Science Advances
AAAS

# Supplementary Materials for

## Specific host-algae relationship, yet flexible bacterial microbiome, in diatom-bearing foraminifera

Elsa B. Girard* *et al.*

*Corresponding author. Email: elsa.girard@naturalis.nl

**This PDF file includes:**

Figs. S1 to S7
Tables S1 to S4

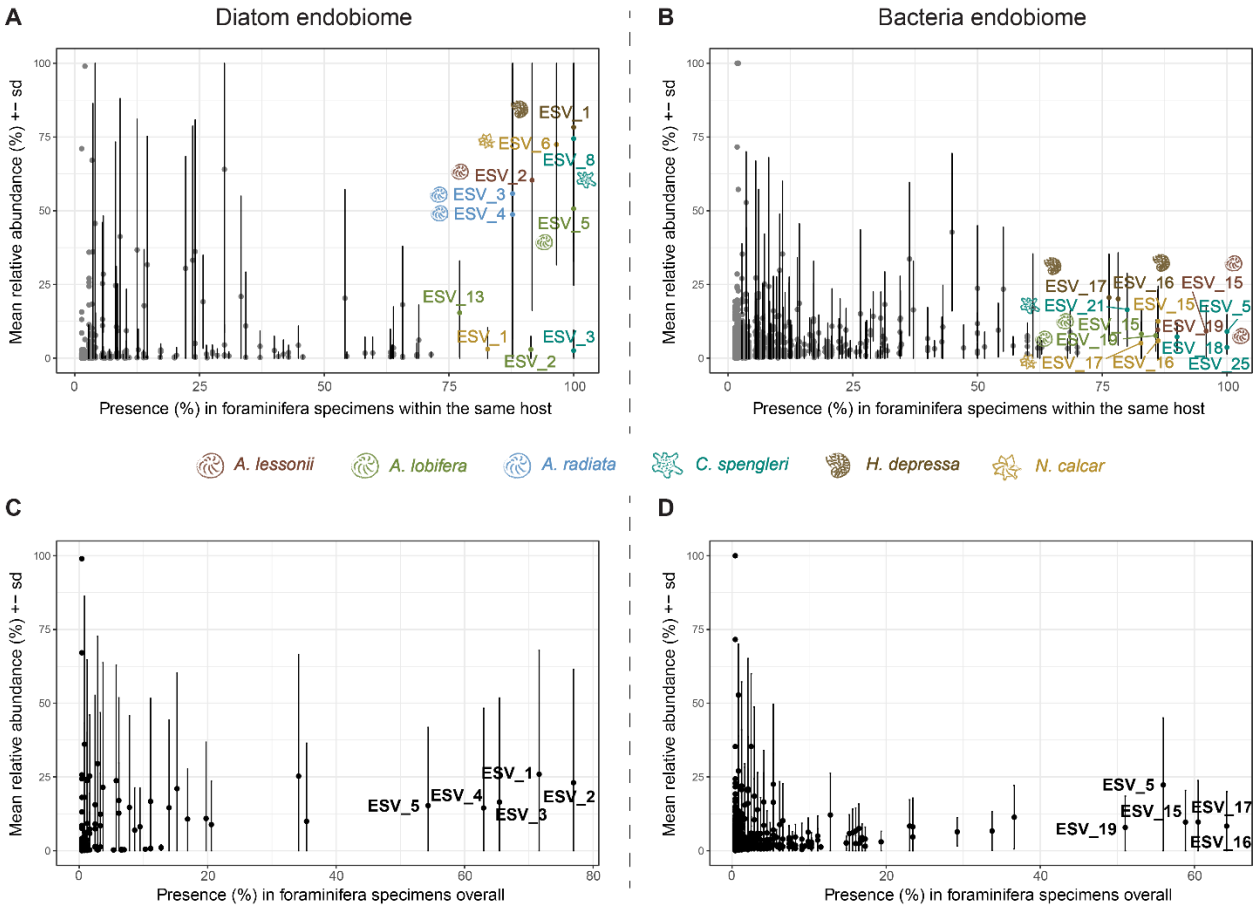


**Fig. S1.**

Distribution of all ESVs among specimens of the same species (**A, B**) and overall (**C, D**) for the diatom endobiome (**A, C**) and the bacteria endobiome (**B, D**). The mean relative abundance (%) and the standard deviation is displayed (values limited between 0 and 100 %). **A, B**) ESVs present in >75 % of the specimens per host species are labelled and coloured based on the host. The primary endobiont is highlighted with the foram host icon if it reaches >50 % mean relative read abundance across its host specimens. **C, D**) ESVs present in >50 % of all specimens are labelled.

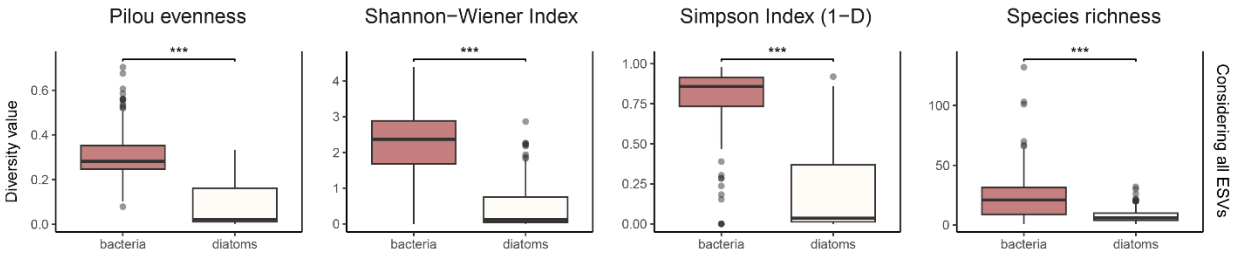


**Fig. S2.**

Diversity indices describing the bacterial and diatom endobiont communities in the foraminifera host. We consider all ESVs after quality filtering and contaminant removal. T-tests were performed to compare the mean diversity values between bacteria and diatom microbiomes, and three asterisks signify a p-value < 0.001.

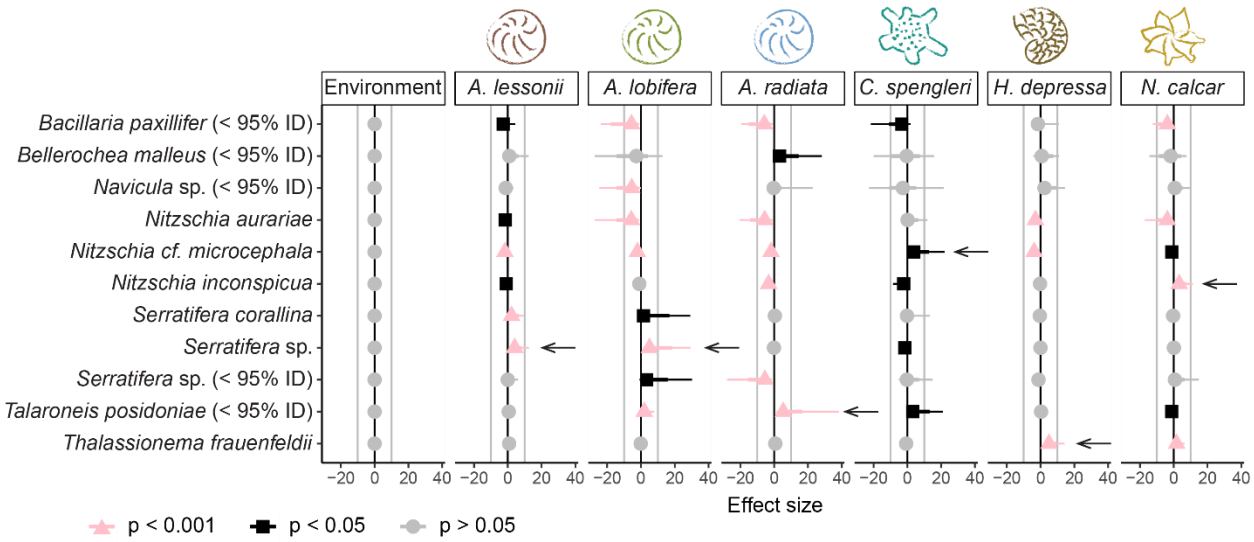


**Fig. S3.**

Species occupancy modelling showing where the primary 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.

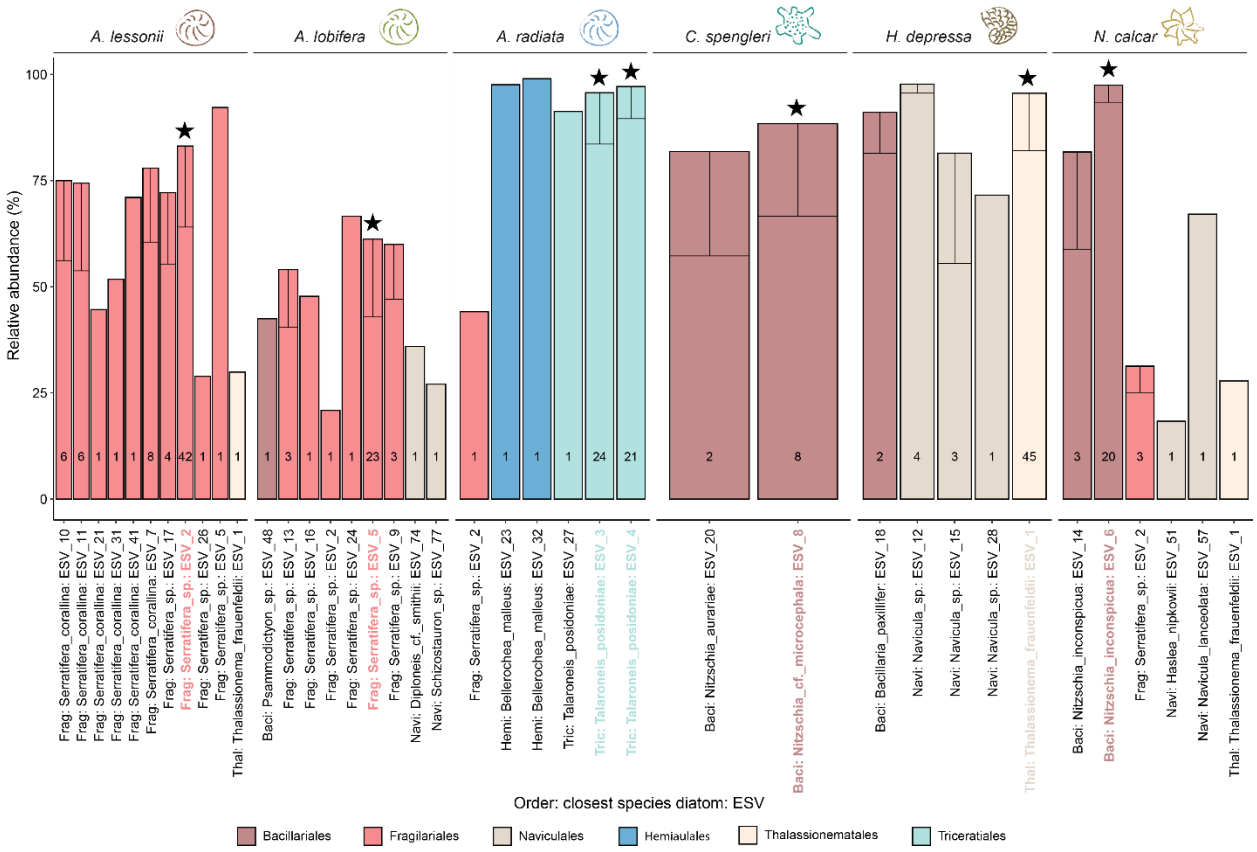


**Fig. S4.**

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 label is colour-coded based on the Order level.

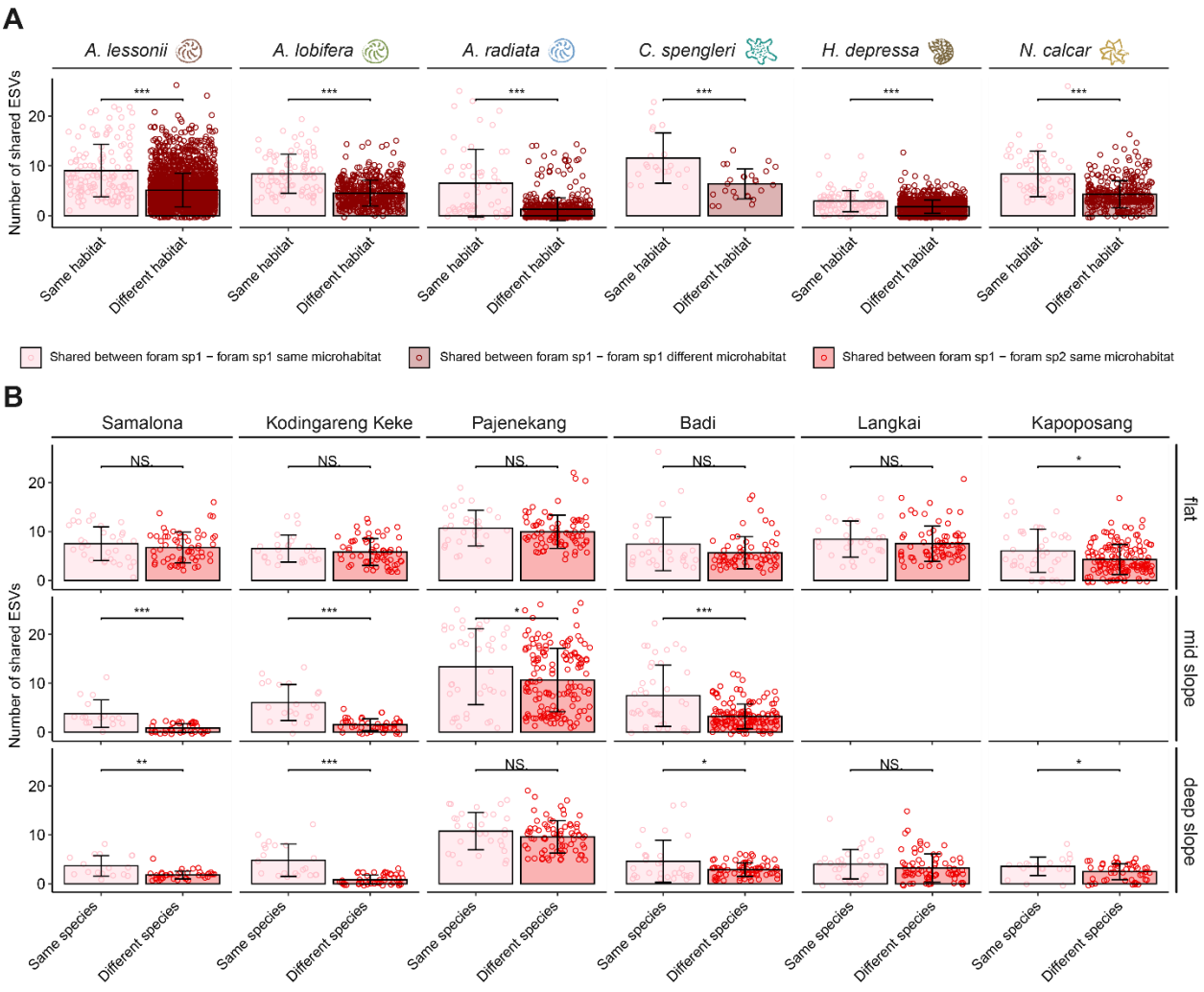


**Fig. S5.**

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.

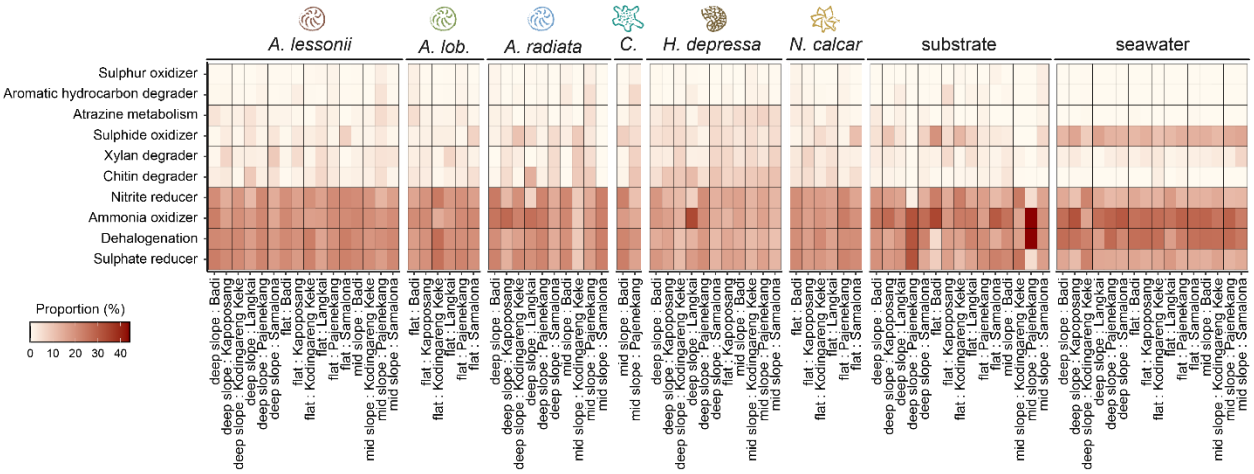


**Fig. S6.**

Ten most abundant bacterial metabolisms inferred from taxonomy. The colour gradient reflects the relative abundance of the bacteria associated to a metabolic phenotype in the bacterial endobiomes from the six foram host species and the environment. The darker the red, the more abundant.

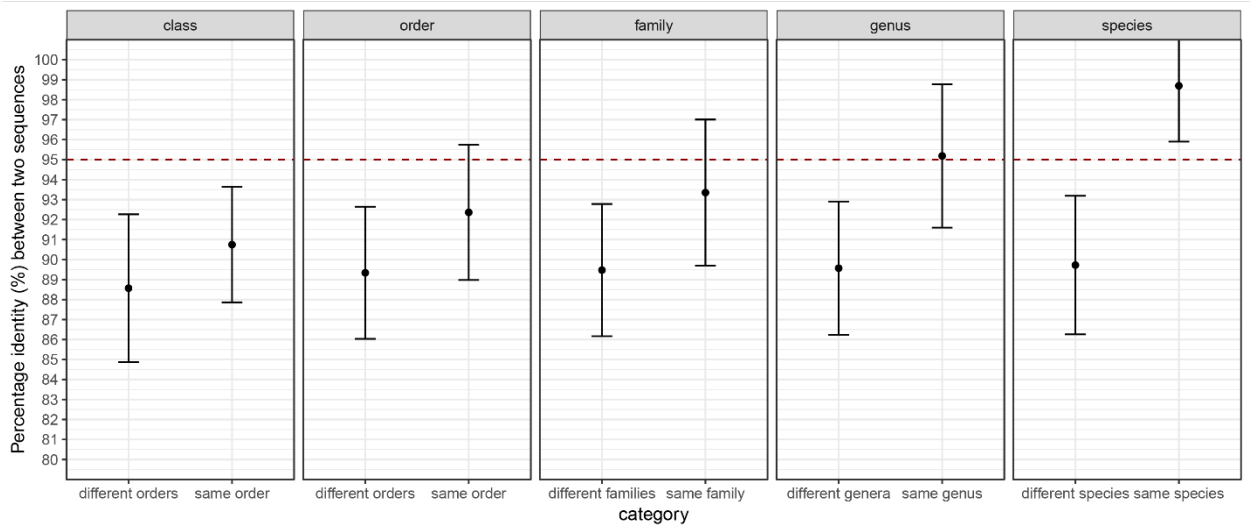


**Fig. S7.**

Using the rbcL reference database from R-syst::diatom version 8 (Rimet et al., 2016) (44), 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.

**Table S1. (separate file)**

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

**Table S2. (separate file)**

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.

**Table S3. (separate file)**

APSCALE report on the sequencing data processing for the rbcL marker, targeted to diatoms.

**Table S4. (separate file)**

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

Note: Supplementary tables are combined into one separate Excel file.