Supporting Information

MICRODIVERSE BACTERIAL CLADES PREVAIL ACROSS ANTARCTIC WETLANDS

María V. Quiroga, James C. Stegen, Gabriela Mataloni, Don Cowan, Pedro H. Lebre, Angel Valverde

Corresponding authors contact: mvquiroga@iib.unsam.edu.ar, angel.valverde@csic.es

Supporting Methods

Supporting Figures S1-S6

Supporting Tables S1

SUPPORTING METHODS

Phyloscore index

For each ASV, its phyloscore (Fodelianakis et al., 2022) was calculated for all pair of communities in which the given ASV was present in one community and absent in the other. Briefly, for a given ASV *i* present in community *j* and absent in community *k*, the nearest phylogenetic distance $d_{i,j,k}$ of ASV *i* to the ASVs present in community *k* was computed. Then, a null community was generated by randomly sampling the number of ASVs present in community *k* from the metacommunity pool of ASVs minus ASV *i*, and the nearest phylogenetic distance $d^0_{i,j,k}$ to our focal ASV *i* was computed. This process was repeated m = 100 times to estimate the natural logarithm-transformed average $\langle \ln d^0_{i,j,k} \rangle_m$ and standard deviation $\sigma^0_{i,j,k}$ of the nearest phylogenetic distance null model distribution. The phyloscore $Z_{i,j,k}$ for ASV *i* in communities *j* and *k* pair was calculated as:

$$Z_{i,j,k} = \frac{\ln\left(d_{i,j,k}\right) - \langle \ln d^0_{i,j,k} \rangle_m}{\sigma^0_{i,j,k}} \qquad (1)$$

Finally, we calculated the sum, mean and median values of the phyloscore for each ASV across all community pairs. These phyloscore values were used as input for phylofactorization, phylofactor R package (Washburne et al., 2019). We implemented a two-sample Wilcoxon test because the phyloscore values (i.e., sum, mean and median) did not show a normal distribution, and ten iterations (i.e, factors). As the choice of metric altered the phylofactorization results (Figure S1), only those clades consistently identified with the three phyloscore values were retained for further analysis (Figure 1d). Finally, we further worked with the average phyloscore values, named hereafter as phyloscore for simplicity (Figure 1c), because the clades detected with this metric were identical or smaller than those detected with the sum or median phyloscore.

Feature-level βNTI (βNTI$_{feat}$) index

βNTI$_{feat}$ is based on the βNTI equation (Danczak et al., 2022). Briefly, βMNTD$_{feat}$ measures the average minimum phylogenetic distance between a given ASV in one community and its nearest ASV in other communities. The main difference from the standard βMNTD equation is that βMNTD$_{feat}$ compares a given ASV from one community to all other communities at a single time. βMNTD$_{feat}$ was calculated for the observed data $\beta\text{MNTD}^{obs}_{feat}$ and randomized communities. The average value $\langle \beta\text{MNTD}^{null}_{feat} \rangle$ and the standard deviation $\beta\text{MNTD}^{sd}_{feat}$ were estimated for the null model distribution with 999 replicates, and βNTI$_{feat}$ was calculated as

$$\beta\text{NTI}_{\text{feat}} = \frac{\beta\text{MNTD}^{obs}_{feat} - \langle \beta\text{MNTD}^{null}_{feat} \rangle}{\beta\text{MNTD}^{sd}_{feat}} \qquad (2)$$

We did not remove conspecifics (rm.conspec = F) because our phylogenetic tree had 0-length branches. We ran the βNTI_{feat} code considering the relative abundance of the ASVs (abund.weig = T) or the presence-absence data (abund.weig = F). Surprisingly, we did not find significantly low βNTI_{feat} values (Figure S2a, b). Considering that our previous study revealed an ASV metacommunity strongly shaped by homogeneous selection (HoS) (Quiroga et al., 2022), and that the phyloscore approach successfully detected several HoS clades, we applied some modifications to the βNTI_{feat} code. To optimize running time, we first calculated the average $\beta \mathrm{MNTD}_{feat}^{obs}$ for each ASV and modified its code following the average $d_{i,j,k}$ index (Fodelianakis et al., 2022), also called nearest taxon distance (NTD). Without removing conspecifics and using only presence-absence data, once comparisons among community pairs where the focal ASV was present in both communities were completely removed from the $\beta \mathrm{MNTD}_{feat}^{obs}$ code, we obtained the same mean values as for the NTD index (Figure S3). Therefore, we applied this first modification to all the $\beta MNTD_{feat}$ calculations in the βNTI_{feat} code yet obtaining none significantly low values of the latter index (Figure S2c). Thus, we compared the phyloscore (1) and $\beta \mathrm{NTI}_{\mathrm{feat}}$ (2) equations and performed a second modification to the code: include a natural logarithm-transformation to the $\beta \mathrm{MNTD}_{feat}^{obs}$, $\langle \beta \mathrm{MNTD}_{feat}^{null} \rangle$ and $\beta \mathrm{MNTD}_{feat}^{sd}$ indexes in equation (2). As we had 0-length branches in the phylogenetic tree, to be able to apply the ln-transformation we had to transform the cophenetic distance object replacing the ceros with the minimum phylogenetic distance across the tree, as suggested by the phyloscore script (Fodelianakis et al., 2022). Remarkably, this modified code estimated significantly low $\beta \mathrm{NTI}_{\mathrm{feat}}$ values and the index showed a less skewed distribution (Figure S2d). Then, to estimate $\beta \mathrm{NTI}_{\mathrm{feat}}$ with a number of null model replicates similar to the phyloscore approach, we randomly subsampled only 100 null communities for the null model distribution (Figure S2e). This final modification of the index will be named hereafter as $\beta \mathrm{NTI}_{\mathrm{feat}}$ for simplicity (Figure 1b), and its code is available at GitHub (https://github.com/mvquiroga/NullModels).

References

Danczak, R. E., Sengupta, A., Fansler, S. J., Chu, R. K., Garayburu-Caruso, V. A., Renteria, L., Toyoda, J., Wells, J., & Stegen, J. C. (2022). Inferring the Contribution of Microbial Taxa and Organic Matter Molecular Formulas to Ecological Assembly. *Frontiers in Microbiology*, *13*, 803420. https://doi.org/10.3389/fmicb.2022.803420

Fodelianakis, S., Washburne, A. D., Bourquin, M., Pramateftaki, P., Kohler, T. J., Styllas, M., Tolosano, M., De Staercke, V., Schön, M., Busi, S. B., Brandani, J., Wilmes, P., Peter, H., & Battin, T. J.

(2022). Microdiversity characterizes prevalent phylogenetic clades in the glacier-fed stream microbiome. *The ISME Journal*, *16*(3), 666–675. https://doi.org/10.1038/s41396-021-01106-6

Quiroga, M. V., Valverde, A., Mataloni, G., Casa, V., Stegen, J. C., & Cowan, D. (2022). The ecological assembly of bacterial communities in Antarctic wetlands varies across levels of phylogenetic resolution. *Environmental Microbiology*, *24*(8), 3486–3499. https://doi.org/10.1111/1462-2920.15912

Washburne, A. D., Silverman, J. D., Morton, J. T., Becker, D. J., Crowley, D., Mukherjee, S., David, L. A., & Plowright, R. K. (2019). Phylofactorization: A graph partitioning algorithm to identify phylogenetic scales of ecological data. *Ecological Monographs*, *89*(2), 1–27. https://doi.org/10.1002/ecm.1353

SUPPORTING FIGURES

Figure S1 The sum (a), mean (b) and median (c) values of phyloscore index for each ASV, and the phylogenetic clades identified with phylofactorization (d, e, f), respectively. ASV order in the barplots aligns to the phylogenetic tree, and barplot colours match those of the distinct clades. Only those clades consistently identified with the three metrics were retained for further analysis, and the average metric was hereafter called phyloscore for simplicity (Figure 1c, d).

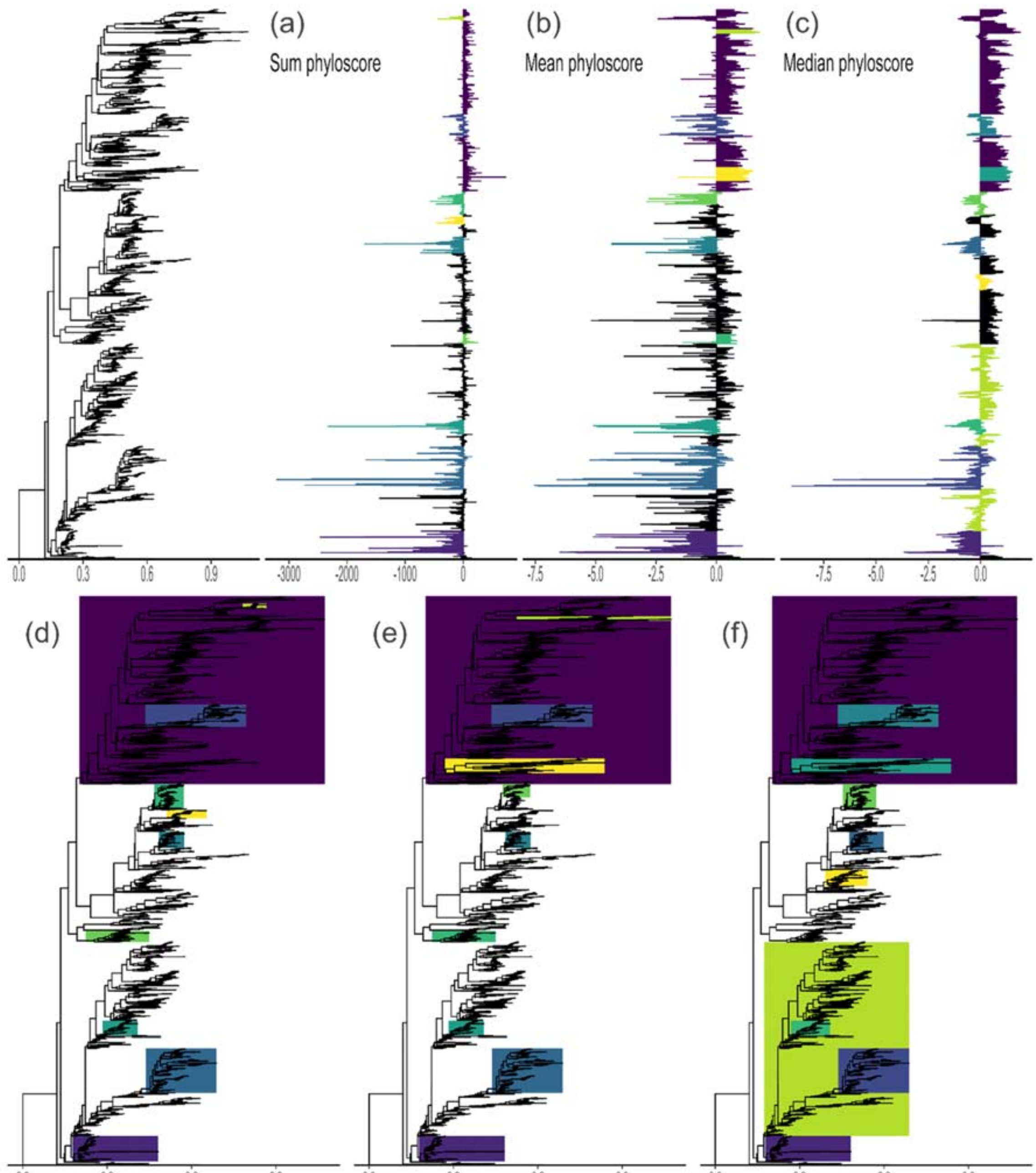


Figure S2 (above) βNTI_{feat} values estimated without removing conspecifics -CF, (a) weighting by relative abundance -AT, (b) considering only presence-absence data -AF, (c) removing all community pair comparisons where the focal ASV was present in both samples -RP, (d) applying natural logarithm-transformation -Ln, with 999 replicate null communities -999rep, or (e) 100 replicate null communities -100rep. Colour scale: significance thresholds (i.e., -2, -1, 1, 2) suggested by Danczak et al. (2022). (below) βNTI_{feat} histograms and their skewness values.

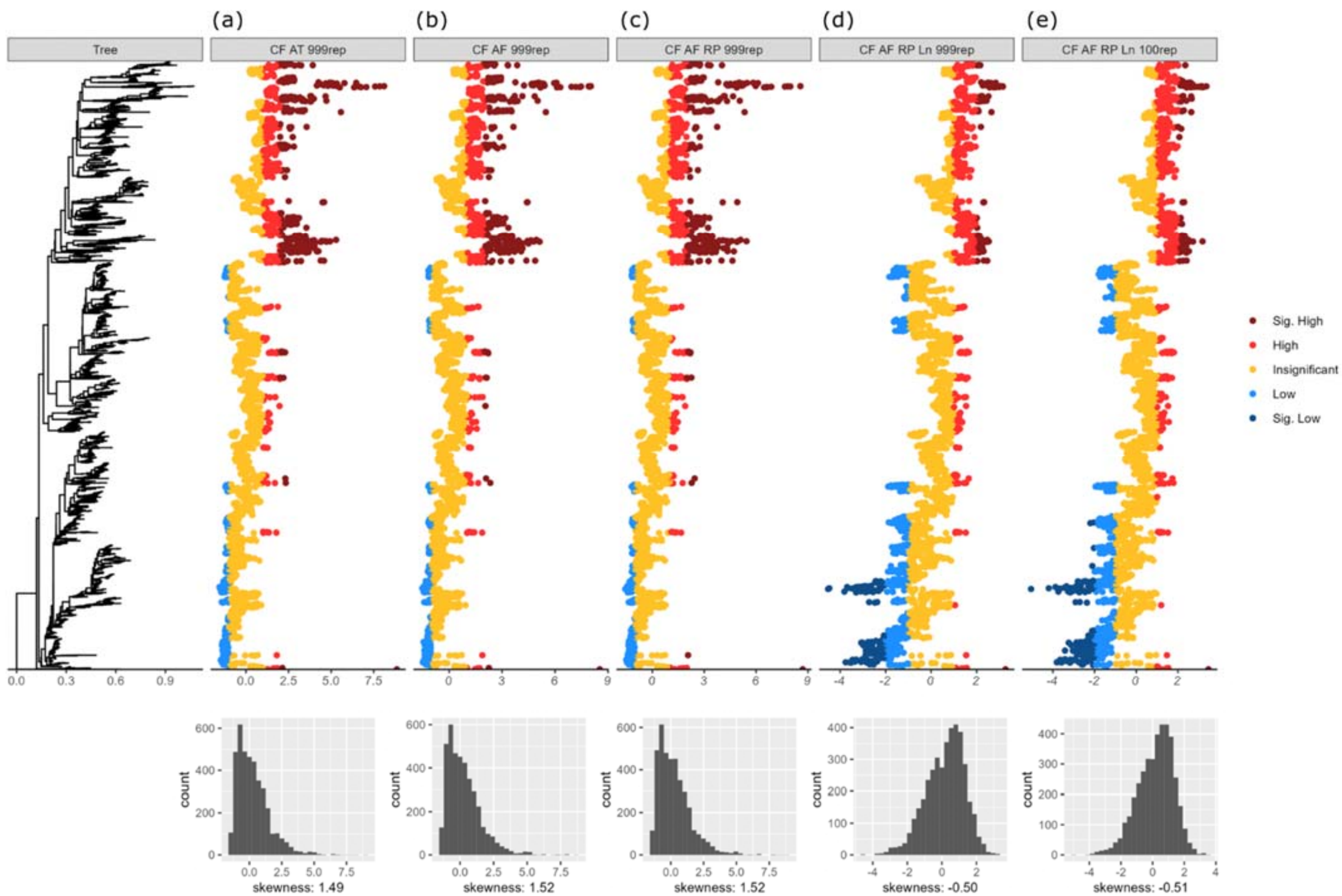


Figure S3 Relationship between $\beta\mathrm{MNTD}_{feat}^{obs}$ and NTD, both calculated as mean values per ASV. This $\beta\mathrm{MNTD}_{feat}$ modified code was implemented in Figure S2c. Red line: 1:1 line.

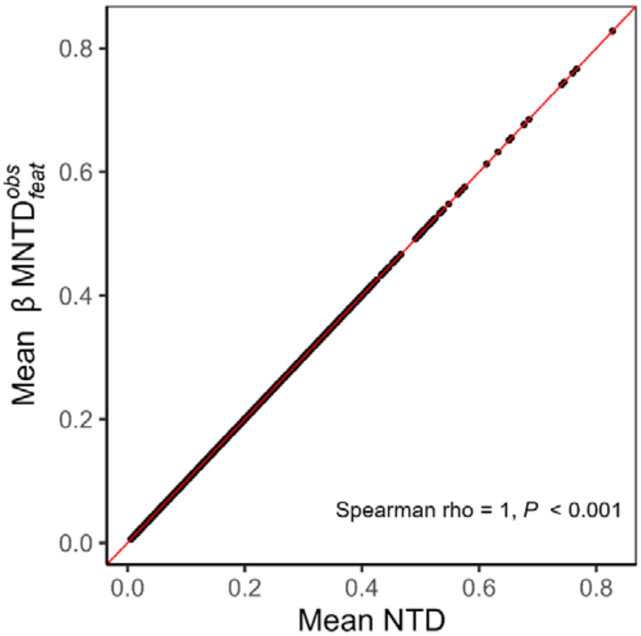


Figure S4 Relationship between the natural logarithm-transformed mean NTD and the phylogenetic depth (i.e., ASV distance from the root). The linear regression model is shown.

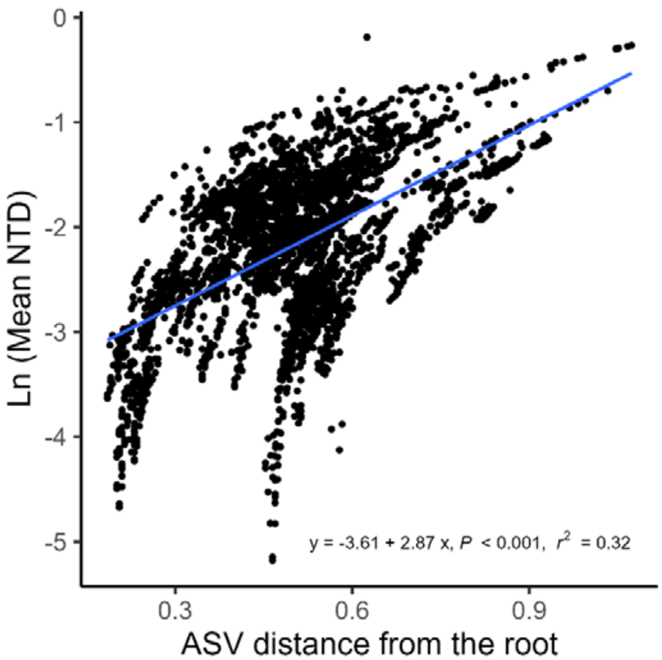


Figure S5 Rarefaction curves for the selected phylogenetic groups and outgroup. Different curves in the same plot represent different samples. #: number of, HeS: heterogeneous selection, HoS: homogeneous selection clade, non-S: non selection clades (outgroup).

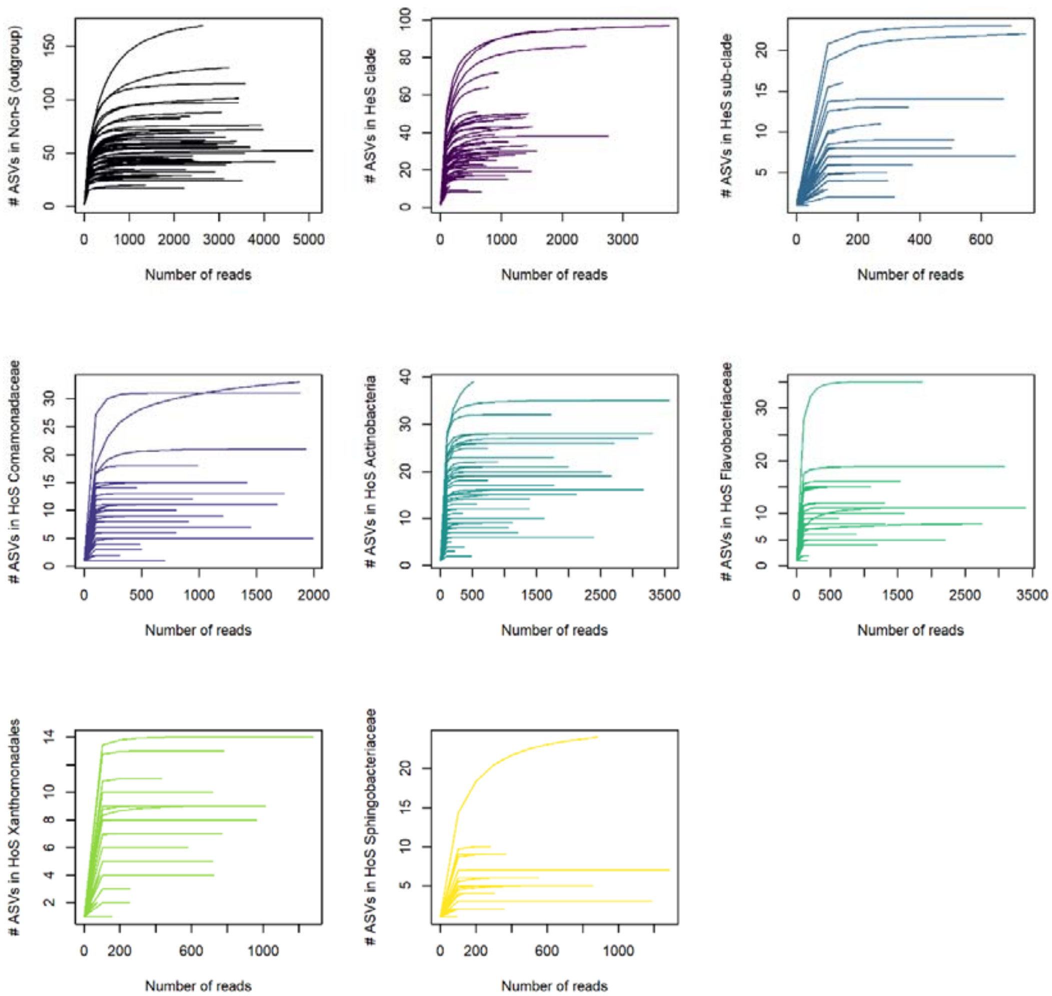


Figure S6 Heatmaps display average sample reads per environment (colour bars). Subtitles indicate the consensus taxonomy of the HoS clade (class, order or family), the genus of the OTU97 and the number of ASVs per OTU97 between brackets. ASVs were clustered using Bray Curtis distance (Hellinger-transformed data) and Ward2 agglomeration method. HoS, homogeneous selection.

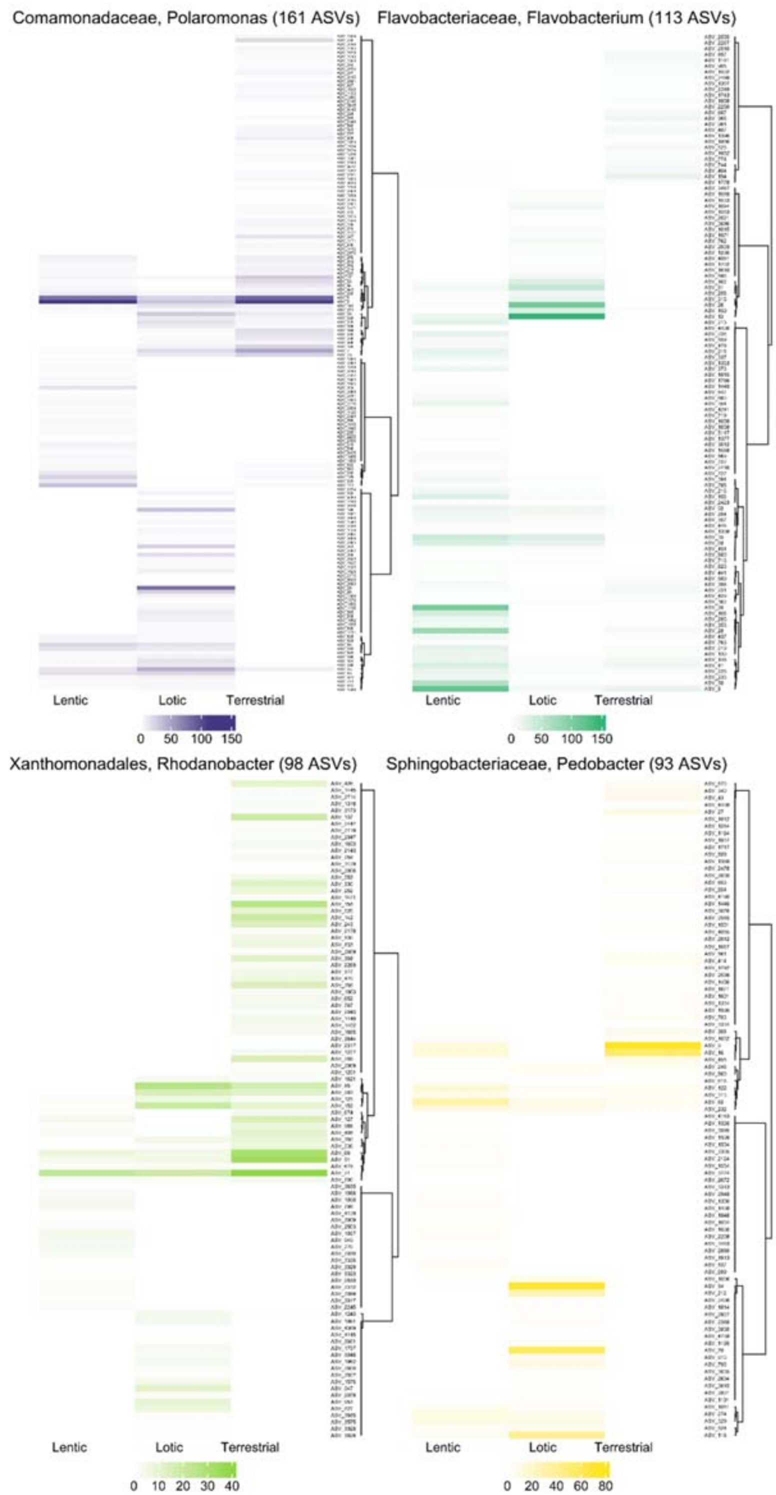
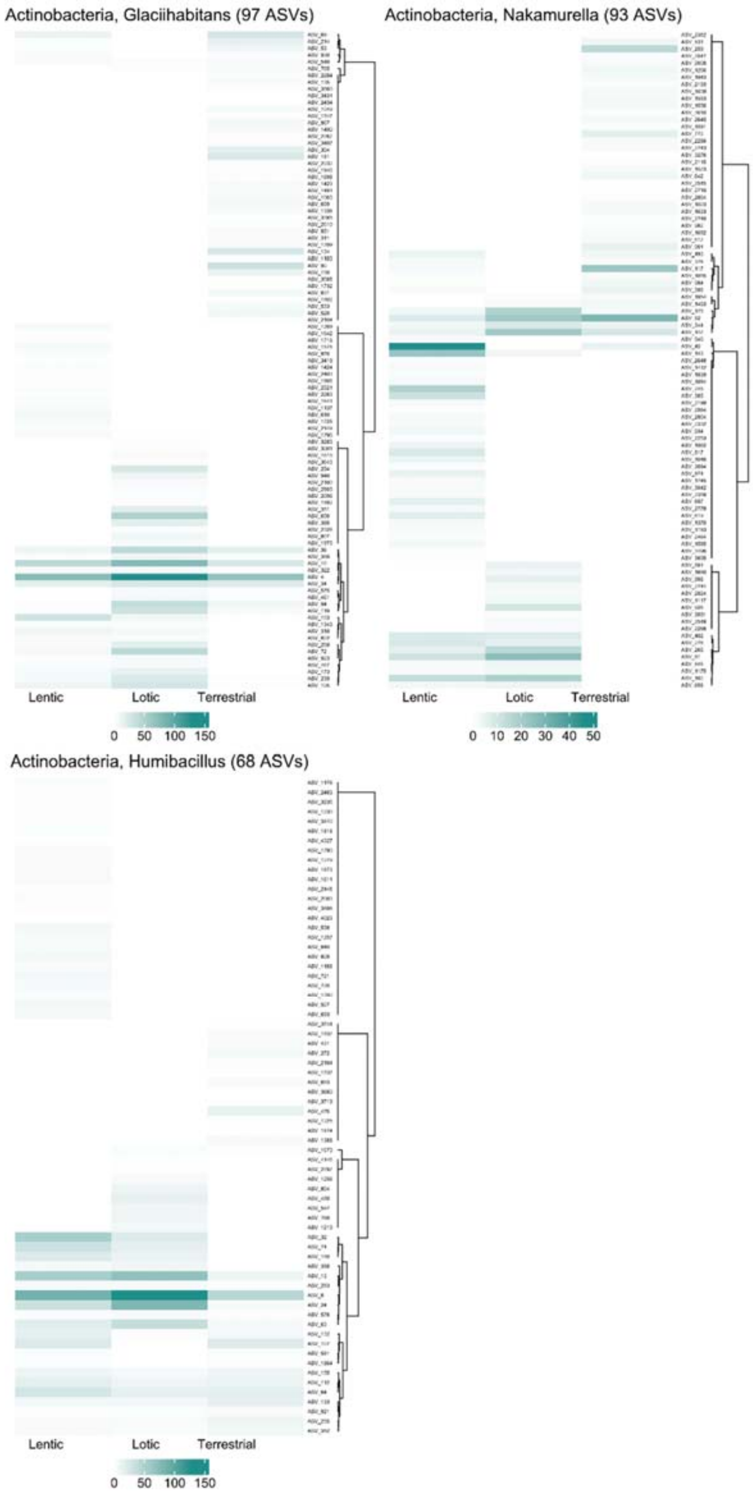


Figure S6 (cont.).



SUPPORTING TABLES

Table S1 Phylogenetic clades consistently identified with phyloscore and βNTI_{feat} approaches.

Phylogenetic clade consensus taxonomy	**No. of ASVs Phyloscore**	**No. of ASVs βNTI_{feat}**	**Phyloscore value** mean (SD)	**βNTI_{feat} value** mean (SD)
HeS clade (27 Phylum)	1311	1311	0.51 (0.53)	1.21 (0.61)
HeS sub-clade, Firmicutes (Phylum)	157	157	-0.09 (0.59)	0.32 (0.45)
HoS clade Actinobacteria (Class)	309	309	-0.73 (1.23)	-1.11 (0.94)
HoS clade Xanthomonadales (Order)	102	80	-0.89 (1.10)	-1.20 (0.42)
HoS clade *Comamonadaceae* (Family)	178	178	-1.39 (1.31)	-1.99 (0.85)
HoS clade *Flavobacteriaceae* (Family)	116	116	-0.78 (0.86)	-0.92 (0.37)
HoS clade *Sphingobacteriaceae* (Family)	93	168	-0.55 (0.57)	-0.74 (0.48)
Non-S clades (Outgroup)	1851	1798	-0.004 (0.602)	0.09 (0.72)