

Ecography

ECOG-00986

Yeh, Y.-C., Peres-Neto, P. R., Hunag, S.-W., Lai, Y.-C., Tu, C.-Y., Shiah, F.-K., Gong, G.-C. and Hsieh, C.-H. 2014. Determinism of bacterial metacommunity dynamics in the southern East China Sea varies depending on hydrography. – *Ecography* doi: 10.1111/ecog.00986

Supplementary material

Supplementary material Appendix 1. Sequencing information for each sampling station.

Table A1. Sequencing information for each sampling station.

Station	Total reads	Total OTUs	Number of reads in subsamples	Average OTUs in subsamples \pm SE
<u>May</u>				
1	44388	760	24730	575 \pm 1.1
3	43653	913	24730	691 \pm 1.2
5	48101	1085	24730	770 \pm 1.3
7	45329	617	24730	461 \pm 0.9
9	24730	489	24730	489 \pm 0.1
11	55262	1069	24730	718 \pm 1.3
13	63133	841	24730	518 \pm 1.2
<u>December</u>				
1	36358	990	36300	986 \pm 0.2
3	47887	955	36300	841 \pm 0.9
5	39837	1053	36300	1006 \pm 0.6
7	73635	990	36300	725 \pm 1.1
9	46198	893	36300	795 \pm 0.9
<u>August</u>				
1	13916	315	12000	288 \pm 0.5
3	35128	735	12000	398 \pm 1.1
5	65363	670	12000	257 \pm 1.0
7	56390	913	12000	375 \pm 1.3
9	12037	718	12000	717 \pm 1.0
11	61667	927	12000	365 \pm 1.2
13	56959	1268	12000	531 \pm 1.2

Supplementary material Appendix 2. Estimation of connectivity

The definition of connectivity for a metacommunity is often unclear in observational research. Here, we followed Kindlmann and Burel (2008) to define metacommunity connectivity as “the ease with which these individuals can move about within the landscape”. Based on this definition, the connectivity function should vary among different dispersal modes (e.g. active versus passive dispersers, see Figure A2). That is because the connectivity for active dispersers relies on migration, but the connectivity for passive dispersers is driven by physical dispersal. The original equation of connectivity presented by Henriques-Silva et al. (2012) was applied in fish metacommunity, which belongs to active dispersers. Henriques-Silva et al. (2012) modified the formula of Hanski (Hanski 1994) to estimate metacommunity connectivity for active dispersers as:

$$Avg.Con.Active = \frac{1}{n} \sum_{i=1}^n Con.Active_i \quad (Eq. A1)$$

$$Con.Active_i = \frac{1}{m} \frac{1}{n-1} \sum_{\substack{j=1 \\ j \neq i}}^n \sum_{k=1}^m P_{jk} \exp(-d_{ij}) \quad (Eq. A2)$$

Avg.Con.Active is the connectivity at the metacommunity scale, which represents the average of patch connectivities. The patch connectivity (*Con.Active_i*) measures the average geographic distance (d_{ij}) between focal patch i and all other $n-1$ patches with the form of negative exponential kernel. Parameter P indicates the presence (1) or absence (0) of the k_{th} species in the j_{th} patch and m represents the total number of species in the patch pair (i.e. the focal and contributing patch). This equation quantifies how easy the $n-1$ patches can contribute the same k_{th} species to the focal patch, with a weighting function depends on the distance between patches, since the distance reflects the resistance of colonization. The contribution of each non-focal patch 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.

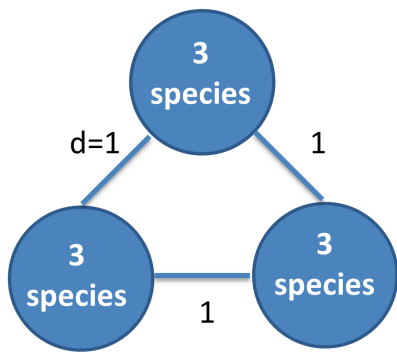
However, marine free-living bacteria belong to passive dispersers; their dispersal relies on physical mechanisms (i.e. the circulation and mixing of water masses). Therefore, the connectivity (contribution of each non-focal patch) depends on the physical dispersal strength; this is fundamentally different from the assumption for active dispersers. For passive dispersers, the patches that are far apart should result in larger resistance of colonization. Therefore, for that patches 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, we change the estimation of patch connectivity to:

$$Con.Passive_i = \frac{1}{m} \frac{1}{n-1} \sum_{\substack{j=1 \\ j \neq i}}^n \sum_{k=1}^m P_{jk} \frac{1}{\exp(-d_{ij})} \quad (\text{Eq. A3})$$

To illustrate the fundamental difference, we used two hypothetical metacommunities (Figure A2) to quantify the connectivity and showed the difference between active and passive dispersers. In this figure, two metacommunities are identical (all patches have the same species), with the only exception that the distance between patches are different. For active dispersers, the estimated connectivity of the metacommunity a is larger than that of metacommunity b; this is obvious, because for active dispersers, the longer distance represents larger resistance of colonization (Eq. A2). By contrast, for passive dispersers, the estimated connectivity of metacommunity b is larger than that of metacommunity a, because for the patches located far apart to have the same species, stronger physical dispersal strength is required to overcome the longer distance (larger resistance) (Eq. A3). As can be seen in Figure A2, even for the identical community settings, estimation of connectivity of metacommunity is fundamentally different between active versus passive dispersers.

Supplementary references:

- Hanski, I. 1994. A practical model of metapopulation dynamics. — *J. Anim. Ecol.* 63: 151-162.
- Henriques-Silva, R. et al. 2012. A community of metacommunities: exploring patterns in species distributions across large geographical areas. — *Ecology* 94: 627-639.
- Kindlmann, P. and Burel, F. 2008. Connectivity measures: a review. — *Landscape Ecol.* 23: 879-890.



Connectivity for each patch

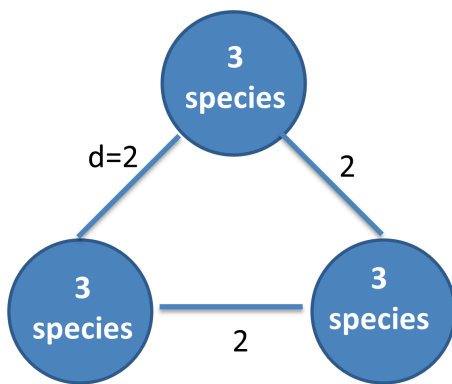
$$Con.Active_i = \frac{1}{3} \times \frac{1}{2} \times \exp(-1) \times 3 \times 2 \cong 0.368$$

$$Con.Passive_i = \frac{1}{3} \times \frac{1}{2} \times \frac{1}{\exp(-1)} \times 3 \times 2 \cong 2.718$$

Average connectivity for metacommunity

$$Avg.Con.Active = \frac{1}{3} \times 3 \times 0.368 = 0.368$$

$$Avg.Con.Passive = \frac{1}{3} \times 3 \times 2.718 = 2.718$$



Connectivity for each patch

$$Con.Active_i = \frac{1}{3} \times \frac{1}{2} \times \exp(-2) \times 3 \times 2 \cong 0.135$$

$$Con.Passive_i = \frac{1}{3} \times \frac{1}{2} \times \frac{1}{\exp(-2)} \times 3 \times 2 \cong 7.389$$

Average connectivity for metacommunity

$$Avg.Con.Active = \frac{1}{3} \times 3 \times 0.135 = 0.135$$

$$Avg.Con.Passive = \frac{1}{3} \times 3 \times 7.389 = 7.389$$

Figure A2. Illustration of the estimation of connectivity for different dispersal modes. d represents the distance between patches.

Supplementary material Appendix 3. Analysis of metacommunity theory based on presence/absence of taxa without considering phylogenetic information

In traditional RDA based on presence/absence of taxa, each taxon was considered as independent. This is apparently a questionable assumption (Lozupone and Knight 2008). By contrast, our variation partitioning approach incorporated the phylogenetic information; that is, the contribution of each taxon to the β -diversity was adjusted according to their phylogenetic position. For comparing the difference between the traditional RDA and our phylogenetically based method, we also do the RDA based on presence/absence of taxa assuming that the taxa are independent. The results showed that the explainable variance of traditional RDA was generally poorer than that of phylogenetically based method (comparing Table A3 with Table 2). The conclusion based on traditional RDA is generally consistent with that based on phylogenetically based method in May and August. However in December, the results of traditional RDA exhibited different patterns from that of phylogenetically based method, indicating that the conclusion sometimes differs dramatically if phylogenetic information is not taken into consideration.

Supplementary references

Lozupone, C. A. and Knight, R. 2008. Species divergence and the measurement of microbial diversity. — *FEMS Microbiol. Rev.* 32: 557-578.

Table A3. Results of variation partitioning against the environmental ([E]), linear spatial ([S1]), and non-linear spatial ([S2]) components based on the presence/absence of taxa without considering phylogenetic information.

	[E]		[S1]		[S2]		[E S1+S2]		[S1 E+S2]		[S2 E+S1]	
	Var.	P	Var.	P	Var.	P	Var.	P	Var.	P	Var.	P
Whole community												
Dec	6.9	0.01	8.5	0.04	5.8	0.02	-	-	-	-	-	-
May	4.6	0.05	7.4	0.04	7.8	0.01	1.6	0.48	0.9	0.48	1.5	0.46
Aug	1.7	0.32	-2.2	0.55	1.9	0.24	-1.3	0.53	-1.8	0.54	-1.9	0.52
Cyanobacteria												
Dec	10.6	0.12	8.5	0.28	10.3	0.16	-	-	-	-	-	-
May	12.7	0.03	17.9	0.06	10.2	0.05	2.4	0.45	3.9	0.42	2.4	0.44
Aug	-1.1	0.66	5.2	0.11	1.6	0.26	1.9	0.49	7.6	0.35	3.6	0.44
Alphaproteobacteria												
Dec	3.7	0.17	0.6	0.48	4.6	0.12	-	-	-	-	-	-
May	9.3	0.02	12.1	0.02	12.8	0.02	-4.4	0.60	1.8	0.47	5.0	0.40
Aug	0.0	0.51	-0.3	0.52	3.0	0.18	2.2	0.49	-3.4	0.57	-0.8	0.51
Gammaproteobacteria												
Dec	12.1	0.04	7.9	0.20	12.8	0.04	-	-	-	-	-	-
May	6.5	0.04	6.3	0.12	9.7	0.02	-4.7	0.60	-8.5	0.68	-3.8	0.55
Aug	6.9	0.04	2.4	0.33	1.4	0.33	1.3	0.50	-1.0	0.52	1.7	0.47
Actinobacteria												
Dec	25.9	0.06	25.5	0.10	27.0	0.08	-	-	-	-	-	-
May	9.0	0.07	9.5	0.15	12.4	0.04	-5.3	0.59	-4.7	0.58	-0.2	0.50
Aug	2.6	0.25	1.7	0.39	2.3	0.28	-6.5	0.61	-4.8	0.60	-1.6	0.52

Var. represents the average explainable variance from RDA. P represents the average P-value. Bold values indicate the results with high bootstrap support; that is, more than 60 out of 100 subsamplings exhibit a significant result ($P < 0.05$).

Supplementary material Appendix 4. Analysis of metacommunity theory based on relative abundance of taxa

Weighted UniFrac distance takes the relative abundance of taxa into consideration; therefore, the community phylogenetic distance is mainly determined by the abundant taxa. Analysis considering weighted UniFrac distance examines the mechanisms influencing number of individuals arriving at habitat patches, which differs from the concept of metacommunity theory. Here, we conducted supplementary analysis for weighted UniFrac distance to investigate the relative importance of environmental and dispersal processes in shaping the relative abundance distribution.

When examining weighted β -diversity, the effects of environmental ([E]), linear spatial ([S1]) and non-linear spatial ([S2]) components were all important at low (December) and intermediate (May) connectivity (non-linear spatial effect in December was marginally significant at $p=0.06$), but the effects became non-significant at high connectivity (August). Note that the unique effects were non-significant.

When examining variation partitioning for the four taxonomic groups, almost all groups exhibited consistent responses between low (December) and intermediate (May) connectivity. However, the communities under high connectivity (August) did not exhibit any significant effect (Table A4). The communities composition of *Cyanobacteria* was explained by the environmental ([E]) and linear spatial ([S1]) predictors, but the unique effects were both non-significant. For the communities composition of *Alphaproteobacteria* and *Actinobacteria*, there was no significant effect. In addition, the communities composition of *Gammaproteobacteria* was significantly explained by the non-linear spatial ([S2]) component at low connectivity (December).

Additionally, the temporal variation of shaping forces was only found in unweighted results, but not in weighted results. The weighted results showed a consistent pattern across sampling months for each taxonomic group, and the reason of this difference has been mentioned above. When concerning unweighted results, in May, almost the four taxonomic groups were significantly explained by the environmental component (except for *Actinobacteria*) (Table 2). That is because the connectivity was high enough (intermediate) to let most of the taxa migrate to all of the patches, and then the local environmental conditions would determine the community composition, suggesting that all dominant taxonomic groups may follow the species sorting model at intermediate connectivity in May. However, in December, during which connectivity was low, only *Gammaproteobacteria* were marginally

explained by non-linear spatial component (Table 2), indicating that the differential driving forces among taxonomic groups can be emphasized at low connectivity.

Table A4. Results of variation partitioning against the environmental ([E]), linear spatial ([S1]), and non-linear spatial ([S2]) components based on the weighted UniFrac distance

	[E]		[S1]		[S2]		[E S1+S2]		[S1 E+S2]		[S2 E+S1]	
	Var.	P	Var.	P	Var.	P	Var.	P	Var.	P	Var.	P
Whole community												
Dec	62.6	0.03	72.4	0.03	66.2	0.06	27.5	0.61	40.1	0.61	25.1	0.60
May	48.2	0.003	49.1	0.01	36.1	0.02	9.7	0.35	2.0	0.65	-3.0	0.75
Aug	13.0	0.27	32.3	0.19	7.6	0.31	-	-	-	-	-	-
Cyanobacteria												
Dec	95.2	0.03	93.7	0.02	98.8	0.06	0.3	0.54	1.3	0.54	1.7	0.55
May	68.8	0.004	58.8	0.04	54.1	0.05	12.7	0.30	-3.1	0.68	-2.8	0.61
Aug	7.9	0.37	39.9	0.16	4.1	0.46	-	-	-	-	-	-
Alphaproteobacteria												
Dec	-13.9	0.98	-10.8	0.79	-8.4	0.87	-	-	-	-	-	-
May	1.0	0.51	11.9	0.33	9.2	0.23	-	-	-	-	-	-
Aug	-6.8	0.69	3.9	0.50	-6.8	0.64	-	-	-	-	-	-
Gammaproteobacteria												
Dec	51.7	0.06	36.5	0.24	53.9	0.01	17.2	0.60	52.0	0.59	20.1	0.60
May	13.8	0.18	22.6	0.21	14.8	0.10	-	-	-	-	-	-
Aug	48.4	0.03	38.9	0.14	21.3	0.14	11.0	0.34	0.3	0.62	-3.0	0.65
Actinobacteria												
Dec	59.4	0.08	85.8	0.05	63.5	0.12	-	-	-	-	-	-
May	18.4	0.16	12.5	0.37	34.5	0.06	-	-	-	-	-	-
Aug	6.4	0.33	5.2	0.51	4.1	0.40	-	-	-	-	-	-

Var. represents the average explainable variance from 3-way PERMANOVA. P represents the average P-value. Bold values indicate the results with high bootstrap support; that is, more than 60 out of 100 subsamplings exhibit a significant result ($P < 0.05$).

Supplementary material Appendix 5. Analysis of metacommunity theory with minimum sequence sample size (12000 sequences).

When concerning the results of variation partitioning based on minimum sequence sample size (12000) (Table A5), the conclusion on the environmental and spatial effects on metacommunity across three months remains qualitatively similar with the results presented in Table 2; the significances of environmental and spatial effects were found in May, but not found in December and August. Moreover, the explanatory power of environmental and spatial effects across three months was consistent with the result presented in Table 2, except for December, in which the significantly environmental effect disappeared. In May, the communities composition of *Alphaproteobacteria* and *Gammaproteobacteria* were significantly explained by the environmental component (marginally for *Gammaproteobacteria* at $p=0.05$ and high bootstrap support). However, the communities composition of *Cyanobacteria* and *Actinobacteria* cannot be explained by either component. In December and August, no statistics were significant. The pattern of variation partitioning across seasons remains the same, but some statistics became non-significant because of small sample size.

Table A5. Results of variation partitioning with minimum sequence sample size (12000 sequences)

	[E]		[S1]		[S2]		[E S1+S2]		[S1 E+S2]		[S2 E+S1]	
	Var.	P	Var.	P	Var.	P	Var.	P	Var.	P	Var.	P
Whole community												
Dec	14.35	0.07	23.80	0.16	51.50	0.65	-	-	-	-	-	-
May	14.83	0.01	18.12	0.02	8.86	0.42	10.79	0.32	18.04	0.30	14.83	0.01
Aug	1.9	0.56	6.3	0.50	5.3	0.20	-	-	-	-	-	-
Cyanobacteria												
Dec	25.96	0.23	30.06	0.35	45.58	0.58	-	-	-	-	-	-
May	14.36	0.20	19.91	0.22	1.56	0.61	-	-	-	-	-	-
Aug	0.3	0.64	8.4	0.43	5.2	0.36	-	-	-	-	-	-
Alphaproteobacteria												
Dec	7.74	0.45	19.34	0.44	52.55	0.50	-	-	-	-	-	-
May	9.84	0.04	6.45	0.20	0.88	0.39	3.73	0.33	-0.29	0.44	9.84	0.04
Aug	2.9	0.50	7.8	0.46	7.4	0.29	-	-	-	-	-	-
Gammaproteobacteria												
Dec	14.03	0.27	22.43	0.38	52.91	0.64	-	-	-	-	-	-
May	16.41	0.05	18.29	0.16	0.16	0.75	13.21	0.37	8.70	0.63	16.41	0.05
Aug	12.1	0.10	15.9	0.20	4.4	0.40	-	-	-	-	-	-
Actinobacteria												
Dec	10.85	0.37	22.40	0.42	40.57	0.59	-	-	-	-	-	-
May	14.07	0.16	30.99	0.13	4.90	0.53	-	-	-	-	-	-
Aug	-0.4	0.59	9.3	0.43	1.8	0.50	-	-	-	-	-	-

Var. represents the average explainable variance from 3-way PERMANOVA. P represents the average P-value. Bold values indicate the results with high bootstrap support; that is, more than 60 out of 100 subsamplings exhibit a significant result (P<0.05).