PNAS
www.pnas.org

Supporting Information for: Emergent dual scaling of riverine biodiversity

Akira Terui*[,1], Seoghyun Kim[1], Christine L. Dolph[2], Taku Kadoya[3], Yusuke Miyazaki[4]

*Corresponding author
Email: hanabi0111@gmail.com

[1] Department of Biology, University of North Carolina at Greensboro, Greensboro, NC 27412, USA
[2] Department of Ecology, Evolution and Behavior, University of Minnesota, St. Paul, MN 55108, USA
[3] National Institute for Environmental Studies, Tsukuba, Ibaraki 305-8506, Japan
[4] Shiraume Gakuen College, Kodaira, Tokyo 187-8570, Japan

This PDF file includes:

- Supplementary text
- Tables S1 - S7
- Figures S1 - S20
- SI References

Contents

Theory **1**

Sensitivity analysis . . . 1

Longitudinal gradient of local species richness . . . 1

Empirical data **2**

Data selection criteria . . . 2

Asymptotic species richness . . . 3

Average predictive comparison . . . 3

Tables **6**

Table S1 Simulation parameter (sensitivity analysis) . . . 6

Table S2 Simulation parameter (main analysis) . . . 7

Table S3 Sensitivity analysis for ecosystem size effects . . . 8

Table S4 Sensitivity analysis for ecosystem complexity effects . . . 9

Table S5 List of fish species in Hokkaido, Japan . . . 10

Table S6 List of fish species in Midwest, US . . . 12

Table S7 Average predictive comparisons . . . 16

Figures **17**

Figure S1 Longitudinal gradient of local species richness ($\sigma_h = 1$, $\sigma_l = 0.01$) . . . 17

Figure S2 Longitudinal gradient of local species richness ($\sigma_h = 1$, $\sigma_l = 1$) . . . 18

Figure S3 Longitudinal gradient of local species richness ($\sigma_h = 0.01$, $\sigma_l = 0.01$) . . . 19

Figure S4 Longitudinal gradient of local species richness ($\sigma_h = 0.01$, $\sigma_l = 1$) . . . 20

Figure S5 Influence of ecosystem size ($p_d = 0.1$, $\sigma_h = 1$, $\sigma_l = 0.01$) . . . 21

Figure S6 Influence of ecosystem size ($p_d = 0.1$, $\sigma_h = 1$, $\sigma_l = 1$) . . . 22

Figure S7 Influence of ecosystem size ($p_d = 0.1$, $\sigma_h = 0.01$, $\sigma_l = 0.01$) . . . 23

Figure S8 Influence of ecosystem size ($p_d = 0.1$, $\sigma_h = 0.01$, $\sigma_l = 1$) . . . 24

Figure S9 Influence of ecosystem size ($p_d = 0.01$, $\sigma_h = 1$, $\sigma_l = 1$) . . . 25

Figure S10 Influence of ecosystem size ($p_d = 0.01$, $\sigma_h = 0.01$, $\sigma_l = 0.01$) . . . 26

Figure S11 Influence of ecosystem size ($p_d = 0.01$, $\sigma_h = 0.01$, $\sigma_l = 1$) . . . 27

Figure S12 Influence of ecosystem complexity ($p_d = 0.1$, $\sigma_h = 1$, $\sigma_l = 0.01$) . . . 28

Figure S13 Influence of ecosystem complexity ($p_d = 0.1$, $\sigma_h = 1$, $\sigma_l = 1$) . . . 29

Figure S14 Influence of ecosystem complexity ($p_d = 0.1$, $\sigma_h = 0.01$, $\sigma_l = 0.01$) . . . 30

Figure S15 Influence of ecosystem complexity ($p_d = 0.1$, $\sigma_h = 0.01$, $\sigma_l = 1$) . . . 31

Figure S16 Influence of ecosystem complexity ($p_d = 0.01$, $\sigma_h = 1$, $\sigma_l = 1$) . . . 32

Figure S17 Influence of ecosystem complexity ($p_d = 0.01$, $\sigma_h = 0.01$, $\sigma_l = 0.01$) . . . 33

Figure S18 Influence of ecosystem complexity ($p_d = 0.01$, $\sigma_h = 0.01$, $\sigma_l = 1$) 34
Figure S19 Correlation structure of explanatory variables 35
Figure S20 Sensitivity analysis of asymptotic species richness 36

References **37**

Theory

Sensitivity analysis

We performed a sensitivity analysis of the metacommunity simulation to identify key simulation parameters that strongly affect the relationships between diversity metrics (α, β, and γ) and ecosystem properties (the number of habitat patches N_p and branching probability P_b). We generated 500 sets of parameter combinations by randomly drawing values of 8 simulation parameters from uniform distributions (**Table S1**). For each parameter combination, we generated 100 branching networks with the gradients of ecosystem size ($N_p \sim Unif(10, 150)$) and complexity ($P_b \sim Unif(0.01, 0.99)$). This results in a total of 50000 simulation replicates. In each simulation replicate, we allowed interspecific variation in niche optimum μ_i and width $\sigma_{niche,i}$ ($\mu_i \sim Unif(-1, 1)$ and $\sigma_{niche,i} \sim Unif(0.1, 1)$, respectively; subscript i represents species) and ran 1400 time steps of metacommunity dynamics. We obtained temporal means of α, β, and γ diversity for the last 1000 time steps. The first 400 time steps were discarded as initialization and burn-in periods. We removed simulation replicates in which no species established populations over the initial 400 time steps.

For each parameter combination, we regressed log-transformed α, β, and γ diversity ($log_{10}\ y_j$ for network replicate j) on the number of habitat patches N_p and branching probability P_b as:

$$\begin{aligned} log_{10}\ y_j &\sim Normal(\mu_j, \sigma^2) \\ \mu_j &= \psi_0 + \psi_1\ log_{10}\ N_{p,j} + \psi_2\ log_{10}\ P_{b,j} \end{aligned} \tag{1}$$

where ψ_k ($k = 0 - 2$) are the intercept (ψ_0) and regression coefficients (ψ_1 and ψ_2). We extracted 500 estimates of ψ_1 and ψ_2, which represent the effects of N_p and P_b on diversity metrics under a given parameter combination. To examine influences of simulation parameters (**Table S1**) on ψ_1 and ψ_2, we developed the following regression model taking ψ_1 or ψ_2 as a response variable u_n (parameter combination n):

$$\begin{aligned} u_n &\sim Normal(\mu_n, \sigma^2) \\ \mu_n &= \zeta_0 + \zeta_1\sigma_{h,n} + \zeta_2\sigma_{l,n} + \zeta_3\sigma_{z,n} + \zeta_4\phi_n + \zeta_5\nu_n + \zeta_6 b_{max,n} + \zeta_7\theta_n + \zeta_8 p_{d,n} \end{aligned} \tag{2}$$

where ζ_k ($k = 0 - 8$) are the intercept (ζ_0) and regression coefficients (ζ_{1-8}). Explanatory variables were standardized to a mean of zero and a standard deviation of one, so that regression coefficients are comparable.

The sensitivity analysis revealed key simulation parameters. For the effects of N_p, the following simulation parameters were influential: the degree of local environmental noise (σ_l; influenced the effects on α and γ diversity), the maximum value of interspecific competition coefficient (b_{max}; influenced the effects on α, β, and γ diversity), dispersal distance (θ; influenced the effects on α and β diversity), and dispersal probability (p_d; influenced the effect on α diversity) (**Table S3**). For the effects of P_b, the following simulation parameters were influential: environmental variation at headwaters (σ_h; influenced the effect on γ diversity), the degree of local environmental noise (σ_l; influenced the effects on α and β diversity), and dispersal distance (θ; influenced the effects on α and β diversity) (**Table S4**).

Based on the results, we identified σ_h, σ_l, b_{max}, θ, and p_d as key parameters. We changed the values of these parameters in the main analysis and examined the relationships between diversity metrics and ecosystem properties.

Longitudinal gradient of local species richness

Longitudinal gradients of local species richness have been extensively studied in rivers, illuminating typical patterns observed in nature. The most common pattern is a downstream increase of local species richness (1–3). However, recent empirical and theoretical studies also showed 'reversed patterns,' in which local species richness decreases downstream (4, 5). We predicted the longitudinal gradient of local species richness to confirm that our simulation scenarios are capable of reproducing the previously observed patterns of local species

richness. We considered 32 simulation scenarios comprising four landscape and eight ecological scenarios, as described in the main text (a set of parameters is described in **Table S2**). Under each simulation scenario, we generated 10 branching networks with fixed parameters of ecosystem size ($N_p = 100$) and complexity ($P_b = 0.5$). This results in a total of 320 simulation replicates. In each simulation replicate, we allowed interspecific variation in niche optimum μ_i and width $\sigma_{niche,i}$ ($\mu_i \sim Unif(-1, 1)$ and $\sigma_{niche,i} \sim Unif(0.1, 1)$, respectively; subscript i represents species) and ran 1400 time steps of metacommunity dynamics. We obtained temporal means of local species richness at each habitat patch for the last 1000 time steps. The first 400 time steps were discarded as initialization and burn-in periods. We removed simulation replicates in which no species established populations over the initial 400 time steps. We evaluated the relationship between local species richness and the number of upstream habitat patches, a proxy for the longitudinal position of a habitat patch.

The simulation reproduced diverse patterns of longitudinal gradients in local species richness (**Figures S1-4**). The downstream increase of local species richness was predicted under a natural landscape scenario, in which environmental variation at headwaters σ_h exceeds the degree of local environmental noise σ_l (**Figure S1**). This pattern was consistent across ecological scenarios except those with long dispersal distance and high dispersal probability (**Figure S1**). Similarly, we observed a downstream increase of local species richness in scenarios with low habitat diversity ($\sigma_h = \sigma_l = 0.01$) and low dispersal probability ($p_d = 0.01$) (**Figure S3**). However, there were cases where local species richness decreased downstream or showed no longitudinal patterns. For example, when local environmental noise exceeds environmental variation at headwaters ($\sigma_l \geq \sigma_h$), local species richness showed a downstream decrease or a vague longitudinal pattern (**Figure S4**). Therefore, the simulation scenarios were capable of reproducing previously observed patterns, suggesting the appropriateness in the choice of parameter combinations.

Empirical data

Data selection criteria

Hokkaido, Japan. In Hokkaido, most data were collected from summer to fall. We screened data through the following procedure:

1. We listed recorded fish species and re-organized species names to make them consistent across data sources. We removed the following species at this stage: (1) identified at the family-level; (2) marine fish species (including species that occasionally use brackish/freshwater habitats).
2. We selected sampling sites based on the following criteria: (1) surveys were conducted with netting and/or electrofishing, (2) surveys were designed to collect a whole fish community, (3) sites contained reliable coordinates (sites with coordinates identical at 3 decimal degrees were treated as the same site), and (4) sites did not involve unidentified species (genus level) that are rarely observed in the data set (< 100 sites occurrence).
3. For sites with multiple visits (i.e., temporal replicates), we used the latest-year observation at each sampling site to minimize variation in sampling efforts among sites. Surveys that occurred in the same year were aggregated into a single observation.
4. We confined sites to those with the latest observation year of ≥ 1990. Although the data set contained observations from 1953, we added this restriction to align the observation period with the data set in the Midwest, US.
5. Four genera (*Lethenteron*, *Pungitius*, *Rhinogobius*, and *Tribolodon*) were treated as species groups (i.e., spp.) as their taxonomic resolutions varied greatly among data sources due to difficulties in identifying species.

Midwest, US. In the Midwest, the data set covered most of Upper Mississippi (Hydrologic Unit Code 2 [HUC 2] , region 07, as defined by U.S. Geological Survey and U.S. Department of Agriculture Natural Resources Conservation Service (6)) and the part of Great Lakes (HUC 2, region 04), Missouri (HUC 2,

region 10), and Ohio (HUC 2, region 05). Fish data were collected from summer to fall with electrofishing (backpack, barge-type, or boat-mounted) and supplemental netting at some locations. We screened data through the following procedure:

1. We used data of the Upper Mississippi (HUC 2, region 07) and Great Lakes basins (HUC 2, region 04) as most sites are included in these regions.
2. We removed records of unidentified species, hybrid species, and commercial species apparently absent in the wild (e.g., goldfish).
3. We used the latest observation at each sampling site to minimize variation in sampling efforts among sites.

Asymptotic species richness

We evaluated sensitivity of asymptotic species richness (Chao 2 estimator) to sample size (i.e., the number of sampling sites in a watershed). We simulated presence-absence data of species with known values of true species richness S_{true} and the number of sampling sites N_{site}. In this simulation, the presence of species i at site x, J_{ix}, was drawn from a Bernoulli distribution as $J_{ix} \sim Bernoulli(p_i \kappa_{ix})$ where p_i is the detection probability for species i and κ_{ix} is the presence probability of species i at site x. Based on the incidence frequency of simulated species $F_i = \sum_{x=1}^{N_{site}} J_{ix}$, we estimated asymptotic species richness using the iNEXT function in the R package 'iNEXT' (7). We calculated % bias of estimated asymptotic species richness S_{est}:

$$\% \; bias = \frac{100(S_{est} - S_{true})}{S_{true}} \tag{3}$$

Positive values of % bias indicate an overestimation of species richness, while negative values indicate an underestimation.

We used the following values for parameters: $S_{true} = 10, 40, 70, 100$ and $N_{site} = 5, 10, 15, 20, 100$. We produced 100 replicates of simulated data sets for each parameter combination, resulting in a total of 2000 simulation replicates. In each simulation replicate, the probabilities of detection and true presence were drawn randomly from uniform distributions as $p_i \sim Unif(0.3, 0.8)$ and $\kappa_{ix} \sim Unif(0, 1)$.

The number of sampling sites influenced the estimation accuracy of asymptotic species richness. The % bias decreased sharply as the number of sampling sites increased (**Figure S20**). In particular, the estimation bias with a small sample size ($N_{site} = 5$) was substantial when the true species richness S_{true} was low; some estimates showed $> 150\%$ bias (**Figure S20**). Given the simulation results, estimates of asymptotic species richness at watersheds with < 10 sampling sites may involve substantial statistical uncertainty.

Average predictive comparison

For regression models, standardized regression coefficients (or its variant) are perhaps the most common summary when comparing effect sizes of explanatory variables. These values are useful if they are directly interpretable. For example, in linear regression models without interactions and variable transformations, regression coefficients have direct interpretations as they represent additive effects. However, there are many cases where regression coefficients are difficult to interpret. In our case, we regressed the log-transformed species richness on explanatory variables, so regression coefficients do not have direct interpretations on the original scale of the response variable y (i.e., species richness).

The average predictive comparison provides an intuitive yet rigorous way to quantify effect sizes for each of explanatory variables in regression models with interactions and/or variable transformations (8). The basic predictive comparison δ_u is an expected change in y on the original scale per a unit difference of the explanatory variable of interest:

$$\delta_u(u^{(1)} \to u^{(2)}, v, \Theta) = \frac{E(y|u^{(2)}, v, \Theta) - E(y|u^{(1)}, v, \Theta)}{u^{(2)} - u^{(1)}} \tag{4}$$

where $E(\cdot)$ is a known function (e.g., exponential), u the input of interest (a value for the explanatory variable of interest), v all the other inputs (a vector in general), Θ a set of parameters in a regression model. In general, δ_u rests on $u^{(1)}$ and $u^{(2)}$ (the lower and higher points of the hypothesized change in the explanatory variable of interest), v, and Θ. Therefore, Gelman and Pardoe (8) defined the *average predictive comparison* Δ_u as "the mean value of δ_u over some specified distribution of the inputs and parameters":

$$\Delta_u = \frac{\int\int_{u^{(2)}>u^{(1)}} du^{(1)}du^{(2)} \int dv \int d\Theta (E(y|u^{(2)}, v, \Theta) - E(y|u^{(1)}, v, \Theta))\ p(u^{(1)}|v)\ p(u^{(2)}|v)\ p(v)\ p(\Theta)}{\int\int_{u^{(2)}>u^{(1)}} du^{(1)}du^{(2)} \int dv \int d\Theta\ (u^{(2)} - u^{(1)})\ p(u^{(1)}|v)\ p(u^{(2)}|v)\ p(v)\ p(\Theta)} \tag{5}$$

However, directly using the above equation is impractical because estimating $p(u^{(1)}|v)$ and $p(u^{(2)}|v)$ from finite data points is challenging (especially when v is continuous). Instead, we estimated Δ_u using the following equation proposed by Gelman and Pardoe (8):

$$\hat{\Delta}_u = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{s=1}^{S} w_{ij}(E(y|u_j, v_i, \Theta_s) - E(y|u_i, v_i, \Theta_s))sign(u_j - u_i)}{\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{s=1}^{S} w_{ij}(u_j - u_i)sign(u_j - u_i)} \tag{6}$$

The summations over n data (i and j are a given data point) and S parameter replicates (s is a given simulation replicate) are a realization of averaging over the distributions of $(u^{(1)}, v)$, $u^{(2)}$, and Θ. The factor w_{ij} is a weight that serves to approximate $p(u^{(1)}|v)$ and $p(u^{(2)}|v)$. In theory, v must be held constant while the input of interest changes from $u^{(1)}$ to $u^{(2)}$. However, there are, if any, few transitions from $u^{(1)}$ to $u^{(2)}$ with a common v. We approximate such exact transitions by assigning each pair of data points with a weight:

$$w_{ij} = w(v_i, v_j) \tag{7}$$

The weight factor should represent the likelihood of u changing from u_i to u_j with a common $v = v_i$ (v_i is in general a vector of explanatory variables for data point i). We used the following weighting function based on Mahalanobis distances following Gelman and Pardoe (8):

$$w(v_i, v_j) = \frac{1}{1 + (v_i - v_j)^T \Omega_v^{-1} (v_i - v_j)} \tag{8}$$

where Ω_v^{-1} is the inverse of the variance-covariance matrix of v. Note that $v = v_i$ when $u = u_i$, and $v = v_j$ when $u = u_j$. The function gives the maximum weight for a pair of u_i and u_j if $v_i = v_j$ while giving a less weight as the Mahalanobis distance between v_i and v_j increases. This property makes sense because our goal is to approximate the probability of transition from u_i to u_j with a common v (i.e., $v_i = v_j$).

We estimated $\hat{\Delta}_u$ for watershed area and branching probability using the estimated coefficients $\hat{\xi}_k$ of the regression model explaining γ diversity (see **Table 1**). A thousand of vectors of simulated parameters Θ_s were drawn from normal distributions with means of $\hat{\xi}_k$ and standard deviations of $\sigma_{\xi,k}$ (the estimated standard errors of parameters ξ_k). The estimated average predictive comparisons are $\hat{\Delta}_u = \frac{1}{S}\sum_{s=1}^{S} \hat{\Delta}_{u,s}$ where

$$\hat{\Delta}_{u,s} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(E(y|u_j, v_i, \Theta_s) - E(y|u_i, v_i, \Theta_s))sign(u_j - u_i)}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(10^{u_j} - 10^{u_i})sign(u_j - u_i)} \tag{9}$$

We exponentiated u_i and u_j in the denominator because we log-transformed watershed area and branching probability in the statistical model. Then

$$S.E.(\hat{\Delta}_{u,s}) = \sqrt{\frac{1}{S-1}\sum_{s=1}^{S}(\hat{\Delta}_{u,s} - \hat{\Delta}_{u})^2} \tag{10}$$

The estimated average predictive comparisons were summarized in **Table S7**.

Tables

Table S1 Simulation parameter (sensitivity analysis)

Table S1 Parameter values used in the sensitivity analysis of the metacommunity simulation. See the main text for model details.

Parameter	Value	Interpretation
σ_h	Unif(0.01, 1)	Environmental variation at headwaters
σ_l	Unif(0.01, 1)	Degree of local environmental noise
σ_z	Unif(0.01, 0.5)	Temporal environmental variability
ρ	1	Strength of spatial autocorrelation in mean environmental condition
ϕ	Unif(0.01, 1)	Extent of spatial autocorrelation in temporal environmental variation
ν	Unif(1, 5)	Cost of a wider niche
b_{max}	Unif(0.5, 1.5)	Maximum value of interspecific competition coefficient
θ	Unif(0.1, 1)	Rate parameter of an exponential dispersal kernel
p_d	Unif(0.01, 0.1)	Dispersal probability
$r_{0,i}$	4	Maximum reproductive rate

Table S2 Simulation parameter (main analysis)

Table S2 Values and interpretation of simulation parameters used in the main simulation. See the main text for model details.

Parameter	Value	Interpretation
σ_h	0.01, 1.00	Environmental variation at headwaters
σ_l	0.01, 1.00	Degree of local environmental noise
σ_z	0.1	Temporal environmental variability
ρ	1	Strength of spatial autocorrelation in mean environmental condition
ϕ	0.05	Extent of spatial autocorrelation in temporal environmental variation
ν	1	Cost of a wider niche
b_{max}	0.75, 1.50	Maximum value of interspecific competition coefficient
θ	0.1, 1.0	Rate parameter of an exponential dispersal kernel
p_d	0.01, 0.10	Dispersal probability
$r_{0,i}$	4	Maxiumum reproductive rate

Table S3 Sensitivity analysis for ecosystem size effects

Table S3 Sensitivity analysis of the metacommunity simulation. Parameter estimates of linear regression models (standard errors in parenthesis) are shown. Response variables are the effects of the number of habitat patches (N_p) on α, β, and γ diversity. Explanatory variables (i.e., simulation parameters) were standardized to a mean of zero and a standard deviation of one prior to the analysis. See **Tables S1 and S2** for interpretation of the simulation parameters.

	Response variable		
	Effect of N_p on α diversity	Effect of N_p on β diversity	Effect of N_p on γ diversity
σ_h	0.008	−0.003	0.005
	(0.002)	(0.001)	(0.002)
σ_l	−0.021	0.004	−0.018
	(0.002)	(0.001)	(0.002)
σ_z	0.0001	−0.013	−0.013
	(0.002)	(0.001)	(0.002)
ϕ	0.001	−0.0002	0.0003
	(0.002)	(0.001)	(0.002)
ν	−0.001	−0.009	−0.009
	(0.002)	(0.001)	(0.002)
b_{max}	0.019	0.028	0.047
	(0.002)	(0.001)	(0.002)
θ	−0.040	0.041	0.001
	(0.002)	(0.001)	(0.002)
p_d	0.017	−0.006	0.010
	(0.002)	(0.001)	(0.002)
Intercept	0.147	0.137	0.284
	(0.002)	(0.001)	(0.002)

Table S4 Sensitivity analysis for ecosystem complexity effects

Table S4 Sensitivity analysis of the metacommunity simulation. Parameter estimates of linear regression models (standard errors in parenthesis) are shown. Response variables are the effects of branching probability (P_b) on α, β, and γ diversity. Explanatory variables (i.e., simulation parameters) were standardized to a mean of zero and a standard deviation of one prior to the analysis. See **Tables S1 and S2** for interpretation of the simulation parameters.

	Response variable		
	Effect of P_b on α diversity	Effect of P_b on β diversity	Effect of P_b on γ diversity
σ_h	0.012	0.007	0.019
	(0.003)	(0.002)	(0.002)
σ_l	0.060	−0.047	0.013
	(0.003)	(0.002)	(0.002)
σ_z	−0.004	−0.002	−0.006
	(0.003)	(0.002)	(0.002)
ϕ	0.002	0.001	0.002
	(0.003)	(0.002)	(0.002)
ν	−0.001	0.001	−0.001
	(0.003)	(0.002)	(0.002)
b_{max}	−0.006	−0.001	−0.006
	(0.003)	(0.002)	(0.002)
θ	0.027	−0.028	−0.001
	(0.003)	(0.002)	(0.002)
p_d	0.007	−0.007	−0.0002
	(0.003)	(0.002)	(0.002)
Intercept	0.145	−0.132	0.013
	(0.003)	(0.002)	(0.002)

Table S5 List of fish species in Hokkaido, Japan

Table S5 List of fish species in Hokkaido, Japan, included in our statistical analysis. 52 species are ordered alphabetically, along with the number of sites present and % occupancy of 2650 sites.

Species	Number of sites present	Occupancy (%)
Acanthogobius lactipes	66	2.49
Anguilla japonica	1	0.04
Carassius buergeri subsp. 2	4	0.15
Carassius cuvieri	24	0.91
Carassius sp.	215	8.11
Channa argus	3	0.11
Cottus amblystomopsis	49	1.85
Cottus hangiongensis	99	3.74
Cottus nozawae	848	32.00
Cottus sp. ME	25	0.94
Cyprinus carpio	50	1.89
Gasterosteus aculeatus	163	6.15
Gnathopogon caerulescens	1	0.04
Gnathopogon elongatus elongatus	2	0.08
Gymnogobius breunigii	34	1.28
Gymnogobius castaneus complex	149	5.62
Gymnogobius opperiens	93	3.51
Gymnogobius petschiliensis	2	0.08
Gymnogobius urotaenia	313	11.81
Hucho perryi	61	2.30
Hypomesus nipponensis	173	6.53
Hypomesus olidus	8	0.30
Lefua nikkonis	21	0.79
Lethenteron spp.	752	28.38
Leucopsarion petersii	2	0.08
Luciogobius guttatus	3	0.11
Misgurnus anguillicaudatus	221	8.34
Noemacheilus barbatulus	1623	61.25
Oncorhynchus gorbuscha	29	1.09
Oncorhynchus keta	153	5.77
Oncorhynchus masou masou	1444	54.49
Oncorhynchus mykiss	466	17.58
Oncorhynchus nerka	6	0.23
Opsariichthys platypus	1	0.04
Osmerus dentex	6	0.23
Phoxinus percnurus sachalinensis	69	2.60
Plecoglossus altivelis altivelis	112	4.23
Pseudorasbora parva	93	3.51
Pungitius spp.	294	11.09
Rhinogobius spp.	181	6.83
Rhodeus ocellatus ocellatus	22	0.83
Salangichthys microdon	10	0.38
Salmo trutta	15	0.57
Salvelinus fontinalis	2	0.08
Salvelinus leucomaenis leucomaenis	640	24.15
Salvelinus malma	274	10.34
Salvelinus malma miyabei	2	0.08

Species	Number of sites present	Occupancy (%)
Silurus asotus	7	0.26
Spirinchus lanceolatus	6	0.23
Tribolodon spp.	1197	45.17
Tridentiger brevispinis	136	5.13
Tridentiger obscurus	7	0.26

Table S6 List of fish species in Midwest, US

Table S6 List of fish species in the Midwest, US, included in our statistical analysis. 159 species are ordered alphabetically, along with the number of sites present and % occupancy of 3999 sites.

Species	Number of sites present	Occupancy (%)
Acipenser fulvescens	7	0.18
Alosa pseudoharengus	1	0.03
Ambloplites rupestris	706	17.65
Ameiurus melas	868	21.71
Ameiurus natalis	663	16.58
Ameiurus nebulosus	30	0.75
Amia calva	95	2.38
Ammocrypta clara	12	0.30
Aphredoderus sayanus	76	1.90
Aplodinotus grunniens	208	5.20
Campostoma anomalum	1345	33.63
Campostoma oligolepis	124	3.10
Carpiodes carpio	128	3.20
Carpiodes cyprinus	234	5.85
Carpiodes velifer	82	2.05
Catostomus commersonii	2928	73.22
Centrarchus macropterus	5	0.13
Chrosomus eos	336	8.40
Chrosomus neogaeus	102	2.55
Clinostomus elongatus	96	2.40
Cottus bairdii	466	11.65
Cottus carolinae	6	0.15
Cottus cognatus	38	0.95
Crystallaria asprella	1	0.03
Ctenopharyngodon idella	19	0.48
Culaea inconstans	1532	38.31
Cyprinella lutrensis	269	6.73
Cyprinella spiloptera	778	19.45
Cyprinella venusta	2	0.05
Cyprinella whipplei	33	0.83
Cyprinus carpio	945	23.63
Dorosoma cepedianum	208	5.20
Erimystax x-punctatus	13	0.33
Erimyzon oblongus	63	1.58
Erimyzon sucetta	10	0.25
Esox americanus vermiculatus	117	2.93
Esox lucius	957	23.93
Esox masquinongy	20	0.50
Etheostoma asprigene	10	0.25
Etheostoma blennioides	1	0.03
Etheostoma caeruleum	194	4.85
Etheostoma chlorosomum	4	0.10
Etheostoma crossopterum	4	0.10
Etheostoma exile	261	6.53
Etheostoma flabellare	843	21.08
Etheostoma gracile	15	0.38
Etheostoma kennicotti	1	0.03

Species	Number of sites present	Occupancy (%)
Etheostoma microperca	20	0.50
Etheostoma nigrum	2544	63.62
Etheostoma proeliare	2	0.05
Etheostoma spectabile	119	2.98
Etheostoma squamiceps	4	0.10
Etheostoma zonale	320	8.00
Fundulus diaphanus	3	0.08
Fundulus dispar	4	0.10
Fundulus notatus	281	7.03
Fundulus olivaceus	42	1.05
Hiodon alosoides	3	0.08
Hiodon tergisus	16	0.40
Hybognathus hankinsoni	575	14.38
Hybognathus nuchalis	24	0.60
Hybopsis amnis	1	0.03
Hypentelium nigricans	680	17.00
Hypophthalmichthys molitrix	14	0.35
Hypophthalmichthys nobilis	3	0.08
Ichthyomyzon castaneus	51	1.28
Ichthyomyzon fossor	35	0.88
Ichthyomyzon gagei	6	0.15
Ichthyomyzon unicuspis	9	0.23
Ictalurus punctatus	413	10.33
Ictiobus bubalus	90	2.25
Ictiobus cyprinellus	129	3.23
Ictiobus niger	41	1.03
Labidesthes sicculus	64	1.60
Lepisosteus oculatus	13	0.33
Lepisosteus osseus	25	0.63
Lepisosteus platostomus	58	1.45
Lepomis cyanellus	1573	39.33
Lepomis gibbosus	290	7.25
Lepomis gulosis	43	1.08
Lepomis humilis	356	8.90
Lepomis macrochirus	1050	26.26
Lepomis megalotis	186	4.65
Lepomis microlophus	19	0.48
Lethenteron appendix	123	3.08
Lota lota	266	6.65
Luxilus chrysocephalus	196	4.90
Luxilus cornutus	1781	44.54
Lythrurus fumeus	6	0.15
Lythrurus umbratilis	222	5.55
Macrhybopsis aestivalis	1	0.03
Macrhybopsis hyostoma	3	0.08
Macrhybopsis storeriana	10	0.25
Micropterus dolomieu	746	18.65
Micropterus punctulatus	15	0.38
Micropterus salmoides	987	24.68
Minytrema melanops	41	1.03
Morone americana	2	0.05
Morone chrysops	65	1.63

Species	Number of sites present	Occupancy (%)
Morone mississippiensis	16	0.40
Moxostoma anisurum	288	7.20
Moxostoma carinatum	8	0.20
Moxostoma duquesni	103	2.58
Moxostoma erythrurum	708	17.70
Moxostoma macrolepidotum	737	18.43
Moxostoma valenciennesi	85	2.13
Neogobius melanostomus	13	0.33
Nocomis biguttatus	1285	32.13
Notemigonus crysoleucas	377	9.43
Notropis anogenus	10	0.25
Notropis atherinoides	207	5.18
Notropis blennius	16	0.40
Notropis boops	9	0.23
Notropis buccatus	47	1.18
Notropis chalybaeus	11	0.28
Notropis dorsalis	1087	27.18
Notropis heterodon	28	0.70
Notropis heterolepis	187	4.68
Notropis hudsonius	81	2.03
Notropis nubilus	58	1.45
Notropis percobromus	167	4.18
Notropis rubellus	62	1.55
Notropis stramineus	973	24.33
Notropis texanus	20	0.50
Notropis volucellus	81	2.03
Notropis wickliffi	10	0.25
Noturus exilis	33	0.83
Noturus flavus	478	11.95
Noturus gyrinus	447	11.18
Noturus nocturnus	38	0.95
Oncorhynchus mykiss	55	1.38
Oncorhynchus tshawytscha	1	0.03
Opsopoeodus emiliae	3	0.08
Perca flavescens	622	15.55
Percina caprodes	337	8.43
Percina carprodes semifasciata	7	0.18
Percina evides	16	0.40
Percina maculata	888	22.21
Percina phoxocephala	259	6.48
Percina sciera	2	0.05
Percopsis omiscomaycus	21	0.53
Phenacobius mirabilis	258	6.45
Phoxinus erythrogaster	417	10.43
Pimephales notatus	1782	44.56
Pimephales promelas	1534	38.36
Pimephales vigilax	84	2.10
Pomoxis annularis	61	1.53
Pomoxis nigromaculatus	376	9.40
Pylodictis olivaris	73	1.83
Rhinichthys atratulus	1118	27.96
Rhinichthys cataractae	627	15.68

Species	Number of sites present	Occupancy (%)
Rhinichthys obtusus	448	11.20
Salmo trutta	400	10.00
Salvelinus fontinalis	372	9.30
Sander canadensis	39	0.98
Sander vitreus	367	9.18
Scaphirhynchus platorynchus	8	0.20
Semotilus atromaculatus	2774	69.37
Umbra limi	1600	40.01

Table S7 Average predictive comparisons

Estimated average predictive comparisons. Regression coefficients were derived from the regression model explaining γ diversity (see **Table 1** in the maintext for the estimated regression parameters). The average predictive comparisons (an expected increase in γ diversity per a unit difference of the explanatory variable of interest) were estimated based on units of 1000 km^2 for watershed area and 0.1 for branching probability.

Input	Estimate	SE	Unit
Watershed area	5.59	2.35	1000 km^2
Branching probability	8.37	2.74	0.1

Figures

Figure S1 Longitudinal gradient of local species richness ($\sigma_h = 1$, $\sigma_l = 0.01$)

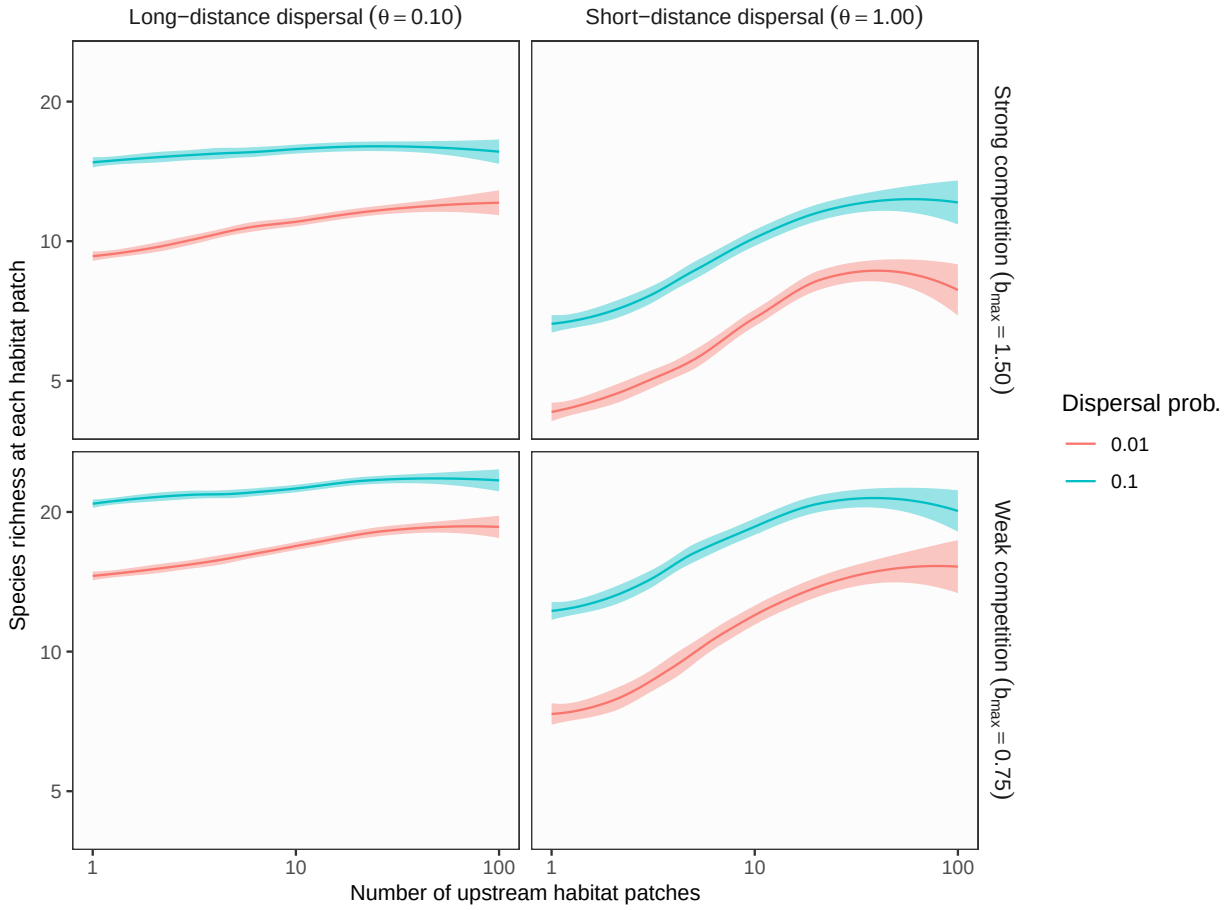

Figure S1 Theoretical predictions for longitudinal gradients of local species richness in branching networks. The longitudinal position of each habitat patch (x-axis) was expressed as the number of upstream habitat patches. In this simulation, environmental variation at headwaters (σ_h) exceeds local environmental noise (σ_l). Lines and shades are loess curves fitted to simulated data and their 95% confidence intervals. Each panel represents different ecological scenarios under which metacommunity dynamics were simulated. Rows represent different competition strength. Competition coefficients (b_{ij}) were varied randomly from 0 to 1.5 (top, strong competition) or 0.75 (bottom, weak competition). Columns and lines represent different dispersal scenarios (dispersal distance and probability). Left and right columns show long-distance (the rate parameter of an exponential dispersal kernel $\theta = 0.10$) and short-distance dispersal ($\theta = 1.0$) scenarios repectively. Red and blue lines show low ($p_d = 0.01$) and high dispersal probabilities ($p_d = 0.10$). Other parameters are as follows: environmental variation at headwaters $\sigma_h = 1$; local environmental noise $\sigma_l = 0.01$; ecosystem size $N_p = 100$; ecosystem complexity $P_b = 0.5$.

Figure S2 Longitudinal gradient of local species richness ($\sigma_h = 1$, $\sigma_l = 1$)

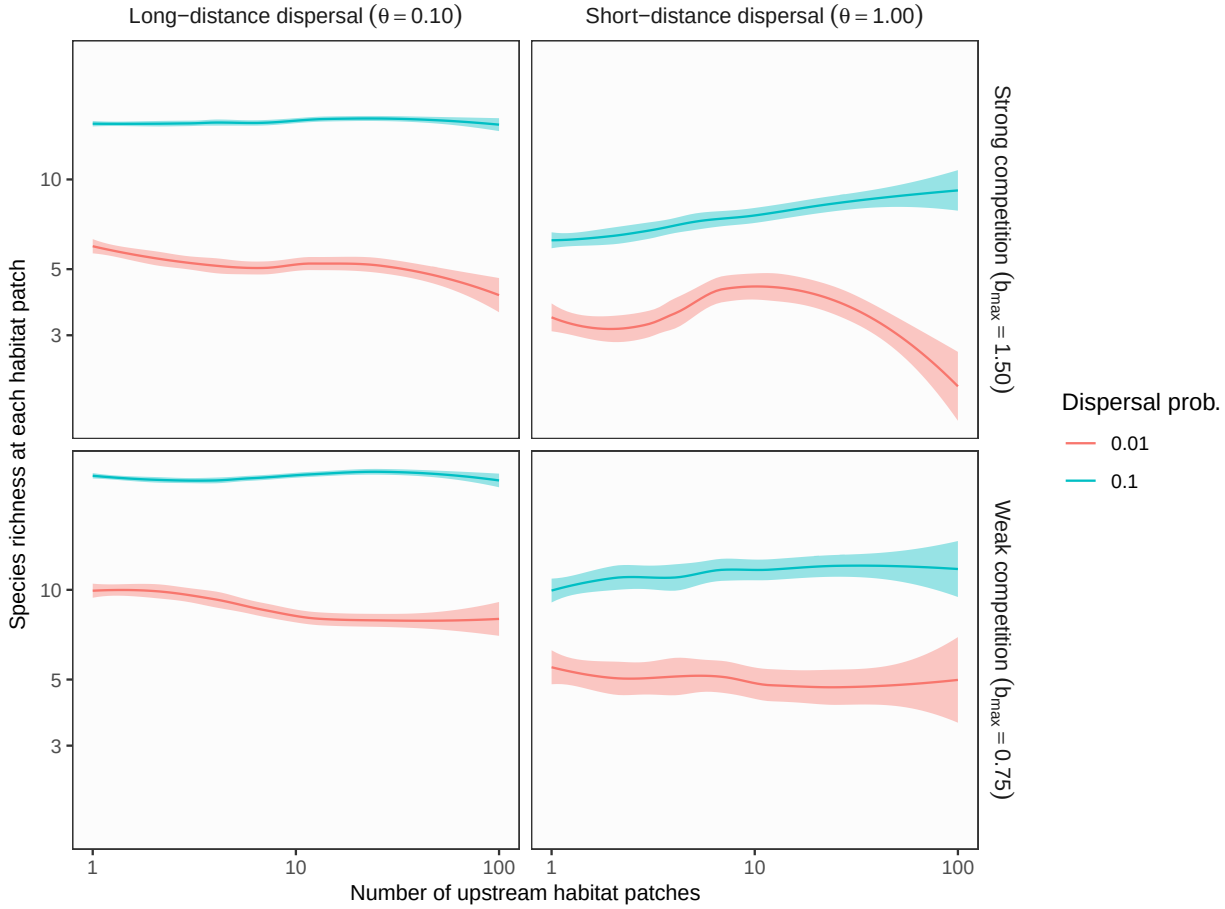

Figure S2 Theoretical predictions for longitudinal gradients of local species richness in branching networks. The longitudinal position of each habitat patch (x-axis) was expressed as the number of upstream habitat patches. In this simulation, environmental variation at headwaters (σ_h) is equal to local environmental noise (σ_l). Lines and shades are loess curves fitted to simulated data and their 95% confidence intervals. Each panel represents different ecological scenarios under which metacommunity dynamics were simulated. Rows represent different competition strength. Competition coefficients (b_{ij}) were varied randomly from 0 to 1.5 (top, strong competition) or 0.75 (bottom, weak competition). Columns and lines represent different dispersal scenarios (dispersal distance and probability). Left and right columns show long-distance (the rate parameter of an exponential dispersal kernel $\theta = 0.10$) and short-distance dispersal ($\theta = 1.0$) scenarios repectively. Red and blue lines show low ($p_d = 0.01$) and high dispersal probabilities ($p_d = 0.10$). Other parameters are as follows: environmental variation at headwaters $\sigma_h = 1$; local environmental noise $\sigma_l = 1$; ecosystem size $N_p = 100$; ecosystem complexity $P_b = 0.5$.

Figure S3 Longitudinal gradient of local species richness ($\sigma_h = 0.01$, $\sigma_l = 0.01$)

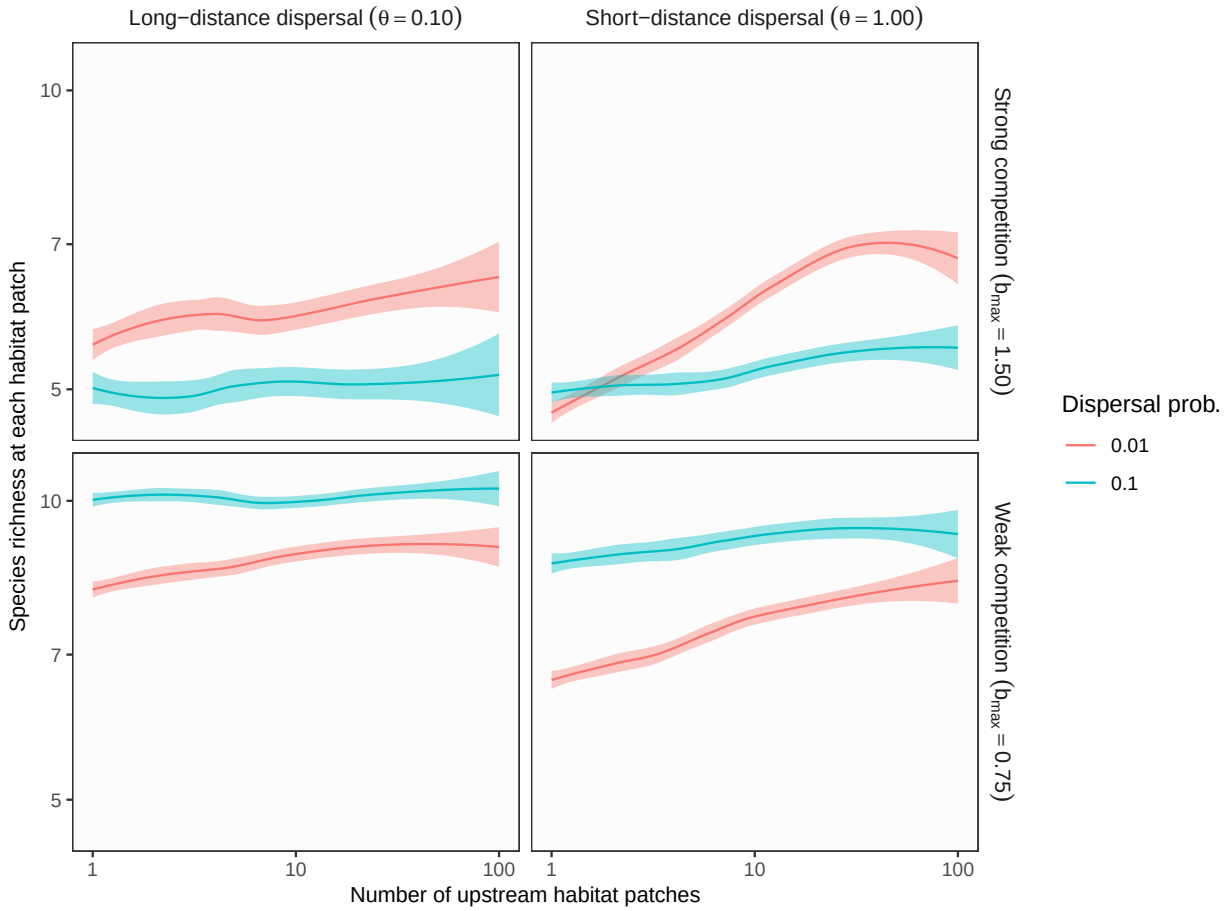

Figure S3 Theoretical predictions for longitudinal gradients of local species richness in branching networks. The longitudinal position of each habitat patch (x-axis) was expressed as the number of upstream habitat patches. In this simulation, environmental variation at headwaters (σ_h) is equal to local environmental noise (σ_l). Lines and shades are loess curves fitted to simulated data and their 95% confidence intervals. Each panel represents different ecological scenarios under which metacommunity dynamics were simulated. Rows represent different competition strength. Competition coefficients (b_{ij}) were varied randomly from 0 to 1.5 (top, strong competition) or 0.75 (bottom, weak competition). Columns and lines represent different dispersal scenarios (dispersal distance and probability). Left and right columns show long-distance (the rate parameter of an exponential dispersal kernel $\theta = 0.10$) and short-distance dispersal ($\theta = 1.0$) scenarios repectively. Red and blue lines show low ($p_d = 0.01$) and high dispersal probabilities ($p_d = 0.10$). Other parameters are as follows: environmental variation at headwaters $\sigma_h = 0.01$; local environmental noise $\sigma_l = 0.01$; ecosystem size $N_p = 100$; ecosystem complexity $P_b = 0.5$.

Figure S4 Longitudinal gradient of local species richness ($\sigma_h = 0.01$, $\sigma_l = 1$)

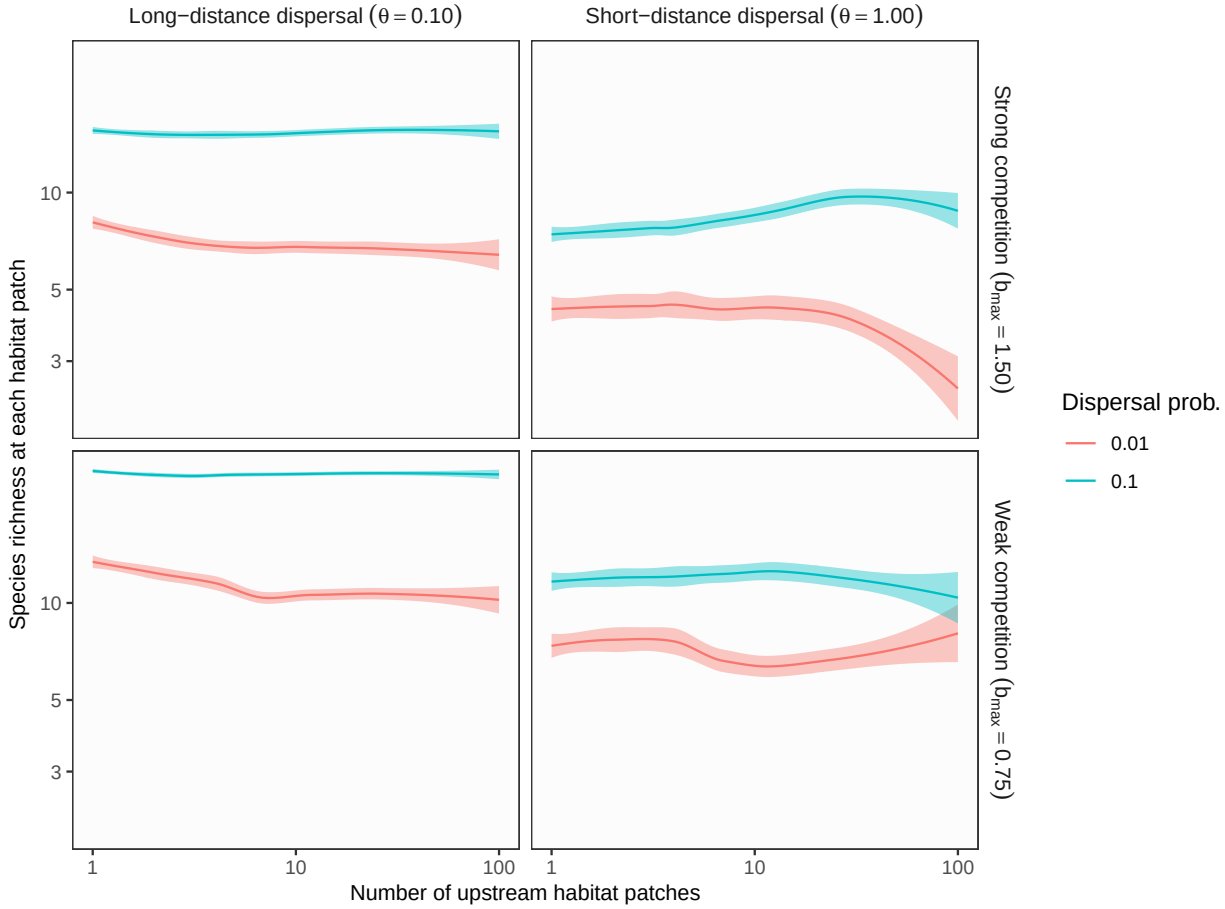

Figure S4 Theoretical predictions for longitudinal gradients of local species richness in branching networks. The longitudinal position of each habitat patch (x-axis) was expressed as the number of upstream habitat patches. In this simulation, environmental variation at headwaters (σ_h) is less than local environmental noise (σ_l). Lines and shades are loess curves fitted to simulated data and their 95% confidence intervals. Each panel represents different ecological scenarios under which metacommunity dynamics were simulated. Rows represent different competition strength. Competition coefficients (b_{ij}) were varied randomly from 0 to 1.5 (top, strong competition) or 0.75 (bottom, weak competition). Columns and lines represent different dispersal scenarios (dispersal distance and probability). Left and right columns show long-distance (the rate parameter of an exponential dispersal kernel $\theta = 0.10$) and short-distance dispersal ($\theta = 1.0$) scenarios repectively. Red and blue lines show low ($p_d = 0.01$) and high dispersal probabilities ($p_d = 0.10$). Other parameters are as follows: environmental variation at headwaters $\sigma_h = 0.01$; local environmental noise $\sigma_l = 1$; ecosystem size $N_p = 100$; ecosystem complexity $P_b = 0.5$.

Figure S5 Influence of ecosystem size ($p_d = 0.1$, $\sigma_h = 1$, $\sigma_l = 0.01$)

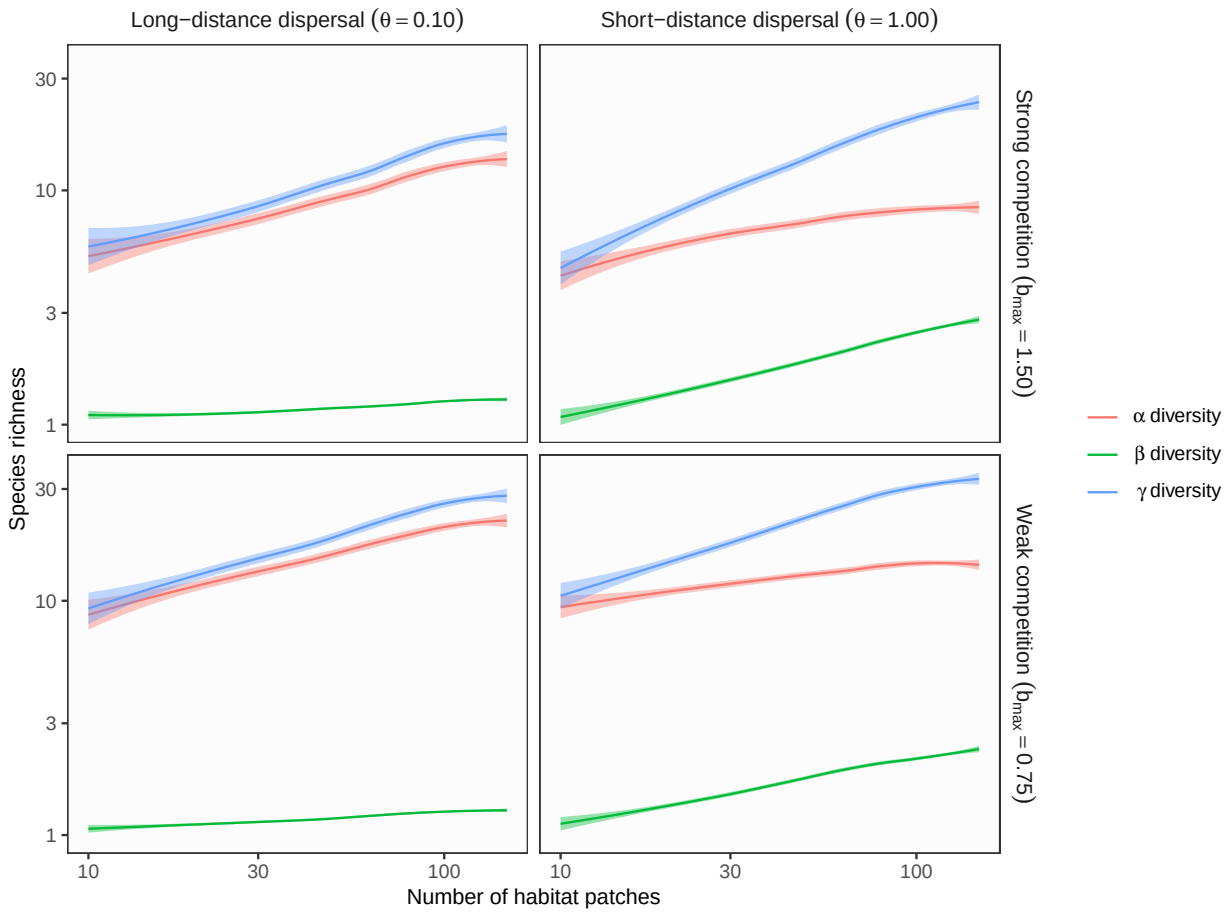

Figure S5 Theoretical predictions for ecosystem size influences (the number of habitat patches) on α, β, and γ diversity in branching networks. In this simulation, environmental variation at headwaters (σ_h) exceeds local environmental noise (σ_l). Lines and shades are loess curves fitted to simulated data and their 95% confidence intervals. Each panel represents different ecological scenarios under which metacommunity dynamics were simulated. Rows represent different competition strength. Competition coefficients (b_{ij}) were varied randomly from 0 to 1.5 (top, strong competition) or 0.75 (bottom, weak competition). Columns represent different dispersal scenarios. Two dispersal parameters were chosen to simulate scenarios with long-distance (the rate parameter of an exponential dispersal kernel $\theta = 0.10$) and short-distance dispersal ($\theta = 1.0$). Other parameters are as follows: dispersal probability $p_d = 0.1$; environmental variation at headwaters $\sigma_h = 1$; local environmental noise $\sigma_l = 0.01$.

Figure S6 Influence of ecosystem size ($p_d = 0.1$, $\sigma_h = 1$, $\sigma_l = 1$)

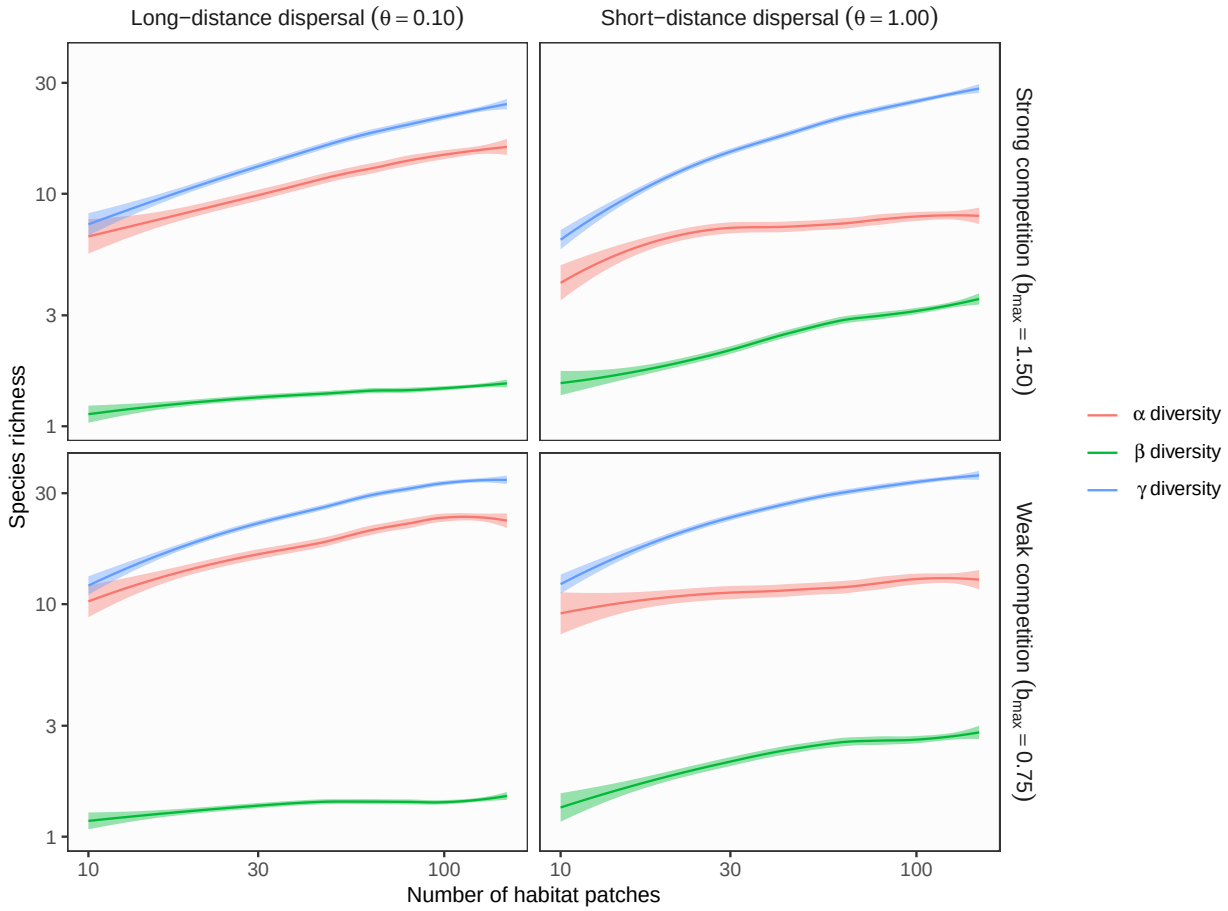

Figure S6 Theoretical predictions for ecosystem size influences (the number of habitat patches) on α, β, and γ diversity in branching networks. In this simulation, environmental variation at headwaters (σ_h) is equal to local environmental noise (σ_l). Lines and shades are loess curves fitted to simulated data and their 95% confidence intervals. Each panel represents different ecological scenarios under which metacommunity dynamics were simulated. Rows represent different competition strength. Competition coefficients (b_{ij}) were varied randomly from 0 to 1.5 (top, strong competition) or 0.75 (bottom, weak competition). Columns represent different dispersal scenarios. Two dispersal parameters were chosen to simulate scenarios with long-distance (the rate parameter of an exponential dispersal kernel $\theta = 0.10$) and short-distance dispersal ($\theta = 1.0$). Other parameters are as follows: dispersal probability $p_d = 0.1$; environmental variation at headwaters $\sigma_h = 1$; local environmental noise $\sigma_l = 1$.

Figure S7 Influence of ecosystem size ($p_d = 0.1$, $\sigma_h = 0.01$, $\sigma_l = 0.01$)

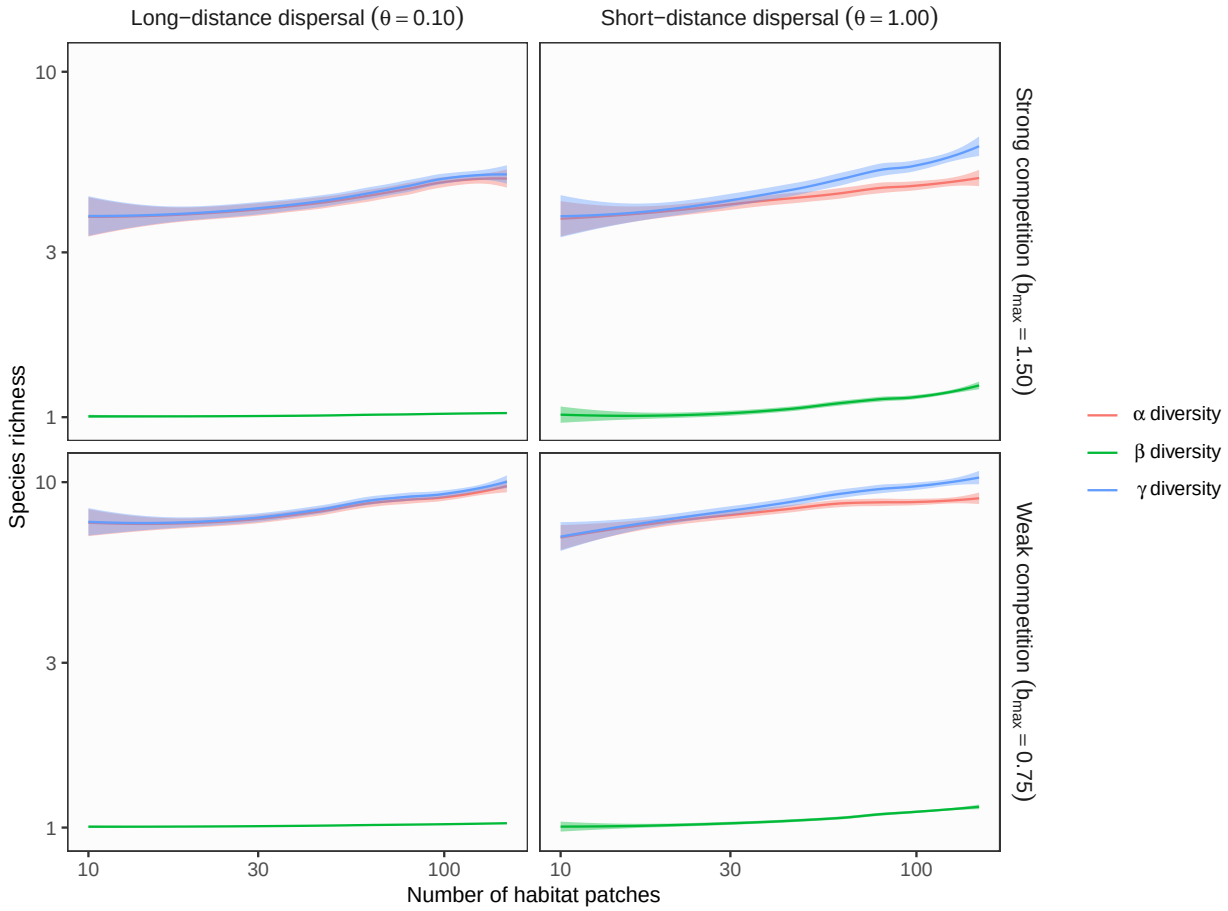

Figure S7 Theoretical predictions for ecosystem size influences (the number of habitat patches) on α, β, and γ diversity in branching networks. In this simulation, environmental variation at headwaters (σ_h) is equal to local environmental noise (σ_l). Lines and shades are loess curves fitted to simulated data and their 95% confidence intervals. Each panel represents different ecological scenarios under which metacommunity dynamics were simulated. Rows represent different competition strength. Competition coefficients (b_{ij}) were varied randomly from 0 to 1.5 (top, strong competition) or 0.75 (bottom, weak competition). Columns represent different dispersal scenarios. Two dispersal parameters were chosen to simulate scenarios with long-distance (the rate parameter of an exponential dispersal kernel $\theta = 0.10$) and short-distance dispersal ($\theta = 1.0$). Other parameters are as follows: dispersal probability $p_d = 0.1$; environmental variation at headwaters $\sigma_h = 0.01$; local environmental noise $\sigma_l = 0.01$.

Figure S8 Influence of ecosystem size ($p_d = 0.1$, $\sigma_h = 0.01$, $\sigma_l = 1$)

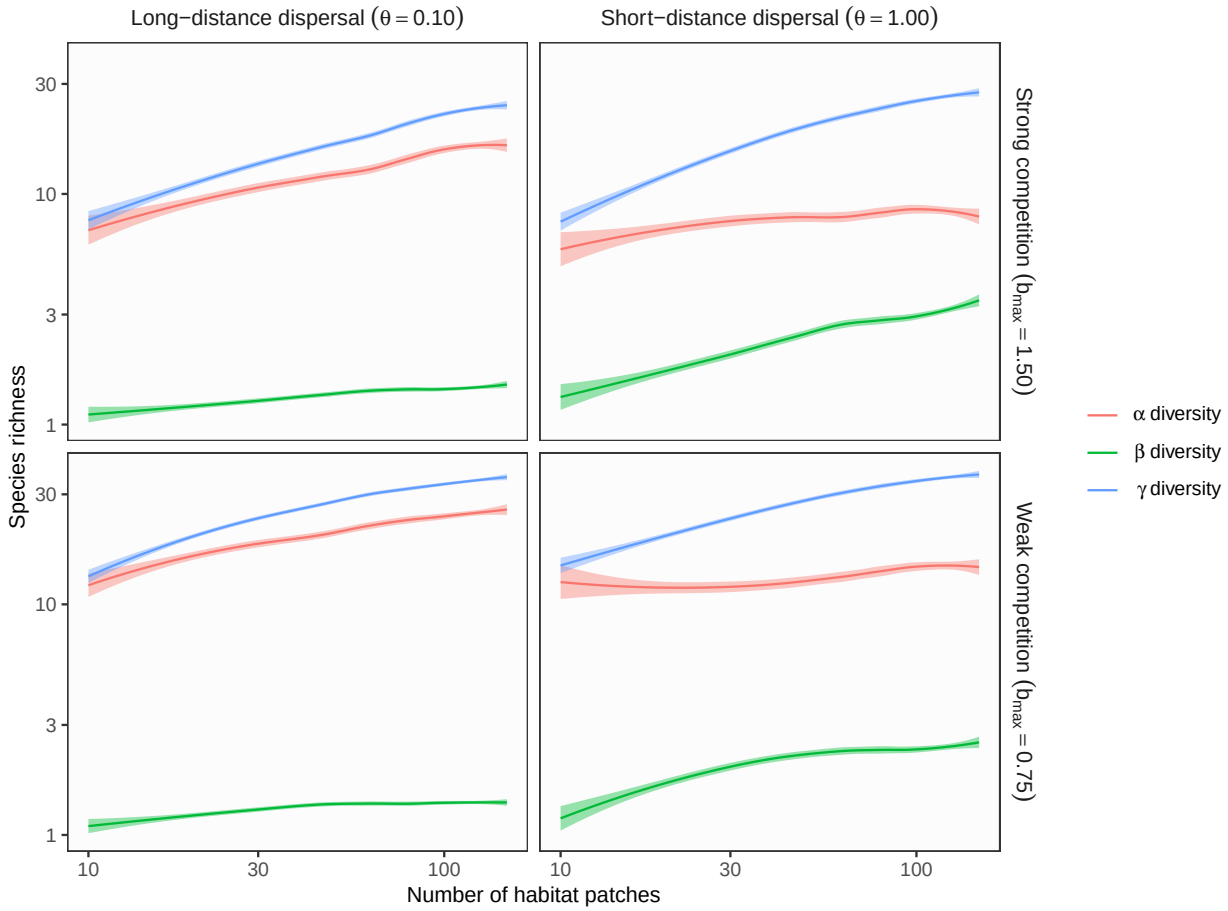

Figure S8 Theoretical predictions for ecosystem size influences (the number of habitat patches) on α, β, and γ diversity in branching networks. In this simulation, environmental variation at headwaters (σ_h) is less than local environmental noise (σ_l). Lines and shades are loess curves fitted to simulated data and their 95% confidence intervals. Each panel represents different ecological scenarios under which metacommunity dynamics were simulated. Rows represent different competition strength. Competition coefficients (b_{ij}) were varied randomly from 0 to 1.5 (top, strong competition) or 0.75 (bottom, weak competition). Columns represent different dispersal scenarios. Two dispersal parameters were chosen to simulate scenarios with long-distance (the rate parameter of an exponential dispersal kernel $\theta = 0.10$) and short-distance dispersal ($\theta = 1.0$). Other parameters are as follows: dispersal probability $p_d = 0.1$; environmental variation at headwaters $\sigma_h = 0.01$; local environmental noise $\sigma_l = 1$.

Figure S9 Influence of ecosystem size ($p_d = 0.01$, $\sigma_h = 1$, $\sigma_l = 1$)

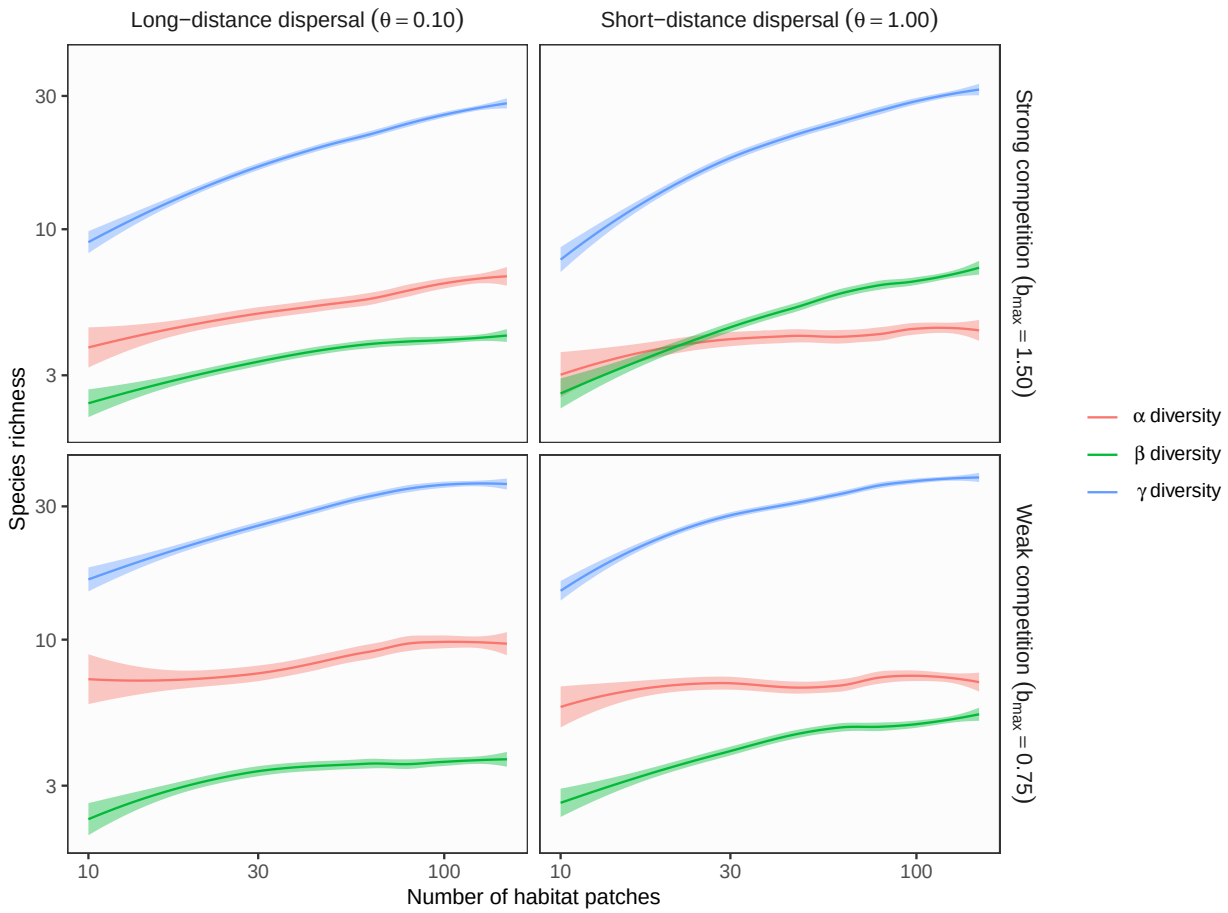

Figure S9 Theoretical predictions for ecosystem size influences (the number of habitat patches) on α, β, and γ diversity in branching networks. In this simulation, environmental variation at headwaters (σ_h) is equal to local environmental noise (σ_l). Lines and shades are loess curves fitted to simulated data and their 95% confidence intervals. Each panel represents different ecological scenarios under which metacommunity dynamics were simulated. Rows represent different competition strength. Competition coefficients (b_{ij}) were varied randomly from 0 to 1.5 (top, strong competition) or 0.75 (bottom, weak competition). Columns represent different dispersal scenarios. Two dispersal parameters were chosen to simulate scenarios with long-distance (the rate parameter of an exponential dispersal kernel $\theta = 0.10$) and short-distance dispersal ($\theta = 1.0$). Other parameters are as follows: dispersal probability $p_d = 0.01$; environmental variation at headwaters $\sigma_h = 1$; local environmental noise $\sigma_l = 1$.

Figure S10 Influence of ecosystem size ($p_d = 0.01$, $\sigma_h = 0.01$, $\sigma_l = 0.01$)

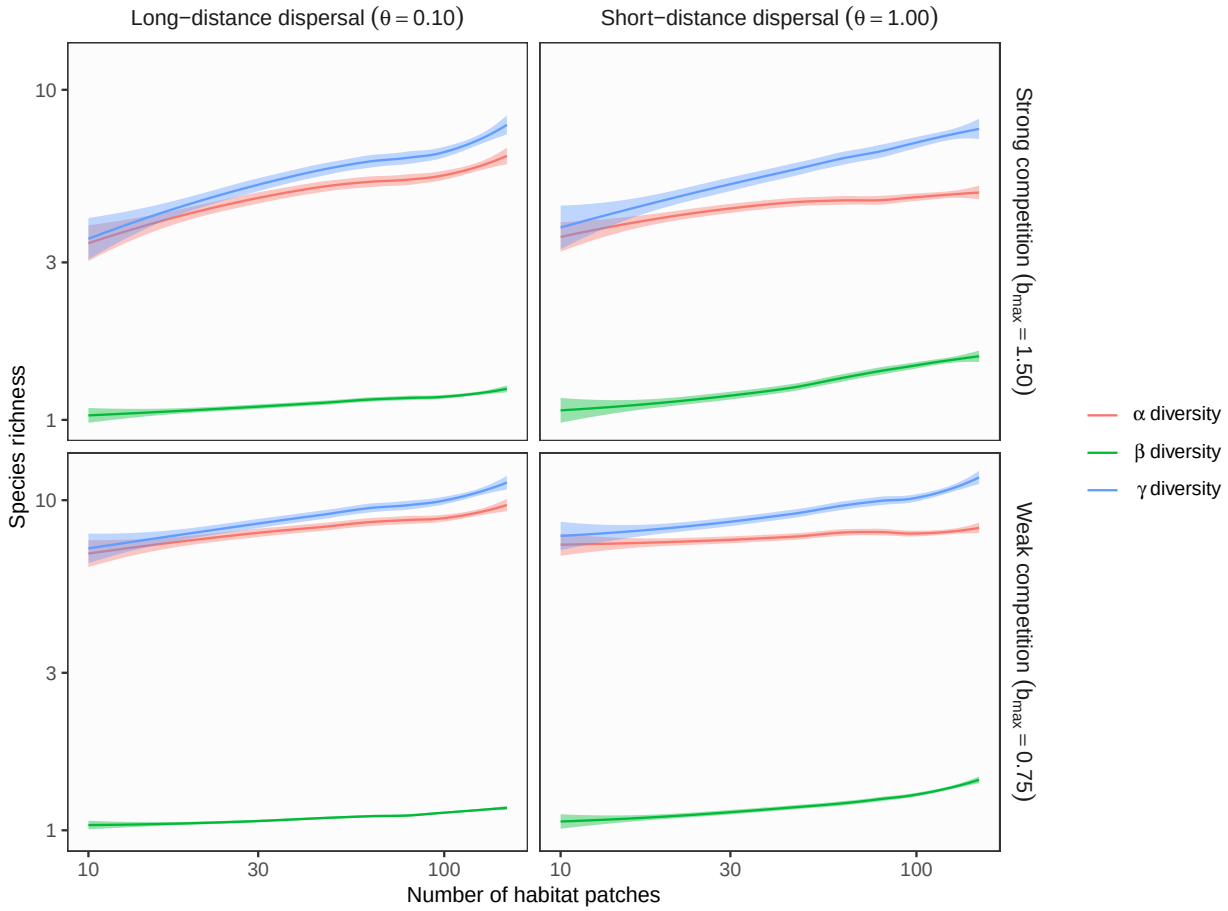

Figure S10 Theoretical predictions for ecosystem size influences (the number of habitat patches) on α, β, and γ diversity in branching networks. In this simulation, environmental variation at headwaters (σ_h) is equal to local environmental noise (σ_l). Lines and shades are loess curves fitted to simulated data and their 95% confidence intervals. Each panel represents different ecological scenarios under which metacommunity dynamics were simulated. Rows represent different competition strength. Competition coefficients (b_{ij}) were varied randomly from 0 to 1.5 (top, strong competition) or 0.75 (bottom, weak competition). Columns represent different dispersal scenarios. Two dispersal parameters were chosen to simulate scenarios with long-distance (the rate parameter of an exponential dispersal kernel $\theta = 0.10$) and short-distance dispersal ($\theta = 1.0$). Other parameters are as follows: dispersal probability $p_d = 0.01$; environmental variation at headwaters $\sigma_h = 0.01$; local environmental noise $\sigma_l = 0.01$.

Figure S11 Influence of ecosystem size ($p_d = 0.01$, $\sigma_h = 0.01$, $\sigma_l = 1$)

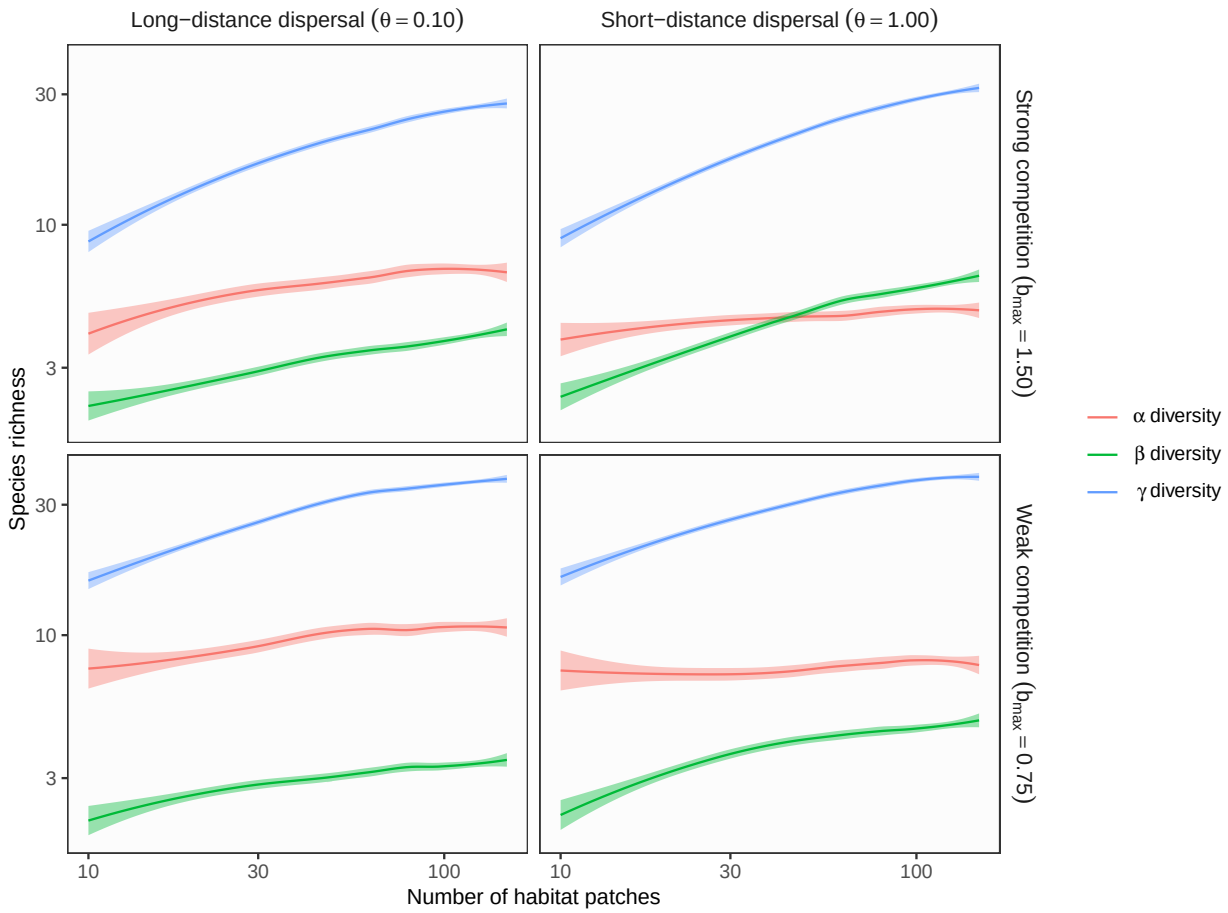

Figure S11 Theoretical predictions for ecosystem size influences (the number of habitat patches) on α, β, and γ diversity in branching networks. In this simulation, environmental variation at headwaters (σ_h) is less than local environmental noise (σ_l). Lines and shades are loess curves fitted to simulated data and their 95% confidence intervals. Each panel represents different ecological scenarios under which metacommunity dynamics were simulated. Rows represent different competition strength. Competition coefficients (b_{ij}) were varied randomly from 0 to 1.5 (top, strong competition) or 0.75 (bottom, weak competition). Columns represent different dispersal scenarios. Two dispersal parameters were chosen to simulate scenarios with long-distance (the rate parameter of an exponential dispersal kernel $\theta = 0.10$) and short-distance dispersal ($\theta = 1.0$). Other parameters are as follows: dispersal probability $p_d = 0.01$; environmental variation at headwaters $\sigma_h = 0.01$; local environmental noise $\sigma_l = 1$.

Figure S12 Influence of ecosystem complexity ($p_d = 0.1$, $\sigma_h = 1$, $\sigma_l = 0.01$)

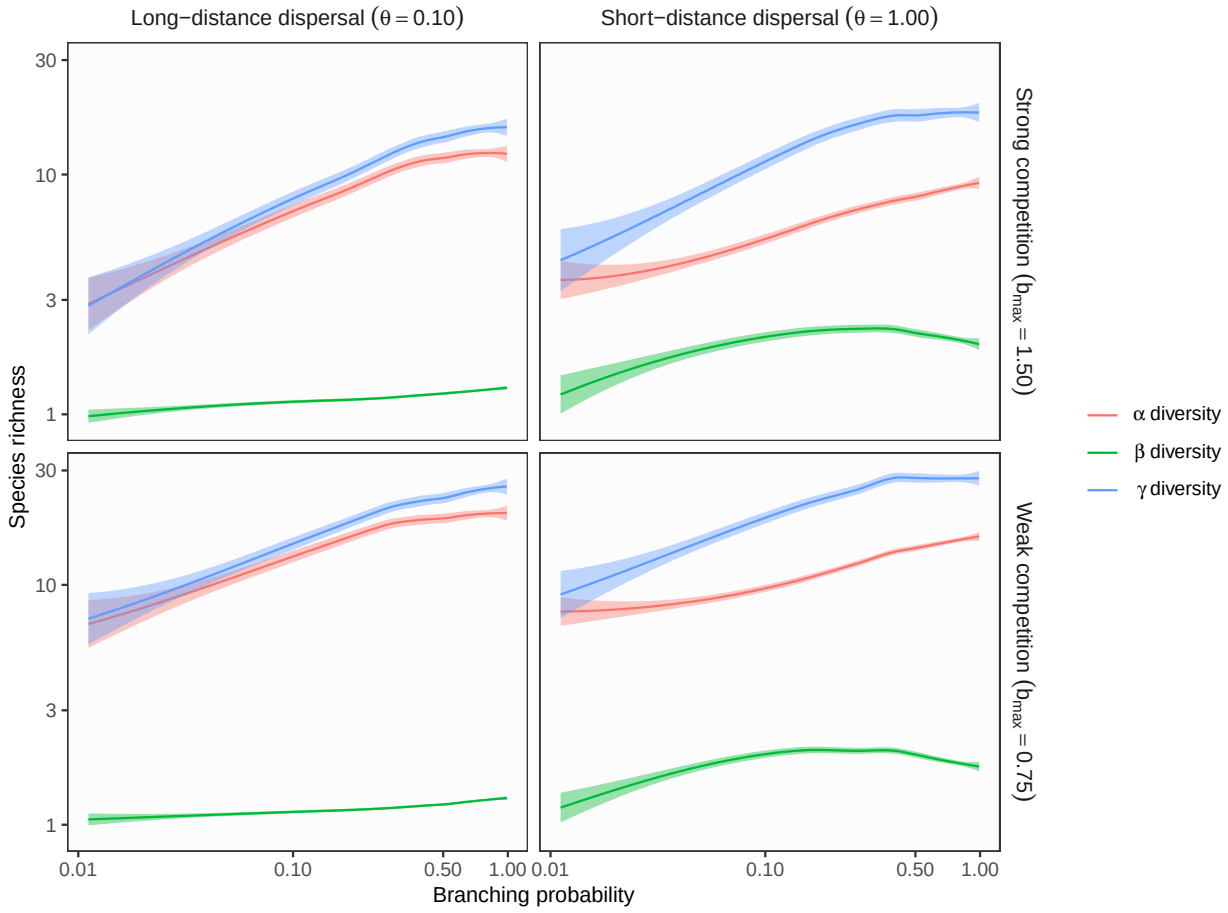

Figure S12 Theoretical predictions for ecosystem complexity influences (branching probability) on α, β, and γ diversity in branching networks. In this simulation, environmental variation at headwaters (σ_h) exceeds local environmental noise (σ_l). Lines and shades are loess curves fitted to simulated data and their 95% confidence intervals. Each panel represents different ecological scenarios under which metacommunity dynamics were simulated. Rows represent different competition strength. Competition coefficients (b_{ij}) were varied randomly from 0 to 1.5 (top, strong competition) or 0.75 (bottom, weak competition). Columns represent different dispersal scenarios. Two dispersal parameters were chosen to simulate scenarios with long-distance (the rate parameter of an exponential dispersal kernel $\theta = 0.10$) and short-distance dispersal ($\theta = 1.0$). Other parameters are as follows: dispersal probability $p_d = 0.1$; environmental variation at headwaters $\sigma_h = 1$; local environmental noise $\sigma_l = 0.01$.

Figure S13 Influence of ecosystem complexity ($p_d = 0.1$, $\sigma_h = 1$, $\sigma_l = 1$)

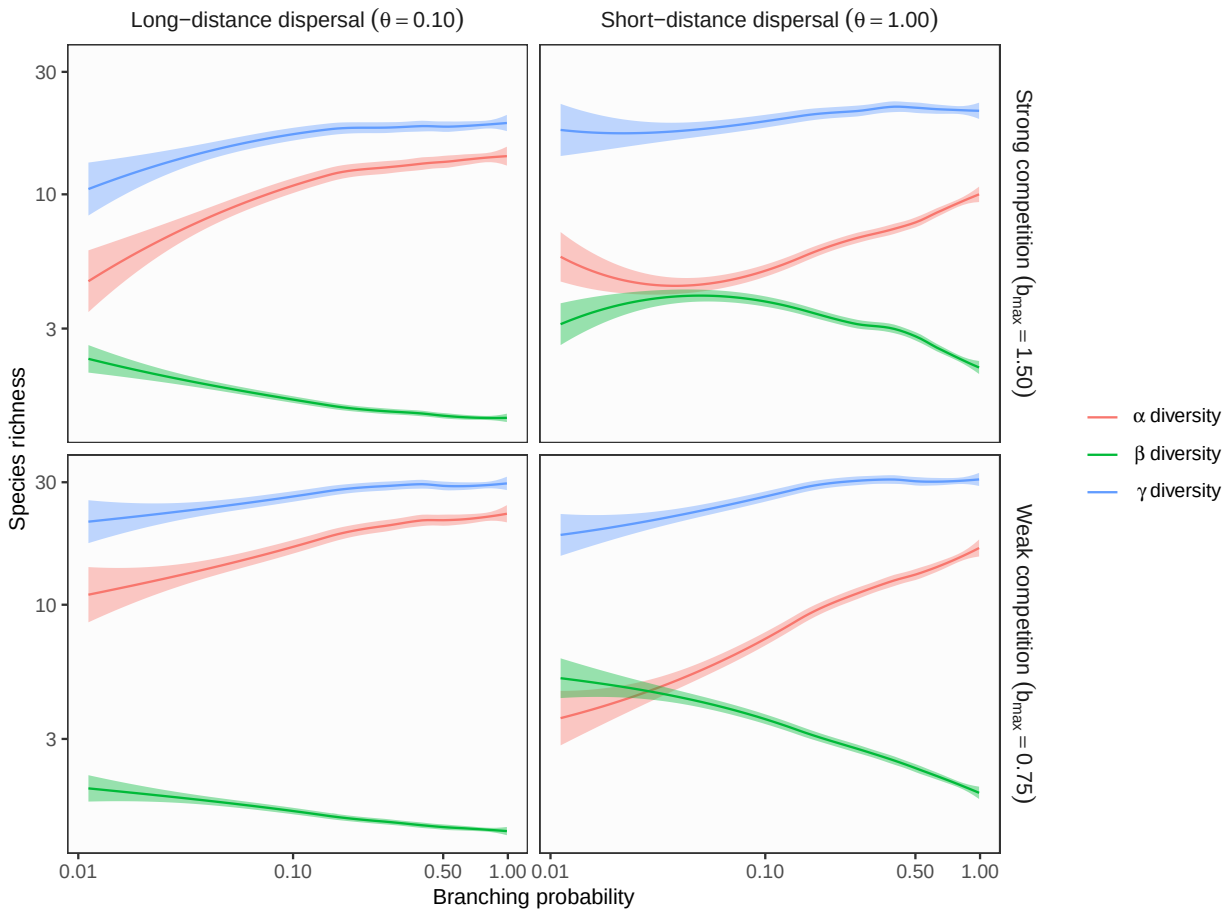

Figure S13 Theoretical predictions for ecosystem complexity influences (branching probability) on α, β, and γ diversity in branching networks. In this simulation, environmental variation at headwaters (σ_h) is equal to local environmental noise (σ_l). Lines and shades are loess curves fitted to simulated data and their 95% confidence intervals. Each panel represents different ecological scenarios under which metacommunity dynamics were simulated. Rows represent different competition strength. Competition coefficients (b_{ij}) were varied randomly from 0 to 1.5 (top, strong competition) or 0.75 (bottom, weak competition). Columns represent different dispersal scenarios. Two dispersal parameters were chosen to simulate scenarios with long-distance (the rate parameter of an exponential dispersal kernel $\theta = 0.10$) and short-distance dispersal ($\theta = 1.0$). Other parameters are as follows: dispersal probability $p_d = 0.1$; environmental variation at headwaters $\sigma_h = 1$; local environmental noise $\sigma_l = 1$.

Figure S14 Influence of ecosystem complexity ($p_d = 0.1$, $\sigma_h = 0.01$, $\sigma_l = 0.01$)

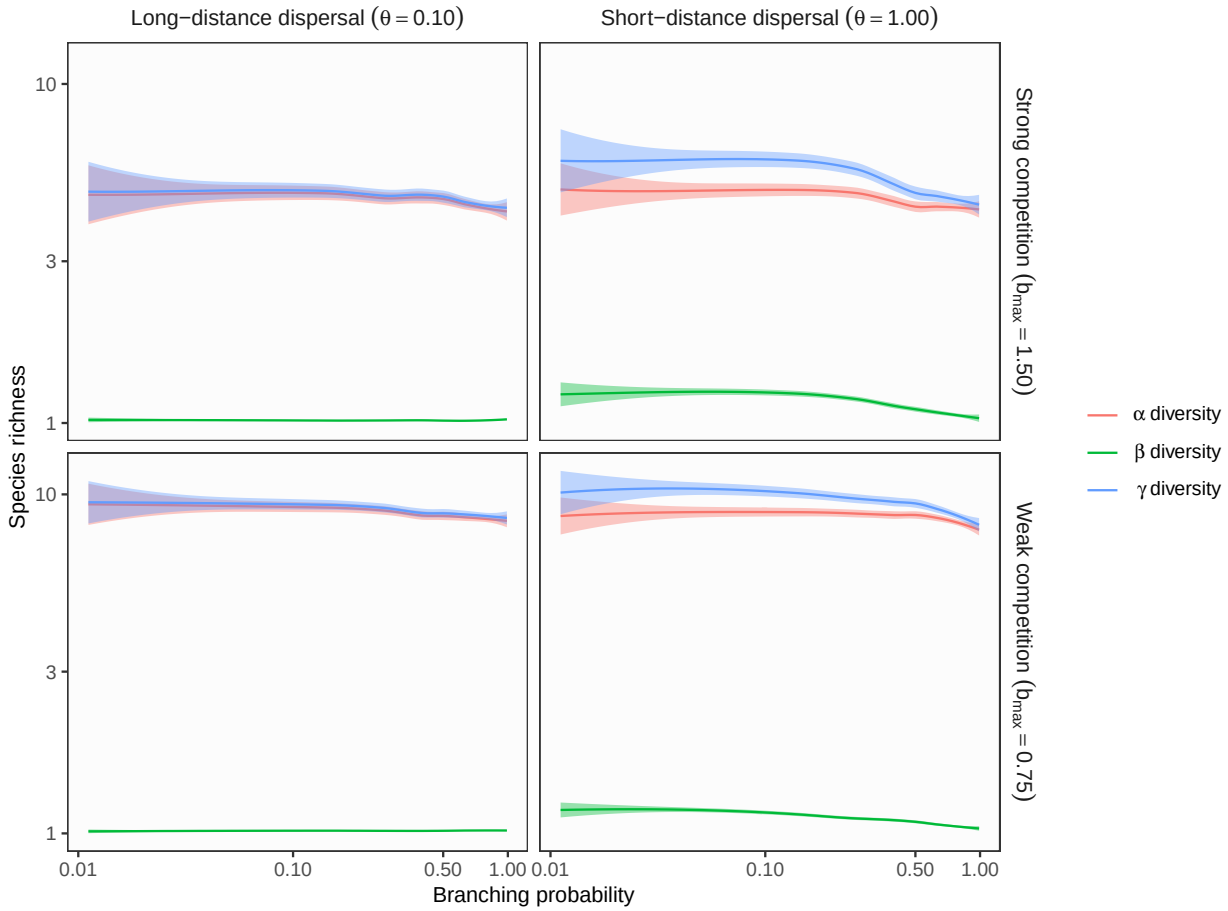

Figure S14 Theoretical predictions for ecosystem complexity influences (branching probability) on α, β, and γ diversity in branching networks. In this simulation, environmental variation at headwaters (σ_h) is equal to local environmental noise (σ_l). Lines and shades are loess curves fitted to simulated data and their 95% confidence intervals. Each panel represents different ecological scenarios under which metacommunity dynamics were simulated. Rows represent different competition strength. Competition coefficients (b_{ij}) were varied randomly from 0 to 1.5 (top, strong competition) or 0.75 (bottom, weak competition). Columns represent different dispersal scenarios. Two dispersal parameters were chosen to simulate scenarios with long-distance (the rate parameter of an exponential dispersal kernel $\theta = 0.10$) and short-distance dispersal ($\theta = 1.0$). Other parameters are as follows: dispersal probability $p_d = 0.1$; environmental variation at headwaters $\sigma_h = 0.01$; local environmental noise $\sigma_l = 0.01$.

Figure S15 Influence of ecosystem complexity ($p_d = 0.1$, $\sigma_h = 0.01$, $\sigma_l = 1$)

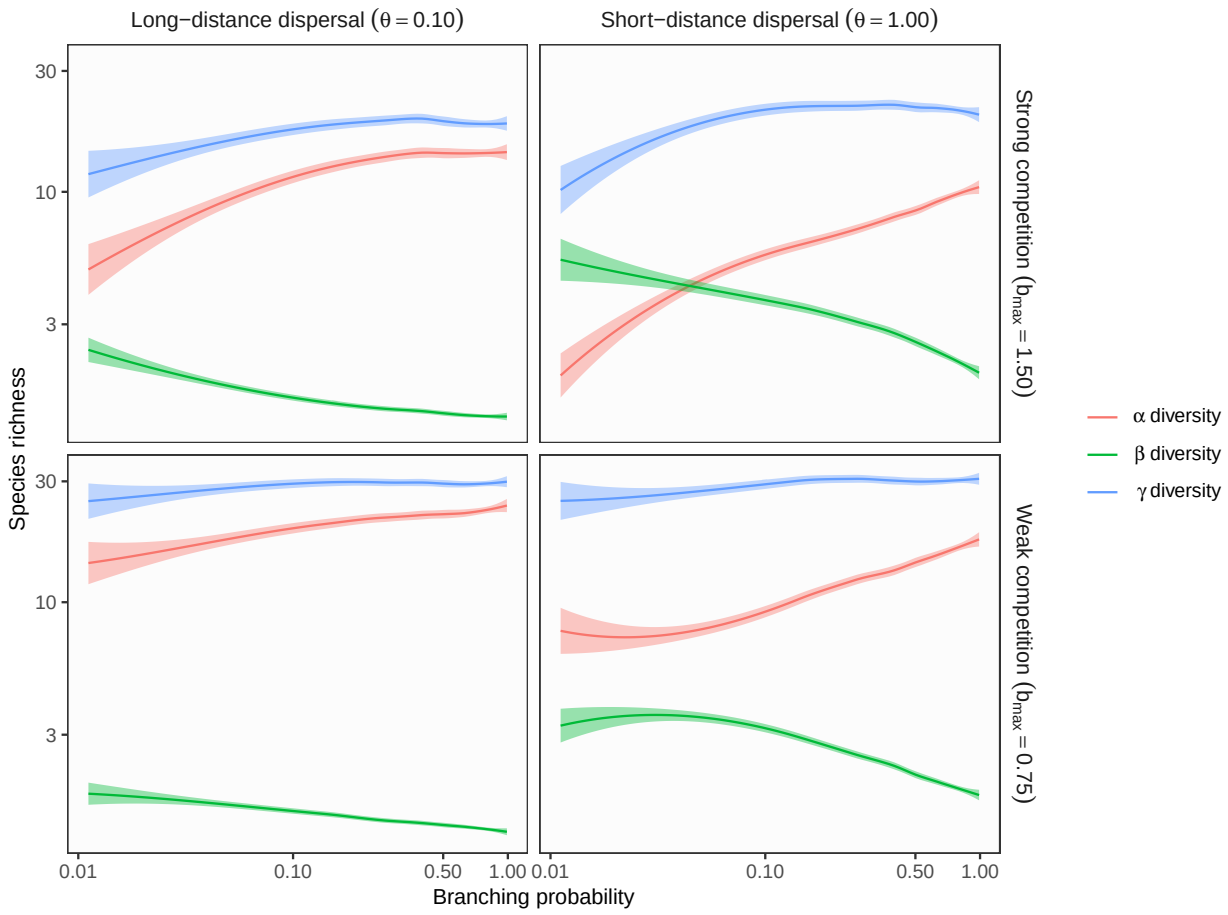

Figure S15 Theoretical predictions for ecosystem complexity influences (branching probability) on α, β, and γ diversity in branching networks. In this simulation, environmental variation at headwaters (σ_h) is less than local environmental noise (σ_l). Lines and shades are loess curves fitted to simulated data and their 95% confidence intervals. Each panel represents different ecological scenarios under which metacommunity dynamics were simulated. Rows represent different competition strength. Competition coefficients (b_{ij}) were varied randomly from 0 to 1.5 (top, strong competition) or 0.75 (bottom, weak competition). Columns represent different dispersal scenarios. Two dispersal parameters were chosen to simulate scenarios with long-distance (the rate parameter of an exponential dispersal kernel $\theta = 0.10$) and short-distance dispersal ($\theta = 1.0$). Other parameters are as follows: dispersal probability $p_d = 0.1$; environmental variation at headwaters $\sigma_h = 0.01$; local environmental noise $\sigma_l = 1$.

Figure S16 Influence of ecosystem complexity ($p_d = 0.01$, $\sigma_h = 1$, $\sigma_l = 1$)

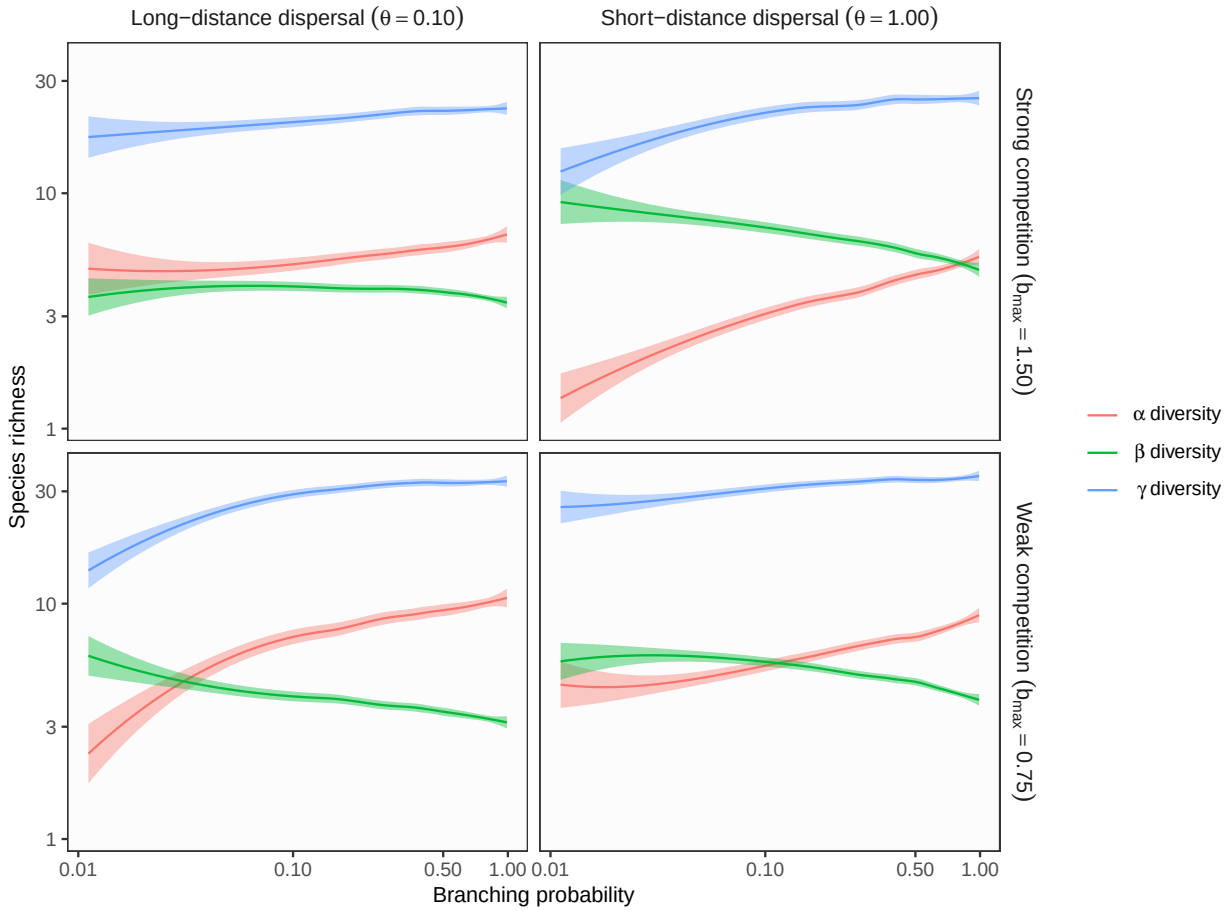

Figure S16 Theoretical predictions for ecosystem complexity influences (branching probability) on α, β, and γ diversity in branching networks. In this simulation, environmental variation at headwaters (σ_h) is equal to local environmental noise (σ_l). Lines and shades are loess curves fitted to simulated data and their 95% confidence intervals. Each panel represents different ecological scenarios under which metacommunity dynamics were simulated. Rows represent different competition strength. Competition coefficients (b_{ij}) were varied randomly from 0 to 1.5 (top, strong competition) or 0.75 (bottom, weak competition). Columns represent different dispersal scenarios. Two dispersal parameters were chosen to simulate scenarios with long-distance (the rate parameter of an exponential dispersal kernel $\theta = 0.10$) and short-distance dispersal ($\theta = 1.0$). Other parameters are as follows: dispersal probability $p_d = 0.01$; environmental variation at headwaters $\sigma_h = 1$; local environmental noise $\sigma_l = 1$.

Figure S17 Influence of ecosystem complexity ($p_d = 0.01$, $\sigma_h = 0.01$, $\sigma_l = 0.01$)

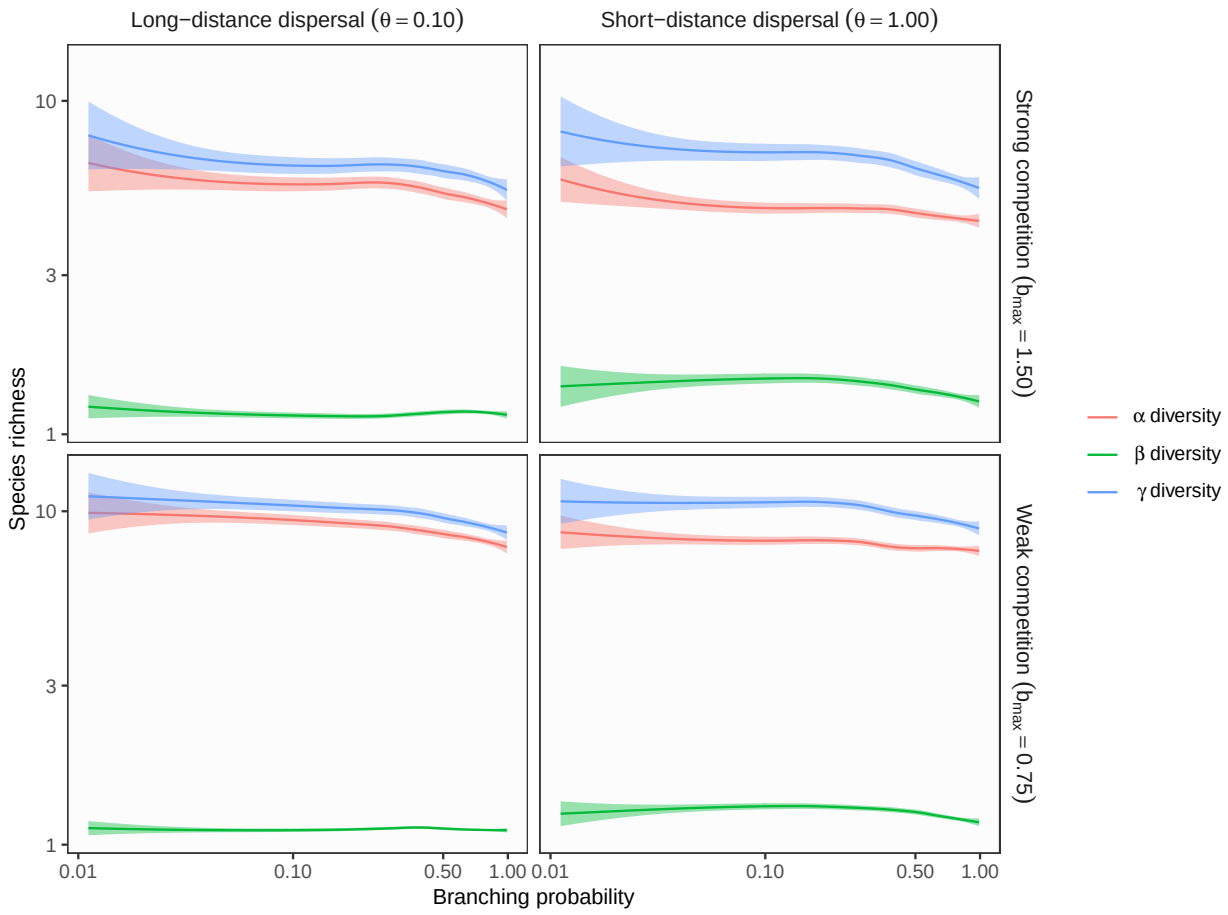

Figure S17 Theoretical predictions for ecosystem complexity influences (branching probability) on α, β, and γ diversity in branching networks. In this simulation, environmental variation at headwaters (σ_h) is equal to local environmental noise (σ_l). Lines and shades are loess curves fitted to simulated data and their 95% confidence intervals. Each panel represents different ecological scenarios under which metacommunity dynamics were simulated. Rows represent different competition strength. Competition coefficients (b_{ij}) were varied randomly from 0 to 1.5 (top, strong competition) or 0.75 (bottom, weak competition). Columns represent different dispersal scenarios. Two dispersal parameters were chosen to simulate scenarios with long-distance (the rate parameter of an exponential dispersal kernel $\theta = 0.10$) and short-distance dispersal ($\theta = 1.0$). Other parameters are as follows: dispersal probability $p_d = 0.01$; environmental variation at headwaters $\sigma_h = 0.01$; local environmental noise $\sigma_l = 0.01$.

Figure S18 Influence of ecosystem complexity ($p_d = 0.01$, $\sigma_h = 0.01$, $\sigma_l = 1$)

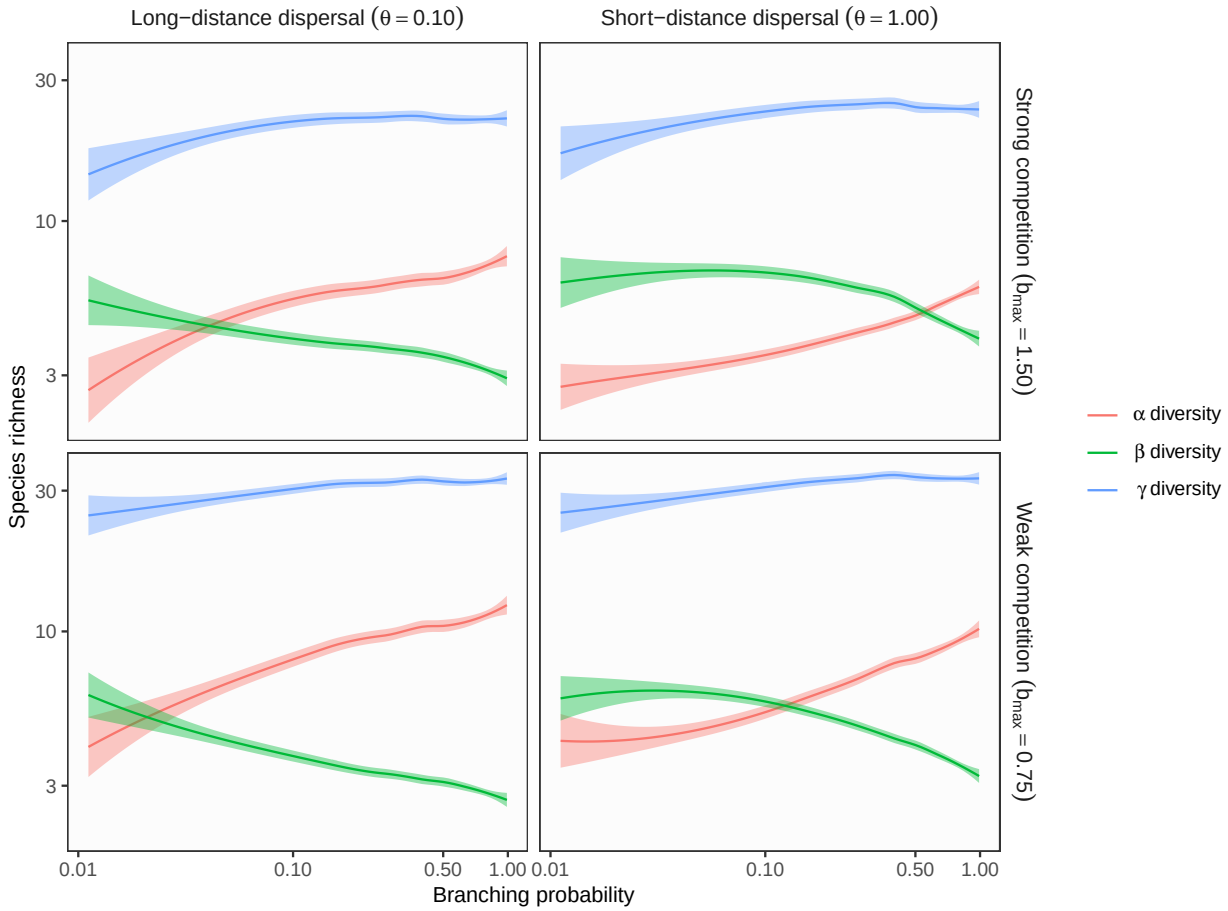

Figure S18 Theoretical predictions for ecosystem complexity influences (branching probability) on α, β, and γ diversity in branching networks. In this simulation, environmental variation at headwaters (σ_h) is less than local environmental noise (σ_l). Lines and shades are loess curves fitted to simulated data and their 95% confidence intervals. Each panel represents different ecological scenarios under which metacommunity dynamics were simulated. Rows represent different competition strength. Competition coefficients (b_{ij}) were varied randomly from 0 to 1.5 (top, strong competition) or 0.75 (bottom, weak competition). Columns represent different dispersal scenarios. Two dispersal parameters were chosen to simulate scenarios with long-distance (the rate parameter of an exponential dispersal kernel $\theta = 0.10$) and short-distance dispersal ($\theta = 1.0$). Other parameters are as follows: dispersal probability $p_d = 0.01$; environmental variation at headwaters $\sigma_h = 0.01$; local environmental noise $\sigma_l = 1$.

Figure S19 Correlation structure of explanatory variables

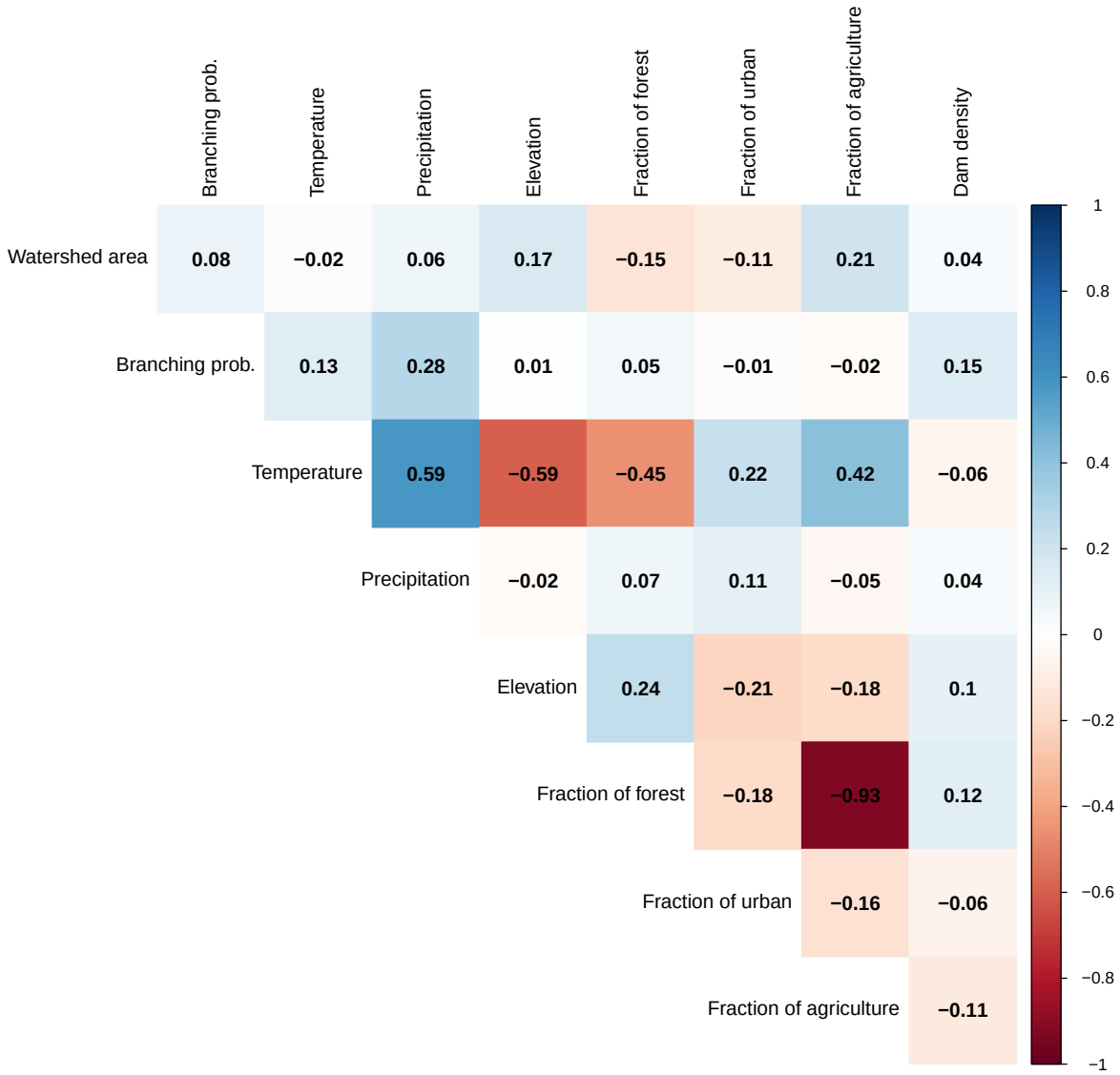

Figure S19 Correlation structure of potential explanatory variables for riverine diversity metrics. Numeric values in each cell are the Pearson's correlation coefficients. Positive and negative correlations were colored in blue and red, respectively, and darker colors indicate stronger correlations. Environmental variables (temperature, precipitation, elevation, fraction of forest, fraction of urban, fraction of agriculture, and dam density) were expressed as deviations from the regional averages to remove any regional effects.

Figure S20 Sensitivity analysis of asymptotic species richness

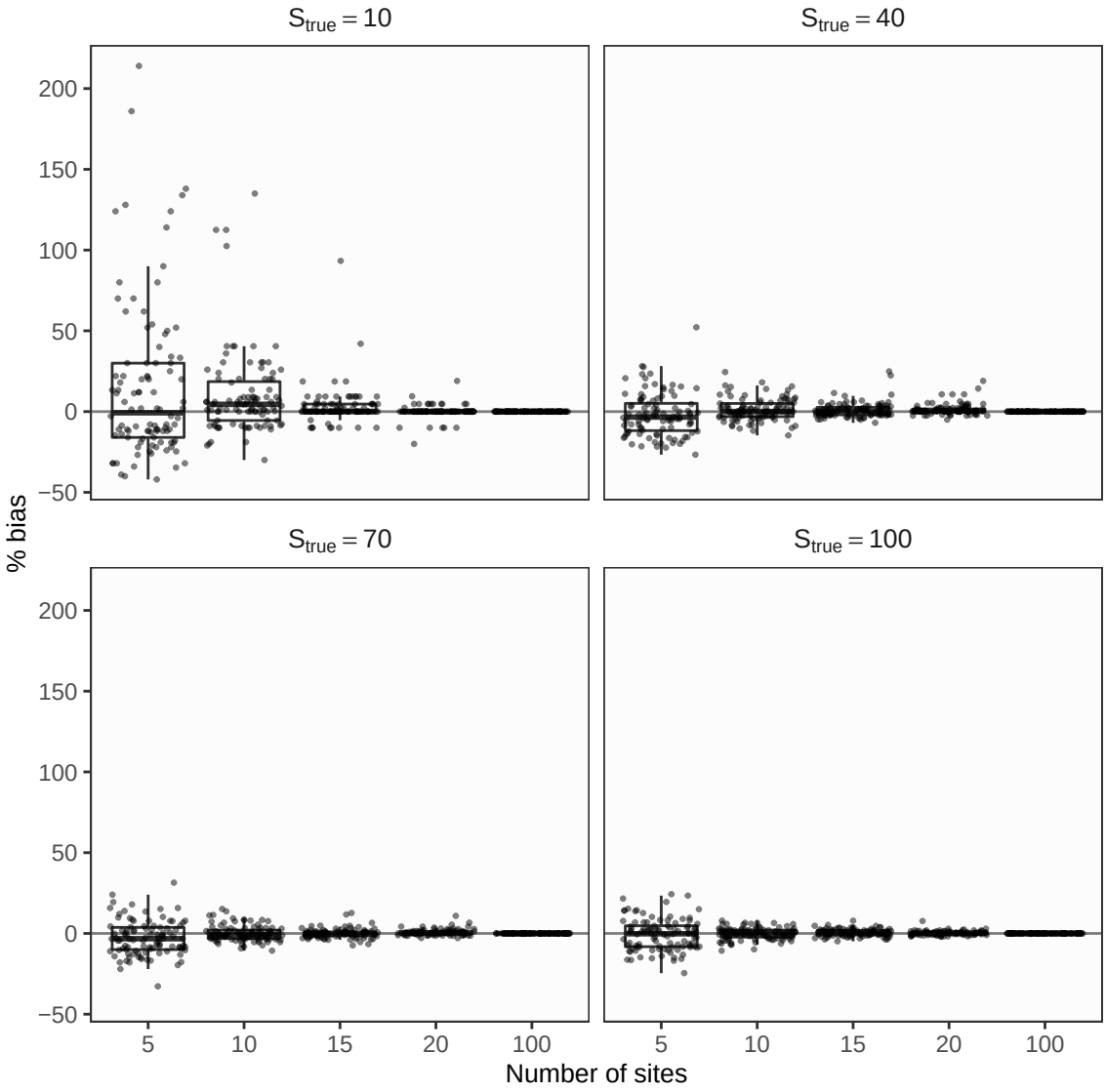

Figure S20 Sensitivity analysis of asymptotic species richness in relation to true species richness S_{true} (panels) and the number of sampling sites N_{site} (x-axis). Positive values of % bias indicate an overestimation of species richness, while negative values indicate an underestimation. Different panels show results with different true species richness. The center lines are median values, the box boundaries 25 and 75 percentiles, and the whiskers 5 and 95 percentiles. Dots represent individual simulation replicates.

References

1. R. M. McDowall, Fighting the flow: Downstream-upstream linkages in the ecology of diadromous fish faunas in West Coast New Zealand rivers. *Freshwater Biology* **40**, 111–122 (1998).

2. F. J. Rahel, W. A. Hubert, Fish assemblages and habitat gradients in a Rocky Mountain-Great Plains stream: Biotic zonation and additive patterns of community change. *Transactions of the American Fisheries Society* **120**, 319–332 (1991).

3. A. Terui, Y. Miyazaki, Three ecological factors influencing riverine fish diversity in the Shubuto River system, Japan: Habitat capacity, habitat heterogeneity and immigration. *Limnology* **17**, 143–149 (2016).

4. K. Bsemer, *et al.*, Headwaters are critical reservoirs of microbial diversity for fluvial networks. *Proceedings of the Royal Society B: Biological Sciences* **280**, 20131760 (2013).

5. E. Harvey, I. Gounand, A. Fronhofer Emanuel, F. Altermatt, Disturbance reverses classic biodiversity predictions in river-like landscapes. *Proceedings of the Royal Society B: Biological Sciences* **285**, 20182441 (2018).

6. U. S. Geological Survey and U.S. Department of Agriculture Natural Resources Conservation Service, Federal standards and procedures for the national watershed boundary dataset (WBD) (4 ed.): Techniques and Methods 11-A3 (2013).

7. T. C. Hsieh, K. H. Ma, A. Chao, iNEXT: An R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods in Ecology and Evolution* **7**, 1451–1456 (2016).

8. A. Gelman, I. Pardoe, Average predictive comparisons for models with nonlinearity, interactions, and variance components. *Sociological Methodology* **37**, 23–51 (2007).