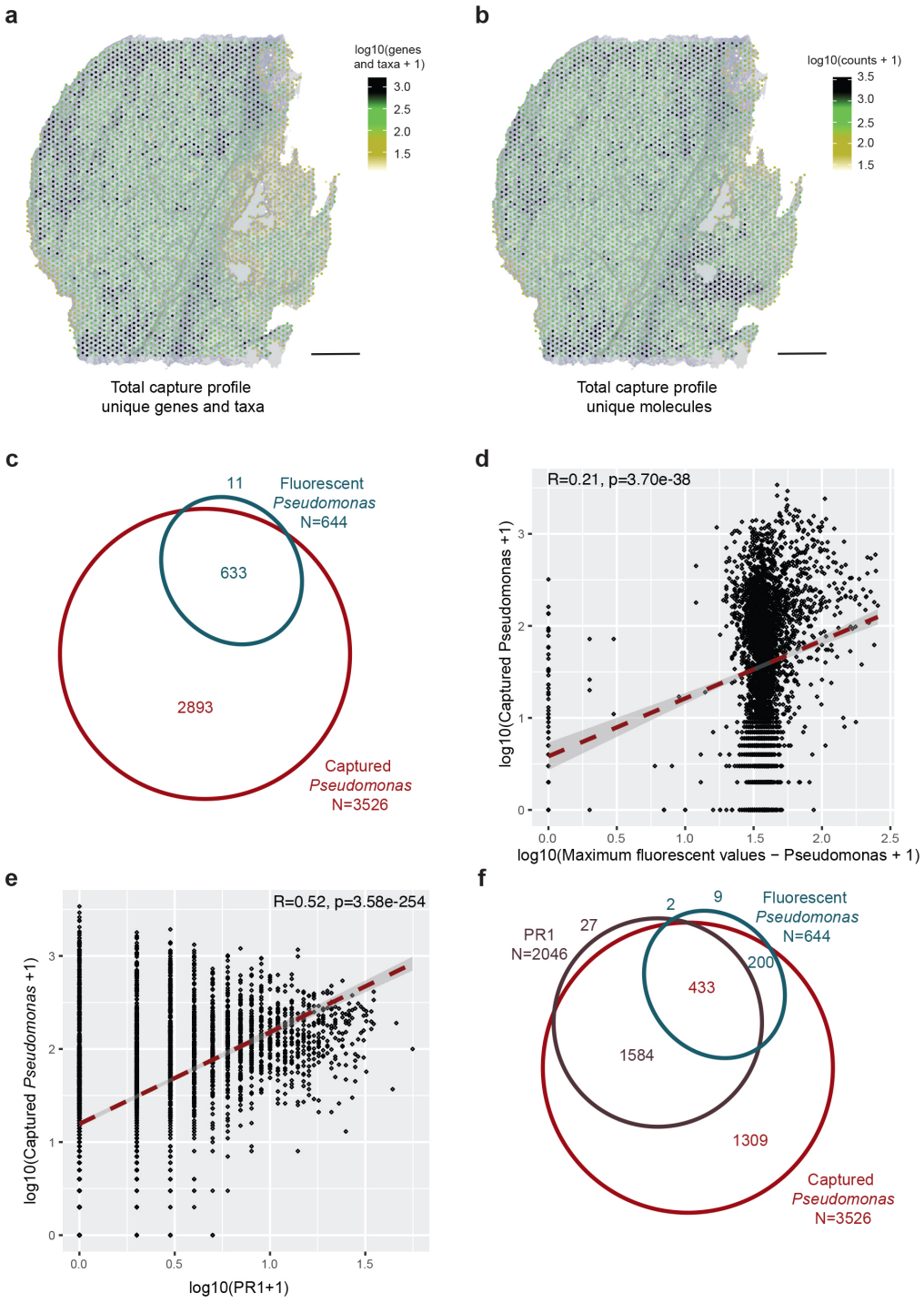
nature biotechnology

Article

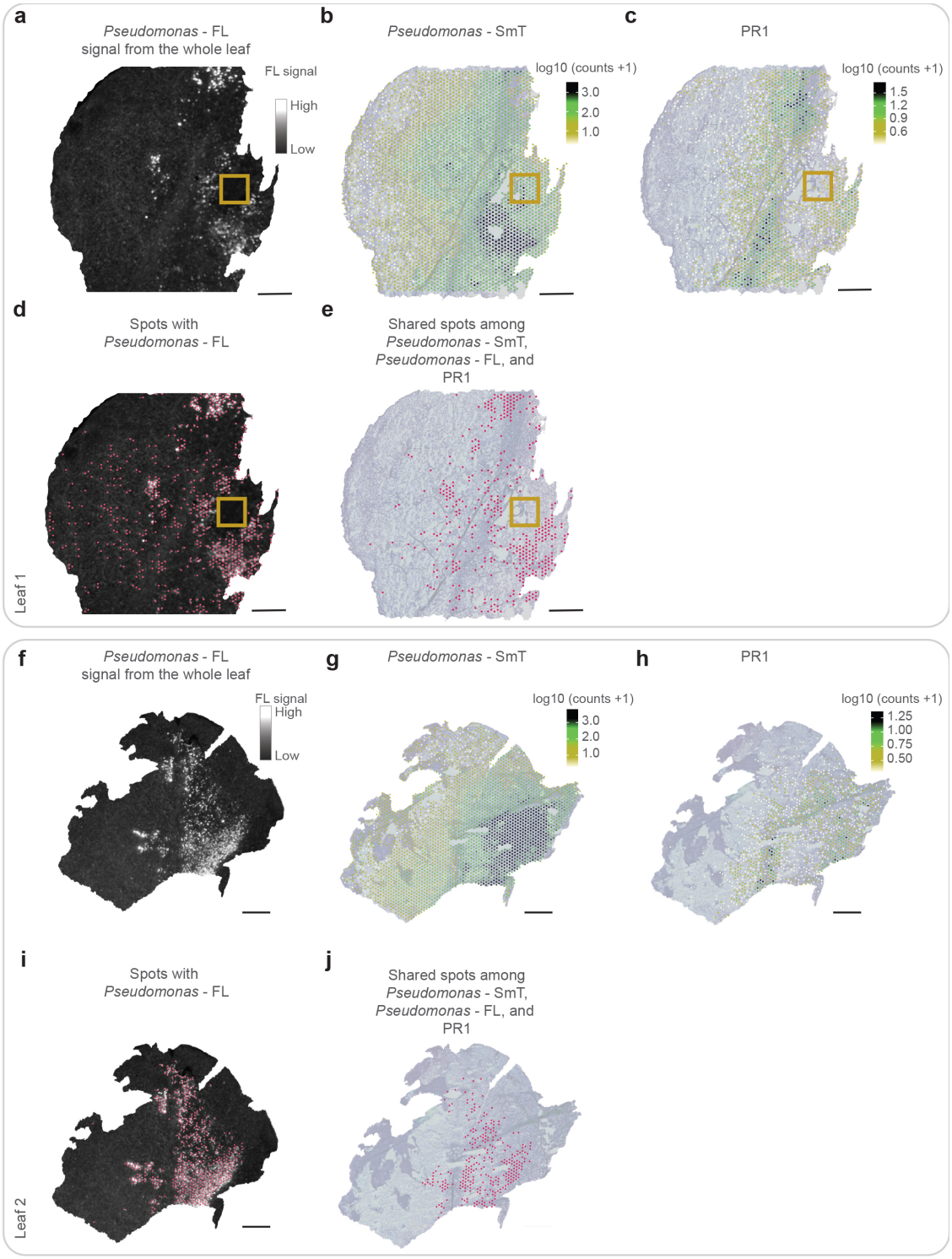
https://doi.org/10.1038/s41587-023-01979-2

Spatial metatranscriptomics resolves host–bacteria–fungi interactomes

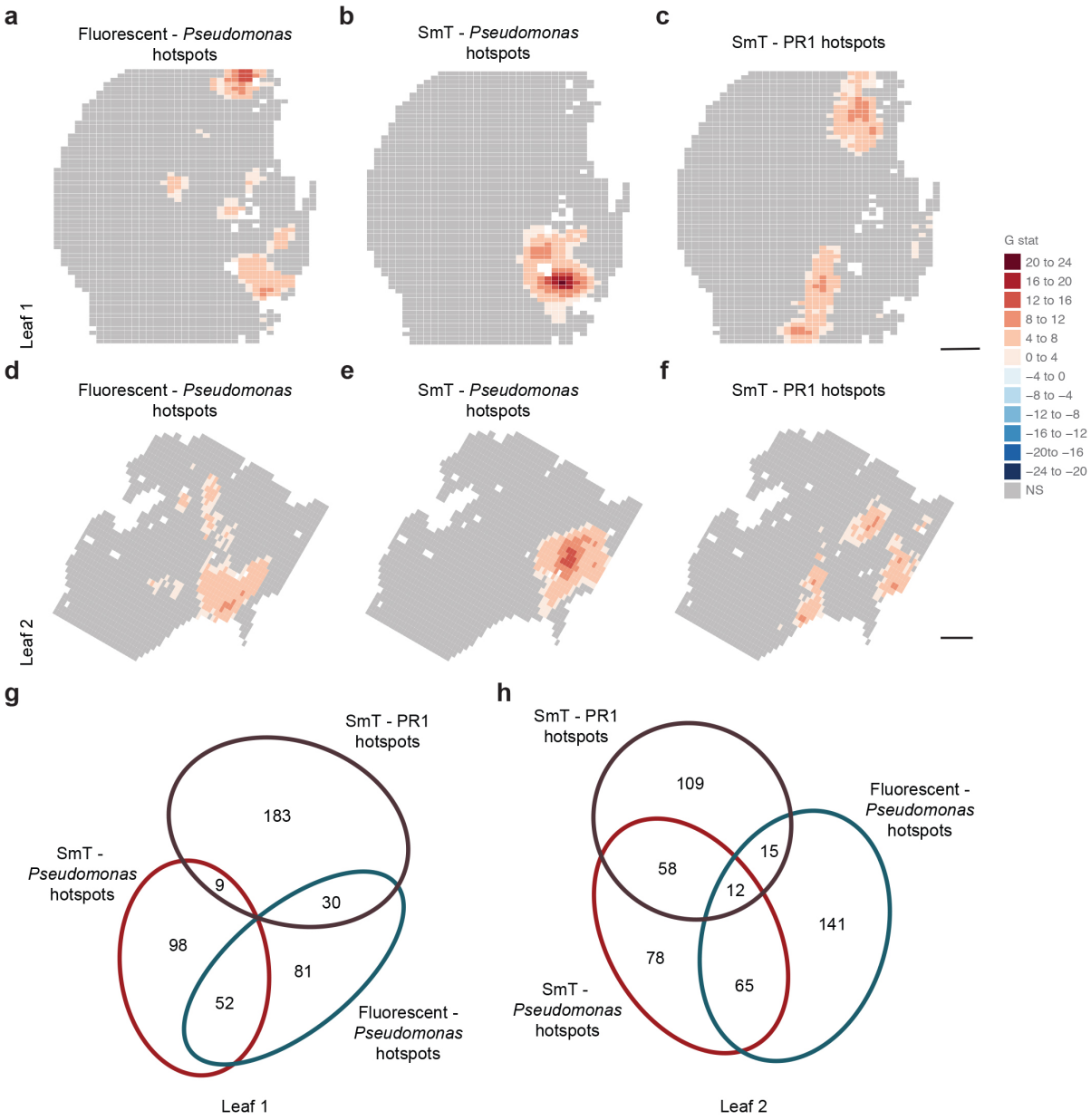
In the format provided by the authors and unedited



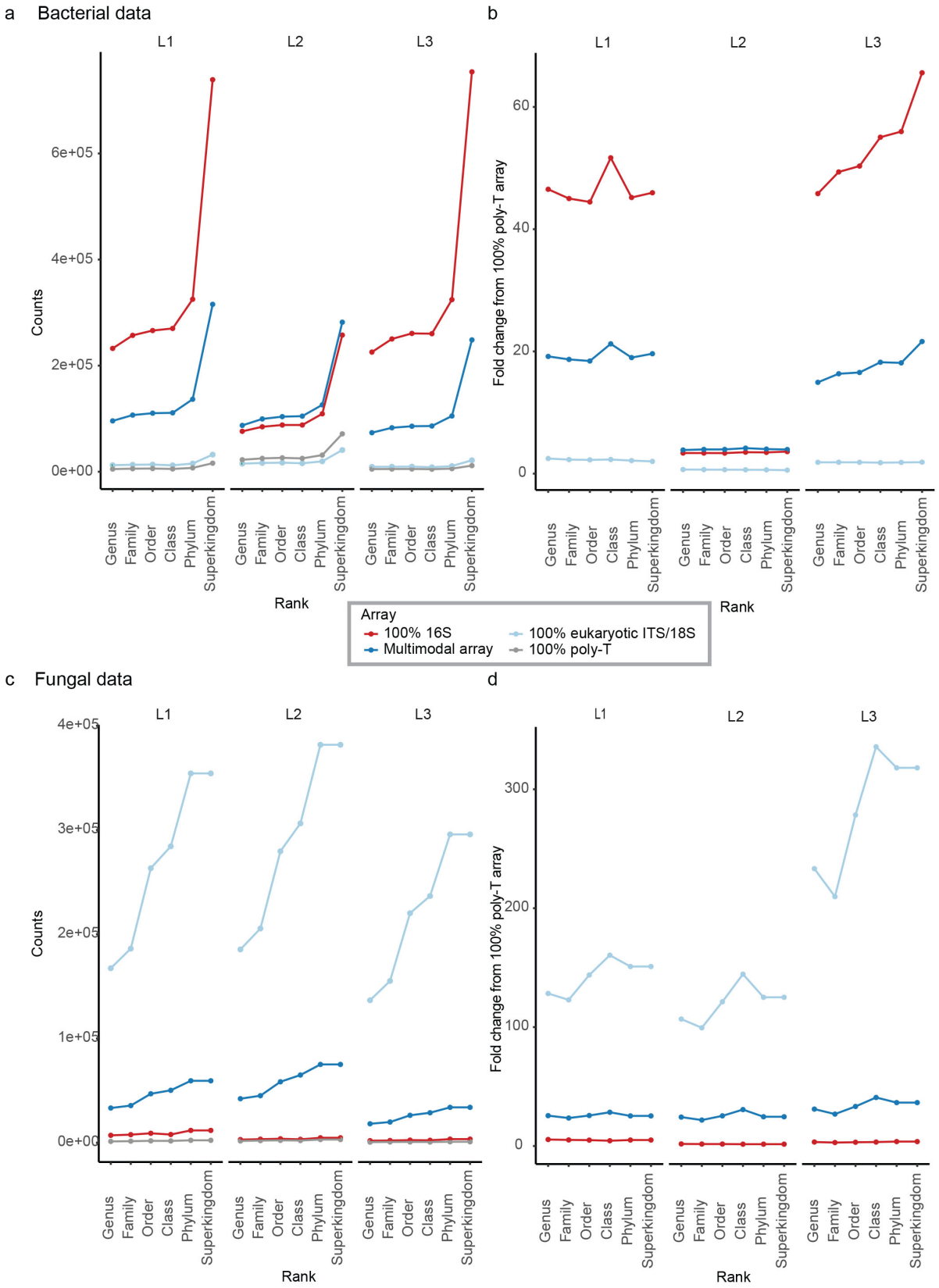
Supplementary Fig. 1: Combined capture profiles of host unique genes and microbial taxa (**a**), and of unique molecules of the host and the microbes (**b**) across the tissue section from *Pseudomonas* infected leaf 1. Scale bar is 1 mm. **c**, Overlap of the spots containing SmT-captured *Pseudomonas* and fluorescence imaging captured *Pseudomonas.* **d**, Scatter plot and the corresponding Pearson correlation (R=0.21, p=3.70e-38, two-tailed test) of log normalised *Pseudomonas* signal captured with the SmT array and the fluorescent imaging (maximum fluorescence intensity per spot area). The red line shows the fitted linear model with 0.95 confidence interval. **e**, Scatter plot and the corresponding Pearson correlation (R=0.52, p=3.58e-254, two-tailed test) of log normalised *Pseudomonas* and *PR1* host gene captured by the SmT array. The red line shows the fitted linear model with 0.95 confidence interval. **f**, Overlap of the spots containing SmT-captured *Pseudomonas*, fluorescence imaging captured *Pseudomonas* and host *PR1* gene. Data presented here is for leaf 1.



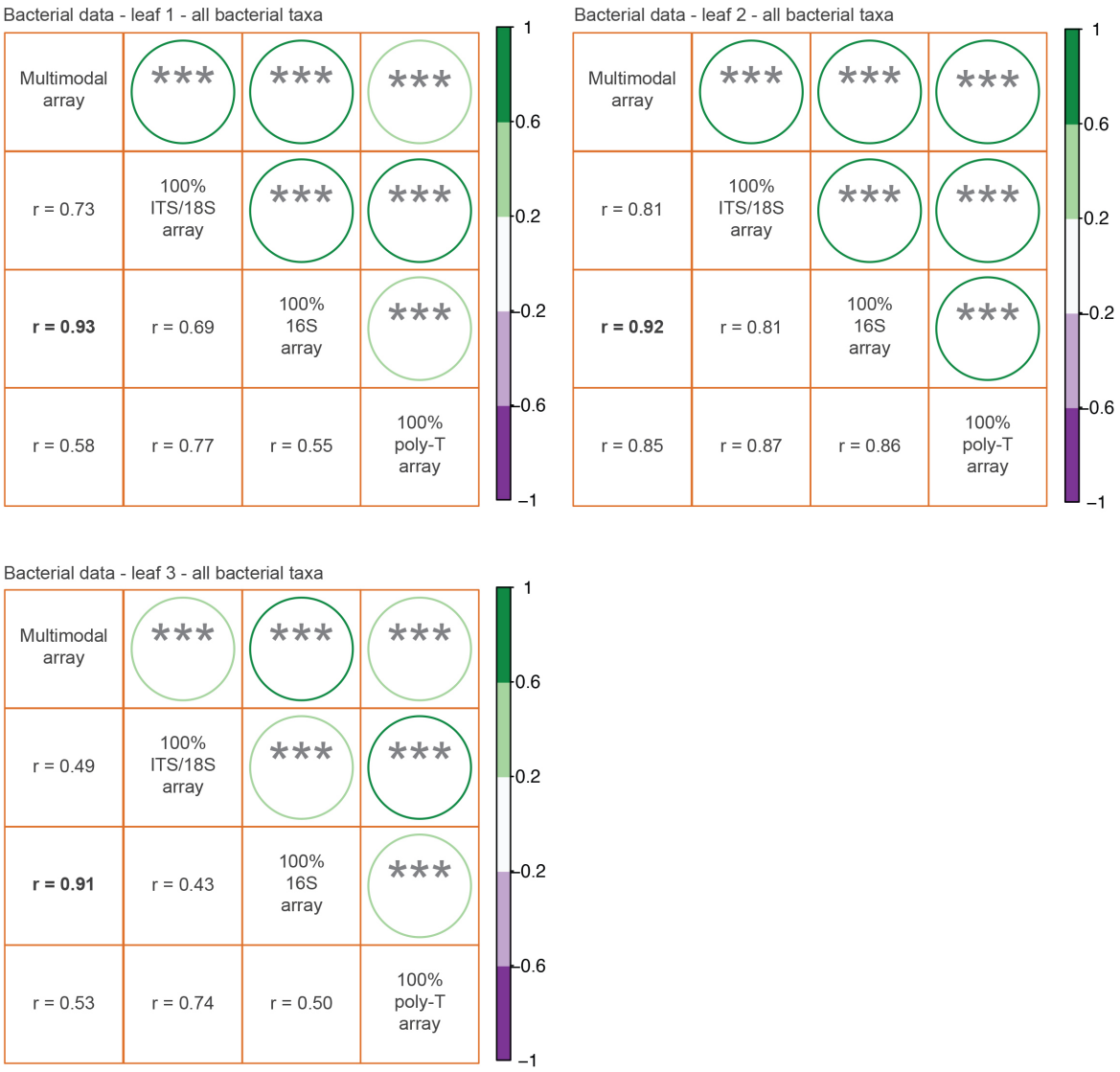
Supplementary Fig. 2: Analysis of two *Arabidopsis* leaves infected with fluorescently labelled *Pseudomonas* DC3000. We used fluorescent microscopy to visualize the signal from *mCherry*-labelled *Pseudomonas* DC3000 bacteria from whole intact leaves (a and f). Afterwards, we sectioned the leaves and profiled separate sections using SmT identifying a higher abundance of *Pseudomonas* DC3000 bacteria in correspondence of the higher fluorescent *Pseudomonas* signal (**b** and **g**) and expression of the host immune gene *PR1* (**c** and **h**). We identified all the spots containing the fluorescent signal (**d** and **i,** see **Methods**) and visualized the spots containing SmT and fluorescence microscopy-derived *Pseudomonas* DC3000 signal and the host *PR1* immune gene (**e** and **j**). Scale bar is 1 mm.



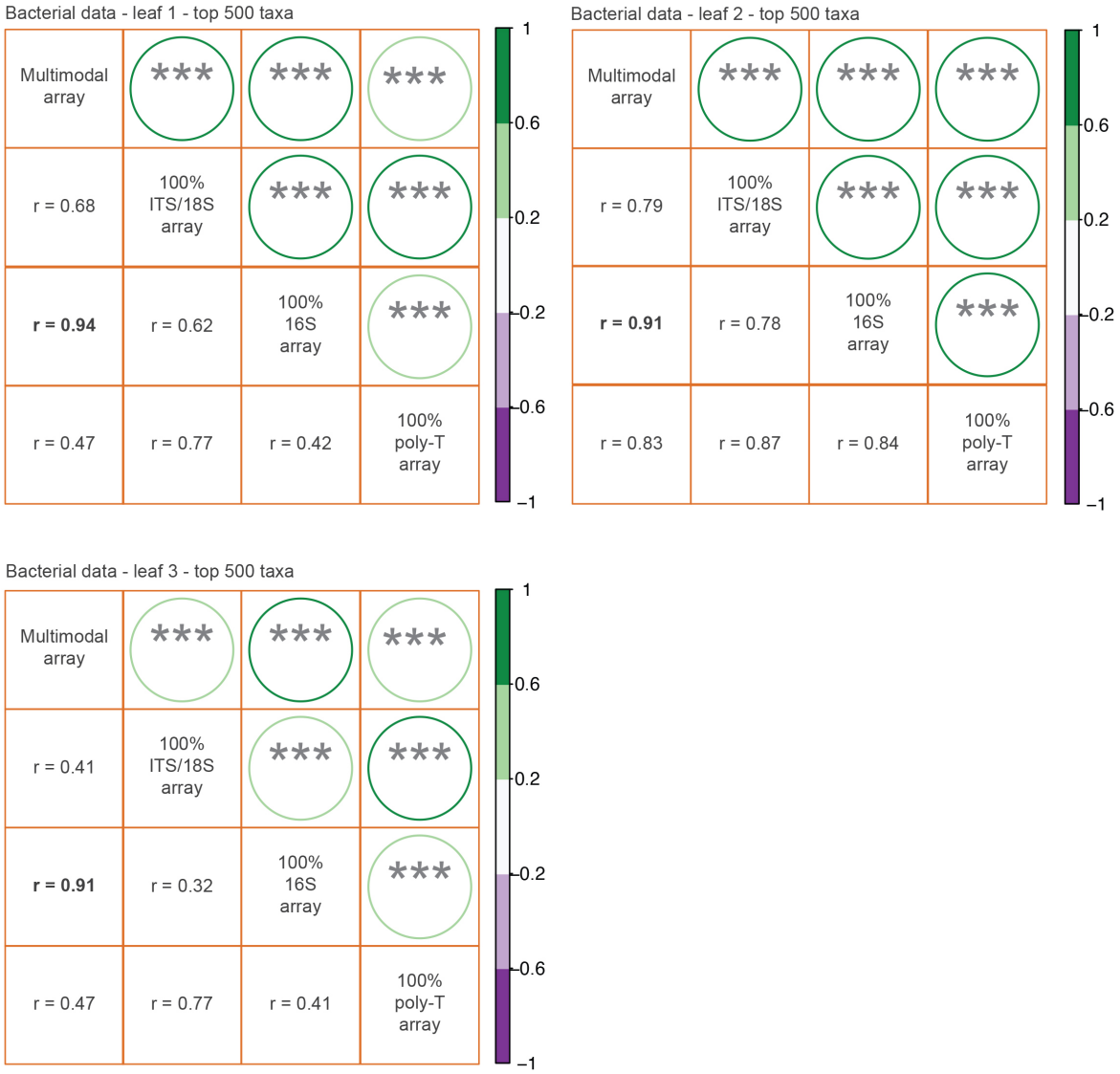
Supplementary Fig. 3: Significant hot- and cold-spots of *Pseudomonas* DC3000 as derived from the fluorescent signal (**a** and **d**) and SmT signal (**b** and **g**), and Significant (BH-FDR corrected p-value$\lesssim$ 0.05) hot- and cold-spots of the expression of the gene *PR1* (**c** and **f**). Results are presented for two separate leaves (**a-c** and **d-f**). Overlap between significant hotspots is presented in panels **g** and **h**. For the fluorescent signals, only values above 45 and 120 were considered for leaf 1 and leaf 2, respectively. Scale bar is 1 mm.



Supplementary Fig. 4: a, Number of unique molecules presented at each taxonomic level for bacterial data in each array and **b**, the fold change between different taxonomic levels from unspecific binding where an array of 100% poly-d(T) probes has been used as a baseline. Fungal unique molecules and the fold change as a function of taxonomic level are presented in **c** and **d**, respectively. L1, L2 and L3 stands for Leaf 1, Leaf 2 and Leaf 3, respectively.

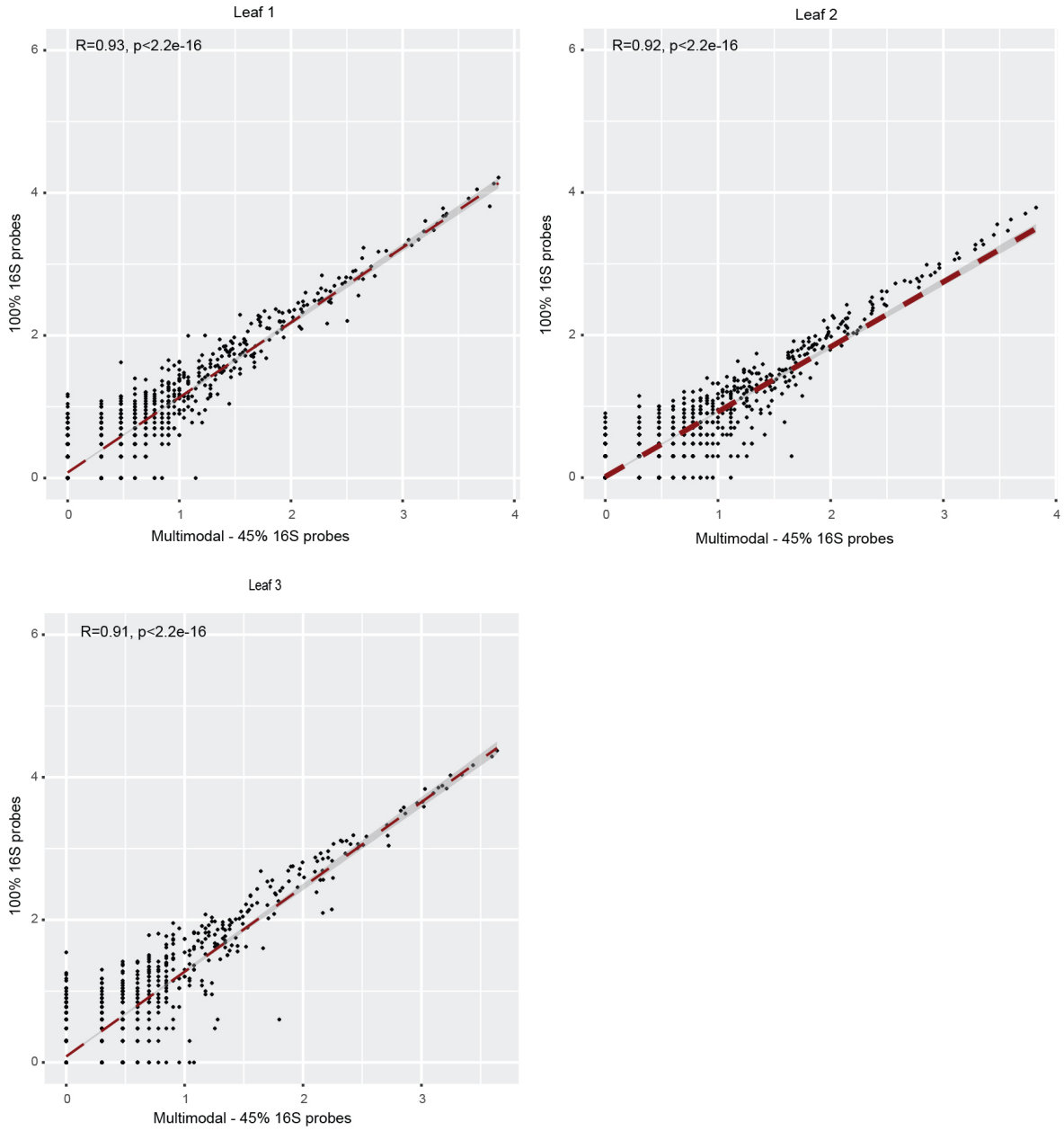


Supplementary Fig. 5: Pairwise Pearson correlation coefficients and the corresponding two-tailed significance test of bacterial 16S components between different array types in three leaves when the full bacterial profile with 1681 taxa is considered. *** indicates the p-value=0. Correlation between the multimodal array and the 100% 16S array is in bold.

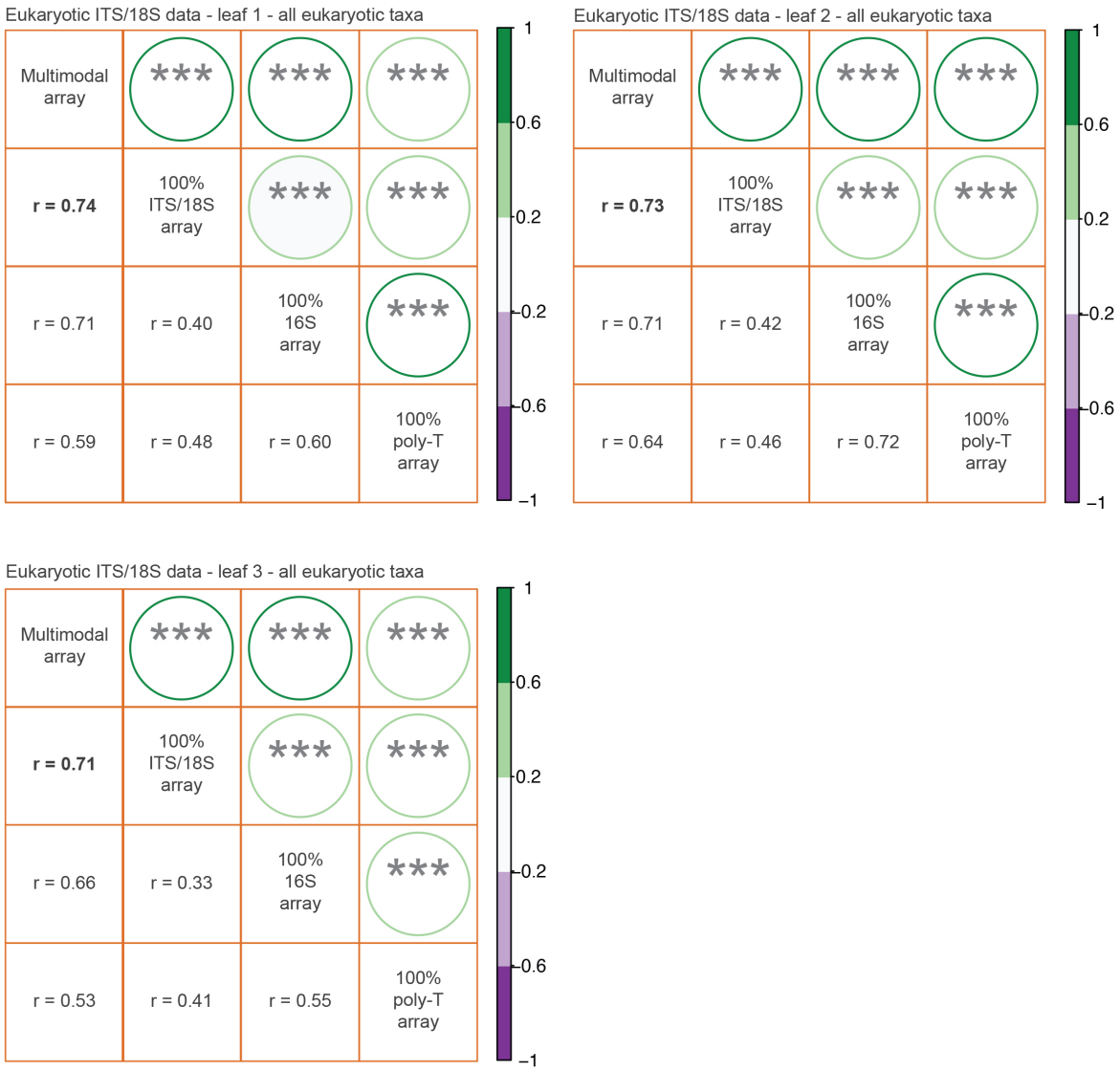


Supplementary Fig. 6: Pairwise Pearson correlation coefficients and the corresponding two-tailed significance test of bacterial 16S components between different array types in three leaves when the 500 most abundant bacterial taxa are considered. *** indicates the p-value=0. Correlation between the multimodal array and the 100% 16S array is in bold.

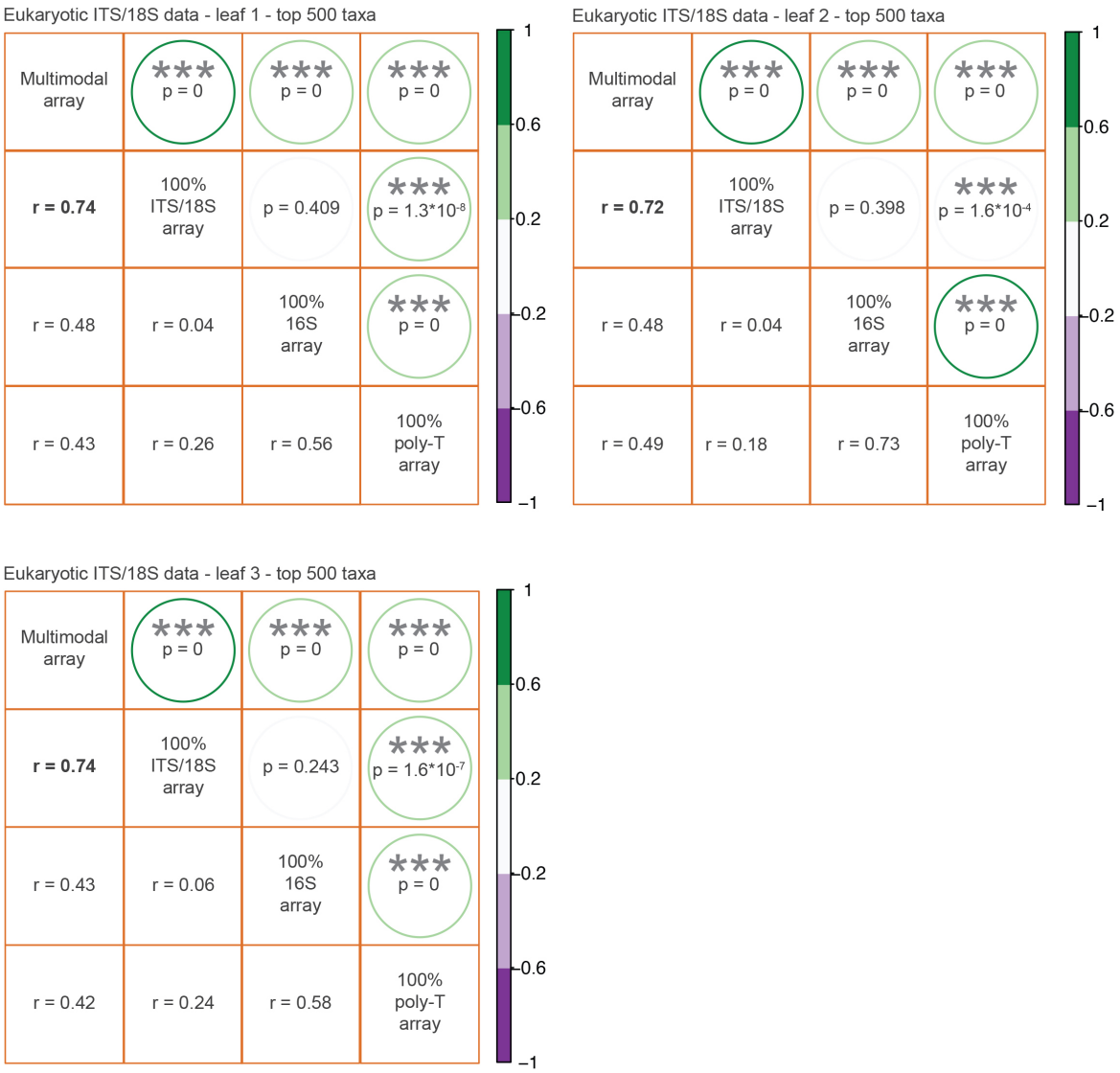
Supplementary Fig. 7: Pairwise Pearson correlation coefficients and the corresponding two-tailed significance test of bacterial 16S components between different array types in three leaves when the 20 most abundant bacterial taxa are considered. *: P<0.05, **: P<0.01, ***: P<0.001. Correlation between the multimodal array and the 100% 16S array is in bold.



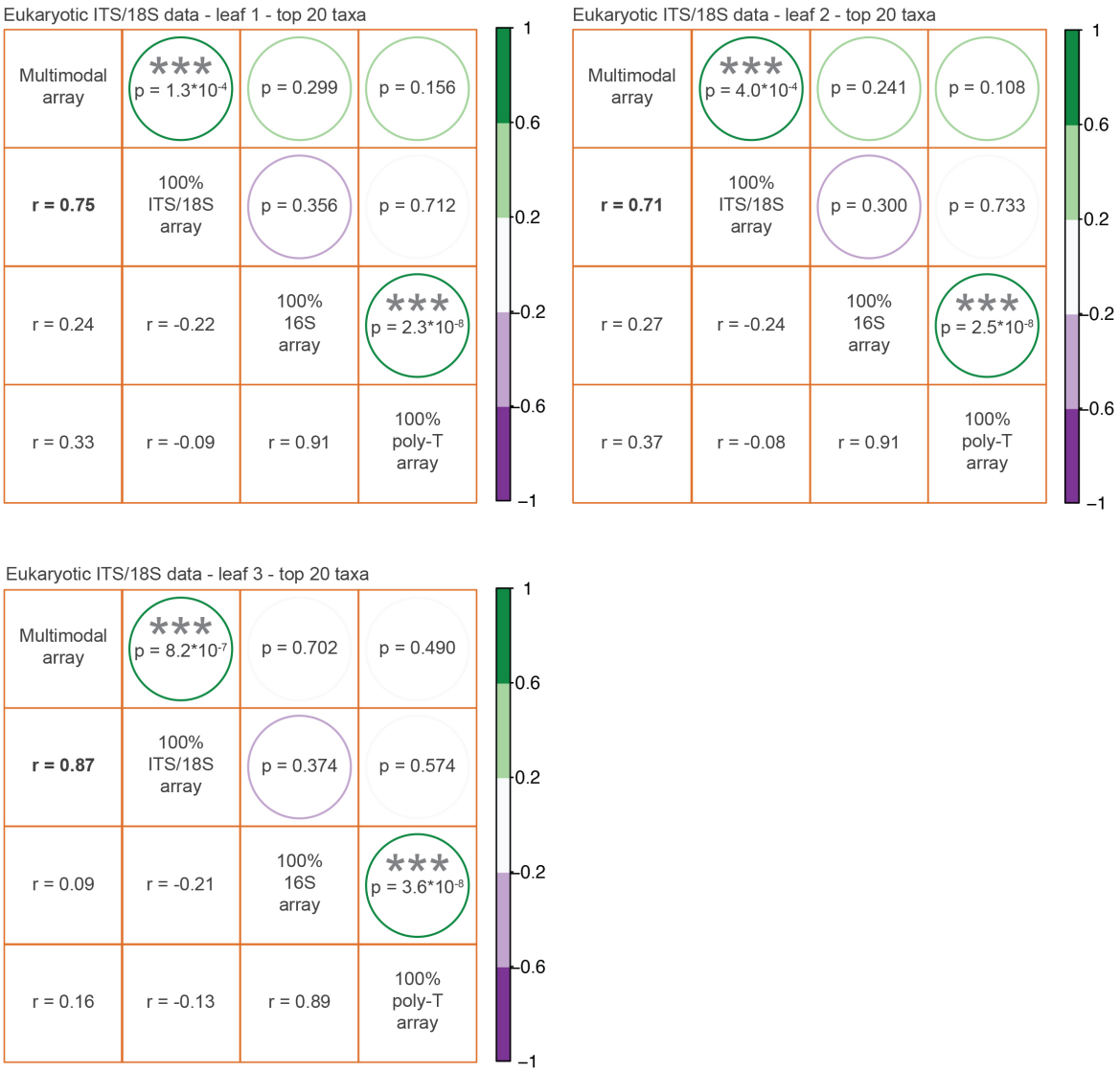
Supplementary Fig. 8. Scatter plots of log normalised unique bacterial molecules and the corresponding Pearson correlation coefficient and the corresponding two-tailed significance test for arrays with 100% and 45% bacterial 16S rRNA probes in leaves 1 to 3. For leaves 1, 2, and 3 the Pearson correlation values are 0.93, 0.92, and 0.91 ($p<2.2e-16$), respectively. The red line shows the fitted linear model with 0.95 confidence interval.



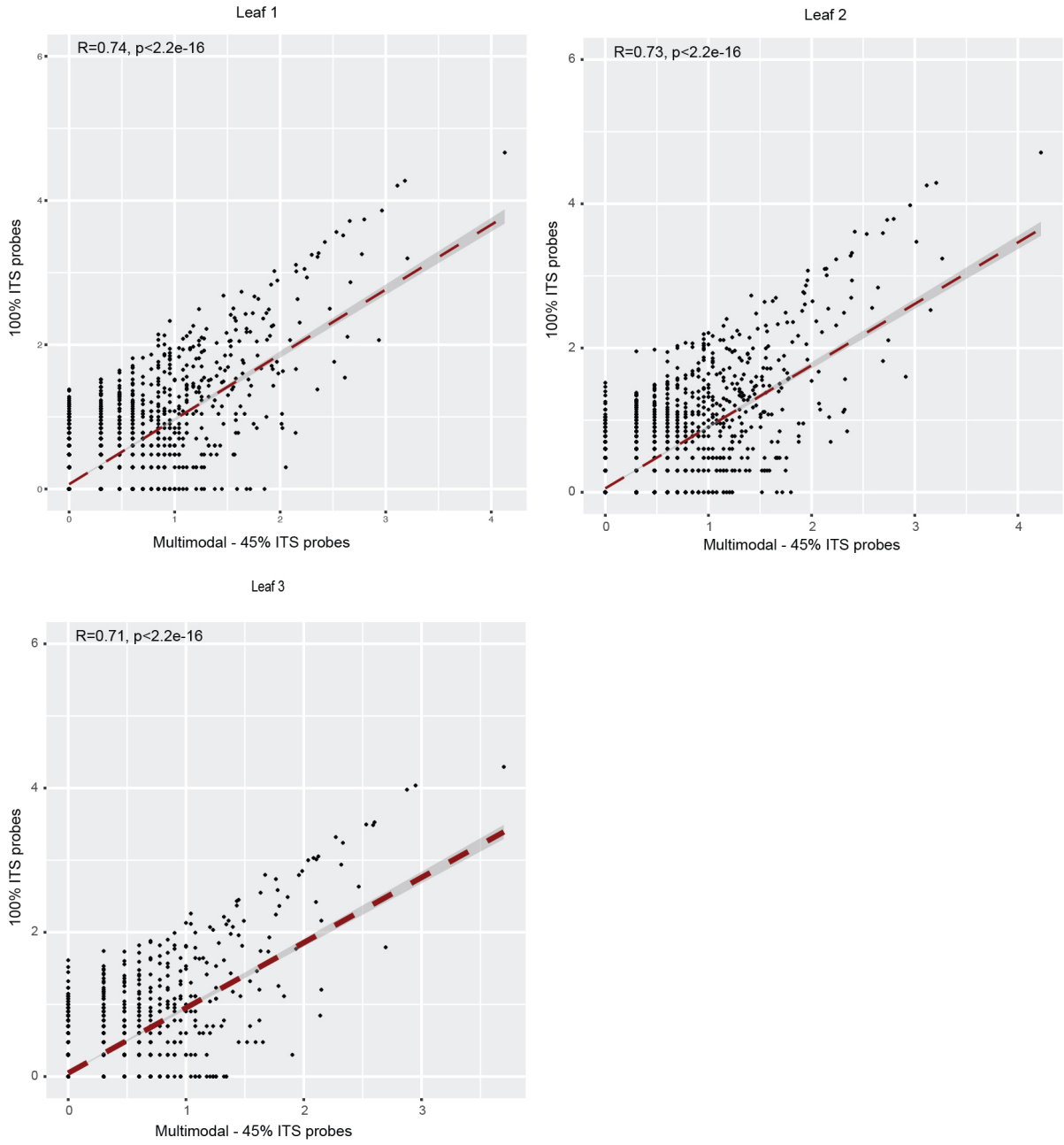
Supplementary Fig. 9: Pairwise Pearson correlation coefficients and the corresponding two-tailed significance test of eukaryotic ITS/18S components between different array types in three leaves when the full eukaryotic profile with 1660 taxa is considered. *** indicates the p-value=0. Correlation between the multimodal array and the 100 % Eukaryotic ITS/18S array is in bold.



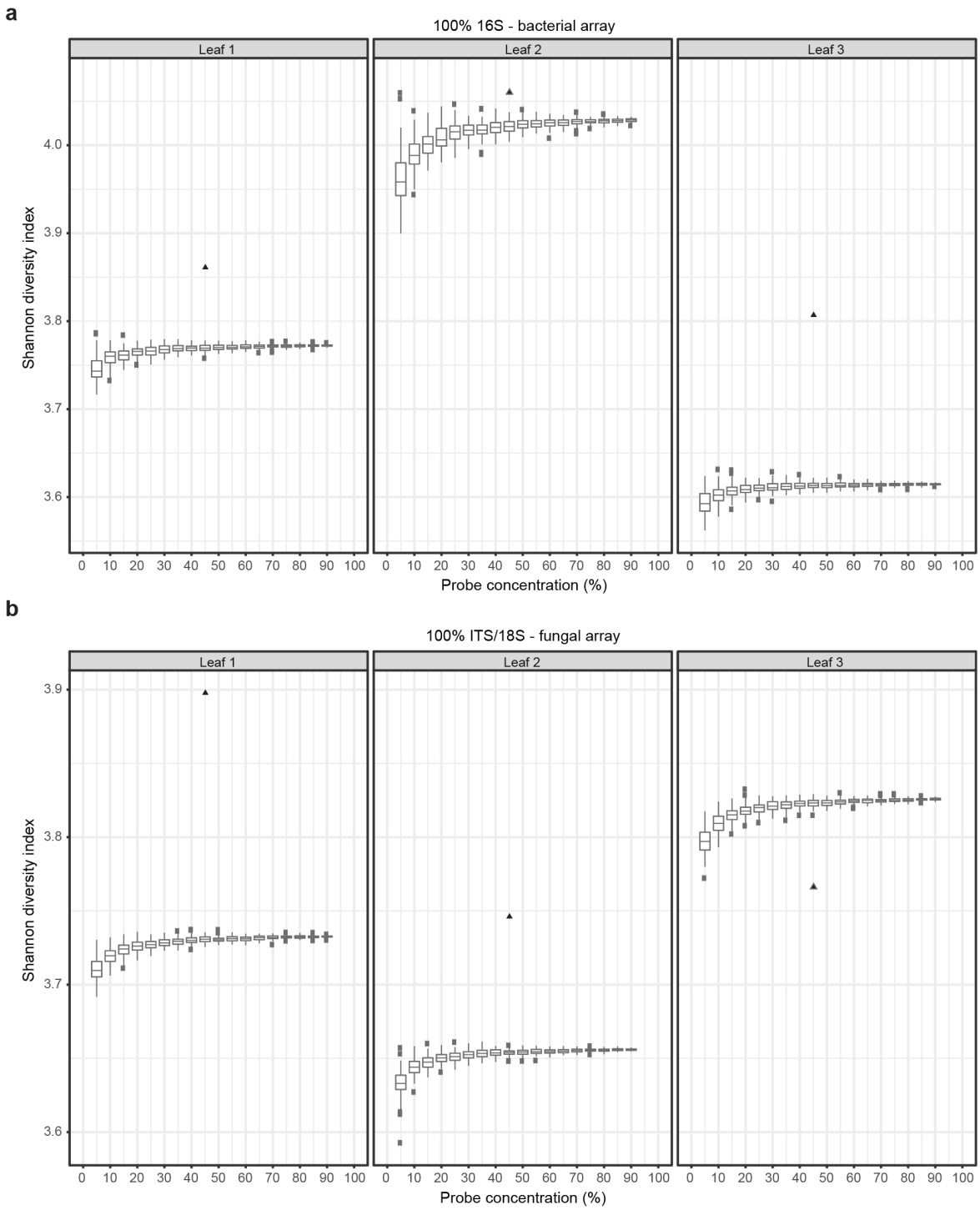
Supplementary Fig. 10: Pairwise Pearson correlation coefficients and the corresponding two-tailed significance test of eukaryotic ITS/18S components between different array types in three leaves when the 500 most abundant eukaryotic taxa are considered. *: P<0.05, **: P<0.01, ***: P<0.001. Correlation between the multimodal array and the 100% Eukaryotic ITS/18S array is in bold.



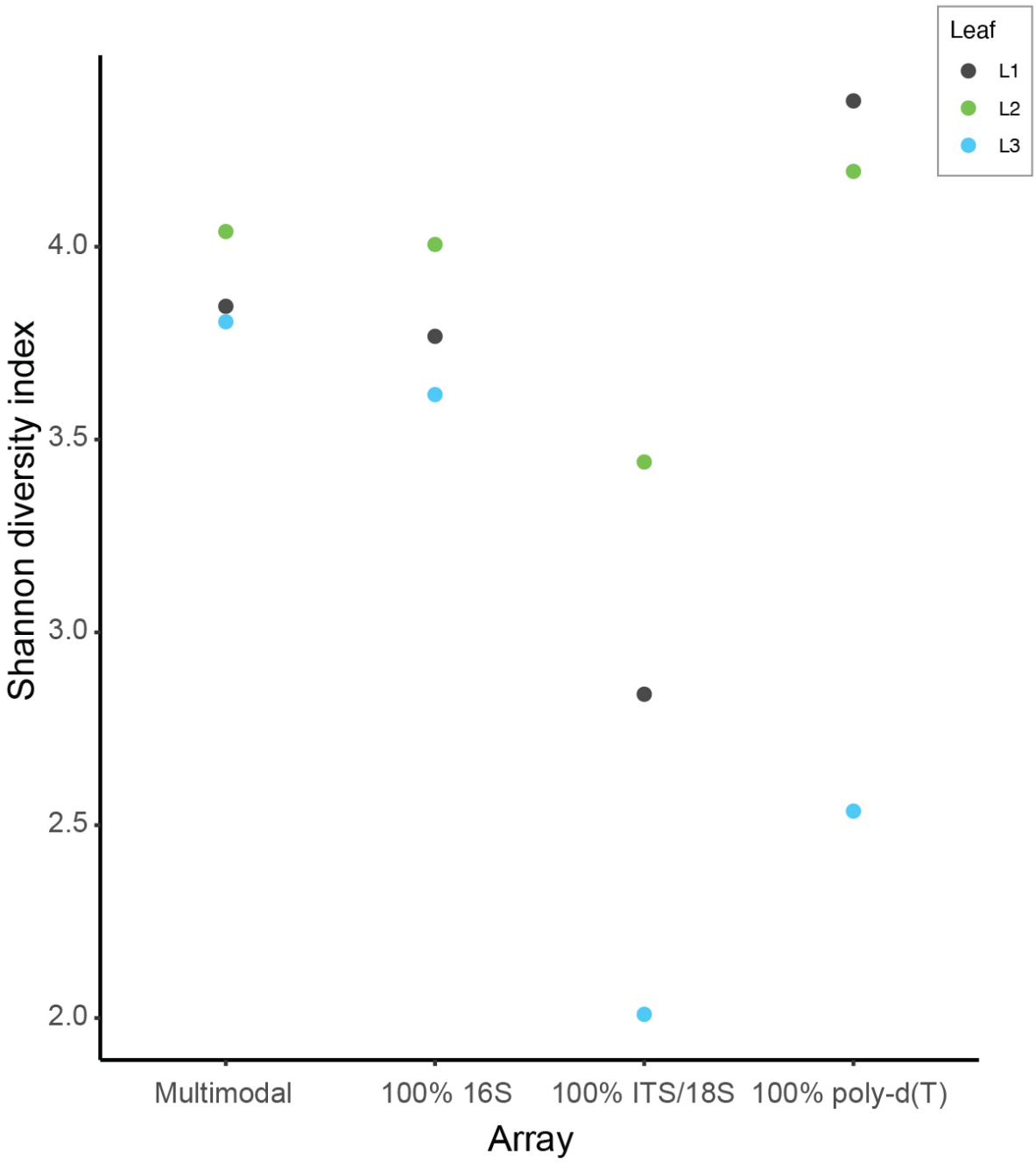
Supplementary Fig. 11: Pairwise Pearson correlation coefficients and the corresponding two-tailed significance test of eukaryotic ITS/18S component between different array types in three leaves when the 20 most abundant eukaryotic taxa are considered. *: P<0.05, **: P<0.01, ***: P<0.001. Correlation between the multimodal array and the 100% Eukaryotic ITS/18S array is in bold.



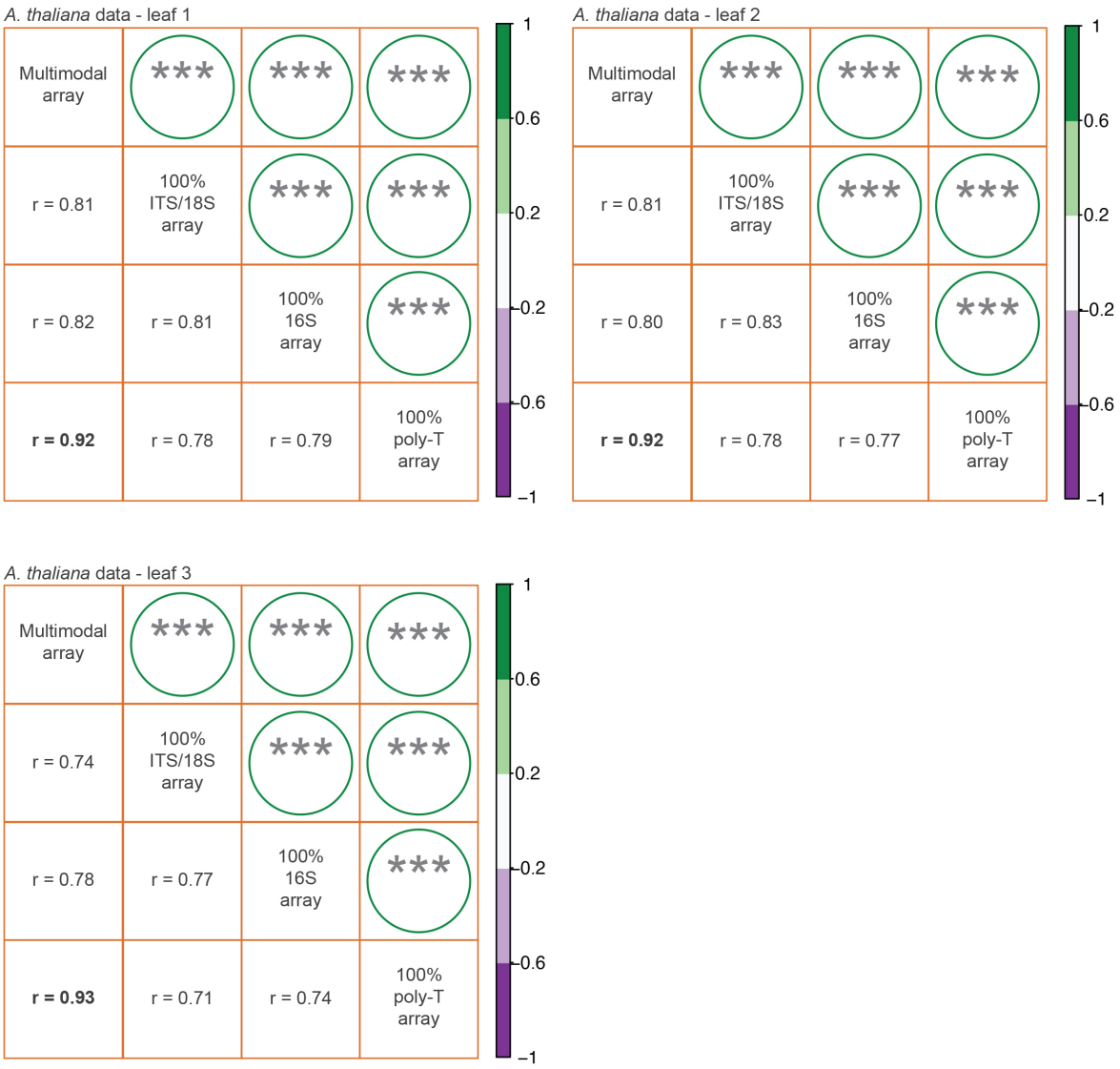
Supplementary Fig. 12: Scatter plots of log normalised unique fungal molecules and the corresponding Pearson correlation and the corresponding two-tailed significance test for arrays with 100% and 45% 18S rRNA/ITS probes in leaves 1 to 3. For leaves 1, 2, and 3 the Pearson correlation values are 0.74, 0.73, and 0.71 ($p<2.2e-16$), respectively. The red line shows the fitted linear model with 0.95 confidence interval.



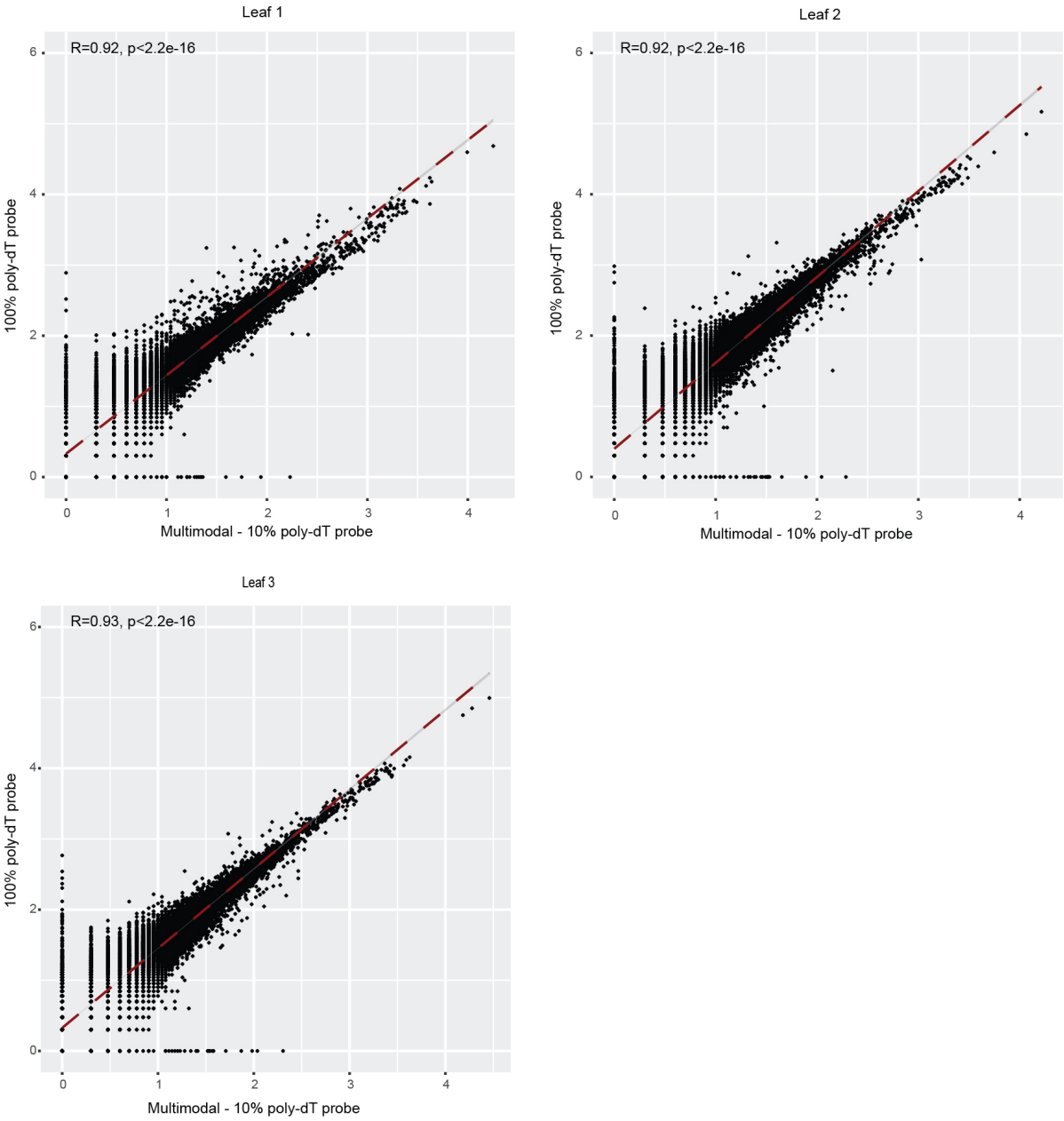
Supplementary Fig. 13: Shannon diversity index as a function of the simulated probe concentration performed by downsampling for 16S-capturing (**a**) and ITS/18S-capturing (**b**) probes. Boxplot for each probe concentration illustrates the median, first and third quartiles of the calculated Shannon diversity index, with whiskers extending 1.5 times the interquartile range. Each boxplot represent 100 repeats of the downsampling simulation (see **Methods** for more details).



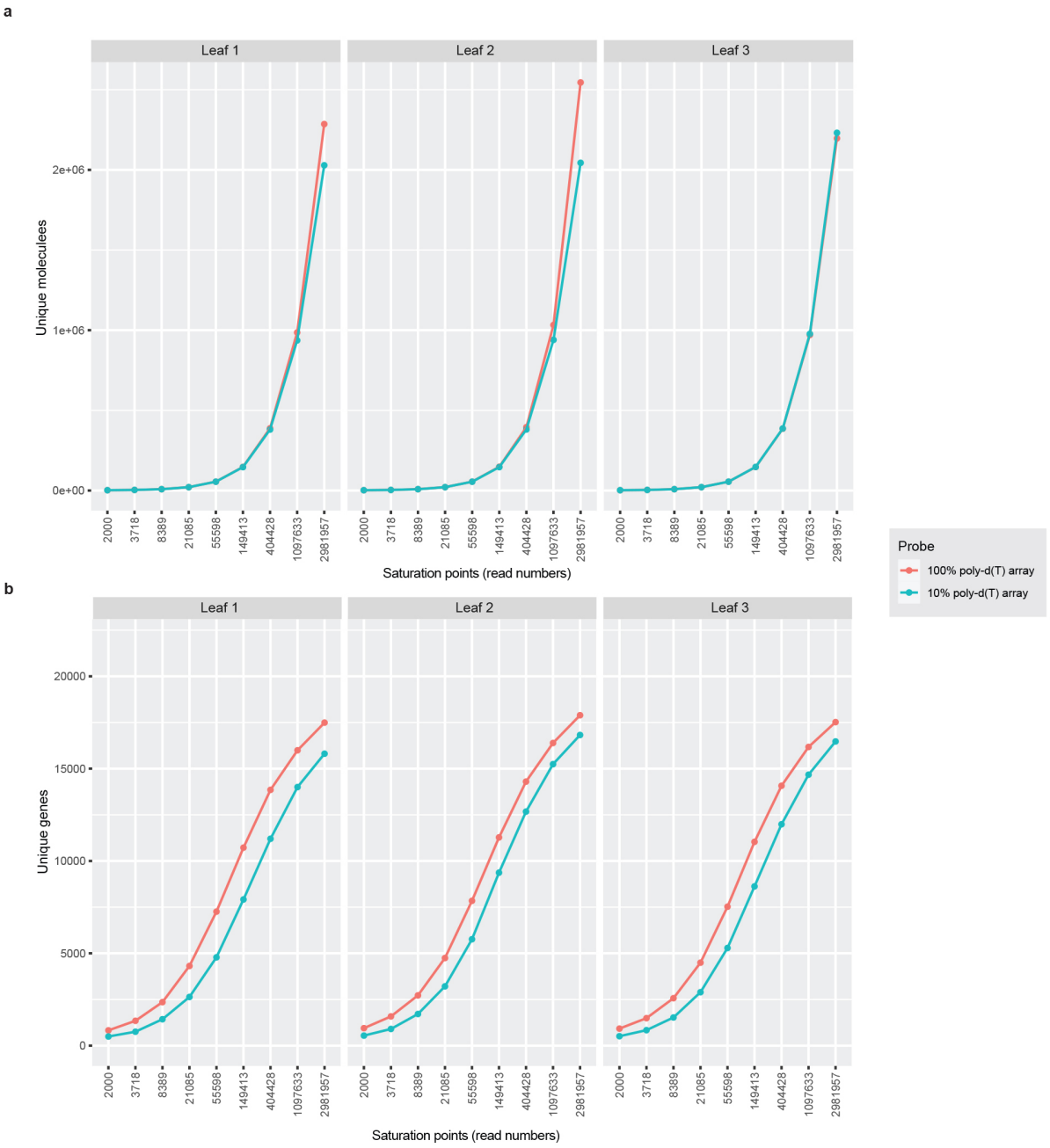
Supplementary Fig. 14: Shannon diversity index between different leaves in different arrays with bacterial assignment. L1, L2 and L3 stands for Leaf 1, Leaf 2 and Leaf 3, respectively.



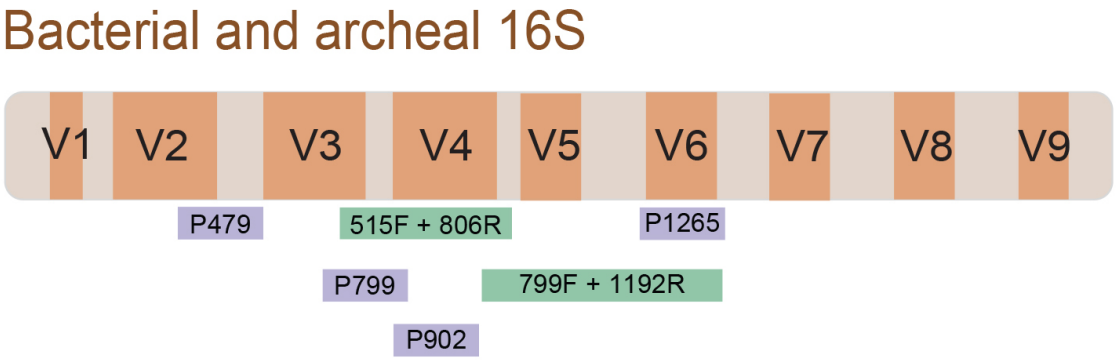
Supplementary Fig. 15: Pairwise Pearson correlation coefficients and the corresponding two-tailed significance test of the ribosomal, chloroplast, mitochondrial and non-coding filtered host *A. thaliana* component between different array types in three leaves. *** indicates the p-value=0. Correlation between the multimodal array and the 100% poly-d(T) array is in bold.



Supplementary Fig. 16. Scatter plots of log normalised unique host molecules and the corresponding Pearson correlation and the corresponding two-tailed significance test for arrays with 100% poly-(d)T and 10% poly-(d)T in leaves 1 to 3. For leaves 1, 2, and 3 the Pearson correlation values are 0.92, 0.92, and 0.93 (p<2.2e-16), respectively. The red line shows the fitted linear model with 0.95 confidence interval.

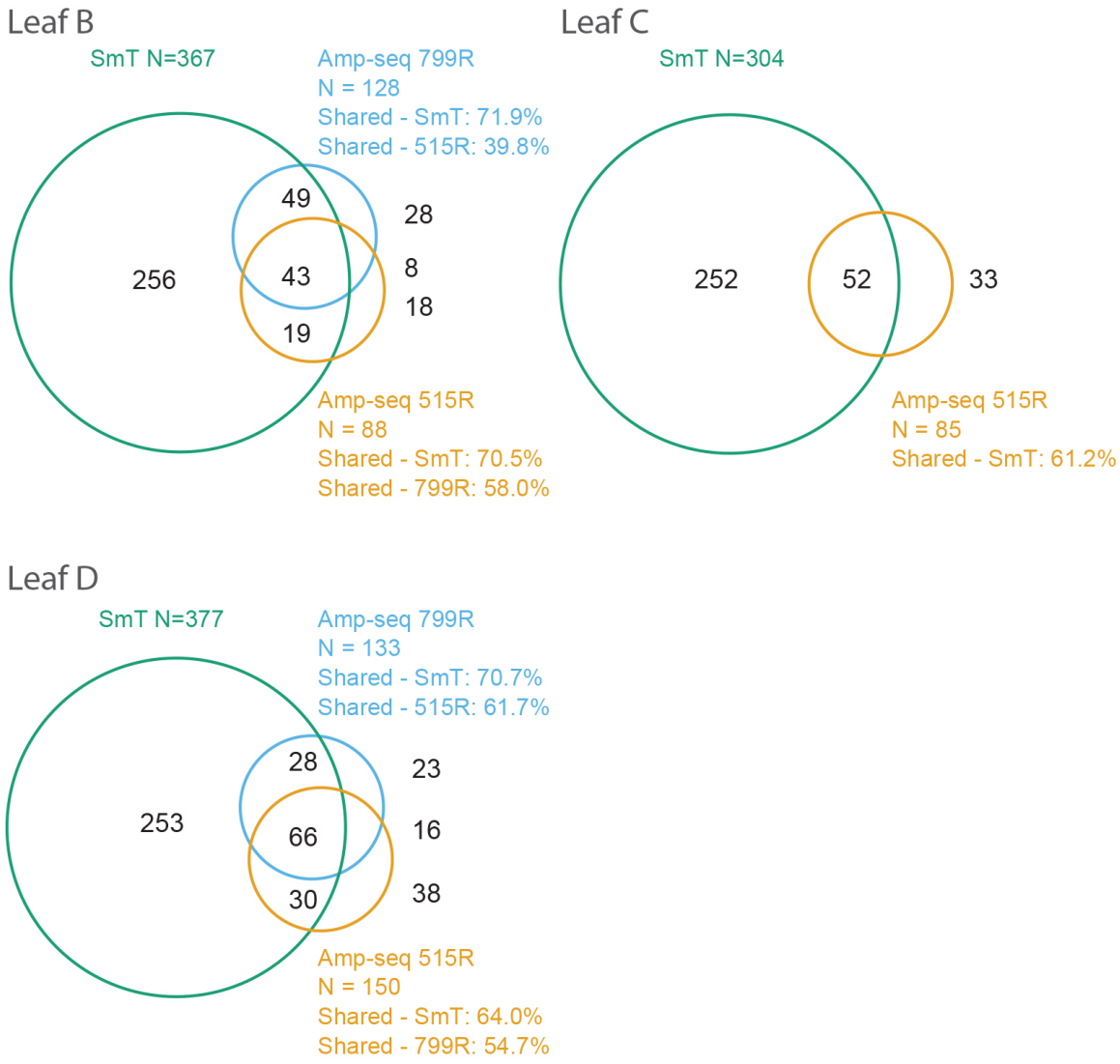


Supplementary Fig. 17: Saturation plots of the unique molecules (**a**) and genes (**b**) in 100% and 10% poly-d(T) (multimodal) array for three replicates.

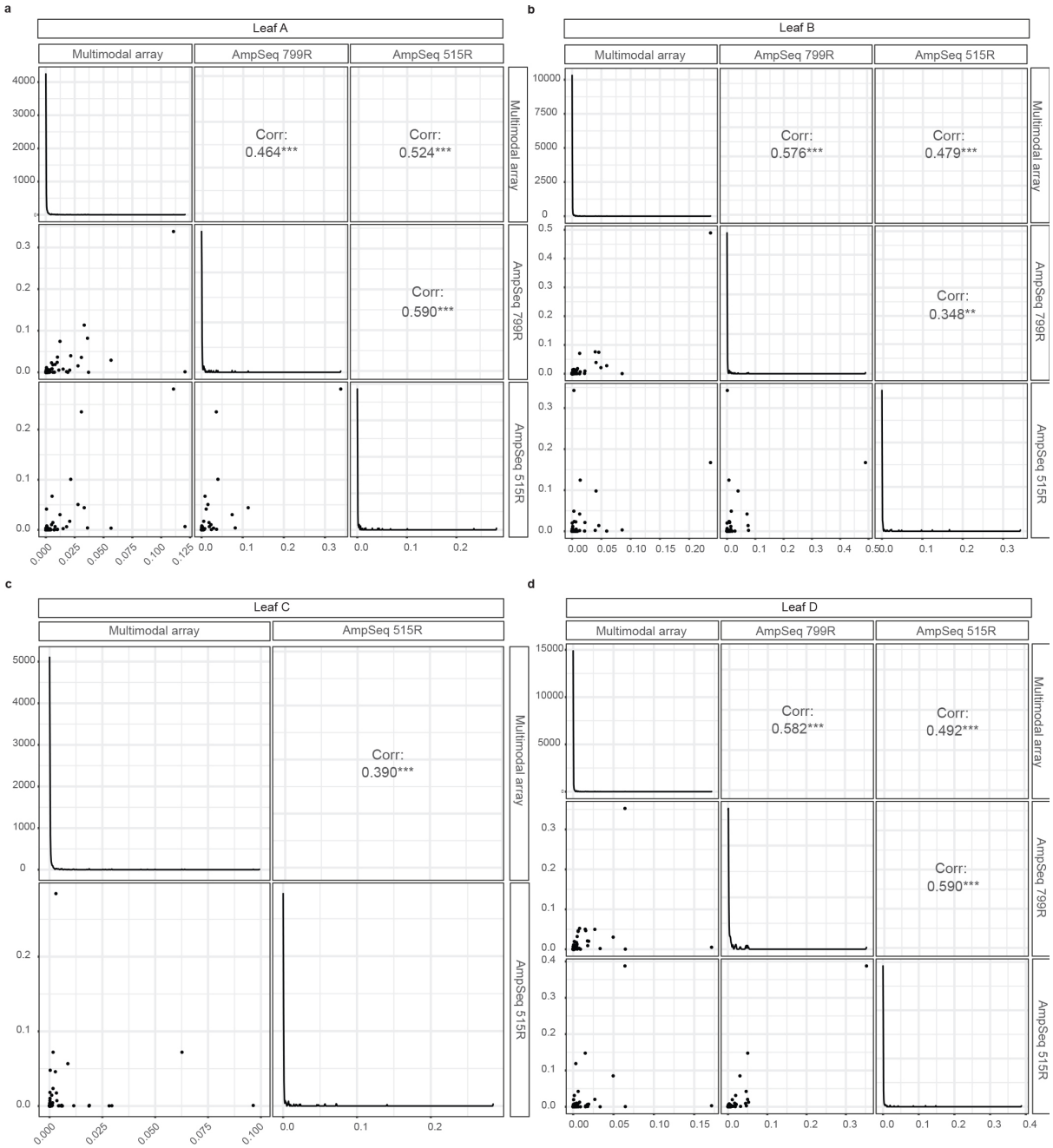


Supplementary Fig. 18: Bacterial and archeal target regions of the SmT probes (purple) and amp-seq primers (green).

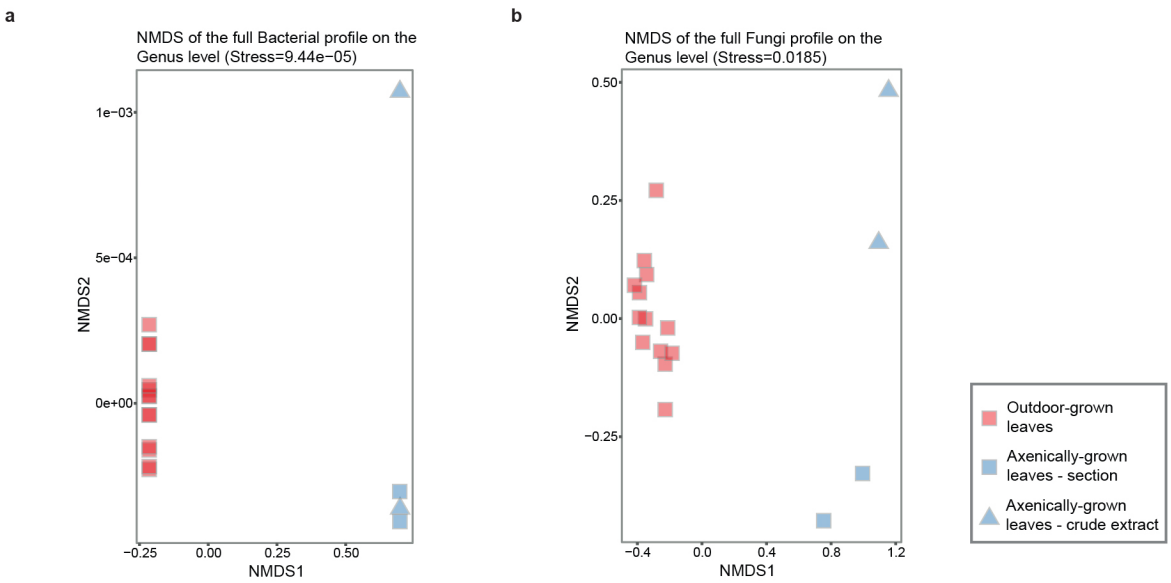
Supplementary Fig. 19: Bacterial relative abundance profiles of four leaf crude extracts (**a-d**), as captured by each of the four different probes (the initial P799 and P902 probes and two additional ones - P479 and P1265) and the unknown probes (reads that their probe of origin could not be detected), and the final profile that derived from all probes.



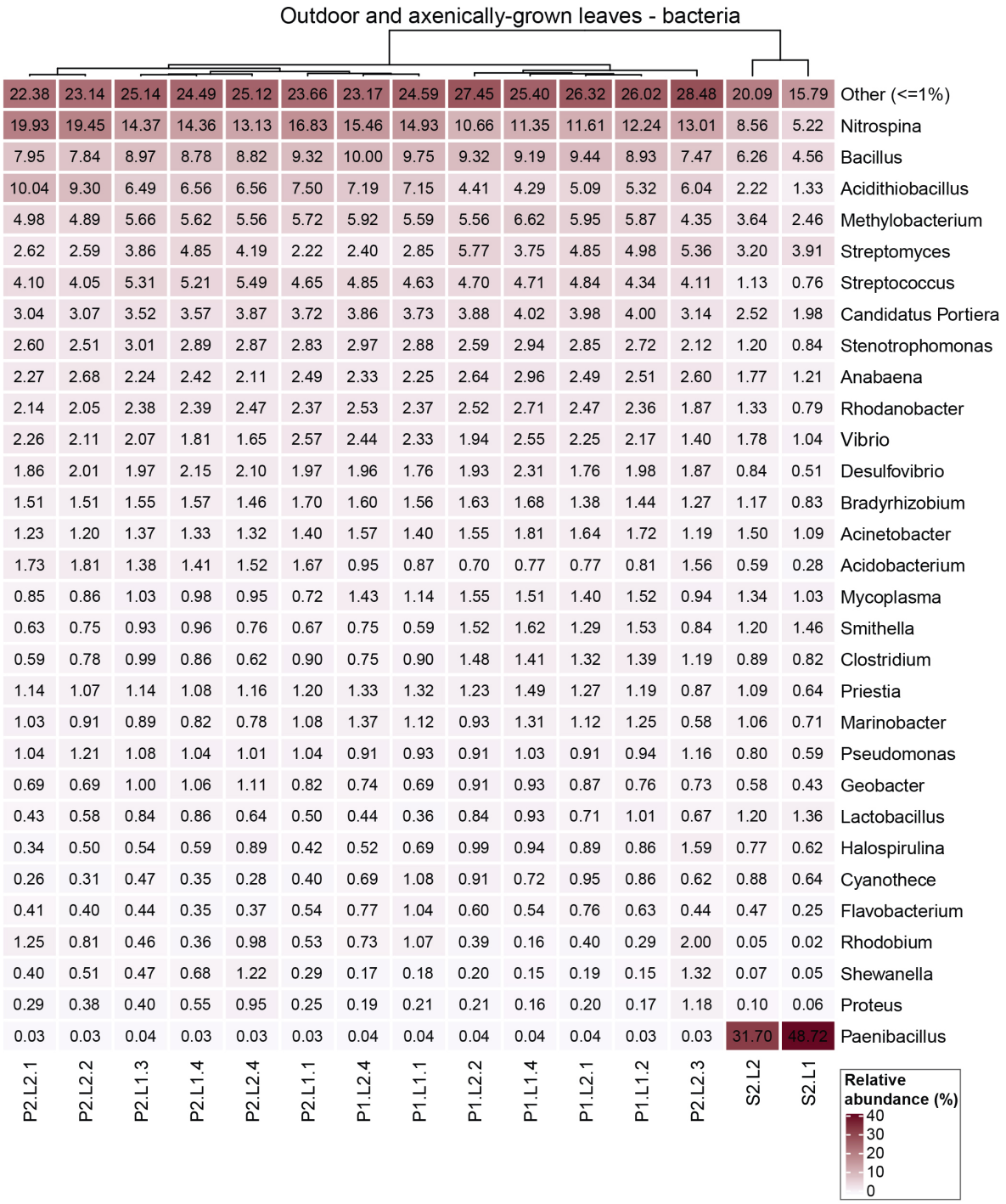
Supplementary Fig. 20: Taxa identified in different biological replicates presented in Venn diagrams presenting average values of subset samples after 100 iterations.



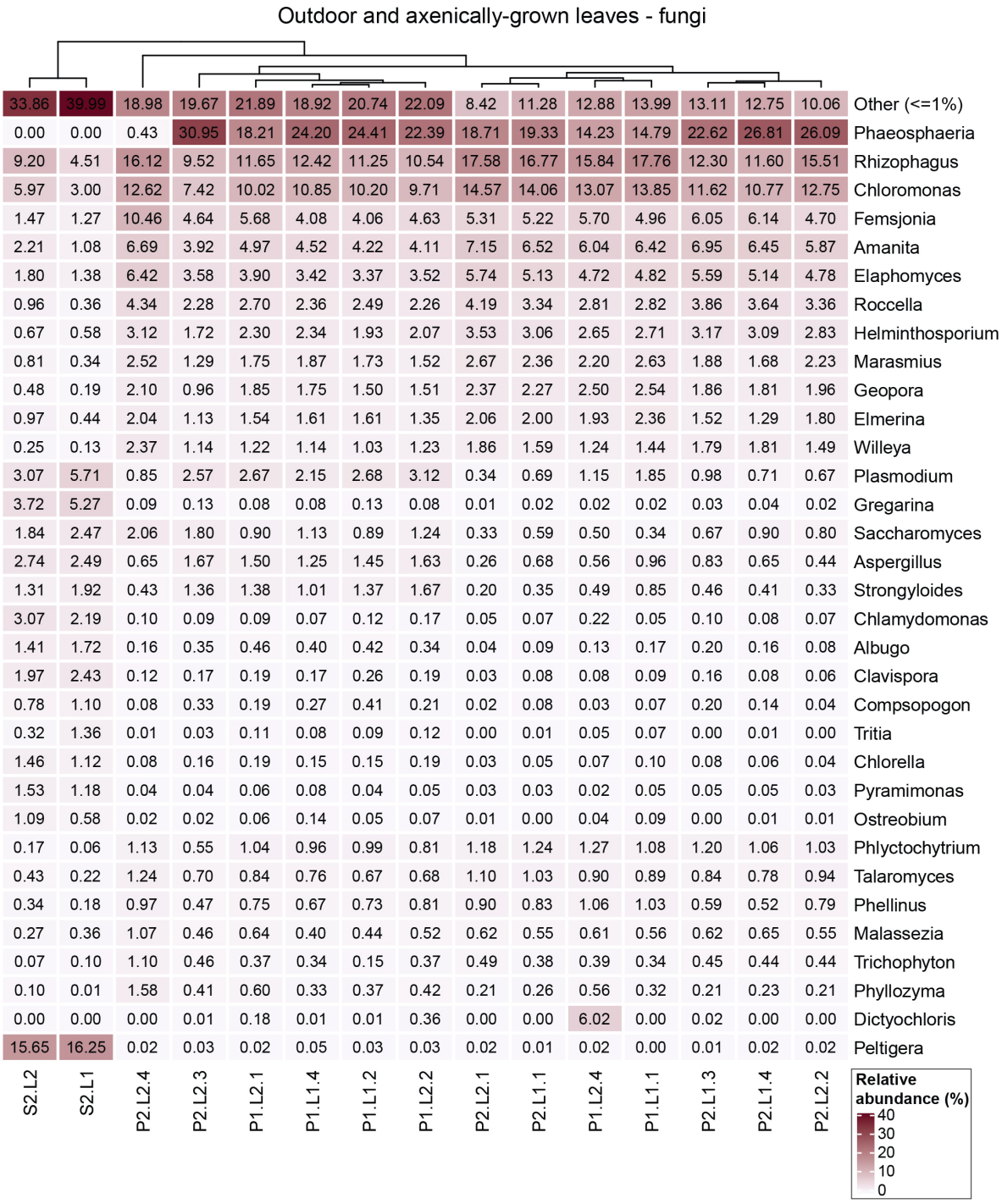
Supplementary Fig. 21: Spearman correlation coefficients of the relative abundance of shared bacterial genera, comparing all pairwise combinations between the multimodal array, amp-seq 799R and amp-seq 515R primer pairs, across four leaf crude extracts (**a-d**). In all cases p-value<0.001 (two-tailed test) denoted by '***'.



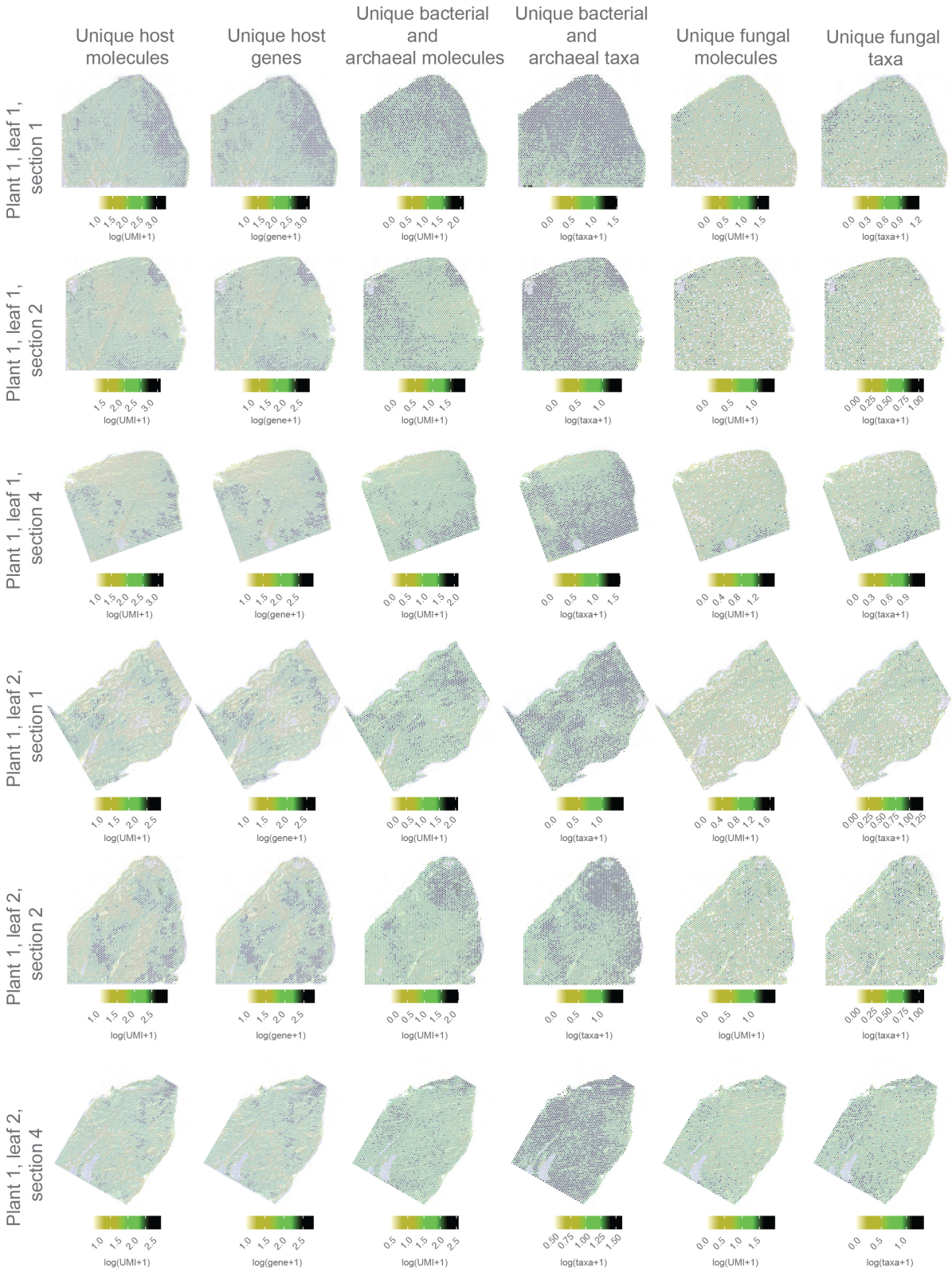
Supplementary Fig. 22: Analyses of axenically-grown leaves as a negative control. Non-metric multidimensional scaling (NMDS) of Bray-Curtis similarity for (a) bacterial and (b) fungal taxa between sections of outdoor-grown leaves and axenically-grown leaves (both sections and crude extracts, as indicated by the different shapes).

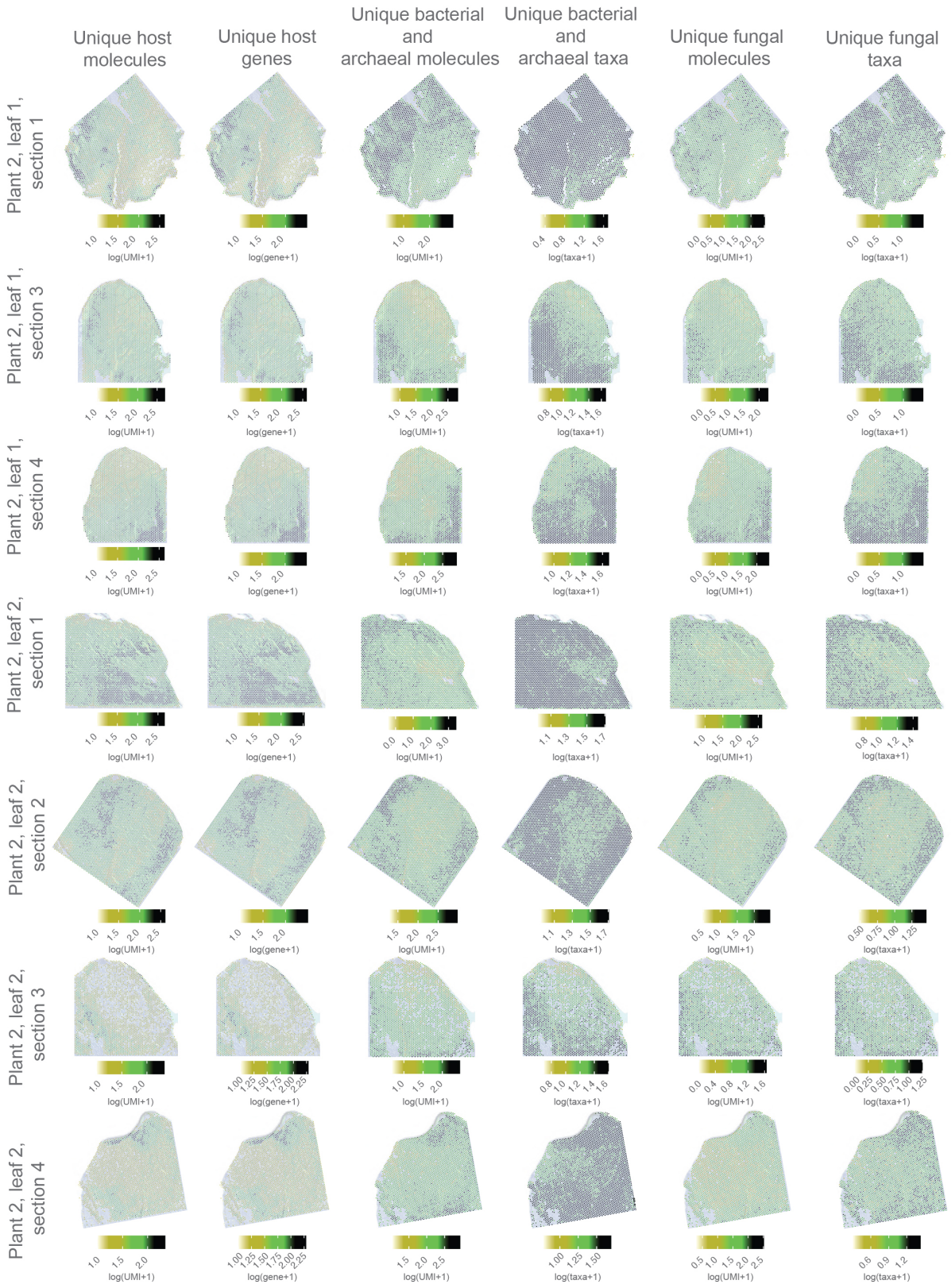


Supplementary Fig. 23: Relative abundance of the bacterial taxa and their hierarchical clustering in the whole tissue area in axenically-grown and outdoor-grown leaves. ‘P’ - plant, ‘L’ - leaf.

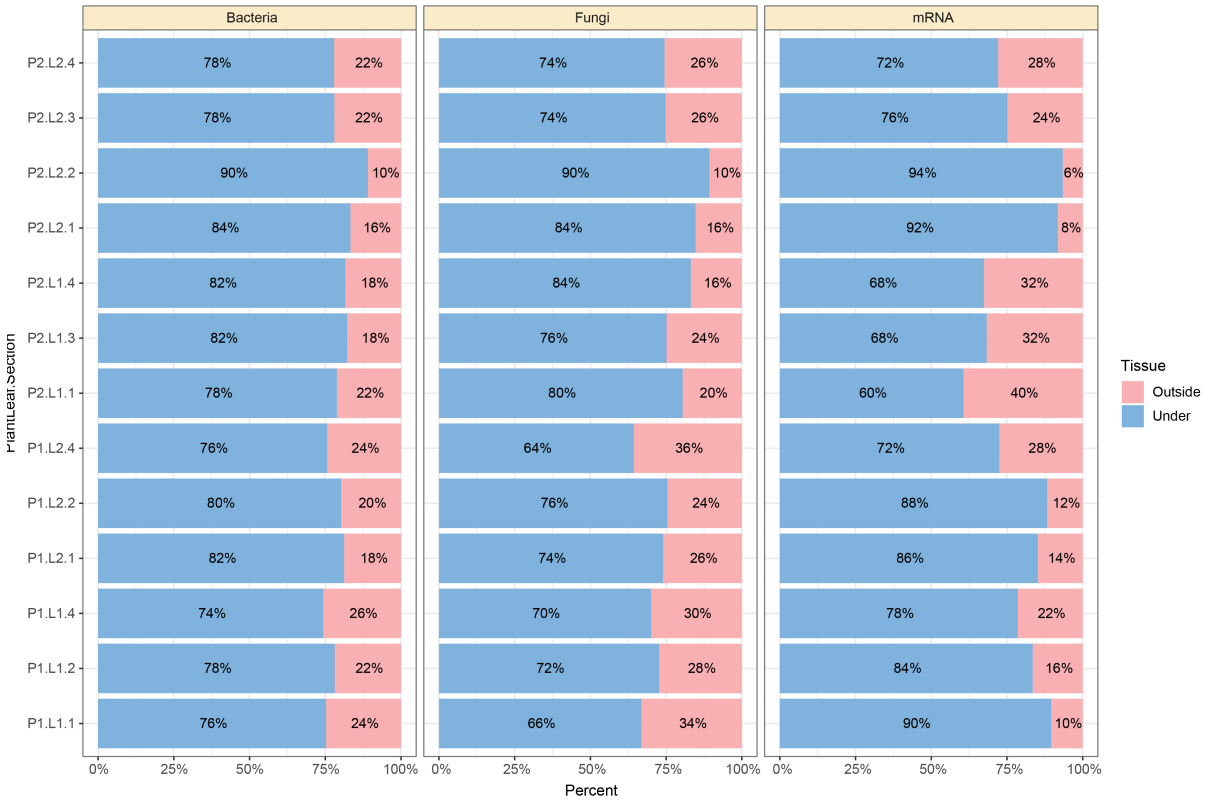


Supplementary Fig. 24: Relative abundance of the bacterial taxa and their hierarchical clustering in the whole tissue area in axenically-grown and outdoor-grown leaves. 'P' - plant, 'L' - leaf.

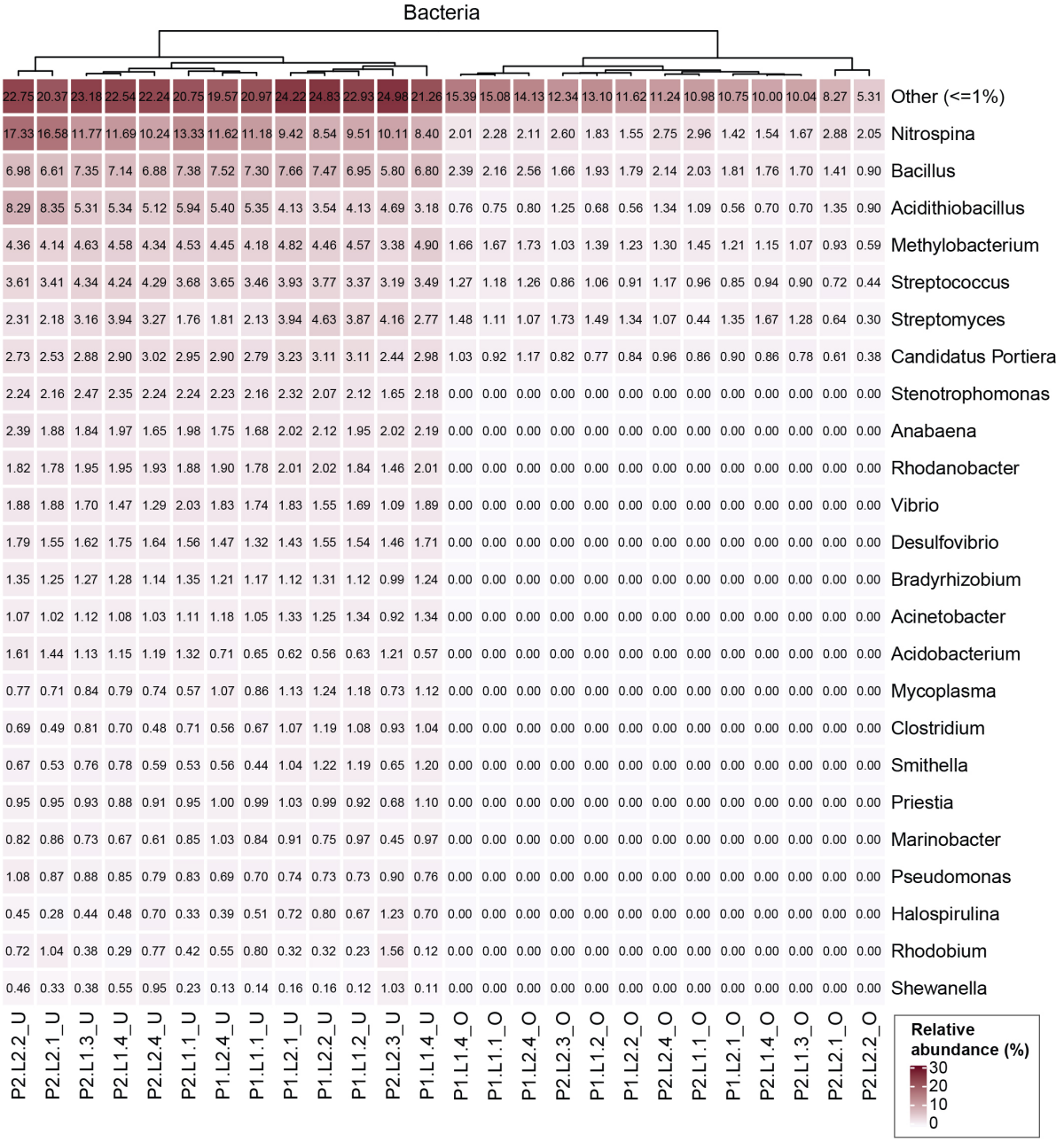

Unique host molecules
Unique host genes
Unique bacterial and archaeal molecules
Unique bacterial and archaeal taxa
Unique fungal molecules
Unique fungal taxa
Plant 1, leaf 1, section 1
Plant 1, leaf 1, section 2
Plant 1, leaf 1, section 4
Plant 1, leaf 2, section 1
Plant 1, leaf 2, section 2
Plant 1, leaf 2, section 4
log(UMI+1)
log(gene+1)
log(taxa+1)



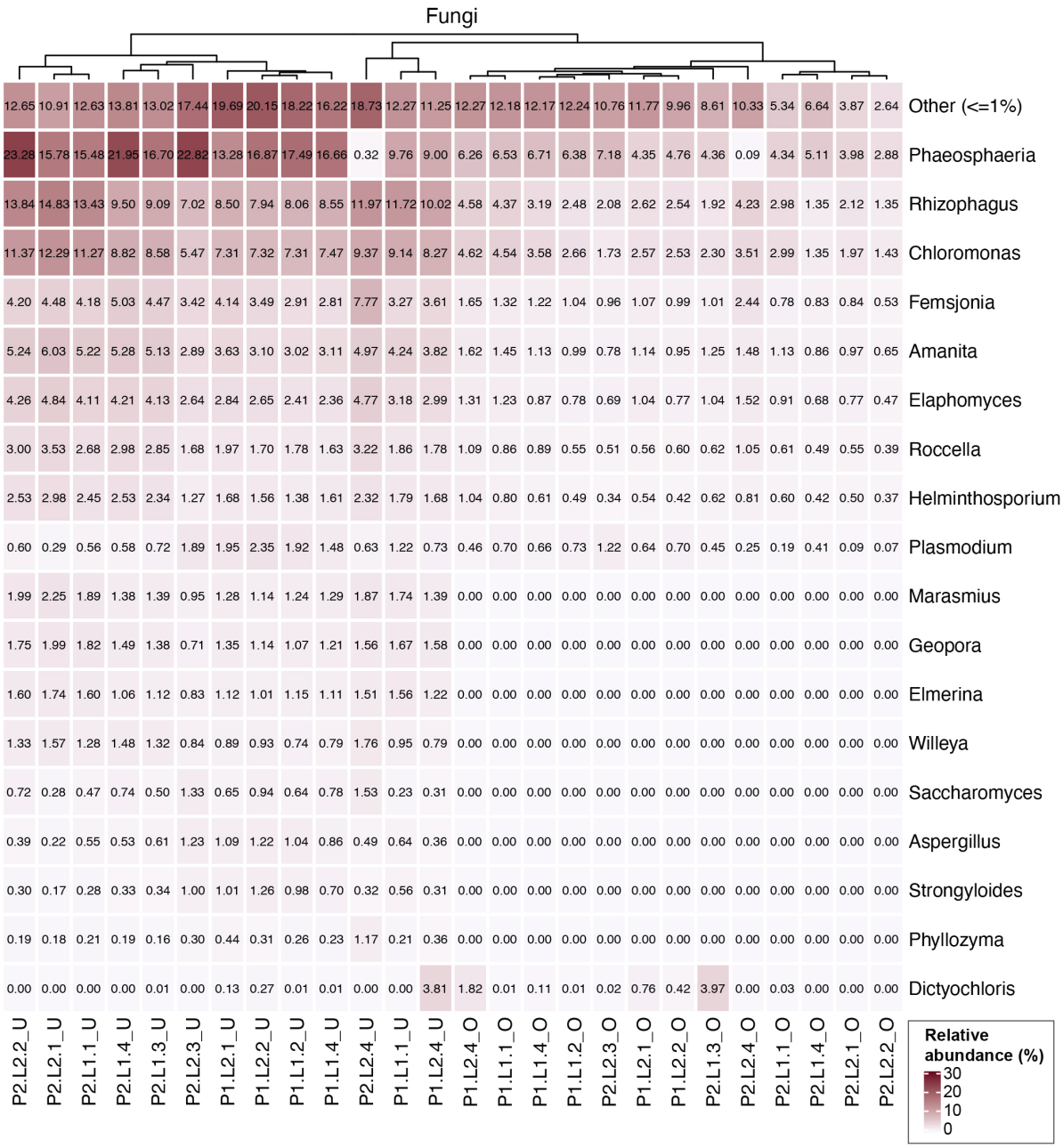
Supplementary Fig. 25: Unique host molecules, unique genes, unique bacterial and archaeal molecules, unique bacterial and archaeal taxa, unique fungal molecules and unique fungal taxa in $\log_{10}$ scale for all the sections in the dataset.



Supplementary Fig. 26: Proportion of unique molecules under the tissue and outside the tissue in each of the leaf sections.

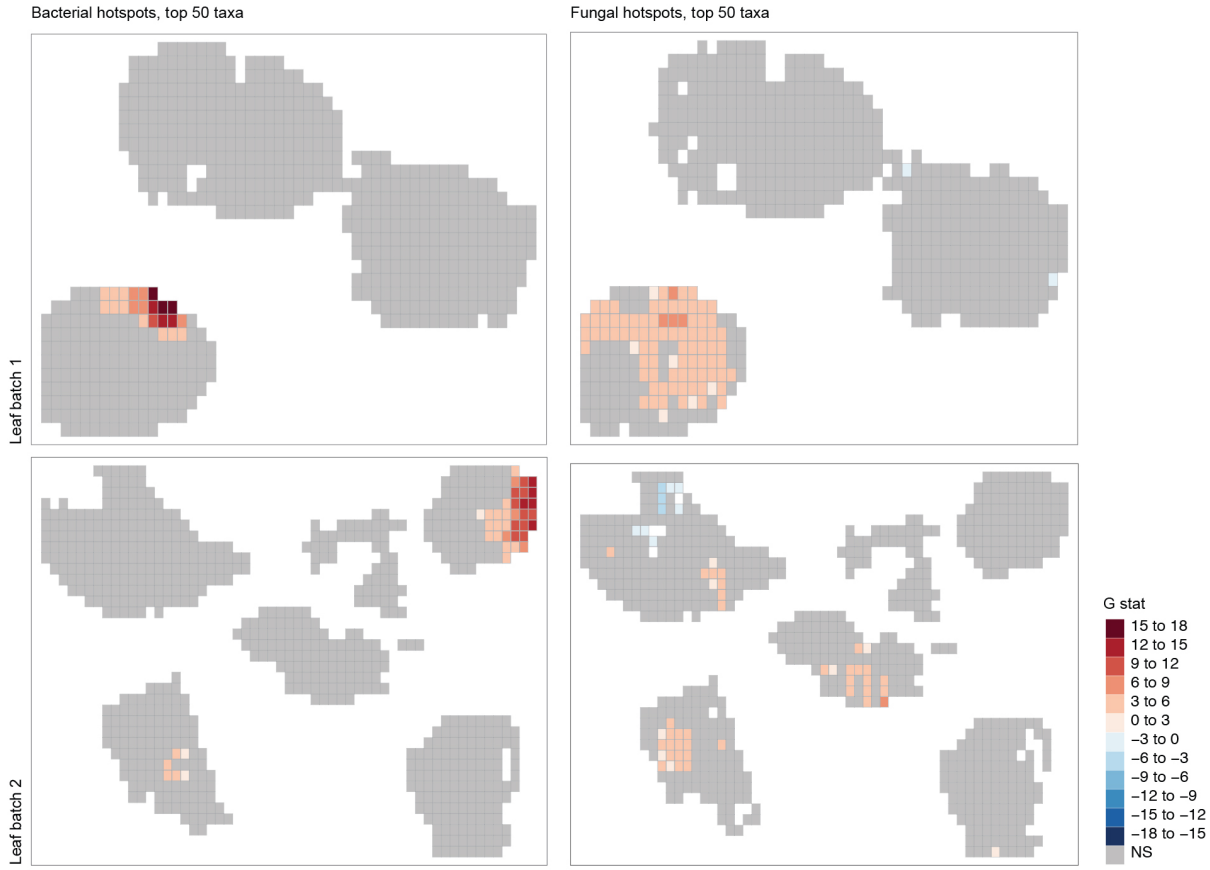


Supplementary Fig. 27: Relative bacterial abundance profiles under- and outside-the-tissue grouped by hierarchical clustering. The suffix 'O' refers to outside-the-tissue and the suffix 'U' refers to under-the-tissue.

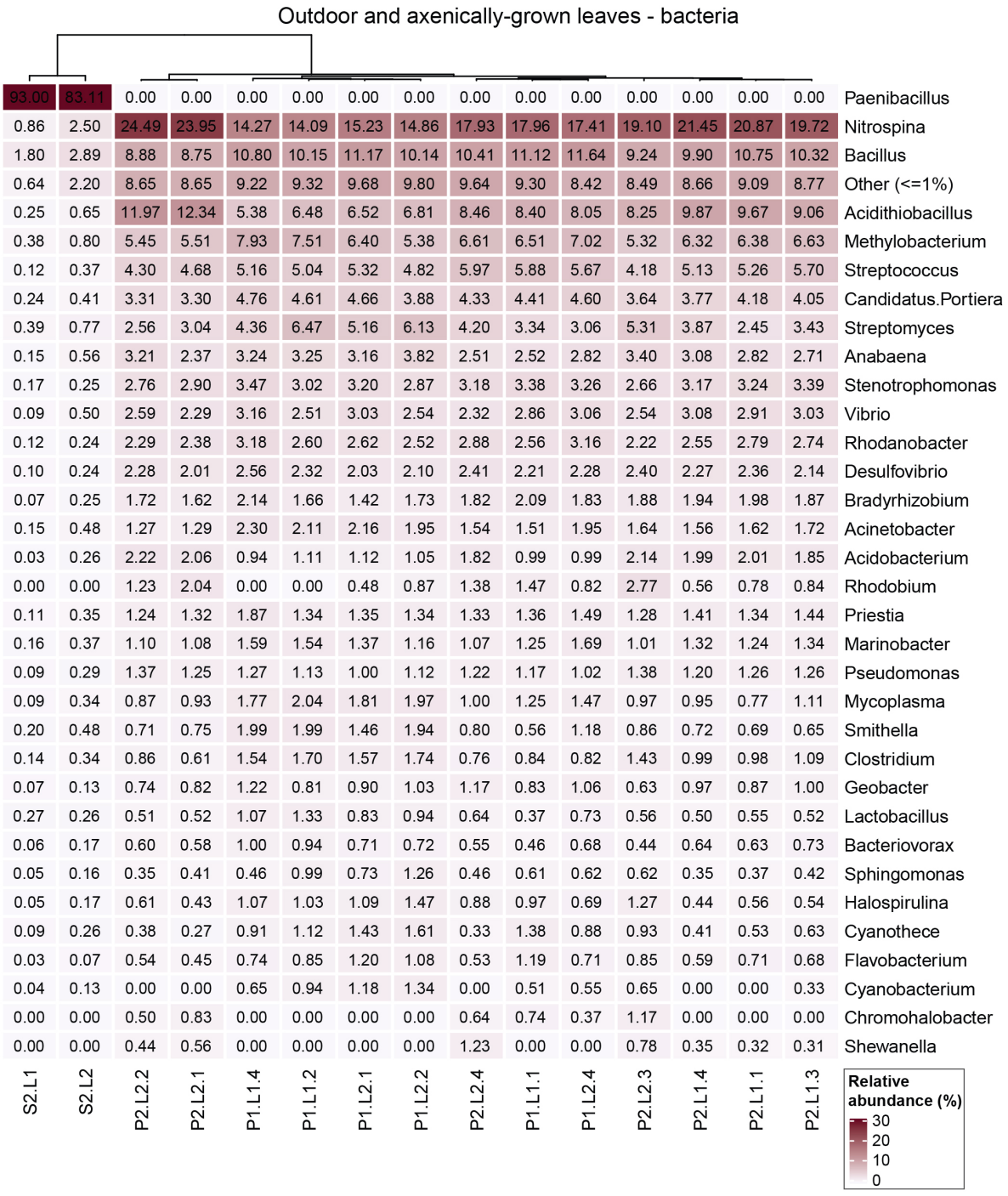


Supplementary Fig. 28: Relative fungal abundance profiles under- and outside-the-tissue grouped with hierarchical clustering. The suffix ‘O’ refers to outside-the-tissue and the suffix ‘U’ refers to under-the-tissue.

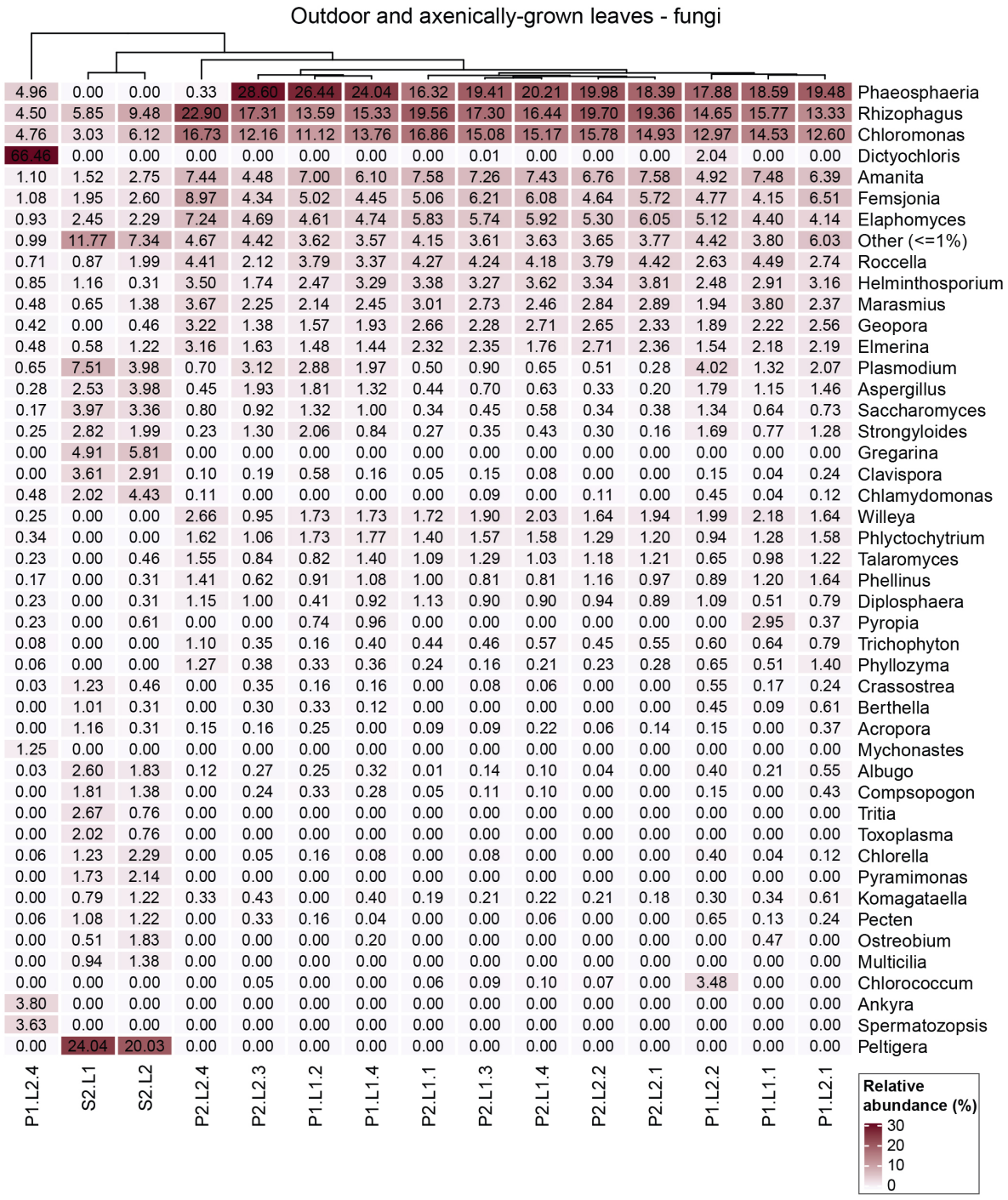
Supplementary Fig. 29: Bacterial and archaeal and fungal hotspots for each of the leaf sections.



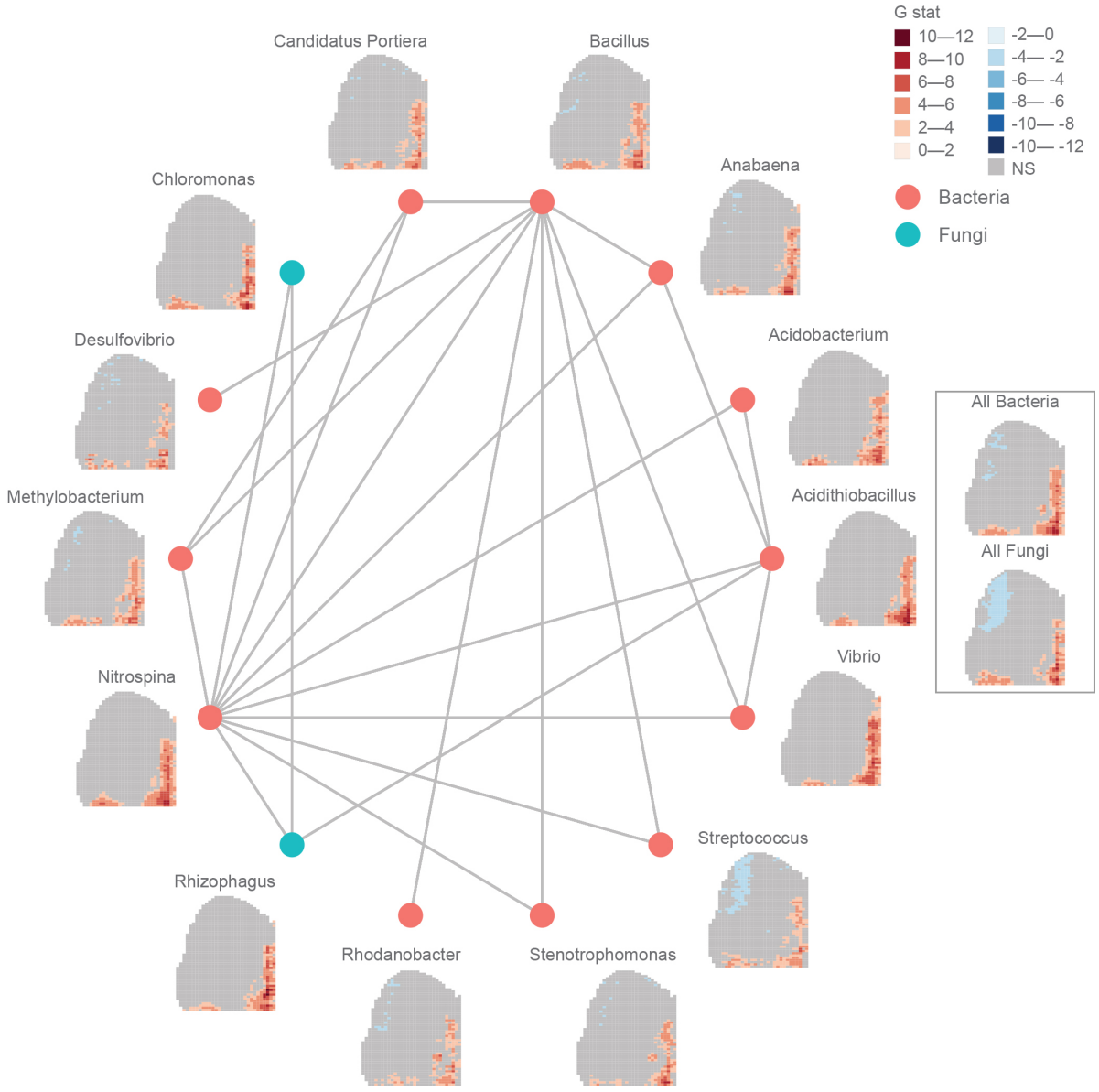
Supplementary Fig. 30: Significant hot- and cold-spots for bacteria and fungi in axenically-grown leaf sections.



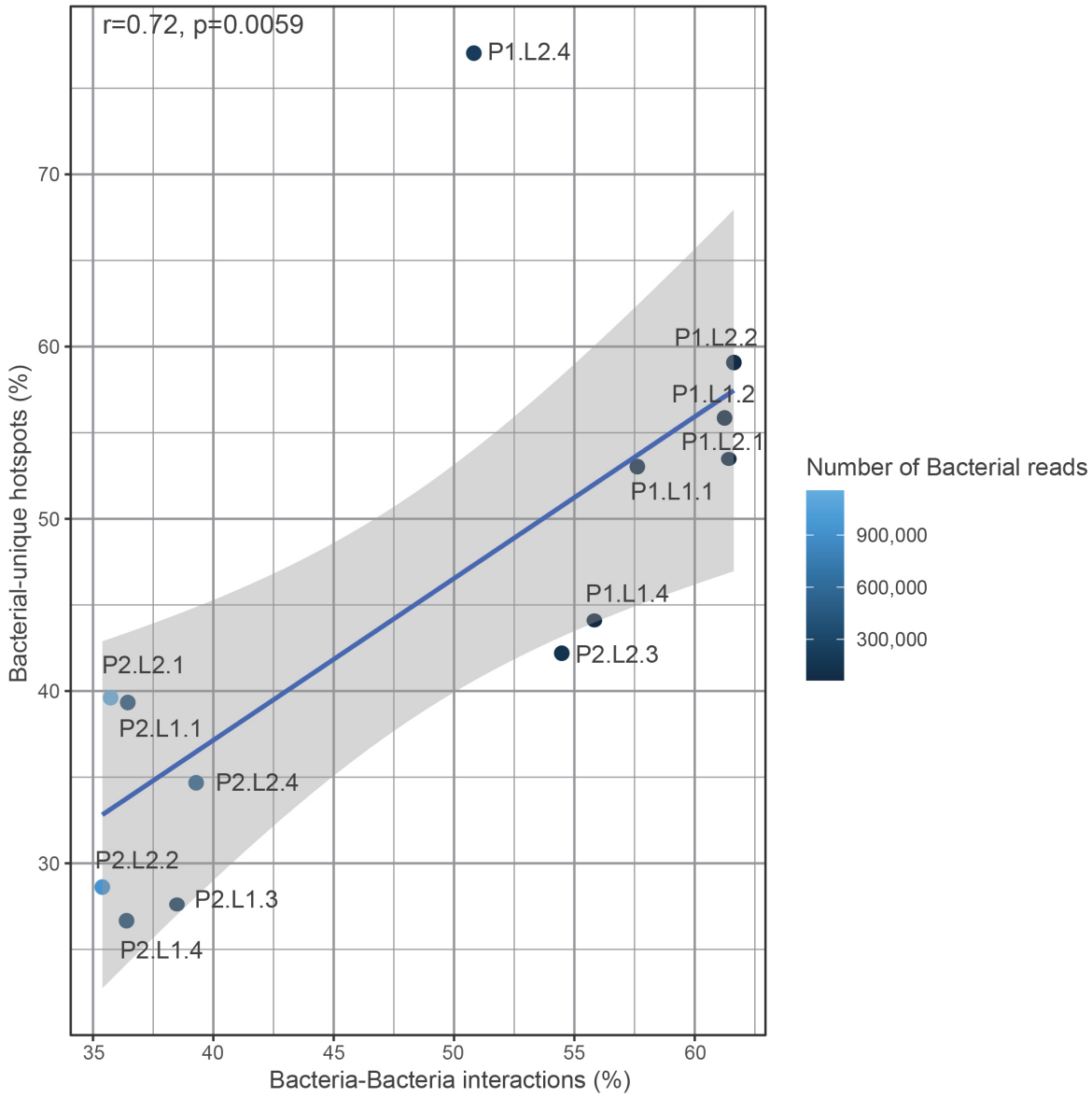
Supplementary Fig. 31: Relative abundance of the bacterial taxa and their hierarchical clustering in the significant hotspots in axenically-grown and outdoor-grown leaves. ‘P’ - plant, ‘L’ - leaf.



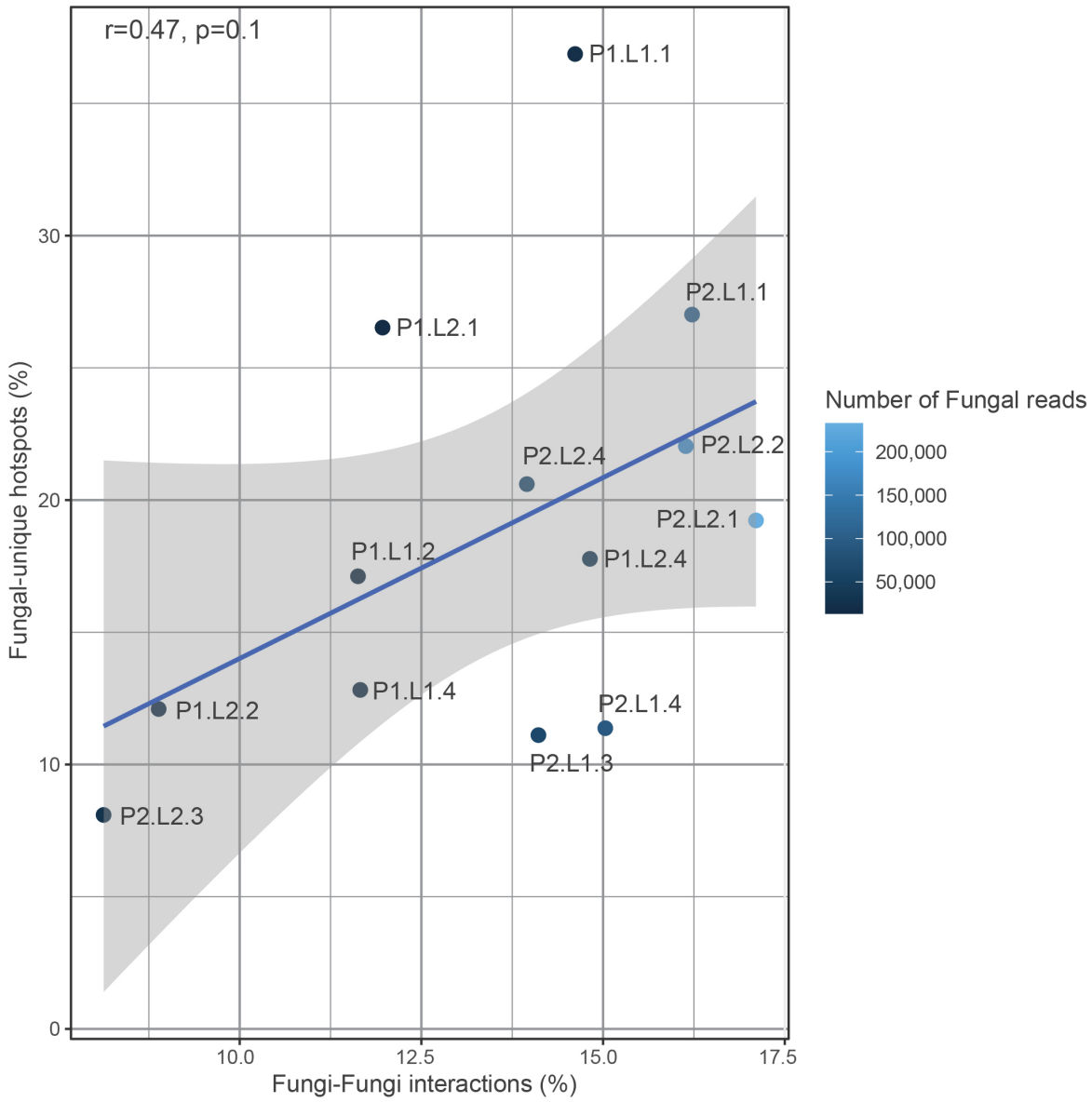
Supplementary Fig. 32: Relative abundance of the fungal taxa and their hierarchical clustering in the significant hotspots in axenically-grown and outdoor-grown leaves. ‘P’ - plant, ‘L’ - leaf.



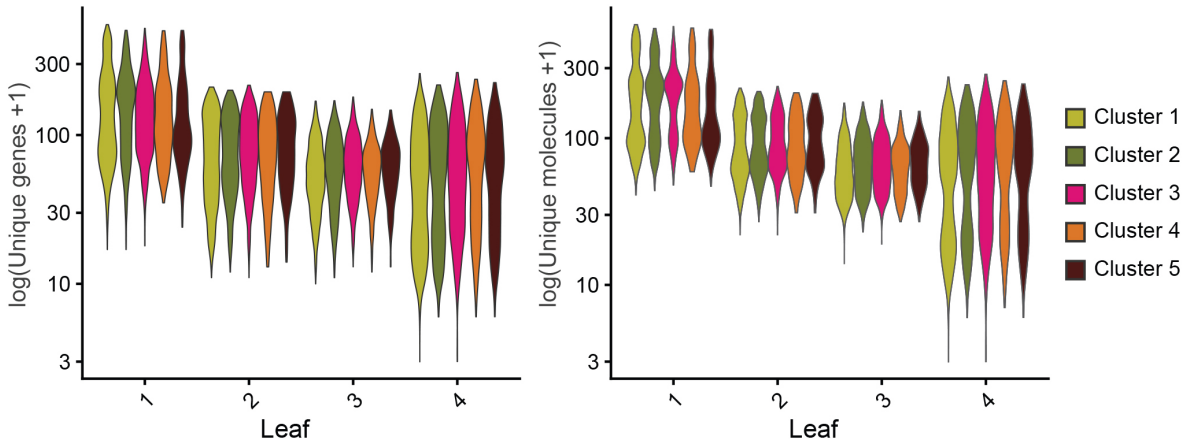
Supplementary Fig. 33: Subnetwork of 14 bacterial and fungal taxa strongly associated with microbial abundance in all leaf sections analysed. An edge connects two taxa if their average pairwise SRCC, across all sections, is above or equal to 0.35. The hotspot pattern for each genus in a representative leaf section (P2.L1.4) is shown next to the network nodes.



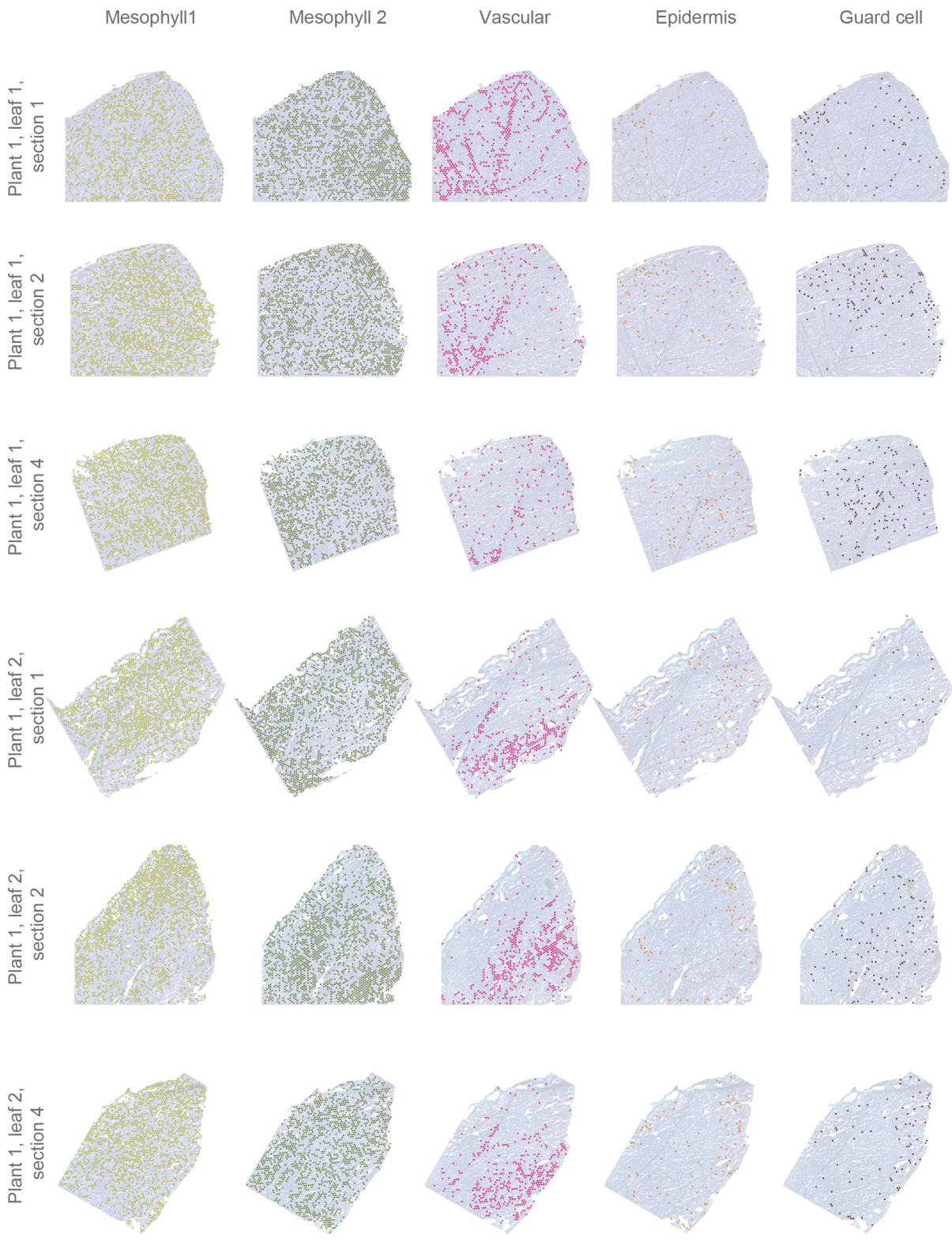
Supplementary Fig. 34: The proportion of bacteria-bacteria interactions as a function of the proportion of bacterial-unique hotspots. Spearman's rank correlation coefficient ρ=0.72, p-value=0.059 (two-tailed test).

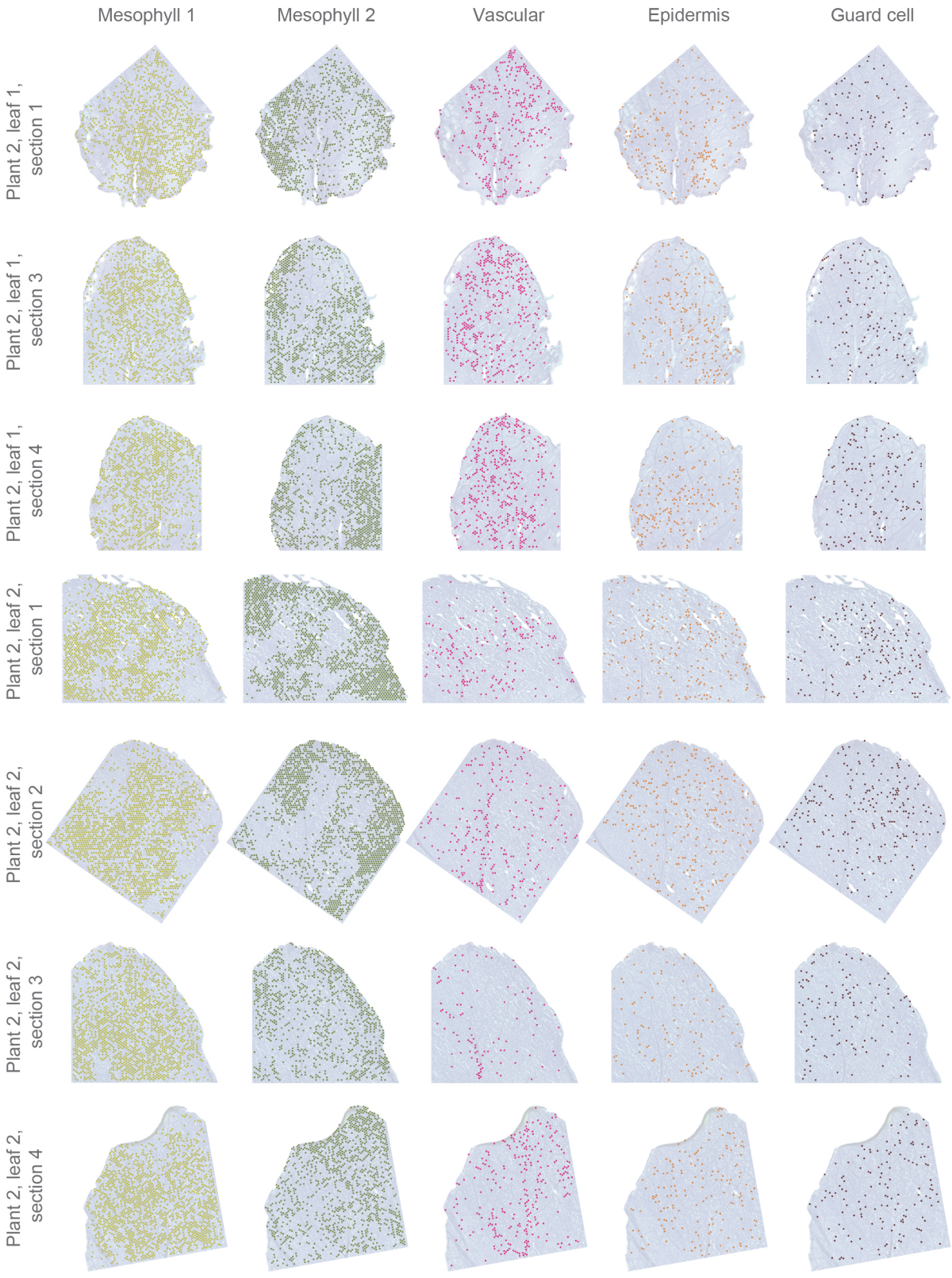


Supplementary Fig. 35: The proportion of fungi-fungi interactions as a function of the proportion of fungi-unique hotspots. Spearman's rank correlation coefficient ρ=0.47, p-value=0.1 (two-tailed test).

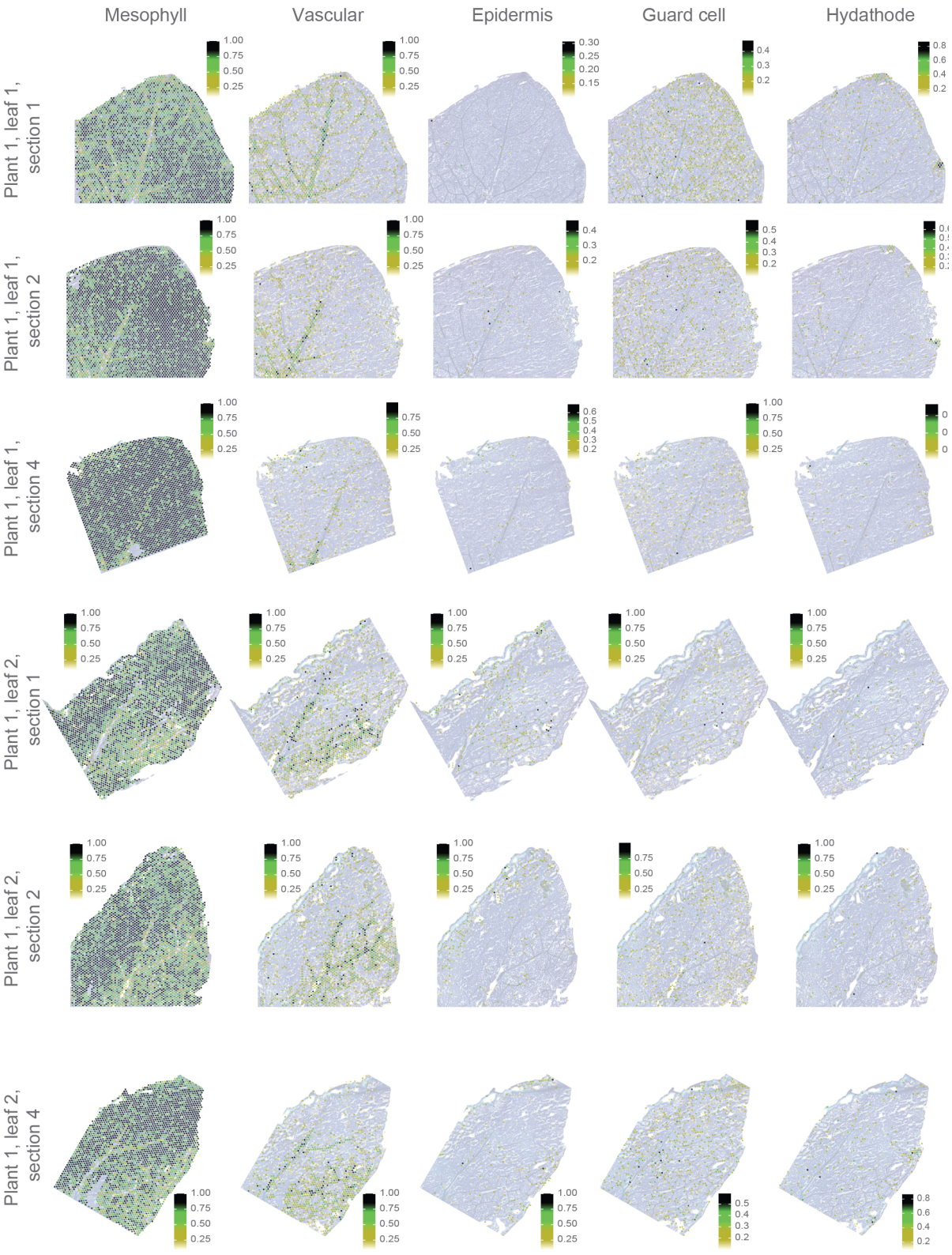


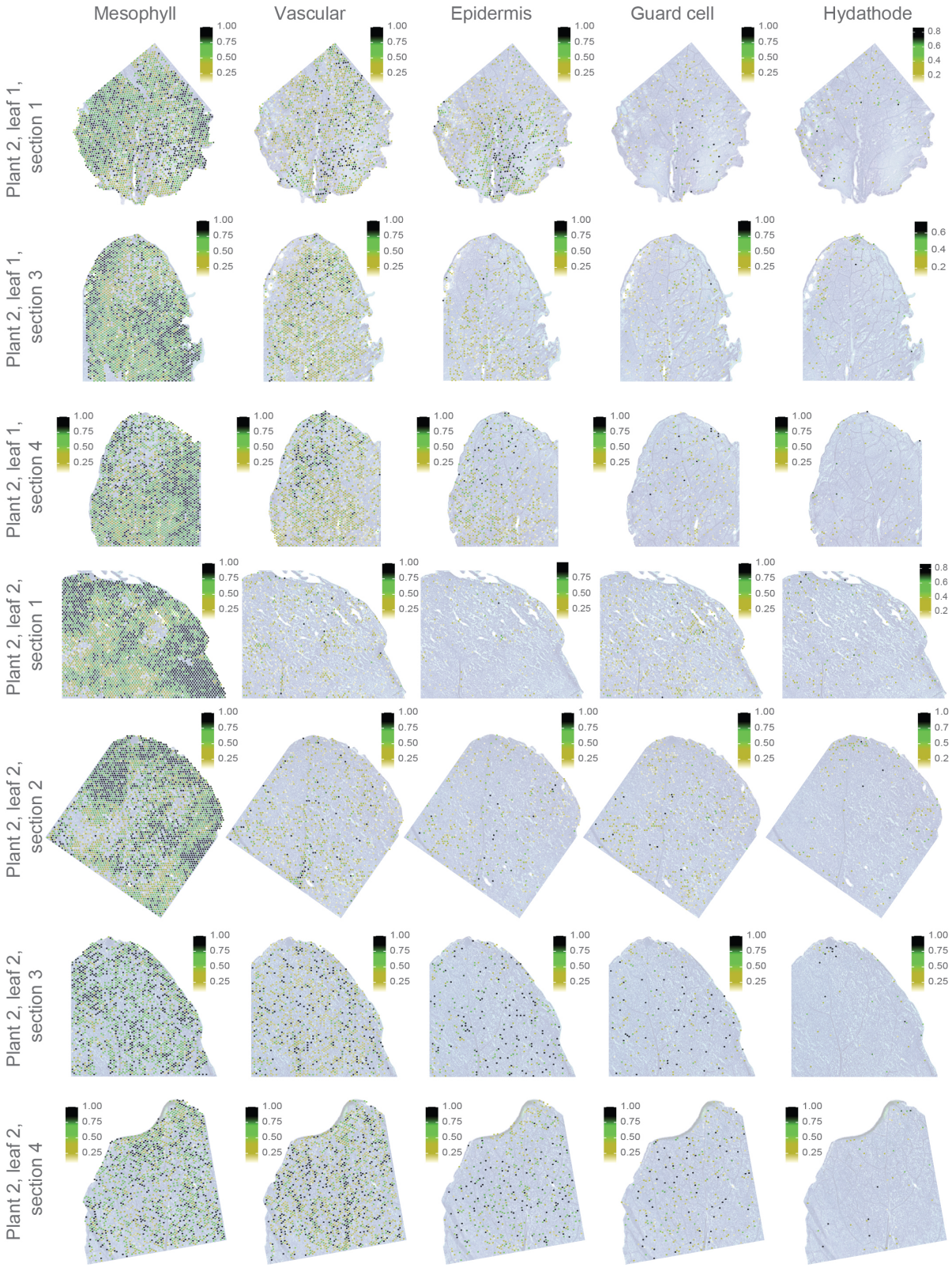
Supplementary Fig. 36: Normalised unique genes and unique molecules for each host gene expression unsupervised cluster in each leaf.


Mesophyll1
Mesophyll 2
Vascular
Epidermis
Guard cell
Plant 1, leaf 1, section 1
Plant 1, leaf 1, section 2
Plant 1, leaf 1, section 4
Plant 1, leaf 2, section 1
Plant 1, leaf 2, section 2
Plant 1, leaf 2, section 4

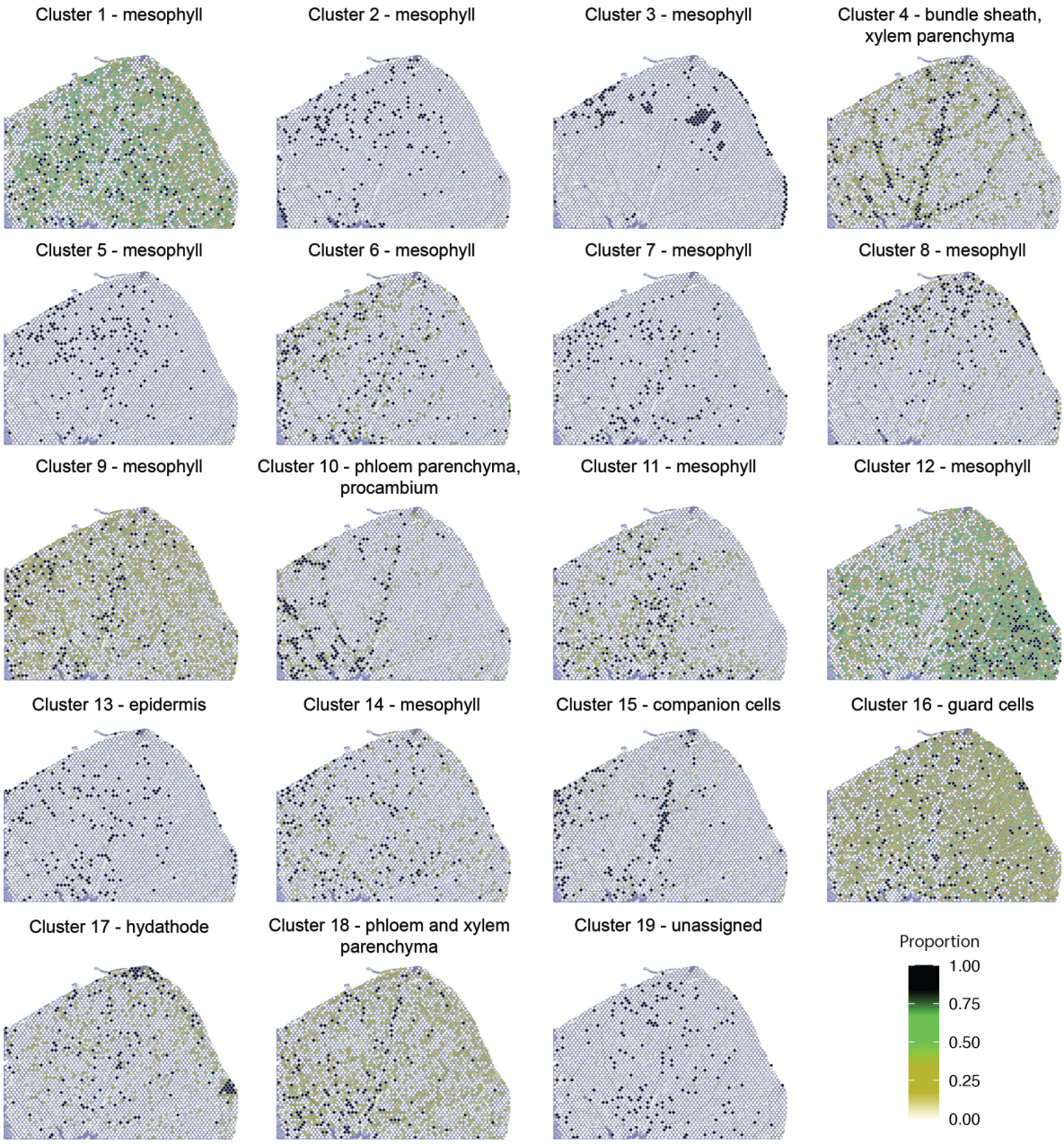


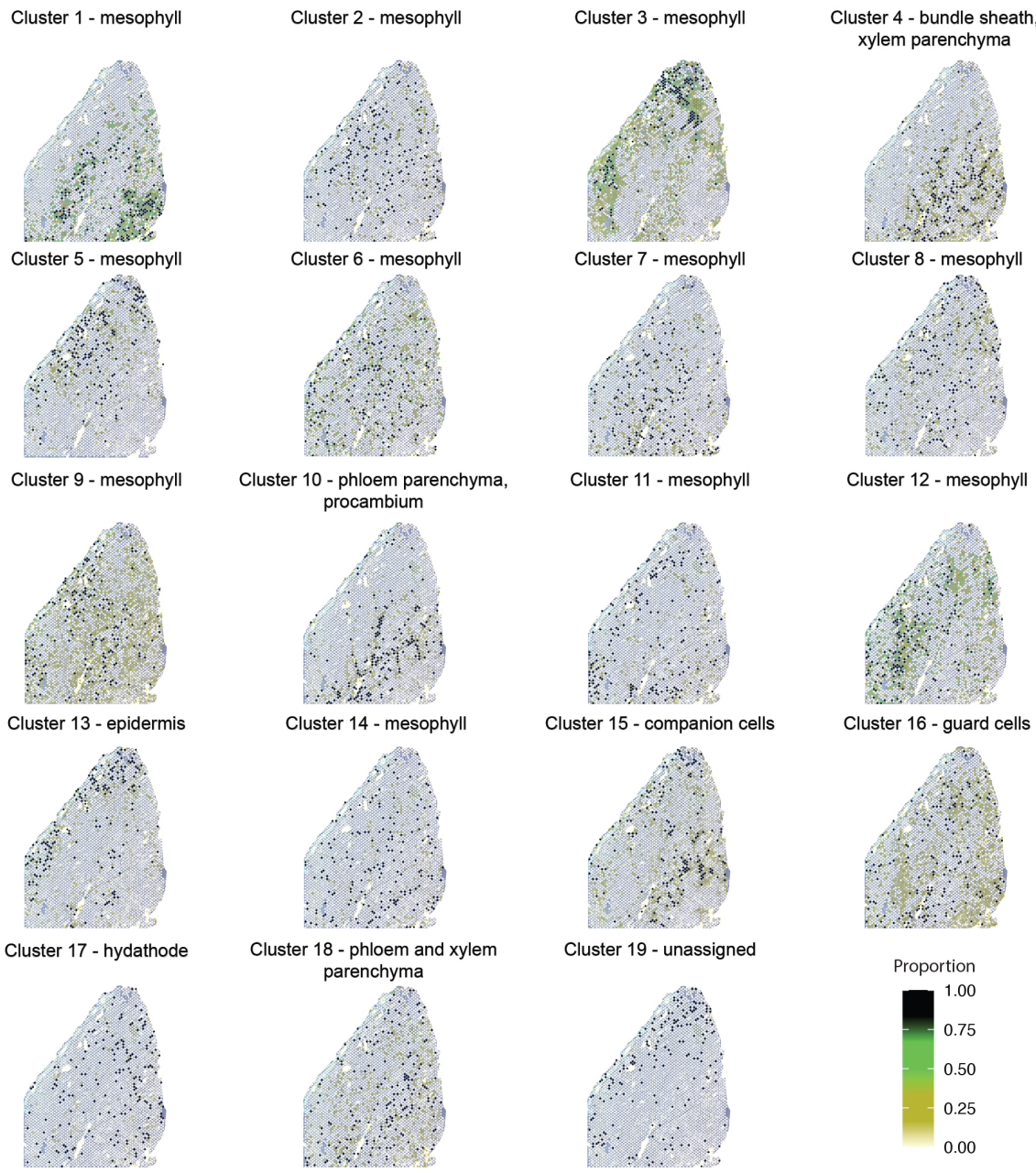
Supplementary Fig. 37: All five UMAP clusters individually visualised in each tissue section.


Mesophyll
Vascular
Epidermis
Guard cell
Hydathode
Plant 1, leaf 1, section 1
Plant 1, leaf 1, section 2
Plant 1, leaf 1, section 4
Plant 1, leaf 2, section 1
Plant 1, leaf 2, section 2
Plant 1, leaf 2, section 4

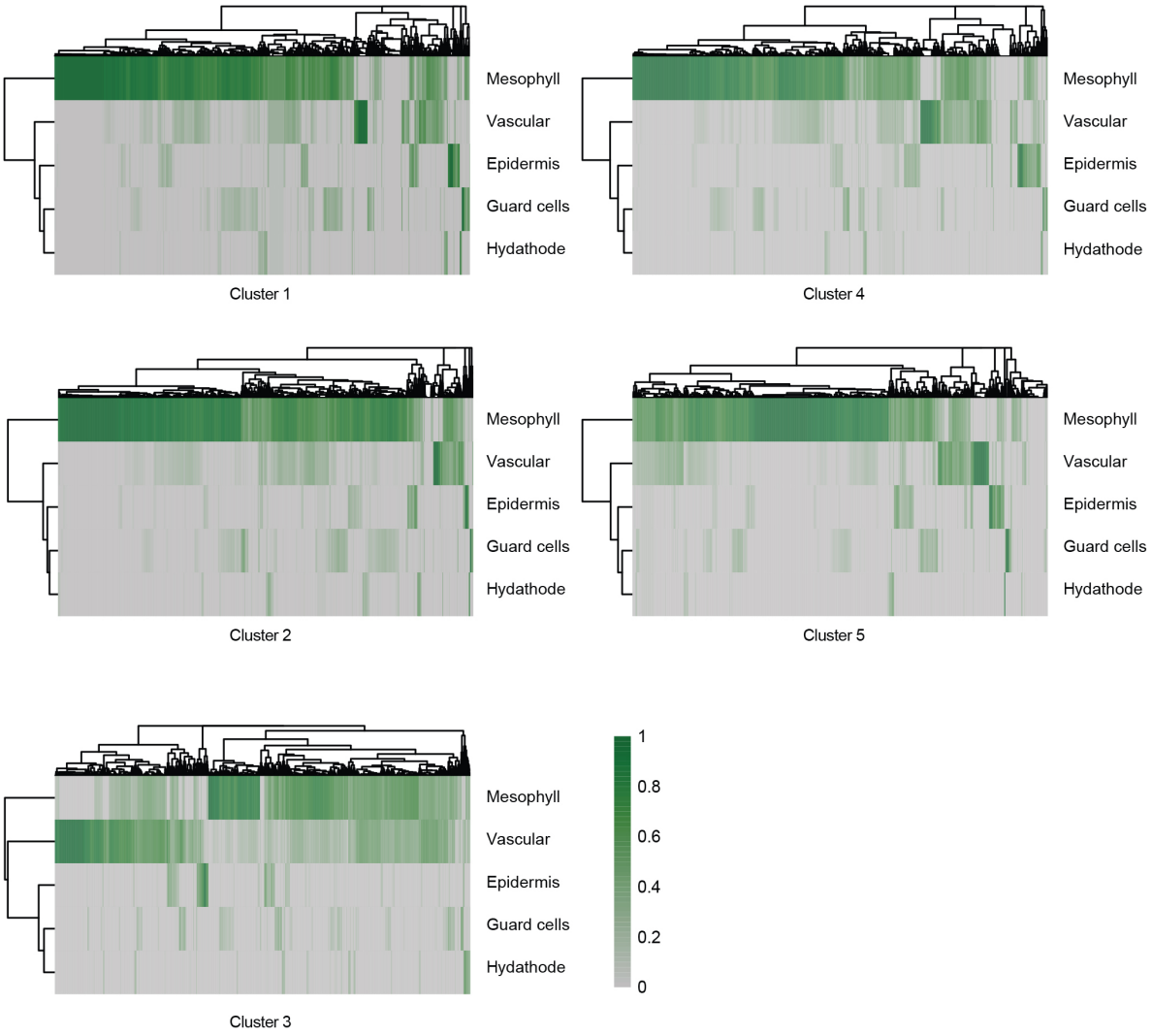


Supplementary Fig. 38: Single-cell proportions of the merged main cell types projected on each of the tissue sections. To improve the visualisation the values presented here are scaled quantiles (see **Methods**), and only spots with a cell type proportion greater than zero are displayed.

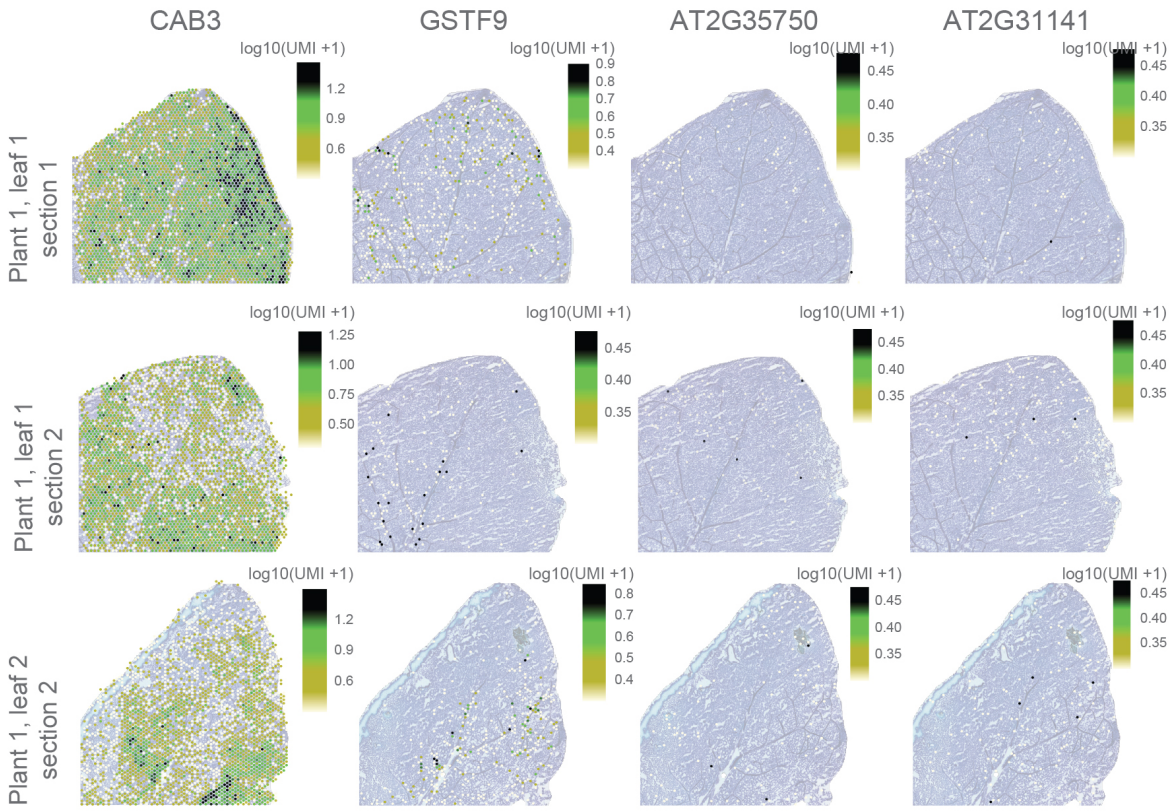

Cluster 1 - mesophyll
Cluster 2 - mesophyll
Cluster 3 - mesophyll
Cluster 4 - bundle sheath, xylem parenchyma
Cluster 5 - mesophyll
Cluster 6 - mesophyll
Cluster 7 - mesophyll
Cluster 8 - mesophyll
Cluster 9 - mesophyll
Cluster 10 - phloem parenchyma, procambium
Cluster 11 - mesophyll
Cluster 12 - mesophyll
Cluster 13 - epidermis
Cluster 14 - mesophyll
Cluster 15 - companion cells
Cluster 16 - guard cells
Cluster 17 - hydathode
Cluster 18 - phloem and xylem parenchyma
Cluster 19 - unassigned
Proportion
1.00
0.75
0.50
0.25
0.00



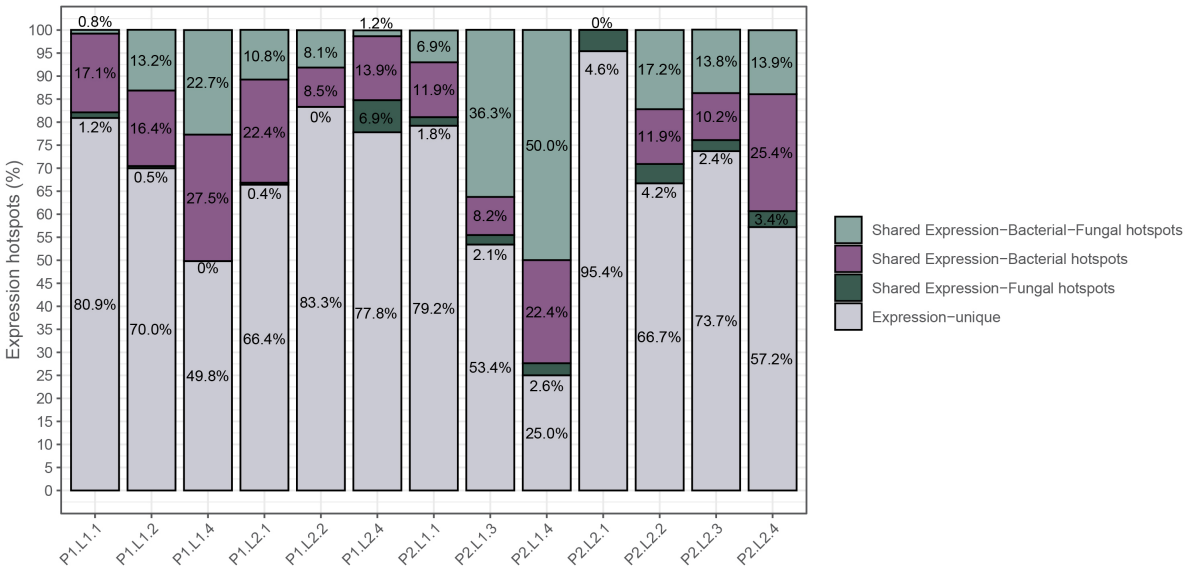
Supplementary Fig. 39: Single-cell proportions of all the cell types projected on two representative tissue sections. To improve the visualisation the values presented here are scaled quantiles (see **Methods**), and only spots with a cell type proportion greater than zero are displayed.



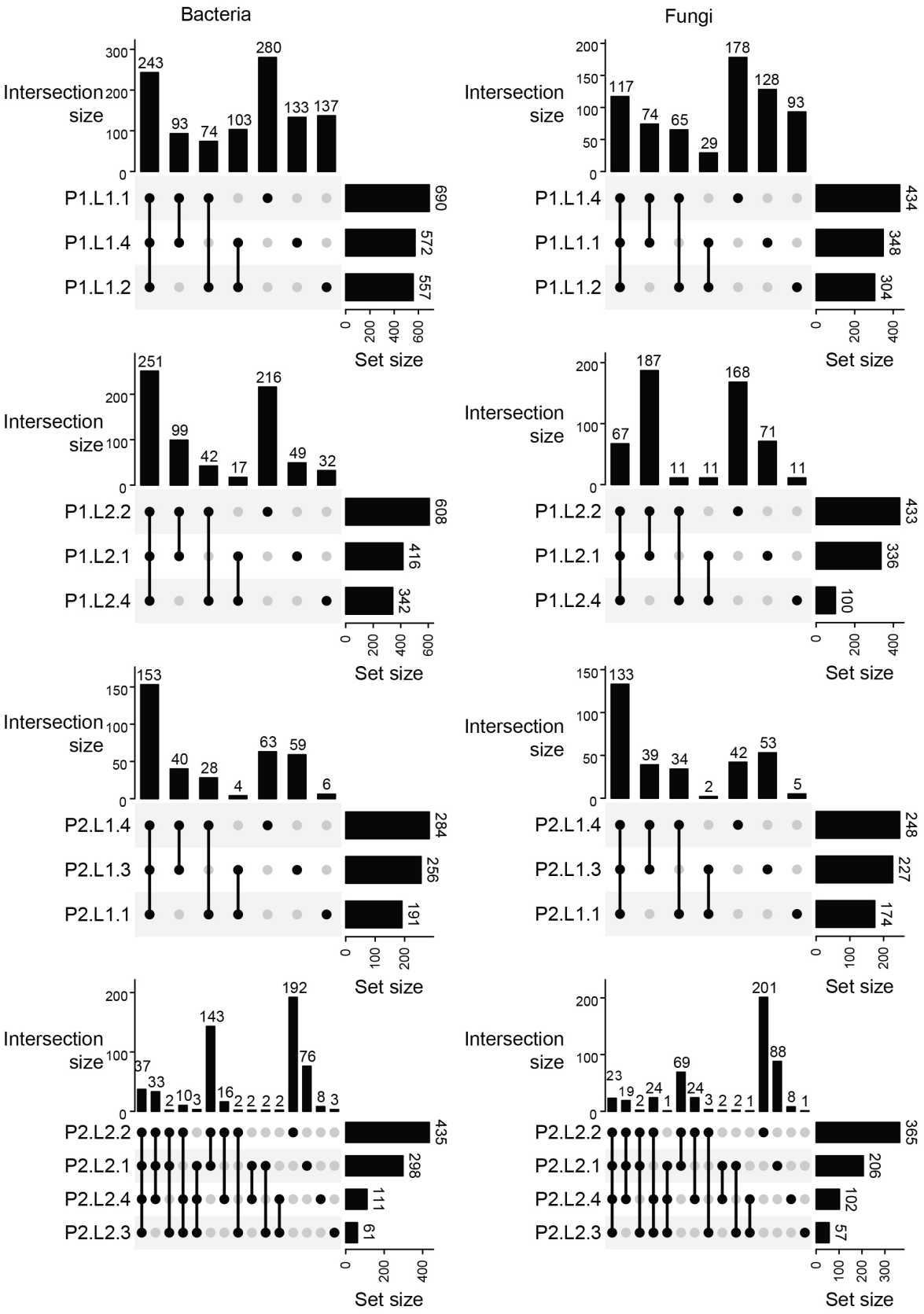
Supplementary Fig. 40: Estimated cell type proportion values from 0-1 for each of the clusters and hierarchical clustering for each spot and cell type.



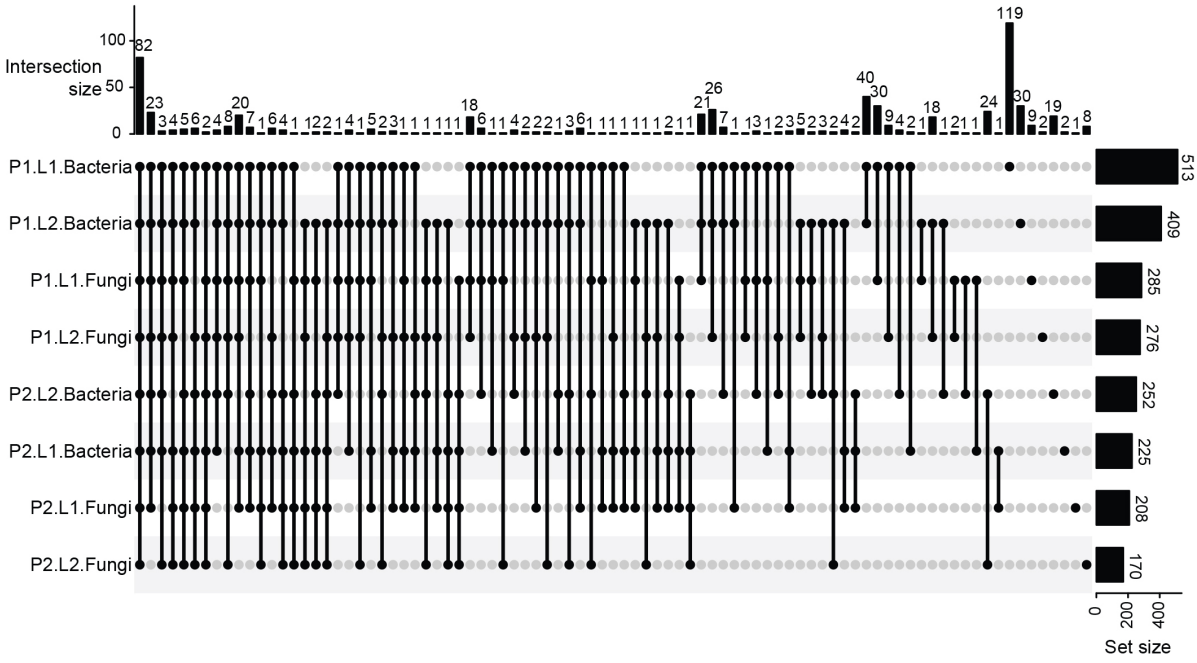
Supplementary Fig. 41: Representative markers for some of the clusters for three representative sections. *CAB3*: CHLOROPHYLL A/B BINDING PROTEIN 3 (Cluster 2 marker), *GSTF9*: GLUTATHIONE S-TRANSFERASE PHI 9 (Cluster 3 marker), AT2G35750: transmembrane protein (Cluster 4 marker), AT2G31141: potential nitrate responsive gene (Cluster 5 marker).



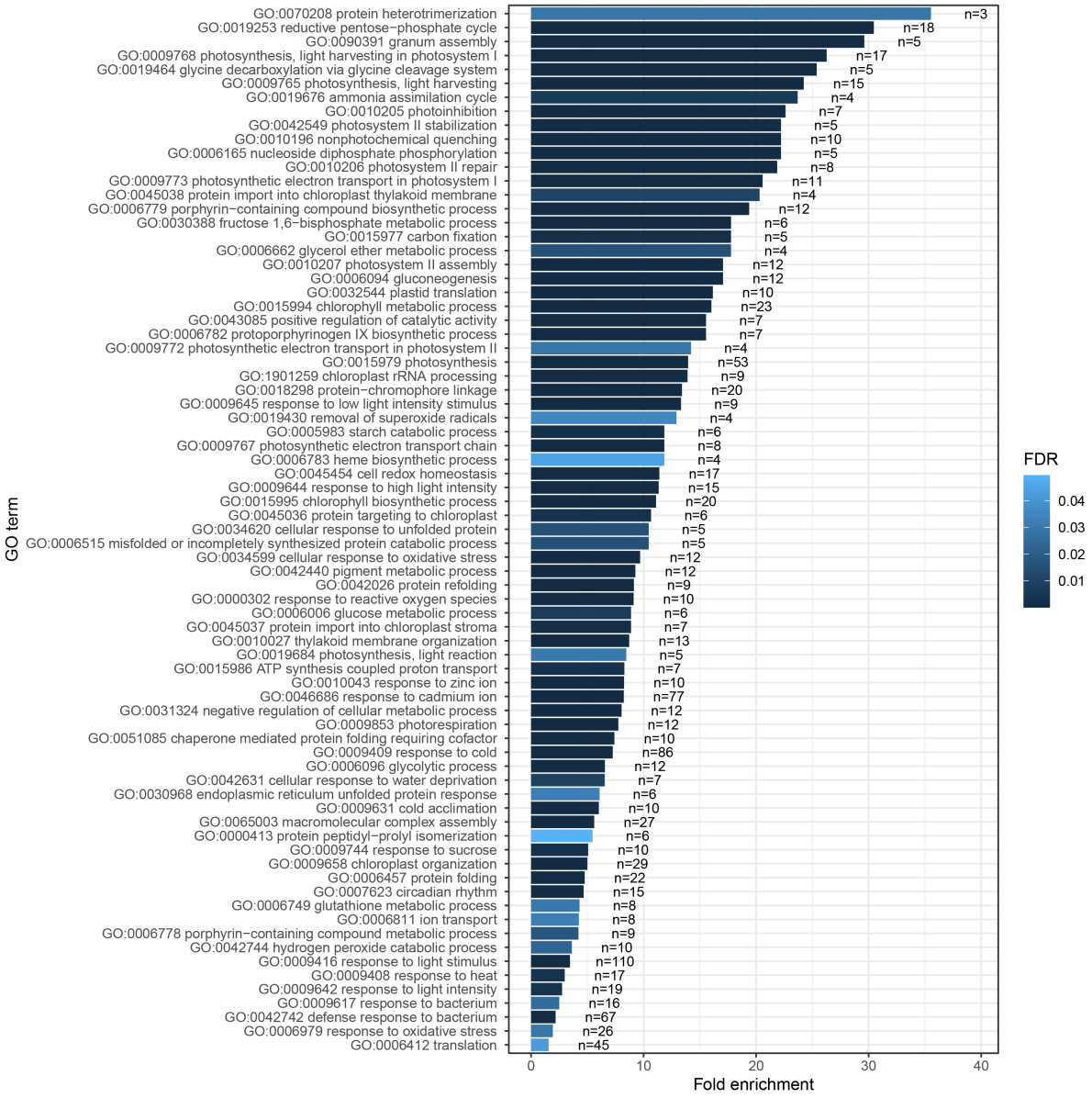
Supplementary Fig. 42: Proportion of hotspots shared between the host *A. thaliana*, bacterial taxa, and fungal taxa in each of the sections.



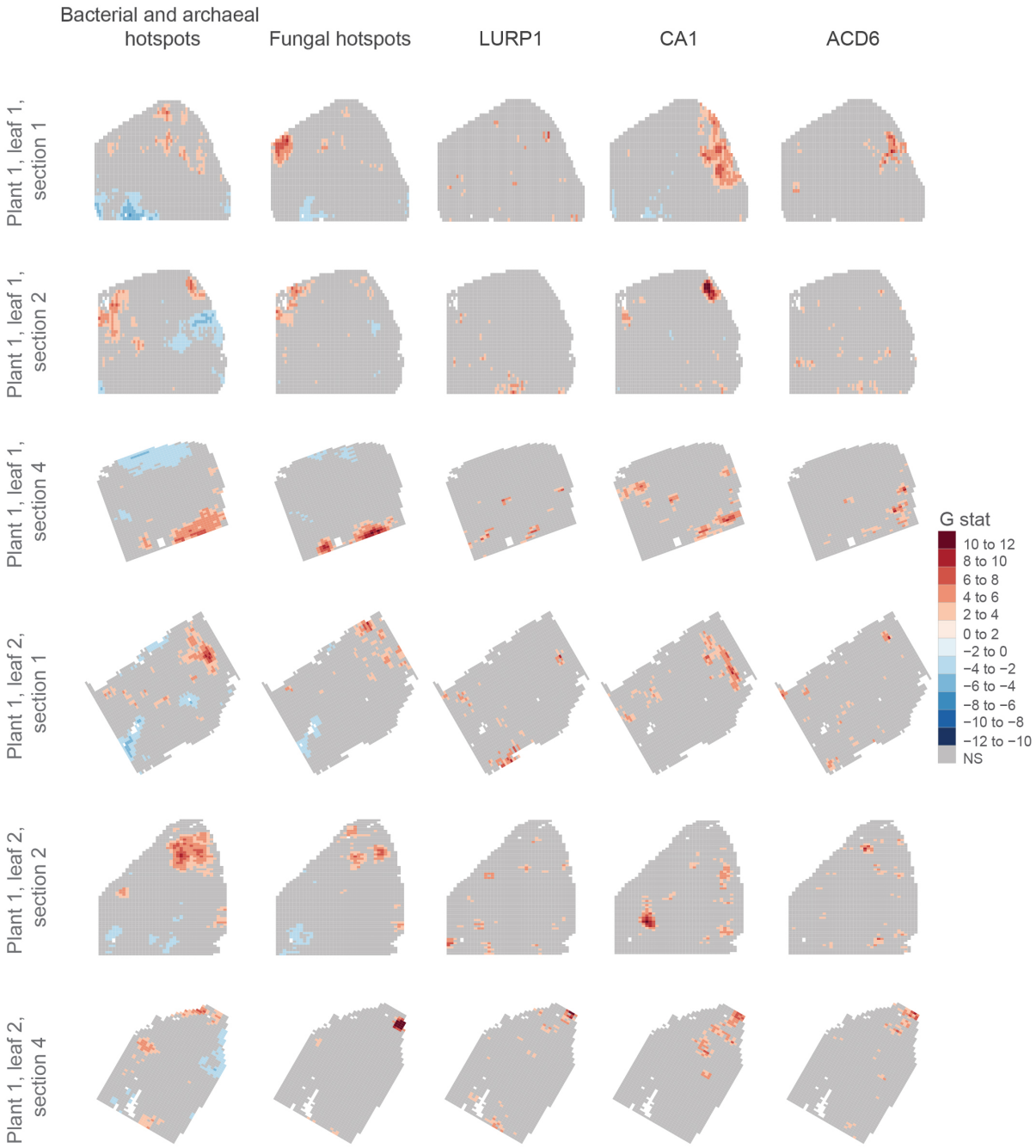
Supplementary Fig. 43: Set of genes selected by Boruta as explanatory of the total bacterial or fungal abundance in a given leaf section were intersected across the leaf sections. Intersection size of the different groups are shown.

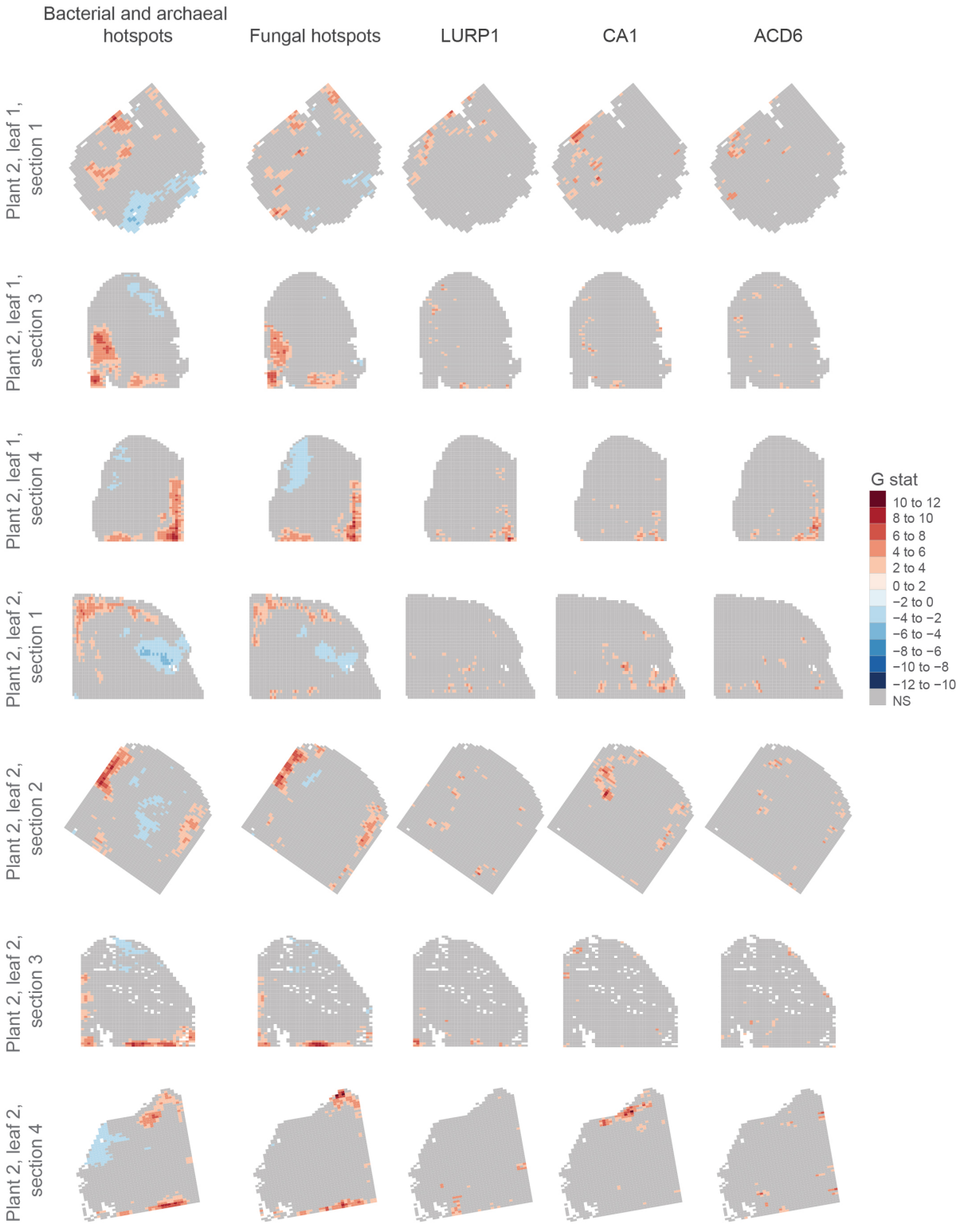


Supplementary Fig. 44: Intersection of bacterial and fungal associated genes across different leaves. Most of the genes are common to bacteria and fungi.

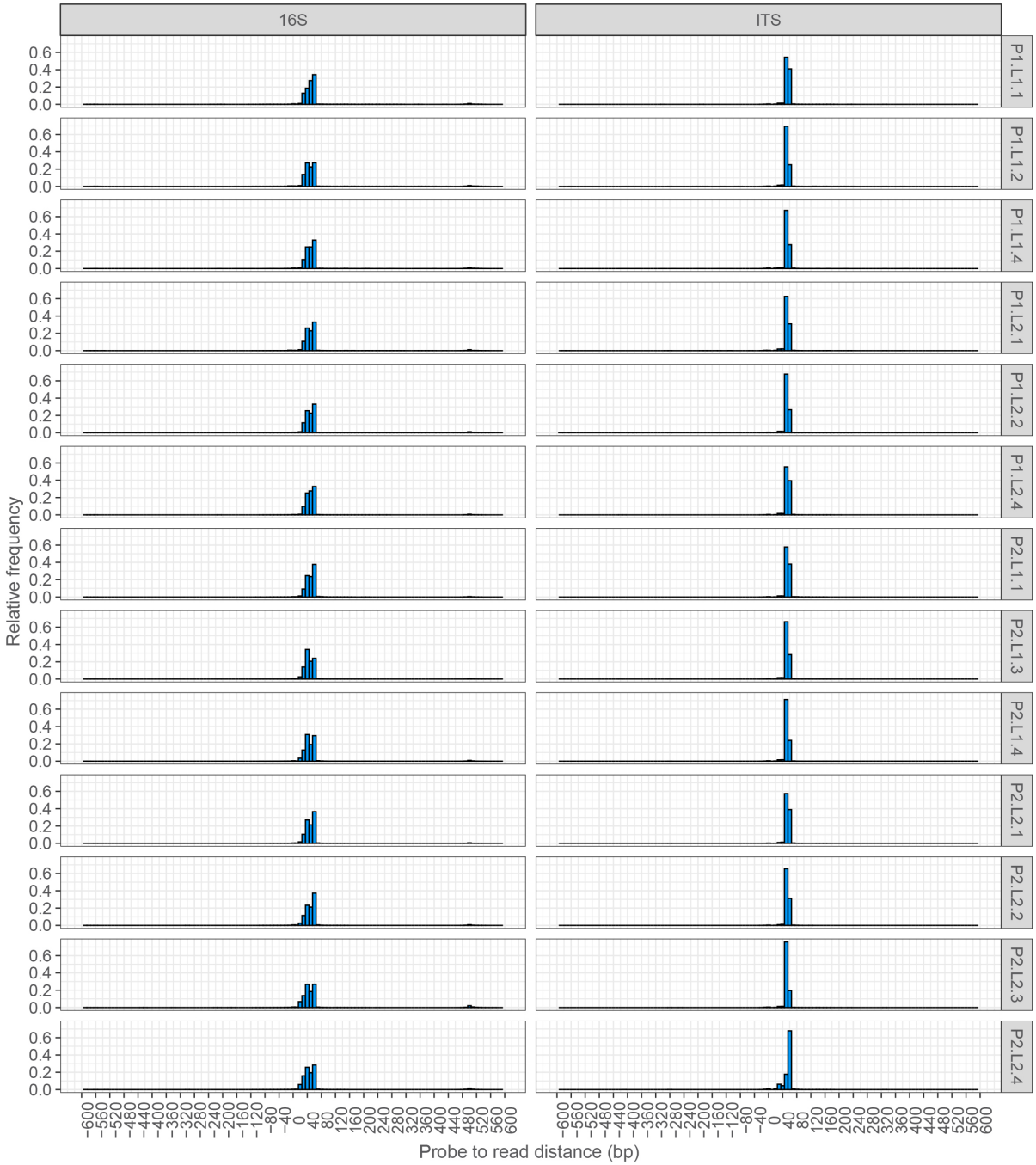


Supplementary Fig. 45: Overrepresented GO terms of all genes significantly associated with microbial abundance (n=the number of genes assigned with a specific GO term).


Bacterial and archaeal hotspots
Fungal hotspots
LURP1
CA1
ACD6
Plant 1, leaf 1, section 1
Plant 1, leaf 1, section 2
Plant 1, leaf 1, section 4
Plant 1, leaf 2, section 1
Plant 1, leaf 2, section 2
Plant 1, leaf 2, section 4
G stat
10 to 12
8 to 10
6 to 8
4 to 6
2 to 4
0 to 2
−2 to 0
−4 to −2
−6 to −4
−8 to −6
−10 to −8
−12 to −10
NS



Supplementary Fig. 46: Spatial distribution of significant hotspots of bacteria, fungi, and three defence-related genes: *CA1* (AT3G01500), *LURP1* (AT2G14560), and *ACD6* (AT4G14400).



Supplementary Fig. 47: The distribution of distance between reads annotated as microbes to their respective probes (after mapping the reads to NCBI 'nt' database). The result of each section among the different outdoor-grown leaves is presented. 16S analysis comprises the pooled four 16S probes used in this study, while ITS comprises the two probes used. The probe-to-read distance for the vast majority of reads is 0-60 bp, confirming that reads were captured by their expected probes.