Supplementary Material

Environmental factors override dispersal-related factors in shaping diatom and macroinvertebrate communities within stream networks in China

Siwen He, Janne Soininen, Kai Chen, Beixin Wang[*]

*Correspondence: <u>wangbeixin@njau.edu.cn</u>

Text S1. Supplementary information on the linear mixed effect model (LME)

We used the linear mixed effect model (LME) to assess the effect of the environmental and geographical distances on community similarity (Sarremejane et al., 2017; He et al., 2020). We used minimum likelihood population effects (MLPE, Clarke et al., 2002) method to account for the non-independence of the distance matrices. In this method, a linear mixed effect model is used as an alternative to traditional linear regression. The random effect (i.e. covariate structure) of the model incorporates a parameter (ρ) that accounts for the proportion of the total variance (σ^2) due to the correlation between distances that originates from the same sampling sites. The covariance between two pairwise distances without a common site is 0. For example, if d_{ij} is the geographical distance between site i and j, then d_{12} and d_{13} have a covariance of $\rho\sigma^2$, whereas d_{12} and d_{34} have a covariance of 0. Residual maximum likelihood (Clarke et al., 2002) is used to obtain unbiased estimates of MLPE model parameters. We assess the amount of variation explained by fixed effect (i.e. environmental and geographical distances) using a marginal R^2 statistic for each model. Here, R^2_β is calculated from the F-statistic and the degrees of the freedom derived from Kenward-Roger's estimates (Edwards et al., 2008) using the difference in explained variation between the model that only includes the fixed and random effects and a null model that only includes random effects. Collinearity between explanatory variables (i.e. environmental and geographical distances) can be a serious problem when one wants to assess the "true" effect of a variable. We tested for the collinearity between explanatory variables prior to LME analysis. In the case of correlations between explanatory variables, we calculated the R^2_β by comparing a model combing the explanatory variable of interest, its significant ($P < 0.05$) collinear variables and the random effects to a null model combing only the collinear variables and the random effects (Sarremejane et al. 2017). We used the R packages "lme4" (Bates et al., 2015) for MLPE, and "pbkrtest" (Halekoh and Højsgaard, 2014) for the Kenward-Roger's estimates.

References not in the main text

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. J. Stat. Softw. 67, 1-48. doi:10.18637/jss.v067.i01.

Clarke, R. T., Rothery, P., and Raybould, A. F. (2002). Confidence Limits for Regression Relationships between Distance Matrices: Estimating Gene Flow with Distance. J. Agric. Biol. Environ. Stat. 7, 361-372. doi:10.1198/108571102320.

Edwards, L. J., Muller, K. E., Wolfinger, R. D., Qaqish, B. F., and Schabenberger, O. (2008). An R2 statistic for fixed effects in the linear mixed model. Stat. Med. 27, 6137-6157. doi:10.1002/sim.3429

Halekoh, U., and Højsgaard, S. (2014). A Kenward-Roger Approximation and Parametric Bootstrap Methods for Tests in Linear Mixed Models - The R Package pbkrtest. J. Stat. Softw. 59, 32. doi:10.18637/jss.v059.i09.

Text S2. Supplementary information on connectivity measures.

We followed Kindlmann and Burel (2008) to define metacommunity connectivity as "the ease with which these individual macroinvertebrates and diatoms can move about within the region". As 1) most of macroinvertebrates can actively dispersal to select suitable habitats, while small diatoms are passively randomly dispersed by wind, stream flow and animals (Farjalla et al., 2012); 2) the connectivity for active dispersers relies on migration, but the connectivity for passive dispersers is driven by physical dispersal (Yeh et al., 2015). We thus estimated the metacommunity connectivity for macroinvertebrates and diatoms differently. Metacommunity connectivity for macroinvertebrates was measured as:

$$Avg.\ Con. = \frac{1}{n}\sum_{i=1}^{n} Con._i$$

$$Con._i = \frac{1}{m}\frac{1}{n-1}\sum_{\substack{j=1 \\ j \neq i}}^{n}\sum_{k=1}^{m} p_{jk}\exp(-d_{ij})$$

Avg. Con. is the average of site connectivities, d_{ij} is the distance between site i (focal site) and j (surrounding site), p indicates the presence or absence of kth taxa in the jth site, n is the total number of sites, and m is the total number of taxa in the site pair (i.e. site i and j). This equation quantifies how easy the n-1 sites can contribute the same kth species to the focal site i, with a weighting function depends on the distance between sites, since the distance reflects the resistance of colonization. The contribution of each surrounding site j is computed by the geographic distance with the negative exponential kernel, meaning that active dispersers which are separated far away should have a larger resistance to colonization, and thus have smaller contribution to the connectivity.

However, diatoms belong to passive dispersers; their dispersal relies on physical mechanisms (i.e. dispersed by wind, stream flow and animals). Therefore, the connectivity (contribution of each surrounding site j) depends on the physical dispersal strength; this is fundamentally different from the assumption for active dispersers. For passive dispersers, the sites that are far apart should result in larger resistance of colonization. Therefore, for that sites located far away to have the same species, it requires stronger physical dispersal strength to overcome the resistance, and thus have large contribution to the connectivity. As such, metacommunity connectivity for diatoms was measured as:

$$Avg.\ Con. = \frac{1}{n}\sum_{i=1}^{n} Con._i$$

$$Con._i = \frac{1}{m}\frac{1}{n-1}\sum_{\substack{j=1 \\ j \neq i}}^{n}\sum_{k=1}^{m} p_{jk}\ 1/\exp(-d_{ij})$$

Analyses of metacommunity connectivity were conducted in R software (R Core Team 2018).

References not in the main text

Kindlmann, P., and Burel, F. (2008). Connectivity measures: a review. Landscape Ecol. 23, 879-890. doi: 10.1007/s10980-008-9245-4.

Yeh, Y. C., Peres-Neto, P. R., Huang, S. W., Lai, Y. C., Tu, C. Y., Shiah, F. K., et al. (2015). Determinism of bacterial metacommunity dynamics in the southern East China Sea varies depending on hydrography. Ecography 38, 198-212. doi:10.1111/ecog.00986.

Table S1. Final model results of multiple regression of distance matrices (MRM) analyses between community dissimilarity and explanatory distance matrices (environmental and geographical distance) for stream macroinvertebrates and diatoms in ITR, QTR and MKR, based on 10,000 permutations. Analyses were conducted for whole network and separately for headwaters and mainstems based on abundance data. - represents matrix was not in included in the final model. Four explanatory matrices are local environmental, climatic, land use and geographical distance matric. *** P < 0.001; ** P < 0.01; * P < 0.05.

Taxon	Basin	Stream type	Standardized Coefficients					Full model	
			Intercept	Local environment	Climate	Land use	Geographical	R2	F
Macroinvertebrates	ITR	Whole network	0.62	-	0.31***	-	-	0.09	44.70***
		Headwaters	0.59	-	0.25*	-	-	0.06	6.61*
		Mainstems	0.50	0.28**	0.31**	-	-	0.21	13.91***
	QTR	Whole network	0.60	-	0.22**	0.45***	-	0.36	118.92/***
		Headwaters	0.42	0.48***	0.34**	-	-	0.53	56.62***
		Mainstems	-	-		-	-	-	-
	MKR	Whole network	0.48	0.33**	-	0.38**	-	0.39	136.96***
		Headwaters	0.65	-	-	0.71***	-	0.50	104.71***
		Mainstems	0.46	0.53***	-	0.39***	-	0.40	34.96***
Diatoms	ITR	Whole network	0.65	-	0.32***	-	-	0.10	48.78***
		Headwaters	0.45	0.43***	0.30**	-	-	0.32	23.76***
		Mainstems	0.52	0.46**	0.36**	-	-	0.42	36.50***
	QTR	Whole network	0.75	0.26*	-	0.29**	-	0.26	77.19***
		Headwaters	0.71	-	0.34*	-	-	0.11	13.53*
		Mainstems	0.59	0.54***	0.20*	-	-	0.37	31.13***
	MKR	Whole network	0.77	0.22**	-	-	0.16**	0.08	19.06***
		Headwaters	0.80	-	0.28*	-	-	0.07	8.76*
		Mainstems	0.79	0.38**	-	-		0.15	17.55***

Table S2. Summary of determination coefficients (R^2_{β}) for the relationships between community dissimilarity and explanatory distance matrices (environmental and geographical distance) for stream macroinvertebrates and diatoms in the ITR, QTR and MKR regions based on the linear mixed effect model (LME). Four explanatory matrices are local environmental, climatic, land use and geographical distance matrices. Analyses were conducted only at the whole network level.

Taxon	Region	R^2_{β}			
		Local environment	Climate	Land use	Geographical
Macroinvertebrates	ITR	0.2798	0.2368	0.0003	0.0004
	QTR	0.1434	0.4316	0.6224	0.0002
	MKR	0.4109	0.0834	0.2568	0.0101
Diatoms	ITR	0.2875	0.0206	0.0090	0.0036
	QTR	0.3200	0.0799	0.1117	0.0614
	MKR	0.2221	0.0024	0.0567	0.1322

Table S3 Results of two-way ANOVA testing for the effects of organism group (i.e., between macroinvertebrates and diatoms) and region on the strength of relationship between community dissimilarities and environmental distance calculated using the coefficient of determination (R^2_β). At Local-scale + Basin-scale, we treated all the obtained determination coefficients (n = 18) as independent estimates in the tests. At Basin-scale, we treated all the obtained land use R^2_β and climate R^2_β (n = 12) as independent estimates in the tests. Df, degrees of freedom; MS, mean squares; P, significance level; SS, sum of squares. Significant P-values are in bold.

Factor	Local-scale + Basin-scale					Basin-scale				
	Df	SS	MS	F	P	Df	SS	MS	F	P
Organism group	1	0.103	0.102	3.786	0.076	1	0.153	0.153	14.636	**0.009**
Region	2	0.070	0.035	1.285	0.312	2	0.140	0.070	6.718	**0.029**
Organism group × Region	2	0.020	0.010	0.362	0.704	2	0.064	0.032	3.074	0.121
Residuals	12	0.325	0.027			6	0.063	0.010		

Table S4. Set of environmental variables identified for stream macroinvertebrates in the ITR, QTR and MKR regions by BIO-ENV analysis. BIO-ENV analyses were conducted for the whole network and separately for headwater and downstream sites based on presence-absence data (N = number of variables identified by BIO-ENV analysis). Abbreviations below axis are as follows. Water temperature, WT; total dissolved solids, TDS; conductivity, Cond; calcium, Ca^{2+}; magnesium, Mg^{2+}; total nitrogen, TN; total phosphorus, TP; ammonia nitrogen, NH_4-N; phosphate, PO_4-P; Mean Diurnal Range (Mean of monthly (max temp - min temp)), BIO2; Isothermality, BIO3; Mean Temperature of Wettest Quarter, BIO8; Mean Temperature of Driest Quarter, BIO9; Mean Temperature of Warmest Quarter, BIO10; Annual Precipitation, BIO12; Precipitation Seasonality (Coefficient of Variation), BIO15; Precipitation of Wettest Quarter, BIO16; Precipitation of Driest Quarter, BIO17; Precipitation of Warmest Quarter, BIO18.

Region	Stream type	Local environment	N	Climate	N	Land use	N
ITR	Whole network	TDS, TP, % sands, % gravels	4	Elevation, BIO15, BIO9	3		0
	Headwater	TDS, PO_4-P, TP, % gravels	4	BIO2, BIO10, BIO17	3		0
	Downstream	WT, TP, % gravels, % cobbles	4		0		0
QTR	Whole network	Wetted width, WT, Cond, Ca^{2+}, PO_4-P, % cobbles	6	Elevation	1	% forest	1
	Headwater	WT, pH	2	Elevation	1		0
	Downstream	Cond, NH_4-N, % cobbles, pH	4		0		0
MKR	Whole network	WT, TN, TP, % sands, % cobbles	5		0	% urban, % farmland	2
	Headwater	WT, Water depth, TN, TP, % sands, % cobbles	6		0	% urban, % farmland	2
	Downstream		0		0		0

Table S5. Set of environmental variables identified for stream diatoms in the ITR, QTR and MKR regions by BIO-ENV analysis. BIO-ENV analyses were conducted for the whole network and separately for headwater and downstream sites based on presence-absence data (N = number of variables identified by BIO-ENV analysis).

Region	Stream type	Local environment	N	Climate	N	Land use	N
ITR	Whole network	TDS, Ca^{2+}, Mg^{2+}, NH_4-N, % gravels, pH	6	Elevation, BIO2, BIO3, BIO12, BIO18	5		0
	Headwater	TDS, Ca^{2+}, NH_4-N, % sands, pH	5	Elevation, BIO18	2		0
	Downstream	Mg^{2+}, % gravels, pH	3		0		0
QTR	Whole network	Water depth, Cond, Mg^{2+}, Ca^{2+}, PO_4-P	5		0	% forest	1
	Headwater	WT, Cond, TP, Mg^{2+}, % gravels, % sands	6		0		0
	Downstream	Water depth, Cond, NH_4-N, Mg^{2+}, Ca^{2+}, PO_4-P, % cobbles, pH	8		0		0
MKR	Whole network	WT, Wetted width, Water depth, TN, TP, % cobbles, % boulders	7		0		0
	Headwater		0	BIO8, BIO9, BIO16	3		0
	Downstream	WT, Wetted width, Water depth, Cond, TN, PO_4-P	6		0		0

Table S6. Results of metacommunity connectivity estimations and mean pairwise beta-diversity (Sørensen dissimilarity) measures for macroinvertebrates and diatoms at different stream locations (i.e., headwater and downstream) in three regions.

Region	Stream location	Macroinvertebrates		Diatoms	
		Connectivity	Beta-diversity	Connectivity	Beta-diversity
ITR	Headwater	9.90×10^{-3}	0.46	6.10×10^{121}	0.57
	Downstream	3.55×10^{-3}	0.52	2.37×10^{131}	0.61
QTR	Headwater	2.02×10^{-3}	0.45	1.27×10^{84}	0.48
	Downstream	7.68×10^{-3}	0.56	1.16×10^{70}	0.56
MKR	Headwater	6.66×10^{-7}	0.56	1.12×10^{68}	0.68
	Downstream	1.10×10^{-3}	0.46	6.45×10^{62}	0.71

Figure S1. The correlations between diatom community composition and macroinvertebrate community composition controlling for environmental distance (white bar) and geographical distance (gray bar) in three basins. * P < 0.05.