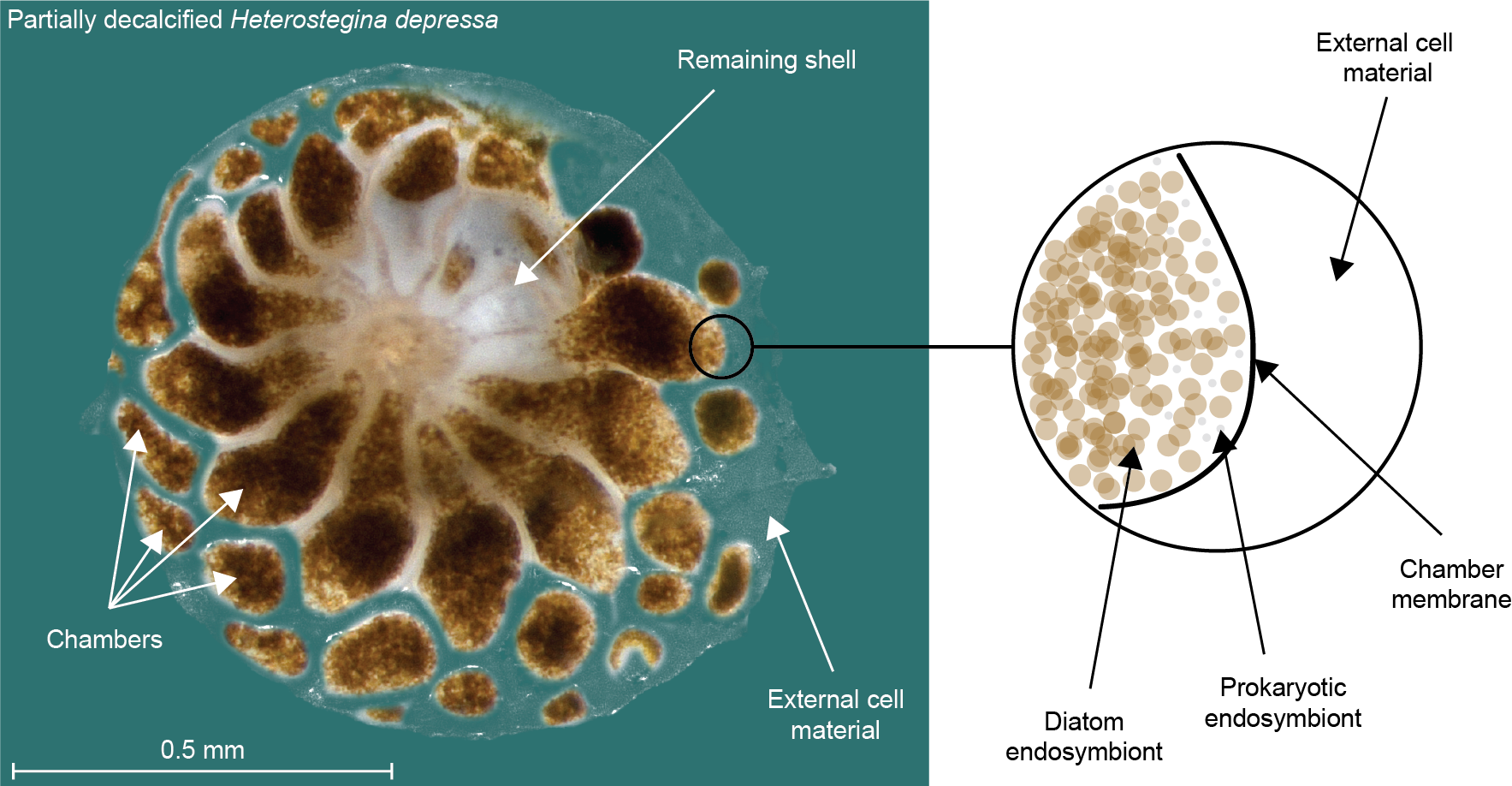
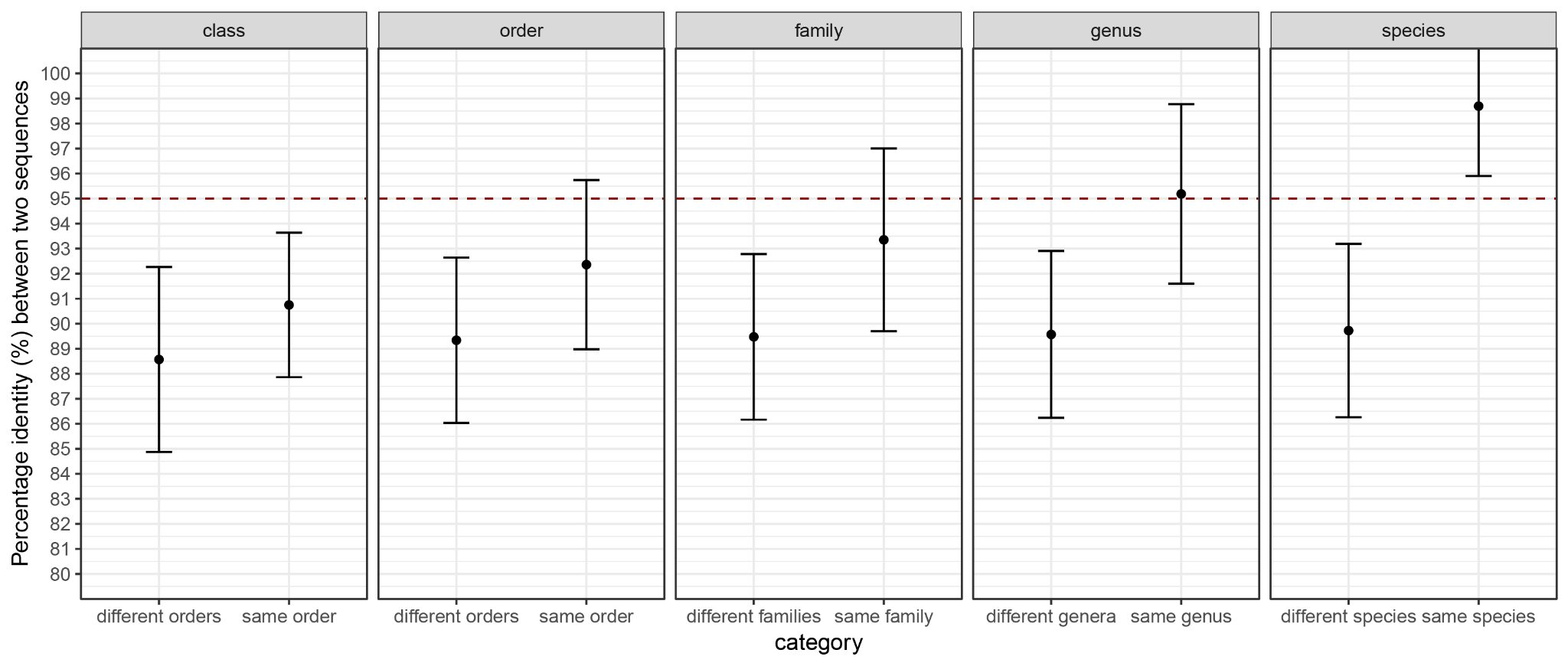
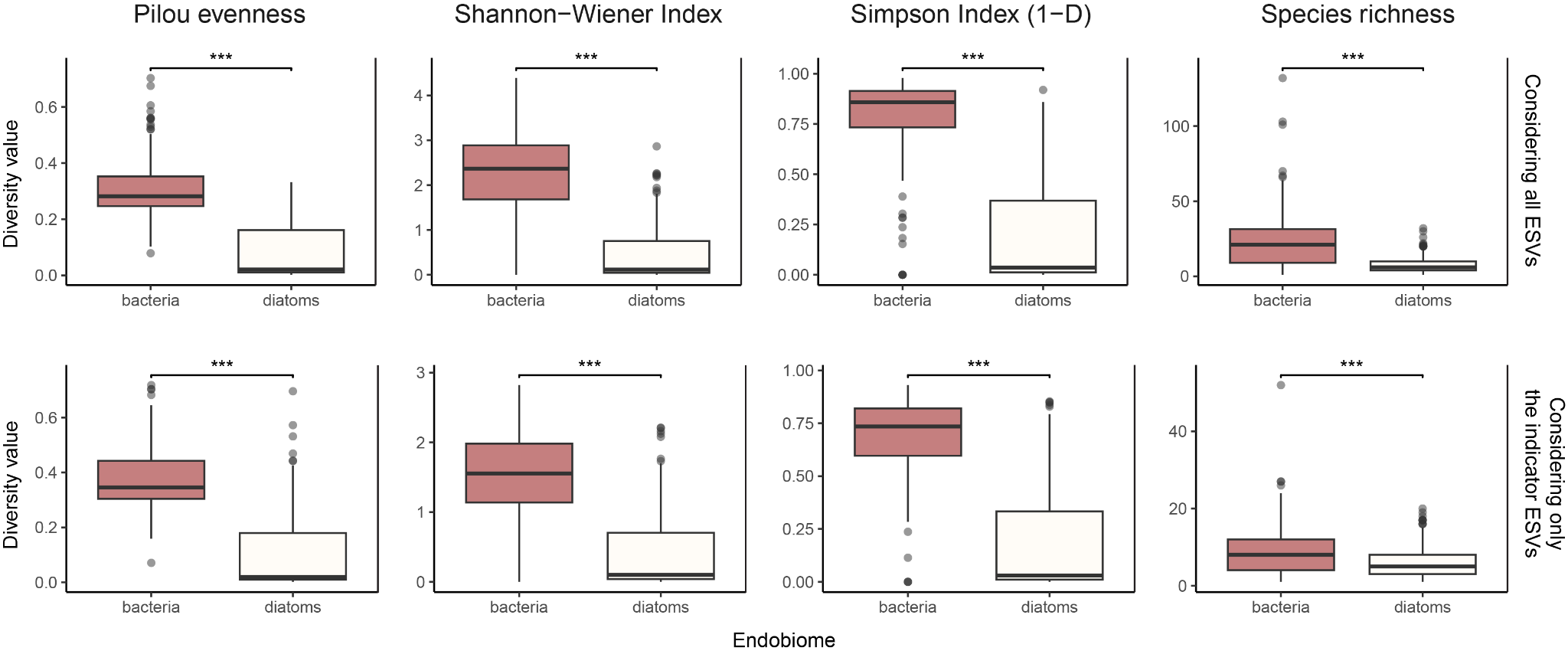
**Supplementary figures**



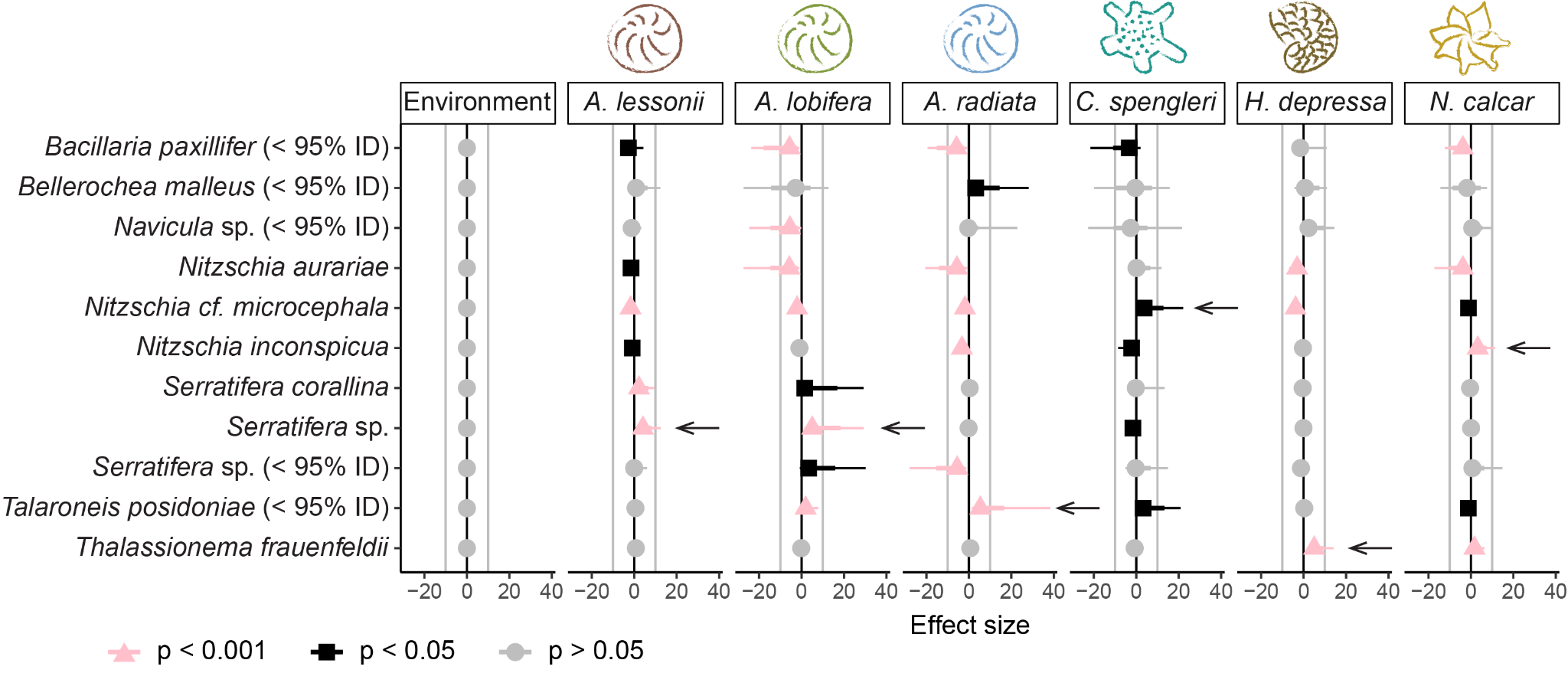
Supplementary Figure S7.1. Schematic representation of the endobionts from a partially decalcified *Heterostegina depressa*, diatom-bearing LBF.

****

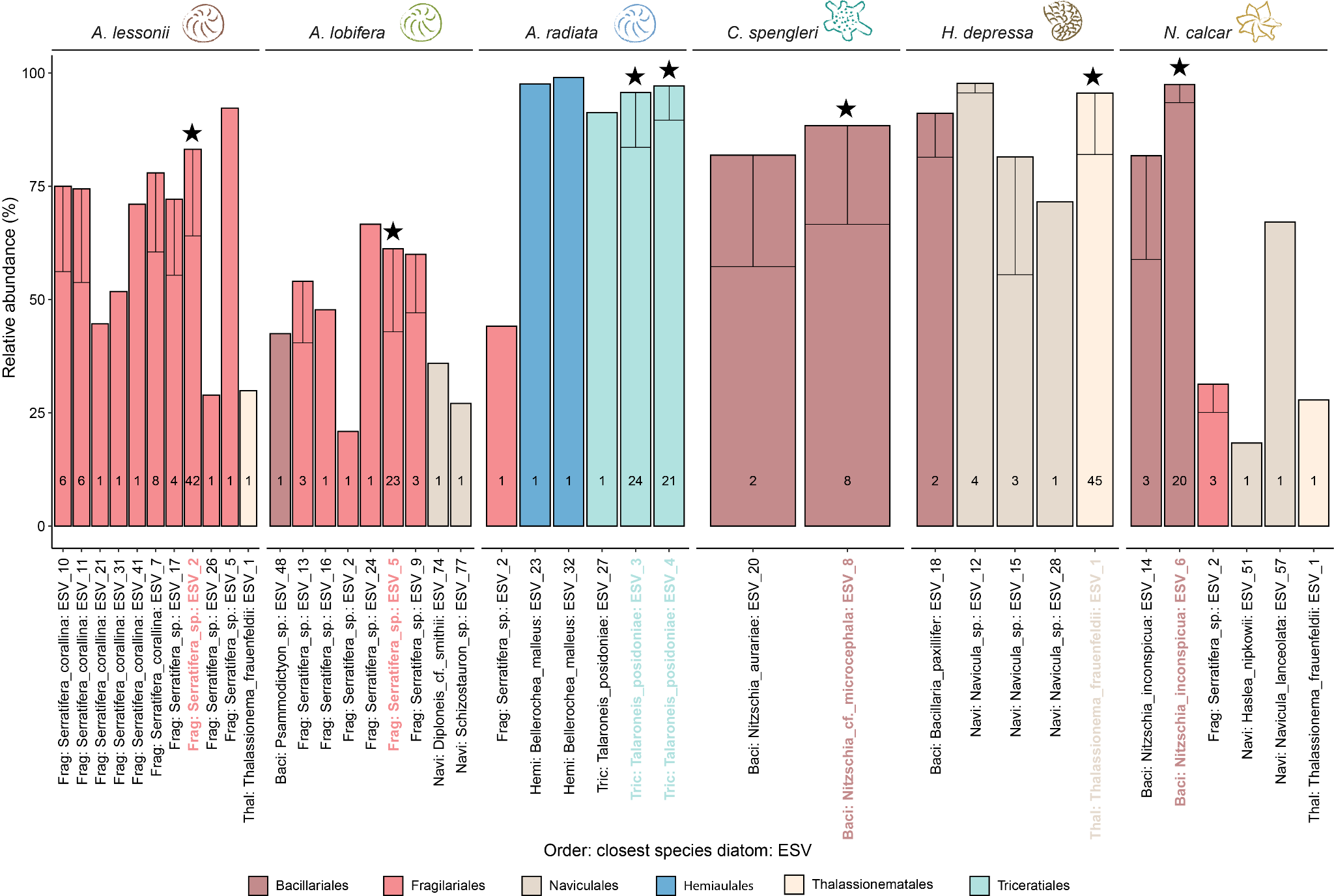
Supplementary Figure S7.2. Using the rbcL reference database from R-syst::diatom version 8 [(Rimet et al., 2016)](https://paperpile.com/c/RAF2Dn/Zhmd), the similarity identity percentage matrix was calculated. The rbcL database contains 4467 sequences covering 1257 species, 267 genera, 88 families, 52 orders and 14 classes within phylum Bacillariophyta. Pairwise comparisons were plotted to confirm the use of 95 % ID as the percentage identity threshold between species. The mean value is displayed and the error bars show the standard deviation from the mean. The red dashed line highlights the threshold commonly used of 95 % ID.

****

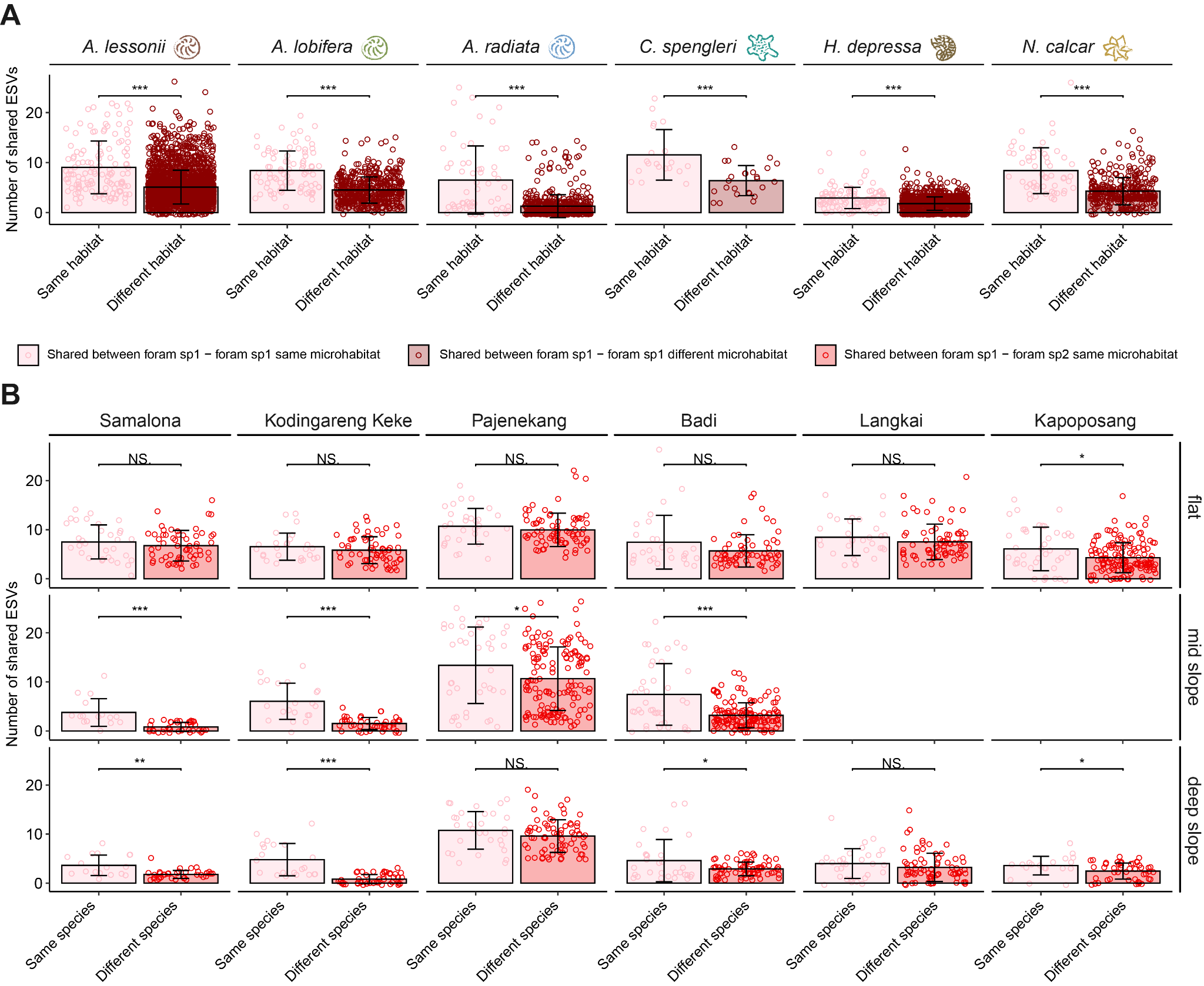
Supplementary Figure S7.3. Diversity indices describing the bacterial and diatom endobiont communities in the foraminifera host. On the top panels, we consider all ESVs after quality filtering and contaminant removal (used for the ANOSIM analysis). On the bottom panels, we consider only the indicator ESVs (used for the analysis of Figure 3 and 4). T-tests were performed to compare the mean diversity values between bacteria and diatom microbiomes, and three asterisks signify a p-value < 0.001.



Supplementary Figure S7.4. Species occupancy modelling showing where the indicator diatom taxon is more likely to be found considering the environmental diatom community as a baseline. Significance is marked by the different colours (pink: highly significant, black: significant, grey: non-significant). This model considers the environmental community (seawater and substrate) a single pool of diatoms. The arrows indicate the primary diatom identified in a majority (> 50%) of specimens for that foraminifera species.



Supplementary Figure S7.5. Mean relative read abundance of the most abundant ESV in every specimen. The number at the bottom of the bars indicates the number of specimens in which this ESV is the most abundant. The error bars, if applicable, display the standard deviation from the mean, applied only towards the lowest value to not exceed 100%. The dominant strain ESV in each foram host is highlighted with a black star and the ESV x label colour-coded based on the Order level.



Supplementary Figure S7.6. Number of share ESVs between two A) specimens from the same species by living either in the same microhabitat (pink) or different microhabitat (dark red), and B) specimens living in the same microhabitat of the same species (pink) or different species (red). T-tests were performed to compare the mean number of shared ESVs between the three categories. ‘\*\*\*’ signifies a p-value < 0.001, ‘\*\*’ signifies a p-value < 0.01, ‘\*’ signifies a p-value < 0.05, ‘NS’ signifies not significant.

**Supplementary tables**

Supplementary Figure S7.1. Sample metadata, including the COI sequence (if available), the COI sequence taxonomical assignment and the percentage identity of the assignment. The samples highlighted in red were removed from the analysis due to sample misidentification or wrong host species revealed by the COI sequence.

Supplementary Figure S7.2. APSCALE report on the sequencing data processing for the rbcL marker, targeted to diatoms.

Supplementary Figure S7.3. APSCALE report on the sequencing data processing for the 16S marker, targeted to bacteria.

Supplementary Figure S7.4. Indicator species analysis results.

Supplementary Figure S7.5. Dunn statistical post-hoc test results. The rows highlighted in red show non-significant results.