

Ecological coherence of diversity patterns derived from classical fingerprinting and Next Generation Sequencing techniques

Angélique Gobet, Antje Boetius, Alban Ramette

Supplementary figures

Fig. S1. Total OTU numbers along sediment depth or over time for ARISA, OTU_{all}, and PyroNoise_{3%} data sets.

Fig. S2. Alpha-diversity comparison of different data sets.

Fig. S3. Percentage of shared OTU of the bacterial community between sediment depth layers or sampling dates after correction and OTU clustering of the 454 MPTS data set and of the ARISA data set.

Fig. S4. Percentage of shared OTU of the bacterial community between sediment depth layers or sampling dates at successive taxonomic levels.

Fig. S5. Percentage of shared OTU of the bacterial community between sediment depth layers or sampling dates after applying MultiCoLA.

Fig. S6. Comparison of extracted variation from the different categories of data sets.

Fig. S7. Partitioning of the biological variation in the bacterial community structure

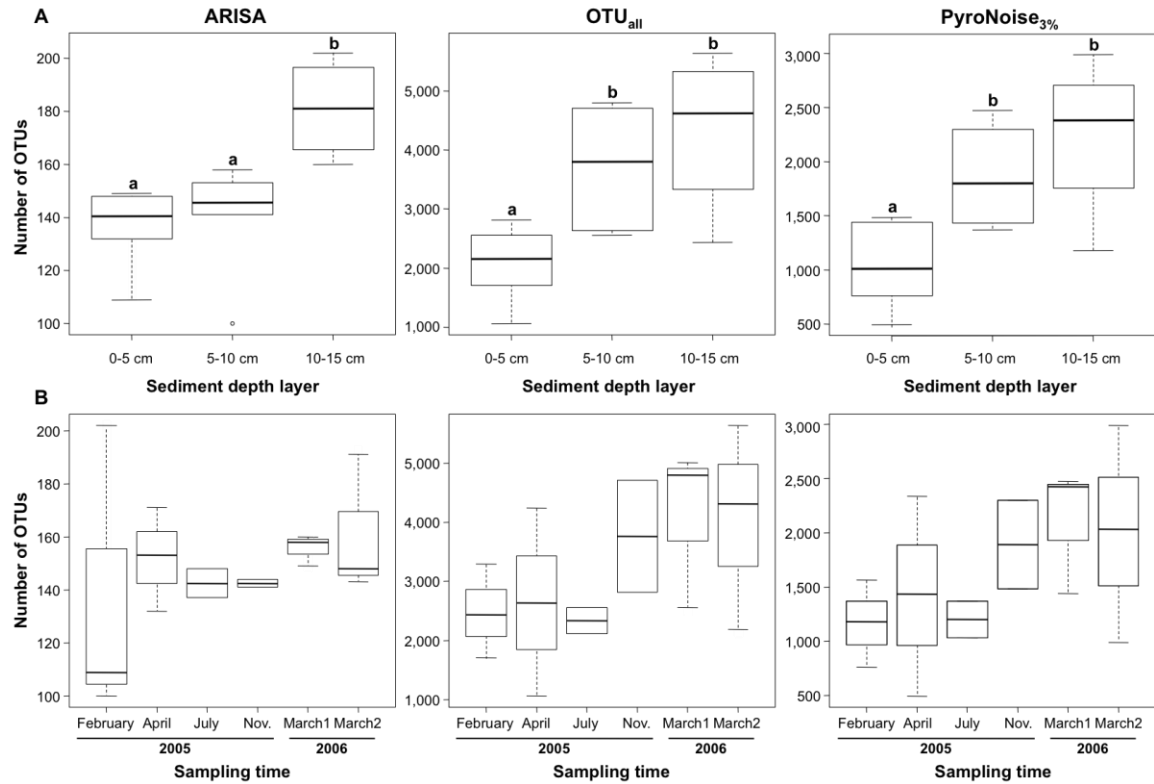


Fig. S1. Total OTU numbers along sediment depth or over time for ARISA, OTU_{all}, and PyroNoise_{3%} data sets. Horizontal lines outside of the box represent the smallest and the largest observations in the data set. The first and third quartiles are indicated by the lowest and highest limits of the box, respectively. The median is indicated by the thick bar in the middle of each box. **(A)** Boxplots were calculated at all sampling times along the three depth layers. A different letter (**a**, **b**) indicates a significant mean difference in OTU number between the selected depth layer and the other(s) (Student t-test, P value ≤ 0.05). There were 6 sampling dates considered for the upper 10 cm layers and 4 sampling times for the 10-15 cm layer. In **(A)**, the circle represents an outlier: a data point inferior or superior to the first or third quartile, respectively, by 1.5 times the interquartile range. **(B)** Boxplots were calculated at all depths over time. The 3 depth layers were considered except for July and November 2005 where 2 upper depth layers were available. There was no significant mean difference in OTU number between sampling times (Student t-test).

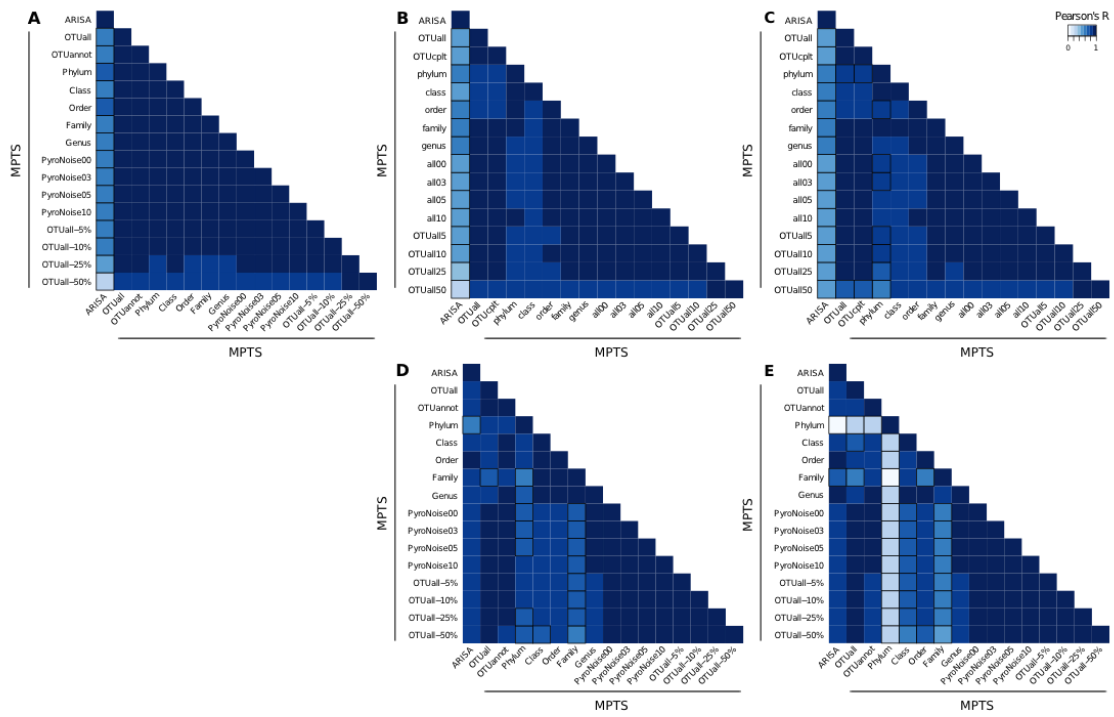


Fig. S2. Alpha-diversity comparison of different data sets. Pearson's correlation coefficient was used to compare (A) the total number of OTUs, (B) Chao's estimators, (C) ACE estimators, (D) Shannon's diversity index, and (E) Simpson's evenness calculated from ARISA and various resolution levels of 454 MPTS data sets. The correlation coefficient was calculated for all samples after 1,000 resampling with replacement. All R values with black edges were not significant after Bonferroni correction for multiple testing. MPTS, Massively Parallel Tag Sequencing.

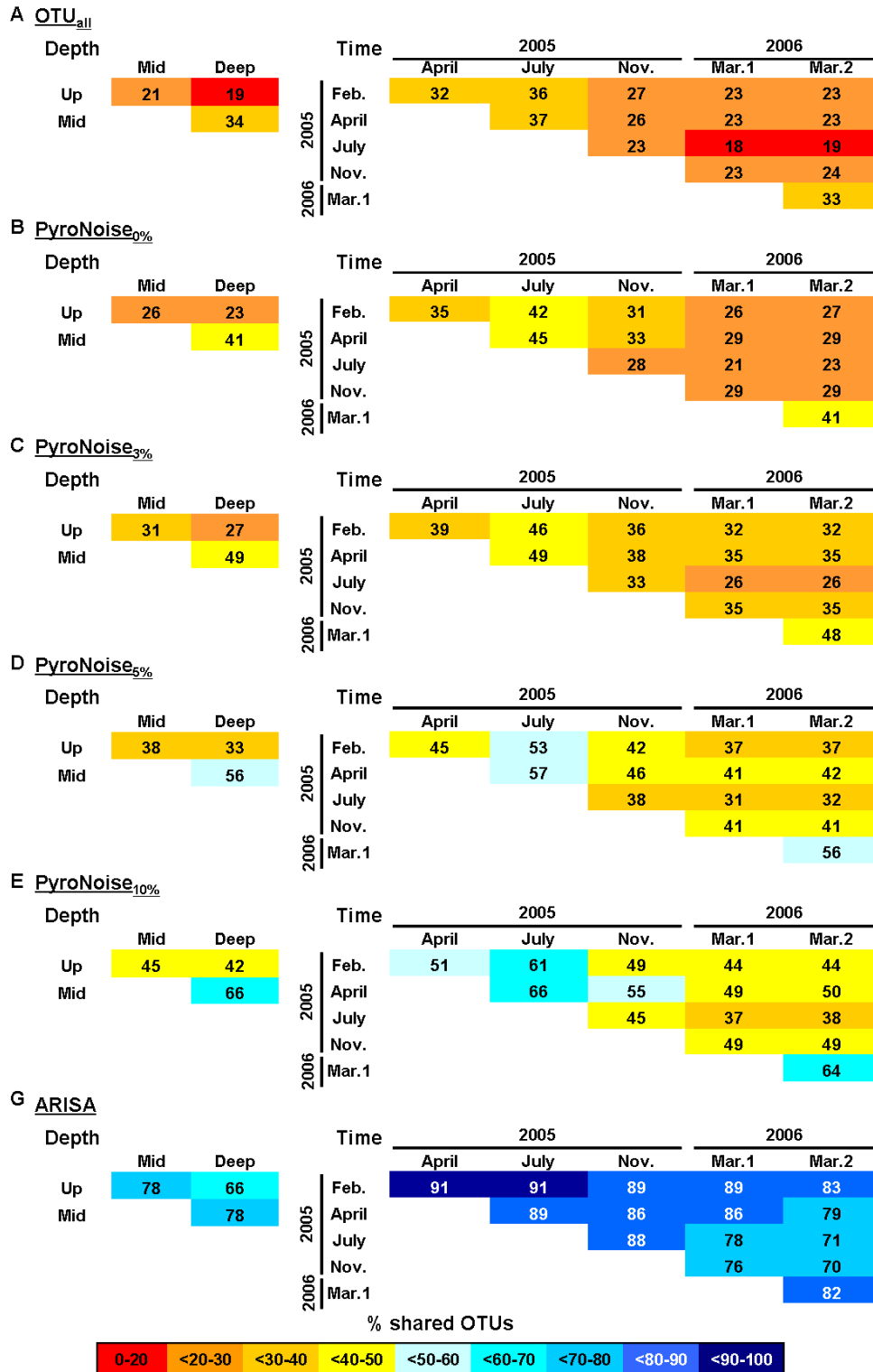


Fig. S3. Percentage of shared OTU of the bacterial community between sediment depth layers or sampling dates after correction and OTU clustering of the 454 MPTS data set and of the ARISA data set. (A) OTU_{all}, (B) PyroNoise unique, (C) PyroNoise 3%, (D) PyroNoise 5%, (E) PyroNoise 10%, and (F) ARISA. The percentage of OTU shared between a sampling depth (or date) and the previous one was calculated and values were represented by heatmap matrices. OTU_{all} represents the original data set with all OTU_{unique}. Figure modified from (Gobet, et al., 2012) for a direct comparison of the techniques.

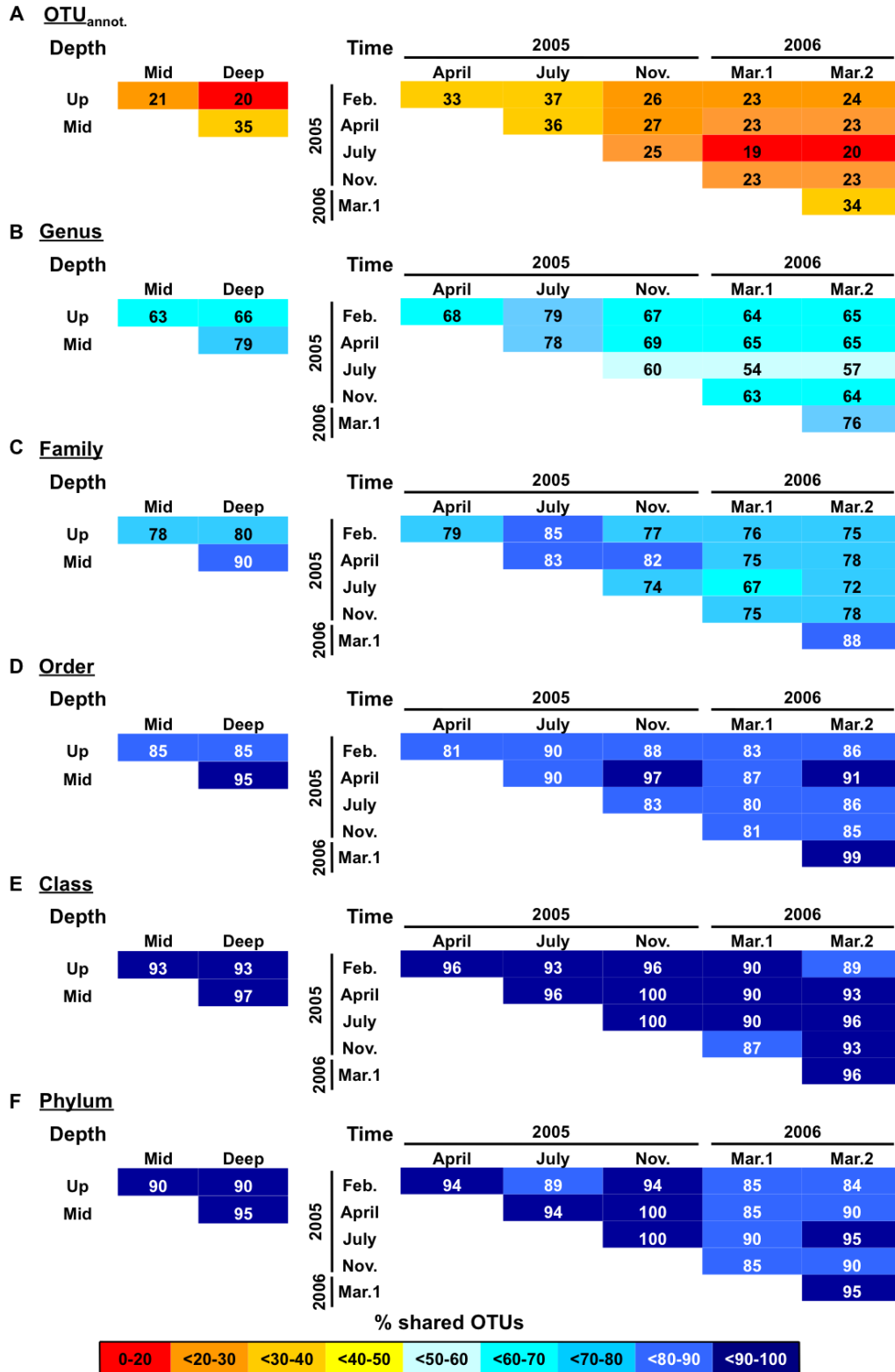


Fig. S4. Percentage of shared OTU of the bacterial community between sediment depth layers or sampling dates at successive taxonomic levels. (A) OTU_{annot.}, (B) Genus, (C) Family, (D) Order, (E) Class, and (F) Phylum levels. OTU_{annotated} represents OTU_{unique} that are completely annotated (phylum to genus levels). See Fig. S2 for further details.

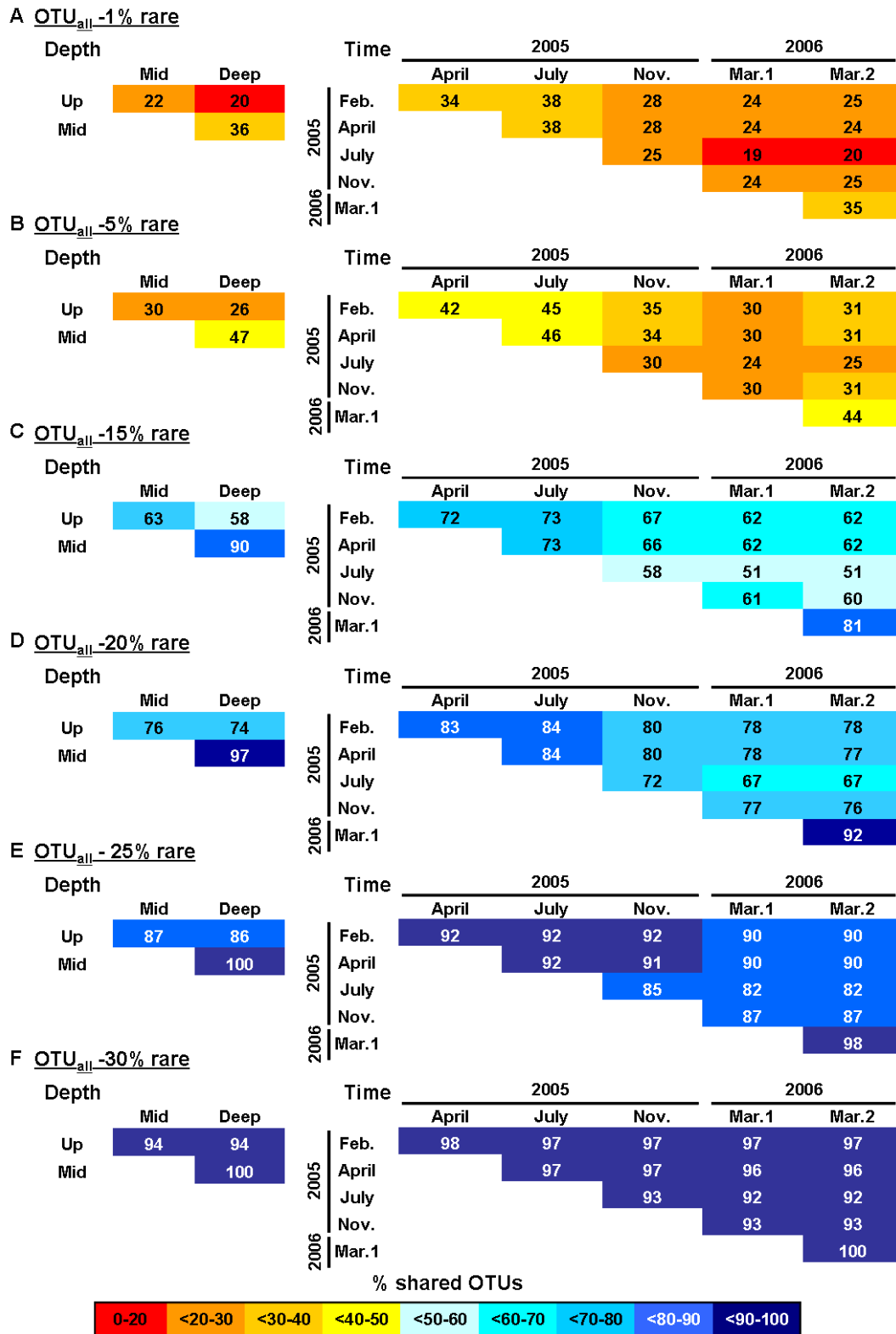


Fig. S5. Percentage of shared OTU of the bacterial community between sediment depth layers or sampling dates after applying MultiCoLA. Successive percentages of rare OTU were removed from the OTU_{all} data set: (A) 1%, (B) 5%, (C) 15%, (D) 20%, (E) 25%, and (F) 30%. See Fig. S2 for further details. Figure modified from (Gobet, et al., 2012) for a direct comparison of the techniques.

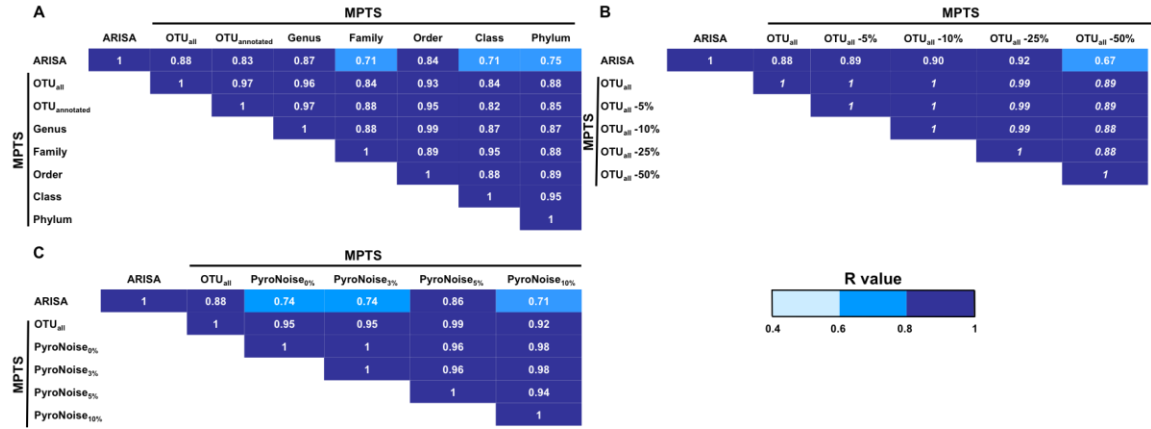


Fig. S6. Comparison of extracted variation from the different categories of data sets. The Procrustes' correlation coefficient was used to compare the ARISA and the OTU_{all} extracted variation with (A) various levels of taxonomic annotation, (B) successive removal of the rare OTU and (C) successive clustering of the PyroNoise-corrected data to define OTU. The Procrustes' correlation coefficient was calculated from the variation in the main axes of extracted variation via non-metric multidimensional scaling [NMDS], based on the distance matrices resulting from the relative sequence abundances. MPTS, Massively Parallel Tag Sequencing. All R values were significant after Bonferroni correction for multiple testing. For (B) values *in italic* indicate no test of significance as the truncated matrices are not statistically independent from each other (see Materials and Methods).

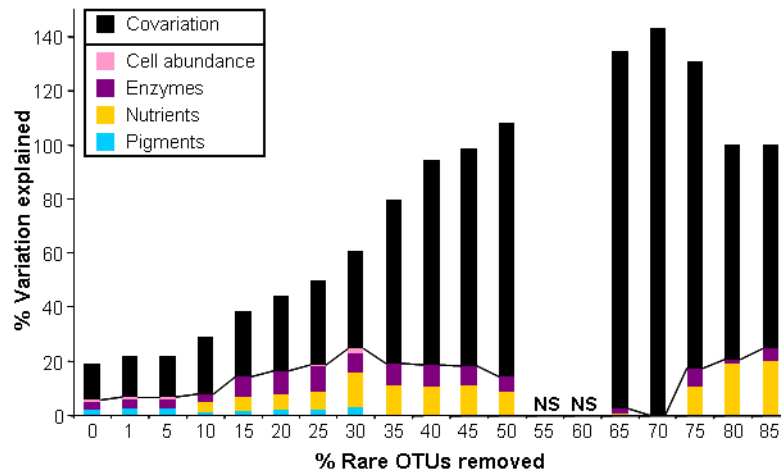


Fig. S7. Partitioning of the biological variation in the bacterial community structure based on the removal of successive percentages (here 0 to 85%) of rare OTU from the OTU_{all} data set by using MultiCoLA. Environmental parameters accounted for include pigments (chlorophyll *a* and pheophytin), nutrients (silicate, phosphate, nitrite, nitrate, ammonium), extra-cellular enzyme activities (chitinase, α -glucosidase, β -glucosidase, lipase, aminopeptidase, phosphatase), cell abundance and their combined effects. The black line in each panel separates the pure factor effects from their covariations. Covariation of any of the 4 environmental factors is represented under the category “covariation”. Negative values, unexplained variation and non-significant (NS) multivariate models are not shown. A non significant (NS) model appears when trying to explain the biological variation from data sets without 55-60% of rare OTU. This is probably not an artifact. Indeed, as an increasing amount of rare OTU is removed, the original structure of the data set may be lost. Hence, as rare OTU are removed from the data set, the amount of explained variation likely varies. The interesting point here is that similar patterns can be kept after a consequent amount of OTU is removed.