## SUPPLEMENTARY INFORMATION

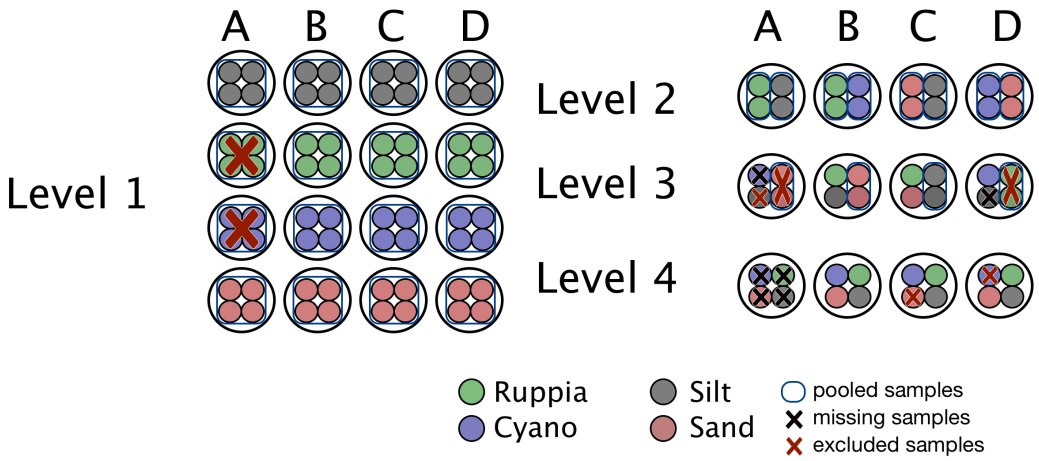


**Figure S1.** A schematic view of the experimental design used in Alsterberg et al. (2017). For that study, a total of 112 cores were sampled in four different habitat types, denominated 'Sandy beach', 'Silty mud', 'Cyanobacterial mats' and '*Ruppia maritima* meadows' for their main characteristic. The cores were placed in large cylinders (25 cm inner diameter) in groups of 4 and within each cylinder the cores shared a common water column. As Alsterberg et al. (2017) investigated the effects of habitat diversity, the cores were grouped in combinations of different habitat richness (Level 1-4). For the measurements of nitrogen fixation and the DNA sampling, the samples from the same habitats within the same cylinder were pooled (as indicated by the blue outer lining). This resulted in a total of 52 samples. From the original 52 samples 6 are missing. Further, 6 samples were excluded as it became evident that they have been mislabeled (see (Fig. S2)), and 1 sample did not have enough DNA left for all analysis and was excluded, too. Hence the current study was based on the remaining 39 samples.

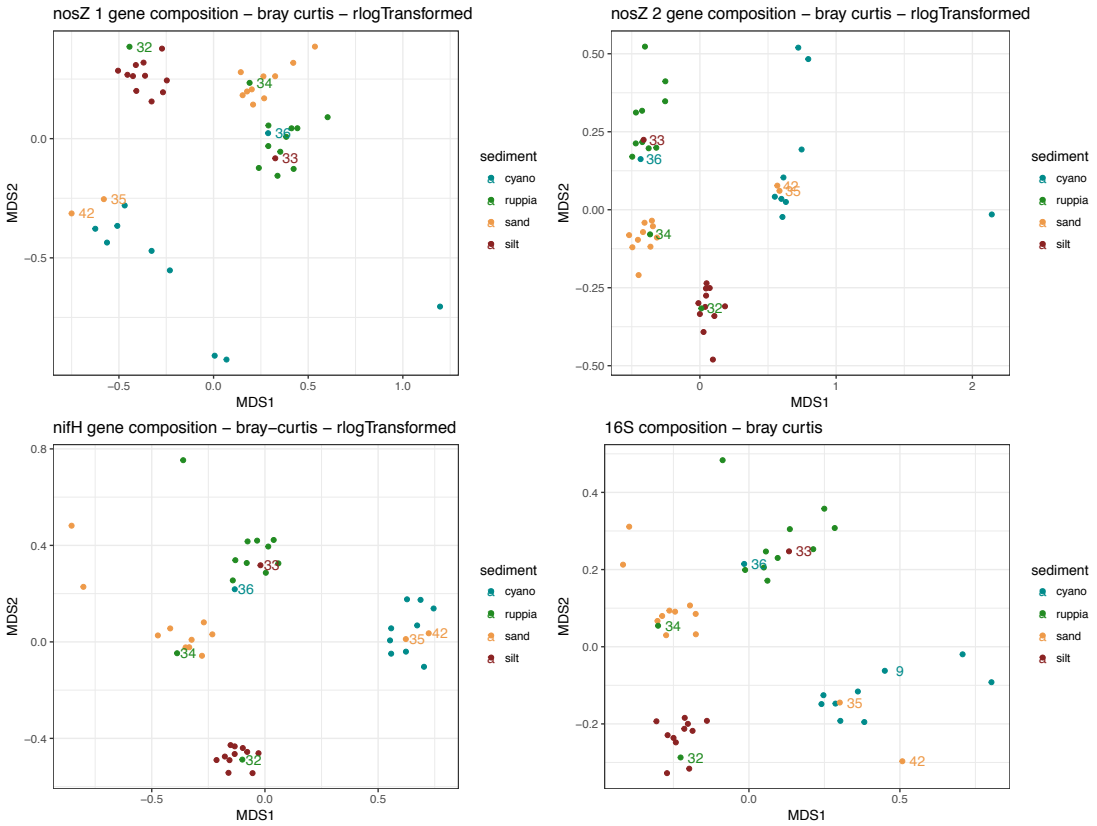


**Figure S2.** Non metric multidimensional scaling of the samples based on 4 marker genes that make it clear that 6 samples have been misassigned. Clustering was done as in (Fig. 3). The four marker genes are 16S, *nifH* (included in this study) and *nosZ I* and *II* not included here. The marker genes have all been extracted, sequenced and analysed separately but show a congruent pattern. Based on the evidence above, the concerned samples have been excluded.

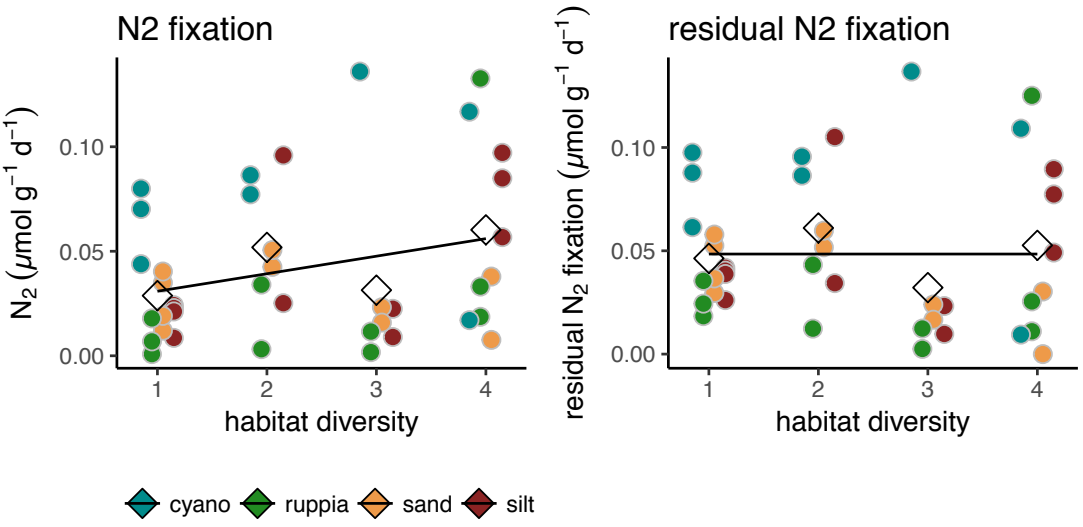


**Figure S3.** Extraction of residual nitrogen fixation. Left: Nitrogen fixation by habitat diversity level as measured in Alsterberg et al. (2017). Right: Residual nitrogen fixation after accounting for habitat diversity by fitting a linear model of the form *nitrogen fixation* $\sim$ *habitat diversity* to the data in the left panel and extracting the residuals from that model. The residuals were brought back to original scale by adding the absolute minimum residual to all values. The solid black line show the linear trend line before (left) and after (right) accounting for habitat diversity, the white diamonds show the mean values by habitat diversity level. All analysis in this study have been performed with the residual nitrogen fixation rate as response variable.

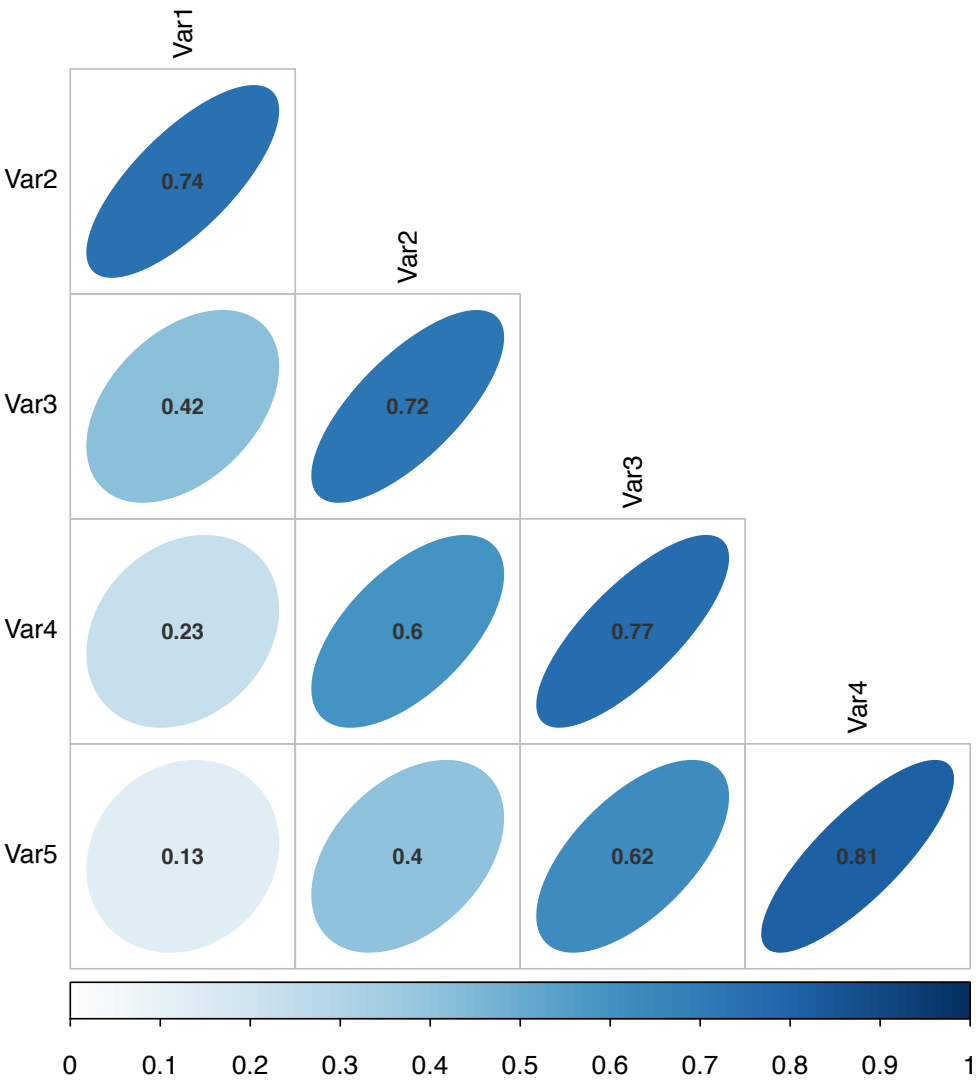


**Figure S4.** Simulation showing that correlation can degrade quickly between linked variables even if each variable is highly correlated to the previous and the next. For this example we simulated 50 values from a normal distribution (*Var*1, mean = 0, standard deviation = 1) and subsequently randomly generated 50 values with a correlation of ∼0.7 to the first set of values (*Var*2). This step was repeated for *Var*3, taking *Var*2 as reference *Var*4 taking *Var*3 as reference and *Var*5, referring to *Var*4. The correlation matrix above gives the pearson correlation coefficient of each pairwise correlation, showing that the first (*Var*1) and the last (*Var*5) variables are no longer tightly linked. The Simulation was performed as a thought experiment for what connection should be expected between two variables (*Var*1 & *Var*5) that are linked by multiple steps as in $Var1 \rightarrow Var2 \rightarrow Var3 \rightarrow Var4 \rightarrow Var5$, where each following step is hypothesised to be closely linked to the previous one.

**Table S1.** Primers and thermal cycling conditions for quantification and sequencing of 16S rRNA and *nifH* genes.

| **Genes** / Primer names | **Sequences** (5'-3') | **References** | **Thermal cycling** | **Efficiency** (qPCR ) |
|---|---|---|---|---|
| ***nifH*** | | | (95°C, 3 min) x 1 | |
| IGK3 | GCIWTHTAYGGIAARGGIGGIATHGGIAA | Ando et al. (2005) | (95°C, 15s; 58°C, 30s; 72°C, 30s; 80°C, 10 s) x 35 | 75% |
| DVV | ATIGCRAAICCICCRCAIACIACRTC | | (95°C, 15s;(60°C to 95°C, 10s, increment 0.5°C)), x 1 | |
| **16S rRNA (sequencing)** | | | (98°C, 3min) x 1 | |
| Pro341F | CCTACGGGNBGCASCAG | Takahashi et al. (2014) | (98°C, 30s; 55°C, 30s; 72°C, 30s) x 25+8 | |
| Pro805R | GACTACNVGGGTATCTAATCC | | (72°C , 10min) x 1 | |
| ***nifH*** **(sequencing)** | | | (95°C, 3min) x 1 | |
| IGK3 | GCIWTHTAYGGIAARGGIGGIATHGGIAA | Ando et al. (2005) | (95°C, 30s; 55°C, 30s; 72°C, 30s) x 30+8 | |
| DVV | ATIGCRAAICCICCRCAIACIACRTC | | (72°C , 10 min) x 1 | |

**Table S2.** statistics for OTUs that have been identified as significantly correlating with nitrogen fixation rates

| **OTU** | **padj** | **baseMean** | **log2FoldChange** |
|---|---|---|---|
| 16S OTU 1005 | 0.0000157 | 13.019125 | 8.128844 |
| 16S OTU 1693 | 0.0452508 | 9.650296 | 4.334838 |
| 16S OTU 2155 | 0 | 21.773348 | 10.350314 |
| 16S OTU 6555 | 0.0065902 | 3.719744 | 5.902401 |
| 16S OTU 1825 | 0.0000103 | 16.289607 | -9.242262 |
| 16S OTU 1869 | 0.0390275 | 4.824531 | -5.632413 |
| 16S OTU 3512 | 0.0390275 | 11.919685 | -6.260393 |
| 16S OTU 291 | 0.0007995 | 2.466452 | 6.625404 |
| *nifH* OTU 2364 | 0.004627 | 8.344701 | 3.844505 |
| 16S allSed OTU 1825 | 0.0096083 | 6.244994 | -8.166978 |