

Multivariate Cutoff Level Analysis (MultiCoLA) of Large Community Datasets. Gobet A., Quince C. & Ramette A.

Supplementary Tables:

Supplementary Table 1. Contribution of environmental parameters to the variation in truncated datasets (dataset-based approach), at the OTU level for all sequences available.

Supplementary Table 2. Contribution of environmental parameters to the variation in truncated datasets (dataset-based approach), at the OTU level for the dataset with a complete annotation.

Supplementary Table 3. Contribution of environmental parameters to the variation in truncated datasets (dataset-based approach), at the Genus level.

Supplementary Table 4. Contribution of environmental parameters to the variation in truncated datasets (dataset-based approach), at the Family level.

Supplementary Table 5. Contribution of environmental parameters to the variation in truncated datasets (dataset-based approach), at the Order level.

Supplementary Table 6. Contribution of environmental parameters to the variation in truncated datasets (dataset-based approach), at the Class level.

Supplementary Table 7. Contribution of environmental parameters to the variation in truncated datasets (dataset-based approach), at the Phylum level.

Supplementary Table 8. Contribution of environmental parameters to the variation in truncated datasets (sample-based approach), at the OTU level for all sequences available.

Supplementary Table 9. Contribution of environmental parameters to the variation in truncated datasets (sample-based approach), at the OTU level for the dataset with a complete annotation.

Supplementary Table 10. Contribution of environmental parameters to the variation in truncated datasets (sample-based approach), at the Genus level.

Supplementary Table 11. Contribution of environmental parameters to the variation in truncated datasets (sample-based approach), at the Family level.

Supplementary Table 12. Contribution of environmental parameters to the variation in truncated datasets (sample-based approach), at the Order level.

Supplementary Table 13. Contribution of environmental parameters to the variation in truncated datasets (sample-based approach), at the Class level.

Supplementary Table 14. Contribution of environmental parameters to the variation in truncated datasets (sample-based approach), at the Phylum level.

Supplementary Table 15. Summary of OTUs numbers after PyroNoise correction of the 454 MPTS dataset.

Supplementary Table 1. Contribution of environmental parameters to the variation in truncated datasets (dataset-based approach), **at the OTU level for all sequences available.**

Cutoff levels ^a	Total number of sequences	R ² , ^b	Individual factor contribution ^c		
			Cell abundance	Nutrients	
				SiO ₂	NH ₄
0%	197,684	17% ***	-0.97		
1%	195,707	17% ***	-0.97		
5%	187,799	19% ***	-0.97		
10%	177,915	23% ***	-0.97	0.42	
15%	168,029	27% ***	-0.96	0.42	
20%	158,143	31% ***	-0.96	0.42	
25%	148,258	34% ***	-0.96	0.43	
30%	138,377	38% ***	-0.96	0.43	
35%	128,459	40% ***	-0.97	0.43	
40%	118,590	42% ***	-0.97	0.43	
45%	108,709	44% ***	-0.97	0.43	
50%	98,762	44% ***	-0.97	0.41	
55%	88,871	45% ***	-0.97	0.38	
60%	78,951	45% ***	-0.98		
65%	68,789	41% ***	-0.97	0.35	
70%	58,637	39% ***	0.95	-0.26	
75%	49,110	22% **	0.96		
80%	38,299	NS			
85%	25,961	NS			
90%	16,852	NS			
95%	6,550				
99%	0				

^aCutoff levels were defined based on the whole dataset strategy (see Supplementary Fig. 1). Cutoff levels were applied until samples were lost due to lack of sequences.

^bAdjusted R² indicate the amount of variation explained by cell abundance and nutrients (SiO₂, silicate; NH₄, ammonium), their significance is indicated as NS (non significant), * (P ≤ 0.05), ** (P ≤ 0.01), and *** (P ≤ 0.001).

^cOnly significant, standardized correlation coefficients to the first redundancy analysis (RDA) axis are indicated for each parameter.

Supplementary Table 2. Contribution of environmental parameters to the variation in truncated datasets (dataset-based approach), **at the OTU level for the dataset with a complete annotation.**

Cutoff levels ^a	Total number of sequences	R ² , ^b	Individual factor contribution ^c		
			Cell abundance	Nutrients	
				SiO ₂	NH ₄
0%	40,660	16% ***	-0.96		
1%	40,253	16% ***	-0.96	0.41	
5%	38,627	18% ***	-0.96	0.41	
10%	36,593	21% ***	-0.96	0.41	
15%	34,560	25% ***	-0.96	0.4	
20%	32,524	29% ***	-0.96	0.4	
25%	30,489	32% ***	-0.96	0.41	
30%	28,453	35% ***	-0.96	0.44	
35%	26,391	38% ***	-0.96	0.46	
40%	24,371	38% ***	-0.97	0.43	
45%	22,315	40% ***	-0.97	0.44	
50%	20,267	41% ***	-0.97	0.39	
55%	18,234	39% ***	-0.96	0.31	
60%	16,106	38% ***	-0.89	0.1	
65%	13,731	31% **	-0.6	-0.28	
70%	11,332	41% ***	-0.29	-0.58	
75%	8,884	54% ***	-0.35	-0.55	
80%	7,181	58% **	-0.54	-0.39	
85%	5,165				
90%	0				
95%	0				
99%	0				

For detailed explanations, see Supplementary Table 1.

Supplementary Table 3. Contribution of environmental parameters to the variation in truncated datasets (dataset-based approach), **at the Genus level**.

Cutoff levels ^a	Total number of sequences	R ² , ^b	Individual factor contribution ^c		
			Cell abundance	Nutrients	
				SiO ₂	NH ₄
0%	40,660	38% ***	-0.96	0.43	
1%	40,251	41% ***	-0.96	0.43	
5%	38,600	48% ***	-0.96	0.44	
10%	36,569	52% ***	-0.96	0.44	0.58
15%	34,549	53% ***	-0.97	0.42	
20%	32,374	58% ***	-0.97	0.41	
25%	30,336	59% ***	-0.97	0.42	
30%	28,367	61% ***	-0.97	0.42	
35%	25,952	62% ***	-0.97	0.4	
40%	23,612	60% ***	-0.97	0.39	
45%	21,013	49% ***	-0.97	0.34	
50%	19,464	39% ***	-0.99	0.41	
55%	17,702	43% ***	-0.95	0.65	
60%	14,982	36% *	-0.42	-0.51	
65%	12,129	NS			
70%	12,129	NS			
75%	7,029				
80%	7,029				
85%	0				
90%	0				
95%	0				
99%	0				

For detailed explanations, see Supplementary Table 1.

Supplementary Table 4. Contribution of environmental parameters to the variation in truncated datasets (dataset-based approach), **at the Family level.**

Cutoff levels ^a	Total number of sequences	R ² , ^b	Individual factor contribution ^c	
			Cell abundance	Nutrients SiO ₂ NH ₄
0%	40,660	41% ***	-0.97	0.47
1%	40,246	44% ***	-0.97	0.47
5%	38,555	50% ***	-0.97	0.48
10%	36,534	53% ***	-0.97	0.48
15%	34,494	53% ***	-0.98	0.5
20%	32,441	53% ***	-0.98	0.49
25%	30,201	59% ***	-0.98	0.47
30%	28,455	61% ***	0.97	-0.46
35%	26,140	56% ***	0.97	-0.59
40%	24,378	58% ***	0.93	-0.65
45%	22,143	59% ***	0.96	-0.55
50%	19,290	62% ***	0.96	
55%	15,089	NS		
60%	15,089	NS		
65%	8,059			
70%	8,059			
75%	8,059			
80%	8,059			
85%	0			
90%	0			
95%	0			
99%	0			

For detailed explanations, see Supplementary Table 1.

Supplementary Table 5. Contribution of environmental parameters to the variation in truncated datasets (dataset-based approach), **at the Order level**.

Cutoff levels ^a	Total number of sequences	R ² , ^b	Individual factor contribution ^c		
			Cell abundance	Nutrients	
				SiO ₂	NH ₄
0%	40,660	47% ***	-0.97	0.46	
1%	40,217	50% ***	-0.97	0.47	
5%	38,446	54% ***	-0.98	0.5	
10%	36,412	54% ***	-0.98	0.49	
15%	34,218	58% ***	-0.98	0.49	
20%	32,241	59% ***	0.98	-0.47	
25%	29,586	63% ***	0.97	-0.48	
30%	28,379	64% ***	-0.97	0.51	
35%	24,963	60% ***	-0.91	0.66	
40%	22,110	63% **	0.94		
45%	22,110	63% **	0.94		
50%	15,089	NS			
55%	15,089	NS			
60%	15,089	NS			
65%	8,059				
70%	8,059				
75%	8,059				
80%	8,059				
85%	0				
90%	0				
95%	0				
99%	0				

For detailed explanations, see Supplementary Table 1.

Supplementary Table 6. Contribution of environmental parameters to the variation in truncated datasets (dataset-based approach), **at the Class level**.

Cutoff levels ^a	Total number of sequences	R ² , ^b	Individual factor contribution ^c		
			Cell abundance	Nutrients	
				SiO ₂	NH ₄
0%	40,660	55%***	-0.97		
1%	40,127	56%***	-0.97		
5%	38,297	58%***	-0.97		
10%	36,176	64%***	-0.97		
15%	34,013	66%***	-0.97		
20%	30,704	65%***	-0.96		
25%	28,065	71%***	-0.94		
30%	28,065	71%***	-0.94		
35%	20,006	78%***	-0.94		
40%	20,006	78%***	-0.94		
45%	20,006	78%***	-0.94		
50%	20,006	78%***	-0.94		
55%	11,937				
60%	11,937				
65%	11,937				
70%	11,937				
75%	0				
80%	0				
85%	0				
90%	0				
95%	0				
99%	0				

For detailed explanations, see Supplementary Table 1.

Supplementary Table 7. Contribution of environmental parameters to the variation in truncated datasets (dataset-based approach), **at the Phylum level.**

Cutoff levels ^a	Total number of sequences	R ² , ^b	Individual factor contribution ^c	
			Cell abundance	Nutrients SiO ₂ NH ₄
0%	40,660	60% ***	-0.97	
1%	40,045	60% ***	-0.97	
5%	38,600	62% ***	-0.97	
10%	36,242	65% ***	-0.97	
15%	33,575	67% ***	-0.93	
20%	25,516	73% ***	-0.93	
25%	25,516	73% ***	-0.93	
30%	25,516	73% ***	-0.93	
35%	25,516	73% ***	-0.93	
40%	13,579			
45%	13,579			
50%	13,579			
55%	13,579			
60%	13,579			
65%	13,579			
70%	0			
75%	0			
80%	0			
85%	0			
90%	0			
95%	0			
99%	0			

For detailed explanations, see Supplementary Table 1.

Supplementary Table 8. Contribution of environmental parameters to the variation in truncated datasets (sample-based approach), **at the OTU level for all sequences available.**

Cutoff levels ^a	Total number of sequences	R ² , ^b	Individual factor contribution ^c		
			Cell abundance	Nutrients	
				SiO ₂	NH ₄
1	197,684	17% ***	-0.97		
2	162,639	23% ***	-0.97		
3	149,499	25% ***	-0.97	0.43	
5	134,258	27% ***	-0.97	0.43	
10	113,172	27% ***	-0.98		
15	101,666	27% ***	-0.98		
20	94,126	27% ***	-0.97	0.38	
30	84,180	28% ***	-0.96	0.30	
55	70,118	28% ***	-0.97	0.36	
80	60,612	20% **	-0.97		
105	54,317	16% *	-0.98		
130	49,187	13% *			
155	44,384	NS			
180	39,872	NS			
208	33,931	NS			

^aCutoff levels were defined based on the sample-based strategy (see Supplementary Fig. 1). Cutoff levels were applied until samples were lost due to lack of sequences.

^bAdjusted R² indicate the amount of variation explained by cell abundance and nutrients (SiO₂, silicate; NH₄, ammonium), their significance is indicated as NS (non significant), * (P ≤ 0.05), ** (P ≤ 0.01), and *** (P ≤ 0.001).

^cOnly significant, standardized correlation coefficients to the first redundancy analysis (RDA) axis are indicated for each parameter.

Supplementary Table 9. Contribution of environmental parameters to the variation in truncated datasets (sample-based approach), **at the OTU level for the dataset with a complete annotation.**

Cutoff levels ^a	Total number of sequences	R ² ^b	Individual factor contribution ^c		
			Cell abundance	Nutrients	
				SiO ₂	NH ₄
1	40,660	16% ***	-0.96	0.41	
2	33,533	22% ***	-0.96	0.42	
3	30,655	25% ***	-0.97	0.44	
5	27,352	25% ***	-0.97	0.48	
10	22,164	23% ***	-0.98		
15	19,591	25% ***	-0.98	0.39	
20	18,010	24% ***	-0.97		
30	16,113	23% ***	-0.76	-0.13	
55	13,314	17% **		-0.80	
80	10,761				
105	9,256				
130	8,331				
155	7,491				
180	6,982				
208	5,436				

For detailed explanations, see Supplementary Table 8.

Supplementary Table 10. Contribution of environmental parameters to the variation in truncated datasets (sample-based approach), **at the Genus level**.

Cutoff levels ^a	Total number of sequences	R ² ^b	Individual factor contribution ^c		
			Cell abundance	Nutrients	
				SiO ₂	NH ₄
1	40,660	38% ***	-0.96	0.43	
2	39,800	40% ***	-0.96	0.42	0.59
3	39,114	41% ***	-0.95	0.43	0.61
5	37,979	42% ***	-0.95	0.43	0.63
10	35,585	36% ***	-0.94		
15	33,526	38% ***	-0.95	0.41	
20	31,993	40% ***	-0.96	0.42	
30	29,405	44% ***	-0.95	0.36	
55	25,952	46% ***	0.95	-0.31	
80	24,252	38% ***	0.98	-0.37	
105	21,847	43% ***	-0.99	0.54	
130	20,689				
155	19,254				
180	18,239				
208	16,501				

For detailed explanations, see Supplementary Table 8.

Supplementary Table 11. Contribution of environmental parameters to the variation in truncated datasets (sample-based approach), **at the Family level.**

Cutoff levels ^a	Total number of sequences	R ² ^b	Individual factor contribution ^c		
			Cell abundance	Nutrients	
				SiO ₂	NH ₄
1	40,660	41% ***	-0.97	0.47	
2	40,233	42% ***	-0.96	0.45	
3	39,841	42% ***	-0.96	0.45	
5	39,087	42% ***	-0.96	0.44	0.60
10	37,588	39% ***	-0.94	0.42	
15	36,328	40% ***	-0.96	0.50	
20	35,245	38% ***	-0.97	0.44	
30	33,077	41% ***	-0.97	0.42	
55	29,427	39% ***	-0.97	0.37	
80	27,830	37% ***	-0.97	0.45	
105	26,303	49% ***	-0.93	0.73	
130	25,053	45% ***	-0.94	0.63	
155	24,346	40% ***	-0.96	0.57	
180	23,357	38% ***	-0.96	0.44	
208	21,827				

For detailed explanations, see Supplementary Table 8.

Supplementary Table 12. Contribution of environmental parameters to the variation in truncated datasets (sample-based approach), **at the Order level**.

Cutoff levels ^a	Total number of sequences	R ² , ^b	Individual factor contribution ^c		
			Cell abundance	Nutrients	
				SiO ₂	NH ₄
1	40,660	47% ***	-0.97	0.46	
2	40,526	47% ***	-0.97	0.45	
3	40,342	47% ***	-0.96	0.44	
5	39,992	48% ***	-0.96	0.42	0.60
10	39,237	45% ***	-0.95	0.39	
15	38,496	43% ***	-0.96	0.44	
20	37,580	38% ***	-0.96	0.40	
30	35,878	41% ***	-0.96	0.39	
55	32,563	37% ***	-0.99	0.44	
80	29,987	35% ***	-0.98		
105	27,960	41% ***	-0.99	0.62	
130	27,358	44% ***	-0.98	0.68	
155	26,382	34% ***	-0.91	0.58	
180	25,541	35% **	-0.94	0.54	
208	24,372	42% ***	-0.94		

For detailed explanations, see Supplementary Table 8.

Supplementary Table 13. Contribution of environmental parameters to the variation in truncated datasets (sample-based approach), **at the Class level**.

Cutoff levels ^a	Total number of sequences	R ² , ^b	Individual factor contribution ^c		
			Cell abundance	Nutrients	
				SiO ₂	NH ₄
1	40,660	55% ***	-0.97		
2	40,636	55% ***	-0.97		
3	40,594	54% ***	-0.96		0.62
5	40,495	54% ***	-0.96		0.61
10	40,144	53% ***	-0.96		0.61
15	39,753	48% ***	-0.95		
20	39,172	42% ***	-0.95		
30	37,932	48% ***	-0.96	0.48	
55	36,156	39% ***	-0.97		
80	34,338	31% **	-0.99		
105	33,168	36% ***	-0.99		
130	32,351	39% **	-0.99		
155	30,658	50% ***	-0.98		
180	30,503	46% ***	-0.97		
208	29,564	34% *			

For detailed explanations, see Supplementary Table 8.

Supplementary Table 14. Contribution of environmental parameters to the variation in truncated datasets (sample-based approach), **at the Phylum level**.

Cutoff levels ^a	Total number of sequences	R ² ^b	Individual factor contribution ^c	
			Cell abundance	Nutrients SiO ₂ NH ₄
1	40,660	60% ***	-0.97	
2	40,644	60% ***	-0.97	
3	40,614	59% ***	-0.97	
5	40,544	58% ***	-0.97	
10	40,358	59% ***	-0.96	0.62
15	40,167	55% ***	-0.96	
20	39,901	49% ***	-0.96	
30	39,285	53% ***	-0.96	
55	37,868	40% ***	-0.96	
80	36,601	39% ***	-0.99	
105	35,579	44% ***	-0.99	
130	34,876	50% ***	-0.99	
155	34,745	45% ***	-0.99	
180	34,580	48% ***	-0.99	
208	33,993	NS		

For detailed explanations, see Supplementary Table 8.

Supplementary Table 15. Summary of OTUs numbers after PyroNoise correction of the 454 MPTS dataset.

Sampling date	Depth layer (cm)	Read number	Filtered number	Unique sequences	3% OTUs	5% OTUs	10% OTUs		
2005	February	0-5	9,526	4,553	784	763	692	591	
		5-10	18,409	8,732	1,639	1,568	1,385	1,102	
		10-15	7,971	3,364	1,207	1,179	1,061	876	
	April	0-5	5,146	2,162	505	496	470	407	
		5-10	10,733	5,740	1,487	1,436	1,314	1,113	
		10-15	15,259	8,862	2,476	2,341	2,083	1,650	
	July	0-5	10,690	5,970	1,068	1,036	927	763	
		5-10	9,632	4,940	1,402	1,373	1,265	1,063	
	November	0-5	14,526	8,950	1,539	1,487	1,326	1,083	
		5-10	21,996	12,981	2,469	2,301	2,072	1,618	
	2006	March1	0-5	10,439	5,839	1,492	1,442	1,324	1,107
			5-10	20,182	11,463	2,643	2,473	2,181	1,737
10-15			19,564	9,981	2,586	2,426	2,154	1,703	
March2		0-5	11,339	5,956	1,035	994	904	772	
		5-10	20,866	11,941	2,133	2,036	1,808	1,467	
		10-15	21,470	11,997	3,210	2,993	2,586	2,008	