Supplementary Data for: Effects of rare taxa on fish index of biotic integrity results in a southeastern US biodiversity hotspot

Justin C. Bagley [1,2,3,*] · Patrick E. O'Neil [4] · Elizabeth S. Huff [1]

[1] *Environmental Indicators Section, Field Operations Division, Alabama Department of Environmental Management, P.O. Box 301463, Montgomery, AL 36130, USA*
justin.bagley@adem.alabama.gov; *https://orcid.org/0000-0001-6737-8380*

[2] *Department of Biology and Environmental Sciences, Auburn University at Montgomery, 7061 Senators Dr., Rm. 209 O GH, Montgomery, AL 36117, USA*

[3] *Department of Biology, Virginia Commonwealth University, 1000 W Cary St., Suite 126, Richmond, VA 23284, USA*

[4] *Geological Survey of Alabama, Walter B. Jones Hall, 420 Hackberry Lane, Tuscaloosa, AL 35487, USA (retired)*
[*] *Corresponding author.*

Data S1. Supplementary Methods, Results, & Discussion

Note: references not cited in the main text are presented in the References section of this document; see main text for all other cited works. Abbreviations follow the main text.

Supplementary Methods

As noted in the main text, Alabama currently has five fish biogeographic provinces, or 'ichthyoregions'. The initial state ichthyoregion classification was developed by O'Neil and Shepard (2007) based on cluster analysis of Alabama fish community survey data, and the classification was later updated in subsequent studies (e.g., O'Neil and Shepard 2010, 2011a,b). The different ichthyoregions of the state generally contain internally cohesive species assemblages, physiographic attributes, and geologic settings, as well as similar stream types (O'Neil and Shepard 2007), and one fish IBI tool has been calibrated for each ichthyoregion (Table 1). A map of Alabama's ichthyoregions is provided for reference in Fig. S1 *below*.

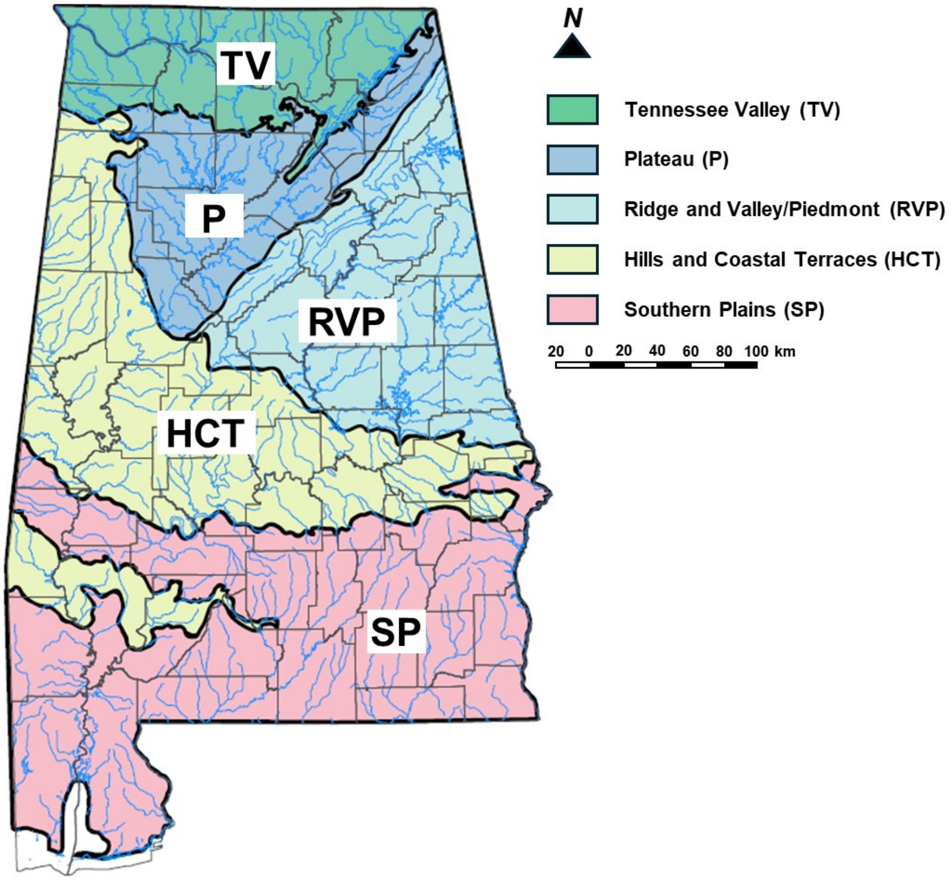

Fig. S1 Map of the five ichthyoregions of Alabama. Ichthyoregions refer to distinct, yet internally cohesive, fish faunal assemblages. The map *above* was modified and partly redrawn after O'Neil and Shepard (2011b), while maintaining their color scheme (see legend).

Wilcoxon signed-rank tests were conducted to evaluate the overall significance of differences in metric-level regression parameters, including goodness-of-fit statistics (i.e., R^2) and slopes, between the SDR and ART treatments, as the data were non-normal (typically left-skewed) and paired (same metric for each treatment). A valid test was possible when comparing post-SDR and post-ART R^2 distributions from the regression models; however, comparisons including SRT R^2-values would not be valid. The reason for this was that the SRT treatment matrix that was used to calculate IBI outcomes was a long-format resampling of the original data matrix with n = 3,007 sites/rows, making the observations violate the common statistical assumption that data are independent and identically distributed. Tests were conducted in the R statistical environment (R Core Team 2023) using the `wilcox.test()` function, and results were considered significant at the α = 0.05 level.

Supplementary Results

Data and treatment characteristics

The sampled fish taxa included 78 North American minnow species (Leuciscidae) including 50 'shiners' from genera *Alburnops*, *Coccotis*, *Cyprinella*, *Ericymba*, *Luxilus*, *Lythrurus*, *Miniellus*, *Notropis*, *Paranotropis*, and *Pteronotropis*; 18 sucker species (Catostomidae); ~18–22 sunfish and black basses (Centrarchidae); 68 darter species (Percidae: Etheostomatinae); and 8 madtom species (Ictaluridae: *Noturus*; Bagley et al. 2023). The 11 species that were globally rare in the fish community data matrix analyzed in this study included one lamprey (*Lethenteron appendix*; Petromyzontidae), one herring (*Alosa chrysochloris*; Dorosomatidae), two minnows (*Cyprinella callitaenia*, *Hybognathus hayi*; Leuciscidae), one sucker (*Carpiodes carpio*; Catostomidae), one madtom (*Noturus* sp. cf. *flavus*; Ictaluridae), four darters—*Etheostoma boschungi*, *Etheostoma trisella*, *Nothonotus* (*Etheostoma*) *wapiti*, and *Percina phoxocephala* (Percidae), and the introduced Asian pond loach (*Misgurnus anguillicaudatus*; Cobitidae; Bagley et al. 2023). The leuciscid minnows of Alabama were all formerly lumped within Cyprinidae, and these taxa are still referred to informally as 'cyprinids' by some biologists as well as in our Alabama fish IBI metric names (e.g., Tables 1 and S1; O'Neil et al. 2006; O'Neil and Shepard 2010), which were published before the taxonomic changes placing North American minnows in Leuciscidae took effect (Tan and Armbruster 2018; Stout et al. 2022; Bagley et al. 2023).

Effects of rare taxon removal on IBI scores and condition ratings

Regarding treatment effects on site IBI scores, a feature presented in the text but not expanded upon precisely was that the IQR of scores increased with treatment intensity, from SRT to ART and SDR (Table 2). Likewise, the ranges of changes in IBI scores varied with treatment intensity. While these changes are visualized in Figs 2A-C of the main text and supplementary Fig. S2 *below*, the raw ranges were from −8 to +6 points around the original IBI scores following the SRT treatment, from −14 to +6 points around the original scores for SDR, and from −12 to +10 points around the original scores for ART. The near-unity regression of post-treatment on

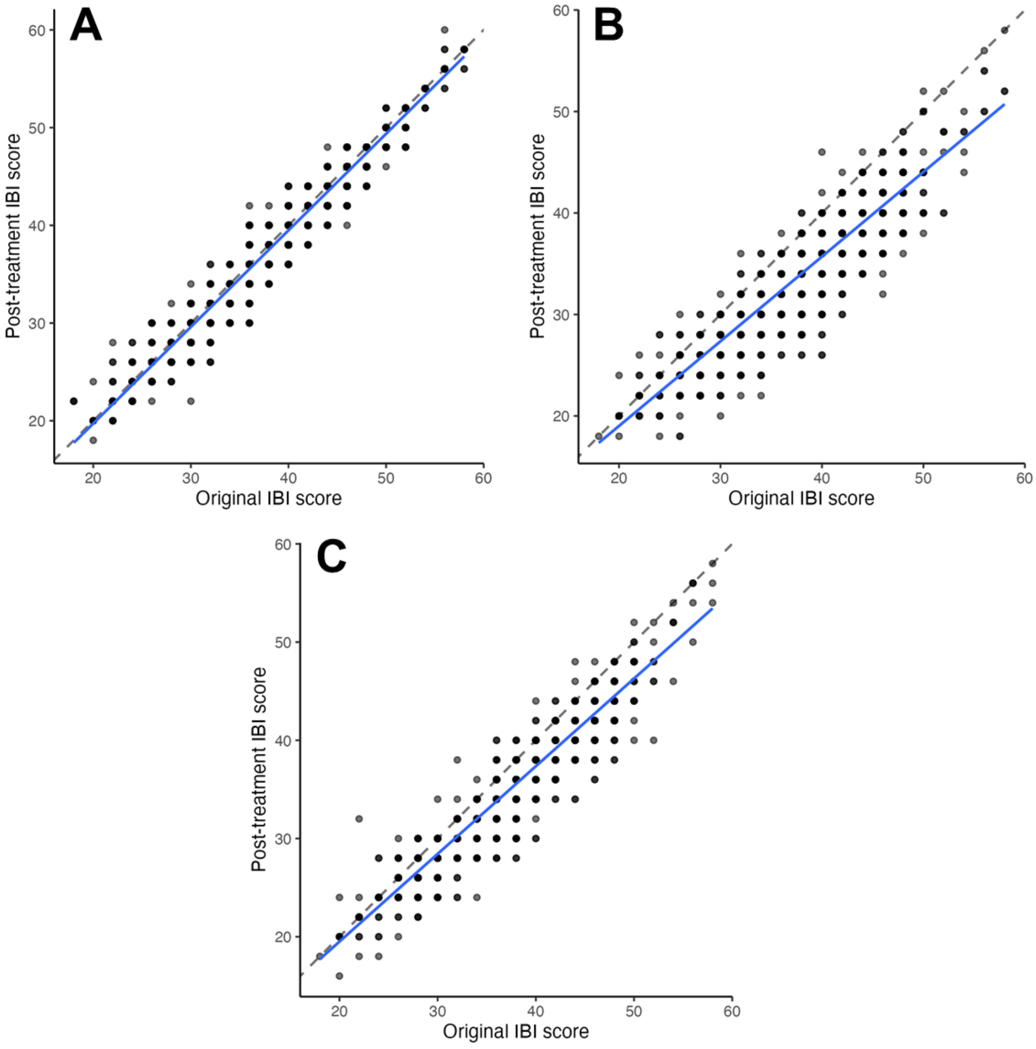

Fig. S2 Scatterplots of post-treatment (y-axis) vs. original site IBI scores (x-axis). Results are shown for each of the three treatments discussed in the main text, including the (A) SRT, (B) SDR, and (C) ART removal treatments. Site results (*transparent gray dots*) are shown alongside the corresponding fitted regression lines (*blue lines*) and a unit (1:1) line for reference (*dashed line*).

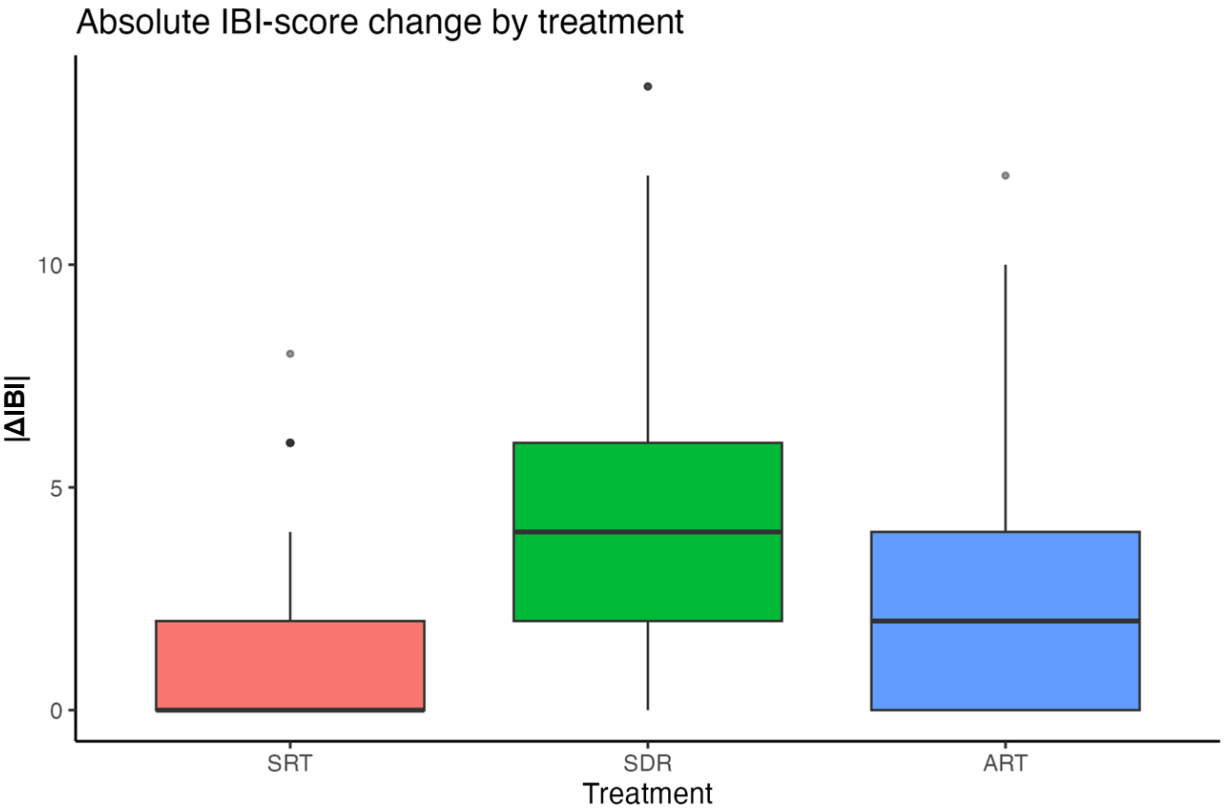

Fig. S3 Box plot summaries of the absolute value of change in IBI score (|ΔIBI|) across sites, presented for each treatment. Here, the raw response variable was calculated as $\Delta\text{IBI} = \text{IBI}_{R,i} - \text{IBI}_{O,i}$, as discussed in the main text. *Vertical lines and dots* represent the range of values, *top and bottom of boxes* indicate the third and first quartiles (75% and 25%), respectively, and the *dark lines* represent median values.

original IBI scores ($R^2 = 0.969$, $a = 0.988$, $b = -0.04$) and a low nRMSE of 2.86% under the SRT treatment confirmed that removing a single rare taxon rarely disrupts the Alabama fish IBIs, overall (e.g., Fig. S2A). In contrast, the IBI score regression following the SDR treatment showed clear compression bias ($R^2 = 0.809$, $a = 0.834$, $b = 2.33$; Fig. S2B), and nRMSE for metric scores after this treatment reached 10.73%. And, last, under the ART treatment, the IBI score regression indicated a modest but consistent downward bias ($R^2 = 0.874$, $a = 0.893$, $b = 1.65$; Fig. S2C), and nRMSE was 7.51%.

Our analyses of the absolute value of changes in IBI scores for sites (|ΔIBI|) provided additional evidence on the effects of rare taxon removal on fish IBI scores and condition ratings. For example, box plots of |ΔIBI| as a function of treatment showed that, as with raw IBI changes and R^2 *above*, the resulting median |ΔIBI| values tracked treatment intensity, increasing from SRT to ART to SDR, and also showed the greatest variance under the SDR treatment (Fig. S3). Furthermore, box plots evaluating |ΔIBI| as a function of our five species richness classes (*S*), six abundance classes (*N*), and five drainage area classes revealed that changes in IBI scores were relatively invariant under the SRT treatment (Figs. S4A,B and S5A), had higher medians and variance under the ART treatment (Figs. S4E,F and S5C), and showed roughly the highest medians and variance under the SDR treatment (Figs. S4C,D and S5B). Within treatments, |ΔIBI|

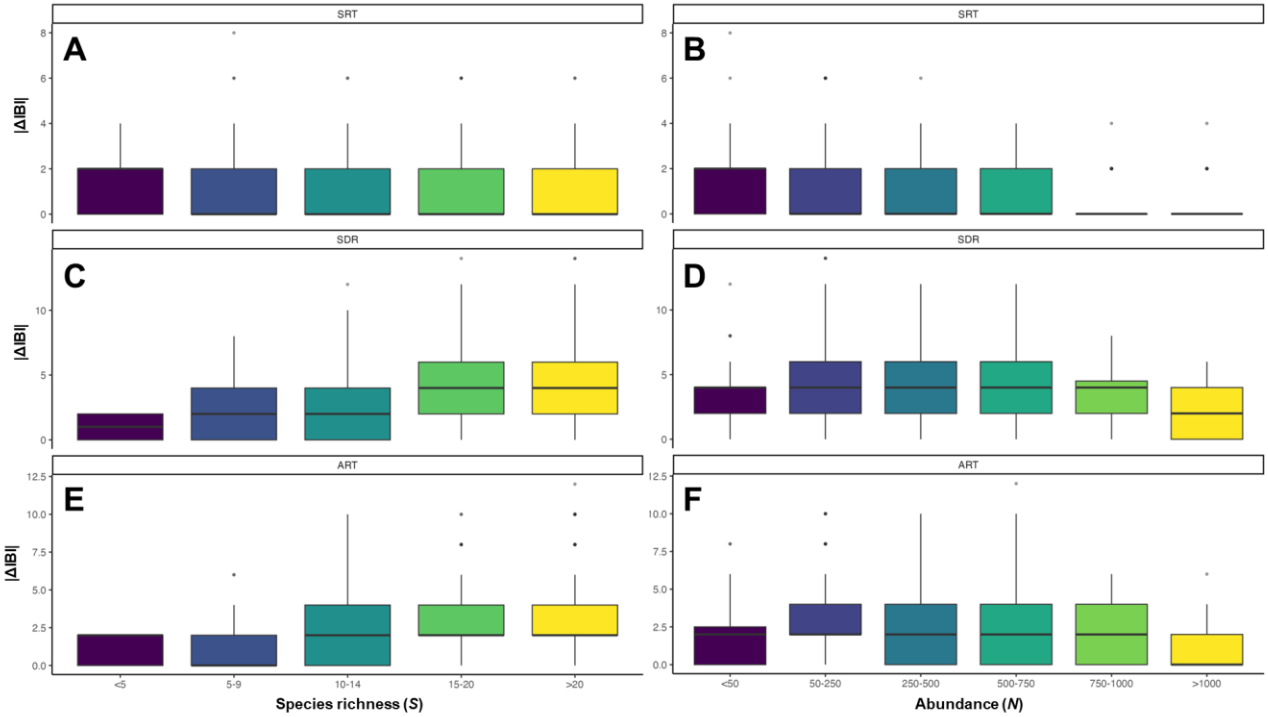

Fig. S4 Summaries of the absolute value of change in IBI score (|ΔIBI|) across sites as a function of sampling intensity classes discussed in the main text (which are the same classes as those shown in Fig. 4). Results are presented as box plots stratified by sample species richness (*S*) and abundance (*N*) for each of the three treatments, including (A, B) SRT removal, (C, D) SDR removal, and (E, F) ART removal. Stratification was based on five species richness classes ($S = \leq 5$, 5–9, 10–14, 15–20, and > 20) and six abundance classes ($N = < 50$, 50–250, 250–500, 500–750, 750–1000, and > 1000). Box representations are the same as those described in the caption to Fig. S3 *above*, except each box is assigned a *different color* by class.

results were particularly robust to variation in drainage area, with little to no change in medians across drainage size classes (Fig. S5). Yet, comparing the *S* and *N* analyses, median |ΔIBI| increased the most for the *S* = 15–20 and *S* > 20 species-richness classes under the SDR treatment (Fig. S4C,D). Compared with the smallest abundance class, median |ΔIBI| rose similarly for the four *N* classes ranging from 50 to 1000 individuals (i.e., *N* =50–250, 250–500, 500–750, and 750–1000 individuals) under the SDR and ART treatments (Fig. S4D,F). and dropped back to lower levels in the largest abundance class (*N* > 1000).

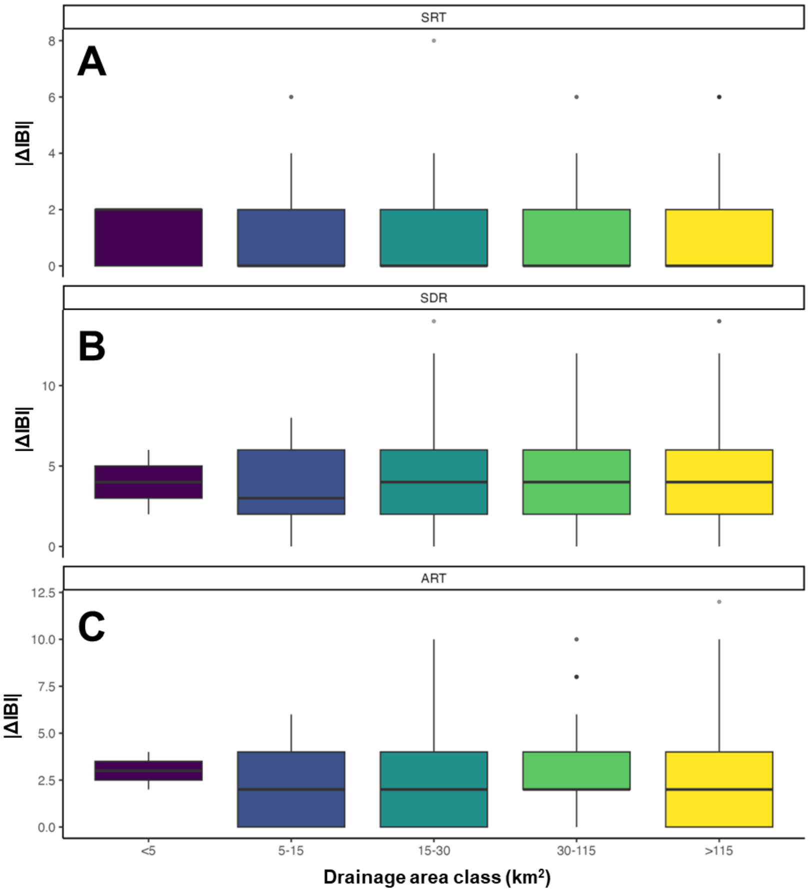

Fig. S5 Summaries of the absolute value of change in IBI score (|ΔIBI|) across sites, presented as a function of drainage area classes discussed in the text (which are the same classes as those shown in plots of nRMSE in Fig. S9 *below*). Results are presented as box plots stratified by site drainage areas for each of the three treatments, including (A) SRT removal, (B) SDR removal, and (C) ART removal. Box representations are the same as those described in the caption to Fig. S3 *above*, except each box is assigned a *different color* based on *x*-axis class.

Per-metric responses and treatment intensity

The full results of linear regression analyses of post-treatment IBI metric scores as a function of original scores are presented in Figs S6–S8 and Table S1 *below*. Linear modeling of metric scores provided additional evidence that the SDR treatment had the greatest effect on IBI metric scores, and the regression results agreed well with the per-metric κ results (Table S2 *below*).

Iteratively removing each rare taxon within each sample in the SRT treatment to create replicate samples (n = 3,007) caused marginal decreases in IBI metric scores, with regression slope parameters and R^2 goodness-of-fit statistics generally falling near 1.0 (mean ± s.d. a = 0.95 ± 0.07, mean ± s.d. R^2 = 0.92 ± 0.08; Fig. S6; Table S1). In line with our a priori expectations, post-SRT removal IBI metric scores more frequently deviated from the original scores for metrics representing measures of species richness (mean ± s.d. R^2 = 0.89 ± 0.06; see metric types in Table 1) when compared with metrics based on species relative abundances (mean ± s.d. R^2 = 0.95 ± 0.09). However, given R^2-values for abundance metric models after SRT were left-skewed, a Welch's t-test was applied (because it allows comparison of means without assuming equal variances) and showed the R^2 distribution for richness metrics was not significantly different than that for abundance-based metrics under this treatment ($t = -1.70$, df = 15.54, p = 0.11). Post-treatment metrics with the greatest deviations from the original scores based on R^2 values were '% DELT + hybrids' followed by several richness measures, including 'Number of centrarchid species', 'Total native species', 'Number of *Lepomis* species', and 'Number of sucker species' (Fig. S6; Table S1). Intercepts, b, for all SRT regressions were close to zero but positive, indicating post-treatment IBI metric scores increased slightly after removals.

After removing singleton/doubleton taxa in the SDR treatment, IBI metric scores decreased to a greater extent than for SRT, causing the regression slope and R^2 statistics to reach

Table S1 Linear regression results showing the fit between post-treatment and original IBI metric scores, following three removal treatments. Regressions were performed separately on results from 1) removing single rare taxa within each sample, 2) removing singleton/doubleton taxa, and 3) removing all rare taxa. The "Metric label" column gives abbreviated names matching Figs S6–S8 *below* (for full names, see Table 1, main text).

	Single-rare-taxon removal				**Singleton/doubleton removal**				**All-rare-taxa removal**			
Metric label	***N***	**R^2**	***a***	***b***	***N***	**R^2**	***a***	***b***	***N***	**R^2**	***a***	***b***
Total native spp.	3007	0.82	0.93	0.082	719	0.34	0.55	0.28	719	0.44	0.66	0.35
No. shiner spp.	1885	0.95	0.97	0.045	444	0.64	0.76	0.26	444	0.79	0.87	0.17
No. cyprinid spp.	1122	0.91	0.95	0.050	274	0.58	0.65	0.33	273	0.70	0.78	0.19
No. sucker spp.	2252	0.90	0.92	0.067	546	0.42	0.47	0.51	545	0.58	0.63	0.34
No. darter+madtom spp.	3007	0.90	0.96	0.037	719	0.53	0.64	0.26	719	0.65	0.77	0.19
No. Lepomis spp.	1877	0.87	0.94	0.16	447	0.32	0.61	0.89	447	0.53	0.75	0.62
No. centrarchid spp.	618	0.78	0.87	0.20	146	0.33	0.48	0.33	146	0.40	0.59	0.33
No. intolerant spp.	1267	0.95	0.96	0.054	298	0.69	0.75	0.31	298	0.79	0.83	0.24
% tolerant spp.	2997	0.98	0.99	0.012	716	0.94	0.98	0.11	717	0.97	0.99	0.04
% dominant spp.	1122	0.99	1.00	-0.001	274	0.90	0.96	0.14	273	0.95	0.98	0.07
% Lepomis	2389	0.98	0.99	0.033	573	0.96	0.99	0.05	573	0.98	1.00	0.03
% GSF+YB	618	0.92	0.95	0.25	145	0.55	0.70	1.61	146	0.72	0.84	0.93
% invertivores	1130	0.99	0.99	0.004	270	0.95	0.99	0.07	271	0.98	0.99	0.02
% omnivores	1695	0.99	1.00	-0.010	433	0.98	1.00	-0.04	433	1.00	1.00	-0.01
% insect. cyprinids	2067	0.98	1.00	0.016	457	0.91	0.99	0.12	458	0.96	1.00	0.05
% top carnivores	3007	0.93	0.98	-0.006	718	0.61	0.82	-0.01	719	0.77	0.92	-0.03
% DELT+hybrids	3007	0.71	0.72	1.39	718	0.19	0.20	4.01	719	0.39	0.42	2.92
No. lithophilic spawners	1885	0.93	0.96	0.037	444	0.61	0.73	0.24	444	0.73	0.83	0.14
% simple lithophils	1122	0.99	0.99	0.008	274	0.97	0.99	0.04	273	0.99	1.00	0.01

Regression abbreviations: *a*, slope; *b*, intercept; *N*, number of samples where metric was applied; R^2, coefficient of determination (goodness-of-fit statistic). Other abbreviation definitions follow the text and Table 1.

their lowest values among any removal treatment results (mean ± s.d. $a = 0.75 \pm 0.23$, mean ± s.d. $R^2 = 0.65 \pm 0.26$; Fig. S7; Table S1). Post-treatment IBI metric scores deviated more frequently from the original scores for richness-based metrics (mean ± s.d. $R^2 = 0.50 \pm 0.15$) as compared to relative abundance metrics (mean ± s.d. $R^2 = 0.80 \pm 0.26$) for this second treatment. Again Welch's *t*-test was appropriate, but in this case differences in R^2 values between these two metric types were statistically significant ($t = -3.13$, df = 14.28, $p < 0.01$). The post-treatment IBI metrics with the greatest deviations from original metric scores after the SDR treatment were the same five listed above for the SRT treatment. Intercept parameters for regressions corresponding to this second treatment were ~1.4-fold to ~99-fold higher than those for the SRT removal treatment, consistent with a greater depression of IBI metric scores for larger original IBI metric scores and/or an increase in post-treatment IBI metric scores for lower original IBI metric scores. Two exceptional cases, with negative regression intercepts, were the '% omnivores' metric ($b = -0.04$) and the '% top carnivores' metric ($b = -0.01$; Fig. S7; Table S1). The highest intercept value was $b = 4.01$ for the '% DELT+ hybrids' metric, indicating a greater decline in post-treatment IBI metric scores for this metric, as compared with other metrics.

After removing all rare taxa in the ART treatment, IBI metric scores decreased such that regression slope and R^2 statistics fell noticeably below 1.0 (mean ± s.d. $a = 0.83 \pm 0.17$, mean ± s.d. $R^2 = 0.75 \pm 0.21$; Fig. S8; Table S1). Again, consistent with our expectations based on prior work (Wan et al. 2010), the post-treatment IBI metric scores more frequently deviated from the original scores for species richness metrics (mean ± s.d. $R^2 = 0.62 \pm 0.15$) as compared with those based on relative abundances (mean ± s.d. $R^2 = 0.87 \pm 0.20$). A Welch's *t*-test showed that differences in R^2 values between these two metric types under ART were highly statistically significant ($t = -3.16$, df = 16.45, $p < 0.01$). The five post-treatment IBI metrics with the greatest

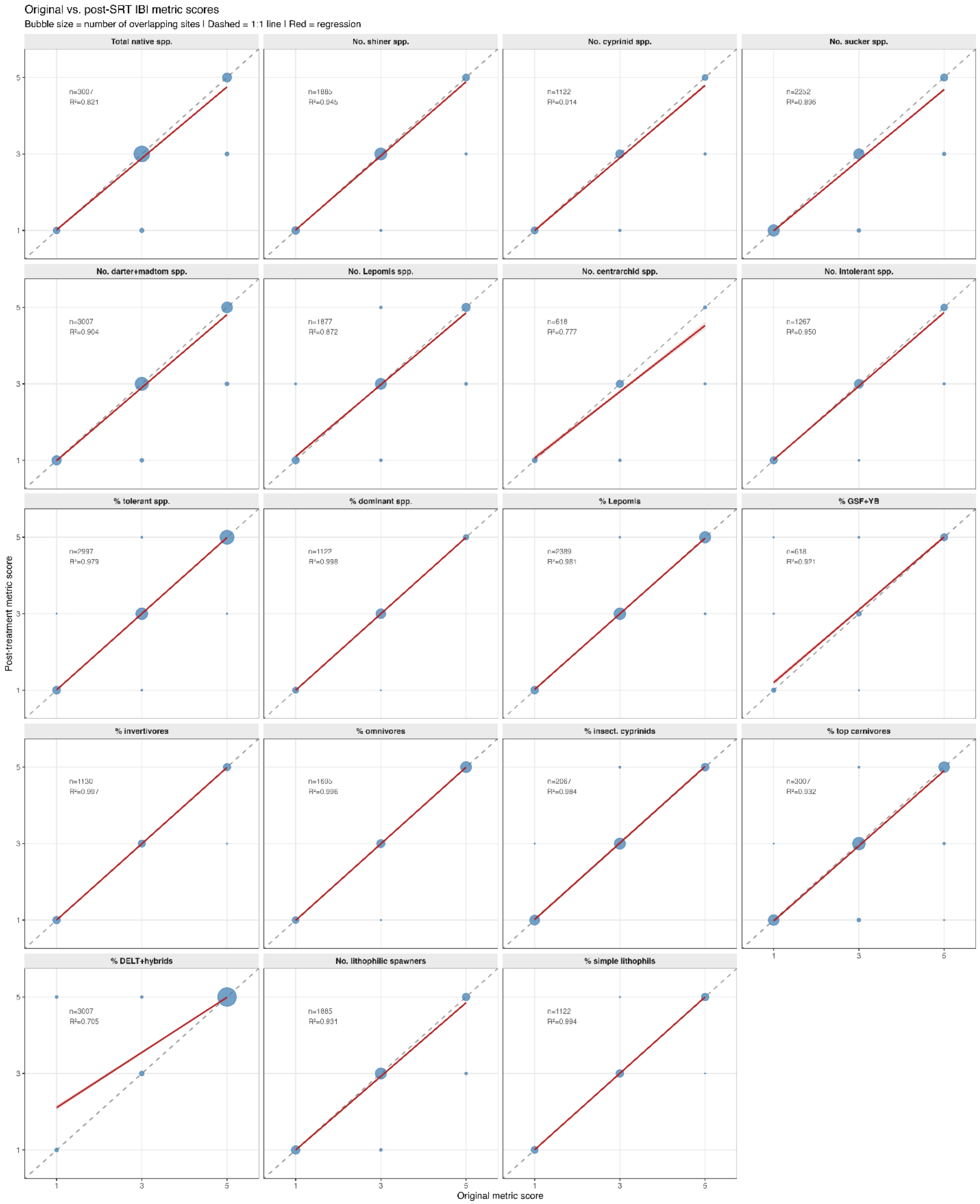

Fig. S6 Plots of post-treatment (*y*-axis) versus original (*x*-axis) IBI metric scores following the single-rare-taxon (SRT) removal treatment. Results are presented as one panel for each of the 19 metrics in this study. Raw values are shown as *transparent blue bubbles*, with bubble size proportional to the number of overlapping observations (sites), while *red lines* display the fitted regression lines, and a *gray dashed 1:1 line* is shown for reference. Sample sizes and coefficients of determination (R^2) are inset in the upper left of each panel, and panel headings give the corresponding metric names.

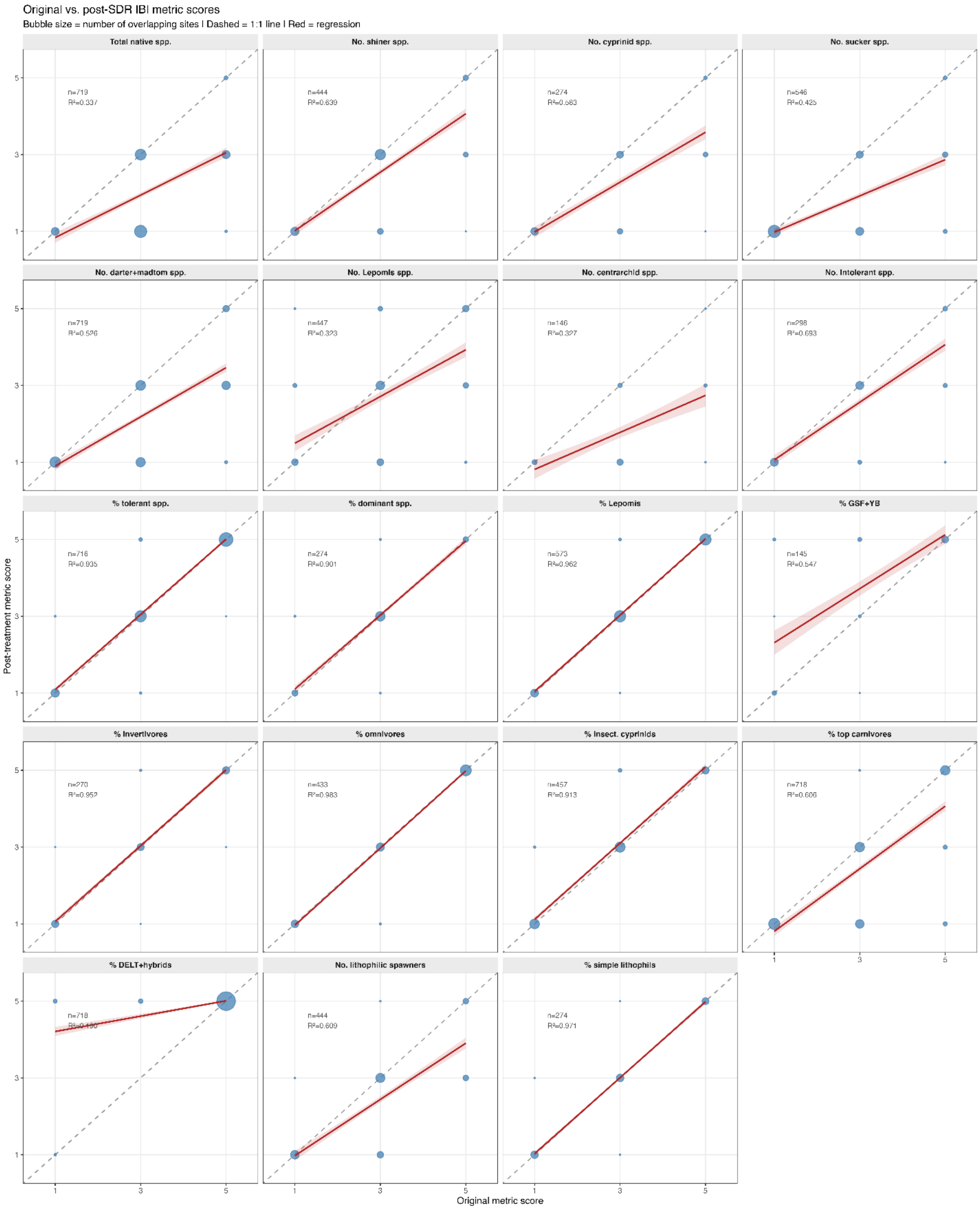

Fig. S7 Plots of post-treatment versus original IBI metric scores following singleton/doubleton removal (SDR). Panel formatting follows that in Fig. S6 *above* (see caption of that figure for additional details).

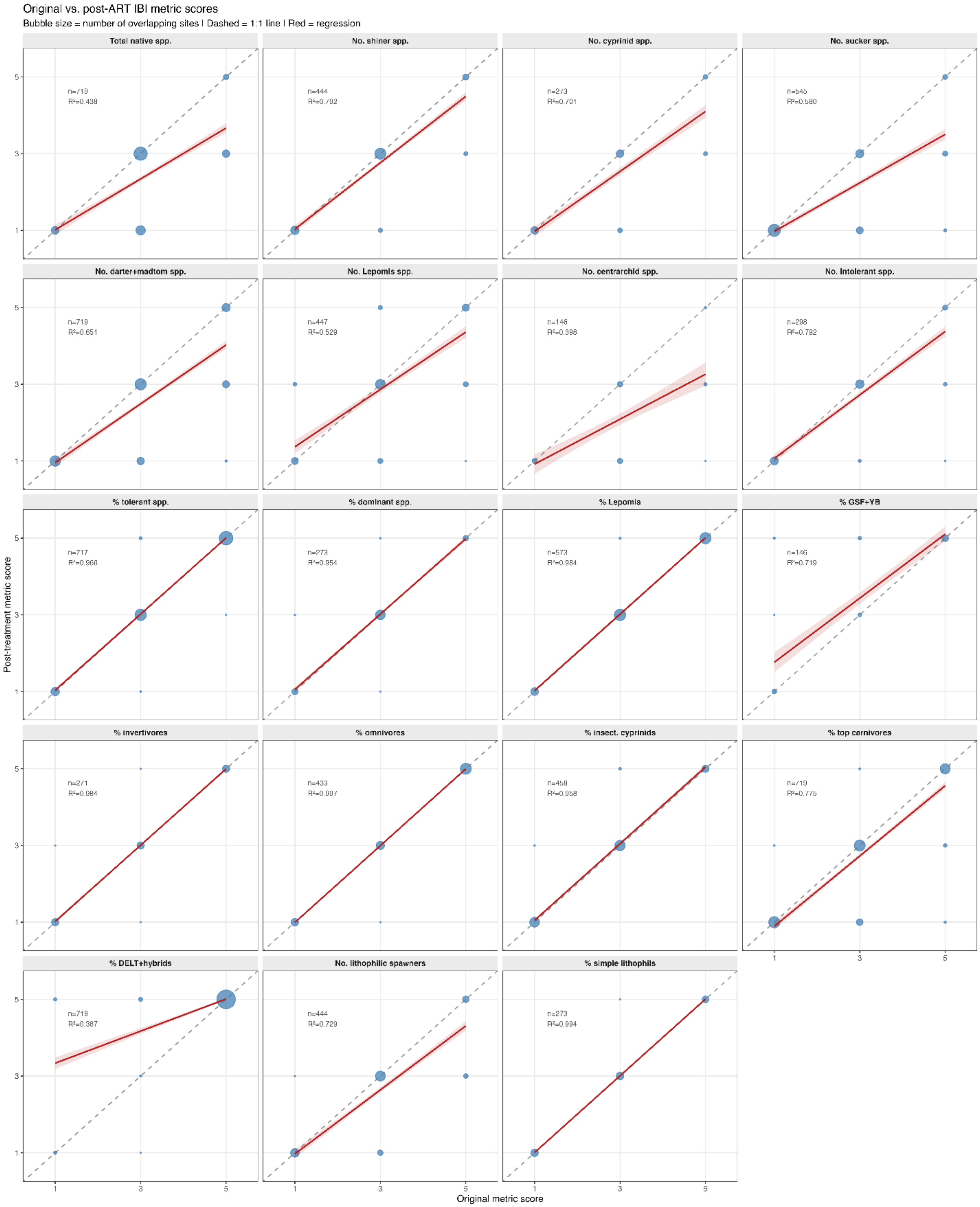

Fig. S8 Plots of post-treatment versus original IBI metric scores following the all-rare-taxa

(ART) removal treatment. Panel formatting follows that in Fig. S6 *above* (see caption of that figure for additional details).

deviations from original metric scores were again '% DELT + hybrids', 'Number of centrarchid species', 'Total native species', 'Number of *Lepomis* species', and 'Number of sucker species' (Table S1). Intercept parameters, *b*, for regressions corresponding to this third removal treatment were up to ~51-fold higher than those for the first removal experiment (median 3.7-fold higher), consistent with a substantial effect on IBI metric scores in essentially all cases. The only exception to this was the '% top carnivores metric', which had a negative intercept ($b = –0.03$) after removal (Table 2). The highest intercept value, $b = 2.92$, was obtained for the regression corresponding to the '% DELT + hybrids' metric.

IBI metric agreement and rare taxon removal

As noted in the main text, Cohen's κ calculations were made overall and for each IBI metric to quantify the agreement between post-treatment versus original IBI metric scores. While Fig. 2D–F of the main text provide visualizations of the distributions of per-metric κ-values, colorized by metric type, the full results for each per-metric κ test are presented *below* in Table S2. Only four metrics had κ = 1.00 indicating perfect agreement with their original values after the SRT treatment, including '% dominant species', '% invertivores', '% omnivores', and '% simple lithophils'. Two of these metrics, '% omnivores' and '% simple lithophils', also showed perfect agreement with their original values following the ART removal treatment. 'Total native species' showed the most reduced agreement with original metric scores across our three treatments (Table S2). Central tendencies of κ under each treatment are given in the main text. The overall pattern of κ-value findings supported the SDR treatment as having the greatest effect on the IBI metrics, and SDR and ART having greater effects on IBI metrics than the SRT treatment.

Table S2 Summary of per-metric weighted Cohen's κ-values for IBI metric scores. Results are presented for all 19 metrics used in Alabama fish IBIs, with the "Metric label" column giving abbreviated names matching Fig. 2 (for full names, see Table 1, main text). The three treatments were: 1) removal of single rare taxa, 2) singleton/doubleton removal, and 3) removal of all rare taxa. *P*-values are not presented, as all tests were highly significant ($p < 0.0001$), thus κ-values took precedent for interpretation.

	Single-rare-taxon removal			**Singleton/doubleton removal**			**All-rare-taxa removal**		
Metric label	***N***	**Cohen's κ**	***z***	***N***	**Cohen's κ**	***z***	***N***	**Cohen's κ**	***z***
Total native spp.	3007	0.88	66.20	719	0.25	12.74	719	0.45	19.45
No. shiner spp.	1885	0.96	57.07	444	0.70	20.24	444	0.85	23.99
No. cyprinid spp.	1122	0.94	41.70	274	0.59	13.54	273	0.74	16.24
No. sucker spp.	2252	0.92	55.89	546	0.47	15.77	545	0.63	19.56
No. darter+madtom spp.	3007	0.93	67.81	719	0.50	20.19	719	0.68	25.41
No. Lepomis spp.	1877	0.91	53.06	447	0.44	12.93	447	0.64	18.21
No. centrarchid spp.	618	0.85	27.94	146	0.26	5.56	146	0.41	7.49
No. intolerant spp.	1267	0.96	45.57	298	0.76	17.50	298	0.85	19.37
% tolerant spp.	2997	0.99	67.84	716	0.95	32.11	717	0.98	32.94
% dominant spp.	1122	1.00	45.81	274	0.94	21.23	273	0.97	21.95
% Lepomis	2389	0.99	62.39	573	0.97	30.12	573	0.99	30.63
% GSF+YB	618	0.95	28.68	145	0.65	9.65	146	0.80	11.56
% invertivores	1130	1.00	42.20	270	0.96	19.97	271	0.99	20.57
% omnivores	1695	1.00	50.80	433	0.99	25.27	433	1.00	25.60
% insect. cyprinids	2067	0.99	59.09	457	0.93	26.12	458	0.97	27.06
% top carnivores	3007	0.95	68.39	718	0.68	24.48	719	0.83	28.90
% DELT+hybrids	3007	0.85	54.61	718	0.27	11.67	719	0.52	17.69
No. lithophilic spawners	1885	0.95	55.41	444	0.63	18.56	444	0.77	22.00
% simple lithophils	1122	1.00	42.67	274	0.98	20.76	273	1.00	21.09

Abbreviations: *N*, sample size; No., number; spp., species (singular or plural); *z*, test statistic used to determine statistical significance of weighted Cohen's κ. All other abbreviations follow the caption information for Table 1.

During metric-level analyses, the results of our Wilcoxon signed-rank test on paired SDR vs. ART regression R^2-values in R showed that these treatments produced significantly different goodness-of-fit statistic distributions across linear regressions of the 19 metrics ($W = 2$, $z = -3.74$, $p < 0.001$). Additionally, paired Wilcoxon tests on SDR vs. ART slope (a) results also showed significantly different slope distributions across linear regressions, between treatments ($W = 0$, $z = -3.82$, $p < 0.001$). Comparisons involving the SRT treatment were not run, even for comparative purposes, due to violations of standard assumptions for statistical tests.

Species composition and drainage area effects

Consistent with Wan et al. (2010), evaluating impacts of upstream drainage area on S, N, and nRMSE values revealed that S and N scaled weakly linearly with drainage area (especially in log–log space; JCB, unpublished results), increasing with each sequential S and N class (Fig S9A, B). However, S showed a slightly more pronounced positive correlation than N, the latter of which was more upwardly skewed for small numbers of samples. Median S increased over our drainage area classes, ranging from ~11 species for the smallest headwater streams (< 5 km^2) to 22 species in the largest streams and rivers (> 115 km^2). Median N also increased with increasing sample abundance, from ~85 for the smallest drainage class (< 5 km^2), to ~215 for the second drainage area class (5–15 km^2) representing large headwater or headwater-adjacent streams, to ~330 for the largest drainage class. Consistent with our a priori hypothesis, and similar to the box plots examining stratification by species richness and abundance (Figs 4A, B), box plots of nRMSE values from the first removal treatment showed essentially no relationship between effects of SRT removal on IBI metric scores and our five drainage area classes (Fig. S9C). Impacts of removing singleton/doubleton taxa on IBI metrics were roughly similar across drainage sizes, but lower quantile and median nRMSE values were higher than for the SRT removal and they also decreased slightly for the third drainage area class but increased in the fourth drainage area class (Fig. S9D), and median values ranged from 14.4 to 20.4 for the SDR removal treatment. By contrast, median values were noticeably higher for both the third and fourth drainage area classes following the ART removal, and median and quantile values of

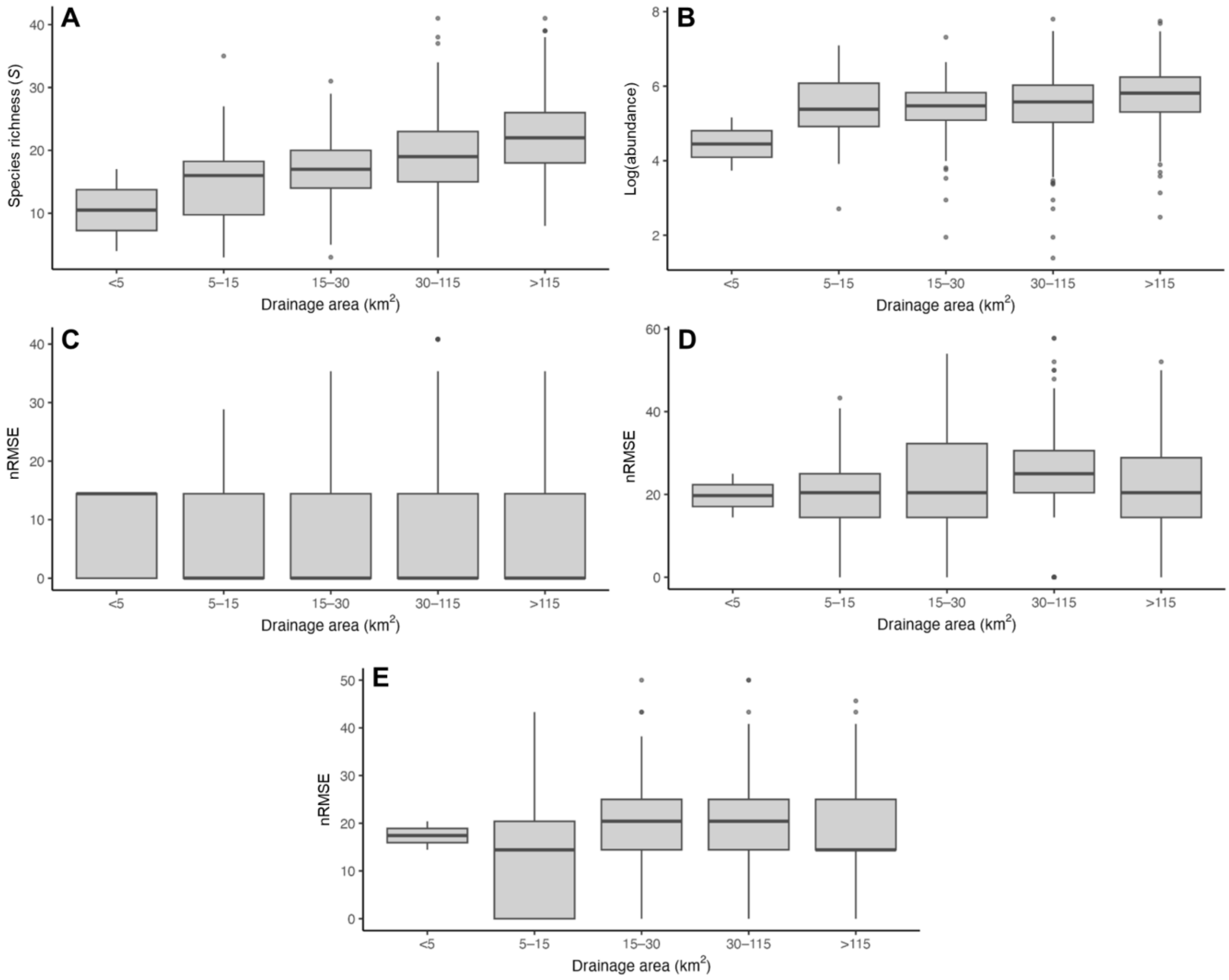

Fig. S9 Box plots evaluating the effects of upstream drainage area on fish community characteristics and IBI metric sensitivity to rare taxon removals in this study. Plotted along the *y*-axis, the panels show (A) species richness (*S*), (B) total abundance (*N*, number of individuals), (C) IBI metric nRMSE values after removal of individual rare taxa within each site (SRT), (D) IBI metric nRMSE values after removal of singleton/doubleton taxa (SDR), and (E) IBI metric nRMSE values after removal of all rare taxa from within each site (ART treatment). Box representations are the same as those described in the caption to Fig. S3 *above*, except each box is filled with *gray shading* and separate boxes are plotted for each drainage are class.

nRMSE from the ART treatment ranged from 20.4 to 22.7 and essentially showed no overall relationship with increasing drainage size (Fig. S9E).

Ichthyoregion-specific effects of rare taxon removal

Ichthyoregion-specific results are discussed in the main text while referencing the supplementary Fig. S10 *below*, which shows effects of our treatments on site IBI metrics and IBI scores, as judged by ΔIBIs and nRMSEs, respectively, plotted as a function of ichthyoregion. The ΔIBIs and nRMSEs had median values at/near zero under the SRT treatment (Fig. S10A,B), indicating limited effects of single rare taxon removals, as well as similar effects of this treatment across space and biogeographic context, as represented by Alabama's ichthyoregions. By contrast, the greatest effects were witnessed in the SP and HCT ichthyoregions, which are located below the Fall Line (O'Neil and Shepard 2007, 2009, 2010), under the SDR and ART treatments. Stronger effects in these regions were primarily registered as effects on metric scores, i.e., higher median nRMSE values, for these treatments, and the significance of differences in nRMSE between ichthyoregions within treatments was evaluated using Kruskal–Wallis omnibus tests discussed in the text. Kruskal–Wallis tests pool all response variable values across groups, sort them from smallest to largest, and substitute them with ranks; ranks belonging to each group are averaged to yield a mean rank; and the null hypothesis, that the difference between the mean ranks of all groups is zero, is tested. Since we obtained significant results in analyses of each treatment, post-hoc Dunn tests with Holm correction were used to see which pairwise relationships between groups (i.e., ichthyoregions) showed significant differences, and thus contributed to the significant test results. Post-hoc Dunn tests (Table S3) indicated that the SP ichthyoregion exhibited significantly higher nRMSE than the P ichthyoregion (ART: $p_{\text{Holm-adjusted}} < 0.001$; SDR: $p_{\text{Holm-adjusted}} < 0.0001$), RVP ichthyoregion (ART: $p_{\text{Holm-adjusted}} = 0.0014$; SDR: $p_{\text{Holm-adjusted}} = <$

0.0001), and TV ichthyoregion (SRT: $p_{\text{Holm-adjusted}} = 0.0073$; SDR: $p_{\text{Holm-adjusted}} = < 0.0001$; ART: $p_{\text{Holm-adjusted}} = 0.0022$) under two or more treatments. The HCT ichthyoregion showed significantly higher nRMSE than the P ichthyoregion (ART: $p_{\text{Holm-adjusted}} = 0.0056$; SDR: $p_{\text{Holm-adjusted}} = 0.0146$), RVP ichthyoregion (ART: $p_{\text{Holm-adjusted}} = 0.0115$; SDR: $p_{\text{Holm-adjusted}} = 0.0260$), and TV ichthyoregion (ART: $p_{\text{Holm-adjusted}} = 0.0125$) under the ART and SDR treatments. By

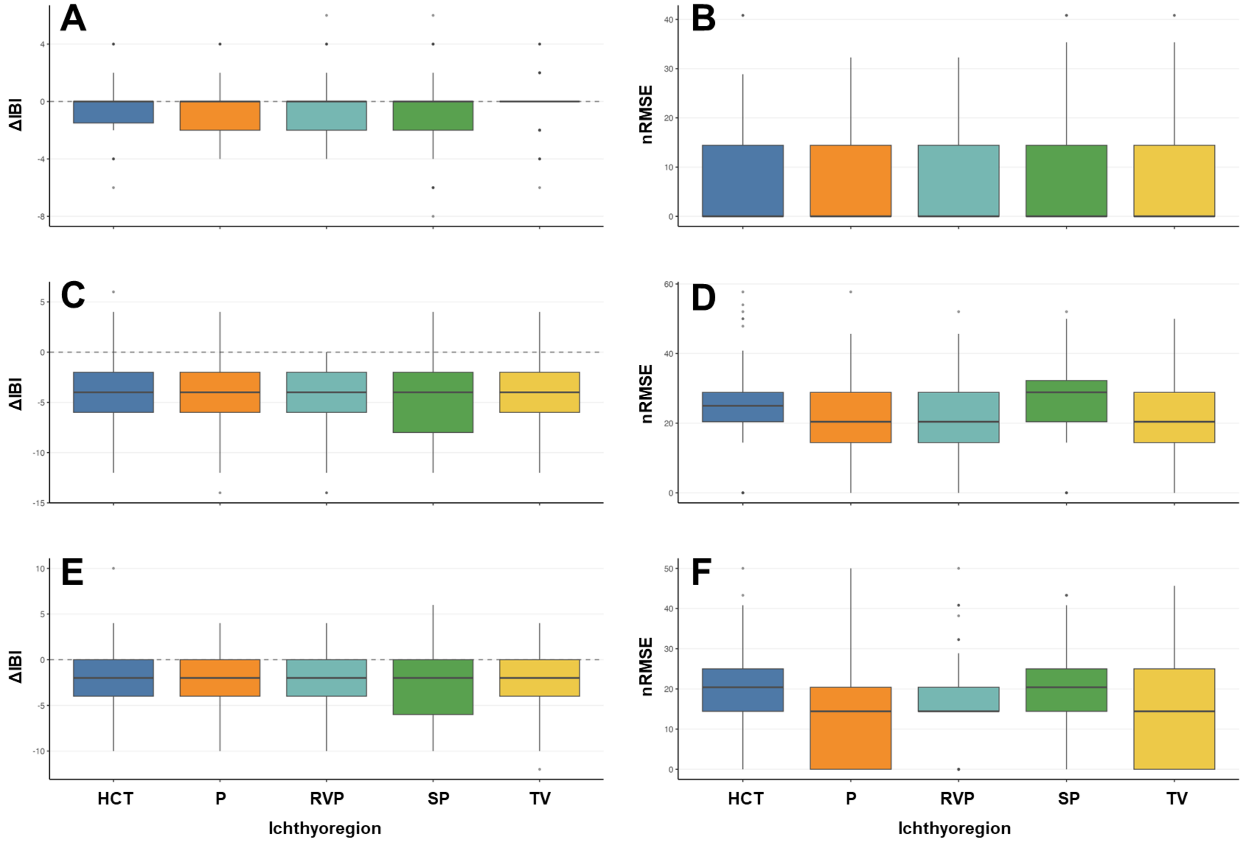

Fig. S10 Summaries of changes in IBI score (ΔIBI) across sites (left column) and IBI metric nRMSE (right column) as a function of ichthyoregion ($n = 5$; Fig. S1). Results are presented as box plots stratified ichthyoregion for our three treatments, including (A, B) SRT removal, (C, D) SDR removal, and (E, F) ART removal. Box representations are the same as those described in the caption to Fig. S3 *above*, except each box is assigned a different *color* based on ichthyoregion (x-axis labels; abbreviations: HCT, Hills and Coastal Terraces; P, Plateau; RVP, Ridge and Valley/Piedmont; SP, Southern Plains; TV, Tennessee Valley).

Table S3 Summary of Dunn post-hoc test results evaluating the magnitude and significance of differences in nRMSEs for all pairs of ichthyoregions, within each treatment. For each treatment, the groups compared are listed along with the resulting *Z* statistic, raw *p*-value of the test, Holm-adjusted *p*-value, and significance level (see inset caption details in text box).

Treatment	**Group 1**	**Group 2**	***Z***	***p* (raw)**	***p* (Holm-adjusted)**	**Significance**
SRT	HCT	P	−0.871	0.384	1.000	
SRT	HCT	RVP	2.037	0.042	0.250	
SRT	P	RVP	2.937	0.003	0.027	*
SRT	HCT	SP	−1.599	0.110	0.439	
SRT	P	SP	−0.706	0.480	0.961	
SRT	RVP	SP	−3.751	$p < 0.001$	0.002	**
SRT	HCT	TV	1.770	0.077	0.383	
SRT	P	TV	2.607	0.009	0.064	
SRT	RVP	TV	−0.122	0.903	0.903	
SRT	SP	TV	3.349	$p < 0.001$	0.007	**
SDR	HCT	P	3.077	0.002	0.015	*
SDR	HCT	RVP	2.853	0.004	0.026	*
SDR	P	RVP	−0.409	0.683	1.000	
SDR	HCT	SP	−1.662	0.097	0.386	
SDR	P	SP	−4.764	$p < 0.0001$	$p < 0.0001$	***
SDR	RVP	SP	−4.644	$p < 0.0001$	$p < 0.0001$	***
SDR	HCT	TV	2.340	0.019	0.096	
SDR	P	TV	−0.678	0.498	1.000	
SDR	RVP	TV	−0.317	0.751	0.751	
SDR	SP	TV	3.987	$p < 0.0001$	$p < 0.0001$	***
ART	HCT	P	3.352	$p < 0.001$	0.006	**
ART	HCT	RVP	3.102	0.002	0.012	*
ART	P	RVP	−0.450	0.652	1.000	
ART	HCT	SP	−0.601	0.548	1.000	
ART	P	SP	−3.989	$p < 0.0001$	$p < 0.001$	***
ART	RVP	SP	−3.778	$p < 0.001$	0.001	**
ART	HCT	TV	3.022	0.003	0.013	*
ART	P	TV	−0.269	0.788	1.000	
ART	RVP	TV	0.158	0.875	0.875	
ART	SP	TV	3.643	$p < 0.001$	0.002	**

In Table S3, ichthyoregion abbreviation definitions follow the main text and Fig. S1. Significance: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ (all Holm-adjusted). Also, positive *Z*-values indicate that Group 1 had a higher nRMSE rank than Group 2.

contrast, no pairwise contrast involving the two least sensitive ichthyoregions, i.e., the TV and P ichthyoregions located above the Fall Line, was statistically significant at the $\alpha = 0.05$ level under any treatment. Taken together, the Kruskal–Wallis and Dunn post-hoc test results indicate that the elevated nRMSE values observed in HCT and SP (Fig. S10) reflect a statistically robust, ichthyoregion-level pattern rather than a sampling artifact, and that this pattern persists across multiple treatments and multiple pairwise comparisons after correction for multiple testing (Table S3).

Supplementary Discussion

Multimetric indices such as the IBI integrate information from biological surveys and drainage characteristics to reveal the structure of local aquatic ecosystems and their responses to human-induced stressors and environmental degradation (Karr 1981; Karr et al. 1986). One key aspect of MMI and IBI applications is component metric selection. Effective metrics must be sensitive, repeatable, capable of distinguishing natural variability in biological condition from human influences, and limited in sensitivity to sampling error (Herricks and Shaefer 1985; Karr et al. 1986; Karr and Chu 1997; Vadas et al. 2022). To date, however, studies of IBI sensitivity to errors and variation in component metrics have been sporadic, leaving the topic understudied. Available studies have also used different methods to analyze widely varying datasets, making comparisons difficult. For example, evaluations of the sensitivity of IBI metrics and outcomes to subsampling (*n* = 9 lakes; Barbour and Gerritsen 1996), repeat sampling (*n* = 12–37 sites; Fore et al. 1994; Hughes et al. 1998), IBI metrics (*n* = 46–84 sites; Angermeier and Karr 1986), rare taxa (*n* = 378 sites; Wan et al. 2010), and DELTs/anomalies and hybrids (*n* = 646 sites; Bagley et al. 2025) have used macroinvertebrate or fish datasets varying over ~70-fold in sample size. We recently showed that fish IBIs for Alabama streams were robust to errors in detecting

DELTs/anomalies and hybridization, but artificial errors altered 13.2% of IBI scores and ~6% of condition ratings, with important management implications (Bagley et al. 2025). While Bagley et al. (2025) focused on the effects of a single metric—the ‘% DELT + hybrids’ metric, they also highlighted the need for broader sensitivity analyses evaluating multiple IBI metrics and outcomes and further exploring the space of IBI responses to sampling errors. In the present study, we took steps towards meeting this need, and our results demonstrate that Alabama fish IBI metrics and outcomes are consistently and significantly disrupted by simulated non-detection of rare taxa, extending the core findings of Wan et al. (2010) to a state with among the most diverse aquatic faunas in the contiguous United States and validating the generality of their findings for Midwestern US fishes across sharply contrasting ecological and biogeographical regions in the southeastern US.

The fact that we found a congruent pattern between our results for Alabama fish IBIs and those that Wan et al. (2010) reported for Minnesota, with species richness-based metrics being more sensitive to rare taxon removal than abundance-based metrics, is particularly noteworthy given that metrics showed only ~44% similarity between our two studies. For example, our dataset containing eight similar IBI metrics (‘Total native species’, ‘Number of darter + madtom species’, ‘Number of intolerant species’, ‘% dominant species’, ‘% DELT + hybrids’, ‘% omnivores’, ‘% simple lithophils’, and ‘% of tolerant species’; Table 1) as compared with those of Wan et al. (2010), plus 11 metrics that differed from the metrics that they used. Such higher-level congruence despite lower-level methodological differences suggests that this trend has generality across space and time and is likely to translate well to other warmwater stream IBI tools and systems in other areas of North America, if not other regions.

Our inferences that singletons/doubletons are commonly encountered during Alabama fish IBI surveys and that singleton/doubleton removal had the greatest effects on IBI metrics and outcomes add important insights to our understanding of fish community ecology and surface water bioassessment in the state. Singleton/doubleton observations are often viewed as indicative of scarce local populations (e.g., Maciel and Arlé 2020), but systematists have shown that singletons are very common in nature and in taxonomic treatments (e.g., Lim et al. 2011). Even so, we do not assume that taxa appearing in our dataset as singleton or doubleton observations are intrinsically rare in nature—only that they were *sampled* in very low numbers during the time and at the location of the corresponding field survey, since 1) detection probabilities vary widely among species and 2) fish species were not exhaustively sampled to determine their true abundances (instead, all observations reflect relative abundances; *cf.* Bayley and Peterson 2001; Kanno et al. 2009). Reasons that singleton/doubleton observations during fish IBI surveys might not reflect fish species true abundances within streams include failures to observe individuals due to depth or turbidity, as well as interspecific variation in susceptibility to electrofishing (e.g., ictalurids vs. other taxa; Kanno et al. 2009).

References*

Bayley PB, Peterson JT (2001) An approach to estimate probability of presence and richness of fish species. Trans Am Fish Soc 130:620–633. https://doi.org/10.1577/1548-8659(2001)130%3C0620:AATEPO%3E2.0.CO;2

Fore LS, Karr JR, Conquest LL (1994) Statistical properties of an index of biological integrity used to evaluate water resources. Can J Fish Aquat Sci 51:1077–1087. https://doi.org/10.1139/f94-107

Herricks EE, Schaeffer DJ (1985) Can we optimize biomonitoring? Environ Manage 9:487–492. https://doi.org/10.1007/BF01867323

Hughes RM, Kaufmann, P. R., Herlihy, A. T., Kincaid, T. M., Reynolds, L., Larsen DP (1998) A process for developing and evaluating indices of fish assemblage integrity. Can J Fish

Aquat Sci 55:1618–1631. https://doi.org/10.1139/cjfas-55-7-1618

Lim GS, Balke M., Meier R (2012) Determining species boundaries in a world full of rarity: singletons, species delimitation methods. Syst Biol 61(1):165–169. https://doi.org/10.1093/sysbio/syr030

Maciel EA, Arlé E (2020) Rare7: An R package to assess the forms of rarity in a community. Ecol Indic 115:106419. https://doi.org/10.1016/j.ecolind.2020.106419

Stout C, Schonhuth S, Mayden R, Garrison NL, Armbruster JW (2022) Phylogenomics and classification of Notropis and related shiners (Cypriniformes: Leuciscidae) and the utility of exon capture on lower taxonomic groups. PeerJ 10: e14072.

Tan M, Armbruster JW (2018) Phylogenetic classification of extant genera of fishes of the order Cypriniformes (Teleostei: Ostariophysi). Zootaxa 4476(1):6–39. https://doi.org/10.11646/zootaxa.4476.1.4

* Only includes references not cited in the main text; see text for additional details.