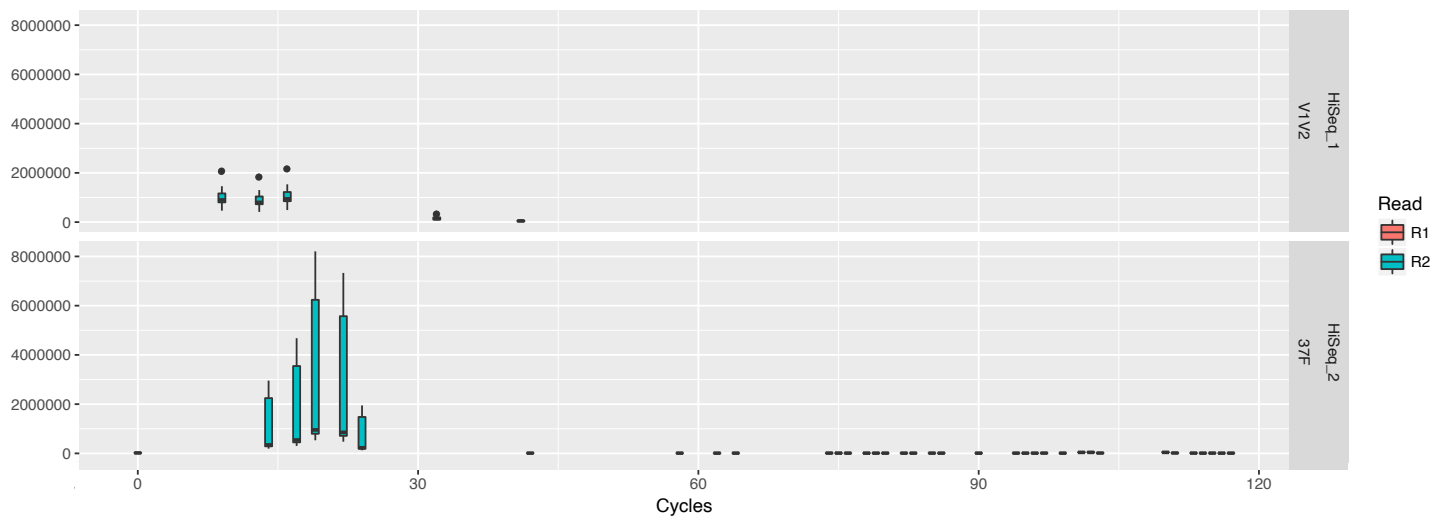
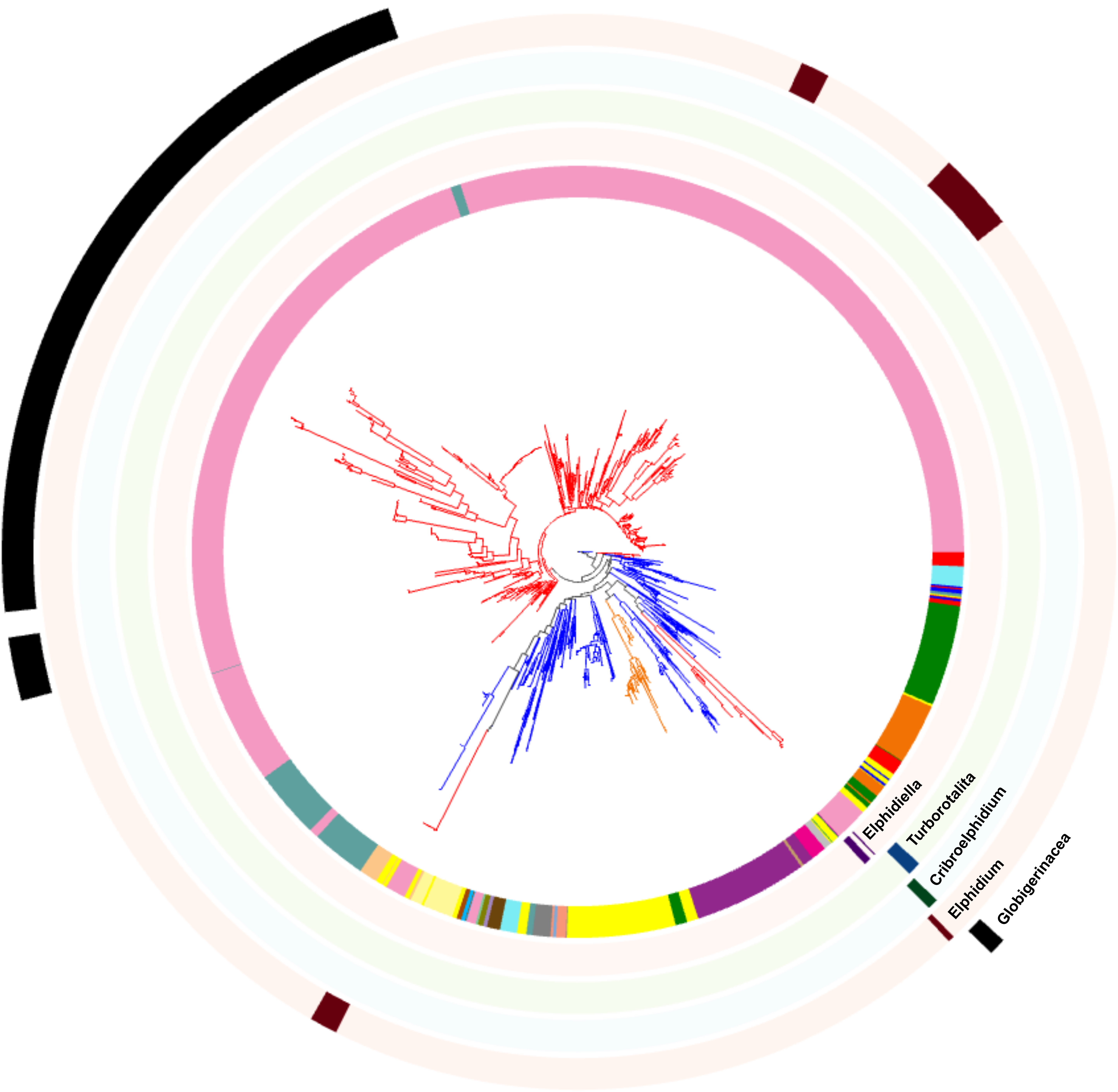


Number of ambiguous bases ("N")



**Supplementary Figure 1:** Frequencies of ambiguous “N” positions per sequencing cycle and HiSeq run. Each run is defined by the sequencing instrument name, the sequencing lane number and the sequenced marker.

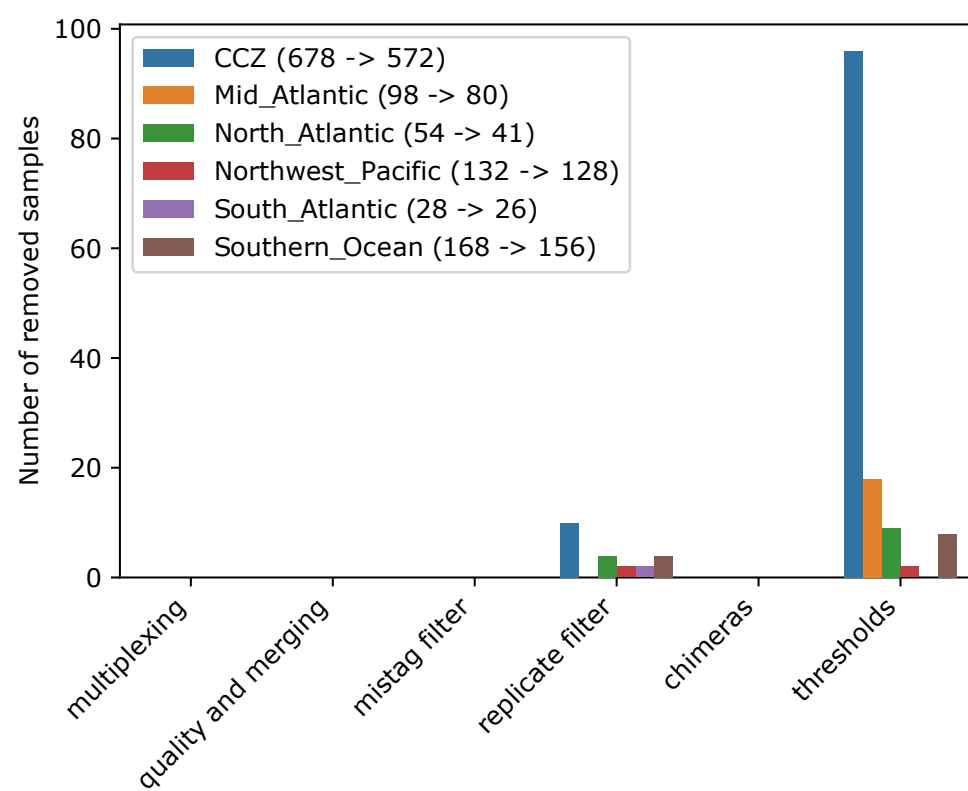


**Supplementary Figure 2:** Foraminifera phylogenetic tree built from the full 18S rRNA fragment and composed of 3,068 non-redundant, non-ambiguous database reference sequences. The branches are colored per Class, highlighting the few Globothalamea sequences branching with the distant clade Monothalamea. These sequences belonging to *Elphidiella*, *Turborotalia*, *Criboelphidium*, *Elphidium* of other planktonic Globigerinacea (see outer rings) have been removed from the reference database.

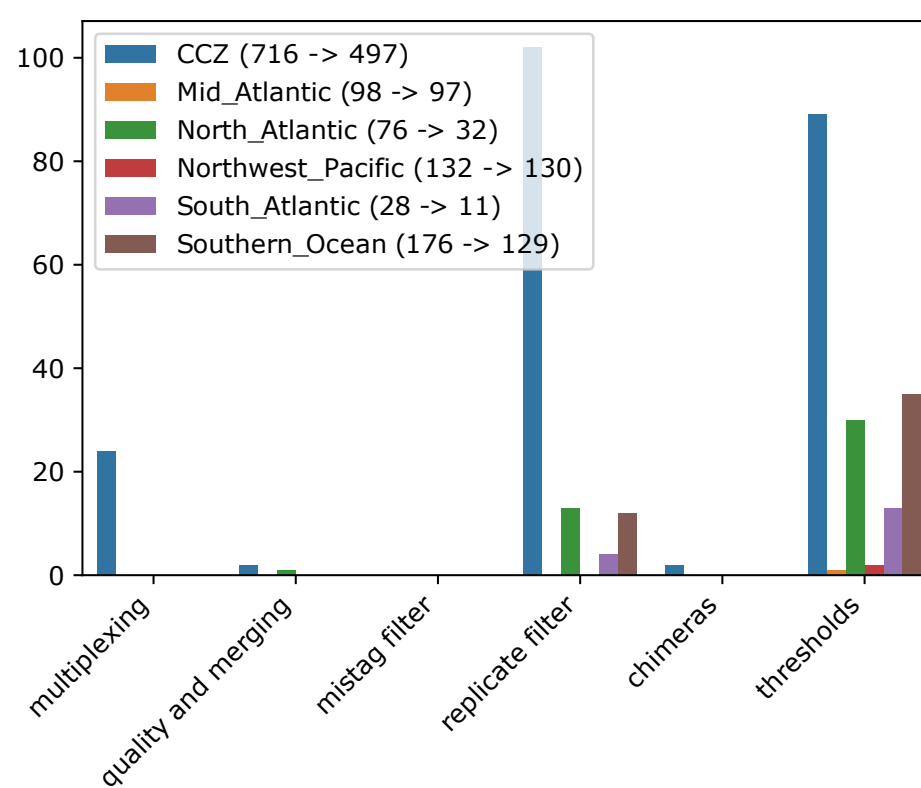


A

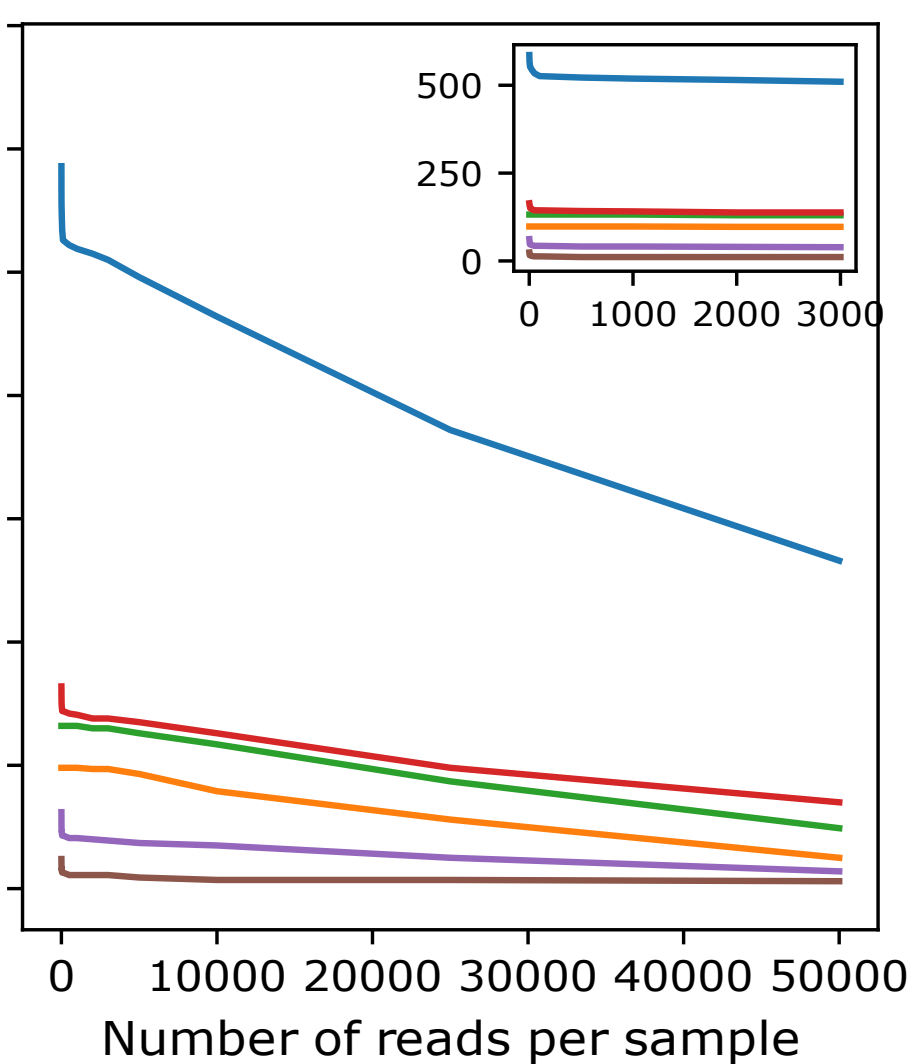
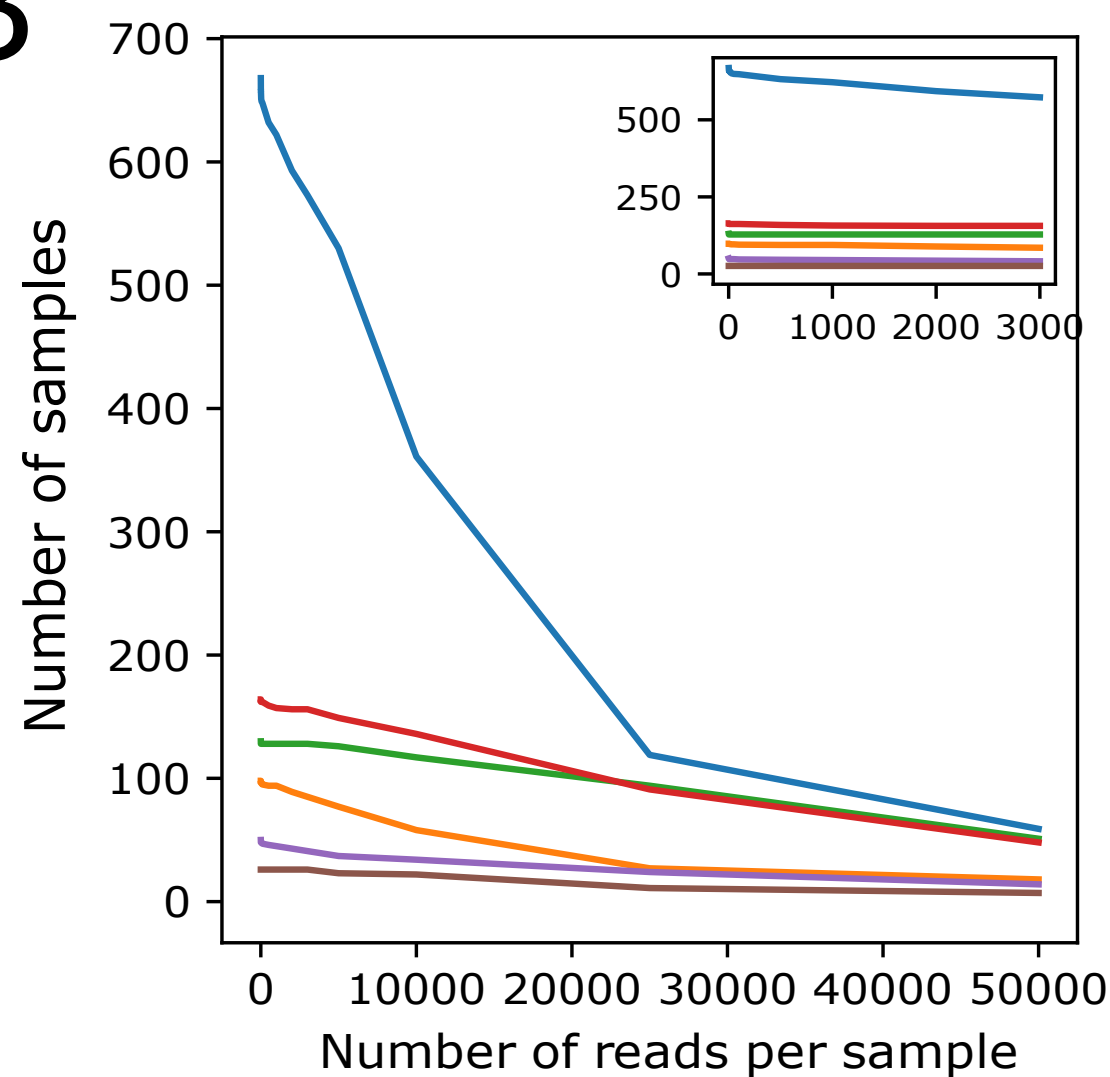
37F



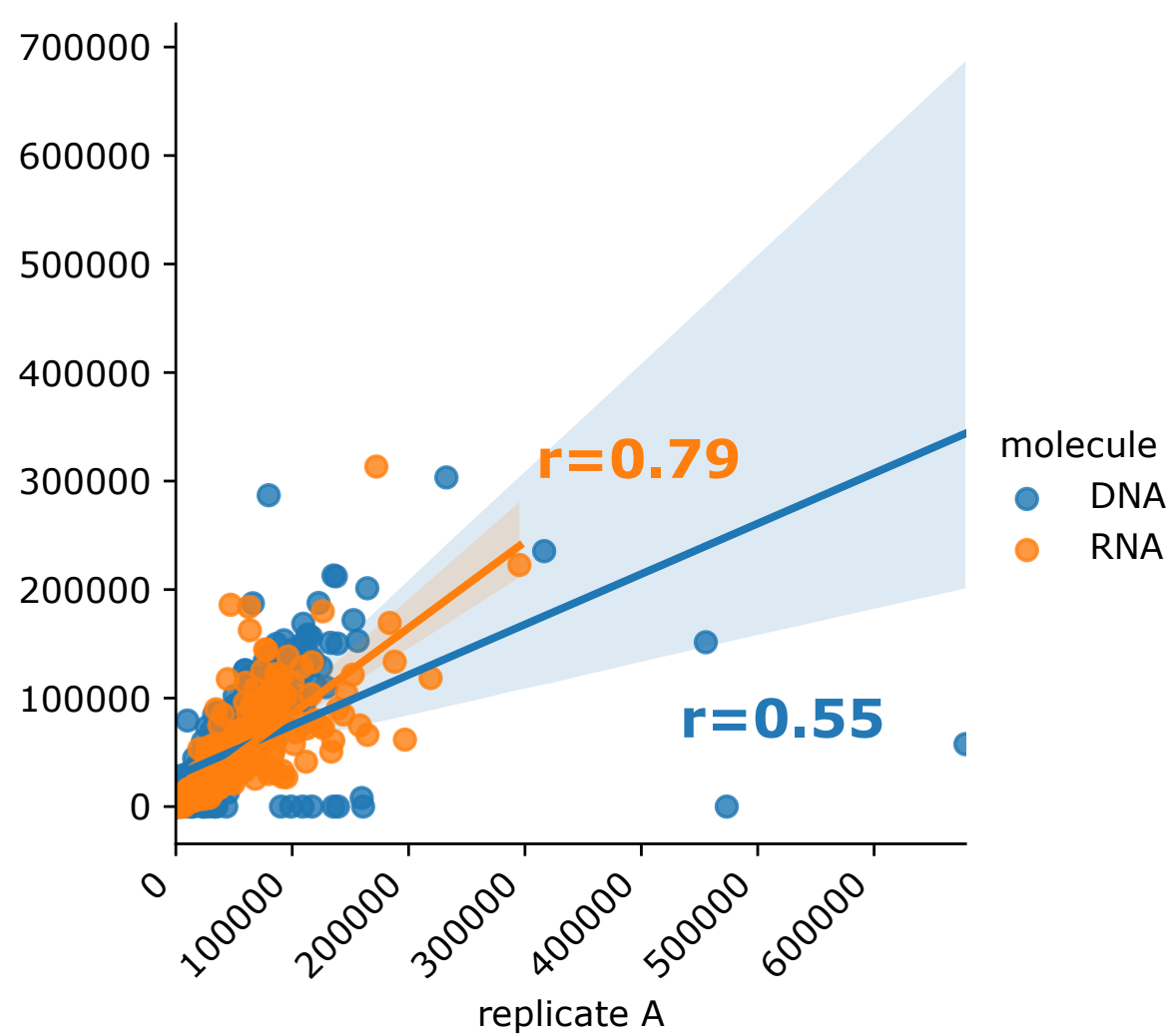
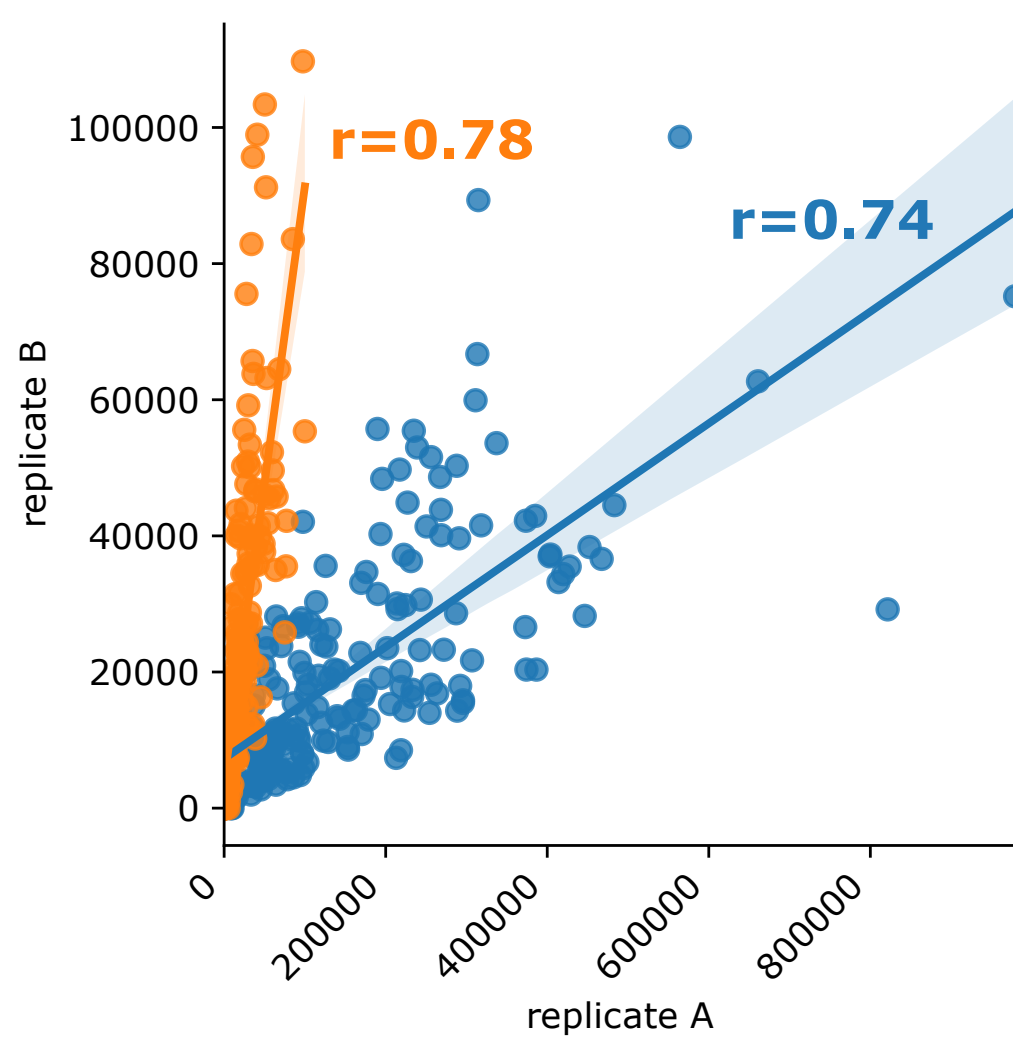
V1V2



B



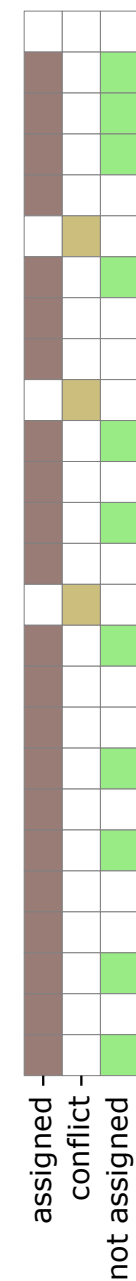
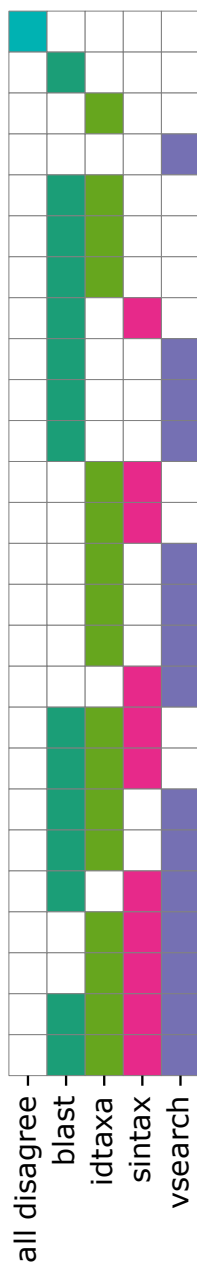
C



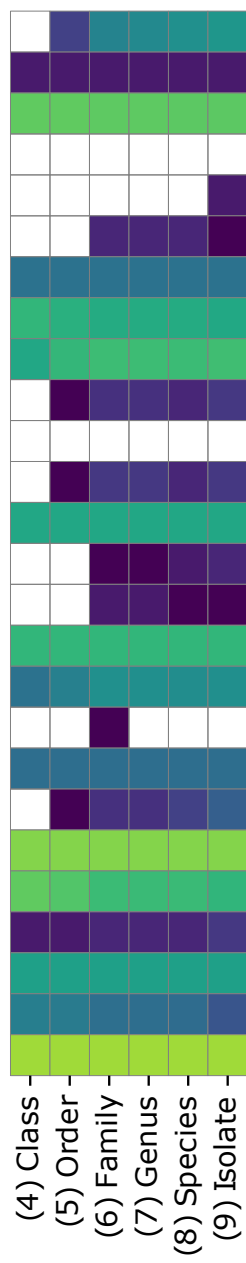
**Supplementary Figure 3:** Raw data filtering and technical features of the 37F (left panels) and V1V2 (right panels) datasets. (A) Number of samples removed after each filtering step. The filtering steps order from left to right corresponds to the sequential filtering procedure. Details are shown per area and the pairs of numbers in parentheses indicate the initial and final numbers of samples per area. (B) Numbers of PCR replicate samples per minimum sequencing depth. The numbers of samples that have sufficient numbers of reads are shown per area. Insets show the numbers of samples remaining for thresholds ranging from 0 to 3,000 reads per sample (the read depth chosen for rarefaction). (C) Numbers of sequence reads per PCR replicates pair. The lines represent the fitted linear regression model (with 95% confidence interval).

37F

full fragment

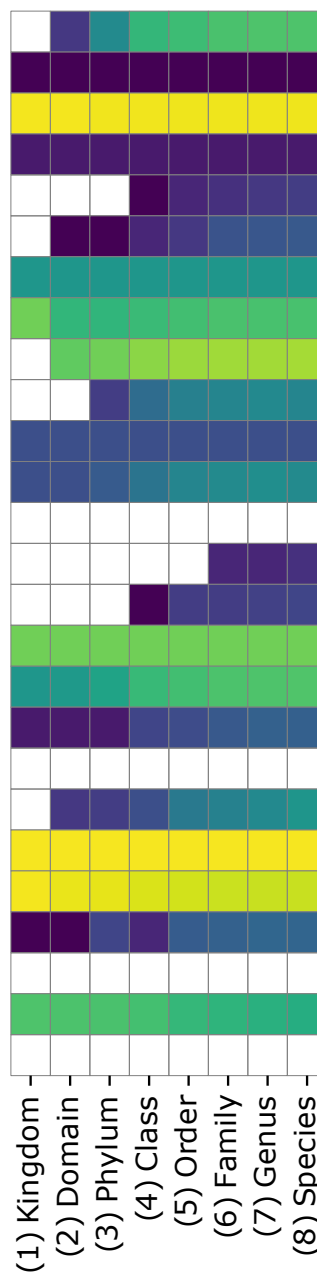


assignment

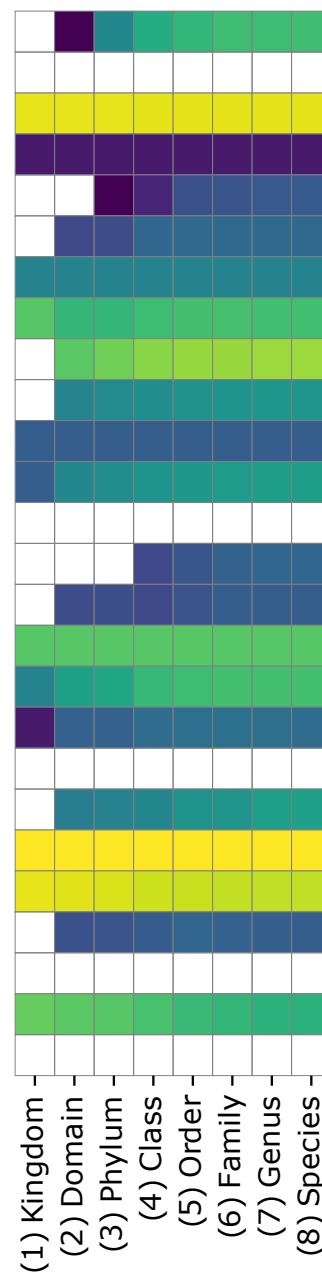


V1V2

PR2



SILVA



log10(Number of OTUs)

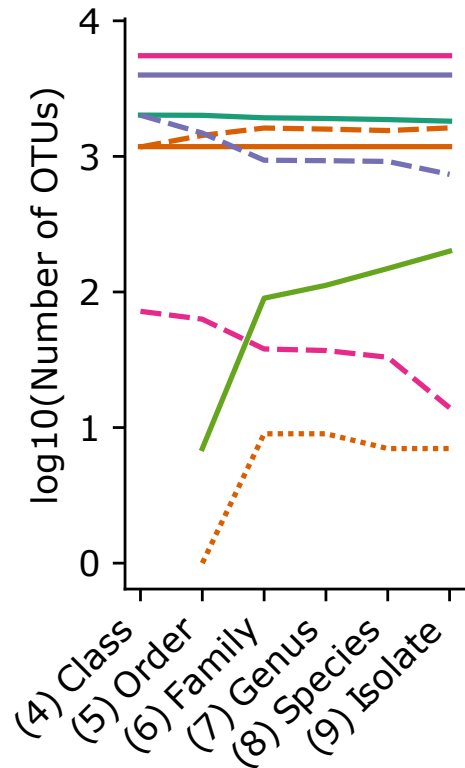


**Supplementary Figure 4:** Summary of the taxonomic assignments. For each marker (main panels, subdivided by database) and for each taxonomic level (x-axis) are shown the number of OTUs unassigned, assigned to a last common ancestor or left unassigned because of equally likely, conflicting assignments (left color encoding “assignment”). As this figure is a summary for different intersections of methods, a single OTU can be counted in multiple cells of this heatmap.



A

37F

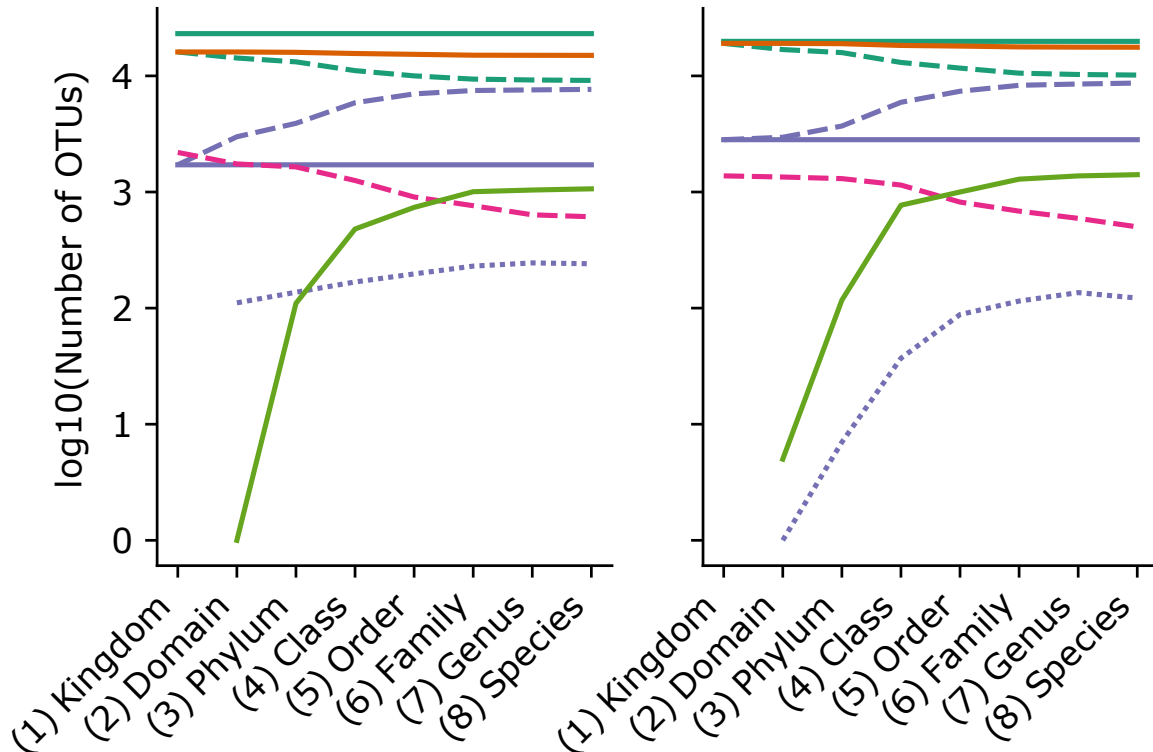


B

SILVA

V1V2

PR2



number of methods

3

1

2

4

all disagree

assignment

— not assigned

--- assigned

..... conflict

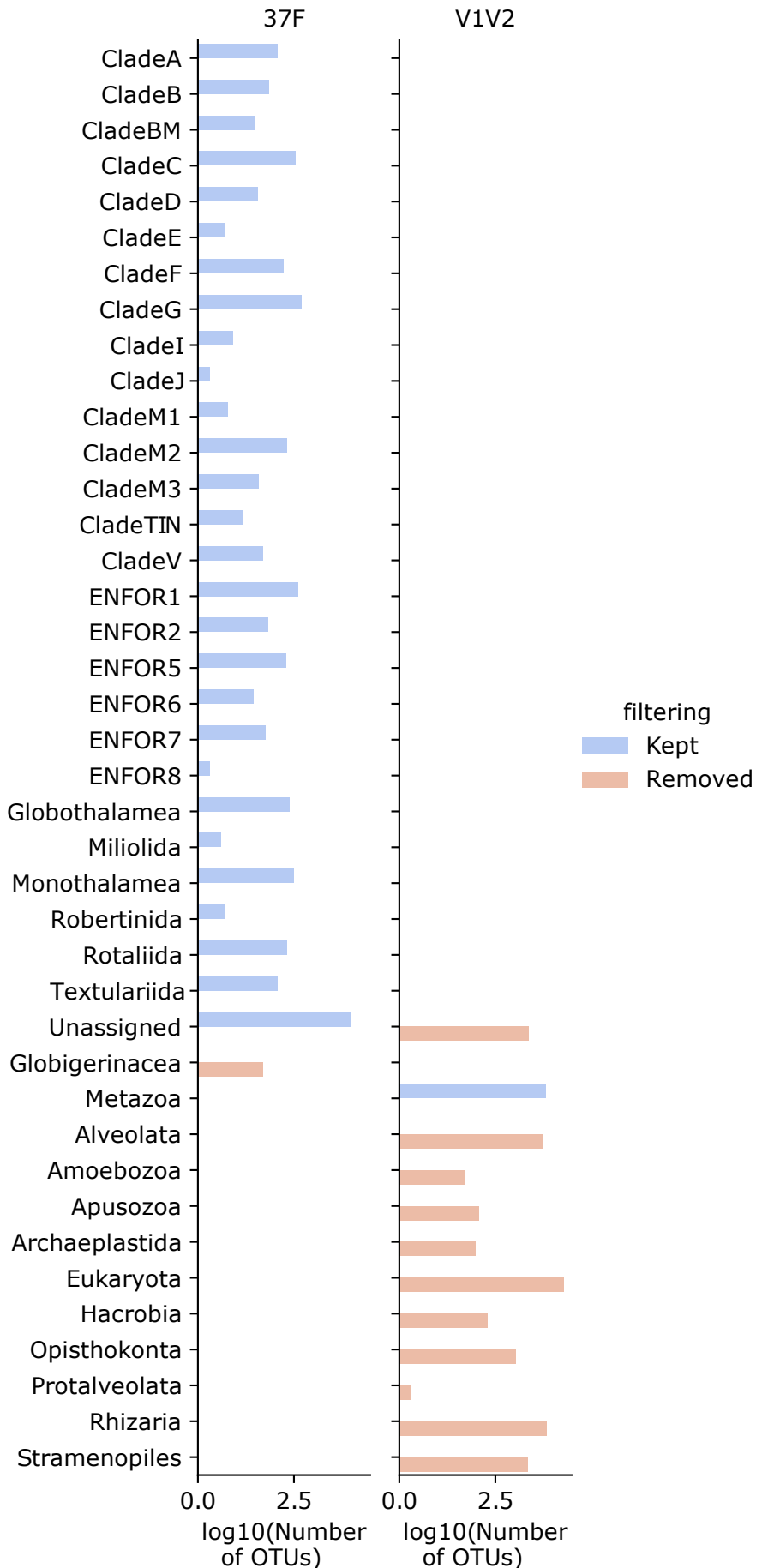
**Supplementary Figure 5: Assignment depths achieved using the different methods and databases.** The log<sub>10</sub>-transformed number of OTUs assigned by one to four assignment methods are shown at each taxonomic level (x-axis) and separately for the two databases (panels) used for (A) 37F and (B) V1V2. The line style indicates whether the OTUs were unassigned, assigned to a last common ancestor or left unassigned because of equally-likely conflicting assignments.



**Supplementary Figure 6:** Taxonomic composition expressed in terms of relative abundances of reads per taxon for the V1V2 “all taxa” dataset but at the Class level (in complement to Figure 2B). Each bar represents the samples collapsed at the area level, and grouped per region.



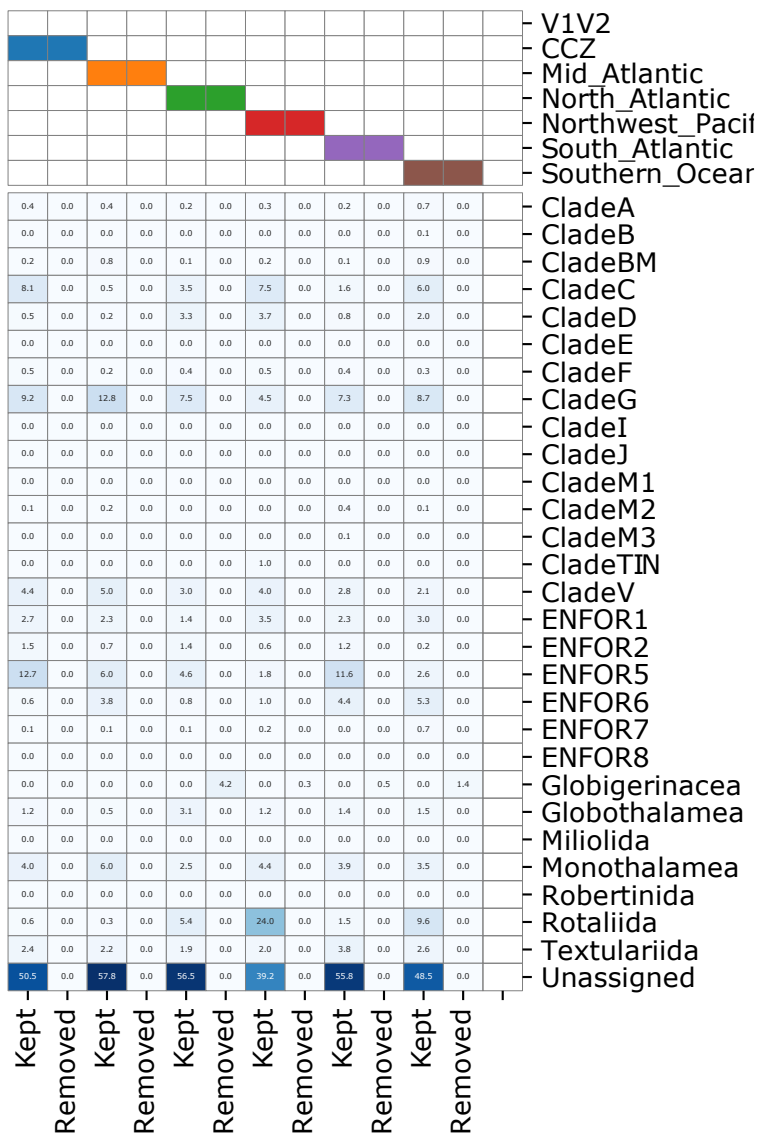
# Metazoa filtering



**Supplementary Figure 7:** Effect of the taxonomic, benthic sequence filtering on the numbers of OTUs per taxonomic marker and per taxon.

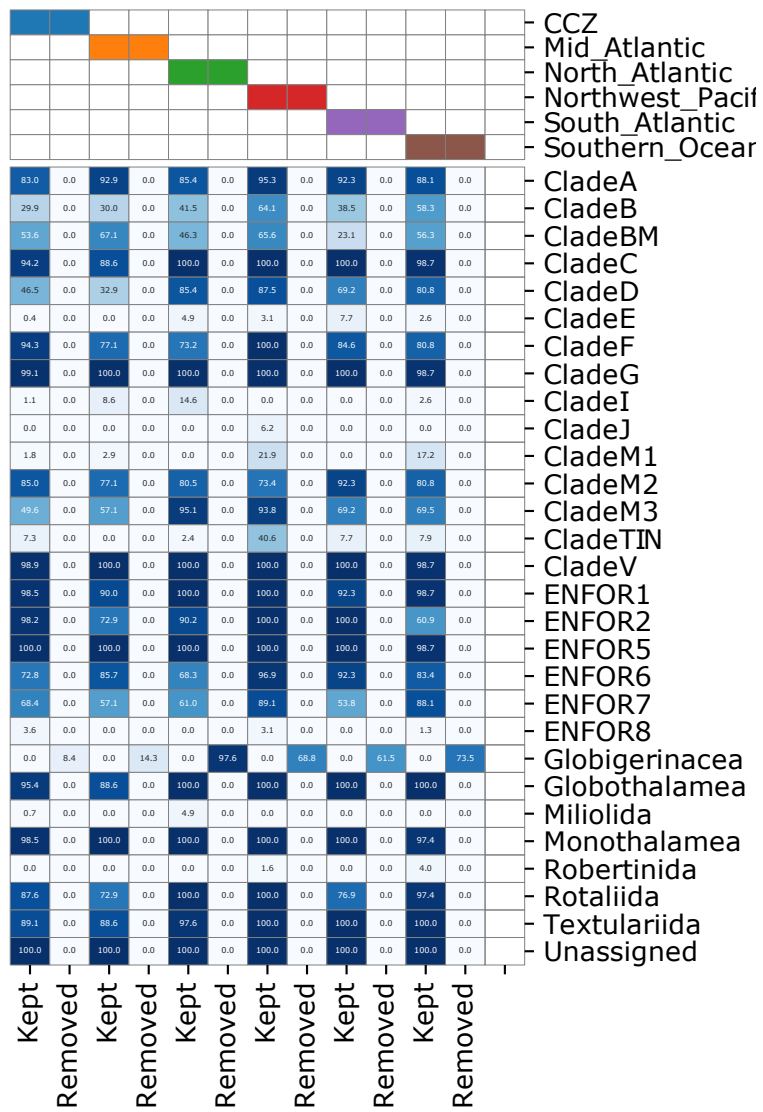
A

37F



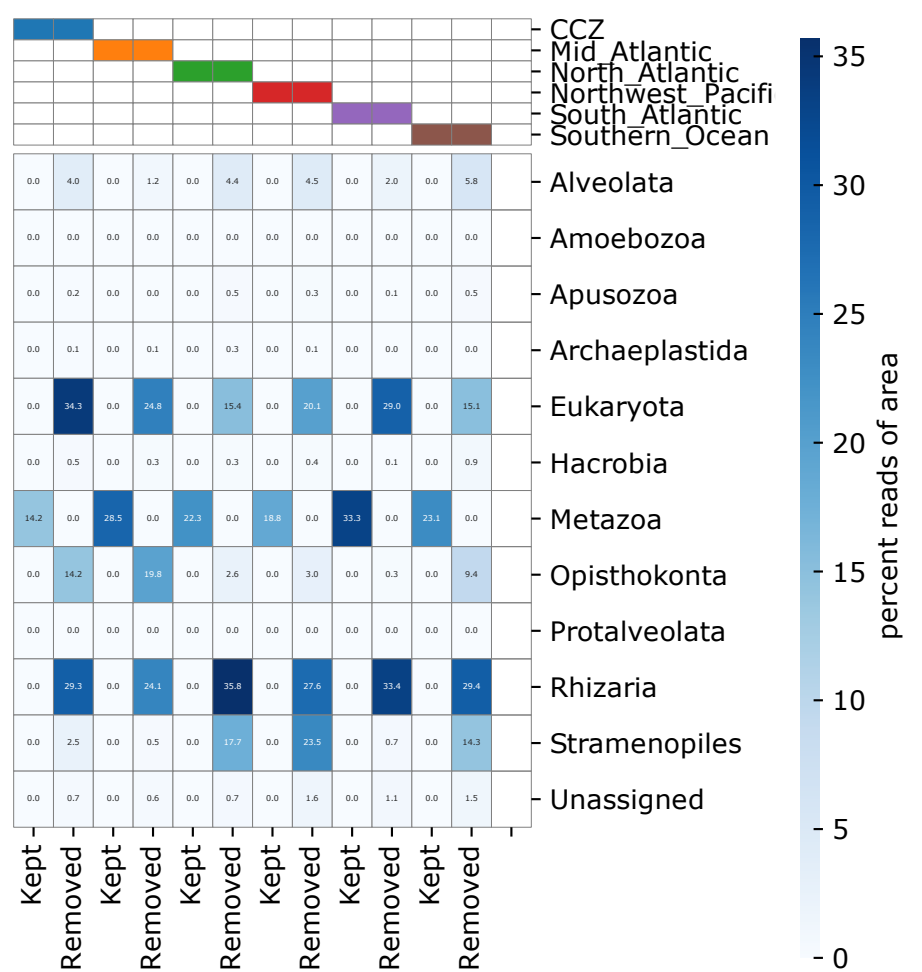
B

37F



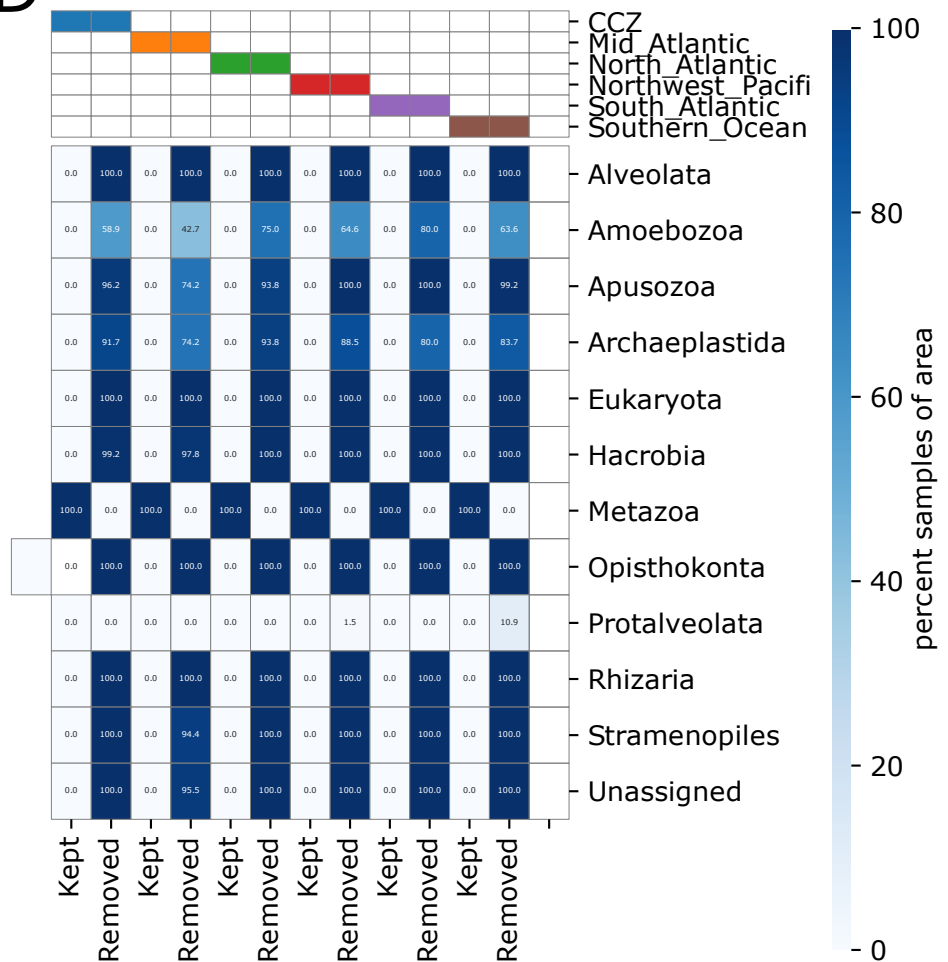
C

V1V2



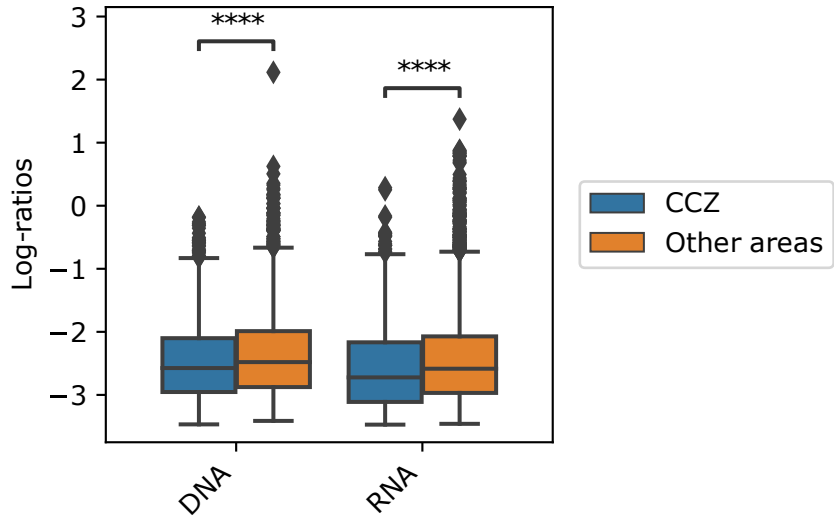
D

V1V2

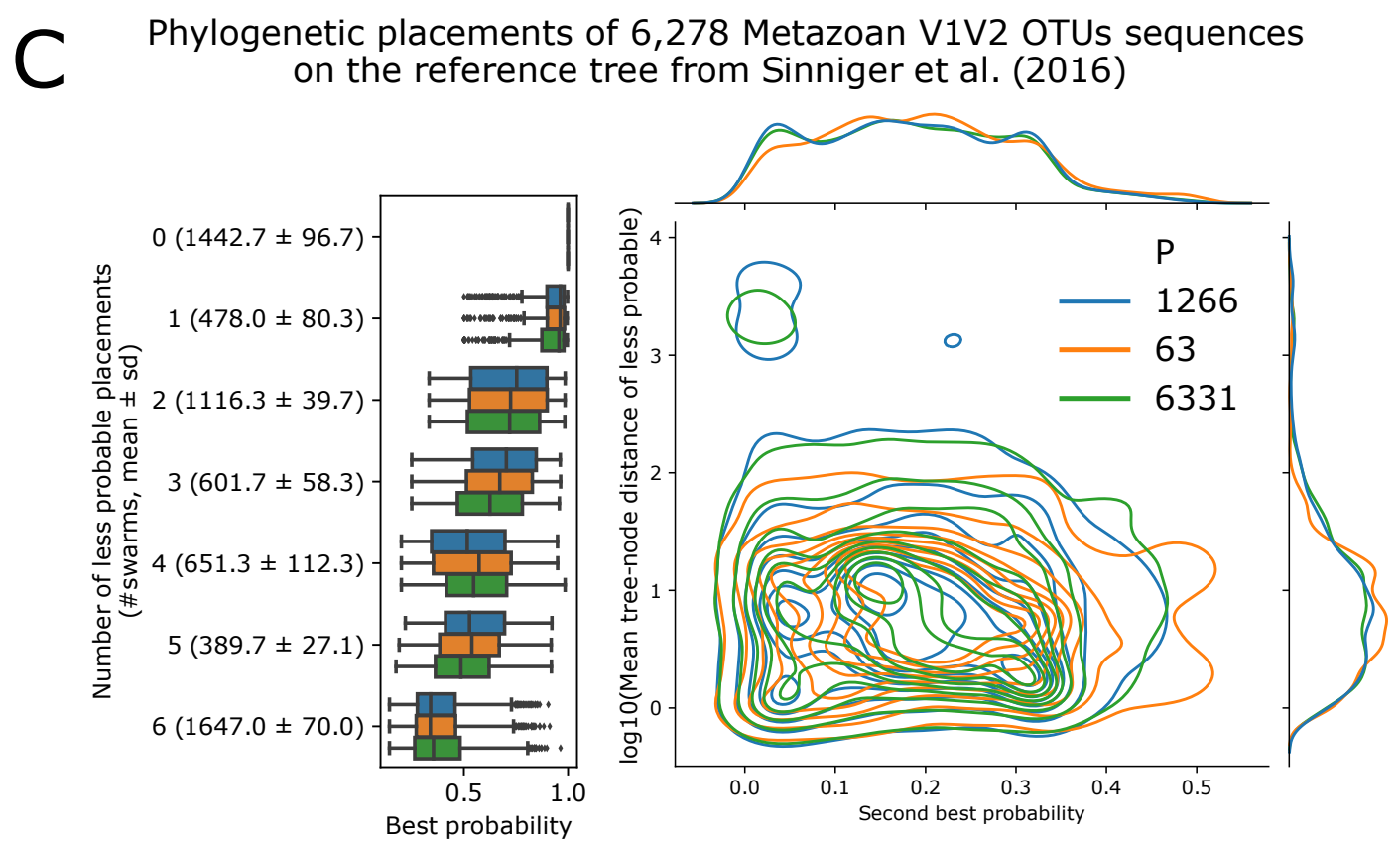
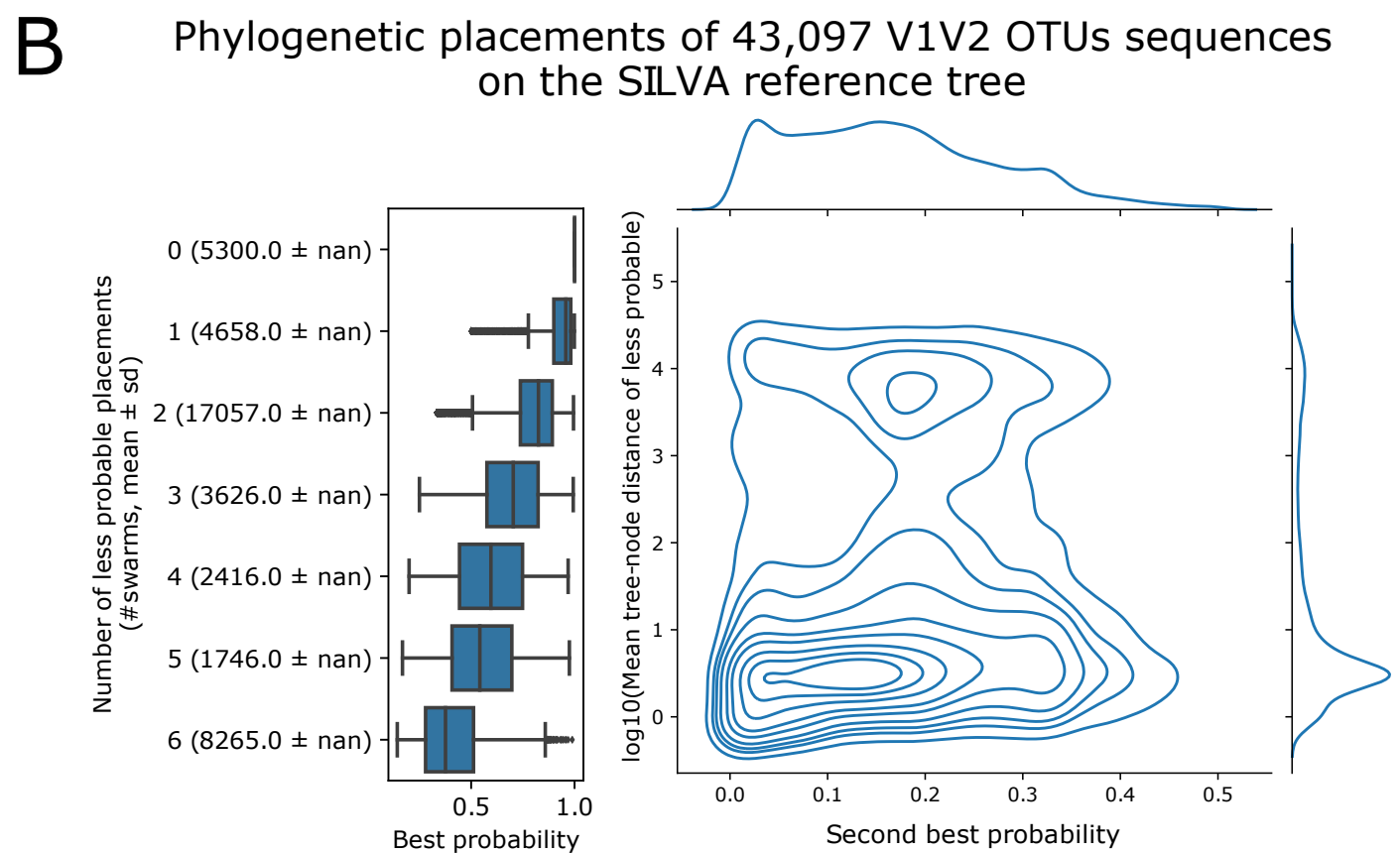
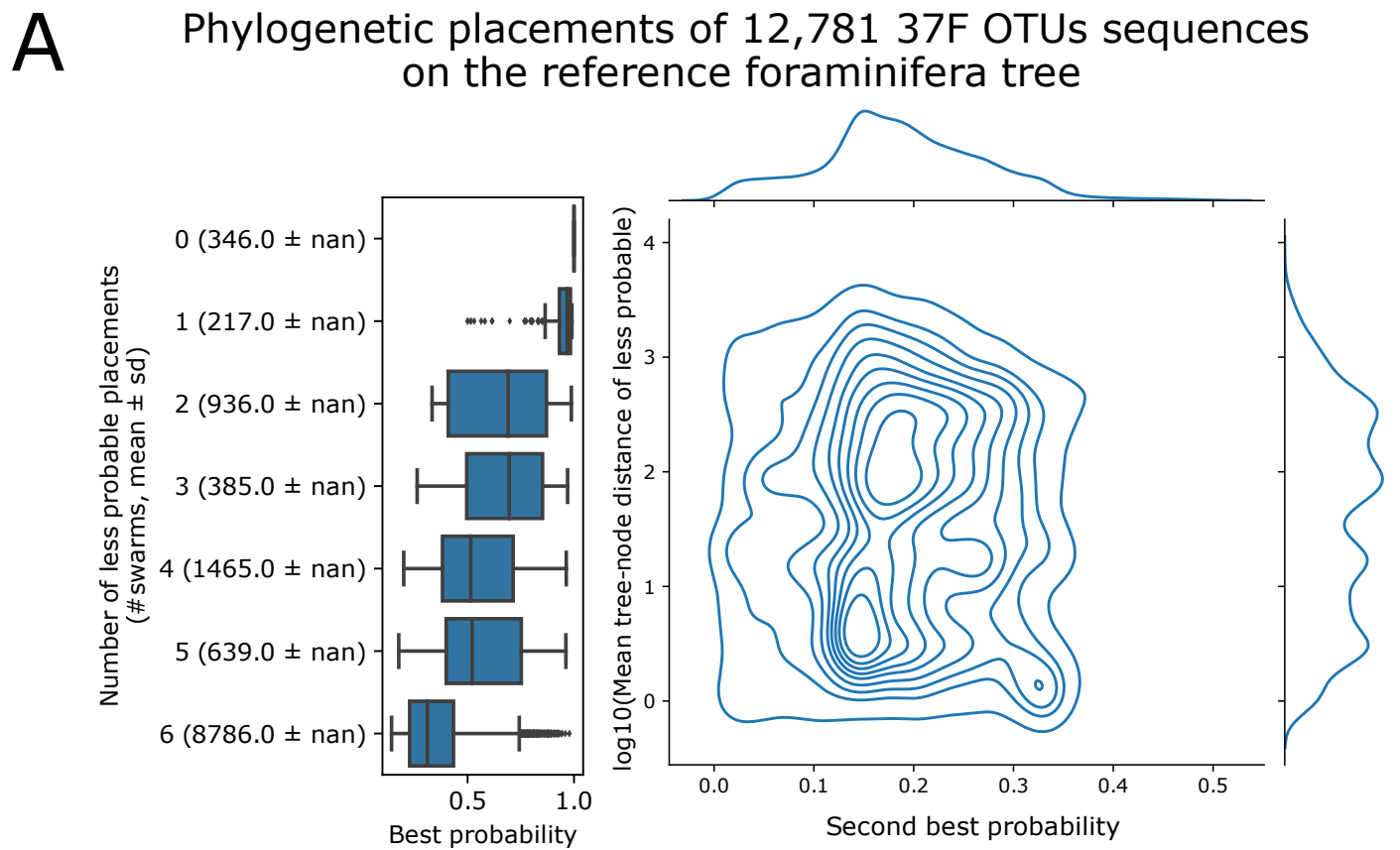


**Supplementary Figure 8:** Effect of the benthic sequence filtering for 37F (A-B) and Metazoa sequence filtering for V1V2 (C-D) detailing for each taxon its percent of reads (A-C) and samples (B-D) per region. The numbers in the cells indicate the exact percentage values.



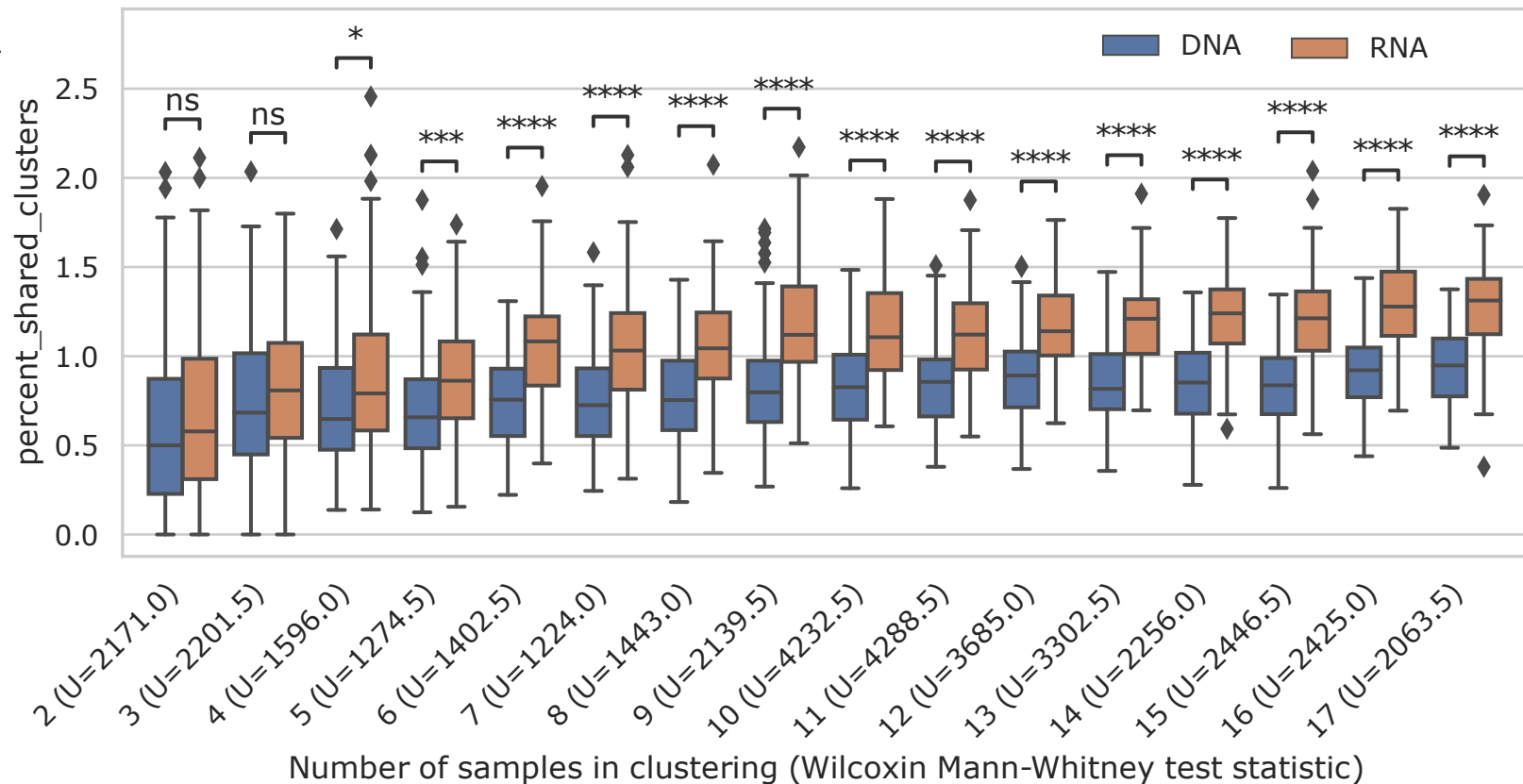
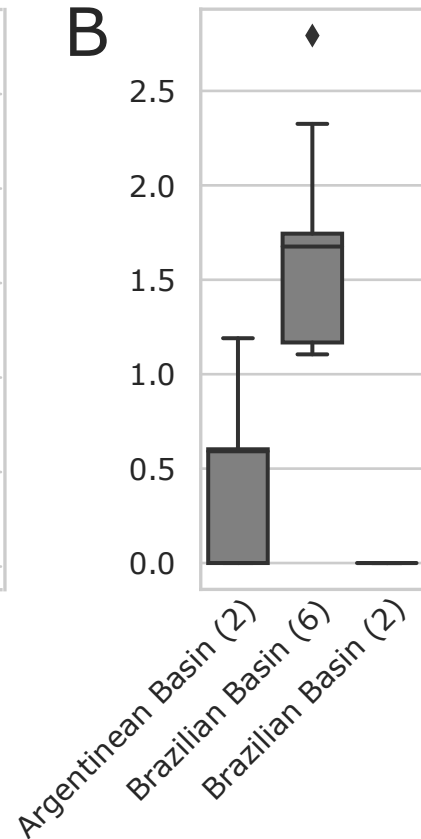


**Supplementary Figure 9: Comparisons of the Globothalamous Foraminifera abundance in the CCZ and other areas based on log-ratios.** For each sample, the rarefied read abundance of each OTU assigned to “Globothalamina” in this sample is divided by the sum of such abundances for other OTUs that are abundant enough to avoid zero in the denominator. We chose to use “Unassigned” OTUs for these internal references as they are found in every sample. Each OTU ratio is then log-transformed (using the python *numpy* `log()` function) and the distributions of the resulting log-ratios compared between that calculated in CCZ samples and in other samples using the Wilcoxon-Mann-Whitney test (as log-ratios are not Normally distributed: Shapiro-Wilk test:  $W=0.927$ ,  $p\text{-value}=0.0$ ).

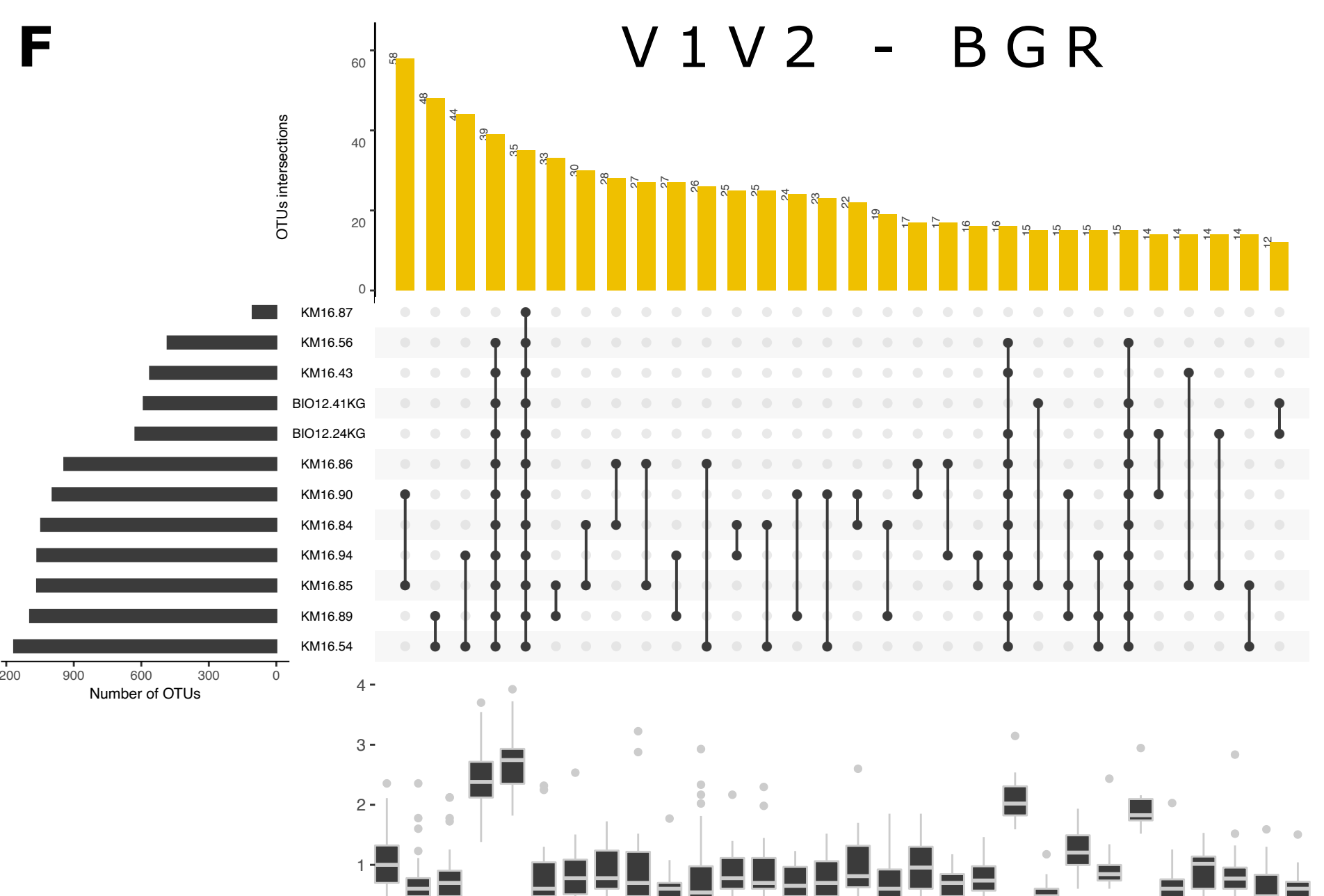
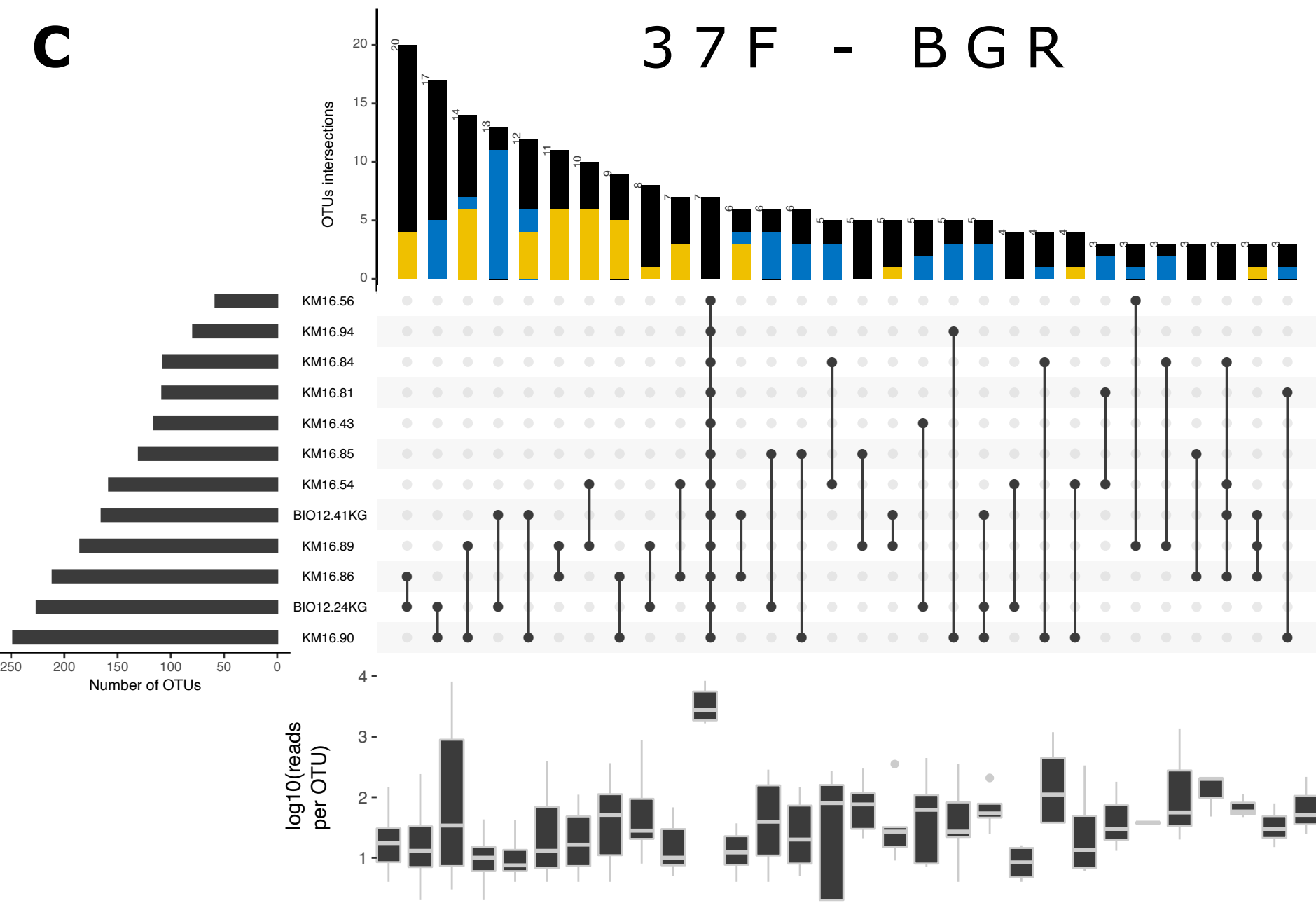
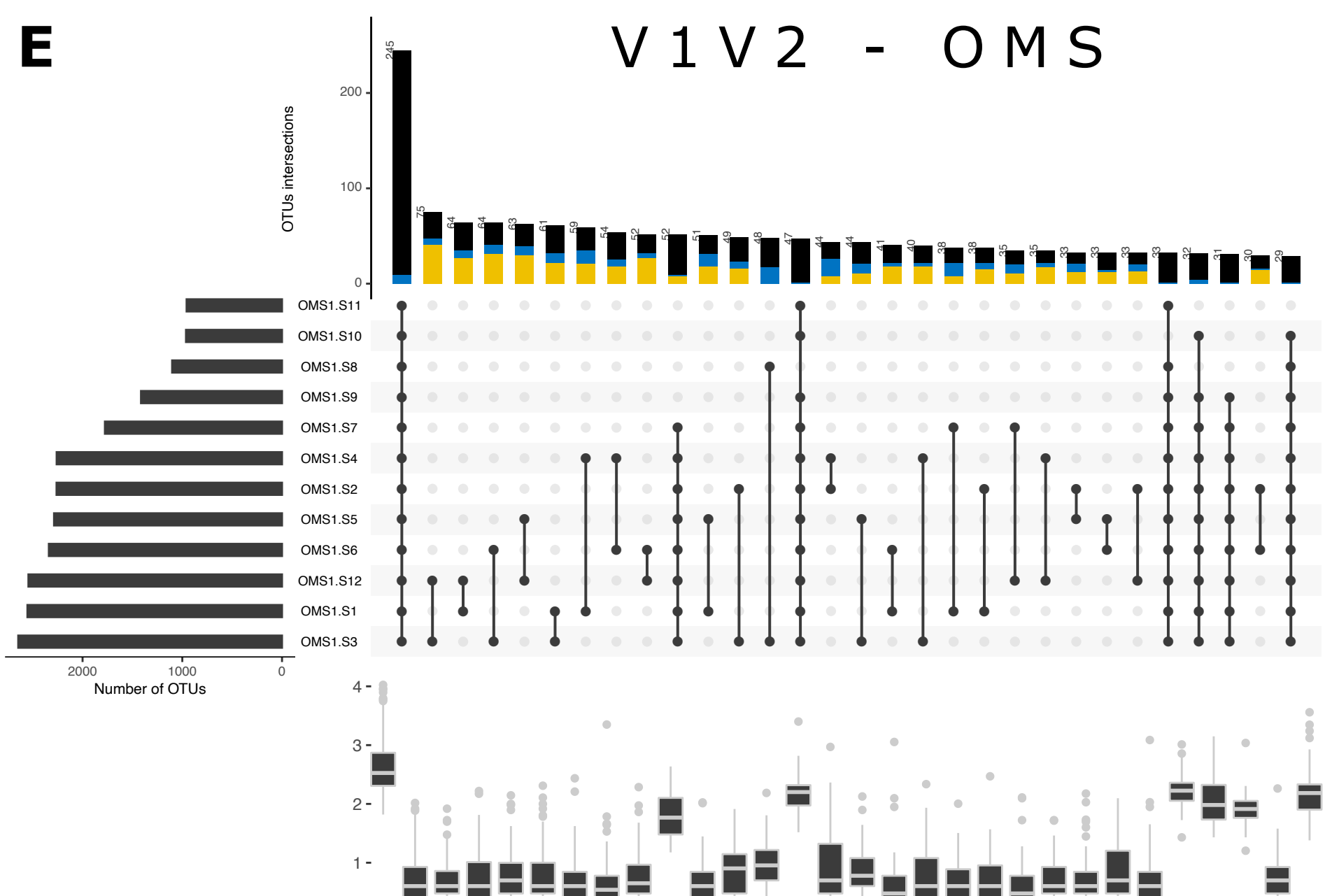
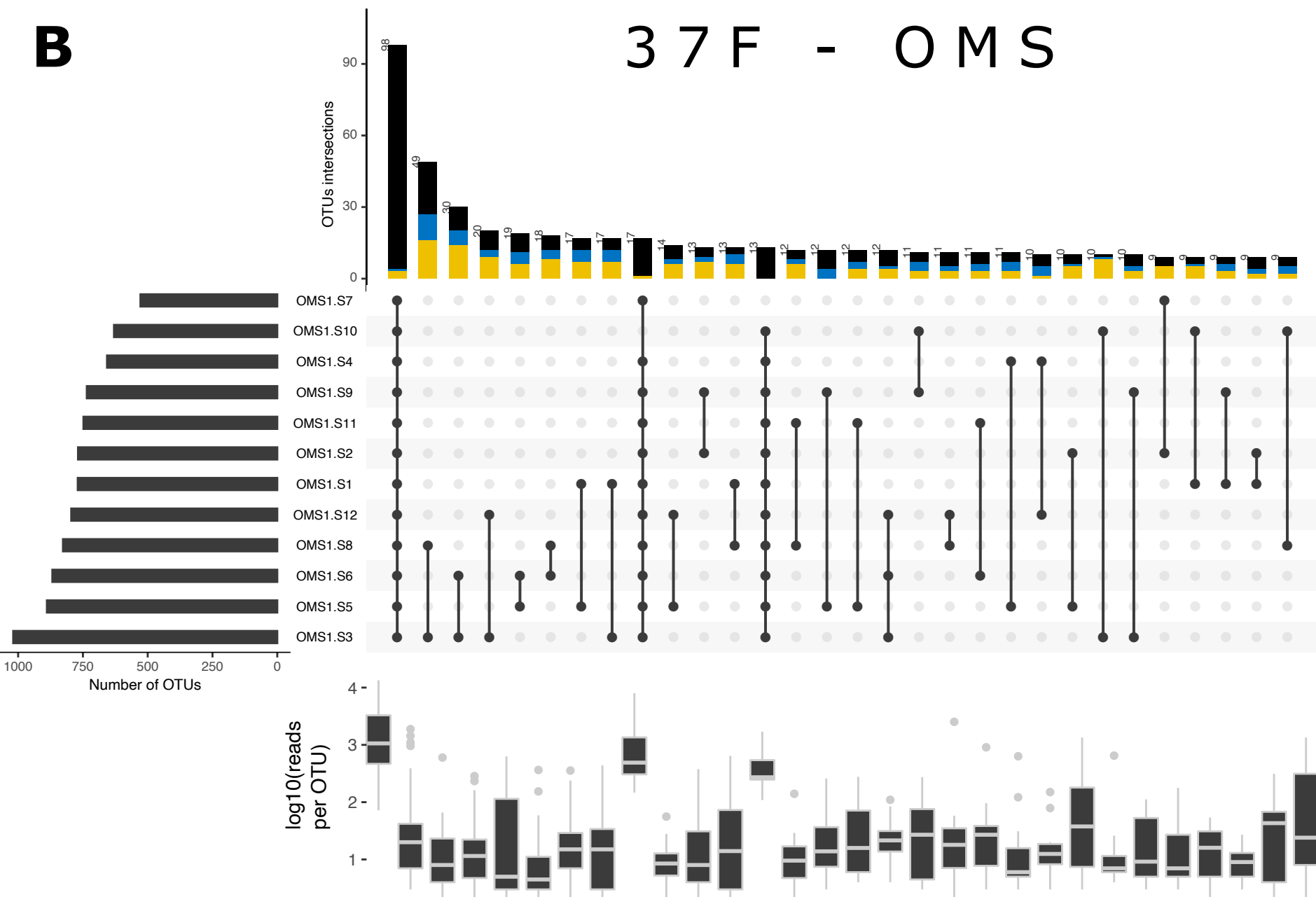
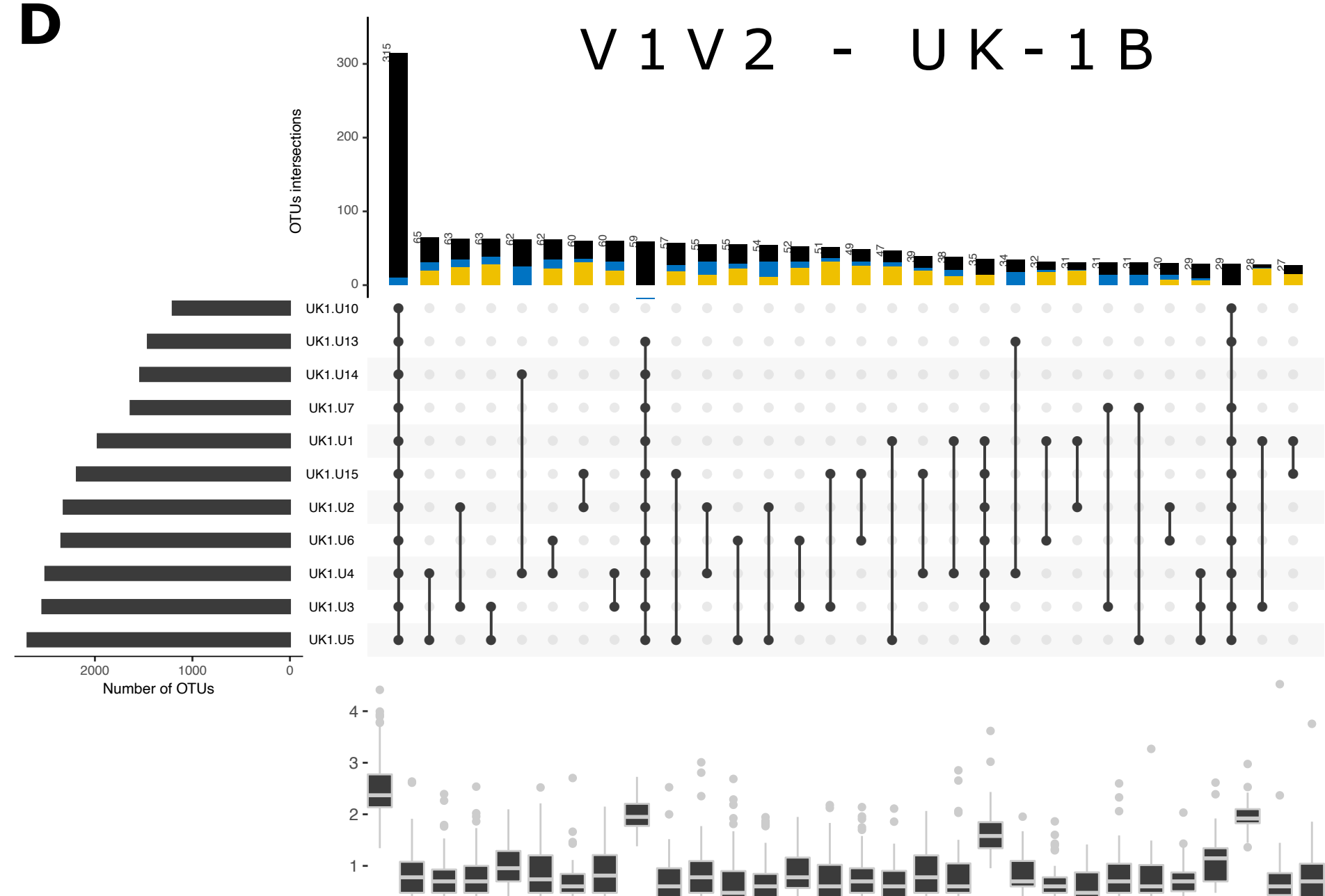
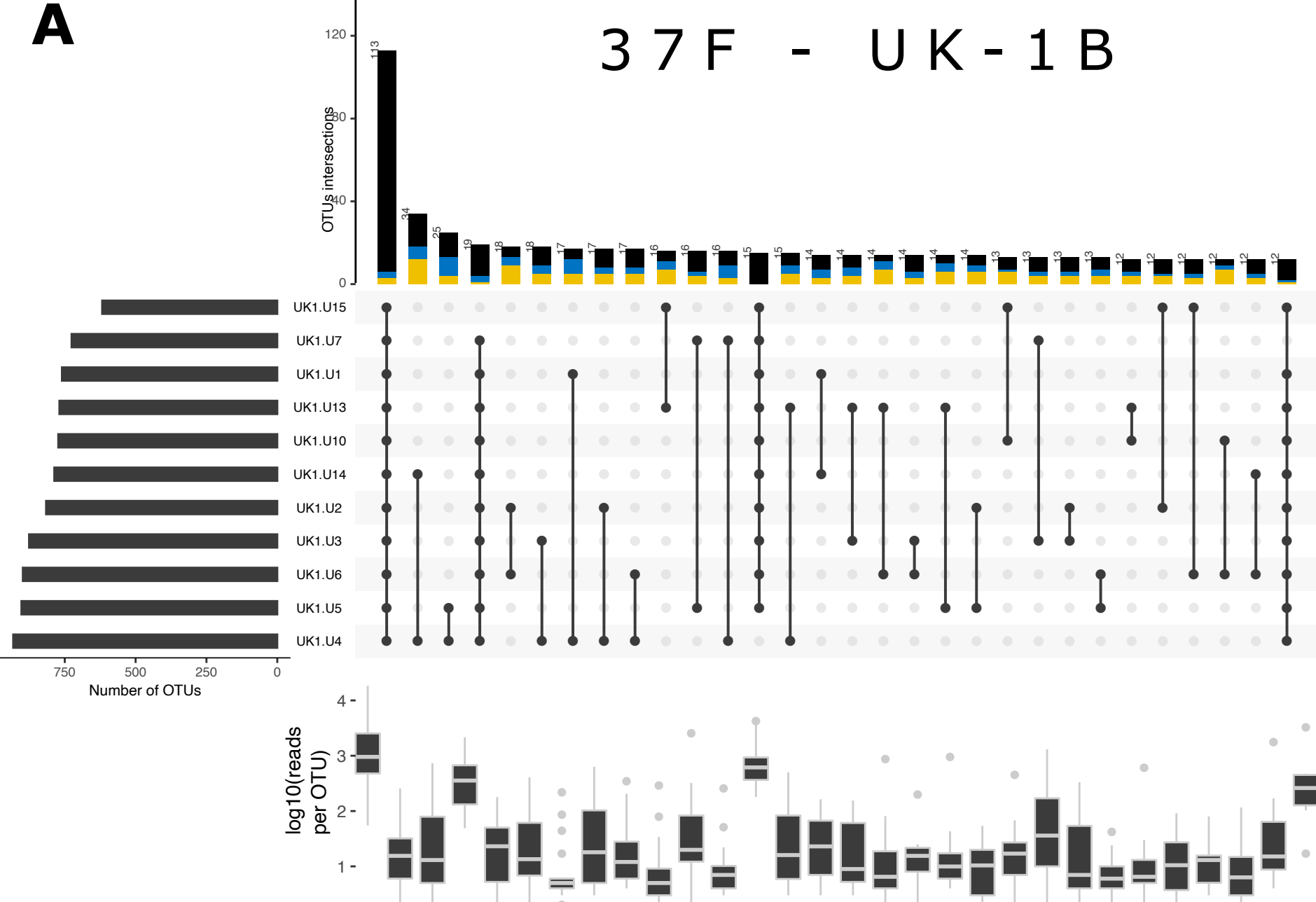


**Supplementary Figure 10:** Quality of the phylogenetic placements for (A) the 37F OTUs onto the reference Foraminifera tree, (B) the V1V2 OTUs onto the SILVA (version 138.1) reference tree and (C) the V1V2 Metazoa OTUs onto the Sinniger et al. (2016) reference tree. The density scatter plot shows the node distance to the best placement for each second-best placement (associated with a maximum probability near 0.5). This node distance corresponds to the difference between node labels, which are incrementally numbered along a post-order traversal of the tree. As a second most probable placement remains in the vicinity of its best placement, this scatter suggests stable placements overall. The side boxplots indicate the distribution of best probabilities for each number of possible placements.



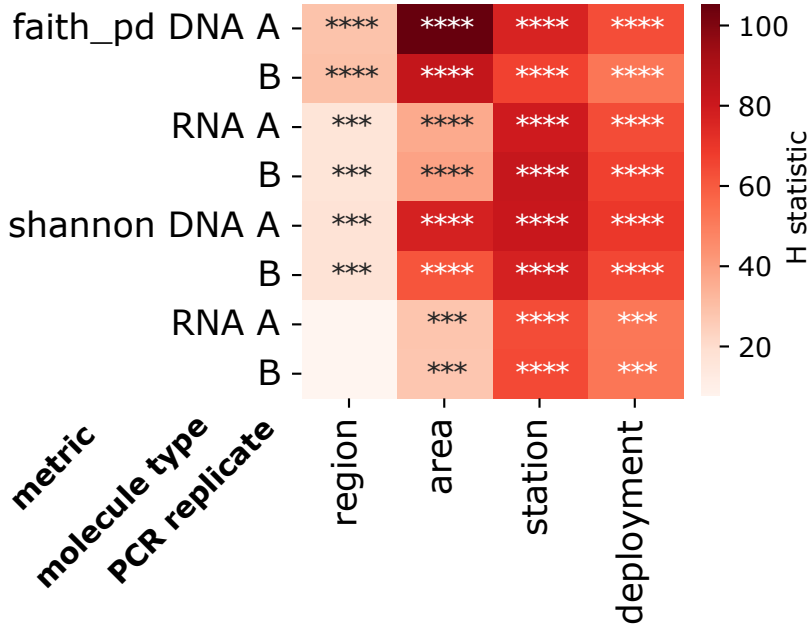
**A****B**

**Supplementary Figure 11:** Co-clustering of V1V2 unique sequences with homologous OTUs from Sinniger et al. (2016). (A) For each number of randomly selected Sinniger et al. samples (x-axis), 100 clustering were realized including similar amounts of V1V2 unique sequences from both studies. From each draw is measured the proportion of studies co-clusters and pairwise Wilcoxon-Mann Whitney tests for differences between the proportions of DNA and RNA co-clusters (p-values adjusted using Bonferroni correction method). (B) Similar co-clustering but performed only on the samples from this study that were taken at the same station as in Sinniger et al. (2016), including 2 samples from the Argentinean Basin area, and another 6 and 2 samples taken at two different stations of the Brazilian Basin area.

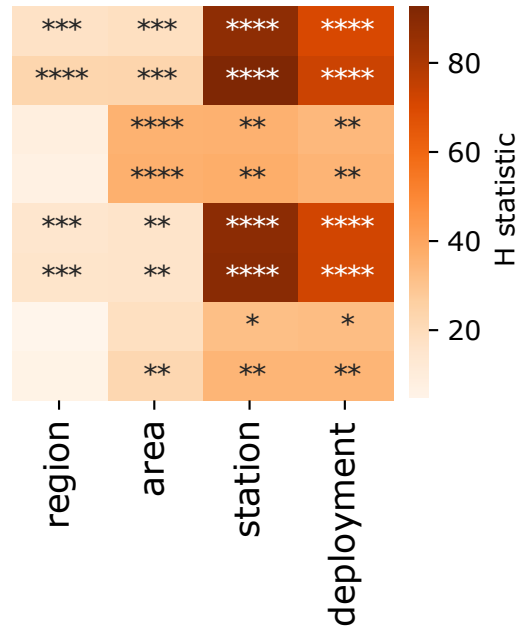


**Supplementary Figure 12:** Same as Figure 5 but the different intersections are shown between well-sampled areas of the CCZ region, including c UK-1B (A and D), area OMS (B and E) and area BGR (C and F), separately for (A-C) the 37F benthic Foraminifera and (D-F) the V1V2 Eukaryotes.

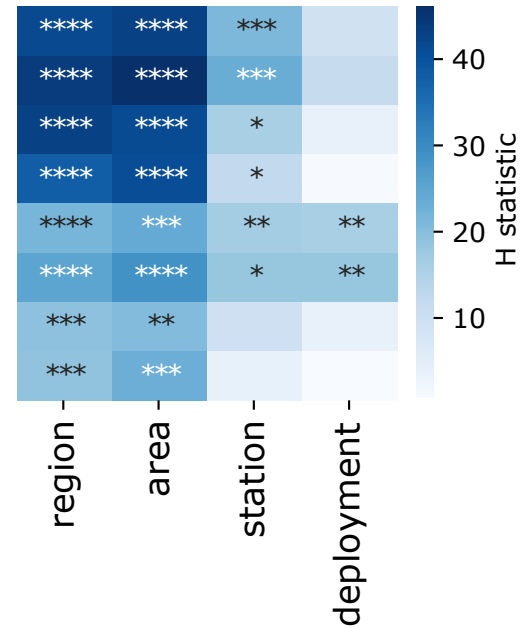
37F benthic



V1V2



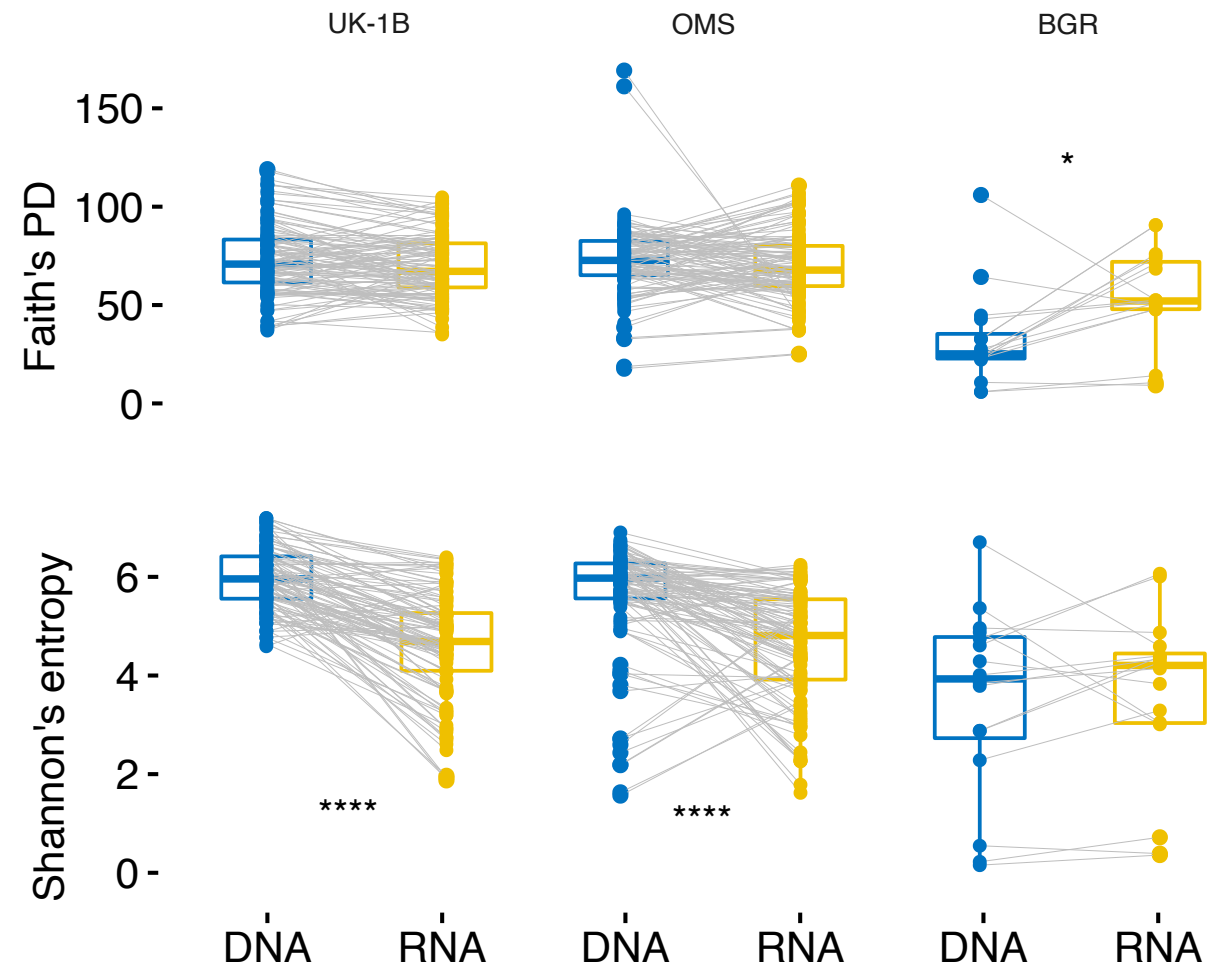
V1V2 Metazoa



**Supplementary Figure 13:** Alpha diversity comparisons (Kruskal-Wallis H test) between regions, areas, stations and deployments. Each test is performed after rarefaction to 3,000 reads and separately for each of the replicates and for DNA and RNA (rows). Cell are colour per effect size (test statistic) and labeled with the level of statistical significance (\*\*\*\*: p-values  $\leq 0.001$ ; \*\*\*: p-value  $\leq 0.01$ ; \*\*: p-value  $\leq 0.03$ ; \*: p-value  $\leq 0.05$ ; no label: p-values  $> 0.05$ ; no color: no testing, adjusted using the Benjamini-Hochberg correction method).

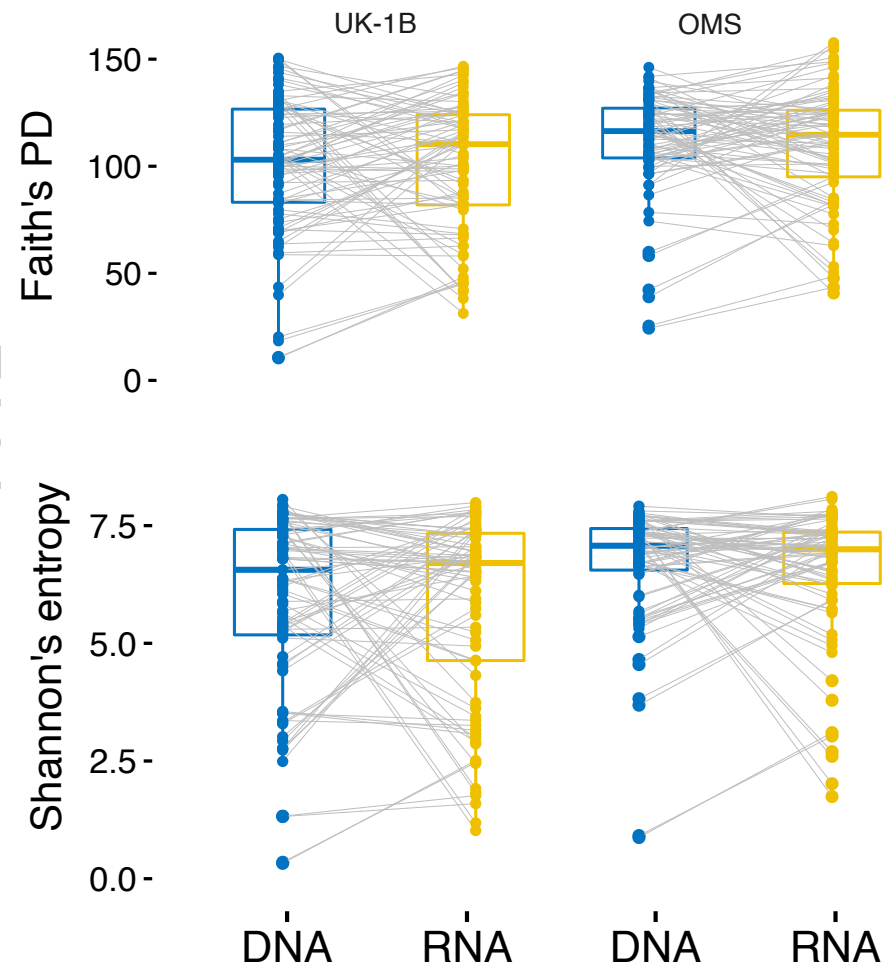
A

37F



B

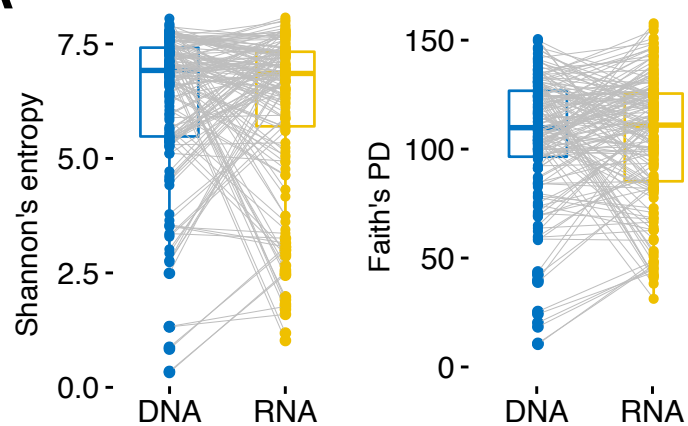
V1V2



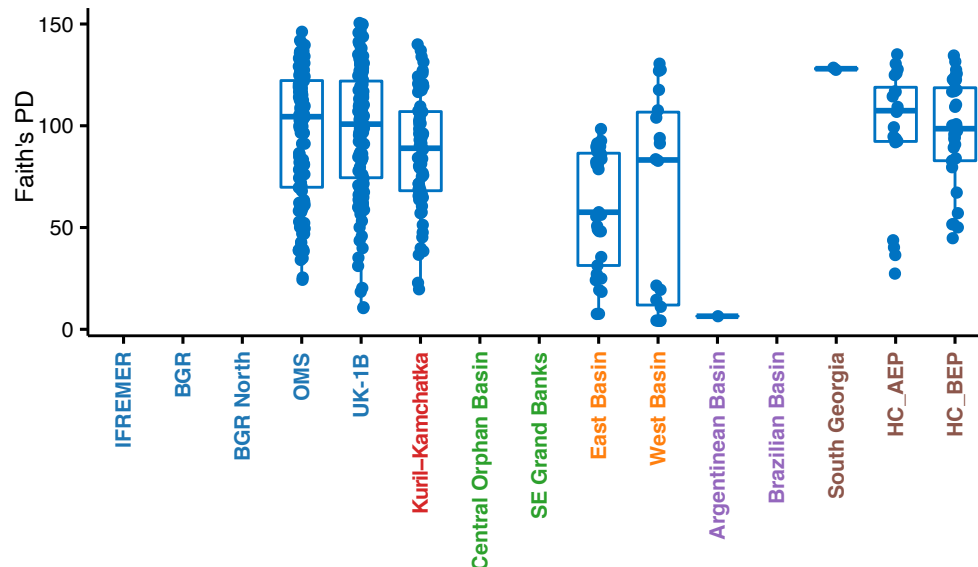
**Supplementary Figure 14.** Difference in Faith's PD and Shannon's entropy alpha diversity between molecules for the 37F benthic Foraminifera (A) and the V1V2 Eukaryota (B) datasets, shown for each area of the CCZ. This is in complement to Figure 4.



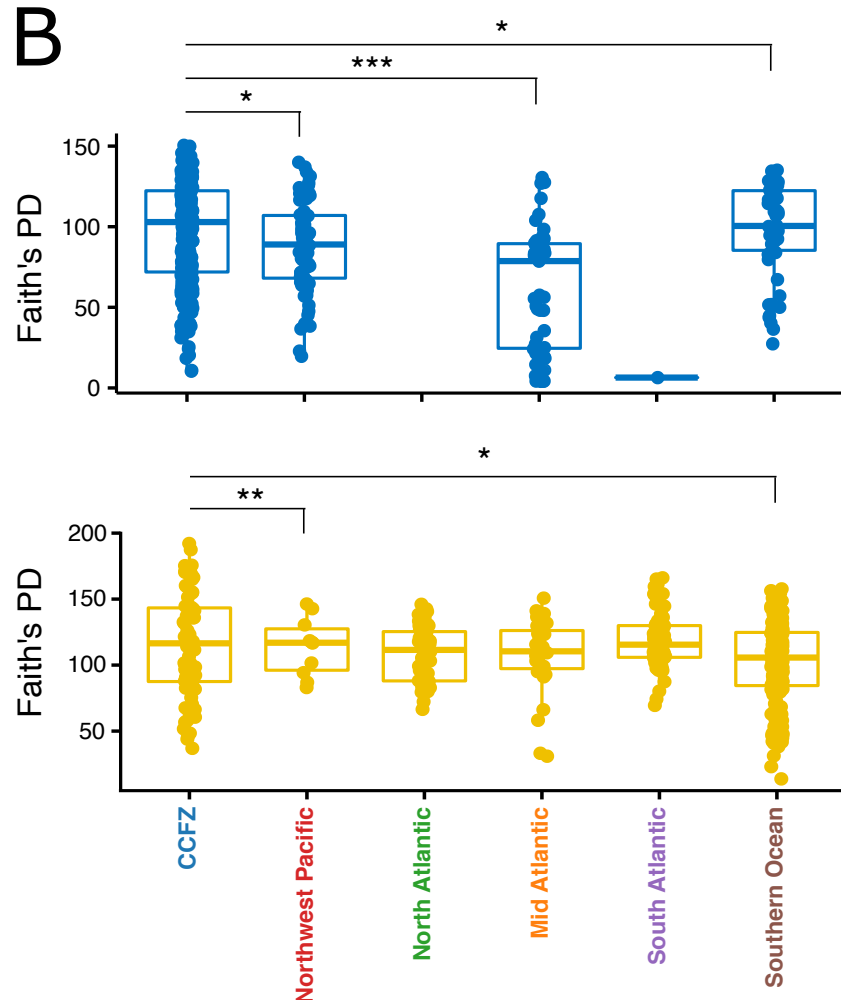
A



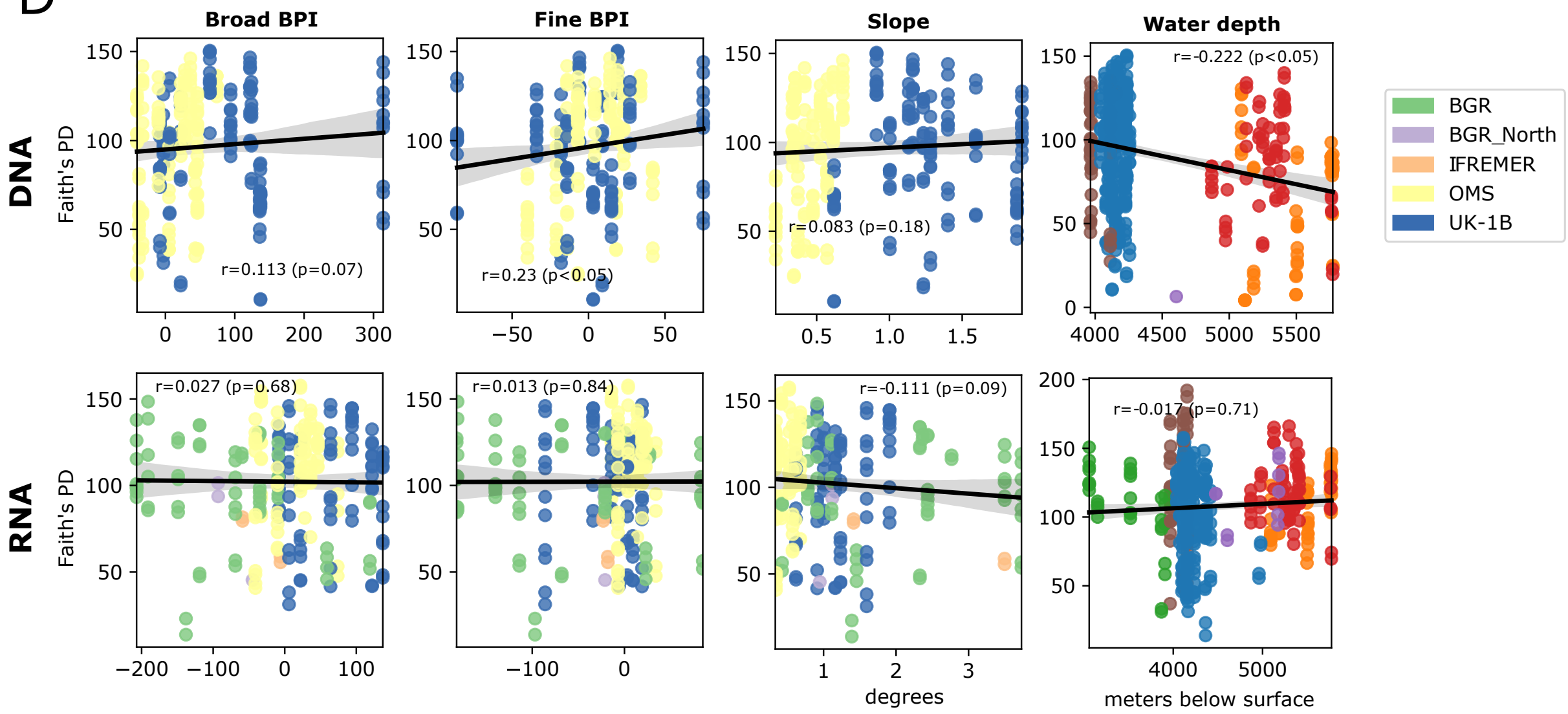
C



B

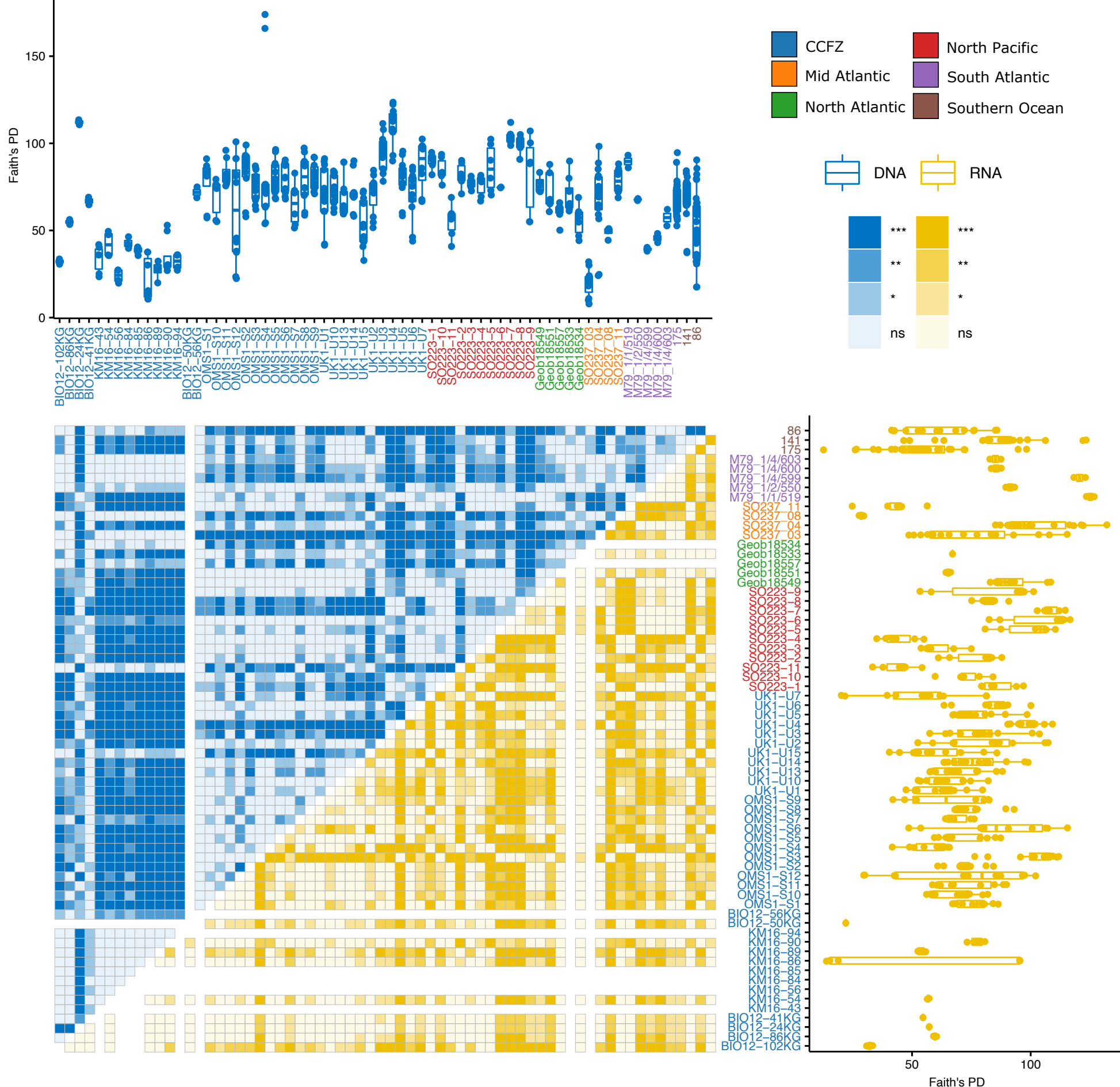


D

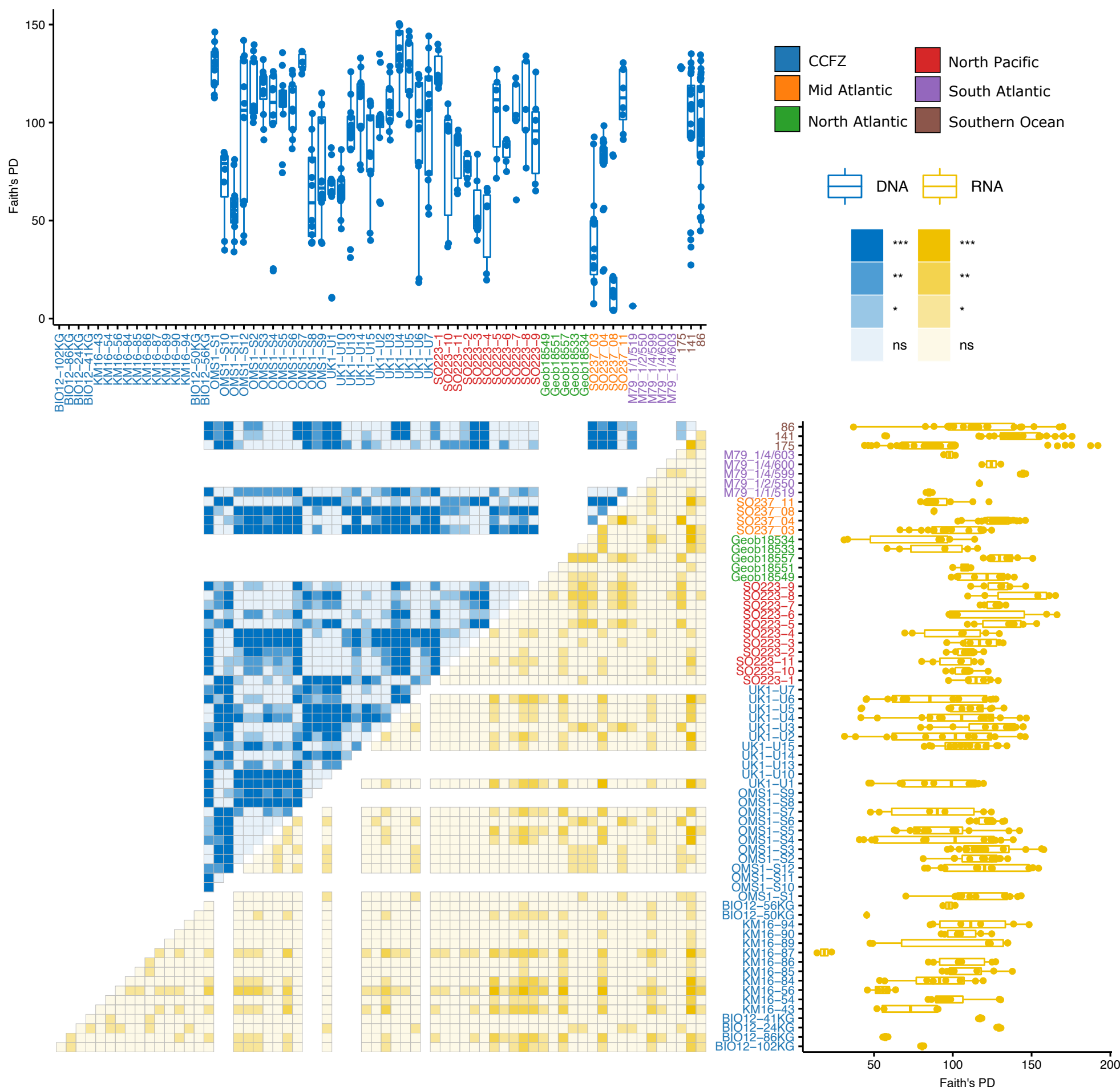


**Supplementary Figure 15:** Same as Figure 5 but for the V1V2 dataset (all taxa).

# A



# B



**Supplementary Figure 16:** Same as Figure 5C but comparisons are made amongst stations (A) for the 37F and (B) for the V1V2 datasets.

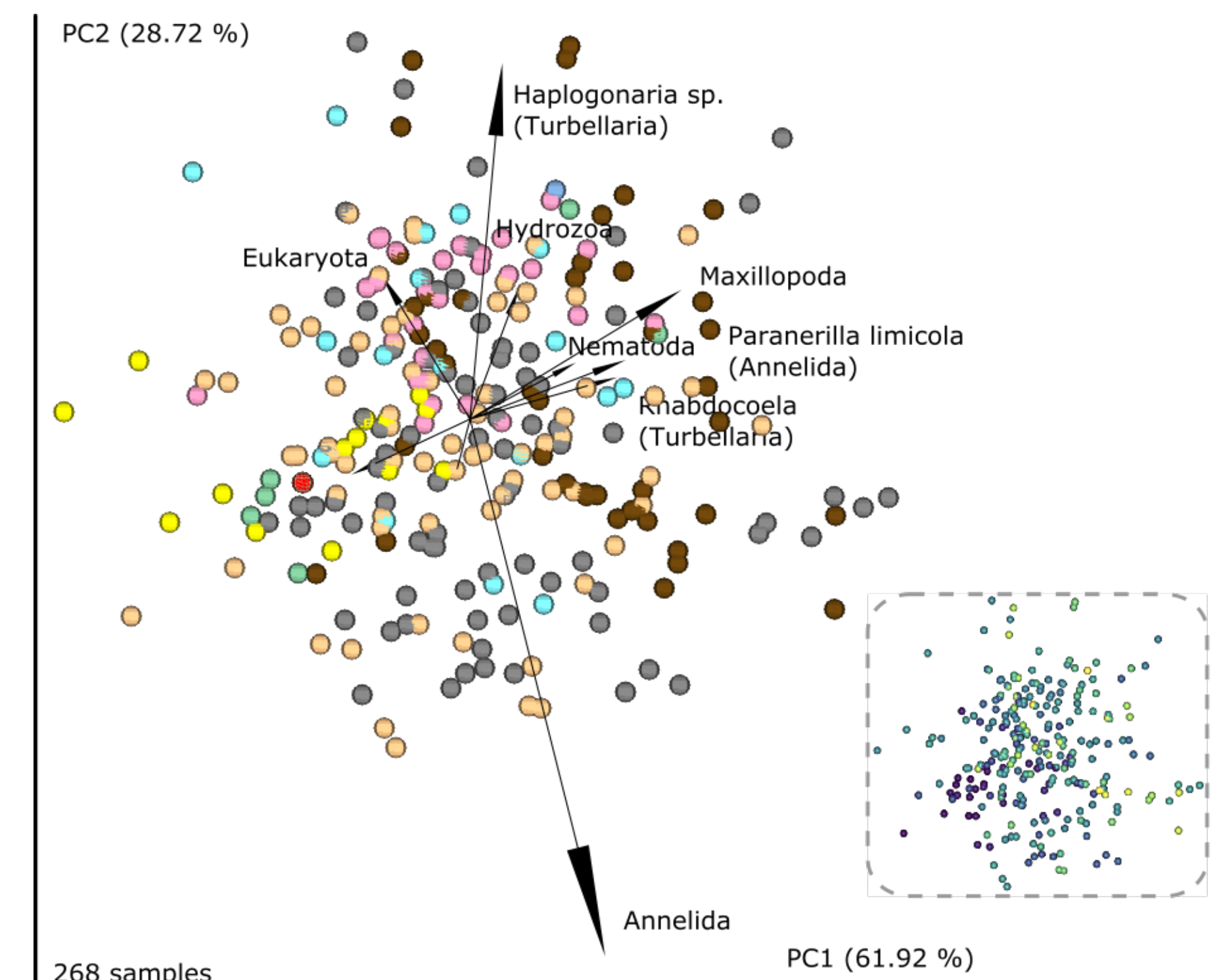
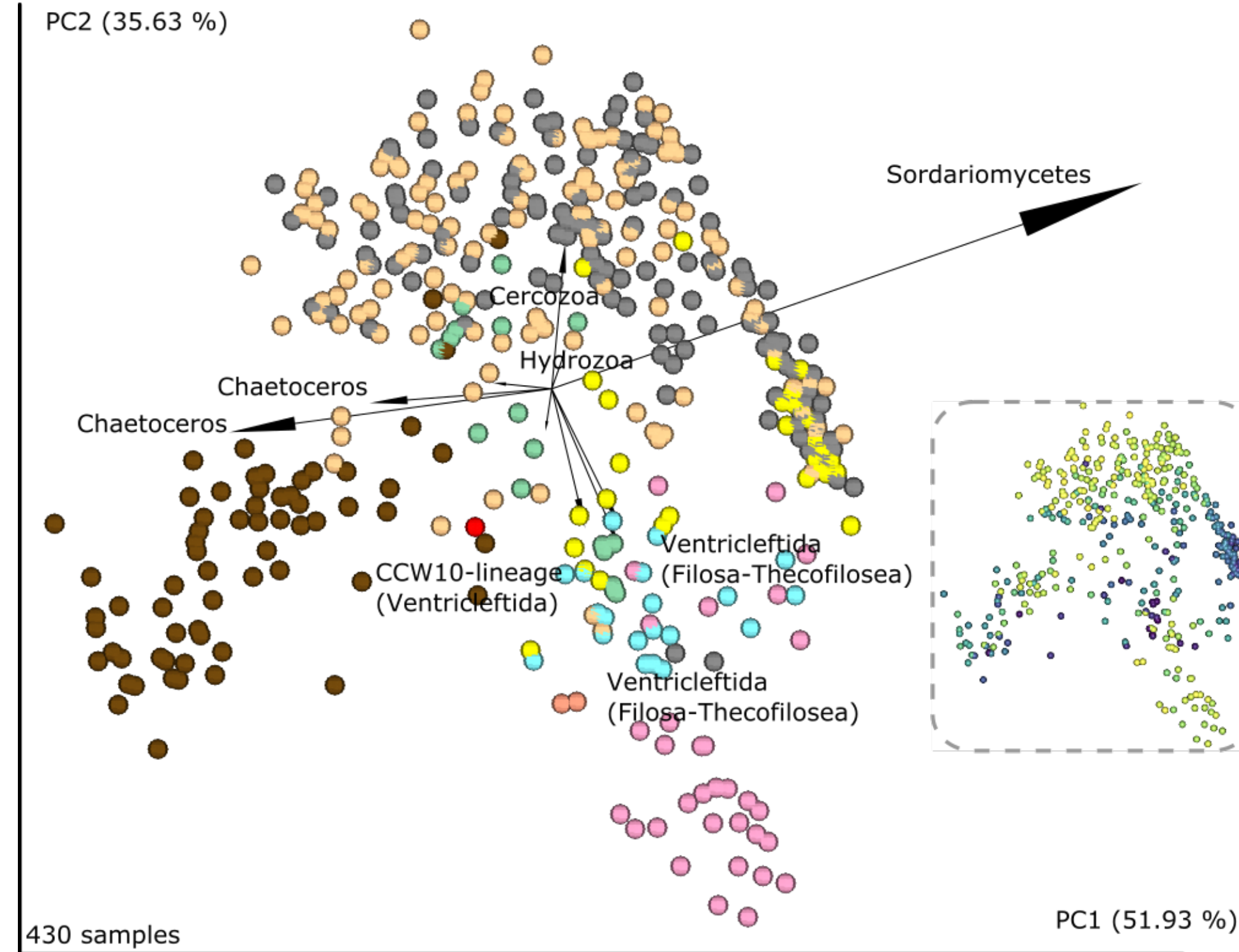
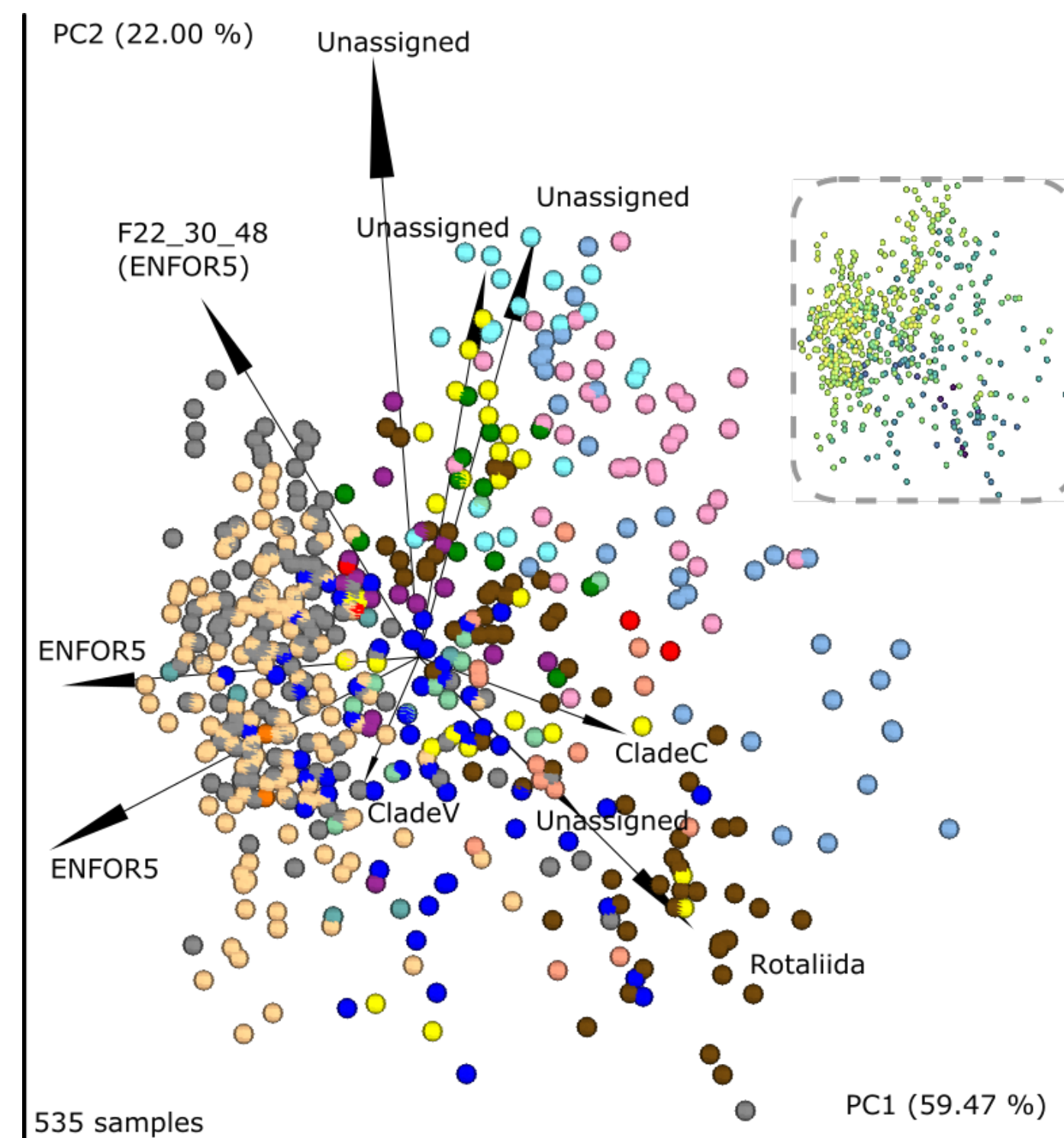


37F

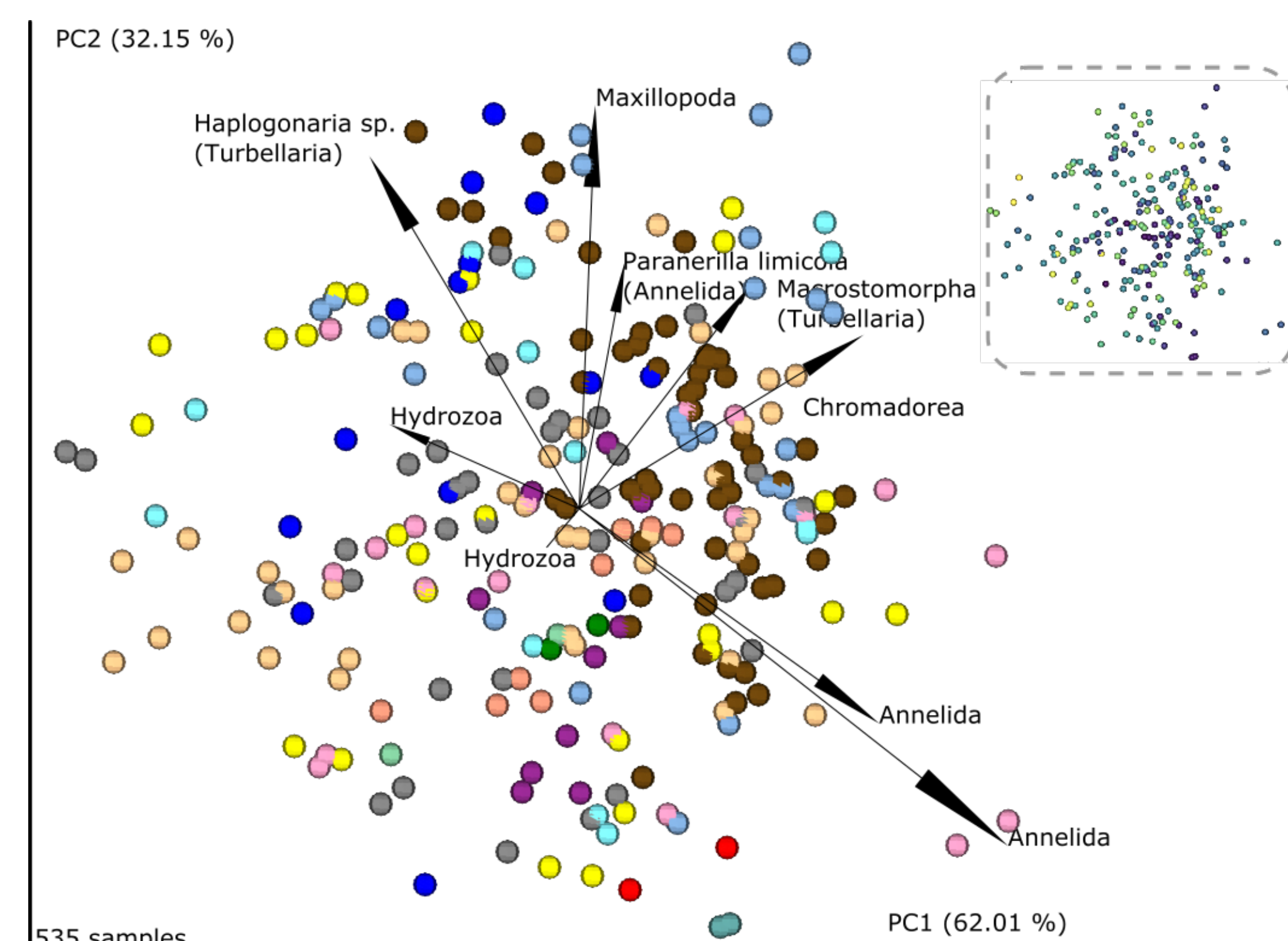
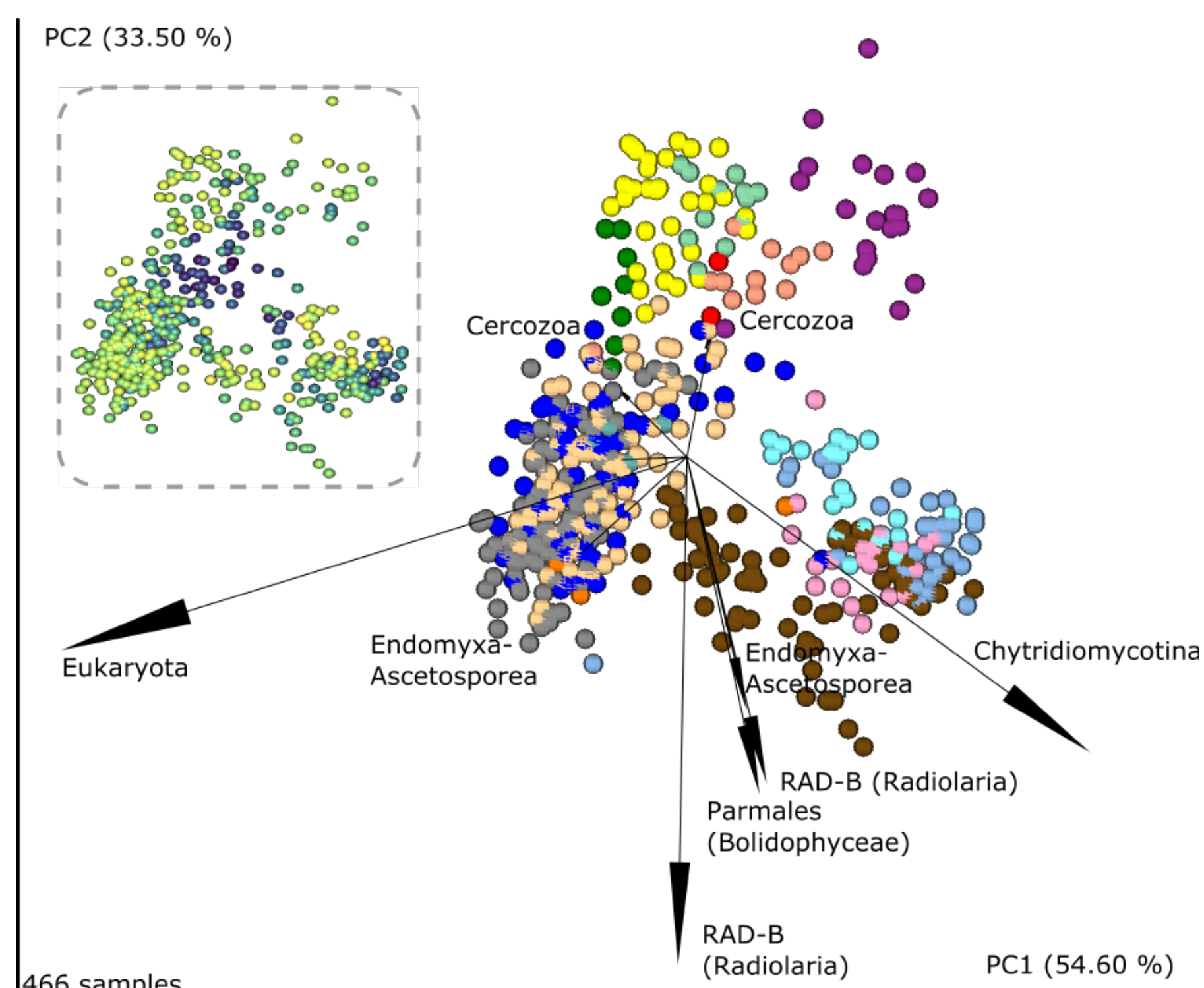
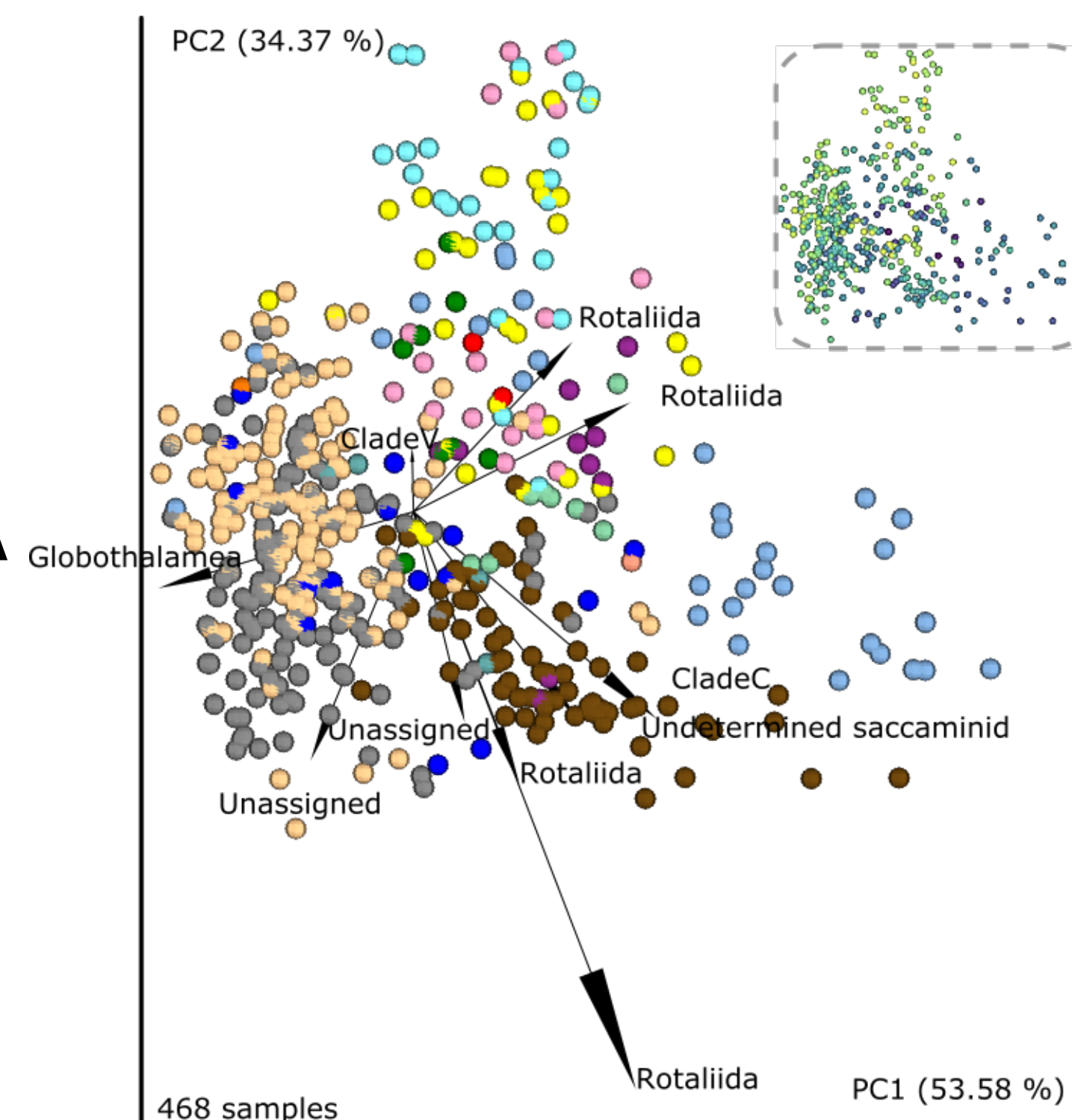
V1V2

V1V2 Metazoa

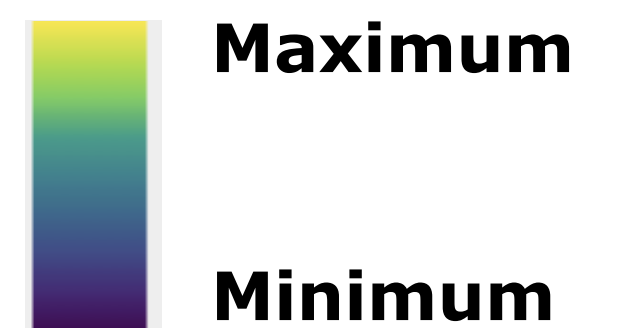
DNA



RNA



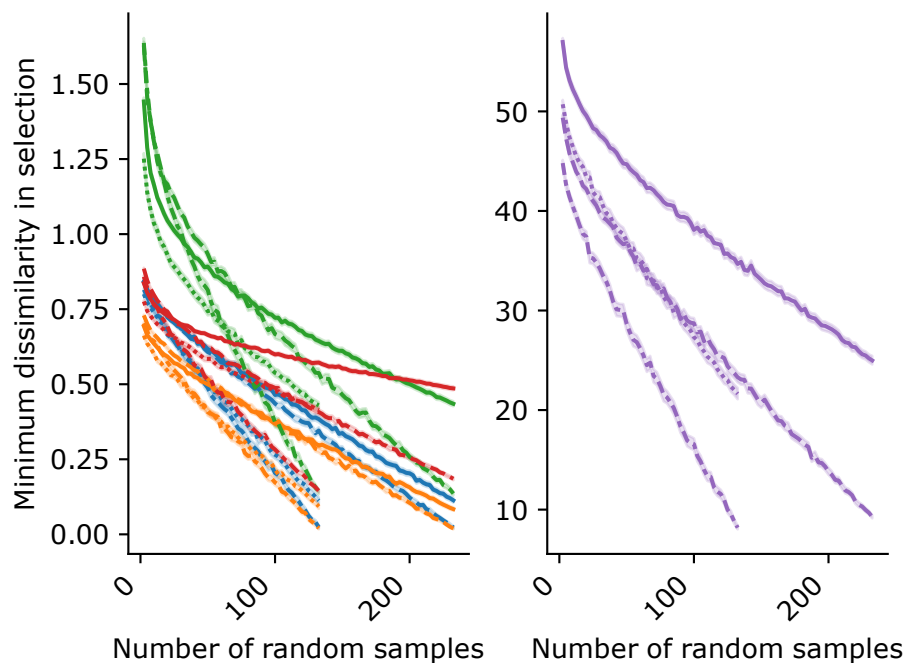
Shannon's entropy



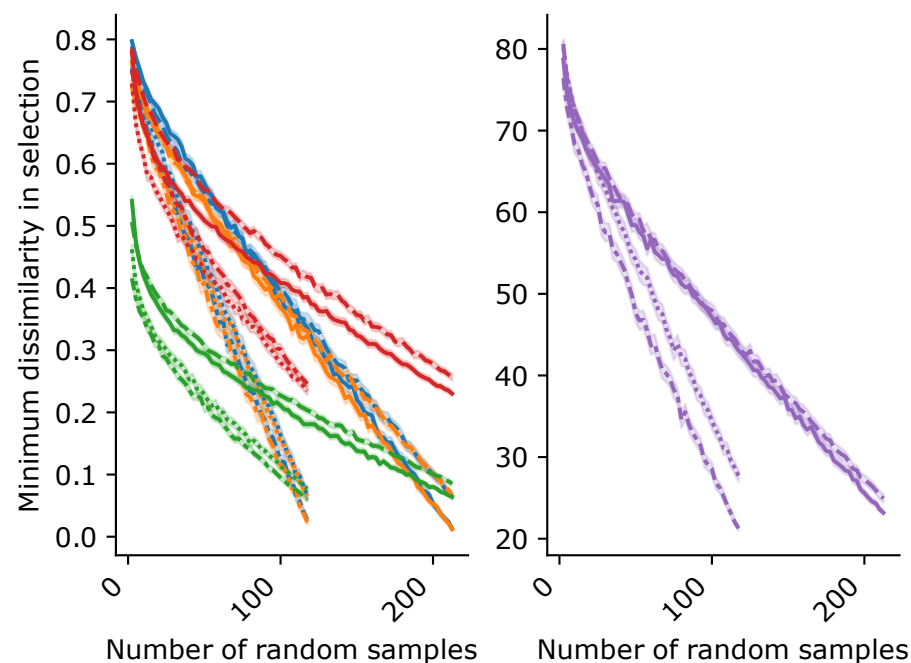


**Supplementary Figure 17: Robust PCAs performed separately for each dataset and for DNA or RNA.** Samples are colored per area (scatter) and the 10 most influential OTUs are indicated (arrows). For each PCA, an inset copy is shown where samples are colored per the value of the Faith's PD alpha diversity index. The numbers of samples in each PCA are indicated in the lower left corners.

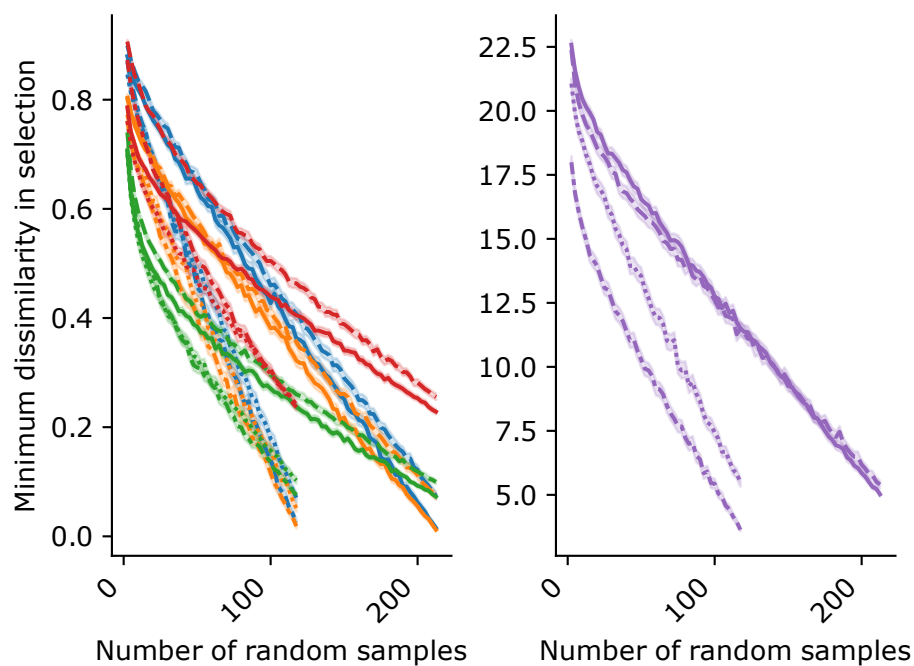
## 37F benthic



## V1V2



## V1V2 Metazoa



### Dissimilarity metric

- Jaccard
- Unweighted UniFrac
- Weighted UniFrac
- Bray-Curtis
- Aitchison

### Subset

- All samples (DNA)
- - - All samples (RNA)
- ..... CCZ (DNA)
- - - CCZ (RNA)

**Supplementary Figure 18:** Minimum-observed distance between samples over increasing numbers of samples for each dataset (panels), metric (colors) and for all or only the CCZ DNA and RNA samples (line style). For each line is represented the 95% confidence interval measured across 10 iterations at each step. The Aitchison distance is shown separately because it does not scale from 0 to 1 (see y-axis).



A

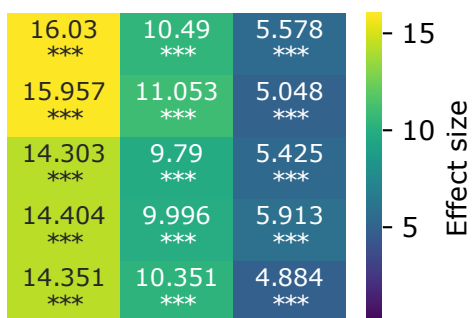
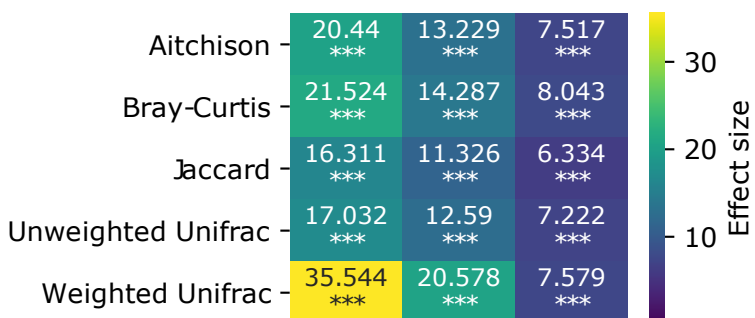
37F

DNA samples (n=535)

RNA samples (n=468)

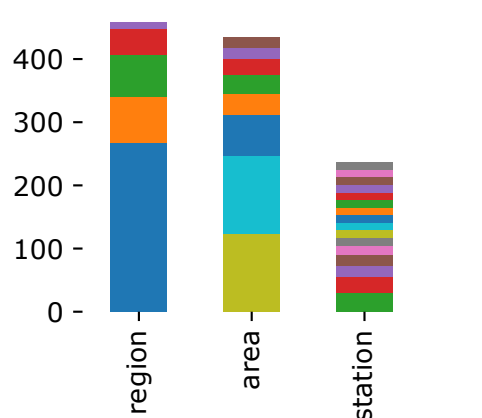
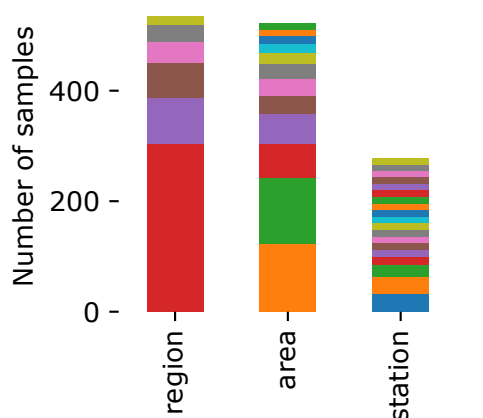
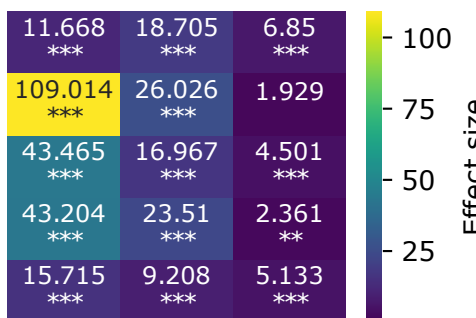
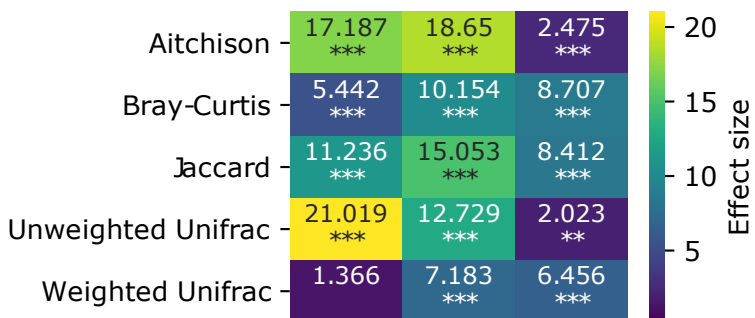
PERMANOVA

PERMANOVA



PERMDISP

PERMDISP



B

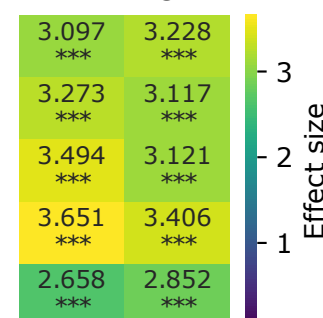
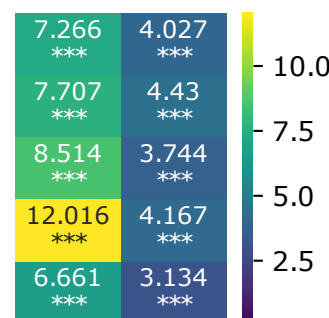
37F - CCZ

DNA samples (n=304)

RNA samples (n=268)

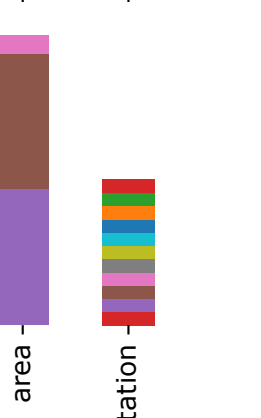
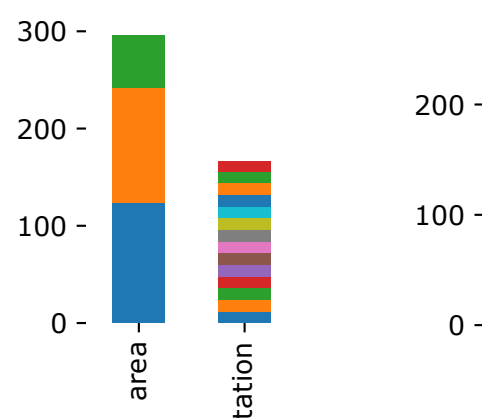
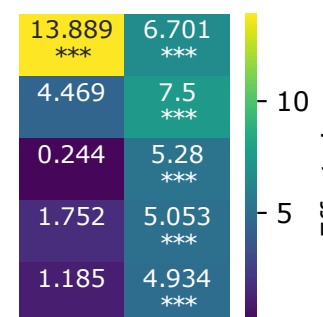
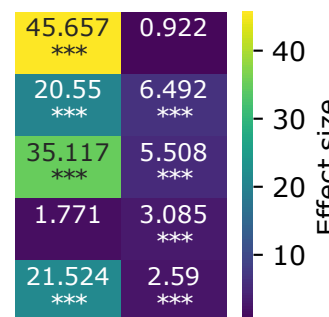
PERMANOVA

PERMANOVA



PERMDISP

PERMDISP



C

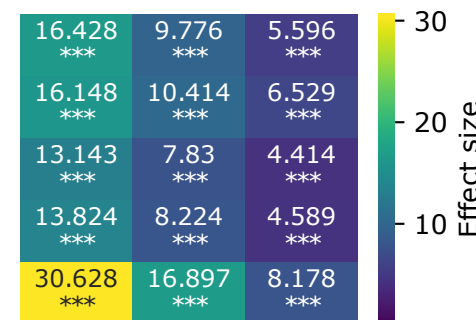
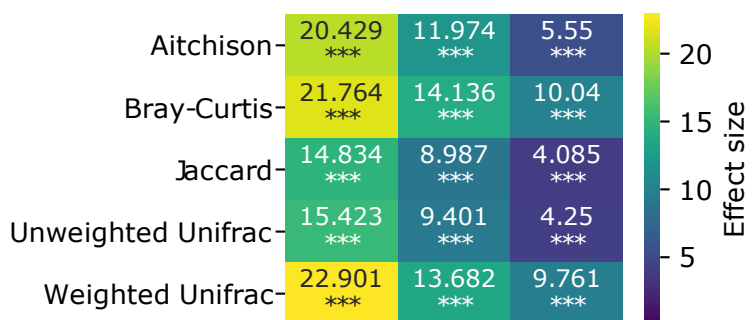
V1V2

DNA samples (n=430)

RNA samples (n=466)

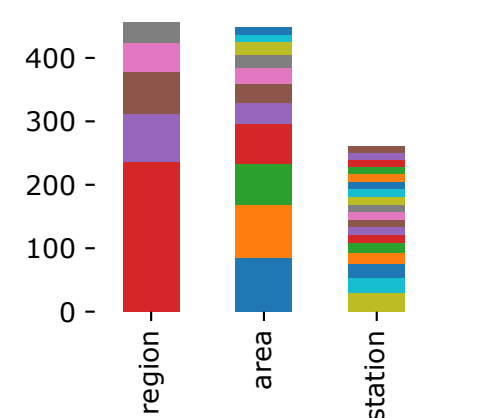
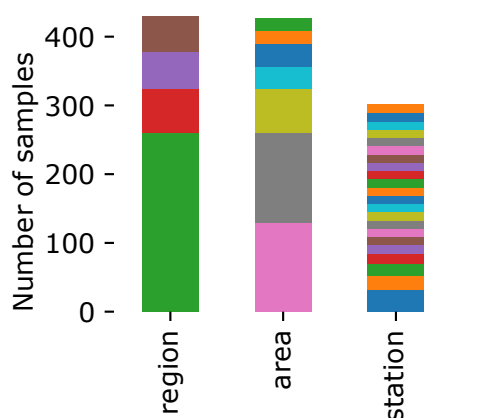
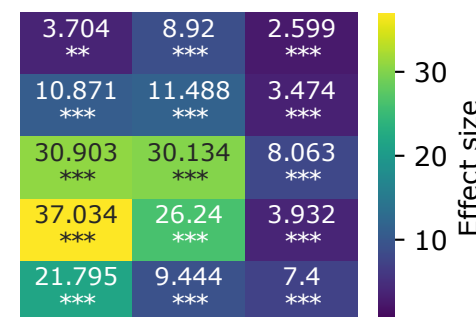
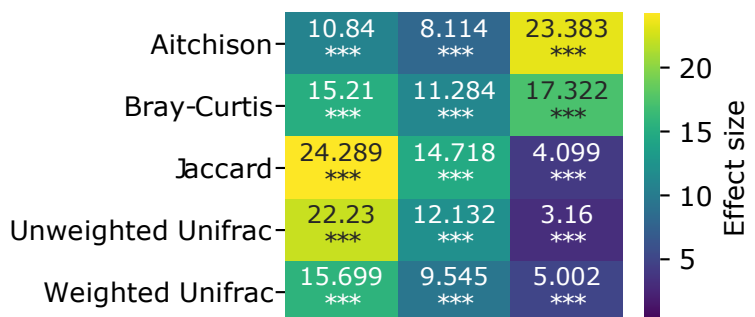
PERMANOVA

PERMANOVA



PERMDISP

PERMDISP



D

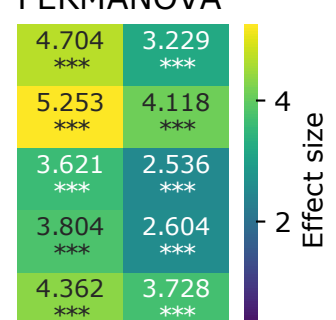
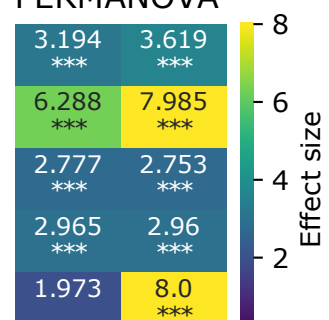
V1V2 - CCZ

DNA samples (n=260)

RNA samples (n=237)

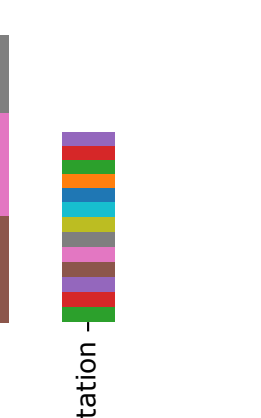
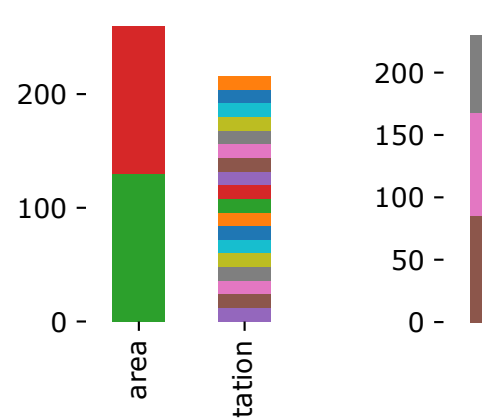
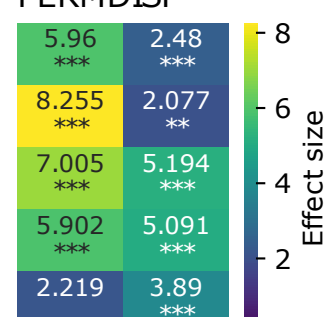
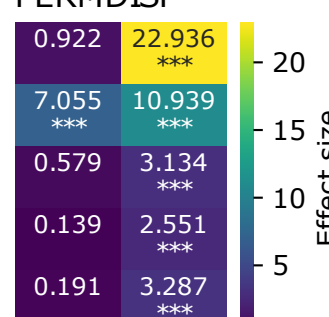
PERMANOVA

PERMANOVA



PERMDISP

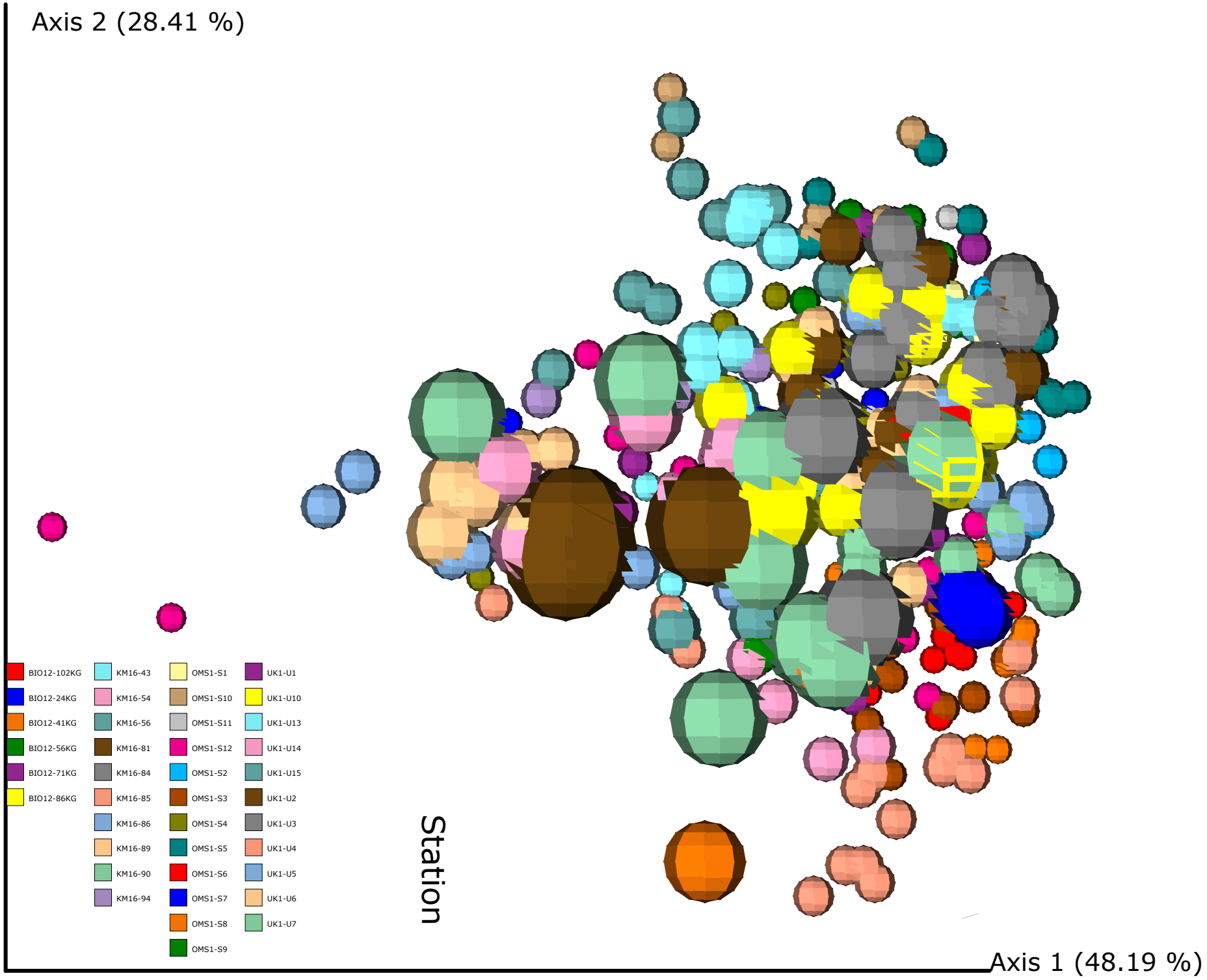
PERMDISP



**Supplementary Figure 19:** Beta-diversity differences between DNA (or RNA) samples groups tested using PERMANOVA and testing for difference in dispersion between groups using PERMDISP. Tests results are shown for (A-B) the 37F benthic Foraminifera, (C-D) the V1V2 (all taxa) and (E-F) the V1V2 Metazoa datasets, and both for (A,C,E) all DNA or all RNA samples, and for (B,D,F) only the CCZ's DNA or RNA samples. The groups of samples interrogated for statistical difference are the columns of each heatmap, with the number of samples in each group indicated by the bar plots (bars colors are not labelled). Heatmaps are color and annotated with the tests' F statistics and the p-values are also annotated (uncorrected): \*\*\*  $\leq 0.01$ , \*\*  $\leq 0.03$ , \*  $\leq 0.05$ , no label  $> 0.05$ . For a PERMANOVA difference to be interpreted, the dispersion of the grouped being compared must not differ.

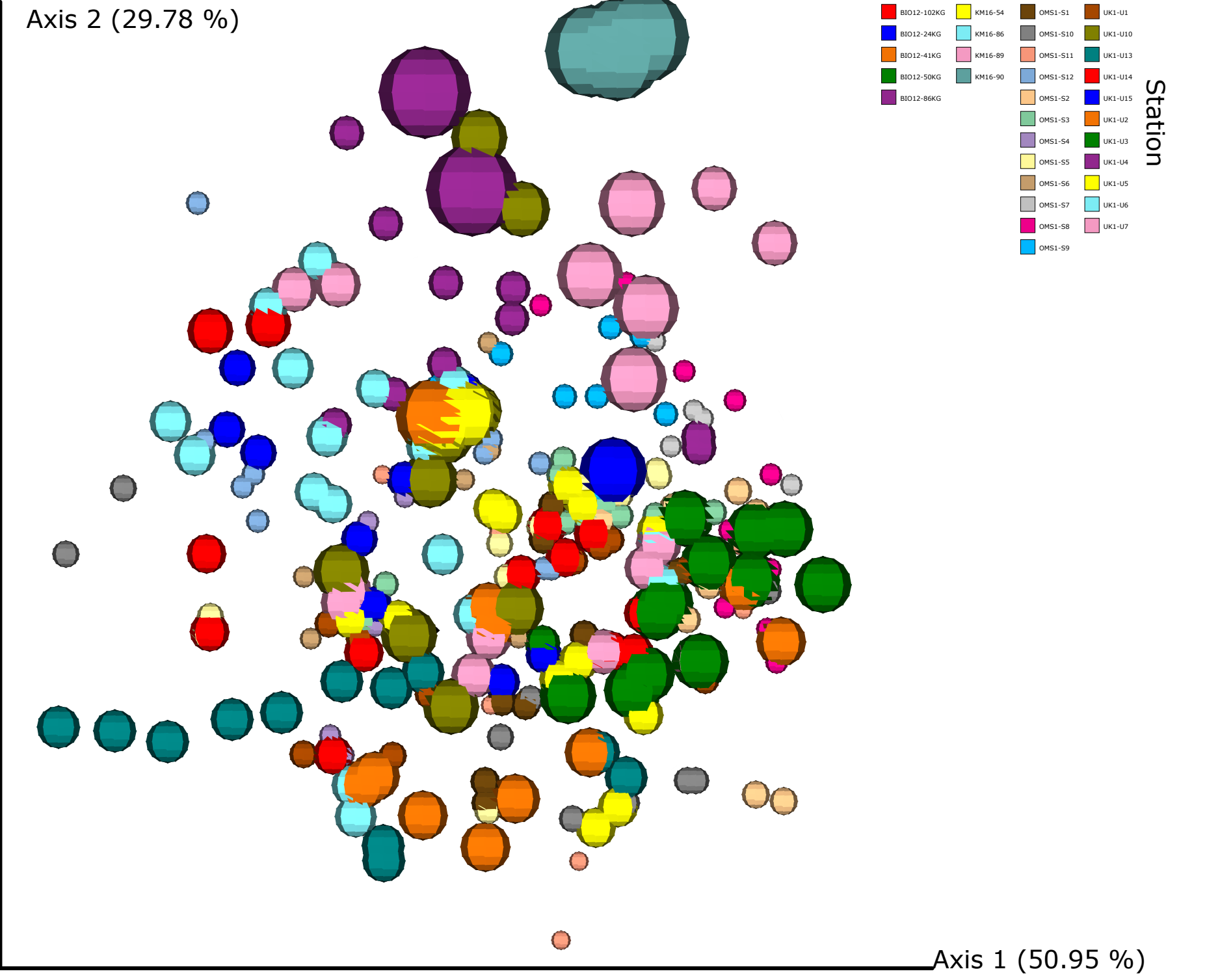
A

37F DNA



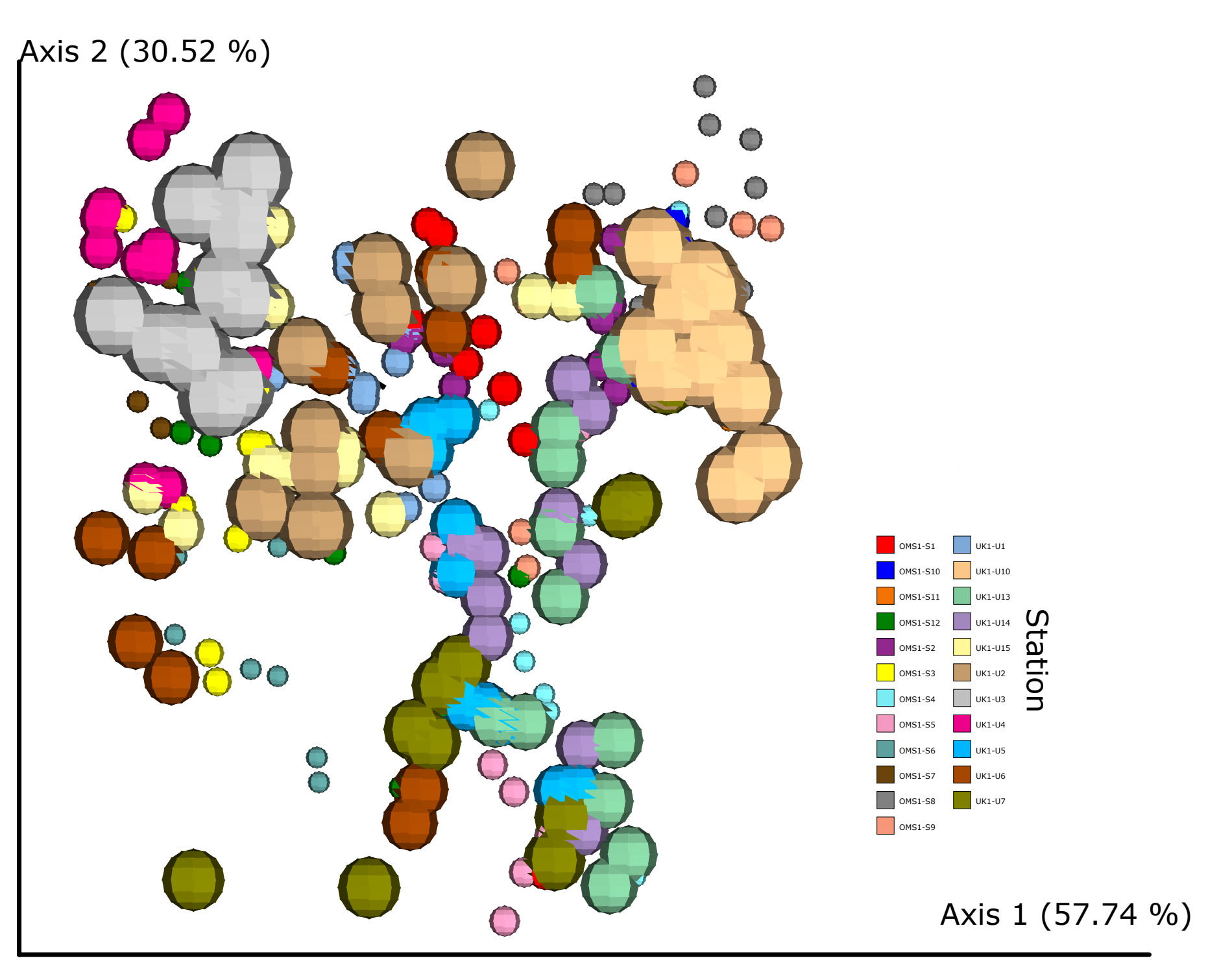
B

37F RNA



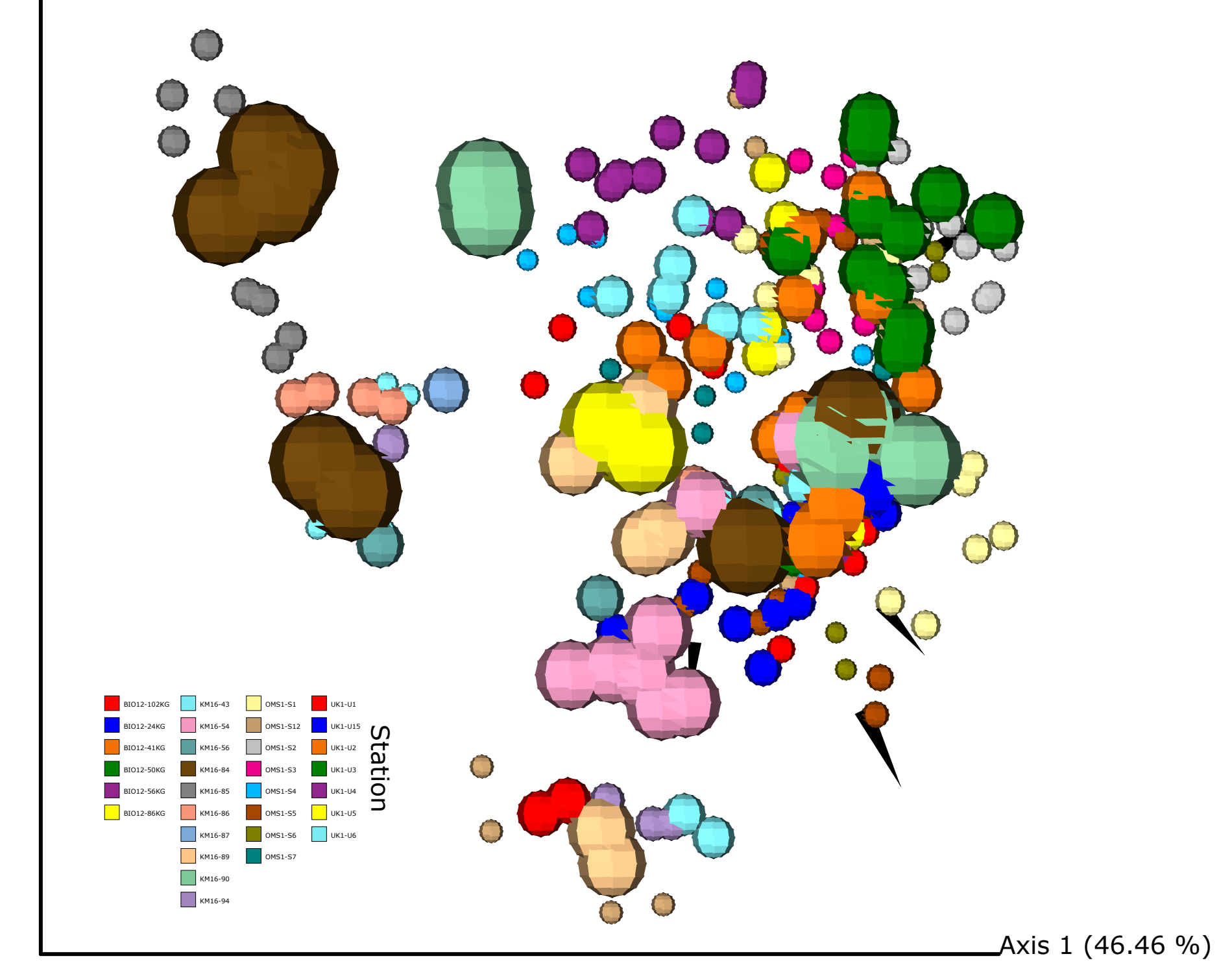
C

V1V2 DNA



D

V1V2 RNA



**Supplementary Figure 20: Robust PCAs performed only on the CCZ samples, separately of each dataset and for DNA or RNA.** RPCA were performed on the CCZ samples separately for (A) the 37F DNA, (B) 37F RNA, (C) V1V2 DNA and (D) V1V2 RNA datasets. Samples are colored per station (scatter) and their sizes indicate the slope of the sea floor (not on the same scale across plots). Inset legends indicate the stations (legend columns per CCZ area). These RPCA were used to measure the distances to centroids for Figure 7 and serve to illustrate groups of station samples for which centroids were calculated and used as references for each sample.