 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.
- Yamashita T., Yamashita K. & Kamimura R. (2007) A stepwise AIC method for variable selection in linear regression. *Communications in Statistics-Theory and Methods*, **36**, 2395–2403.

### Supporting Information

Additional Supporting Information may be found in the online version of this article:

**Fig. S1** Plot results from fourfold cross-validation of models predicting local species richness, with entire assemblages as well as each of the four major families, based on original (a–e) and rarefied (f–j) datasets.

**Table S1** Loadings of each hydrological index for the first three principle components (PCs) for the four major families.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

(Manuscript accepted 16 August 2012)