

**Supplementary Tables**

Table S1. Bioinformatic data of each dataset.

Dataset	# of samples	# of samples used	# of seqs raw	# of seqs preprocessed	Mean read length	# of OTUs	# of OTUs annotated: phylum level	# of OTUs annotated: family level
Río de la Plata	30	29	1,475,757	632,565	359.11	36,189	35,421	34,689
Delaware	128	120	1,114,734	412,850	365.69	33,139	32,681	31,217
Krka	28	28	380,107	224,062	414.1	10,194	10,171	9,802
Pearl	17	16	106,586	35,189	451.72	5,180	5,159	4,916

Table S2. Bacterial indicators of the four environmental groups defined for the Río de la Plata estuary, when OTUs were defined at the 99% level of similarity. The candidate species are those who meet the criterion of being present in at least 50% of the samples composing each group. A and B denote the IndVal components of specificity and sensitivity, respectively. The indicator OTUs are the combination of candidate OTUs that were selected as the best indicators of each group among all possible significant combinations ( $p < 0.05$ ).

	n candidate species	IndVal <sup>a</sup>		Indicator OTUs and their taxonomic affiliation
		A	B	
G1	55	1 1	0.88 0.63	OTU18710 (Flavobacteriaceae) + OTU58010 (SAR11) OTU26066 (Actinomarinaceae) + OTU18710 (Flavobacteriaceae) + OTU28349 (NS4)
G2	92	1	1	OTU75750 (SAR11) + OTU58056 (!Ketogulonicigenium)
G3	83	1 1	.0.89 0.89	OTU55318 (SAR11) + OTU25119 (Verrucomicrobiae) OTU55318 (SAR11) + OTU76715 (Verrucomicrobiae)
G4	63	1 1	0.83 0.67	OTU12330 (SAR86) + OTU75566 (SAR11) + OTU76034 (Planktomarina) OTU57973 (SAR11) + OTU65826 (Synechococcus) + OTU75522 (SAR11)

<sup>a</sup>All indicators achieved 100% of group coverage.

Table S3. True group membership and predictions of sample assignment to each one of the three groups of the Río de la Plata based on the bacterial indicators (OTUs at 99% of similarity). The predictions were run using leave one out validation, the probability of the assignment of each sample by each indicator is given.

Sample	p indicator G1	p indicator G2	p indicator G3	p indicator G4	True Group	Assigned to
RdlP_1	0	0	1	0	G3	G3
RdlP_2	1	0	0	0	G1	G1
RdlP_3	0	0	0	1	G4	G4
RdlP_4	0	0	1	0	G3	G3
RdlP_5	1	0	0	0	G1	G1
RdlP_6	0	0	0	1	G4	G4
RdlP_8	0	1	0	0	G2	G2
RdlP_9	0	0	0	1	G4	G4
RdlP_10	0	0	1	0	G3	G3
RdlP_11	0	1	0	0	G2	G2
RdlP_12	0	0	0	1	G4	G4
RdlP_13	0	0	1	0	G3	G3
RdlP_14	0	1	0	0	G2	G2
RdlP_15	0	0	0	1	G4	G4
RdlP_16	0	0	1	0	G3	G3
RdlP_17	0	1	0	0	G2	G2
RdlP_18	0	0	0	1	G4	G4
RdlP_19	0	0	1	0	G3	G3
RdlP_20	0	1	0	0	G2	G2
RdlP_21	1	0	0	0	G1	G1
RdlP_22	0	0	1	0	G3	G3
RdlP_23	0	1	0	0	G2	G2
RdlP_24	1	0	0	0	G1	G1
RdlP_25	0	0	1	0	G3	G3
RdlP_26	1	0	0	0	G1	G1
RdlP_27	1	0	0	0	G1	G1
RdlP_28	0	0	1	0	G3	G3
RdlP_29	1	0	0	0	G1	G1
RdlP_30	1	0	0	0	G1	G1

Table S4. Non-parametric pairwise comparison of the physicochemical characteristics of the four estuaries. The Euclidean distance was used to calculate the dissimilarity matrix, Bonferroni correction was employed to account for multiple comparisons.

Pair	Df	Sums of Sqs	F Model	R <sup>2</sup>	p value	p adjusted for multiple comparisons
RdIP vs Dlwr	1	50720.607	18.410682	0.10616809	0.001	0.006*
RdIP vs Pearl	1	28504.227	18.172794	0.29707314	0.001	0.006*
RdIP vs Krka	1	7824.371	25.323187	0.31526622	0.001	0.006 *
Dlwr vs Pearl	1	7783.266	2.314249	0.01603618	0.107	0.642
Dlwr vs Krka	1	85352.838	30.774241	0.16655050	0.001	0.006 *
Pearl vs Krka	1	42346.713	26.329841	0.38533444	0.001	0.006 *

Table S5. True group membership and predictions of sample assignment to each one of the four estuaries based on the bacterial indicators found for each estuary in the combined dataset containing the BCC of all estuaries. The predictions were run using leave one out validation, the probability of the assignment of each sample by each indicator is given.

Sample	p indicator RdIP	p indicator Dlwr	p indicator Krka	p indicator Pearl	True estuary	Assigned to
RdIP_1	0.89	0	0	0	RdIP	RdIP
RdIP_2	0.89	0	0	0	RdIP	RdIP
RdIP_3	0.90	0	0	0	RdIP	RdIP
RdIP_4	0.90	0	0	0	RdIP	RdIP
RdIP_5	0.89	0	0	0	RdIP	RdIP
RdIP_6	0.90	0	0	0	RdIP	RdIP
RdIP_8	0.90	0	0	0	RdIP	RdIP
RdIP_9	0.85	0	0	0	RdIP	RdIP
RdIP_10	0.89	0	0	0	RdIP	RdIP
RdIP_11	0.90	0	0	0	RdIP	RdIP
RdIP_12	0.85	0	0	0	RdIP	RdIP
RdIP_13	0.90	0	0	0	RdIP	RdIP
RdIP_14	0.84	0	0	0	RdIP	RdIP
RdIP_15	0.85	0	0	0	RdIP	RdIP
RdIP_16	0.90	0	0	0	RdIP	RdIP
RdIP_18	0.85	0	0	0	RdIP	RdIP
RdIP_19	0.88	0	0	0	RdIP	RdIP
RdIP_20	0.85	0	0	0	RdIP	RdIP
RdIP_21	0.84	0	0	0	RdIP	RdIP
RdIP_22	0.89	0	0	0	RdIP	RdIP
RdIP_23	0.90	0	0	0	RdIP	RdIP
RdIP_24	0.85	0	0	0	RdIP	RdIP
RdIP_25	0.90	0	0	0	RdIP	RdIP
RdIP_26	0.90	0	0	0	RdIP	RdIP
RdIP_27	0.84	0	0	0	RdIP	RdIP
RdIP_28	0.90	0	0	0	RdIP	RdIP
RdIP_29	0.90	0	0	0	RdIP	RdIP
RdIP_30	0.85	0	0	0	RdIP	RdIP

Dlwr_1	0.91	0	0	0	Dlwr	RdIP
Dlwr_2	0	1	0	0	Dlwr	Dlwr
Dlwr_3	0	1	0	0	Dlwr	Dlwr
Dlwr_4	0	1	0	0	Dlwr	Dlwr
Dlwr_5	0	1	0	0	Dlwr	Dlwr
Dlwr_6	0	1	0	0	Dlwr	Dlwr
Dlwr_7	0	1	0	0	Dlwr	Dlwr
Dlwr_8	0	1	0	0	Dlwr	Dlwr
Dlwr_9	0	0	0	0	Dlwr	NA
Dlwr_10	0	1	0	0	Dlwr	Dlwr
Dlwr_11	0	1	0	0	Dlwr	Dlwr
Dlwr_12	0	1	0	0	Dlwr	Dlwr
Dlwr_13	0	1	0	0	Dlwr	Dlwr
Dlwr_14	0	1	0	0	Dlwr	Dlwr
Dlwr_15	0	1	0	0	Dlwr	Dlwr
Dlwr_16	0.91	1	0	0	Dlwr	Dlwr
Dlwr_17	0	1	0	0	Dlwr	Dlwr
Dlwr_18	0	1	0	0	Dlwr	Dlwr
Dlwr_19	0	1	0	0	Dlwr	Dlwr
Dlwr_20	0	0	0	0	Dlwr	NA
Dlwr_21	0	1	0	0	Dlwr	Dlwr
Dlwr_22	0	1	0	0	Dlwr	Dlwr
Dlwr_23	0	1	0	0	Dlwr	Dlwr
Dlwr_24	0	1	0	0	Dlwr	Dlwr
Dlwr_25	0	1	0	0	Dlwr	Dlwr
Dlwr_26	0.87	1	0	0	Dlwr	Dlwr
Dlwr_27	0	1	0	0	Dlwr	Dlwr
Dlwr_28	0	1	0	0	Dlwr	Dlwr
Dlwr_29	0.87	1	0	0	Dlwr	Dlwr
Dlwr_30	0	1	0	0	Dlwr	Dlwr
Dlwr_31	0	1	0	0	Dlwr	Dlwr
Dlwr_32	0	0	0	0	Dlwr	NA
Dlwr_33	0	1	0	0	Dlwr	Dlwr
Dlwr_34	0	1	0	0	Dlwr	Dlwr
Dlwr_35	0	1	0	0	Dlwr	Dlwr

Dlwr_36	0.87	1	0	0	Dlwr	Dlwr
Dlwr_37	0	1	0	0	Dlwr	Dlwr
Dlwr_38	0	1	0	0	Dlwr	Dlwr
Dlwr_39	0.91	0	0	0	Dlwr	RdIP
Dlwr_40	0	1	0	0	Dlwr	Dlwr
Dlwr_41	0.87	1	0	0	Dlwr	Dlwr
Dlwr_42	0	1	0	0	Dlwr	Dlwr
Dlwr_43	0	1	0	0	Dlwr	Dlwr
Dlwr_44	0	1	0	0	Dlwr	Dlwr
Dlwr_45	0.87	1	0	0	Dlwr	Dlwr
Dlwr_46	0	1	0	0	Dlwr	Dlwr
Dlwr_47	0	1	0	0	Dlwr	Dlwr
Dlwr_48	0	1	0	0	Dlwr	Dlwr
Dlwr_49	0	1	0	0	Dlwr	Dlwr
Dlwr_50	0	1	0	0	Dlwr	Dlwr
Dlwr_51	0	1	0	0	Dlwr	Dlwr
Dlwr_52	0.92	1	0	0	Dlwr	Dlwr
Dlwr_53	0	1	0	0	Dlwr	Dlwr
Dlwr_54	0	1	0	0	Dlwr	Dlwr
Dlwr_55	0	1	0	0	Dlwr	Dlwr
Dlwr_56	0	1	0	0	Dlwr	Dlwr
Dlwr_57	0	1	0	0	Dlwr	Dlwr
Dlwr_58	0.87	1	0	0	Dlwr	Dlwr
Dlwr_59	0	1	0	0	Dlwr	Dlwr
Dlwr_60	0	1	0	0	Dlwr	Dlwr
Dlwr_61	0	0	0	0	Dlwr	NA
Dlwr_62	0	1	0	0	Dlwr	Dlwr
Dlwr_63	0	1	0	0	Dlwr	Dlwr
Dlwr_64	0	1	0	0	Dlwr	Dlwr
Dlwr_65	0	1	0	0	Dlwr	Dlwr
Dlwr_66	0	1	0	0	Dlwr	Dlwr
Dlwr_67	0	1	0	0	Dlwr	Dlwr
Dlwr_68	0	1	0	0	Dlwr	Dlwr
Dlwr_69	0	1	0	0	Dlwr	Dlwr
Dlwr_70	0	0	0	0	Dlwr	NA

Dlwr_71	0	1	0	0	Dlwr	Dlwr
Dlwr_72	0.91	1	0	0	Dlwr	Dlwr
Dlwr_73	0	1	0	0	Dlwr	Dlwr
Dlwr_74	0	1	0	0	Dlwr	Dlwr
Dlwr_75	0	1	0	0	Dlwr	Dlwr
Dlwr_76	0	1	0	0	Dlwr	Dlwr
Dlwr_78	0	1	0	0	Dlwr	Dlwr
Dlwr_79	0	1	0	0	Dlwr	Dlwr
Dlwr_80	0	1	0	0	Dlwr	Dlwr
Dlwr_81	0	1	0	0	Dlwr	Dlwr
Dlwr_82	0	1	0	0	Dlwr	Dlwr
Dlwr_83	0	1	0	0	Dlwr	Dlwr
Dlwr_84	0	1	0	0	Dlwr	Dlwr
Dlwr_85	0	0	0	0	Dlwr	NA
Dlwr_86	0	1	0	0	Dlwr	Dlwr
Dlwr_87	0	1	0	0	Dlwr	Dlwr
Dlwr_88	0	1	0	0	Dlwr	Dlwr
Dlwr_89	0	1	0	0	Dlwr	Dlwr
Dlwr_90	0	1	0	0	Dlwr	Dlwr
Dlwr_91	0.90	1	0	0	Dlwr	Dlwr
Dlwr_92	0	1	0	0	Dlwr	Dlwr
Dlwr_93	0	1	0	0	Dlwr	Dlwr
Dlwr_94	0	1	0	0	Dlwr	Dlwr
Dlwr_95	0	0	0	0	Dlwr	NA
Dlwr_96	0	0	0	0	Dlwr	NA
Dlwr_97	0.93	0	0	0	Dlwr	RdIP
Dlwr_98	0	1	0	0	Dlwr	Dlwr
Dlwr_99	0	1	0	0	Dlwr	Dlwr
Dlwr_100	0	1	0	0	Dlwr	Dlwr
Dlwr_101	0	1	0	0	Dlwr	Dlwr
Dlwr_102	0.87	1	0	0	Dlwr	Dlwr
Dlwr_103	0	1	0	0	Dlwr	Dlwr
Dlwr_104	0	1	0	0	Dlwr	Dlwr
Dlwr_105	0.91	0	0	0	Dlwr	RdIP
Dlwr_106	0	1	0	0	Dlwr	Dlwr



Dlwr_107	0	1	0	0	Dlwr	Dlwr
Dlwr_108	0	1	0	0	Dlwr	Dlwr
Dlwr_109	0.87	1	0	0	Dlwr	Dlwr
Dlwr_110	0.87	1	0	0	Dlwr	Dlwr
Dlwr_111	0	0	0	0	Dlwr	NA
Dlwr_112	0	1	0	0	Dlwr	Dlwr
Dlwr_113	0	1	0	0	Dlwr	Dlwr
Dlwr_114	0	1	0	0	Dlwr	Dlwr
Dlwr_115	0	1	0	0	Dlwr	Dlwr
Dlwr_116	0	1	0	0	Dlwr	Dlwr
Dlwr_117	0	1	0	0	Dlwr	Dlwr
Dlwr_118	0	1	0	0	Dlwr	Dlwr
Dlwr_119	0	1	0	0	Dlwr	Dlwr
Dlwr_120	0	1	0	0	Dlwr	Dlwr
Dlwr_121	0	1	0	0	Dlwr	Dlwr
Dlwr_122	0	1	0	0	Dlwr	Dlwr
Dlwr_123	0.91	0	0	0	Dlwr	RdIP
Dlwr_124	0	1	0	0	Dlwr	Dlwr
Dlwr_125	0	1	0	0	Dlwr	Dlwr
Dlwr_126	0	1	0	0	Dlwr	Dlwr
Dlwr_127	0	1	0	0	Dlwr	Dlwr
Dlwr_128	0.87	1	0	0	Dlwr	Dlwr
Krka_14	0	0	1	0	Krka	Krka
Krka_15	0	0	1	0	Krka	Krka
Krka_16	0	0	1	0	Krka	Krka
Krka_17	0	0	1	0	Krka	Krka
Krka_19	0	0	1	0	Krka	Krka
Krka_20	0	0	1	0	Krka	Krka
Krka_22	0	0	1	0	Krka	Krka
Krka_23	0	0	1	0	Krka	Krka
Krka_25	0	0	1	0	Krka	Krka
Krka_27	0	0	1	0	Krka	Krka
Krka_28	0	0	1	0	Krka	Krka
Krka_29	0	0	1	0	Krka	Krka
Krka_30	0	0	1	0	Krka	Krka

Krka_44	0	0	1	0	Krka	Krka
Krka_45	0	0	1	0	Krka	Krka
Krka_47	0	0	1	0	Krka	Krka
Krka_48	0	0	1	0	Krka	Krka
Krka_50	0	0	1	0	Krka	Krka
Krka_51	0	0	1	0	Krka	Krka
Krka_52	0	0	1	0	Krka	Krka
Krka_53	0	0	1	0	Krka	Krka
Krka_54	0	0	1	0	Krka	Krka
Krka_57	0	0	1	0	Krka	Krka
Krka_58	0	0	1	0	Krka	Krka
Krka_59	0	0	1	0	Krka	Krka
Krka_60	0	0	1	0	Krka	Krka
Krka_61	0	0	1	0	Krka	Krka
Krka_63.	0	0	1	0	Krka	Krka
Pearl_1	0	0	0	1	Pearl	Pearl
Pearl_2	0	0	0	1	Pearl	Pearl
Pearl_3	0	0	0	1	Pearl	Pearl
Pearl_4	0	0	0	1	Pearl	Pearl
Pearl_5	0	0	0	1	Pearl	Pearl
Pearl_6	0	0	0	1	Pearl	Pearl
Pearl_7	0	0	0	1	Pearl	Pearl
Pearl_8	0	0	0	1	Pearl	Pearl
Pearl_9	0	0	0	1	Pearl	Pearl
Pearl_10	0	0	0	1	Pearl	Pearl
Pearl_11	0	0	0	1	Pearl	Pearl
Pearl_12	0	0	0	1	Pearl	Pearl
Pearl_13	0	0	0	1	Pearl	Pearl
Pearl_14	0	0	0	1	Pearl	Pearl
Pearl_15	0	0	0	1	Pearl	Pearl
Pearl_16	0	0	0	1	Pearl	Pearl

Table S6. Detailed performance of each ML method for prediction of samples belonging to each estuary, when combined in a single dataset.

	RdIP	Dlwr	Krka	Pearl
NNet				
Sensitivity	0.8000	1.0000	0.8000	1.00000
Specificity	1.0000	0.8462	1.0000	1.00000
Pos Pred Value	1.0000	0.9259	1.0000	1.00000
Neg Pred Value	0.9706	1.0000	0.9706	1.00000
Prevalence	0.1316	0.6579	0.1316	0.07895
Detection Rate	0.1053	0.6579	0.1053	0.07895
Detection Prevalence	0.1053	0.7105	0.1053	0.07895
Balanced Accuracy	0.9000	0.9231	0.9000	1.00000
RF				
Sensitivity	0.8000	0.9600	1.0000	1.00000
Specificity	0.9697	0.9231	1.0000	1.00000
Pos Pred Value	0.8000	0.9600	1.0000	1.00000
Neg Pred Value	0.9697	0.9231	1.0000	1.00000

Prevalence	0.1316	0.6579	0.1316	0.07895
Detection Rate	0.1053	0.6316	0.1316	0.07895
Detection Prevalence	0.1316	0.6579	0.1316	0.07895
Balanced Accuracy	0.8848	0.9415	1.0000	1.00000
C5.0				
Sensitivity	0.8000	1.0000	1.0000	0.66667
Specificity	1.0000	0.8462	1.0000	1.00000
Pos Pred Value	1.0000	0.9259	1.0000	1.00000
Neg Pred Value	0.9706	1.0000	1.0000	0.97222
Prevalence	0.1316	0.6579	0.1316	0.07895
Detection Rate	0.1053	0.6579	0.1316	0.05263
Detection Prevalence	0.1053	0.7105	0.1316	0.05263
Balanced Accuracy	0.9000	0.9231	1.0000	0.83333
LDA				
Sensitivity	1.0000	1.0000	0.8000	0.66667
Specificity	1.0000	0.8462	1.0000	1.00000

Pos Pred Value	1.0000	0.9259	1.0000	1.00000
Neg Pred Value	1.0000	1.0000	0.9706	0.97222
Prevalence	0.1316	0.6579	0.1316	0.07895
Detection Rate	0.1316	0.6579	0.1053	0.05263
Detection Prevalence	0.1316	0.7105	0.1053	0.05263
Balanced Accuracy	1.0000	0.9231	0.9000	0.83333
SVM				
Sensitivity	1.0000	0.9600	0.8000	1.00000
Specificity	0.9697	0.9231	1.0000	1.00000
Pos Pred Value	0.8333	0.9600	1.0000	1.00000
Neg Pred Value	1.0000	0.9231	0.9706	1.00000
Prevalence	0.1316	0.6579	0.1316	0.07895
Detection Rate	0.1316	0.6316	0.1053	0.07895
Detection Prevalence	0.1579	0.6579	0.1053	0.07895
Balanced Accuracy	0.9848	0.9415	0.9000	1.00000

## Supplementary Figures

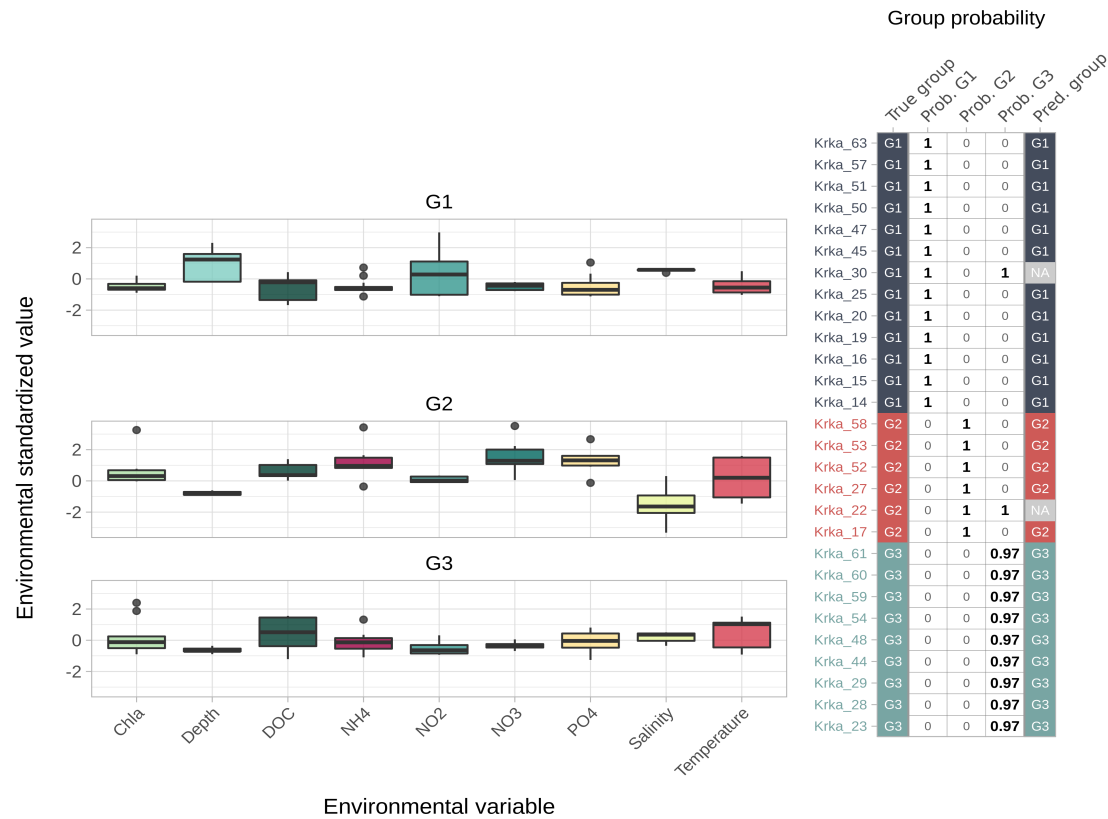


Figure S1. A) Physicochemical variables characterising each of the four groups of the Krka estuary. The lower and upper hinges of the boxes correspond to the first and third quartiles. Data beyond the end of the whiskers are the outlier points. The horizontal line represents the median value, B) assignment to the groups according to the bacterial indicators, using leave one out cross-validation.

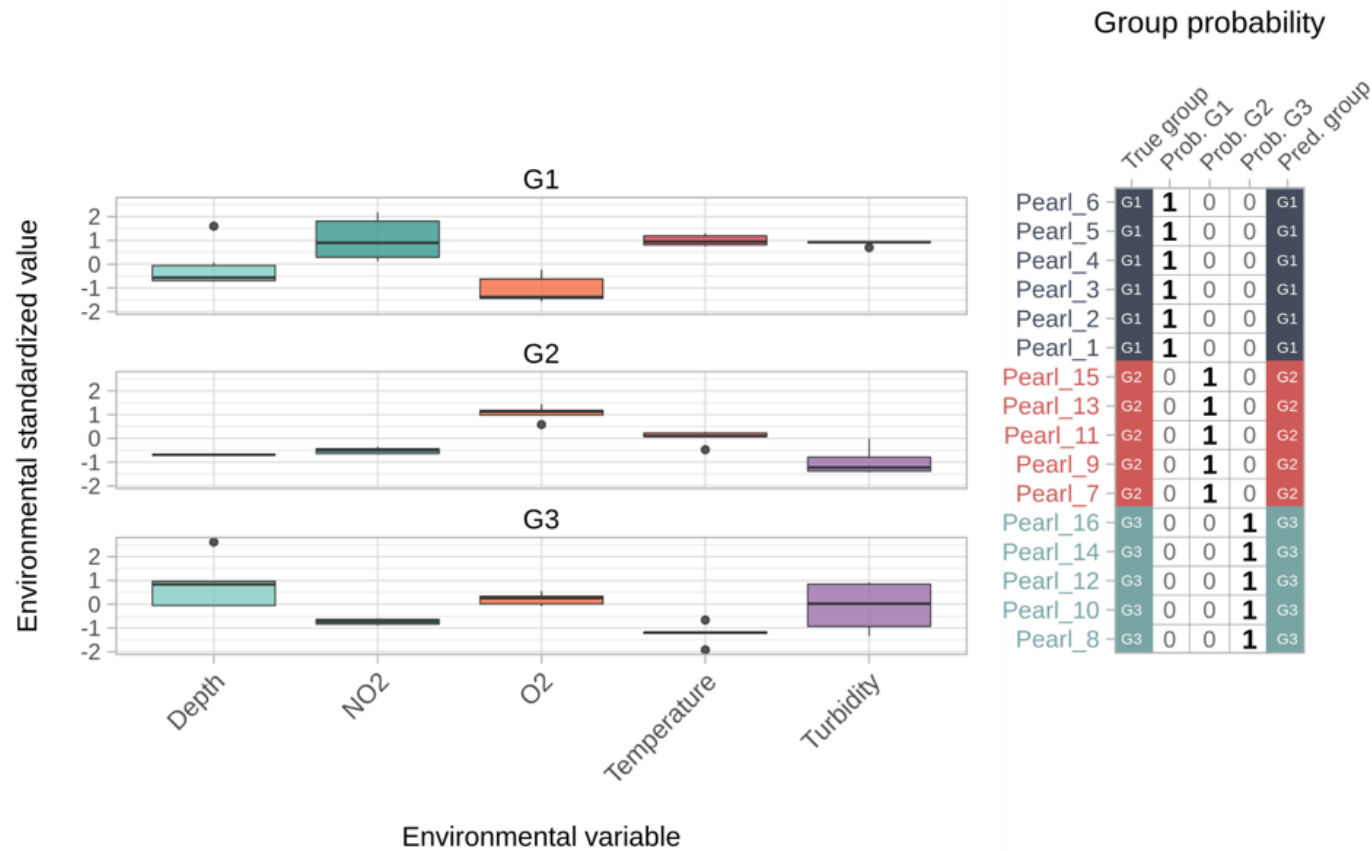


Figure S2. A) Physicochemical variables characterising each of the four groups of the Pearl estuary. The lower and upper hinges of the boxes correspond to the first and third quartiles. Data beyond the end of the whiskers are the outlier points. The horizontal line represents the median value, B) assignment to the groups according to the bacterial indicators, using leave one out cross-validation.

!

!

