

Spatial metatranscriptomics resolves host–bacteria–fungi interactomes

Received: 27 June 2022

Accepted: 6 September 2023

Published online: 20 November 2023

 Check for updatesSami Saarenpää ^{1,7}, Or Shalev^{2,5,7}, Haim Ashkenazy ^{2,7}, Vanessa Carlos^{2,3}, Derek Severi Lundberg^{2,6}, Detlef Weigel ^{2,4} & Stefania Giacomello ¹✉

The interactions of microorganisms among themselves and with their multicellular host take place at the microscale, forming complex networks and spatial patterns. Existing technology does not allow the simultaneous investigation of spatial interactions between a host and the multitude of its colonizing microorganisms, which limits our understanding of host–microorganism interactions within a plant or animal tissue. Here we present spatial metatranscriptomics (SmT), a sequencing-based approach that leverages 16S/18S/ITS/poly-d(T) multimodal arrays for simultaneous host transcriptome- and microbiome-wide characterization of tissues at 55- μm resolution. We showcase SmT in outdoor-grown *Arabidopsis thaliana* leaves as a model system, and find tissue-scale bacterial and fungal hotspots. By network analysis, we study inter- and intrakingdom spatial interactions among microorganisms, as well as the host response to microbial hotspots. SmT provides an approach for answering fundamental questions on host–microbiome interplay.

Advances in spatially resolved transcriptomics technologies have greatly improved the understanding of eukaryotic host gene expression mechanisms in animal and plant tissues^{1–4}. These technologies have been designed to capture targeted^{3,5,6} or untargeted^{1,2,4} RNA information based on imaging or sequencing of unique molecules, enabling the study of hundreds of genes or the whole transcriptome, respectively.

Spatial variation is also prominent in host–microorganism interactions, and single-cell RNA-sequencing (scRNA-seq) of the host has been used to understand how this affects host cellular responses during infection⁷. However, integrated spatially resolved analyses of microbial identity and the host response remain rare and are typically focused on individual microbial taxa within a host⁸. With existing technology, it has not been possible to simultaneously resolve the spatial interactions between a host and the multitude of microorganisms colonizing it. This has considerably limited our understanding of host–microorganism interactions at the tissue level.

Microorganisms often live in diverse communities surrounded by other microorganisms. Both cooperative and antagonistic interactions between microorganisms are known to be important for the functionality and health of ecosystems, plants, animals and humans^{9–11}. Moreover, the success of microbial colonization and infection depends strongly on the spatial structure of microbial interactions with other microorganisms and with multicellular species, and several pioneering studies have revealed clear and functionally significant spatial organization in host-associated microbial communities^{12–14}. Much broader knowledge of the spatial organization of microorganisms within hosts, and the associated local host responses, is therefore needed to fully understand the biology of the host–microorganism–microorganism interactome.

Fluorescence in situ hybridization (FISH)-based techniques provided the first insights into microbial spatial organization in different environments¹⁵ and in host tissues, including mouse gut¹⁶, human plaque microfilms¹⁶ and *Arabidopsis thaliana* roots¹⁷. A limitation of these targeted methods is that they use a set of predesigned probes,

¹SciLifeLab, Department of Gene Technology, KTH Royal Institute of Technology, Stockholm, Sweden. ²Max Planck Institute for Biology Tübingen, Tübingen, Germany. ³Cluster of Excellence Physics of Life, TU Dresden, Dresden, Germany. ⁴Institute for Bioinformatics and Medical Informatics, University of Tübingen, Tübingen, Germany. ⁵Present address: Systems Biology of Microbial Communities, University of Tübingen, Tübingen, Germany. ⁶Present address: Swedish University of Agricultural Sciences, Uppsala, Sweden. ⁷These authors contributed equally: Sami Saarenpää, Or Shalev, Haim Ashkenazy. ✉e-mail: stefania.giacomello@scilifelab.se

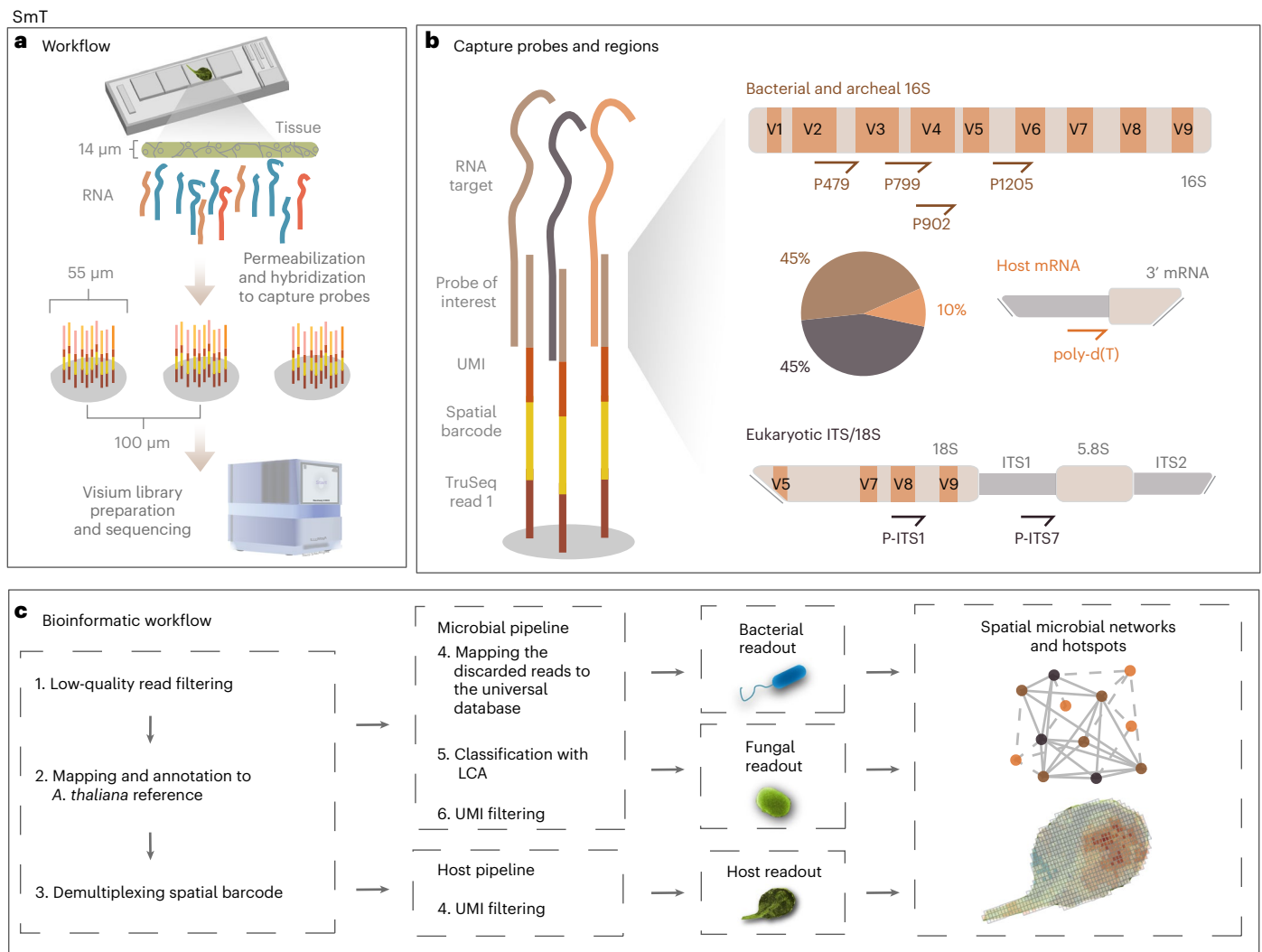


Fig. 1 | Overview of the method. a, SmT uses capture arrays on glass slides. Each capture array contains 4,992 spots that are 55 μm in diameter and 100 μm from center to center. Cells are permeabilized to release RNA molecules that hybridize to the barcoded capture probes in the spots. The captured molecules are then processed into a sequencing library. **b**, Capture probes consist of a sequencing adapter, a spatial barcode, a UMI and a capture moiety. Polyadenylated mRNAs are captured with poly-d(T) probes that comprise 10% of all the capture probes. Ribosomal RNAs from fungi are captured with P-ITS7 and P-ITS1 probes targeting the 18S rRNA and ITS regions, respectively. Ribosomal RNAs from bacteria and archaea are captured with P479, P799, P902 and P1205 probes targeting bacterial 16S rRNA. Bacterial and archaeal probes and fungal probes each comprise 45%

of the capture probes. **c**, A bioinformatic workflow designed to assign the reads to host or microbial modalities. First, low-quality reads are filtered out, the remaining reads are mapped against the *A. thaliana* TAIR10 reference genome, and spatial barcodes are demultiplexed. Second, mapped *A. thaliana* reads are filtered based on their UMI and compiled to obtain a gene-count matrix. Third, the reads not mapping to *A. thaliana* are mapped to a universal database to remove those that are not clearly of microbial origin. The remaining microbial reads are classified with LCA based on their identity and UMIs, and unique taxa are counted to generate separate unique taxa-count matrices for fungi and for bacteria and archaea.

each specific to a single microbial taxon. Current FISH-based technologies thus cannot provide comprehensive spatial descriptions of unknown microbiomes. Moreover, despite recent advances, these methods cannot yet achieve complete spatial resolution of the host's expression patterns due to their limited capacity and overfitting to specific hosts¹⁸.

Plants are colonized by a heterogeneous set of microorganisms whose diversity is comparable to that of the human gut's microbial population¹⁹. Similar to gut microorganisms, plant colonizing microorganisms affect the host's health and physiology in various ways, ranging from beneficial²⁰ to harmful^{21,22}. Plant microbial communities are shaped in an environment-dependent manner by the intertwined forces of host–microorganism and microorganism–microorganism interactions, which ultimately determine the fitness of the host and the associated microorganisms²³.

Because of the limitations of current analytical methods, microbial interactions within plants are often deduced from complete tissues or whole plants, based on 16S rDNA abundance data^{9,24,25}. This approach inevitably makes it impossible to resolve microscale differences in abundance. Hence, bulk RNA-seq can only be used to study average plant–microorganism interactions in a tissue^{26,27}. Given the tremendous variation of unique RNA profiles found within tissues, demonstrated repeatedly by spatial transcriptomic (ST) and scRNA-seq analyses^{2,28}, it is very likely that important information has been obscured by the limited spatial resolution of the techniques used to study plant–microorganism interactions.

Here we present spatial metatranscriptomics (SmT; Fig. 1), an untargeted approach that allows simultaneous interrogation of bacterial and fungal communities, and the corresponding host transcriptional responses with a spatial resolution of 55 μm . By capturing the

spatial distribution of bacterial and archaeal 16S rRNA sequences, together with fungal internal transcribed spacer (ITS) and 18S rRNA sequences and the host mRNAs, we link local changes in host gene expression to the size and composition of local microbial populations in *A. thaliana* leaves. We resolve the organization of microbial communities along tissue sections and demonstrate the presence of microbial hotspots at the leaf scale, and how these locally impact host responses.

Results

Spatial detection of bacterial infection and host response

To determine whether mRNA molecules could be captured from the host *A. thaliana* leaf sections while preserving the tissue's morphology, we applied an optimized ST protocol to leaves grown under laboratory conditions²⁹. To this end, we permeabilized a 14- μm thick longitudinal leaf section on a glass surface uniformly coated with poly-d(T) capture probes. Following cDNA synthesis with fluorophores, we obtained a fluorescent cDNA footprint (Fig. 2a) whose morphology matched to that of the original leaf, demonstrating that spatial host gene expression patterns can be obtained from longitudinal leaf sections. Next, because bacterial communities are typically characterized based on 16S rDNA sequences, we hypothesized that capturing 16S rRNA molecules could provide information on the spatial distribution of bacteria in host tissues. To prove this concept, we analyzed leaves of lab soil-grown *A. thaliana* plants infiltrated with the model pathogen *Pseudomonas syringae* pv. *tomato* DC3000 (Pst DC3000), which was genetically labeled to enable its fluorescence imaging in whole leaves (Fig. 2b). The array used in the analysis contained two degenerate probes (P799 and P902) to capture bacterial and archaeal (hereafter 'bacterial') diversity from 16S rRNA hypervariable regions, together with poly-d(T) probes to capture host mRNA, mixed in the following proportions: 50% poly-d(T), 25% P799 and 25% P902.

We imaged intact infected leaves 3 d postinfiltration to record the fluorescent spatial pattern of the bacterial infiltration and analyzed corresponding 14- μm -thick tissue sections with the array described above (Fig. 2b). We detected a uniform host and bacterial molecular capture throughout the tissue section (Supplementary Fig. 1a,b), indicating successful tissue permeabilization and RNA hybridization. We identified 512,779 unique bacterial molecules, of which 92.4% corresponded to *Pseudomonas*, indicating the controlled infection system of our lab-grown leaves. The density of *Pseudomonas* 16S rRNA molecules was the highest around the infiltration site (yellow squares in Fig. 2b) and gradually declined toward distal regions, thus providing a more comprehensive picture than the fluorescence imaging, which had missed the spatial component of the infection gradient (Supplementary Figs. 1c and 2a–c). This is confirmed by the positive correlation ($r = 0.21$, $P < 2.2 \times 10^{-16}$; Supplementary Fig. 1d) between the *Pseudomonas* array signal and the *Pseudomonas* fluorescence signal, which indicates higher sensitivity of the array in capturing the decreasing gradient of microbial content from the infection site that could not be recognized based on the fluorescence signal (Fig. 2b and Supplementary Figs. 1d and 2d). A similar pattern was seen in another leaf replicate (Supplementary Fig. 2f,g,i).

We next investigated the expression patterns of host genes in relation to *Pseudomonas* localization. Following a machine learning-based analysis (Boruta³⁰), we found the expression of *pathogenesis-related gene 1* (*PR1*, typical marker of the plant immune response³⁰) as the most associated with *Pseudomonas* localization in both replicates (Supplementary Table 1). As expected, the *PR1* spatial gene expression pattern closely matched the distribution of SmT-derived *Pseudomonas* signal (Fig. 2b and Supplementary Fig. 2c,h), significantly correlated with the *Pseudomonas* signal detected by the array ($r = 0.52$, $P < 2.2 \times 10^{-16}$; Supplementary Fig. 1e) and was nearly fully contained within the region where *Pseudomonas* was detected by the array (Supplementary Fig. 1f). Finally, we found spatial colocalization of fluorescent microscopy-derived *Pseudomonas* signal with the SmT-derived

Pseudomonas signal and host *PR1* immune gene expression (Supplementary Fig. 2e,j). These spatial patterns, which are obvious from visual inspection, were validated by statistical hotspot analysis, in which only significant spatial heterogeneities are considered (Supplementary Fig. 3). Taken together, these results show that we are able to simultaneously capture bacterial taxonomic information and host transcripts.

Simultaneous detection of microbial and host spatial data

Having demonstrated that bacterial information can be specifically captured together with information on host gene expression, we aimed to add a third modality to our arrays, capturing information from eukaryotic microorganisms, specifically fungi. We designed 18S rRNA/ITS probes specific for fungi and tested their performance in both separate arrays and a multimodal array. For this purpose, we created arrays with 100% poly-d(T) probes, 100% 16S rRNA probes and 100% 18S rRNA/ITS probes, as well as a multimodal array containing all three probe types (10% poly-d(T), 45% 16S rRNA and 45% 18S rRNA/ITS). We dissected three leaves of outdoor-grown *Arabidopsis* plants into four 14- μm thick longitudinal sections and analyzed consecutive sections from each leaf using the four array types. The multimodal and unimodal arrays greatly enriched the proportion of captured reads for the corresponding taxa when compared to the unspecific poly-d(T) probes (Supplementary Fig. 4). Specifically, at the genus level, the multimodal array enriched bacterial and fungal unique molecules up to -19- and -31-fold, respectively. At the superkingdom level, the 100% 16S rRNA array enriched bacterial-unique molecules up to -47-fold when compared to the 100% poly-d(T) array, and the 100% 18S rRNA/ITS array enriched fungal unique molecules up to -233-fold. As expected, the multimodal array enriched microbial signals to a lesser degree than the 100% 16S rRNA and 100% 18S rRNA/ITS arrays, given the lower concentration of microorganism-specific probes in the multimodal arrays (Supplementary Fig. 4).

Importantly, the bacterial information captured using the multimodal arrays was almost identical to that captured from consecutive tissue sections using 100% 16S rRNA arrays (both qualitatively and quantitatively). The multimodal array captured up to 962 bacterial taxa and 179 fungal taxa at the genus level (Supplementary Table 2), and recapitulated the profile of 100% 16S rRNA arrays independently if full bacterial components ($r = 0.91$ – 0.93 , $P < 0.001$), top 500 bacterial taxa ($r = 0.92$ – 0.93 , $P < 0.001$) or top 20 bacterial taxa ($r = 0.96$ – 0.99 , $P < 0.001$) were considered (Fig. 2c and Supplementary Figs. 5–8). Similarly, the multimodal array recapitulated the profile of 100% 18S rRNA/ITS arrays if full fungal components ($r = 0.71$ – 0.74 , $P < 0.001$) were considered, while the correlations obtained for the top 500 and 20 fungal taxa were on average 0.71 and 0.77 ($P < 0.001$), respectively (Fig. 2c and Supplementary Figs. 9–12).

Bray–Curtis similarity showed that the bacterial profile obtained using the bacterial 16S rRNA array was most similar to that of the multimodal array, while the fungal profile obtained with the multimodal array clustered with that for the eukaryotic 18S rRNA/ITS array (Fig. 2d). Conversely, the bacterial profile obtained with the eukaryotic 18S rRNA/ITS array and the poly-d(T) array differed markedly from that obtained with the bacterial 16S rRNA array, and the fungal profile obtained with the bacterial 16S array and the poly-d(T) array differed markedly from that obtained with the 18S rRNA/ITS array. By downsampling the 100% 16S rRNA and 18S rRNA/ITS arrays, simulating various probe concentrations, we identified that the Shannon diversity index was almost entirely saturated at 45% simulated probe concentration in all samples for both array types, showing that no new information could be captured by increasing the probe microbial concentrations (Supplementary Fig. 13). When a kingdom-specific array was used to analyze a kingdom other than that for which it was designed, it failed to do so (Fig. 2d). We confirmed this result by calculating the Shannon diversity index across leaves, revealing that the multimodal and 100% 16S rRNA arrays captured similar levels of diversity ($H' = 3.62$ – 4.01

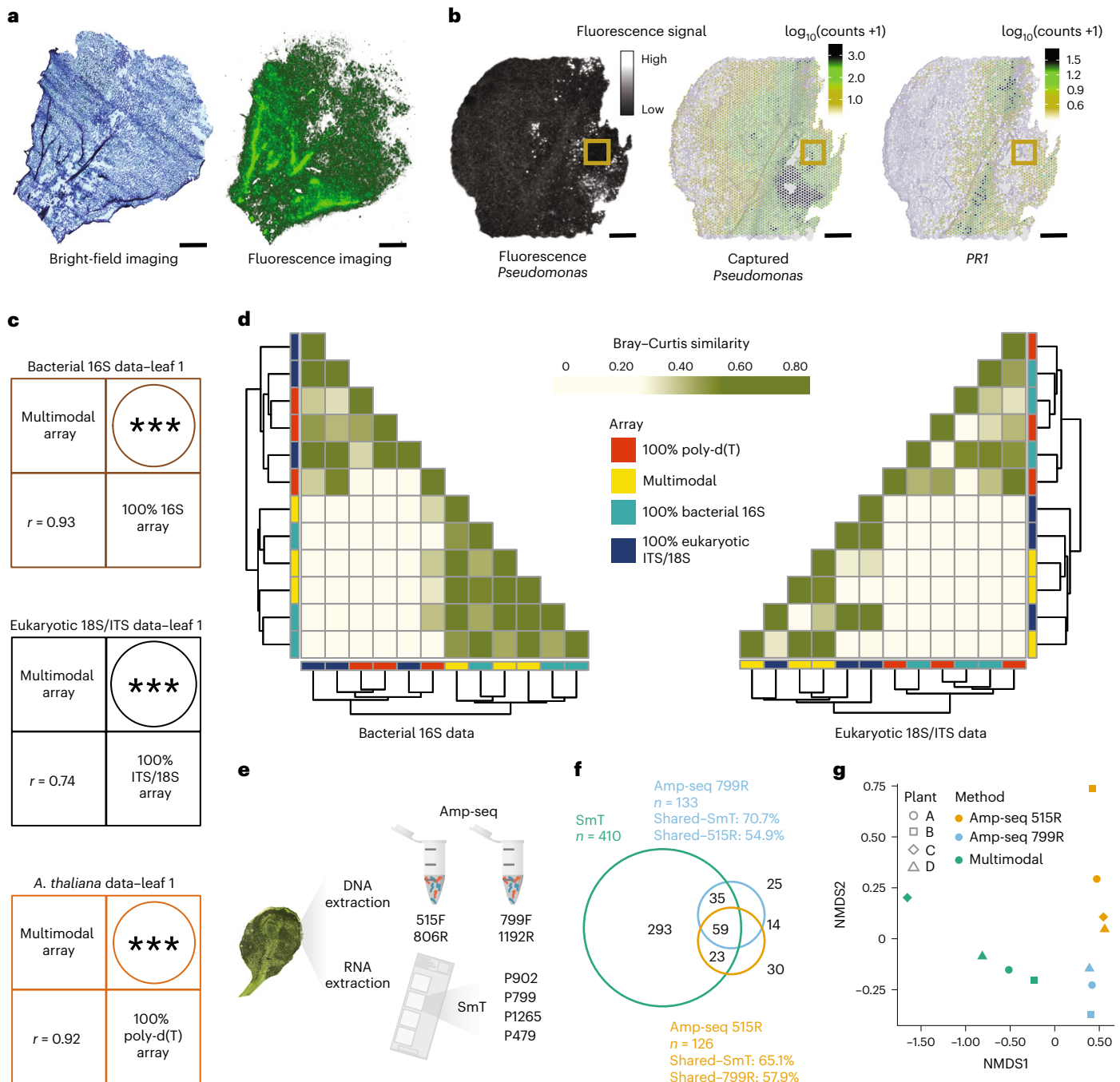


Fig. 2 | SmT resolves the microbial profile and host transcriptome at microscopic resolution. **a**, A Toluidine blue-stained bright-field image of a 14- μm thick longitudinal *A. thaliana* leaf section (left) and the fluorescent cDNA footprint (right) of the same section from the tissue optimization experiment. **b**, Left, fluorescence image of an intact *A. thaliana* leaf syringe-infiltrated (yellow square) with mCherry-tagged Pst DC3000 bacteria. Middle, a 14- μm thick longitudinal section from the same leaf was analyzed using a 50% poly-d(T), 25% P799 and 25% P902 array, revealing the spatial capture of Pst DC3000 16S rRNA molecules. Right, spatial distribution of *PR1* gene expression in the same leaf section. Scale bars: 1 mm. **c**, Pearson correlation coefficient and the

corresponding two-tailed significance test of bacterial 16S rRNA, eukaryotic 18S rRNA/ITS and *A. thaliana* molecules captured in leaf 1 with a multimodal array containing 10% poly-d(T), 45% 16S rRNA and 45% 18S rRNA/ITS probes and with 100% 16S rRNA or 18S rRNA/ITS and poly-d(T) arrays. In all correlations, $P = 0$. **d**, Bray-Curtis similarity for bacterial and fungal taxa captured on different arrays, organized by hierarchical clustering. **e**, Experimental validation of SmT by amplicon sequencing. **f, g**, Numbers of bacterial and archaeal taxa detected using the two methods in a representative sample of four leaves from two plants are compared qualitatively using a Venn diagram (**f**) and quantitatively using NMDS (**g**). NMDS, non-metric multidimensional scaling.

and $H' = 3.81\text{--}4.04$, respectively), different from the 100% 18S rRNA/ITS and 100% poly-d(T) arrays ($H' = 2.76$ and $H' = 3.70$, respectively; Supplementary Fig. 14). Overall, the bacterial profile captured by the 16S rRNA array and the fungal profile captured by the 18S rRNA/ITS array could only be recapitulated by the multimodal array and not

by any of the unspecific probes (Fig. 2d). These results imply that the multimodal array quantitatively enriched microbial counts and accurately profiled microbial populations within tissue sections, unlike the unspecific poly-d(T) probes (Fig. 2d, Supplementary Figs. 4–14 and Supplementary Table 2).

Finally, we confirmed that the multimodal array correctly captured the transcriptomic profile of the host as well (Fig. 2c and Supplementary Figs. 15–17) by comparing the *A. thaliana* gene expression pattern captured with the multimodal array to that obtained with the 100% poly-d(T) array. The multimodal array captured 16,368 *Arabidopsis* genes on average and its correlation with the 100% poly-d(T) array was high ($r = 0.92–0.93$, $P < 0.001$). Overall, these results show that multimodal arrays enable accurate simultaneous capture of the host transcriptome, the bacterial profile and the fungal profile.

Validation of the multimodal array with amplicon sequencing

As each of the two 16S rRNA probes captures a slightly different bacterial community, we introduced two additional 16S rRNA probes, P479 and P1265 (Supplementary Figs. 18 and 19 and Fig. 1), thus improving the ability of the multimodal array to capture the bacterial taxonomic range. We compared the results of this multimodal array with those from 16S rDNA amplicon sequencing (amp-seq)—current gold standard for bacterial profiling. Amp-seq involves PCR amplification of crude DNA extracts using a primer pair. Conversely, our multimodal array captures RNA fragments that are targeted by individual probes. To be able to directly compare the multimodal array to amp-seq—which is conducted on crude extracts—we sampled four leaves from field-grown *A. thaliana* plants and simultaneously extracted their RNA and DNA. We then analyzed the crude RNA extracts with the multimodal array containing the additional P479 and P1265 probes and used the extracted DNA for amp-seq of two 16S loci with V3–V4 (primers 515F + 806R) and V4–V6 (primers 799F + 1192R; Fig. 2e).

We first qualitatively compared the bacterial profiles obtained using the multimodal array to those obtained with the two single pairs of 16S rDNA amp-seq primers by analyzing the presence or absence of every genus found by at least one of the three processes (Fig. 2f and Supplementary Fig. 20). SmT detected more than three times the total number of bacterial taxa detected by the two amp-seq primer pairs (Fig. 2f), including ~71% of the taxa detected by the amp-seq V4–V6 primers and ~65% of the taxa detected by the amp-seq V3–V4 primers. The two amp-seq primer pairs overlapped in ~56% of detected taxa.

We obtained similar results for the other three biological replicates (Supplementary Fig. 20), and a similar trend in quantitative analyses, comparing the Bray–Curtis distances, based on relative abundances (Fig. 2g). Furthermore, pairwise Spearman correlations calculated on bacterial profiles of genera shared across each pair of possible comparisons between the three profiles (Supplementary Fig. 21) showed that SmT delivers an accurate quantitative microbial profile, comparable to amp-seq. In summary, these results confirm that our multimodal array accurately profiles bacteria in *A. thaliana* leaves and captures a more diverse taxonomic range than standard amplicon sequencing.

Microbial hotspots in the leaf govern microbial interactions

The spatial distribution of the members of natural microbial communities within host leaves has been largely unknown. Therefore, we used SmT to investigate the microbial profiles of different leaf sections in outdoor-grown *A. thaliana* leaves. The microbial profiles of the different sections were similar, reflecting the similarity of the environments in which the source plants were grown (Methods) and the reproducibility of our method (Fig. 3a). We ensured that this similarity is not driven by any environmental contaminants by quantitatively comparing the observed microbial profiles with those of axenically-grown leaves. Despite the axenically-grown leaves presented microorganisms that probably survived the seed surface sterilization (for example, sporulating microorganisms; Methods), we found that both the bacterial and fungal profiles of outdoor- and axenically-grown leaves largely differ (Supplementary Figs. 22–24). In fact, 42% of the axenically-grown leaf microbial relative abundance alone was characterized by one bacterial genus, that is *Paenibacillus* (highly resistant spore-forming bacteria³¹), while the same bacterial genus had an average relative abundance

of only 0.035% in the outdoor-grown leaves. Among outdoor-grown leaves, considering only taxa with relative abundances above 1%, we identified 29 bacterial taxa and 23 fungal taxa at the genus level (Supplementary Tables 3 and 4). The relative abundances of different microorganisms did not vary greatly across sections, leaves or whole plants. Analysis of the overall spatial distributions of bacterial and fungal genera (Fig. 3b and Supplementary Fig. 25) revealed that microorganisms were present across almost the entire leaf surface—unique bacterial molecules were detected in 99.9% of sampling spots at an average density of ~277 molecules per million reads, while unique fungal molecules were detected in 97.5% of sampling spots at an average density of ~261 molecules per million reads (Supplementary Table 5 and Supplementary Fig. 25). We validated that this pattern is not a technical artifact of lateral diffusion by comparing the reads under and outside the tissue, finding that for both the microbial and host profiles, the vast majority of reads was derived from under-the-tissue, while also showing a different microbial profile than outside-the-tissue (Supplementary Figs. 26–28).

We next analyzed the geography of microbial colonization. Although we detected both bacteria and fungi across the entire leaf surface, they were concentrated in hotspots rather than being homogeneously distributed (Fig. 3c and Supplementary Fig. 29). Some leaf regions, in 100% of the outdoor-grown leaf sections analyzed, were highly colonized with microorganisms, while others were uncolonized or colonized at very low levels. This complex spatial pattern, instead, could not be observed in sections of axenically-grown leaves where less than half of the tissue sections presented a few small highly delimited hotspots (Supplementary Fig. 30). Moreover, these were almost completely related to only one bacteria, that is *Paenibacillus*³¹ (93% and 83% of the hotspot microbial composition in leaf batches 1 and 2, respectively), in contrast to the mixed and diverse hotspot microbial composition found in outdoor-grown leaves (Supplementary Figs. 31 and 32).

Further investigation of outdoor-grown leaves revealed that some hotspots were shared between bacteria and fungi (Fig. 3d and Supplementary Fig. 29). The relative abundance of shared and unique hotspots varied widely across the 13 leaf sections (Fig. 3d). Because microbial interactions are constrained by physical proximity³², we hypothesized that the relative abundance of shared and unique hotspots controls the proportion of interkingdom and intrakingdom interactions. To test this, we computed the interaction network of the 50 most abundant bacterial and fungal taxa using an algorithm that accounts for the spatial structure of our data (Methods). We exemplify this approach by focusing on a subnetwork of 14 taxa (12 bacterial and 2 fungal), which are strongly associated (average pairwise Spearman's rank correlation coefficient (SRCC) ≥ 0.35) in all tested leaf sections (Supplementary Fig. 33). We then tested the association between the relative abundance of shared hotspots and the magnitude of interkingdom (bacteria–fungi) interactions across the leaf sections, revealing a positive correlation between the two features (SRCC = 0.72, $P = 0.0058$; Fig. 3e). This implies that microbial interactions are driven by their spatial organization, and specifically by their presence in shared hotspots. We found a similar association for the magnitude of bacteria–bacteria interactions and the fraction of bacterial-unique hotspots (SRCC = 0.72, $P = 0.059$; Supplementary Fig. 34), but lower for fungi–fungi interactions and the fraction of fungal-unique hotspots (SRCC = 0.47, $P = 0.1$; Supplementary Fig. 35).

Together, these results demonstrate a considerable spatial organization of microorganisms within the leaf.

Microbial hotspots and host gene expression associations

Because microorganism–microorganism interactions are driven by spatial relatedness, we hypothesized that microbial organization might also drive host–microorganism interactions. We therefore investigated the effects of microbial hotspots on the host transcriptome by reducing

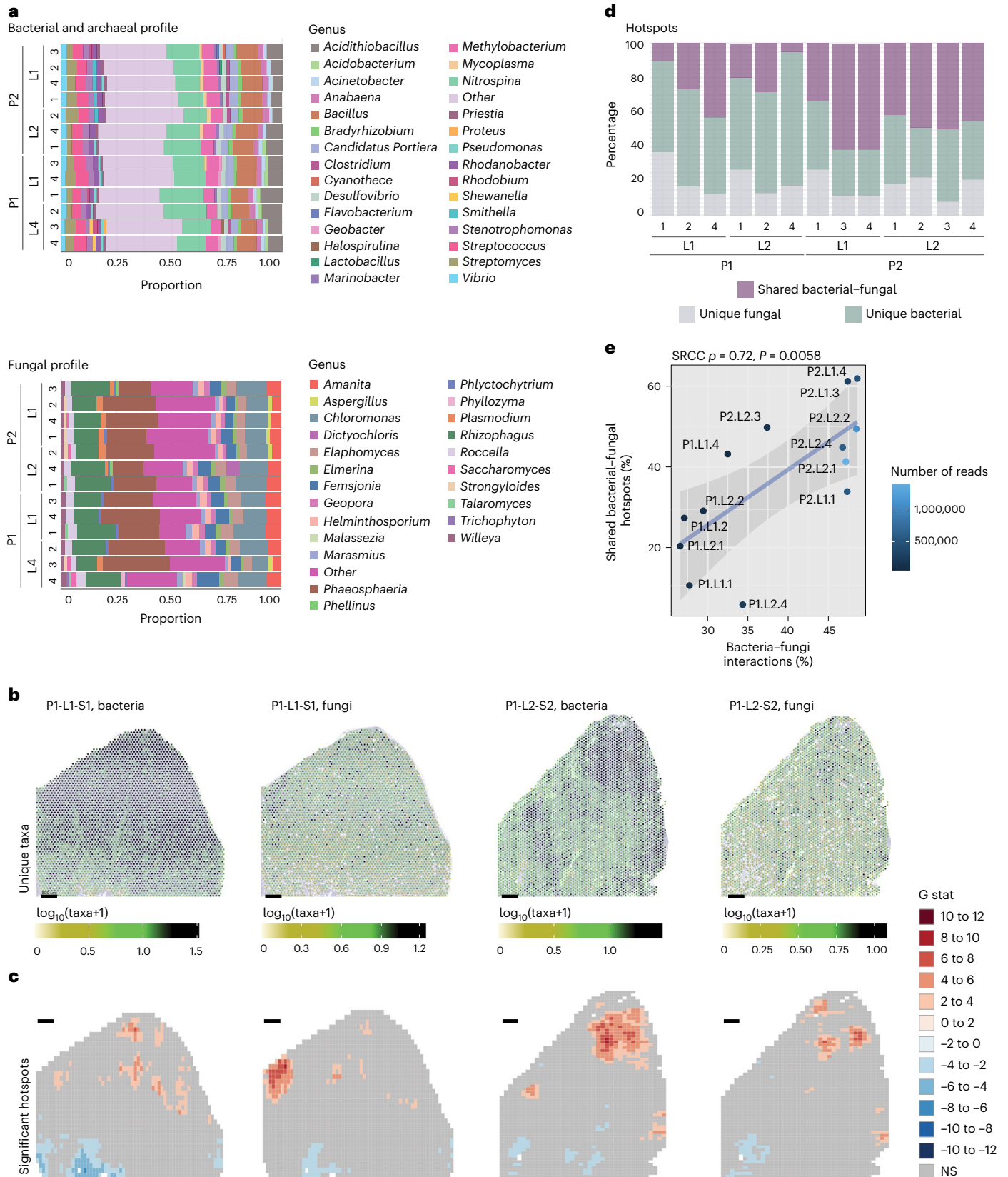


Fig. 3 | Microbial interactions are driven by spatial organization. **a**, Bacterial and fungal profiles for each of the sections of four leaves (‘L’) from two plants (‘P’). ‘Other’ denotes binned bacterial and archaeal genera and fungal genera having $\leq 1\%$ abundance. **b**, Numbers of unique microbial taxa per capture spot. **c**, Significant hot- and cold-spots for bacteria and fungi in a representative leaf section. NS, not significant. Scale bars: 500 μ m. **d**, Percentages of shared and

unique hotspots among bacteria and fungi across 13 different leaf sections. Of note is the variance in shared regions across the sections. The sections were taken from two leaves from two different plants (four leaves in total); ‘P’ denotes plant and ‘L’ denotes leaf. **e**, The proportion of interkingdom (bacteria–fungi) interactions as a function of the proportion of shared interkingdom hotspots and the number of reads. SRCC $\rho = 0.72, P = 0.0058$ (two-tailed test).

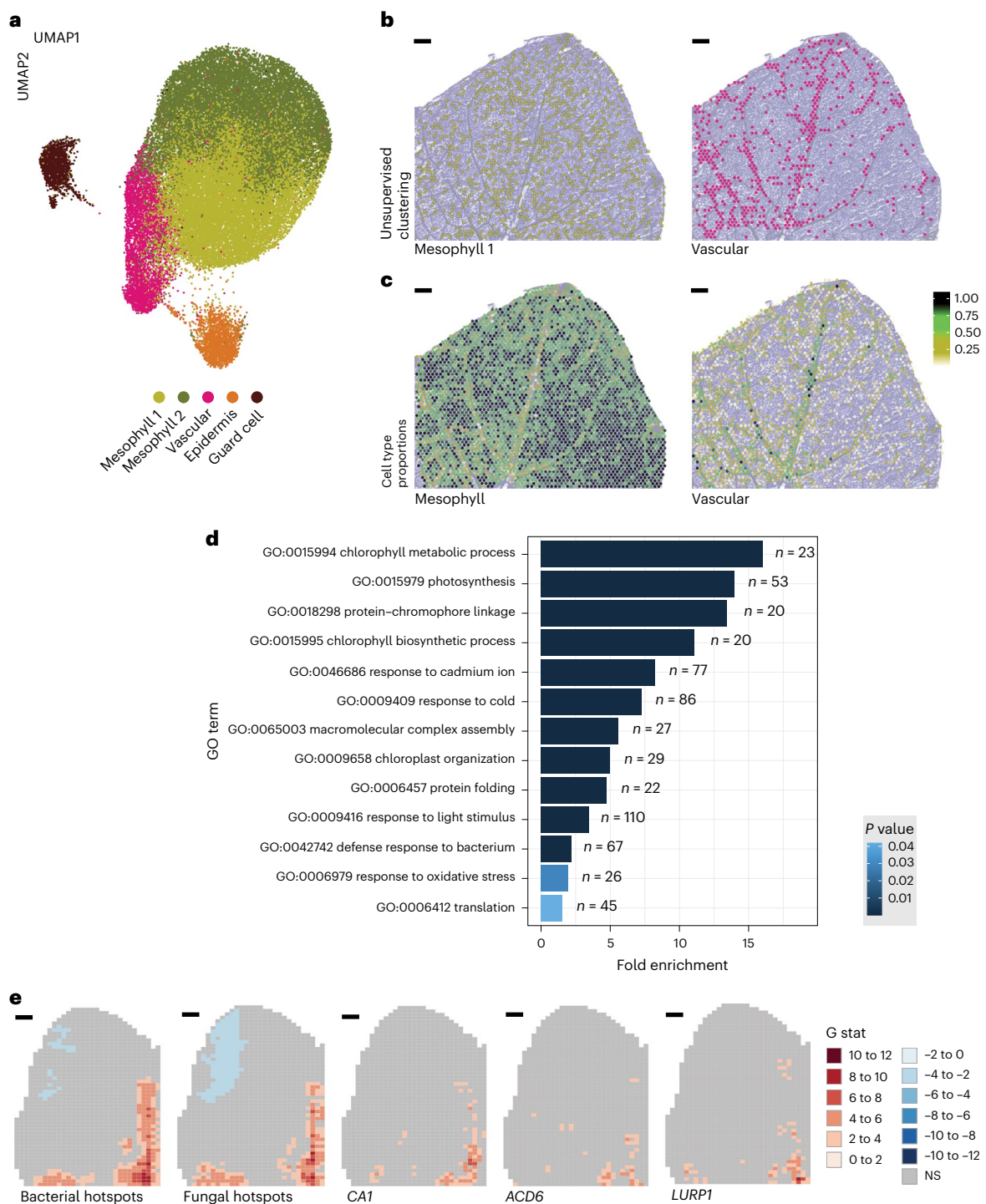


Fig. 4 | Host response is associated with microbial colonization pattern.

a, UMAP clustering of host gene expression. **b**, Projection of UMAP clusters mesophyll 1 and vascular on a representative leaf section. **c**, Mesophyll and vascular cell-type proportions projected on a representative leaf tissue section. **d**, Overrepresented GO terms for microbial-associated genes (n denotes the

number of genes labeled with the indicated GO term). **e**, Spatial distributions of significant bacterial and fungal hotspots together with hotspots for the expression of the defense-related genes *CA1* (AT3G01500), *LURP1* (AT2G14560) and *ACD6* (AT4G14400). NS, not significant. Scale bars: 500 μ m.

the host expression patterns into five cell clusters using uniform manifold approximation and projection (UMAP³³; Fig. 4a and Supplementary Figs. 36 and 37). As expected, the clustered spots reflected the leaf's tissue structure, in which different cell types are distributed fairly evenly with the exception of vascular tissue (Fig. 4b and Supplementary Fig. 37). The close proximity of clusters 1 and 2 in UMAP indicates that these cells have similarities in their gene expression

patterns as confirmed by the spot deconvolution analysis, which identified most of the spots populated by mesophyll cell types (Fig. 4c, Supplementary Figs. 38–40 and Supplementary Table 6). For example, *chlorophyll a/b binding protein 3* (*CAB3*), a common marker gene for mesophyll cells³⁴, is upregulated in cluster 2 (avg. \log_2 (fold change(FC)) = 0.33; Supplementary Fig. 41). Instead, cluster 3 is populated by both mesophyll and vascular cell types, as its spatial location

clearly suggests (Fig. 4b), and in agreement with the spatial expression of the gene *glutathione s-transferase phi 9* (*GSTF9*; Supplementary Fig. 41). Clusters 4 and 5, in addition to mesophyll cell types, presented epidermal cell types (Supplementary Fig. 40), while cluster 5 contained the putative guard cell-type-marker gene *AT2G31141* (ref. 35; $\text{avg. log}_2(\text{FC}) = 1.05$; Supplementary Fig. 41).

Overall, these results show that our system accurately resolves spatial host expression profiles in leaves. However, gene annotation analysis revealed no strong association between any of the five clusters and microbial colonization, and there was no obvious visual overlap between the clusters and the microbial hotspots (Supplementary Figs. 29 and 37).

To further investigate the host response to microbial hotspots, we first tested what fraction of expression hotspots overlapped with the microbial hotspots. We found that it highly varies across leaf sections, ranging from 4.6 to 75% shared expression-microbial hotspots (Supplementary Fig. 42). Next, we performed a machine learning-based analysis ('Boruta') to associate the host's spatial gene expression pattern with bacterial and fungal abundance (Methods). This revealed 1,323 and 954 host genes that were significantly associated with bacteria and fungi, respectively (Supplementary Table 7). To test how general our results are, we asked how often genes were associated with microbial abundance in at least two sections of the same leaf (Supplementary Fig. 43). While moving from one section per leaf to two sections per leaf substantially reduced the number of genes significantly associated with microbial abundance, the size of this gene set was more moderately reduced when requiring that a gene was significant in three sections per leaf (Supplementary Table 8). This behavior implies that the chosen cutoff enriches real biological signal. This conservative approach reduced the number of genes associated with bacteria and fungi to 645 and 442, respectively, thus filtering singular hits. The vast majority of these (63% of 667 genes in total) was associated with both kingdoms, indicating involvement in a general microbial response by the host, rather than a kingdom-specific one (Supplementary Fig. 44). A gene ontology (GO) analysis revealed enrichment of biological process terms associated with plant immune responses, including [GO:0042742](#)—'defense response to bacterium' and [GO:0006979](#)—'response to oxidative stress' (Fig. 4d, Supplementary Fig. 45 and Supplementary Table 9). In total, 73 (11%) of the associated genes had GO terms associated with defense responses to bacteria and/or fungi (Supplementary Table 10). The spatial correlation between gene expression and microbial abundance is well illustrated by the expression patterns of the following three genes: *ACD6*, *CAI* and *LURP1* (Fig. 4e and Supplementary Fig. 46). All three genes are related to basal plant immunity—*ACD6* is broad-spectrum disease resistance gene activated by diverse microorganisms³⁶, the *CAI* gene product binds the immune-related hormone salicylic acid³⁷ and regulates stomatal opening during pathogen invasions³⁸ and *LURP1* is required for resistance to the pathogen *Hyaloperonospora parasitica*³⁹. Overall, these results reveal a connection between the spatial organization of microorganisms within the leaf and the host expression signature.

Discussion

We present SmT, a multimodal untargeted sequencing method to investigate host–microorganism–microorganism interactions in tissue sections at a resolution of 55 μm . Numerous spatially resolved transcriptomics methods have been introduced so far⁴⁰ based on either targeted^{3,41} or untargeted^{1,4,42} capture of the transcriptional information and characterized by different spatial resolutions ranging from subcellular^{43–45} to multiple cells^{1,46}. These methods have been applied to a wide range of tissues from humans^{47–49} to plants^{2,45,50,51}. Recently, methods have been developed that are capable of detecting multiple modalities such as protein and transcriptional information^{41,46,52–54} or chromatin accessibility and transcriptional information⁵⁵. In addition, ref. 56 presented spatial capture of bacterial information in human

cancer tissue. Our SmT approach extends these recent efforts by capturing information not only from the host and its colonizing prokaryotic microorganisms but also from its colonizing eukaryotic ones, thus achieving to retrieve spatial information from three different coexisting organisms simultaneously by using a diverse set of probes specific for polyadenylated transcripts, 16S rRNA and 18 rRNA/ITS regions.

SmT captures fungal, bacterial and host signals from a tissue section while preserving their spatial structure and thus enabling integrated network analysis of gene expression by the host and its microbiota. Recent advances in smFISH techniques for microbiome analysis support the spatially resolved capture of over 1,000 bacterial taxa at the single-cell level⁴⁶ or the detection of bacterial metabolic activities¹⁵. However, smFISH is a laborious technique and requires the design of highly sensitive probes to capture a sample's full bacterial diversity. Extending it to simultaneous detection of host whole-transcriptome information and potentially another microbial kingdom will likely be very challenging. SmT provides a straightforward approach, by sequencing the 16S rRNA and 18S rRNA/ITS variable regions together with polyadenylated transcripts. Our validation of SmT with amplicon sequencing, the gold standard method for bacterial profiling, revealed that SmT can more sensitively capture bacterial diversity than amplicon sequencing. This improvement is probably related to the usage of four individual probes simultaneously, compared with a set of two primers, providing a more diverse set of captured molecules. Similarly to our results, 16S amplicon sequencing was unable to detect many rare bacterial taxa in soil samples, probably due to primers bias⁵⁷. Nevertheless, like any other emerging technologies, SmT presents limitations. The higher sensitivity of SmT comes with an increased risk of capturing signals from environmental contamination. As we have shown, this risk can be mitigated by exploiting the spatial information associated with each read. Specifically, by focusing on hotspots, contrasting the profiles under and outside the tissue and comparing different sections of the same sample, we were able to highlight fundamental differences between plants from different environments. A further limitation of the current implementation of SmT is that it does not yet achieve single-cell resolution. However, at least for the host, spot deconvolution allowed us to resolve the cell-type composition of spots.

We showcased SmT on *A. thaliana* leaves, which are an important model system for phyllosphere microbiology. We found microbial hotspots within plant leaves, reminiscent of microbial microniches in the human mouth^{56,58}. An important question for future research will be whether there are specific leaf locations that favor a specific spatial organization of microorganisms within the leaf. We hypothesize that the invasion point at which the epiphytes entered the leaf is one factor governing the location of hotspots³⁹, while the boundaries of hotspots may be set by the host response or simple ecological factors such as a local lack of nutrients in specific microenvironments⁶⁰. An ecologically important aspect of microbial hotspots is that interactions are the strongest between microorganisms in close physical proximity^{32,61}. New knowledge of interkingdom microbial interactions will be particularly valuable, given that interkingdom interactions can be associated with plant health^{9,24}.

As for microbial interactions, studies of plant responses to microbial colonization have mainly been limited to analyses of homogenized whole tissues^{26,27,62}. SmT now allows us to link microbial abundance at the micrometer level to host transcriptional responses. We found a high degree of overlap between the sets of genes associated with bacteria and fungi, implying a general response of leaf cells to microorganisms, although this generality could be driven by the extensive colocalization of bacteria and fungi in the sampled leaves. Furthermore, it may relate to the quantitative rather than qualitative difference in plant gene expression profiles to a diverse set of microorganisms⁶³. Among the gene functions highly associated with microorganisms, chloroplast-related functions showed the greatest enrichment. This is

consistent with reports linking chloroplasts to plant defense and pathogen invasion as well as photosynthesis⁶⁴. This non-self-host-response profile we describe is less immune-centered than that recently described for the non-self *A. thaliana* response⁶³. This difference is unsurprising given that (1) our study examined outdoor-grown plants instead of plants infected with individual microorganisms in a controlled environment, (2) we profiled host expression at a very late stage of the host–microbiota interaction (after a few months of growth) instead of just 9 d postinfection and (3) we describe the host response at the micrometer scale in different regions of individual leaves rather than the average response among homogenized leaves. Despite these methodological and conceptual differences, both studies revealed some similarities, such as the association between microbial infection and the immune-related gene *GSTF6* (AT1G02930), which was among the 24 general non-self-response genes that were discovered.

In conclusion, the versatility of SmT bodes well for its potential application to the many other tissue types ranging from plants to animals, including humans, where local differences in microbial colonization are an important determinant of health or disease.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41587-023-01979-2>.

References

- Ståhl, P. L. et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).
- Giacomello, S. et al. Spatially resolved transcriptome profiling in model plant species. *Nat. Plants* **3**, 17061 (2017).
- Eng, C.-H. L. et al. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. *Nature* **568**, 235–239 (2019).
- Rodrigues, S. G. et al. Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).
- Ke, R. et al. In situ sequencing for RNA analysis in preserved tissue and cells. *Nat. Methods* **10**, 857–860 (2013).
- Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).
- Tian, Y. et al. Single-cell immunology of SARS-CoV-2 infection. *Nat. Biotechnol.* **40**, 30–41 (2022).
- Sounart, H. et al. Dual spatially resolved transcriptomics for SARS-CoV-2 host-pathogen colocalization studies in humans. Preprint at *bioRxiv* <https://doi.org/10.1101/2022.03.14.484288> (2022).
- Durán, P. et al. Microbial interkingdom interactions in roots promote *Arabidopsis* survival. *Cell* **175**, 973–983 (2018).
- Fan, Y. & Pedersen, O. Gut microbiota in human metabolic health and disease. *Nat. Rev. Microbiol.* **19**, 55–71 (2021).
- Logares, R. et al. Disentangling the mechanisms shaping the surface ocean microbiota. *Microbiome* **8**, 55 (2020).
- Kim, D. et al. Spatial mapping of polymicrobial communities reveals a precise biogeography associated with human dental caries. *Proc. Natl Acad. Sci. USA* **117**, 12375–12386 (2020).
- Mark Welch, J. L., Rossetti, B. J., Rieken, C. W., Dewhirst, F. E. & Borisy, G. G. Biogeography of a human oral microbiome at the micron scale. *Proc. Natl Acad. Sci. USA* **113**, E791–E800 (2016).
- Mark Welch, J. L., Hasegawa, Y., McNulty, N. P., Gordon, J. I. & Borisy, G. G. Spatial organization of a model 15-member human gut microbiota established in gnotobiotic mice. *Proc. Natl Acad. Sci. USA* **114**, E9105–E9114 (2017).
- Dar, D., Dar, N., Cai, L. & Newman, D. K. Spatial transcriptomics of planktonic and sessile bacterial populations at single-cell resolution. *Science* **373**, eabi4882 (2021).
- Shi, H. et al. Highly multiplexed spatial mapping of microbial communities. *Nature* **588**, 676–681 (2020).
- Cao, Z. et al. Spatial profiling of microbial communities by sequential FISH with error-robust encoding. *Nat. Commun.* **14**, 1477 (2023).
- Xia, C., Fan, J., Emanuel, G., Hao, J. & Zhuang, X. Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. *Proc. Natl Acad. Sci. USA* **116**, 19490–19499 (2019).
- Hacquard, S. et al. Microbiota and host nutrition across plant and animal kingdoms. *Cell Host Microbe* **17**, 603–616 (2015).
- Finkel, O. M., Castrillo, G., Herrera Paredes, S., Salas González, I. & Dangl, J. L. Understanding and exploiting plant beneficial microbes. *Curr. Opin. Plant Biol.* **38**, 155–163 (2017).
- Mansfield, J. et al. Top 10 plant pathogenic bacteria in molecular plant pathology. *Mol. Plant Pathol.* **13**, 614–629 (2012).
- Dean, R. et al. The top 10 fungal pathogens in molecular plant pathology. *Mol. Plant Pathol.* **13**, 414–430 (2012).
- Penczykowski, R. M., Laine, A.-L. & Koskella, B. Understanding the ecology and evolution of host-parasite interactions across scales. *Evol. Appl.* **9**, 37–52 (2016).
- Aglar, M. T. et al. Microbial hub taxa link host and abiotic factors to plant microbiome variation. *PLoS Biol.* **14**, e1002352 (2016).
- Shalev, O. et al. Commensal *Pseudomonas* strains facilitate protective response against pathogens in the host plant. *Nat. Ecol. Evol.* **6**, 383–396 (2022).
- Nobori, T. et al. Multidimensional gene regulatory landscape of a bacterial pathogen in plants. *Nat. Plants* **6**, 883–896 (2020).
- Vogel, C., Bodenhausen, N., Gruissem, W. & Vorholt, J. A. The *Arabidopsis* leaf transcriptome reveals distinct but also overlapping responses to colonization by phyllosphere commensals and pathogen infection with impact on plant health. *New Phytol.* **212**, 192–207 (2016).
- Tabula Muris Consortium. A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. *Nature* **583**, 590–595 (2020).
- Giacomello, S. & Lundeberg, J. Preparation of plant tissue to enable spatial transcriptomics profiling using barcoded microarrays. *Nat. Protoc.* **13**, 2425–2446 (2018).
- Kursa, M. B. & Rudnicki, W. R. Feature selection with the Boruta package. *J. Stat. Softw.* **36**, 1–13 (2010).
- Grady, E. N., MacDonald, J., Liu, L., Richman, A. & Yuan, Z.-C. Current knowledge and perspectives of *Paenibacillus*: a review. *Microb. Cell Fact.* **15**, 203 (2016).
- Esser, D. S., Leveau, J. H. J., Meyer, K. M. & Wiegand, K. Spatial scales of interactions among bacteria and between bacteria and the leaf surface. *FEMS Microbiol. Ecol.* **91**, fiu034 (2015).
- McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: uniform manifold approximation and projection. *J. Open Source Softw.* **3**, 861 (2018).
- Berkowitz, O. et al. RNA-seq analysis of laser microdissected *Arabidopsis thaliana* leaf epidermis, mesophyll and vasculature defines tissue-specific transcriptional responses to multiple stress treatments. *Plant J.* **107**, 938–955 (2021).
- Obulareddy, N., Panchal, S. & Melotto, M. Guard cell purification and RNA isolation suitable for high-throughput transcriptional analysis of cell-type responses to biotic stresses. *Mol. Plant. Microbe Interact.* **26**, 844–849 (2013).
- Lu, H., Rate, D. N., Song, J. T. & Greenberg, J. T. ACD6, a novel ankyrin protein, is a regulator and an effector of salicylic acid signaling in the *Arabidopsis* defense response. *Plant Cell* **15**, 2408–2420 (2003).

37. Poque, S. et al. Potyviral gene-silencing suppressor HCPro interacts with salicylic acid (SA)-binding protein 3 to weaken SA-mediated defense responses. *Mol. Plant. Microbe. Interact.* **31**, 86–100 (2018).
38. Zhou, Y. et al. Carbonic anhydrases CA1 and CA4 function in atmospheric CO₂-modulated disease resistance. *Planta* **251**, 75 (2020).
39. Knoth, C. & Eulgem, T. The oomycete response gene LURP1 is required for defense against *Hyaloperonospora parasitica* in *Arabidopsis thaliana*. *Plant J.* **55**, 53–64 (2008).
40. Moses, L. & Pachter, L. Museum of spatial transcriptomics. *Nat. Methods* **19**, 534–546 (2022).
41. Merritt, C. R. et al. Multiplex digital spatial profiling of proteins and RNA in fixed tissue. *Nat. Biotechnol.* **38**, 586–599 (2020).
42. Vickovic, S. et al. High-definition spatial transcriptomics for in situ tissue profiling. *Nat. Methods* **16**, 987–990 (2019).
43. Chen, W.-T. et al. Spatial transcriptomics and in situ sequencing to study Alzheimer's disease. *Cell* **182**, 976–991 (2020).
44. Chen, A. et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* **185**, 1777–1792 (2022).
45. Xia, K. et al. The single-cell stereo-seq reveals region-specific cell subtypes and transcriptome profiling in *Arabidopsis* leaves. *Dev. Cell* **57**, 1299–1310 (2022).
46. Liu, Y. et al. High-spatial-resolution multi-omics sequencing via deterministic barcoding in tissue. *Cell* **183**, 1665–1681 (2020).
47. Asp, M. et al. A spatiotemporal organ-wide gene expression and cell atlas of the developing human heart. *Cell* **179**, 1647–1660 (2019).
48. Hildebrandt, F. et al. Spatial transcriptomics to define transcriptional patterns of zonation and structural components in the mouse liver. *Nat. Commun.* **12**, 7046 (2021).
49. Berglund, E. et al. Automation of spatial transcriptomics library preparation to enable rapid and robust insights into spatial organization of tissues. *BMC Genomics* **21**, 298 (2020).
50. Duncan, S., Olsson, T. S. G., Hartley, M., Dean, C. & Rosa, S. A method for detecting single mRNA molecules in *Arabidopsis thaliana*. *Plant Methods* **12**, 13 (2016).
51. Giacomello, S. A new era for plant science: spatial single-cell transcriptomics. *Curr. Opin. Plant Biol.* **60**, 102041 (2021).
52. Wang, G., Moffitt, J. R. & Zhuang, X. Author correction: multiplexed imaging of high-density libraries of RNAs with MERFISH and expansion microscopy. *Sci Rep.* **8**, 6487 (2018).
53. Liu, Y. et al. High-plex protein and whole transcriptome co-mapping at cellular resolution with spatial CITE-seq. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-023-01676-0> (2023).
54. Ben-Chetrit, N. et al. Integration of whole transcriptome spatial profiling with protein markers. *Nat Biotechnol.* **41**, 788–793 (2023).
55. Deng, Y. et al. Spatial-CUT&Tag: spatially resolved chromatin modification profiling at the cellular level. *Science* **375**, 681–686 (2022).
56. Galeano Niño, J. L. et al. Effect of the intratumoral microbiota on spatial and cellular heterogeneity in cancer. *Nature* **611**, 810–817 (2022).
57. Liu, H. et al. FACS-iChip: a high-efficiency iChip system for microbial 'dark matter' mining. *Mar. Life Sci. Technol.* **3**, 162–168 (2021).
58. Mark Welch, J. L., Ramirez-Puebla, S. T. & Borisy, G. G. Oral microbiome geography: micron-scale habitat and niche. *Cell Host Microbe* **28**, 160–168 (2020).
59. Melotto, M., Underwood, W., Koczan, J., Nomura, K. & He, S. Y. Plant stomata function in innate immunity against bacterial invasion. *Cell* **126**, 969–980 (2006).
60. Geier, B. et al. Spatial metabolomics of in situ host-microbe interactions at the micrometre scale. *Nat. Microbiol.* **5**, 498–510 (2020).
61. Tecon, R., Ebrahimi, A., Kleyer, H., Erev Levi, S. & Or, D. Cell-to-cell bacterial interactions promoted by drier conditions on soil surfaces. *Proc. Natl Acad. Sci. USA* **115**, 9791–9796 (2018).
62. Finkel, O. M. et al. A single bacterial genus maintains root growth in a complex microbiome. *Nature* **587**, 103–108 (2020).
63. Maier, B. A. et al. A general non-self response as part of plant immunity. *Nat. Plants* **7**, 696–705 (2021).
64. Lu, Y. & Yao, J. Chloroplasts at the crossroad of photosynthesis, pathogen infection and plant defense. *Int. J. Mol. Sci.* **19**, 3900 (2018).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

Methods

Bacterial leaf-infiltration assay for microscopy

Seeds of *A. thaliana* (accession Col-0) were surface sterilized by an overnight incubation at -80°C followed by washing with ethanol (5–15 min shaking in a solution of 75% EtOH (Sigma-Aldrich) and 0.5% Triton X-100 (Sigma-Aldrich), followed by a 95% EtOH wash and drying in a laminar flow hood). Stratification was done in a 0.1% agar solution at 4°C for 7 d before planting. Seeds were sown on potting soil (CLT Topferde; www.einheitserde.de), in 60-pot trays (Herkuplast Kubern). During the first 2 d after sowing (the germination period), the trays were covered with a transparent lid to reduce the likelihood of pest infection. Indoor growing conditions were as follows: Cool White Deluxe fluorescent bulbs (25 to $175\ \mu\text{mol m}^{-2}\ \text{s}^{-1}$), 23°C and 65% relative humidity. Plants were grown under long-day conditions (16 h of light) for 15 d before syringe-infiltration with mCherry-tagged Pst DC3000 at $\text{OD}_{600} = 0.001$. Only half of the leaf was infiltrated (in relation to the main vein). A 3xmCherry construct had been inserted at the atn7 site and was a kind gift from Brian Kvitko.

Pst DC3000 was grown overnight in Luria Broth with the appropriate antibiotics (gentamicin and nitrofurantoin, $5\ \mu\text{g ml}^{-1}$ each), then diluted 1:10 on the following morning, and was grown for an additional 4 h to initiate the log phase, after which the bacteria were centrifuged at $3500g$ for 90 s, and resuspended in $10\ \text{mM MgSO}_4$.

Three days after infections, leaves were dissected and placed on $0.5\times$ MS medium with agar (Duchefa, M0255), inspected under a Zeiss Axio Zoom.v16 fluorescence stereomicroscope to verify that the mCherry signal was present, and immediately flash-frozen in liquid N_2 . The leaves were stored at -80°C before cryosectioning.

Imaging of bacterial-infected leaves

Infected *A. thaliana* leaves were imaged on a Zeiss Axio Zoom.v16 fluorescence stereomicroscope, equipped with an LED array for transmitted illumination and an X-Cite XYLIS LED (Excelitas Technologies) for epi-illumination. All leaves were imaged using a PlanNeoFluar Z $1\times/0.25$ dry objective and a Hamamatsu ORCA-Flash4.0 digital CMOS camera (c11440-22C) with 2×2 binning. mCherry-tagged Pst DC3000 was detected using the Zeiss filter set 45 (00000-1114-462), which includes a $560/40\ \text{nm}$ excitation filter, a $630/75\ \text{nm}$ emission filter and a $585\ \text{nm}$ dichroic mirror. Bright-field images were acquired as references for the outline of the leaves for the analysis. The camera exposure time was 220 ms at 5% of light intensity. The images of infected leaves have a pixel size of $18.6\ \mu\text{m}^2$ and were acquired at a $\times 7$ magnification. Image acquisition was done using the ZEN 2.1 software package.

Outdoor-grown plants

For the analysis of microbial hotspots, microbial interactions and host responses to wild microbiomes, seeds of *A. thaliana* (accession Col-0) were germinated and grown indoors for seven short days (8 h of light). On 27 February 2019, the trays were placed outdoors near the Max Planck Institute for Biology Tübingen in a naturalized environment surrounded by other plants. Plants were irrigated weekly with regular tap water. Twenty-seven days after outdoor planting, individual leaves were sampled and immediately flash-frozen in liquid N_2 . Leaves from different plants were stored separately at -80°C before cryosectioning.

Axenically-grown plants

To grow axenic plants, *Arabidopsis* Col(0) seeds were pretreated at 37°C for 24 h, followed by a cold treatment at -20°C for 24 h. The seeds were then rinsed in 70% ethanol for 5 min, 20% chlorine for 20 min, and washed in sterile water three times before being transferred to Murashige Skoog plant agar plates. Subsequently, the seeds were vernalized at 4°C for 48 h before allowing them to germinate and grow into seedlings, still on the MS plates, at 20°C under long daylight conditions (16 h of light and 8 h of darkness). Ten days after the vernalization, individual leaves were sampled and immediately frozen in liquid N_2 .

Leaves from different plants were stored separately at -80°C before cryosectioning.

Two batches ('1' and '2') of axenically-grown leaves were analyzed in the SmT experiments. Leaves were prepared as described in the subsection Sample preparation and sectioning. Three leaf sections from batch 1 and five leaf sections from batch 2 were cryosectioned and attached onto two multimodal array capture areas, respectively. In addition, from the same leaves, four sections per leaf were collected to a Lysing Matrix D tube (MP Biomedicals) for total RNA extraction, which was performed using the RNAqueous-Micro Total RNA Isolation kit (Invitrogen, Thermo Fisher Scientific) using minor modification. Specifically, the leaf sections were disrupted using a Fastprep-24 instrument (MP Biomedicals) in $50\ \mu\text{l}$ of Lysis Buffer at $6.0\ \text{m s}^{-1}$ for 40 s. Subsequently, the homogenized tissue lysate was centrifuged and transferred to the binding column followed by washes with wash buffers according to the manufacturer's protocol. Finally, the total RNA was eluted in $20\ \mu\text{l}$ of elution solution, and $10\ \mu\text{l}$ from each of the two samples was added onto two multimodal array capture areas, respectively, during the cDNA synthesis (instead of tissue sections).

Multimodal array structure

SmT uses multimodal slides (10x Genomics) with capture areas of $6.5\times 6.5\ \text{mm}$. Each capture area comprises 4,992 spots, with diameters of $55\ \mu\text{m}$ each. The spots are covered with capture probes in the following proportion: 45% 16S rRNA probes, 45% 18S rRNA/ITS probes and 10% poly-d(T) probes.

Probe design

Probes were designed using the following two approaches: one based on established primers of the relevant marker genes (P799 (ref. 65) and P902 (ref. 66)) and a de novo approach (P1265 and P479) (Supplementary Fig. 18). On average, the probing sites were 100 nt upstream of the target site. In general, we aimed to maximize the following two variables: the conservation of the probe sites and the variability of the 100 nt downstream target sites. The de novo design process was adopted because previously designed primers were suboptimal with respect to these criteria.

Previously designed primers were used as templates due to their wide usage in the field, which is indicative of useful specificity—it implies that they have a wide taxonomic range and good ability to exclude host reads such as those originating from 16S chloroplast rRNA. Four probes were designed based on the following previous primers; the 16S probes 16S:P799 (5'-TTA VVG CRT GGA CWM CCM GGG TAT CTA ATC CKG TT-3') and 16S:P902 (5'-CSS YTG TGY GSG GSC CCC CGT CAA TTC MTT TGA GTT TYA RYC-3') were based on the mainstream primers 799F⁶⁵ and 902R⁶⁶, respectively. Additionally, the eukaryotic capture probes 18S:P-ITS1 (5'-CCT ACG GAA ACC TTG TTA CGA CTT TTT ACT TCC TCT AAA TGA CCA AG-3') and ITS:P-ITS7 (5'-RRG CGC AAK RTG CGT TCA AAG ATT CGA TGA YTC AC-3') were based on the mainstream primers ITS1F⁶⁷ and ITS7F⁶⁸, respectively. To fit the primers to the annealing conditions of the array, we reversed-complemented all forward-oriented primers (that is, all of them but 902R; the target RNA is single stranded, so reversal of the primer orientation was needed to capture it) and elongated them to obtain 35–45 bp long sequences, as recommended for microbial profiling in microarray systems⁶⁹. To this end, 16S rRNA and ITS custom databases were downloaded (on 29 April 2020) from NCBI GenBank and the sequences downstream of the primer (up to 100 nt, including the primer) were extracted. These sequences were then aligned using the software Clustal Omega (v1.2.4) and the sequence profiles were plotted using weblogo (v3.7.5). The primers were elongated by manual inspection of the resulting weblogo. The length and degeneracy level were limited to obtain fewer than 35,000 unique probe sequences.

In addition to these probes, the following two de novo 16S probes were designed to complement the primer-based probes (as shown in

Supplementary Fig. 18): 16S:P1265 (5'-GGT AAG GTT YYK CGC GTT GCD TCG AAT TAA ACC RCAT-3') and 16S:P479 (5'-TCT CAG THC CAR GTT GGC YBD YCD YCC TCT CARR-3'). To design these probes, representative sequences were selected from the SILVA 16S database (v138.1) using CDHIT (v4.8.1) to the level of 99% sequence identity. Representative sequences were aligned using MAFFT (v7.245), and the sequence profile was plotted using weblogo (v3.7.5). In this process, we targeted highly variable regions with a conserved matching probing site.

Sample preparation and sectioning

The leaves stored at -80°C were immersed in 50% Optimal Cutting Compound (OCT, Sakura) in PBS (Medicago). Embedded samples were frozen in a cryostat (Cryostar NX70, Thermo Fisher Scientific) and sectioned to obtain 14- μm longitudinal sections. Tissue sections were then laid over the multimodal capture areas of the arrays.

Tissue optimization experiment

Tissue permeabilization conditions were identified using a modified variant of a previously reported protocol²⁹. Briefly, after attaching of the tissue section to the slide surface containing 100% poly-d(T) capture probes, the tissue was fixed in methanol (VWR) at -20°C for 40 min and stained with 0.05% Toluidine Blue (Sigma-Aldrich) at room temperature for 2 min. Tissue sections were imaged using a Zeiss Axio-Imager 2T and a Metafer slide scanning system (v. 3.14.2, MetaSystems). They were then permeabilized with pepsin (Sigma-Aldrich) in 0.1% per 0.1 M HCl (Fluka) at 37°C for 30 min. The plant mRNA molecules that had hybridized to the capture probes were reverse transcribed to cDNA using SuperScript III (Invitrogen, ThermoFisher Scientific) and Cy3-dCTP-nucleotides (PerkinElmer) at 42°C overnight. Tissue sections were removed from the slide surface by incubation for 1 h at 37°C in a hydrolytic enzyme mixture consisting of pectate lyase (Megazyme), xyloglucanase (Megazyme), xylanase 10A (Nzytech), β -mannanase 26A (Nzytech) and cellulase (Worthington) in monobasic sodium citrate (Sigma-Aldrich), pH 6.6. They were then incubated with 2% β -mercaptoethanol (Calbiochem) in RLT buffer (Qiagen) and proteinase K (Qiagen) in PKD buffer (Qiagen) for 1 h each. Finally, the fluorescent cDNA footprint was imaged using an Innoscan 910 (Innosys) slide scanning system and Mapix image analysis software (v9.1.0, Innosys) with a pixel size of 5.0 and a gain of 50.

Sequencing library preparation

Sequencing libraries were prepared according to the Visium protocol (10x Genomics) with the following modifications: multimodal slides with leaf sections attached to the capture areas were incubated for 2 min at 37°C followed by a 40-min fixation in methanol (VWR) at -20°C . Capture areas were washed with PBS (Medicago) and incubated for 2 min at 37°C . Tissue sections were stained for 2 min with 0.05% Toluidine Blue (Sigma-Aldrich) at room temperature followed by two washes with ultrapure water and warming at 37°C for 2 min. The slides were mounted with 85% glycerol (Merck) and the bright-field images were acquired with a Zeiss AxioImager 2X microscope and a Metafer slide scanning system (v. 3.14.2, Metasystems) at $\times 20$ magnification. To increase permeabilization efficiency and reduce the effect of secondary metabolites, the slides were incubated in 2% (wt/vol) polyvinylpyrrolidone 40 (PVP-40, Sigma-Aldrich) at room temperature for 10 min. Host plant and eukaryotic microbial cells were permeabilized using the permeabilization enzyme (10x Genomics) at 37°C for 30 min. Bacterial organisms were permeabilized using 10 mg ml^{-1} lysozyme from chicken egg white (Sigma-Aldrich) in 0.05 M EDTA pH 8.0 (Invitrogen) and 0.1 M Tris-HCl, pH 7.0 (Invitrogen) for 30 min at 37°C .

The rest of the SmT workflow followed the procedure described in the Visium Spatial Gene Expression user guide with the following modification: reverse transcription was performed using 2% (wt/vol) PVP-40 instead of nuclease-free water to reduce adverse impacts due to secondary metabolites and cDNA was amplified by performing

12–15 PCR cycles. Libraries were sequenced using the Illumina Nextseq 2000 and Nextseq 1000/2000 P2 or P3 Reagents (200 cycles) kit.

Preprocessing of the reads and bright-field images

Template switch oligo and long poly-A stretches were removed from Read 2 using cutadapt v. 2.9 (ref. 70). The location of the tissue was determined using the Loupe Browser v. 5.1.0 (10x Genomics), in which all the spots containing at least 25% of the tissue were selected and their locations (that is, x and y coordinates) were recorded.

Read alignment

TSO- and poly-A trimmed reads were analyzed using the ST Pipeline⁷¹ (v. 1.7.9, https://github.com/jfnavarro/st_pipeline), which enables simultaneous analysis of the spatial location, unique molecular identifier (UMI) and mRNA molecule. First, the pipeline trimmed poly-N stretches that are longer than 15 bp. Read 2 was then mapped against the *A. thaliana* TAIR10 genome release⁷² using the STAR v. 2.7.7a⁷³ mapping tool and annotated with htseq-count 1.0 (ref. 74). The spatial barcode in read 1 was demultiplexed using Taggd (v. 0.3.6)⁷⁵ and the information from read 1 and read 2 was combined. The ST Pipeline then grouped the reads based on the spatial barcode, gene and genomic location. Finally, the unique molecules were identified using a UMI and the counts were compiled into the gene-count matrix.

Taxonomic assignment of microbial reads

Reads were mapped against the *A. thaliana* reference genome using STAR v. 2.7.7a⁷³ and all reads aligning to the genome were discarded, leaving putative microbial reads. Next, read datasets were demultiplexed based on their probe types (that is, 16S rRNA and ITS/18S rRNA). For each probe dataset, the reads were first clustered into representative sequences by the fastx_uniques module of usearch v. 11.0.667 (ref. 76). Next, the representative sequences (query) were searched for the best homolog (hit) in the NCBI NT database (downloaded on January 2021)⁷⁷ using MMseqs2 v. 1f30213 (refs. 78,79). For each query, all of the best hits (that is, those with the highest identical bit score and a taxonomic assignment on the genus level) were selected for further consideration. Next, the taxonomic assignment for a query was set as the lowest common ancestor (LCA) among the best hits as calculated by TaxonKit v. 0.7.2 (ref. 80) using the NCBI Taxonomy database (downloaded on January 2021)⁸¹. For 18S rRNA/ITS probes, reads were further considered if they were classified as Eukaryota but not as unclassified, chloroplast, mitochondria, uncultured, *Streptophyta*, *Chordata* or *Arthropoda* on the genus or the phylum levels. Similarly, for 16S rRNA probes, reads were considered if they were classified as bacteria but not as unclassified, chloroplast, mitochondria or uncultured on the genus level. Finally, reads were further filtered by their UMI, such that for each spatial location, only one representative read with a given UMI was retained. The number of reads considered for each dataset is provided in Supplementary Table 11.

The annotation of the sequences used for taxonomic assignment was assessed to confirm that they originated from the expected locus. On average, 93.7 and 96.8% of the sequences captured by the 16S and ITS probes, respectively, were annotated as 16S and ITS rDNA loci, and most of the rest were annotated as full genomes, which include 16S and ITS rDNA loci (Supplementary Table 12). We further validated the observed microbial profiles by confirming that the reads containing each of the targeted probes fell within the expected range when aligned against the corresponding sequences in the NCBI 'nt' database (Supplementary Fig. 47). This further confirms that the reads originated from the expected targeted region.

Pst DC3000 infection experiment—data processing

Processed, aligned reads were analyzed using STUtility (v. 0.1.0)⁸². To exclude low-quality spots, the *A. thaliana* host data and bacterial-unique molecules were summed together and every spot with fewer than 20 unique molecules was discarded. Each spot containing

less than 10 unique genes/taxa was discarded. The visualized genes and taxa were \log_{10} normalized and projected on a bright-field image of the tissue section with an opacity of 0.75.

The maximum fluorescence intensities for each spot location were performed by manually aligning the fluorescence image and bright-field image. Then Matlab (2022a) was used to identify the centers of the spots and the k -nearest-neighbor algorithm was implemented to identify pixels that are a maximum of 27.5 μm away from the center. Maximum fluorescence values for each of the spots were extracted.

To generate the scatter plots and Pearson correlation, \log_{10} -normalized SmT captured unique *Pseudomonas* molecules were plotted against the \log_{10} -normalized host *PRI* gene expression and the \log_{10} -normalized maximum *Pseudomonas* fluorescence values from the fluorescence imaging using ggplot2 (v. 3.3.5.)⁸³. The Eulerr⁸⁴ package in R was used to generate the Venn-diagrams with a cutoff of 45 and 120 for leaves 1 and 2, respectively, to remove the background fluorescence signal and minimum of 1 unique molecule per spot for SmT captured *Pseudomonas* and *PRI*. Hotspots analysis using the fluorescence signal values with the applied cutoffs, as well as using the SmT reads, was performed as described in the subsection Analysis of microbial hotspots.

Enrichment experiment

Glass slides bearing a multimodal capture array (10% poly-d(T) probes, 45% bacterial 16S rRNA probes and 45% eukaryotic 18S rRNA/ITS probes), a 100% poly-d(T) array, a 100% bacterial 16S rRNA array and a 100% eukaryotic 18S rRNA/ITS array were used. Three leaves were sectioned on each of these capture slides meaning every leaf had a consecutive section on each array type. Sequencing libraries were prepared as per the above protocol and sequenced with Nextseq 2000 (Illumina). The reads were annotated as described above and analyzed using R (v. 4.0.5).

STUtility (v. 0.1.0) was used to read the *A. thaliana* data to an object and sums of gene values were \log_{10} -transformed. Pairwise Pearson correlation coefficients were calculated and visualized with the corrplot package (v. 0.92) function corrplot.mixed using significance levels of 0.001, 0.01 and 0.05, with hierarchical clustering permitted. The scatter plots are visualized using ggplot2 (v. 3.3.5.)⁸³.

For bacterial 16S rRNA and eukaryotic 18S rRNA/ITS data, unique molecules were summed together per taxon, generating a table containing the sum of unique molecules, phylogenetic paths and metadata relating to section identification. Any annotations to phylum *Streptophyta* were removed, after which the data were divided into bacterial and fungal datasets based on their superkingdom. For taxonomic rank plots, the unique molecules for the different taxonomic levels were counted and compared with the 100% poly-d(T) array to calculate the fold change for microbial taxa at each of the taxonomic levels. Pairwise correlations, and unique molecules for each taxonomic rank, were only calculated for classified reads. We performed the analysis three times—first with all taxa and then with only the most highly expressed 500 and 20 taxa. Shannon diversity and Bray–Curtis similarity were calculated using vegan R package (v. 2.5-7)⁸⁵.

Simulation of probe concentration and effect on diversity. Different proportions of reads—ranging from 5 to 95%—were sampled of samples analyzed on a 100% 16S rRNA or 18S/ITS rRNA array to simulate the effect of different probe concentration on the captured microbial diversity (Shannon diversity index). The procedure was repeated 100 times. The distribution of this simulated Shannon diversity is presented together with the diversity observed in the 45% probe concentration multimodal SmT array.

Saturation of the host information was calculated by subsampling the annotated reads to the saturation point (2,000; 3,718; 8,389; 21,085; 55,598; 149,413; 404,428; 1,097,633 and 2,981,957 reads), and unique molecules and genes were counted and plotted against the saturation points.

Validation of SmT with amplicon sequencing

To compare the performance of SmT to that achieved with amplicon sequencing, seeds of *A. thaliana* (accession Est-1) were surface sterilized and stratified at 4 °C for 1 week in a refrigerator, and then sown in plastic trays (Herkuplast) filled with wild soil from the Heuberger Tor experimental site of the University of Tübingen (Germany). The seeds were left outside to germinate in the same field in late September. The plants developed and overwintered without supplemental watering. Additional plants in each pot were thinned in January 2020 with tweezers, and individual plants were sampled at the end of March 2020 before flowering. The sampling protocol involved cutting the mature rosettes with sterile scissors, placing them in sterile 50 ml centrifuge tubes, and vigorously shaking them in sterile water. The water was then dumped and new water was added until the leaves released no further dirt. After washing, plants were immediately flash-frozen in liquid N₂, and subsequently stored at –80 °C prior to nucleic acid extraction. Both DNA and RNA were extracted from each plant. The entire rosette was lysed in a buffer containing 2% β -mercaptoethanol to extract all nucleic acids while preserving RNA. One proportion of the lysate was used for RNA extraction by the phenol/chloroform protocol, while another portion was used to extract DNA following a previously described potassium acetate and SPRI bead protocol⁸⁶. The DNA moiety was used for 16S rDNA amplicon sequencing. The following two sets of primers were used: (1) 515F-806R (V3-V4) in combination with plastid-blocking clamps⁸⁷ and (2) 799F-1192R (V4-V6), which does not amplify chloroplasts and for which the mitochondrial amplicon was removed by gel extraction⁸⁸. The RNA moiety was used for SmT, using the same pipeline as for all other samples with the exception that crude extracts were used in place of leaf samples (so spatial information was not extracted). A total of 300 μg of RNA was used for the array. In total, four plant samples were used for 16S rRNA profiling, comparing two amplicon sequencing primer sets to the SmT array, with the exception of leaf C for which amp-seq 799F-1192R was not performed. The reads obtained by amplicon sequencing were analyzed in the same way as the array reads, excluding the initial mapping to the *A. thaliana* TAIR10 database. For both of the methods, the reads were subsampled to the same sequencing depth. See the ‘Read alignment’ and ‘Taxonomic assignment of microbial reads’ subsections for information about the full pipeline.

Spearman correlation was calculated between all taxonomic profiles at the genus level (each amp-seq-primer-pair-profile with the SmT-profile and with the other primer-pair derived profile). In all comparisons, only taxa that were detected in both profiles were accounted for. The analysis was performed and plotted using the ggpairs function which is part of R GGally package (version 2.1.0)⁸⁹.

Analysis of microbial hotspots

Microbial hotspots (based on 16S rRNA/ITS reads) were identified using the Getis-Ord G^* statistic⁹⁰ as implemented in the localG function of the R spdep package (v. 1.1.11)⁹¹. The calculation was performed using a 2 \times 2 grid applied to the count matrix resulting from the sum of reads belonging to the 50 most abundant genera (separately for 16S rRNA/ITS reads). A similar calculation was done for individual host genes so that the association between microbial and G -values for individual host genes could be done. The p.adjustSP function of the R spdep package was used with the BH-FDR⁹² method to correct the G stats P values while accounting for the number of neighbors of each region. Hotspot spatial maps were plotted using the R tmap package (v. 3.3-2).

Microbial interaction network analysis

Microbial interactions were inferred based on the Spearman rank correlation coefficient (SRCC) values of the reads count associated with each pair of genera. Specifically, for each pair of microbial genera, in each leaf section, SRCC was calculated accounting for all spots of the array (that is, each spot on the array was considered as a ‘sample’ for each genus). We considered pairs of genera to be interacting if their SRCC-corrected

P value (BH-FDR) was below 0.05. Next, to account also for the spatial organization of microorganisms in the array, we computed the SRCC value of each candidate pair based on shuffled abundance matrices. This step, repeated 1,000 times, results in an empirical null distribution of expected SRCC values where the spatial association between paired genera is random. The shuffled count matrix was generated by using the `permatfull` function implemented in the R `vegan` package (v. 2.5.6) while keeping the total number of reads associated with each genus across all samples (spots) constant (that is argument `fixedmar = 'columns'`). Finally, the significance of each candidate pair of genera was calculated by comparing the SRCC value based on the unshuffled count matrix to the empirical null pair distribution⁹³ following a BH-FDR correction. Microbial interactions were considered to be also spatially significant if their corrected-empirical *P* value was below 0.05. The network was created based on these microbial pairwise correlation values using the R `igraph` package (v. 1.2.6) and plotted using the R `ggraph` package (v. 2.0.5).

Host mRNA clustering

For the *A. thaliana* host data, the counts were filtered using `STUtility` (v. 0.1.0)⁸² by removing the low-quality spots and genes containing at least 10 and 30 counts, respectively. In addition, each spot was required to have at least 10 genes and each gene was required to cover at least 20 spots. Chloroplast, mitochondrial, ribosomal and noncoding genes were filtered from the data set because many of them are not polyadenylated and might contain genes captured with 16S rRNA and 18S rRNA/ITS probes. Finally, after the filtering steps, the spots with fewer than 10 genes were removed because they were considered to be of low quality.

Each section was normalized individually using the `Seurat` (v. 4.1.0)⁹⁴ function `sctransform`⁹⁵ to eliminate intrasection batch effects. To reintegrate the sections back together, anchor features were selected and the whole data was scaled based on these features. Principal component analysis (PCA) was performed on this data using identified variable features. Based on the results of the PCA, the intersection batch effects (experiment date, plant and leaf) were removed with `Harmony` (v. 0.1.0)⁹⁶ using a diversity clustering penalty of 4 and PCA dimensions of 1 to 8.

Normalized gene counts were projected onto 2D leaf section images using `UMAP`³³ with the eight first dimensions from `Harmony` and a resolution of 0.22. To identify cluster-specific markers, raw counts were normalized using the `NormalizeData` function with `LogNormalize` as a normalization method and the `FindAllMarkers` function with the parameters of `test.use = 'poisson'` and `logfc.threshold = 0.15`.

Spot cell-type deconvolution

Cell-type proportions in the spatial host data were analyzed using `Stereoscope` (v. 0.3)⁹⁷ with the Single Cell Leaf Atlas data⁹⁸, who kindly provided the raw count data and cell-specific annotations. `Stereoscope` used raw gene-count matrices from single-cell data and raw spatial data from which spots outside the tissues had been removed. The `stereoscope` was run with a `-gpu` setting using batch sizes of 2,048 and epoch sizes of 50,000 for spatial and single-cell dataset and 5,000 most highly expressed genes from the single-cell dataset.

The single-cell data contained 19 clusters, which were reduced to the following five: mesophyll (11 clusters), vascular (4 clusters), epidermis (1 cluster), guard cell (1 cluster) and hydathode (1 cluster). These collapsed as well as the 19 original clusters were projected on tissue using `STUtility` (v. 0.1.0)⁸² and heatmaps for each of the clusters were generated with `pheatmap` (v. 1.0.12)⁹⁹. To aid the visual interpretation, the cell-type proportions were scaled by quantiles using the 95th percentile of the data in each section and cell type.

Host-response analyses

We used the `Boruta` algorithm³⁰ to determine which set of *A. thaliana* genes is important to explain the microbial load on each spot of the array. Briefly, we modeled the relationship between the expression profile of all *A. thaliana* genes— $G_1 \dots G_n$ and M —the sum of

the 50 most abundant bacterial/fungal genera in each spot of the array (M)— $M - G_1 \dots G_n$. We treated the task as a regression problem and used the random forest algorithm¹⁰⁰ to calculate the importance of each gene in the model. Next, we used `Boruta` to assign a significance score for each gene based on its importance for the model's accuracy. For this purpose, we used the R implementation of the `Boruta` package (v. 7.0.0) with 1,000 trees. This procedure was performed for each leaf section, once using the un-normalized read counts and once using the `Getis-Ord G` statistic value, treating each spot as an observation. Overall, a gene was considered further if it was found to be significant by `Boruta` for at least one measure (that is, reads count or `G` statistics), and if its SRCC *P* value (after FDR correction) was below 0.01. GO enrichment analyses were performed with the DAVID web server with the DAVID knowledgebase v2022q1 (refs. 101,102).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Sequencing data have been deposited at NCBI-SRA under the BioProject [PRJNA784452](https://doi.org/10.1038/s41587-023-01979-2). Source data files (bright-field images, alignment matrices, putative microbial reads and annotation files and gene/taxa matrices) for each of the experiments have been deposited to Zenodo (<https://doi.org/10.5281/zenodo.8308137>)¹⁰³.

Code availability

Scripts written for the analyses described in this paper are available on GitHub (<https://github.com/giacomellolab/SpatialMetaTranscriptomics>)¹⁰⁴.

References

- Hanshew, A. S., Mason, C. J., Raffa, K. F. & Currie, C. R. Minimization of chloroplast contamination in 16S rRNA gene pyrosequencing of insect herbivore bacterial communities. *J. Microbiol. Methods* **95**, 149–155 (2013).
- Hodkinson, B. P. & Lutzoni, F. A microbiotic survey of lichen-associated bacteria reveals a new lineage from the Rhizobiales. *Symbiosis* **49**, 163–180 (2009).
- Gardes, M. & Bruns, T. D. ITS primers with enhanced specificity for basidiomycetes—application to the identification of mycorrhizae and rusts. *Mol. Ecol.* **2**, 113–118 (1993).
- Ihrmark, K. et al. New primers to amplify the fungal ITS2 region—evaluation by 454-sequencing of artificial and natural communities. *FEMS Microbiol. Ecol.* **82**, 666–677 (2012).
- Gardner, S. N., Thissen, J. B., McLoughlin, K. S., Slezak, T. & Jaing, C. J. Optimizing SNP microarray probe design for high accuracy microbial genotyping. *J. Microbiol. Methods* **94**, 303–310 (2013).
- Martin, M. `Cutadapt` removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12 (2011).
- Navarro, J. F., Sjöstrand, J., Salmén, F., Lundeberg, J. & Ståhl, P. L. `ST Pipeline`: an automated pipeline for spatial mapping of unique transcripts. *Bioinformatics* **33**, 2591–2593 (2017).
- Berardini, T. Z. et al. The *Arabidopsis* information resource: making and mining the 'gold standard' annotated reference plant genome. *Genesis* **53**, 474–485 (2015).
- Dobin, A. et al. `STAR`: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
- Anders, S., Pyl, P. T. & Huber, W. `HTSeq`—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
- Costea, P. I., Lundeberg, J. & Akan, P. `TagGD`: fast and accurate software for DNA Tag generation and demultiplexing. *PLoS ONE* **8**, e57521 (2013).

76. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
77. Sayers, E. W. et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **49**, D10–D17 (2021).
78. Mirdita, M., Steinegger, M. & Söding, J. MMseqs2 desktop and local web server app for fast, interactive sequence searches. *Bioinformatics* **35**, 2856–2858 (2019).
79. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).
80. Shen, W. & Ren, H. TaxonKit: a practical and efficient NCBI taxonomy toolkit. *J. Genet. Genomics* **48**, 844–850 (2021).
81. Schoch, C. L. et al. NCBI Taxonomy: a comprehensive update on curation, resources and tools. *Database* **2020**, baaa062 (2020).
82. Bergenstråhle, J., Larsson, L. & Lundeberg, J. Seamless integration of image and molecular analysis for spatial transcriptomics workflows. *BMC Genomics* **21**, 482 (2020).
83. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2016).
84. Larsson, J. et al. Area-proportional Euler and Venn diagrams with ellipses, R package version 7.0.0. <https://cran.r-project.org/web/packages/eulerr/eulerr.pdf> (2022).
85. Oksanen, J. et al. vegan: community ecology package, R package version 2.5-7. https://rstudio-pubs-static.s3.amazonaws.com/754046_aa2efba458b54204bbe06d3a0468a4e2.html (2020).
86. Regalado, J. et al. Combining whole-genome shotgun sequencing and rRNA gene amplicon analyses to improve detection of microbe-microbe interaction networks in plant leaves. *ISME J.* **14**, 2116–2130 (2020).
87. Lundberg, D. S., Yourstone, S., Mieczkowski, P., Jones, C. D. & Dangl, J. L. Practical innovations for high-throughput amplicon sequencing. *Nat. Methods* **10**, 999–1002 (2013).
88. Bulgarelli, D. et al. Structure and function of the bacterial root microbiota in wild and domesticated barley. *Cell Host Microbe* **17**, 392–403 (2015).
89. Schloerke, B. et al. GGally: extension to ‘ggplot2’, R package version 2.1.0. <https://cran.r-project.org/web/packages/GGally/index.html> (2021).
90. Ord, J. K. & Getis, A. Local spatial autocorrelation statistics: distributional issues and an application. *Geogr. Anal.* **27**, 286–306 (1995).
91. Bivand, R. S. & Wong, D. W. S. Comparing implementations of global and local indicators of spatial association. *Test* **27**, 716–748 (2018).
92. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **57**, 289–300 (1995).
93. North, B. V., Curtis, D. & Sham, P. C. A note on the calculation of empirical *P* values from Monte Carlo procedures. *Am. J. Hum. Genet.* **71**, 439–441 (2002).
94. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587 (2021).
95. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).
96. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
97. Andersson, A. et al. Single-cell and spatial transcriptomics enables probabilistic inference of cell type topography. *Commun. Biol.* **3**, 565 (2020).
98. Kim, J.-Y. et al. Distinct identities of leaf phloem cells revealed by single cell transcriptomics. *Plant Cell* **33**, 511–530 (2021).
99. Kolde, R. pheatmap: pretty heatmaps, R package version 1.0.12. <https://cran.r-project.org/web/packages/pheatmap/index.html> (2019).
100. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
101. Sherman, B. T. et al. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res.* **50**, W216–W221 (2022).
102. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
103. Saarenpää, S. et al. Spatial metatranscriptomics resolves host-bacteria-fungi interactomes. *Zenodo* <https://doi.org/10.5281/ZENODO.8308137> (2023).
104. Saarenpää, S. et al. giacomellolab/SpatialMetaTranscriptomics. *GitHub* <https://github.com/giacomellolab/SpatialMetaTranscriptomics> (2023).

Acknowledgements

We thank Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) for providing computational infrastructure and the Bio-Optics Facility of the Max Planck Institute for Biology Tübingen for assistance with microscopy. We thank 10x Genomics for providing multimodal arrays. The authors thank J. Sundström and colleagues at the Department of Plant Biology at the Swedish University of Agricultural Sciences (SLU), Uppsala, Sweden, for providing axenically-grown *A. thaliana* leaves, L. Larsson for discussions on data analysis, J. Wu for help with fluorescence image data analysis and Hailey Sounart for help in the lab. S.G. was supported by Formas grant 2017-01066 and VR grant 2020-04864. H.A. was supported by a fellowship from the Alexander von Humboldt Foundation. O.S. was supported by a fellowship from DAAD. D.W. was supported by ERC-SyG PATHOCOM 951444 and the Max Planck Society.

Author contributions

S.G. and D.W. were responsible for the conceptualization of the study. S.G. designed the SmT experiments; S.S. and O.S. contributed to the design of the SmT; O.S. designed the experiments involving Pst DC3000 and collected the samples and H.A. developed the workflow for hotspot and network analysis. S.S. conducted SmT experiments; O.S. inoculated *A. thaliana* leaves with Pst DC3000 and collected wild *Arabidopsis* leaves; V.C. performed leaf fluorescence imaging; D.S.L. generated amp-seq data and S.S., O.S., H.A., D.W. and S.G. interpreted the results. S.S. performed read alignments and unsupervised clustering of the host data; O.S. and S.S. analyzed the enrichment level of the different probe percentages; H.A. and O.S. designed bacterial and fungal capture probes. H.A. performed taxonomical annotation of the sequencing data, calculated microbial hotspots and spatial networks and performed GO and Boruta analyses. S.S., O.S., H.A., D.W. and S.G. validated the results. S.S. and H.A. handled the data curation process. S.S., O.S., H.A., D.W. and S.G. wrote the original draft of the manuscript. All authors contributed to the review and editing of the manuscript. S.S. created the visualizations. Study supervision was jointly carried out by S.G. and D.W. S.G., D.W., S.S. and O.S. managed the project administration. S.G. and D.W. secured the necessary funding for the project.

Funding

Open access funding provided by Royal Institute of Technology

Competing interests

S.G. and S.S. are scientific advisors to 10x Genomics, which holds IP rights to the ST technology. S.G. is an inventor on patent filings relating to this work. S.G. holds 10x Genomics stock options. D.W. holds equity

in Computomics, which advises plant breeders. D.W. also consults for KWS SE, a plant breeder and seed producer with activities throughout the world. All other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41587-023-01979-2>.

Correspondence and requests for materials should be addressed to Stefania Giacomello.

Peer review information *Nature Biotechnology* thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

For the whole leaves infected with bacteria ZEN 2.1 software package was used for the bright field and fluorescent image acquisition. Bright field images for all the sections were acquired using Metafer Slide Scanning system (v. 3.14.2, Metasystems). The fluorescent signal for tissue optimization experiment was acquired using Mapix image analysis software (v. 9.1.0, Innopsys).

Data analysis

The novel probe sequences were designed using the SILVA 16S database (v138.1) using CDHIT (v4.8.1). Representative sequences were aligned using MAFFT (v7.245), and the sequence profile was plotted using weblogo (v3.7.5).

Raw reads were quality trimmed and filtered with cutadapt (v. 2.9) and spot locations were identified with Loupe Browser (v. 5.1.0). Gene expression matrices were generated by ST Pipeline (version 1.7.9) and aligned to Arabidopsis thaliana reference genome (TAIR10) with STAR (v. 2.7.7a). Within ST Pipeline the annotation was performed using htseq-count (v. 1.0) and the demultiplexing with Taggd (v. 0.3.6)

Taxonomic assignment of microbial reads was performed using usearch (v. 11.0.667) and MMseqs2 (version 1f30213) against the NCBI NT database [downloaded January 2021]. This was followed by applying the LCA algorithm part of TaxonKit (version 0.7.2) R package using the NCBI taxonomy database and followed by UMI filtering. The procedure is detailed at the 'Taxonomic assignment of microbial reads' section of the method. The result of this analysis is a spatial matrix with reads count assigned to each Bacterial/Fungi genus.

The probe concentration experiments and the experiments with fluorescently labeled Pseudomonas DC3000 were analyzed with R (version 4.0.5) and STUtility (v. 0.1.0), Seurat (v. 4.1.0, major dependence in ST Pipeline) and ggplot2 (v. 3.3.5). Microbial content was analyzed with vegan (v. 2.5-7). Correlation plots were generated with corrplot (v. 0.92). Matlab (2022a) and k-nearest-neighbor algorithm was used to extract the maximum fluorescent values per spot. R (v. 4.2.2) and Eulerr (v. 7.0.0) package were used to generate the venn-diagrams. Comparison between SmT and amp-seq was analyzed in R (version 4.0.3) using ggpairs function part of GGally package (v. 2.1.0).

Microbial hotspots (based on 16S rRNA/ITS reads) were analyzed in R (version 4.0.3) using the Getis-Ord G statistic as implemented in the localG function of the R spdep package (version 1.1.11). The p.adjustSP function of the R spdep package was used with the BH-FDR method to correct the G stats p-values while accounting for the number of neighbors of each region. Hotspot spatial maps were plotted using the R tmap package (v. 3.3-2).

Microbial interaction network were inferred based on the Spearman rank correlation coefficient (SRCC) corrected p-value (BH-FDR) values of the reads count associated with each pair of genera. The procedure to account also for the spatial organization of microbes in the array is described in the 'Microbial interaction network analysis' section of the methods section. The microbial network was calculated using R package igraph (v. 1.2.6) and visualized with R package ggraph (v. 2.0.5).

Host, bacterial and fungal modalities were visualized in R (version 4.0.5) and STUtility (v. 0.1.0) and Seurat (v. 4.1.0). Harmony (version 0.1.0). For host expression each section was normalized with Seurat (v. 4.1.0) function SCTransform and Harmony (version 0.1.0). Markers were identified using Seurat's commands LogNormalize and FindAllMarkers. Spot cell type deconvolution for the host data was performed with Stereoscope (v. 0.3) and further visualized in R (v. 4.0.5) with STUtility (v. 0.1.0) and pheatmap (v. 1.0.12) packages. Host saturation points were calculated with STPipeline (v. 1.7.9) and visualised in R (v. 4.0.5) using ggplot2 (v. 3.3.5).

Host-response analyses was performed by applying the Boruta algorithm to determine which set of *A. thaliana* genes is important to explain the microbial load on each spot of the array. For this purpose we used the R (version 4.0.3) implementation of the Boruta package (v. 7.0.0) with 1000 trees. Overall, a gene was considered further if it was found to be significant by Boruta and if its SRCC p-value (after FDR correction) was below 0.01. GO enrichment analyses were performed with the DAVID webserver with the DAVID knowledgebase v2022q1. The procedure is detailed at the 'Host-response analyses' section of the method.

All internal scripts and code written to analyze the data and produce the figures of this manuscript are available on Github (<https://github.com/giacomellolab/SpatialMetaTranscriptomics>).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All sequencing data have been deposited at NCBI-SRA under the Bioproject PRJNA784452. Source data files (bright field images, alignment matrices, putative microbial reads and annotation files, and gene/taxa matrices) have been deposited to Zenodo (DOI:). Scripts written for the analyses described in this paper are available on Github (<https://github.com/giacomellolab/SpatialMetaTranscriptomics>).

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender

Population characteristics

Recruitment

Ethics oversight

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

Sample size	<ul style="list-style-type: none"> - The proof-of-concept experiment with fluorescently labelled <i>Pseudomonas</i> DC3000 was conducted on 2 biological replicates. - The validation experiment comparing amp-seq vs SmT was conducted on 4 biological replicates. - The probe concentration experiments were conducted on 3 leaves with 4 sections from each leaf for different probe concentrations. - The axenically-grown leaves derive from experiments from 8 leaves one section from each, profiled in two separate arrays. - The detection of microbial populations and analysis of <i>A. thaliana</i> gene expression patterns were carried out on a total of 42,389 capture spots deriving from 13 tissue sections across 4 leaves of 2 different plants.
Data exclusions	No data was excluded from the analysis.
Replication	<p>Multiple plants and leaves and sections were included to enhance the reproducibility of the method. Multiple consecutive sections from each leaf sample were run with the same experimental conditions, where possible, to enhance the reproducibility of the approach.</p> <ul style="list-style-type: none"> - The proof-of-concept experiment with fluorescently labelled <i>Pseudomonas</i> DC3000 was conducted on two biological replicates on two multimodal capture areas. - The validation experiment comparing amp-seq vs SmT was conducted on four biological replicates on multimodal SmT arrays and compared to one (Leaf C) or two (Leaves A, B, D) primer pairs of amplicon sequencing. - The probe concentration experiments were conducted on three leaves with four sections from each leaf for different probe concentrations all providing expected results. - The axenically-grown leaves derive from experiments from 8 leaves one section from each, profiled in two separate arrays and compared with crude RNA extracts from the same leaf batches on two separate arrays. - The detection of microbial populations and analysis of <i>A. thaliana</i> gene expression patterns were carried out on 13 tissue sections across 4 leaves of 2 different plants. Cell type proportions per spot were calculated to identify similar regions from the tissue sections in an unsupervised manner. <p>The attempts at replication were successful.</p>
Randomization	Each experiment was processed in their own batch since Spatial metaTranscriptomics allows us to analyze all the samples (leaf sections) and the spatial locations (spots) in the same manner, using the same criteria. Comparisons between the methods were done for all the data without exclusions in an unsupervised manner so randomization was not needed.
Blinding	The researchers were not blinded as the results did not include subjective measurements and were performed by individual researchers. However, all analyses were performed independently on each sample (leaf section). Comparisons between the different leaves and plants studied were done only in retrospect.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Included in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Included in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging