The following supplements accompany the article

Diversity patterns of benthic bacterial communities along the salinity continuum of the Humber estuary (UK)

Andrea Vidal-Durà*, Ian T. Burke, Robert J. G. Mortimer, Douglas I. Stewart

*Corresponding author: andrea.vVidal@uab.cat

Aquatic Microbial Ecology 81: 277–291 (2018)

Supplement 1 Salinity zones of the Humber Estuary and correlation with other physicochemical characteristics

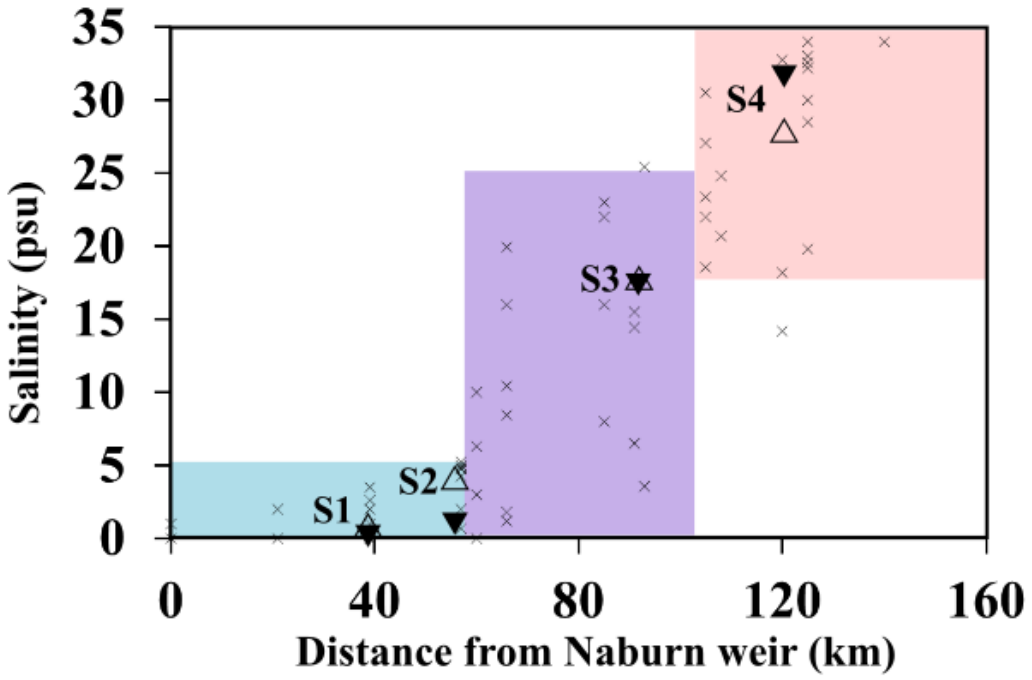

Figure S1: Salinity zonation based on salinity records of different sites along the Humber estuary (x) (Freestone, 1987; Prastka and Malcolm, 1994; NRA, 1995, 1996; Sanders *et al.*, 1997; Barnes and Owens, 1998; Mitchell, 1998; Uncles *et al.*, 1998; Mortimer *et al.*, 1999; Williams and Millward, 1999; ABP Research 2000; Millward *et al.*, 2002; Burke *et al.*, 2005; Uncles *et al.*, 2006; Fujii and Raffaelli, 2008; Garcia-Alonso *et al.*, 2011). Salinity ≤5 (blue area); 0-25 salinity range (purple area); and 18-35 salinity range (pink area).The triangle markers indicate the porewater salinity measurements of this study (S1-S4) (empty and coloured markers for surface and subsurface porewater salinity respectively).

Salinity was correlated (R>0.80) with porewater nitrate (positively and negatively correlated in surface and subsurface porewaters respectively), sulfate and ammonium (only in surface porewater), and also with total iron in the solids and with 0.5 N HCl extractable iron (II) in solids only in subsurface sediments. Salinity was also negatively correlated with particle size (R>0.9), but was not correlated with the TOC content in the solids.

Supplement 2 Supplementary Taxonomical Data

Table S1: Summary of the taxonomical classification of the four main Bacteria phyla to class level (0.7 confidence at class level) in the different samples (S1-S4). The suffixes *s* and *d* refer to surface and subsurface sediments respectively. The percentages of reads at the Bacteria phylum level are respective to the total reads per sample, and the percentages at class level are respective to the total reads of the specific phyla per sample. The error is ±1σ.

% of reads		**S1s**	**S1d**	**S2s**	**S2d**	**S3s**	**S3d**	**S4s**	**S4d**	**Average**
% Acidobacteria		**17**	**16**	**14**	**15**	**8**	**1**	**6**	**10**	**11±5**
	Classes Acidobacteria									
	Acidobacteria_Gp2	0	0	0	0	0	0	0	0	0±0
	Acidobacteria_Gp3	2	2	2	2	2	0	1	0	1±1
	Acidobacteria_Gp4	2	3	3	3	3	1	2	0	2±1
	Acidobacteria_Gp5	0	0	0	0	0	0	0	0	0±0
	Acidobacteria_Gp6	42	41	36	38	25	10	8	3	25±16
	Acidobacteria_Gp7	4	7	4	4	5	2	4	1	4±2
	Acidobacteria_Gp9	1	2	2	2	2	4	2	7	3±2
	Acidobacteria_Gp10	1	1	2	2	2	1	2	1	1±1
	Acidobacteria_Gp11	0	0	0	0	0	0	0	0	0±0
	Acidobacteria_Gp16	15	16	17	17	19	13	27	9	17±5
	Acidobacteria_Gp17	13	12	14	14	10	18	8	13	13±3
	Acidobacteria_Gp18	4	2	4	3	2	5	1	4	3±1
	Acidobacteria_Gp19	0	0	0	0	0	0	0	0	0±0
	Acidobacteria_Gp21	1	1	2	2	4	10	10	18	6±6
	Acidobacteria_Gp22	2	1	3	2	7	2	14	10	5±5
	Acidobacteria_Gp23	1	1	1	1	3	6	5	10	4±3
	Acidobacteria_Gp26	0	0	0	0	0	0	1	1	0±0
	Holophagae	1	1	0	0	0	0	0	0	0±0
	Not classified	11	10	11	11	16	29	14	23	16 ±7
% Proteobacteria		**40**	**40**	**45**	**45**	**47**	**92**	**49**	**47**	**51±17**
	Classes Proteobacteria									
	Alphaproteobacteria	15	15	13	13	14	1	18	7	12±5

	Betaproteobacteria	23	25	20	20	10	0	4	3	13±10
	Gammaproteobacteria	40	36	45	45	58	4	54	45	41±17
	Deltaproteobacteria	15	19	15	15	11	1	21	36	17±10
	Epsilonproteobacteria	0	0	1	0	2	94	0	3	13±33
	Other classes (Zetaproteobacteria and Oligoflexia)	0	0	0	0	0	0	0	0	0±0
	Not classified	6	5	6	6	4	0	3	6	5±2
% Bacteroidetes		**8**	**8**	**10**	**9**	**15**	**1**	**21**	**13**	**10±6**
	Classes Bacteroidetes									
	Bacteroidetes_incertae_sedis	0	0	0	0	0	0	0	0	0±0
	Bacteroidia	4	3	4	3	3	5	4	4	4±1
	Cytophagia	8	4	10	9	11	3	3	2	6±4
	Flavobacteriia	35	35	38	39	59	79	77	88	56±22
	Sphingobacteriia	20	25	18	19	13	4	8	2	14±8
	Not classified	10	10	5	5	1	0	0	0	20±13
% Chloroflexi		**12**	**11**	**11**	**11**	**7**	**3**	**6**	**16**	**9±4**
	Classes Chloroflexi									
	Anaerolineae	19	16	21	22	19	36	22	32	23±7
	Caldilineae	35	37	33	31	35	13	36	11	29±11
	Chloroflexia	0	0	0	0	0	0	0	0	0±0
	Dehalococcoidotes	0	0	0	0	0	0	0	0	0±0
	Thermomicrobia	0	0	0	0	0	0	0	0	0±0
	Not classified	45	46	46	47	46	51	41	57	47±5

Table S2: Summary of the taxonomical classification of the phyla *Proteobacteria* to order level (0.7 confidence at class level) in the different samples (S1-S4). The suffixes *s* and *d* refer to surface and subsurface sediments respectively. The percentages of reads at order level are calculated respectively to the total reads of that class per sample. When all the reads in an order were classified to the same family, this has also been specified in brackets.

% of reads		**S1s**	**S1d**	**S2s**	**S2d**	**S3s**	**S3d**	**S4s**	**S4d**
Class	**Order (Family)**								
Alphaproteobacteria									
	Alphaproteobacteria_incertae_sedis	5	6	4	3	2	1	1	2
	Caulobacterales	2	1	1	0	1	0	0	0
	Kordiimonadales	0	0	0	0	0	0	0	0
	Parvularculales	0	0	0	0	0	0	0	0
	Rhizobiales	34	34	36	35	23	44	17	42
	Rhodobacterales	21	23	27	29	50	38	72	44
	Rhodospirillales	16	15	14	14	12	9	4	7
	Rickettsiales	1	1	1	0	0	0	0	0
	Sneathiellales	0	0	0	0	0	0	0	0
	Sphingomonadales	21	20	17	18	11	8	6	5
	Not classified	0	0	0	0	0	0	0	0
	Total	100	100	100	100	100	100	100	100
Betaproteobacteria									
	Burkholderiales	39	36	35	35	38	27	25	31
	Gallionellales	4	3	4	4	1	2	1	0
	Hydrogenophilales	5	3	6	6	2	27	1	34
	Methylophilales	13	14	18	16	29	12	52	8
	Neisseriales	4	5	5	5	5	3	4	4
	Nitrosomonadales	16	17	16	17	14	19	14	14
	Rhodocyclales	19	22	14	16	11	10	3	9
	Not classified	1	1	1	1	0	0	0	0
	Total	100	100	100	100	100	100	100	100
Gammaproteobacteria									
	Aeromonadales	1	0	1	1	1	0	0	0
	Alteromonadales	11	10	10	10	11	5	10	2

	Chromatiales	6	6	7	7	6	10	5	5
	Enterobacteriales	0	0	0	0	1	0	0	0
	Gammaproteobacteria_incertae_sedis	54	52	59	59	64	43	77	83
	Legionellales	2	2	1	1	0	0	0	0
	Methylococcales	2	2	2	2	1	1	0	1
	Oceanospirillales	4	3	3	3	5	2	1	1
	Pseudomonadales	3	4	2	2	2	0	1	0
	Thiotrichales	1	0	0	0	0	30	2	0
	Vibrionales	0	0	0	0	0	0	0	0
	Xanthomonadales	14	18	10	12	4	7	3	9
	Not classified	3	2	4	3	3	1	1	0
	Total	100	100	100	100	100	100	100	100
Deltaproteobacteria									
	Bdellovibrionales	4	1	3	2	4	0	1	0
	Desulfarculales	0	0	0	0	0	1	0	1
	Desulfobacterales	22	15	30	27	38	74	46	75
	Desulfovibrionales	0	1	1	1	1	0	0	0
	Desulfuromonadales	39	55	33	36	32	11	44	19
	Myxococcales	23	20	18	20	17	6	7	3
	Syntrophobacterales	12	8	14	14	8	7	2	2
	Syntrophorhabdaceae	0	0	1	1	0	0	0	0
	Not classified	0	0	0	0	0	0	0	0
	Total	100	100	100	100	100	100	100	100
Epsilonproteobacteria									
	Campylobacterales *(Campylobacteraceae)*	100	100	100	100	100	100	100	100

Table S3: Summary of the taxonomical classification of the phyla *Bacteroidetes* to order level (0.7 confidence at class level) in the different samples (S1-S4). The suffixes *s* and *d* refer to surface and subsurface sediments respectively. The percentages of reads at order level are calculated respectively to the total reads of that class per sample. When all the reads in an order were classified to the same family, this has also been specified in brackets.

% of reads		**S1s**	**S1d**	**S2s**	**S2d**	**S3s**	**S3d**	**S4s**	**S4d**
Class	**Order (Family)**								
Bacteroidia									
	Bacteroidales *(Bacteroidaceae)*	2	0	0	0	0	0	0	0
	Bacteroidales *(Marinilabiaceae)*	20	27	15	16	1	0	0	0
	Bacteroidales *(Porphyromonadaceae)*	5	1	2	2	9	1	0	0
	Bacteroidales *(Prevotellaceae)*	0	0	0	0	0	0	0	0
	Bacteroidales *(Prolixibacteraceae)*	74	72	82	83	89	99	100	100
	Bacteroidales *(Rikenellaceae)*	0	0	0	0	0	0	0	0
	Not classified	0	0	1	0	1	0	0	0
	Total	100	100	100	100	100	100	100	100
Cytophagia									
	Cytophagales *(Cytophagaceae)*	68	79	58	58	32	23	18	6
	Cytophagales *(Flammeovirgaceae)*	32	21	42	42	68	77	82	94
	Not classified	0	0	0	0	0	0	0	0
	Total	100	100	100	100	100	100	100	100
Flavobacteriia									
	Flavobacteriales *(Flavobacteriaceae)*	93	96	96	96	97	99	99	100
	Flavobacteriales *(Cryomorphaceae)*	7	4	4	4	3	1	1	0
	Not classified	0	0	0	0	0	0	0	0
	Total	100	100	100	100	100	100	100	100
Sphingobacteriia									
	Sphingobacteriales *(Chitinophagaceae)*	26	19	15	11	3	7	3	9
	Sphingobacteriales *(Cyclobacteriaceae)*	1	0	0	0	0	2	0	0
	Sphingobacteriales *(Rhodothermaceae)*	6	6	9	10	9	14	12	13
	Sphingobacteriales *(Saprospiraceae)*	57	65	71	74	87	77	84	77

	Sphingobacteriales *(Sphingobacteriaceae)*	10	10	5	5	1	0	0	0
	Not classified	0	0	0	0	0	0	0	0
	Total	100	100	100	100	100	100	100	100

Table S4: Summary of the taxonomical classification of the phyla *Chloroflexi* to order level (0.7 confidence at class level) in the different samples (S1-S4). The suffixes *s* and *d* refer to surface and subsurface sediments respectively. The percentages of reads at order level are calculated respectively to the total reads of that class per sample. When all the reads in an order were classified to the same family, this has also been specified in brackets.

% of reads		**S1s**	**S1d**	**S2s**	**S2d**	**S3s**	**S3d**	**S4s**	**S4d**
Class	**Order (Family)**								
Anaerolineae									
	Anaerolineales *(Anaerolineaceae)*	19	16	21	22	19	36	22	32
	Not classified	81	84	79	78	81	64	78	68
	Total	100	100	100	100	100	100	100	100
Caldilineae									
	Caldilineales *(Caldilineaceae)*	35	37	33	31	35	13	36	11
	Not classified	65	63	67	69	65	87	64	89
	Total	100	100	100	100	100	100	100	100
Chloroflexia									
	Chloroflexales	6	24	50	73	100	100	100	100
	Herpetosiphonales	94	76	50	28	0	0	0	0
	Not classified	0	0	0	0	0	0	0	0
	Total	100	100	100	100	100	100	100	100

Supplement 3 Diversity Data

In this study "Hill numbers" (Hill, 1973) were used to analyse bacterial diversity. Hill numbers (D_q) were proposed as a unified family of diversity indices that compensate for the disproportionate impact of rare taxa by weighting taxa based on abundance (Hill, 1973; Jost, 2006, 2007). As a result, D_q are more suitable for working with the large datasets produced by amplicon sequencing technologies than traditional diversity measures (Kang *et al.*, 2016). The basic expression for the Hill numbers (D_q) is represented in Equation 1.

$$D_q = \left(\sum_{i=1}^{S} p_i^q\right)^{\frac{1}{1-q}} \qquad \text{(Eq 1)}$$

Where S is the total number of species (OTUs in this study) and p_i is the proportion of individuals belonging to the i^{th} species in the dataset. The parameter q, is the "order of the diversity measure" and determines how the abundance is weighted. By increasing the index q the diversity measurement places progressively more weight on the more abundant OTUs within a population. The unweighted Hill number, D_0, is exactly equivalent to the species richness. D_1 is a measure of the number of common species and is equivalent to the exponential of Shannon entropy. D_2 is a measure of the number of dominant species and is equivalent to the inverse of Simpson concentration (Hill, 1973; Jost, 2006, 2007). The conversion of traditional diversity indices to D_q of different order is presented in Supplementary Table 6. Complete information about the diversity results in this study is presented in Supplementary Table 7.

Table S5: Conversion of traditional diversity indices to "Hill numbers" (D_q) for q=0 q=1 and q=2 (D_0, D_1, and D_2) (modified from Jost (2007)).

Order of the diversity measurement (q)	**Traditional Diversity Index (D)**	**To convert diversity indices (D) to diversity measurement (D_q)**	**Diversity measurements in terms of p_i (D_q)**
0	Species Richness $D \equiv \sum_{i=1}^{S} p_i^0$	D	$D_0 = \sum_{i=1}^{S} p_i^0$=S
1	Shannon entropy $D \equiv -\sum_{i=1}^{S} p_i \ln p_i$	exp (D)	$D_1 = \exp\left(-\sum_{i=1}^{S} p_i \ln p_i\right)$
2	Simpson concentration $D \equiv \sum_{i=1}^{S} p_i^2$	1/D	$D_2 = {}^{1}/_{\sum_{i=1}^{S} p_1^2}$

Hills numbers represent measures of a number of species, the *effective number of species*, i.e."the number of equally abundant species that would be needed to give the same value of a diversity measure" (Chao *et al.*, 2014, p.46). They are symbolised by D_q (Eq 1). The sum in Equation 1 is symbolised in Jost (2007) by ${}^{q}\lambda$, and it is the key of these calculations:

$$\lambda^q = \sum_{i=1}^{S} p_i^q \qquad \text{(Eq.2)}$$

So Equation 1 will look like:

$$D_q = \left(\sum_{i=1}^{S} p_i^q\right)^{1/(1-q)} = (\lambda^q)^{1/(1-q)} \qquad \text{(Eq. 3)}$$

To analyse regional diversity (gamma diversity D_q^{γ}), we need to calculate its different components (alpha and beta). The alpha (single community), beta (between the different communities considered), and gamma (regional) components of a diversity index, can be individually converted to diversity measurements (D_q^{α}, D_q^{β}, and D_q^{γ}). Following Whittaker's multiplicative law (Whittaker, 1972), alpha, beta and gamma diversities are related like so:

$$D_q^{\gamma} = D_q^{\alpha} \times D_q^{\beta}$$ (Eq. 4)

For the alpha component of any diversity index (D^{α}):

$$D^{\alpha} \equiv D\left({}^{\alpha}_{q}\lambda\right) = D\left[\frac{w_1^q \sum_{i+1}^{S} p_{i1}^q + w_2^q \sum_{i=1}^{S} p_{i2}^q + \cdots}{w_1^q + w_2^q + \cdots}\right]$$ (Eq. 5)

w_j is the statistical weight of community j (number of individuals (valid reads) in the community j (sample j) divided by the total number of reads in the region). Therefore, the alpha diversity measurement of order q ($D_q{}^{\alpha}$) equivalent to that diversity index (D^{α}) is:

$$D_q^{\alpha} \equiv D\left({}^{\alpha}_{q}\lambda\right) = \left[\frac{w_1^q \sum_{i+1}^{S} p_{i1}^q + w_2^q \sum_{i=1}^{S} p_{i2}^q + \cdots}{w_1^q + w_2^q + \cdots}\right]^{1/(1-q)}$$ (Eq. 6)

That expression is undefined at q=1, but the limit exists as q approaches 1 ($\lim_{q\to 1}$) being the exponential of alpha Shannon entropy:

$$D_1^{\alpha} \equiv exp\left[-w_1 \sum_{i=1}^{S}(p_{i1} \ln p_{i1}) + -w_2 \sum_{i=1}^{S}(p_{i2} \ln p_{i2}) + \cdots\right]$$ (Eq.7)

When different communities (samples in our case) are considered, $D_1{}^{\alpha}$ is not the average of the diversity indices of the individual communities. We must average the basics sums ($^{q}\lambda$) (Eq. 1) of the individual communities and then calculate the diversity index of that average ($^{*}D_1{}^{\alpha}$) (Jost, 2007). So community weights are considered in this calculation. For our samples:

$$^{*}D_1{}^{\alpha}=438$$

Regional diversity measurement of order 1 ($D_1{}^{\gamma}$) of all the pooled samples equals:

$$D_1^{\gamma} \equiv exp\left[\sum_{i=1}^{S} -(w_1 p_{i1} + w_2 p_{i2} + \cdots) \times \ln(w_1 p_{i1} + w_2 p_{i2} + \cdots)\right]$$ (Eq. 8)

$$D_1{}^{\gamma}=934$$

Then we can calculate the beta diversity ($D_1{}^{\beta}$) which is the measurement of the relative change in species composition between locations or communities by using Equation 4:

$$D_1^{\beta} = D_1^{\gamma} / D_1^{\alpha}$$

$$D_1{}^{\beta} = 934/438 = 2.13$$

$D_1{}^{\beta}$ has been described as the number of distinct communities or samples in the region (Jost, 2007). This measurement can be converted into MacArthur's (1965) homogeneity measure (Equation 9). This ratio answers the question of "*what proportion of total diversity is found within the averaged community or sample?"*(Jost, 2007). According to this homogeneity measure, 47% of the total diversity is found in the average community.

$$M = 1 \Big/ D_1^{\beta} = \exp(D^{\alpha}) \Big/ \exp(D^{\beta}) \text{ (Eq. 9)}$$

$$M = 1/D_1{}^{\beta} = 1/2.13 = 0.47$$

Table S6: Number of reads and bacterial diversity measurements of the samples (S1-S4). The suffixes *s* and *d* refer to surface and subsurface sediments respectively. The total number of reads include sequences classified as Archaea and Bacteria (including poorly classified reads, <0.7 confidence). Number of reads per sample is the sum of all the reads that exceed the quality thresholds (after quality control, and classified to the Bacteria phylum level at a confidence level >0.7).

	S1		**S2**		**S3**		**S4**	
	S1s	**S1d**	**S2s**	**S2d**	**S3s**	**S3d**	**S4s**	**S4d**
Total number of paired-end reads	712,402	976,928	638,252	821,137	577,701	1,208,696	721,187	522,816
Number of reads classified to the Bacteria phylum level	556,621	802,492	492,132	641,797	454,121	1,113,761	633,444	370,056
OTUs richness or Hill numbers of order 0 (D_0^{α})	5,262	5,968	5,599	6,004	5,541	3,873	4,126	3,488
Average Richness	4,983							
Shannon Entropy	7.1	7.3	7.1	7.2	3.3	6.7	6.0	6.3
Hill numbers of order 1 (D_1^{α})	1,436	1,174	1,309	1,250	833	28	546	412
Hill numbers of order 2 (D_2^{α})	487	378	385	374	175	10	154	120
Alpha (averaged community, ${}^{*}D_1^{\alpha}$), gamma (regional, D_1^{γ}), and beta (between communities, D_1^{β}) diversities of order 1	${}^{*}D_1^{\alpha} = 438$ $D_1^{\gamma} = 934$ $D_1^{\beta} = 2.13$							

Supplement 4 Taxa accumulation Curves (TAC)

Taxa accumulation curves (TAC) (Supplementary Figure 2) have been calculated for the full raw dataset, without replacement (reads are picked at random and a given read can only be picked once and results are averaged over multiple trials; here 8 iterations). TACs cannot be calculated from the dataset after removal of artefacts due to the way the sequence analysis pipeline operates. Therefore the dataset used to calculate the TAC contained the unfiltered-reads pool from the eight samples used, i.e. reads from before quality checks were applied in the sequence analysis pipeline. Some of the reads that were clustered into the OTUs generated were later removed from the diversity analysis (e.g. OTUs identified as Archaea and OTUs which were not classified to the Bacteria phylum level with a confidence >0.7).

TAC for the regional OTUs richness shows that D_0^{γ} varies by less than 0.1 once >60% of the dataset is subsampled. However, this does not mean that more "rare" taxa would not be found when the sequencing densities were higher.

TAC for common and dominant OTUs show that both D_1^{γ} and D_2^{γ} vary by less than 0.1% once 20% of the dataset is subsampled. Further, because these metrics characterise the number of common and dominant OTUs it is inconceivable that the values calculated would be any different if deeper sequencing had been undertaken.

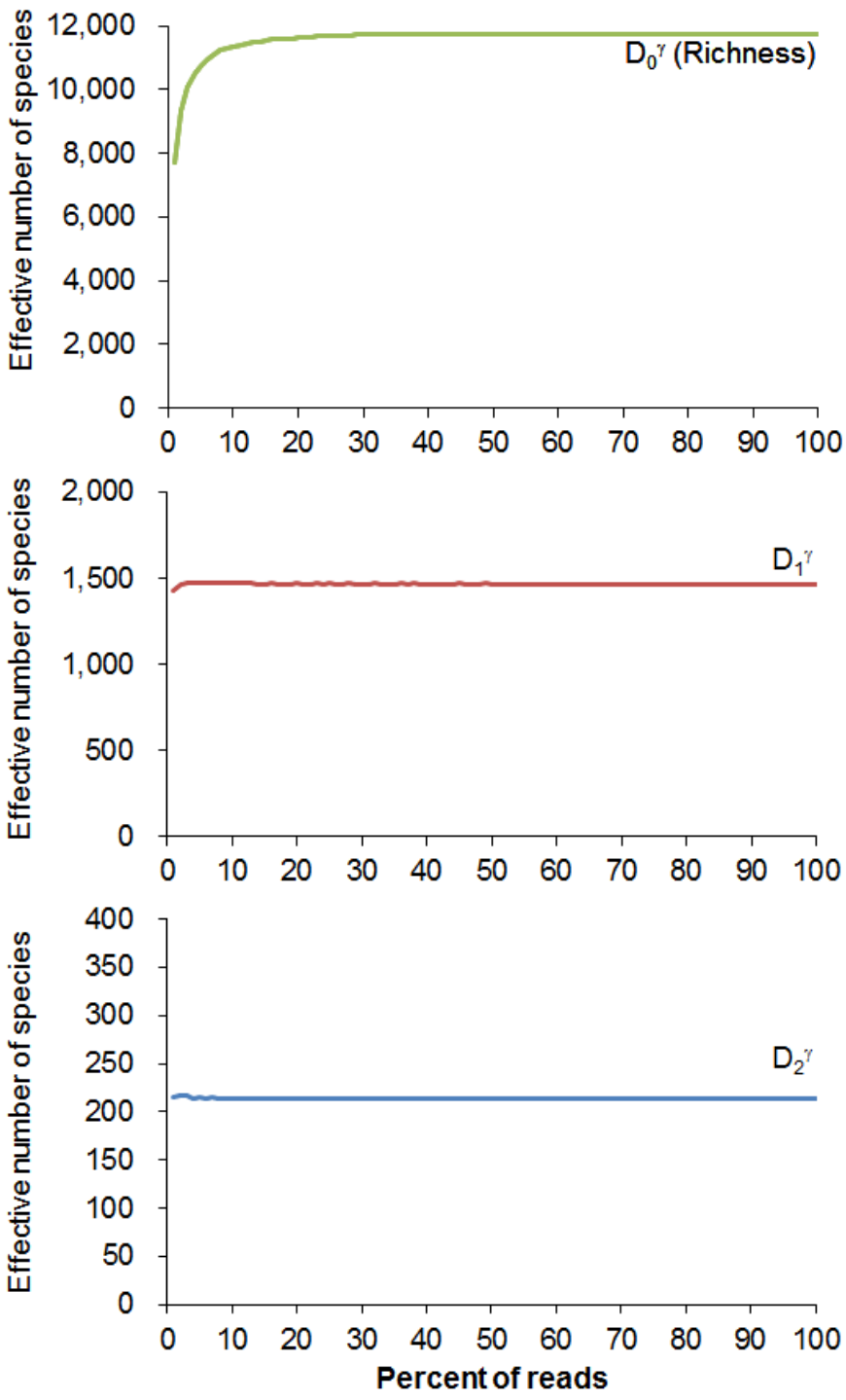

Figure S2: Taxa accumulation curves for the unfiltered regional dataset subsampled without replacement (average of 8 replications) indicating that D_1^γ and D_2^γ converge very rapidly, and D_0^γ converges when >60% of the dataset is subsampled. The unfiltered dataset contains OTUs later removed from the diversity analysis, such as OTUs identified as archaea and OTUs which were

not classified to the Bacteria phylum level with a confidence > 0.7. In the diversity analysis, taxa represented unique OTUs at 97% similarity cutoff.

Supplement 5 Bray-Curtis dissimilarity Matrix

To obtain the matrix we used the package "vegan" (Oksanen *et al.*, 2013). First we import the data of the bacterial community (relative abundance data):

```
>community_data <-read.csv("relativeabundancetable.csv",row.names = 1,
check.names = FALSE)
```

Then we can obtain the matrix by applying the following command:

```
>vegdist("community_data",method="bray", binary=FALSE, diag=FALSE, upper=FALSE,
na.rm=FALSE)
```

Table S7: Bray-Curtis dissimilarity matrix

	S1s	**S1d**	**S2s**	**S2d**	**S3s**	**S3d**	**S4s**	**S4d**
S1s	0							
S1d	0.239319	0						
S2s	0.265464	0.28351	0					
S2d	0.239564	0.245803	0.130322	0				
S3s	0.552836	0.553242	0.442159	0.446962	0			
S3d	0.911113	0.913001	0.901849	0.901164	0.900012	0		
S4s	0.705474	0.705811	0.635151	0.638155	0.424329	0.913353	0	
S4d	0.762316	0.772682	0.704974	0.704546	0.635732	0.873585	0.55026	0

Supplement 6 Heat map

To obtain the heat map we used the packages "gplots" and "RColorBrewer".

First we import the data of the bacterial community (relative abundance data):

```
>community_data <-read.csv("relativeabundancetable.csv",row.names = 1,
check.names = FALSE)
> matrix_community <-community_data(x[,2:ncol(x)])
```

The colours are defined by the `colorRampPalette` command (example of 5 colours). The intervals are defined as well.

```
>my_palette<- colorRampPalette(c("antiquewhite3", "skyblue1", "yellowgreen", "sal
mon", "red3")
> col_breaks = c(seq(0,0.001,length=100), seq(0.001,0.01,length=100), seq(0.01,0.
1, length=100), seq(0.1,1, length=100), seq(1,100, length=100))
```

Finally we use the `heatmap2` function to create the heat map and we define separately the legend:

```
> heatmap.2(matrix_community, Rowv= FALSE, main= "Heatmap", dendogram="none", col
=my_palette, breaks=col_breaks, trace="none", density.info = "none", key= TRUE, s
ymkey = FALSE, scale = "none", rowsep = 1:nrow(matrix_communinity), sepcolor = "w
hite", sepwidth = c(0.05, 0.05))
>legend("left", fill = my_palette(5), legend = c("0 to 0.001", "0.001 to 0.01", "
0.01 to 0.1", "0.1 to 1", ">1"))
```

The heat map is another graphical representation of the similarities and dissimilarities of the bacterial community composition along the salinity gradient. The green and red bands (from 0.01 to >1) are the important ones to look at. We can interpret the grey bands as absence or extremely low abundances. Samples from the inner estuary (S1s, S1d, S2s, and S2d) share the bands with the greatest abundances (right part of the heat map). S3s also shares a similar pattern although the green areas are a modestly more spread than at the inner samples. The S3d shows again the most unalike community composition. It was also the less diverse of the whole set of samples when D_1^{α} and D_2^{α} were applied. The S4s and S4d samples in the Heat map vary slightly from the rest of the samples (see some green bars that are not seen in other samples), which may indicate the differences in the bacterial community composition in the outer most estuary.

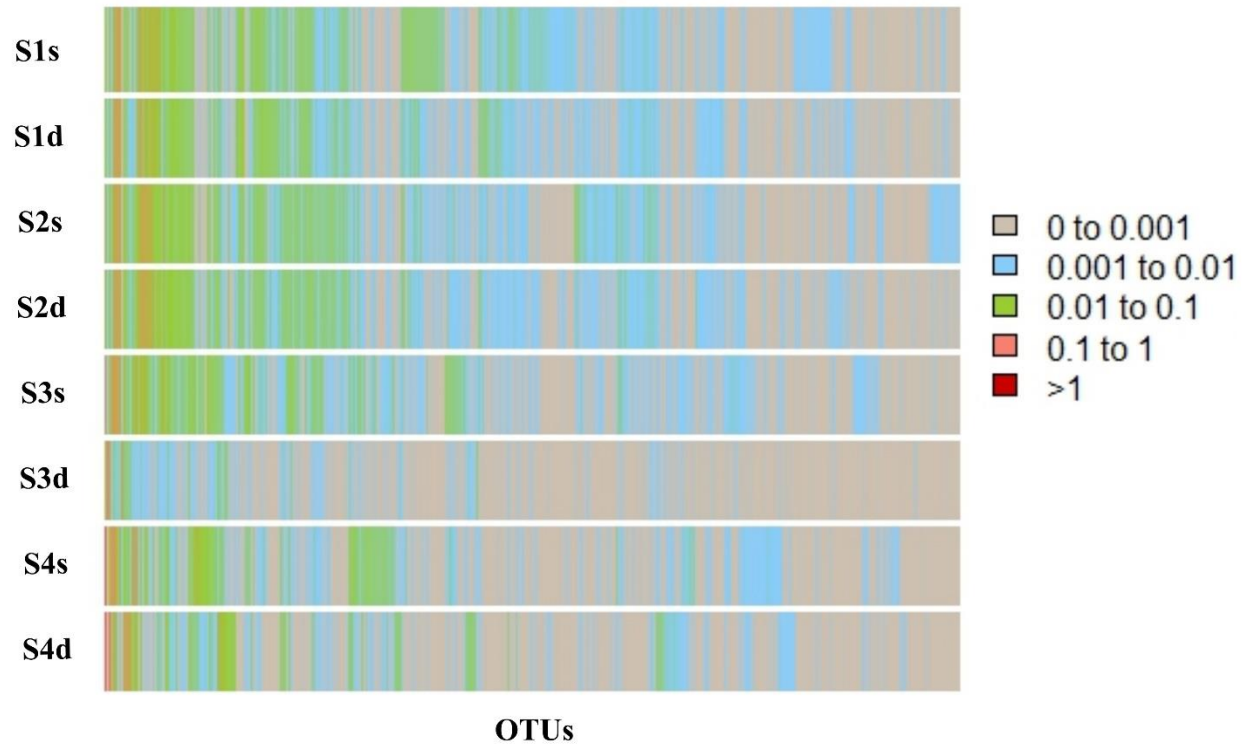

Figure S3: Heat map displaying relative abundance (%) of the OTUs (7656 OTUs) for the eight estuarine sediment samples (S1-S4). The suffixes *s* and *d* refer to surface and subsurface sediments respectively.

Supplement 7 BIOENV test

We use the function `bioenv`. The environmental parameters used were the following: Salinity; concentration in porewater of nitrate, ammonium, sulfate, Fe^{2+} and Mn^{2+}; 0.5 N HCl extractable $Fe^{2+}_{(s)}$; total iron in sediments; Fe-pyrite; Fe-FeS_2; Total Organic Carbon (TOC); and sediment grain size (D_{50}).

The function `bioenv` was applied as follows:

```
> bioenv_solution <-bioenv(community_data, environmental_variables,
fix.dist.method="bray", var.dist.method="euclidean",scale.fix=FALSE,
scale.var=TRUE, var.max=ncd(var.mat))
4095 possible subsets (this may take time...)
> bioenv_solution
```

```
Call:
bioenv(comm = community_matrix, env = env_variables,      fix.dist.method =
"bray", var.dist.method = "euclidean",      scale.fix = FALSE, scale.var = TRUE,
var.max = ncd(var.mat))

Subset of environmental variables with best correlation to community data.

Correlations:      spearman

Dissimilarities:   bray

Best model has 2 parameters (max. 12 allowed):


PW_Sulfate PW_Ammonium
with correlation  0.9447181
```

This was the output solution of the BIOENV, in which we can see that the ammonium and sulfate concentrations in porewater were the environmental parameters with the best correlation (0.94) with the bacterial community data. However, the third option included ammonium, salinity, and acid extractable Fe(II) in the sediments (0.94).

```
> summary(bioenv_solution)
```

	size	correlation
PW_Ammonium	1	0.8318
PW_Sulfate PW_Ammonium	**2**	**0.9447**
Salinity PW_Ammonium Fe2_solids	3	0.938
Salinity PW_Sulfate PW_Ammonium Fe2_solids	4	0.9288
Salinity PW_Sulfate PW_Ammonium Fe2_solids PW_Mn2	5	0.8872
Salinity PW_Nitrate PW_Sulfate PW_Ammonium Fe2_solids D50	6	0.8697
Salinity PW_Nitrate PW_Sulfate PW_Ammonium Fe2_solids PW_Fe2 D50	7	0.8522
Salinity PW_Nitrate PW_Ammonium Fe2_solids PW_Fe2 PW_Mn2 FeS2 D50	8	0.8363
Salinity PW_Nitrate PW_Ammonium Fe2_solids PW_Fe2 PW_Mn2 FeTOT FeS2 D50	9	0.8112
Salinity PW_Nitrate PW_Sulfate PW_Ammonium Fe2_solids PW_Mn2 FeTOT Fe_Pyrite TOC D50	10	0.7750

```
Salinity PW_Nitrate PW_Sulfate PW_Ammonium Fe2_solids
PW_Fe2 PW_Mn2 FeTOT Fe_Pyrite FeS2 D50                    11      0.7499

Salinity PW_Nitrate PW_Sulfate PW_Ammonium Fe2_solids
PW_Fe2 PW_Mn2 FeTOT Fe_Pyrite FeS2 TOC D50                12      0.6864
```

In order to test the significance of bioenv results, we have used function mantel. The result for the Mantel statistic was R=0.72 which indicates strong positive correlation between the two distance matrices.

```
> veg.dist <- vegdist(community_matrix, method = "bray", binary = FALSE, diag =
FALSE, upper = FALSE, na.rm = FALSE)

> env.dist<-vegdist(scale(env_variables), "euclid")

> mantel(veg.dist, env.dist)

Mantel statistic based on Pearson's product-moment correlation


Call:

mantel(xdis = veg.dist, ydis = env.dist)


Mantel statistic r: 0.7238

      Significance: 0.002

Upper quantiles of permutations (null model):
  90%   95% 97.5%   99%
0.343 0.441 0.507 0.616

Permutation: free
Number of permutations: 999
```

Supplement 8 Statistical tests

Non-parametric multivariate analysis of variance (PERMANOVA) (Anderson, 2001) was used to assess the similarity in the microbial abundance among samples. The test was performed with the function adonis in R.

- To test if there are differences in the composition of the bacterial communities (OTUs relative abundance) in samples from different depths (permutations constrained within sites). The null

hypothesis was true, there were no significant differences in the bacterial community composition in samples from different depths ($p>0.05$).

```
> adonis_depth <- adonis(community.matrix ~ depth, strata = variables$Site, data
= variables, permutations = 999, method = "bray")
> adonis_depth
Call:
adonis(formula = community.matrix ~ depth, data = variables, permutations = 999,
method = "bray", strata = variables$Site)

Permutation: free
Number of permutations: 999

Terms added sequentially (first to last)

          Df SumsOfSqs MeanSqs F.Model      R2 Pr(>F)
depth      1   0.18477 0.18477 0.84573 0.12354  0.622
Residuals  6   1.31085 0.21848         0.87646
Total      7   1.49562                 1.00000
```

- To test if there are differences in the composition of the bacterial communities (OTUs relative abundance) in samples from different sites/zones of the estuary. We test different options combining depths, sites and zones (with two groups, i.e. inner/outer, and with three groups, i.e. inner/middle/outer) and the results were consistent. All the PERMANOVA tests indicated that there were significant differences in the bacterial community composition in samples from different sites (or zones of the estuary) ($p<0.05$).

```
> adonis_site&depth <-adonis(community.matrix ~ site+depth, data = variables, per
mutations =999, method = "bray", strata = NULL )
> adonis_site&depth
Call:
adonis(formula = community.matrix ~ site + depth, data = variables, permutations
= 999, method = "bray", strata = NULL)

Permutation: free
Number of permutations: 999

Terms added sequentially (first to last)
```

```
          Df SumsOfSqs MeanSqs F.Model      R2 Pr(>F)
site       3   0.90033 0.30011  2.1932 0.60198  0.021 *
depth      1   0.18477 0.18477  1.3503 0.12354  0.233
Residuals  3   0.41052 0.13684         0.27448
Total      7   1.49562                1.00000
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
> adonis_zone&depth <-adonis(community.matrix ~ zone+depth, data=variables_zone,
permutations =999, method = "bray", strata = NULL)
> adonis_zone&depth
Call:
adonis(formula = community.matrix ~ zone + depth, data = variables_zone,      per
mutations = 999, method = "bray", strata = NULL)
Permutation: free
Number of permutations: 999
Terms added sequentially (first to last)
             Df SumsOfSqs MeanSqs F.Model      R2 Pr(>F)
   zone       1   0.58058 0.58058  3.9752 0.38819  0.004 **
   depth      1   0.18477 0.18477  1.2651 0.12354  0.247
   Residuals  5   0.73027 0.14605         0.48827
   Total      7   1.49562                 1.00000
   ---
   Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
> adonis_zone <- adonis(community.matrix ~ zone, data = variables_zone,
permutations = 999, method = "bray")
> adonis_zone
Call:
adonis(formula = community.matrix ~ zone, data = variables_zone, permutations =
999, method = "bray")

Permutation: free
Number of permutations: 999

Terms added sequentially (first to last)
```

```
          Df SumsOfSqs MeanSqs F.Model      R2 Pr(>F)
zone       1   0.58058 0.58058   3.807 0.38819  0.029 *
Residuals  6   0.91504 0.15251         0.61181
Total      7   1.49562                1.00000
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Function adonis considering 3 zones:

```
> adonis_zones3 (formula = community.matrix ~ zone, data = variables_zone3, permu
tations = 999, method = "bray")
> adonis_zones3

Permutation: free
Number of permutations: 999


Terms added sequentially (first to last)


          Df SumsOfSqs MeanSqs F.Model     R2 Pr(>F)
zone       2   0.85161  0.4258  3.3059 0.5694   0.01 **
Residuals  5   0.64401  0.1288         0.4306
Total      7   1.49562                1.0000
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1



> adonis_zones3&depth <- adonis(community.matrix ~ zone+depth, data = variables_z
one3, permutations = 999, method = "bray", strata = NULL)
> adonis_zones3&depth

Call:
adonis(formula = community.matrix ~ zone + depth, data = variables_zone3,      pe
rmutations = 999, method = "bray", strata = NULL)

Permutation: free
Number of permutations: 999

Terms added sequentially (first to last)

          Df SumsOfSqs MeanSqs F.Model      R2 Pr(>F)
zone       2   0.85161 0.42580  3.7088 0.56940  0.007 **
depth      1   0.18477 0.18477  1.6094 0.12354  0.153
Residuals  4   0.45924 0.11481         0.30706
Total      7   1.49562                 1.00000
---
```

```
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Supplement 9 References

ABP Research & Consultancy, L. (2000). "Humber geomorphological studies - Stage 2: 3-D Modelling of Flows, Salinity and Sediment Transport in the Humber Estuary". (Southampton).

Anderson, M.J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology* 26**,** 32-46. doi: 10.1111/j.1442-9993.2001.01070.pp.x

Barnes, J., and Owens, N.J.P. (1998). Denitrification and nitrous oxide concentrations in the Humber estuary, UK, and adjacent coastal zones. *Marine Pollution Bulletin* 37**,** 247-260. http://dx.doi.org/10.1016/S0025-326X(99)00079-X

Burke, I.T., Boothman, C., Lloyd, J.R., Mortimer, R.J.G., Livens, F.R., and Morris, K. (2005). Effects of Progressive Anoxia on the Solubility of Technetium in Sediments. *Environmental Science & Technology* 39**,** 4109-4116.

Chao, A., Gotelli, N.J., Hsieh, T.C., Sander, E.L., Ma, K.H., Colwell, R.K., and Ellison, A.M. (2014). Rarefaction and extrapolation with Hill numbers: a framework for sampling and estimation in species diversity studies. *Ecological Monographs* 84**,** 45-67. doi: 10.1890/13-0133.1

Freestone, D., Jones, N., North, J.,Pethick, J., Symes, D., and Ward, R. (1987). "The Humber Estuary, Environmental Background". Institute of Estuarine and Coastal Studies).

Fujii, T., and Raffaelli, D. (2008). Sea-level rise, expected environmental changes, and responses of intertidal benthic macrofauna in the Humber estuary, UK. *Marine Ecology Progress Series* 371**,** 23-35. doi: 10.3354/meps07652

Garcia-Alonso, J., Greenway, G.M., Munshi, A., Gomez, J.C., Mazik, K., Knight, A.W., Hardege, J.D., and Elliott, M. (2011). Biological responses to contaminants in the Humber Estuary: Disentangling complex relationships. *Marine Environmental Research* 71**,** 295-303. doi: 10.1016/j.marenvres.2011.02.004

Hill, M.O. (1973). Diversity and evenness: a unifying notation and its consequences. *Ecology* 54**,** 427-432. doi: 10.2307/1934352

Jost, L. (2006). Entropy and diversity. *Oikos* 113**,** 363-375. doi: 10.1111/j.2006.0030-1299.14714.x

Jost, L. (2007). Partitioning diversity into independent alpha and beta components. *Ecology* 88**,** 2427-2439. doi: 10.1890/06-1736.1

Kang, S., Rodrigues, J.L.M., Ng, J.P., and Gentry, T.J. (2016). Hill number as a bacterial diversity measure framework with high-throughput sequence data. *Scientific Reports* 6. doi: 10.1038/srep38263

MacArthur, R.H. (1965). Patterns of species diversity. *Biological reviews* 40**,** 510-533.

Millward, G.E., Sands, T.K., Nimmo, M., Turner, A., and Tappin, A.D. (2002). Nickel in the Humber plume: Influences of particle dynamics and reactivity. *Estuarine Coastal and Shelf Science* 54**,** 821-832. doi: 10.1006/ecss.2001.0859

Mitchell, S.B., West, R.J., Arundale, A.M.W., Guymer, I., Couperthwaite, J.S. (1998). Dynamic of the Turbidity Maxima in the Upper Humber Estuary System, UK. *Marine Pollution Bulletin* 37**,** 190-205.

Mortimer, R.J.G., Davey, J.T., Krom, M.D., Watson, P.G., Frickers, P.E., and Clifton, R.J. (1999). The effect of macrofauna on porewater profiles and nutrient fluxes in the intertidal zone of the Humber Estuary. *Estuarine Coastal and Shelf Science* 48**,** 683-699. doi: 10.1006/ecss.1999.0479

NRA (1995). "Sea Vigil Water Quality Monitoring : the Humber Estuary 1992-1993", (ed.) M. Section. (Peterborough: National Rivers Authority Anglian Region). Retrieved from http://www.environmentdata.org/archive/ealit:2851

NRA (1996). "Sea Vigil Water Quality Monitoring: The Humber Estuary 1994", (ed.) M. Section. (Peterborough: National Rivers Authority Anglian Region). Retrieved from http://www.environmentdata.org/archive/ealit:2868

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., and Wagner, H. (2013). *vegan: Community Ecology Package. R package version 2.0-10* [Online]. Available: http://CRAN.R-project.org/package=vegan [Accessed 22-08-2016].

Prastka, K.E., and Malcolm, S.J. (1994). Particulate phosphorus in the Humber estuary. *Journal of Aquatic Ecology* 28**,** 397:403. doi: 10.1007/BF02334209

Sanders, R.J., Jickells, T., Malcolm, S., Brown, J., Kirkwood, D., Reeve, A., Taylor, J., Horrobin, T., and Ashcroft, C. (1997). Nutrient Fluxes through the Humber estuary. *Journal of Sea Research* 37**,** 3-23.

Team, R. (2015). "RStudio: Integrated Development Environment for R". .0.99.486 ed. (Boston, MA: RStudio, Inc). Retrieved from http://www.rstudio.com/

Uncles, R.J., Joint, I., and Stephens, J.A. (1998). Transport and retention of suspended particulate matter and bacteria in the Humber-Ouse Estuary, United Kingdom, and their relationship to hypoxia and anoxia. *Estuaries* 21**,** 597-612. doi: 10.2307/1353298

Uncles, R.J., Stephens, J.A., and Law, D.J. (2006). Turbidity maximum in the macrotidal, highly turbid Humber Estuary, UK: Flocs, fluid mud, stationary suspensions and tidal bores. *Estuarine, Coastal and Shelf Science* 67**,** 30-52. doi: 10.1016/j.ecss.2005.10.013

Whittaker, R.H. (1972). Evolution and measurement of species diversity. *Taxon* 21**,** 213-251. doi: 10.2307/1218190

Williams, M.R., and Millward, G.E. (1999). Dynamics of particulate trace metals in the tidal reaches of the Ouse and Trent, UK. *Marine Pollution Bulletin* 37**,** 306-315. doi: 10.1016/s0025-326x(99)00054-5