Contrasting patterns of microbial dominance in the *Arabidopsis thaliana* phyllosphere

Derek S. Lundberg[1], Roger de Pedro Jové[*1,2], Pratchaya Pramoj Na Ayutthaya[*1], Talia L. Karasov[1,3], Or Shalev Skriptchak[1], Karin Poersch[1,4], Wei Ding[1], Anita Bollman-Giolai[5], Ilja Bezrukov[1], Detlef Weigel[1†]

[1]Department of Molecular Biology, Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany
[2]Present address: Centre for Research in Agricultural Genomics, 08193 Barcelona, Spain
[3]Present address: School of Biological Sciences, University of Utah, 84112 Salt Lake City, USA
[4]Present address: University of Freiburg, 79085 Freiburg im Breisgau, Germany
[5]Department of Evolutionary Biology and Environmental Studies, University of Zürich, 8057 Zürich, Switzerland
[*]These authors contributed equally.
[†]Corresponding author weigel@tue.mpg.de (D.W.)

Abstract

Pseudomonas and *Sphingomonas* are among the most abundant bacterial genera in the phyllosphere of wild *Arabidopsis thaliana.* Relative to *Pseudomonas*, the ecology of *Sphingomonas* and its interaction with plants remains elusive, despite its global ubiquity and known representatives of plant-beneficial strains. We analyzed the genomic features of over 400 endophytic *Sphingomonas* isolates collected from *A. thaliana* and neighboring plants, revealing high intergenomic diversity compared to much more homogenous *Pseudomonas* populations on the same plants. Variation in plasmid complement and additional genomic features suggest high adaptability, and the widespread presence of protein secretion systems hints at frequent biotic interactions. While some of the isolates showed plant-protective properties, this was a rare trait. To begin to understand the bacterial populations at the investigated *A. thaliana* sites and the alternate hosts of these strains when *A. thaliana* becomes limiting, we employed amplicon sequencing and a novel bulk-culturing metagenomics approach on *A. thaliana* and neighboring plants, both in spring when *A. thaliana* was flowering and in late summer before the emergence of the *A. thaliana* winter cohort. Our data reveal that *Sphingomonas* and *Pseudomonas* strains from *A. thaliana* not only survive, but also thrive on other diverse local plant hosts, suggesting that leaf-to-leaf transmission from these biotic reservoirs may be a major source of microbes to the next generation.

Introduction

Most ecosystems, including host-associated microbiomes, are composed of a handful of common species and a wide assortment of rarer species (*1*, *2*). Perhaps unsurprisingly, common species are frequently implicated in direct interactions with the host and other microbes. For example in the human gut, the genera *Bacteroides* or *Prevotella*, which often occupy 20% or more of the entire community (*3*), both modulate the host immune system and are particularly successful competitors of other microbes (*4*). Similarly, on and inside plant leaves, the Proteobacterial genera *Sphingomonas* and *Pseudomonas* are among the most common bacterial taxa, not only on the model plant *Arabidopsis thaliana* (*5–9*), but also on many other species across continents (*10–17*).

Pseudomonas-leaf interactions are well studied (*12*, *18–20*), largely due to interest in important plant pathogens in the genus (*12*). *Sphingomonas*-leaf interactions, however, are not nearly as well-understood at a genetic level or population level, despite their ubiquity on plant leaves and increasing reports of plant beneficial members (*21–23*). *Sphingomonas* derives its name from its membrane-bound sphingolipids (*24*), common structural and signalling molecules in eukaryotes that are found in few bacterial taxa (*3*, *25*). Curiously, these taxa include the previously mentioned *Bacteroides* and *Prevotella* of the gut, whose sphingolipids interact with the mammalian host and even influence host nutrition (*3*, *26*). Whether *Sphingomonas* sphingolipids also interact with plants remains unknown (*25*). *Sphingomonas* also associate with plant roots (*27*) and seeds (*21*) and are common in soil and freshwater (*28*) among other habitats. Some strains can improve plant growth and abiotic stress tolerance in contaminated soils (*29*), and various others can promote plant growth through the production of growth regulators (*30*). *Sphingomonas* strains affect the abundances of other microbes (*5*, *21*, *31*, *32*), and some protect against pathogenic bacteria (*21*, *33*) or fungi (*22*).

To relate leaf-associated *Sphingomonas* colonization and ecology, we used more familiar patterns of *Pseudomonas* colonization as a benchmark. In our previous studies of microbes colonizing local *A. thaliana* in southwest (SW) Germany, *Pseudomonas* varied widely in bacterial load on individual plants, ranging from nearly absent to very high titres (*7*, *9*, *18*). At the genomic level, 1,524 cultured *Pseudomonas* isolates from local *A. thaliana* plants consisted primarily of a group of closely-related *P. viridiflava* (hereafter referred to as *Pv-ATUE5*) that shared at least 99.9% nucleotide identity in their core genomes and the same diagnostic partial 16S rRNA gene sequence (*18*). Despite their similarity to each other, *Pv-ATUE5* strains differed in pathogenicity in a gnotobiotic system, and phylogenetic analysis suggested that subgroups of *Pv-ATUE5* diverged around 300,000 years ago, consistent with complex selective pressures that have not favored conquest by a single isolate (*18*).

In this work, we sought to characterize local *Sphingomonas* populations at the strain level to ask if, like *Pv-ATUE5*, a single lineage rules the local *A. thaliana* bacterial community, and

to determine general genetic features of plant-associated *Sphingomonas*. We also extended our survey onto neighboring plants of various species at the site and asked to what extent *Sphingomonas* and *Pseudomonas* strains common to *A. thaliana* are generalists, with selective pressures shaped by life on multiple host plants. This work reinforces the notion that individual *Sphingmonas* strains likely have broad host ranges and that they have genetic features equipping them for a multitude of biotic interactions, suggesting major roles not only in the assembly of leaf communities but also in the general plant perception of microbes.

Results

Sphingomonas colonizes leaves to similar levels as *Pseudomonas*, but more consistently

A previous dataset of sequenced metagenomes for 176 *A. thaliana* rosette samples (*9*) allowed us to calculate the bacterial load on the leaf as the ratio of bacterial to plant sequence counts in each sample. We used this ratio to scale, for every sample, the relative abundances of amplicon sequence variants (ASVs) of the V4 region of the 16S rRNA gene (hereafter rDNA) that had been prepared from the same DNA extraction, giving bacterial loads for each ASV. We first analyzed these ASVs at the genus level. *Sphingomonas* and *Pseudomonas* were approximately equal in abundance overall, and twice as abundant as the third most common genus (*Pedobacter*). However, *Pseudomonas* loads varied widely between samples, with a coefficient of variation (standard deviation / mean abundance) more than double that of *Sphingomonas*, which was among the most constant genera (Figure 1a). Impressively, the reliability of *Sphingomonas* colonization also applied to different *Sphingomonas* ASVs that independently colonized most plants (Figure 1b), suggesting that strains from each ASVs may occupy a different niche in the leaf rather than competing with each other to fill up a general "*Sphingomonas* niche" on a given plant. We named the most abundant *Sphingomonas* and *Pseudomonas* ASVs, whose average abundance was several fold higher than the next most abundant ASVs in each genus, SphASV1-V4 and PseASV1-V4, respectively.

Major *Sphingomonas* and *Pseudomonas* ASVs are enriched endophytically

We reanalysed a previous dataset of slightly longer rDNA sequences in the V3V4 region (*18*) that partitioned *A. thaliana* leaf epiphytes from endophytes using thorough surface sterilization, and located ASVs containing an exact match to the SphASV1-V4 and PseASV1-V4 described above. We found for each a single match to a highly abundant V3V4 ASV, and we refer to the longer V3V4 rDNA sequences henceforth as SphASV1 and PseASV1. PseASV1 uniquely matched the representative sequence previously used to define the *Pv-ATUE5* lineage (*18*). SphASV1 was more abundant than PseASV1 in both the epiphytic and endophytic fractions of the leaf, and both were well-represented in the endophytic compartment (Figure 1c), suggesting the capability to evade or tolerate the plant immune system and persist inside the leaf. SphASV1 was in fact the most abundant single ASV in the dataset. The combination of

numerical dominance, endophytic enrichment, and high consistency of colonization drove our interests to characterize genetic and ecological features of this ASV and local *Sphingomonas* more generally.

Sphingomonas has a wider host range than *Pseudomonas* in the phyllosphere

An abandoned train station in Eyach, Germany, supports a population of hundreds to thousands of *A. thaliana* plants, and has been a source of material for several recent plant-associated microbiome studies (*5*, *7*, *18*, *34*). We collected *A. thaliana* and neighboring plants in spring, when most *A. thaliana* plants were green, mature, and flowering. We also collected plants in late summer when there were no living *A. thaliana* because all plants of the species had senesced and dried out by mid summer, and the next cohort had not yet germinated. We reasoned that by late summer, the springtime *A. thaliana*-associated microbial community must have either migrated to survive in other plant hosts or non-plant habitats, or died (*12*, *35*, *36*), and thus at this later time point we could rule out that any isolates collected from other species or soil were direct migrants from *A. thaliana* hosts.

For each plant species, we pooled at least one entire leaf from at least 6 independent individuals per sample, collecting 7 independent samples per species. Although this pooling strategy that spanned many plants precluded analysis of variation between individual hosts, it maximized our ability to broadly survey the site (Supplementary Figure 1). After bringing leaf samples back to the lab, we surface-sanitized them in 70% ethanol for 45-60 seconds and macerated them in buffer with a mortar and pestle. We then mixed the fresh lysate with glycerol to make -80°C freezer stocks that cryo-protected the live bacteria (Methods).

We first extracted metagenomic DNA directly from the frozen lysates using a custom bead-beating based protocol as previously described (*9*, *18*), and prepared 16S rDNA amplicon libraries spanning the V3 and V4 regions (Methods) using peptide nucleic acids (PNAs) to reduce organelle amplification (*37*). The residual chloroplast sequences in each sample corresponded to the expected plant host species, confirming that we had correctly identified species for harvest (Supplementary Figure 2). We could not obtain sufficient bacterial reads from dandelion and thistle due to a natural mutation in their chloroplast sequences that made our PNAs ineffective (*38*), and 16S rDNA bacterial abundance data from these species are therefore absent due to an overabundance of organelle sequences. To contrast bacterial communities across the sampled plant hosts, we binned the ASVs into bacterial families, and clustered samples by their pairwise Bray-Curtis dissimilarity (Figure 1d). This revealed substantial similarities in bacterial family membership between groups of samples, with clear plant clustering both by sampling season and also by some plant taxa, consistent with previous publications linking plant genotype and seasonal effects to bacterial community composition (*13*, *39*). The family Pseudomonadaceae, of which 92.9% was the genus *Pseudomonas*, reached high relative abundances in *A. thaliana*, but also in some other hosts—particularly the other Brassicaceae *Draba verna* and *Cardamine hirsuta* (Figure 1e). Sphingomonadaceae, of which

80.7% corresponded to the genus *Sphingomonas*, was ubiquitous and abundant across all plant hosts (Figure 1e, Supplementary Figure 3). As on *A. thaliana*, PseASV1 and SphASV1 accounted for a substantial fraction of reads from *Pseudomonas* and *Sphingomonas* (44.7% and 39.4%, respectively) (Figure 1f, 1g).

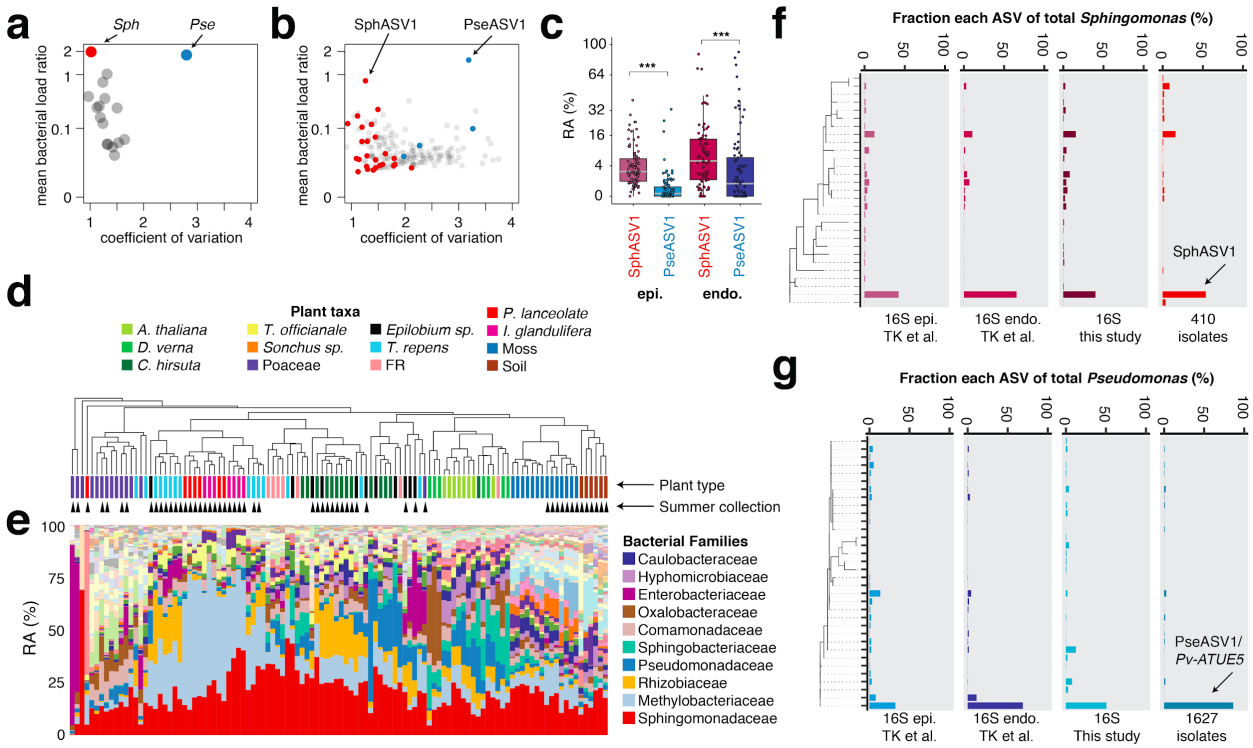

Figure 1. *Sphingomonas* and *Pseudomonas* are abundant and culturable, but differ in consistency of colonization | a) *Sphingomonas* (red, S) has a similarly high bacterial load to *Pseudomonas* (blue, P) across 176 *A. thaliana* plants from (*9*), but among the lowest coefficients of variation (standard deviation / mean abundance) of other major genera (grey). **b)** Similar to **(a)**, but showing the 16S rDNA sequences that make up the most abundant genera. *Sphingomonas* ASVs (including SphASV1) have on average low coefficients of variation, and *Pseudomonas* ASVs have on average higher coefficients of variation. **c)** The median relative abundance (RA) of SphASV1 is higher than PseASV1 both epiphytically (epi.) and endophytically (endo.) (*** $p < 0.001$, FDR-adjusted Mann-Whitney U-test). **d)** Plant samples are clustered by the Bray-Curtis dissimilarity of the relative abundance of bacterial families. Summer collection indicated with black triangles. **e)** Stacked bars represent the relative abundances (RA) of bacterial families in the present study, with the key to the top 10 bacterial families at right. **f)** Relative abundances of *Sphingomonas* ASVs detected epiphytically (epi.) and endophytically (endo.) in TK et al. (*18*), this study, and in 410 locally-cultured isolates from multiple plant species. SphASV1 is indicated. **g)** Same as in **f**, but for *Pseudomonas*, with 1,627 isolates combining 1,524 from *(18)* and 103 additional isolates from this study. PseASV1 is indicated.

Cultured bacterial populations resemble those on wild leaves

Pseudomonas isolates from local *A. thaliana* populations, the majority of which were associated with the PseASV1 16S rDNA sequence, have been previously characterized (*18*). To investigate *Sphingomonas* genome diversity across host species, and likewise to characterize the genomic features associated with highly abundant *Sphingomonas* groups, we cultured endophytic

Sphingomonas from frozen plant lysates using both remaining *A. thaliana* lysates from (*18*) as well as lysates from the diverse plant hosts in the present study. We enriched for *Sphingomonas* using LB supplemented with 100 mg/mL streptomycin, an antibiotic to which most *Sphingomonas* is resistant due to a natural mutation in the *rpsL* gene (*40*), and isolated 410 *Sphingomonas* colonies. Using LB supplemented with 100 mg/mL nitrofurantoin as previously described (*18*), we also isolated an additional 103 Pseudomonads. We generated draft genome assemblies with an average of 65-fold genome coverage and BUSCO (*41*) completeness of at least 85% (Methods). We first extracted from the assemblies the V3 and V4 regions of the 16S rDNA, allowing us to relate the genomes to our existing ASVs. Critically, for both *Pseudomonas* and *Sphingomonas*, we recovered isolates in relative abundances consistent with those from culture-independent surveys (Figure 1f, 1g), demonstrating that culturing did not introduce substantial biases in recovery rates, and that we had sequenced a sufficient number of isolates to capture broad patterns of diversity on leaves.

Sphingomonas 16S rDNA sequence similarity belies high genomic diversity

Previously we observed low genomic diversity among PseASV1/*Pv-ATUE5* strains isolated from *A. thaliana* (*18*). To similarly evaluate genomic diversity for an analogous set of *Sphingomonas*, we selected all the *Sphingomonas* isolates in our collection from *A. thaliana* that also had the SphASV1 16S rDNA sequence (174 SphASV1 out of 340 total *A. thaliana* isolates). We compared the genomes to each other using MASH (*42*), which decomposes genomes into *k*-mers and calculates a distance based on the fraction of shared *k*-mers. As an additional comparator, we also included a representative set of 99 diverse PseASV1/*Pv-ATUE5* genomes isolated from *A. thaliana*, comprising all 82 in a condensed tree of 165 strains from (*18*) that were at least 0.1% different across the core genome, and an additional 17 strains from this study. We converted the MASH distances to a similarity score between 0 (least similar) and 100 (identical) (Methods). When PseASV1/*Pv-ATUE5* genomes were compared to each other, all comparisons had MASH similarities > 95. However, the majority of pairwise comparisons between SphASV1-associated *Sphingomonas* genomes had similarities of 90 or less, indicating that the V3V4 region of the 16S rDNA sequence was a relatively poor predictor of genome similarity in *Sphingomonas* (Figure 2a, 2b). We extracted full-length 16S rDNA sequences from the genomes and clustered them into operational taxonomic units (OTUs) at 99.5% identity, which yielded no additional subgroups for PseASV1/*Pv-ATUE5* strains, but partitioned SphASV1-associated strains into 6 subgroups (Figure 2a). Genomes within these subgroups were more similar, but pairwise MASH similarities within one such subgroup of 46 isolates still averaged less than 95, further confirming that 16S rDNA was less informative for local *Sphingomonas* than for local *Pseudomonas* (Figure 2b).

We initially suspected the relatively high overall genome diversity of SphASV1-associated isolates compared to *Pv-ATUE5* isolates might be due to high variation in the accessory

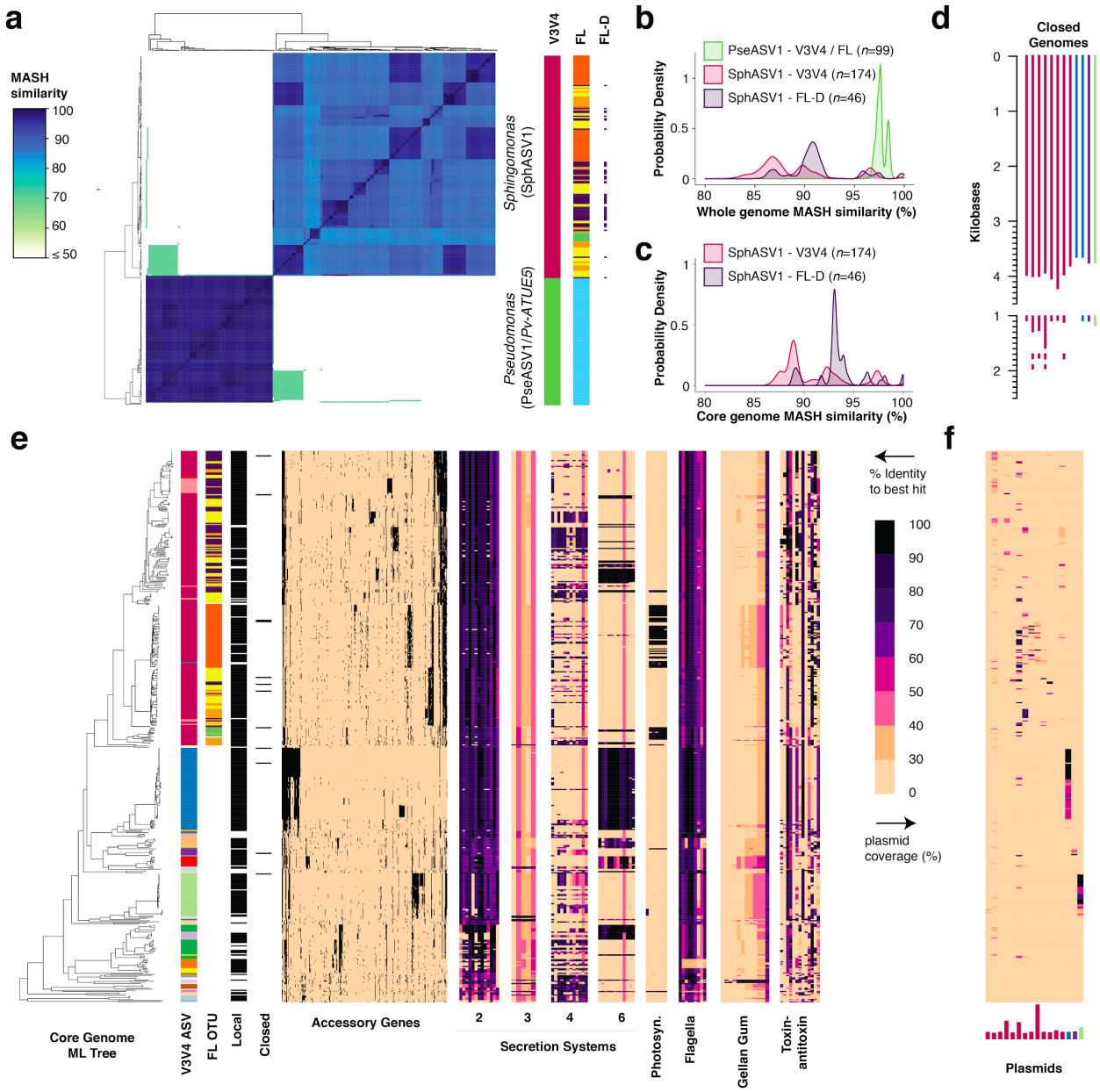

Figure 2. *Sphingomonas* 16S rDNA sequences belie high genomic diversity | **a)** Heatmap of MASH similarity between 99 *Pseudomonas* PseASV1/*Pv-ATUE5* genomes and 174 *Sphingomonas* SphASV1 genomes. Color bars to the right indicate which genomes contain the V3V4 16S rDNA ASV used to define PseASV1 and SphASV1 ('V3V4') and which belong to different full length 16S rDNA OTU ('FL'). A diverse group of genomes sharing a full length OTU is also indicated (FL-D) **b)** Density plot showing the distribution of MASH similarities shown in the heatmap in **(a)**, subset by bacterial genus and 16S rDNA grouping. V3V4, variable regions 3 and 4; FL, a full length OTU; FL-D, the group of highly diverse genomes sharing a full length OTU sequence as indicated in **(a)**. **c)** similar to **(b)** but showing MASH similarity calculated only on an alignment of 274 core genes. **d)** Base map of the 12 closed genomes, colored by their V3V4 16S group, showing plasmids to scale. **e)** Core genome maximum likelihood tree of 410 *Sphingomonas* isolates from this study and 70 from RefSeq based on 274 core genes, with genomic features indicated by aligned heatmaps. 'V3V4', color map of the different V3V4 ASV sequences associated with each genome (white if unknown); 'FL OTU', color map of the 99.5% identity OTUs associated with SphASV1 and and closely-related ASVs; 'Local', 410 genomes sequenced and isolated in this study are indicated by black; 'Accessory Genes', presence absence matrix of all genes not used to compute the ML tree; 'Secretion Systems',

BLAST results of each genome against genes of the type 2, 3, 4, and 6 secretion systems, with percent identities as indicated by the color legend. 'Photosyn', genes involved in anoxygenic photosynthesis; 'Flagella', genes involved in building an operating flagellum; 'Toxin-antitoxin', genes associated with toxin-antitoxin pairs. **f)** 'Plasmids', percent coverage of plasmid sequences from the closed genomes, with the relative size and ASV background of each plasmid indicated by the length and color of the bars.

genome—specifically the differential presence of plasmids. We noted that some publicly-available complete genomes of *Sphingomonas* contained several plasmids, and after closing 12 of our genomes with long-read sequencing (Methods), we detected a total of 16 plasmids with up to three per genome, comprising up to 14% of the total genome size (Figure 2d). We therefore recalculated genome similarities using 274 shared genes identified as the soft core genome of our *Sphingomonas* dataset after annotating open reading frames using Prokka (*43*), and using PanX (*44*) to identify genes present in at least 70% of genomes (Methods). We compared the aligned core genomes of SphASV1-associated isolates again using MASH (Figure 2c). SphASV1-associated core-genomes were more diverse than even whole genomes (which include accessory genes) of *Pv-ATUE5*. Thus, despite the pervasiveness of the SphASV116S rDNA sequence in our dataset, we cannot currently extrapolate what this ASV sequence means at the level of genomic content.

Sphingomonas genomes reveal adaptations for competitive life in the phyllosphere

To explore relatedness of SphASV1-associated genomes, and how they compare with other *Sphingomonas* genomes from our collection, we calculated maximum-likelihood (ML) core genome phylogenies (*44*) from all 410 local *Sphingomonas* isolates (340 from *A. thaliana* and 70 from other local plant hosts), along with 70 sequence-related isolates from NCBI RefSeq (*45*) (Figure 2e). Gene presence or absence in the accessory genome appeared to correspond to divisions in the core genome tree. We compiled a short list of *Sphingomonas* genes that had a high likelihood, based on literature, to improve survival among competing bacteria, or to facilitate interaction with a plant host, and made a custom database of *Sphingomonas* protein sequences. We searched for these features in our genomes by aligning the assemblies using BLASTX (*46*). We considered matches with at least 30% amino acid identity over at least 60% of the protein length as positive hits. While nearly all isolates haddiagnostic genes for the type 2 secretion system, and genes for the type 4 and type 6 secretion systems were common among some clades including those of SphASVI, only a handful of less abundant strains seemed to potentially have a type 3 secretion system (Figure 2e). Flagellar motility is common. A group of SphASV1 genomes has a full suite of genes for anoxygenic photosynthesis, a fascinating feature that can supplement heterotrophic energy production and likely improves survival in well-illuminated and nutrient-poor conditions such as the phyllosphere (*47–50*). All genomes were rife with toxin-antitoxin systems, likely to stabilize plasmids or superintegrons (*51*). Indeed, we found widespread evidence of plasmids in our draft genomes by using minimap2 (*52*) to search for alignments to the 16 circularized plasmid sequences from our 12 complete

genomes (Figure 2d). Considering alignments covering at least 30% of the length of a plasmid as a positive hit, 250 (60%) of our isolates showed signatures of one or more of these 16 plasmids (Figure 2f).

To increase confidence in the preceding gene searches in our draft genomes, we repeated the analysis comparing the 12 closed genomes to their draft genome counterparts. Both draft and complete versions of each genome clustered tightly in a ML tree, and showed essentially the same presence/absence patterns (Supplementary Figure 4), suggesting that draft genomes were sufficient for analysis of gene content at this level of detail.

Some *Sphingomonas* attenuate *Pseudomonas* virulence in *A. thaliana*

Besides reaching similarly high abundances in the same leaves (Figure 1a-b), both *Sphingomonas* and *Pseudomonas* grow on many similar substrates *in vitro (33)*, suggesting potential niche overlap in the phyllosphere. Previous work revealed that certain strains of *Sphingomonas*, in particular *S. melonis* Fr1, can ameliorate symptoms in *A. thaliana* leaves caused by pathogenic *Pseudomonas* and *Xanthomonas* in a gnotobiotic system (*23*, *33*, *53*), and more recently a strain of *S. melonis* was demonstrated to protect rice against seedling blight pathogen *Burkholderia plantarii*. We first sought to screen some of our local isolates for potential plant-protective activity against virulent *Pv-ATUE5* strains in a gnotobiotic 24-well plate agar system, as we used previously (*18*).

We germinated an *A. thaliana* accession endemic to our field site, Ey15-2, on ½ strength MS media in the presence of each of 19 strains of local *Sphingomonas* that represented much of the diversity of our collection, as well as *S. melonis* Fr1. Following 10 days of co-cultivation, we challenged the plants with *Pseudomonas*, either the virulent local *P. viridiflava* strain *Pv-ATUE5*:p25c2 (*18*) or the model pathogen *P. syringae* (*Pst*) DC3000, and monitored plant health over the next 6 days by measuring green pixels (Methods). Generally, *Pst* DC3000 slowed plant growth compared to buffer control, while *Pv-ATUE5*:p25c2 killed plant tissues (Figure 3a-d). However, in both this experiment and in a replicate experiment, our *Sphingomonas* strain *SphATUE*:S139H133 and *S. melonis* Fr1 improved plant health in the presence of *Pv-ATUE5*:p25c2 (Figure 3a-b, Supplementary Figure 5), with infected plants not dying, but instead growing slowly at a rate not different than plants treated with the less deadly *Pst* DC3000 (FDR-adjusted Mann-Whitney U-test, $p > 0.05$).

Surprisingly, seedlings germinated in the presence of *S. melonis* Fr1 in these conditions, although green even after *Pseudomonas* infection, were consistently stunted compared to those germinated in buffer or those germinated with any other *Sphingomonas,* including the protective *SphATUE*:S139H133 (Figure 3a, 3c). We also observed this stunting phenotype in the Col-0 accession (not shown) for which a protective effect of *S. melonis* Fr1 had been previously reported. Although the authors reported that the plant transcriptomic response to protective *S. melonis* Fr1 involved induction of a set of defense genes overlapping those induced by *Pst* DC3000 (*53*), they did not observe negative effects on plant growth due to *S. melonis* (*23*, *53*).

We suspected that our result might be due to differences in our gnotobiotic system. Because gnotobiotic systems are entirely different environments than what plants experience in nature, we also tested plant protection for plants grown on soil. We therefore set up a similar experiment in which 2-week old seedlings grown on potting soil were sprayed first with *Sphingomonas* and 4 days later challenged with *Pseudomonas* sprayed at high concentrations ($O.D._{600}$ = 1.0), using environmental conditions known to produce robust symptoms (*54*). At 5 days post infection (dpi), we observed well-reported *Pst* DC3000 symptoms on most plants including chlorotic leaves, stunted growth, and increased anthocyanin at the apical meristem (*54*). However *Pv-ATUE5*:p25c2, which was consistently deadly on agar plates, did not produce symptoms on soil-grown *A. thaliana* of a similar age (Methods) and had no clear effect on rosette size (Figure 3e). Because some disease resistance responses break down at very high humidity (*55*), we reasoned that we might not see *Pv-ATUE5*:p25c2 symptoms on soil due to natural resistance mechanisms that are suppressed in the ~100% relative humidity of the gnotobiotic system. We therefore also infected soil-grown plants of the immune-compromised *A. thaliana* mutant *enhanced disease susceptibility 1* (*eds1*-1), which is defective for numerous defense responses mediated by salicylic acid (SA) (*56*, *57*) and highly susceptible to *Pst* DC3000 infection (*58*). We also observed no symptoms on *eds1*-1, with infected plants indistinguishable from those sprayed with boiled *Pseudomonas* control (Figure 3e). *Pst* DC3000, in contrast, induced strong disease symptoms on *eds1*-1. In the future, it will be important to extend these results to mutants defective in jasmonic acid (JA) signalling, which is more important than SA in defense against other isolates of *P. viridiflava (59)*.

We did not observe a clear protective effect of any *Sphingomonas* strain against *Pst* DC3000 symptoms on soil (Figure 3f), and furthermore *S. melonis* Fr1-treated plants grown on soil were not stunted. To confirm that *Sphingomonas* was still present on leaves at the end of the experiment, and if pathogen titres had been affected, we quantified bacterial communities in the leaves of 5 plants per condition at 5 dpi using hamPCR (*60*), a quantitative amplicon sequencing approach that derives bacterial load by relating abundance of 16S rDNA to a single-copy host gene). Encouragingly, we observed the expected *Sphingomonas* 16S rDNA sequences enriched in each endpoint sample (Figure 3g). Supporting a protective role for *S. melonis* Fr1 on soil, samples pre-treated with this strain supported less *Pst* DC3000 than those pre-treated with others, and this was true on both the Ey15-2 and *eds1*-1 backgrounds, although the difference was only significant in *eds1*-1 (Mann-Whitney U-test, $p<0.05$). However, we found no robust support that any of the local isolates, including Sph139H133 that was protective on agar, reduced pathogen titres for plants grown on soil. Consistent with the plant phenotypes, *Pst* DC3000, but not *Pv-ATUE5*:p25c2, grew to higher titres in *eds1*-1 than in Ey15-2 (Supplementary Figure 6).

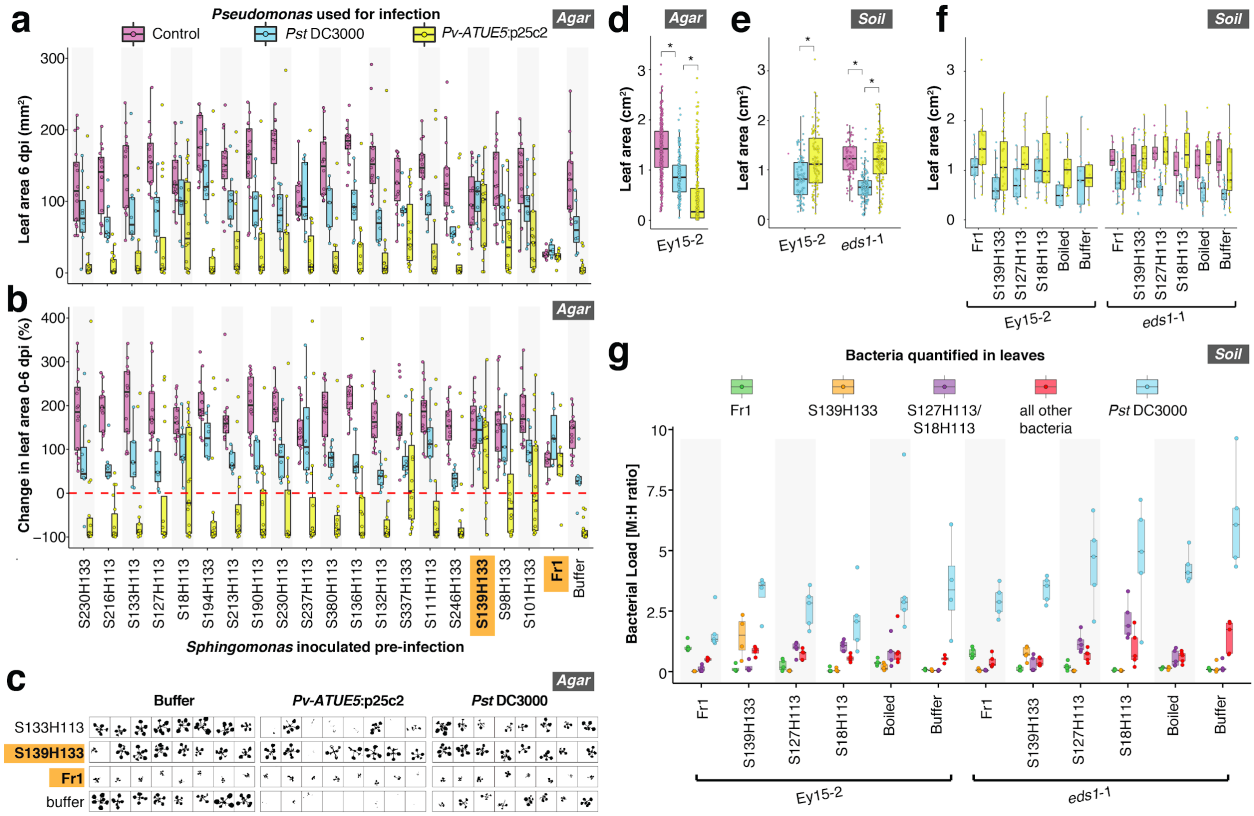

Figure 3. Some *Sphingomonas* offer protection against *Pseudomonas* | a-f) 19 local *Sphingomonas* isolates, *S. melonis* Fr1 (Fr1), or buffer were used to pre-treat *A. thaliana* Ey15-2 seeds germinating in 24-well agar plates. On day 10, seedlings were challenged with a buffer control, *Pst* DC3000, or *Pv-ATUE5:p25c2* and monitored for 6 days after infection ($n \geq 6$ per condition). Orange-highlighted *Sphingomonas* insignificant difference between infected and buffer-treated plants (Mann-Whitney U-test with FDR adjustment, $p > 0.05$) **a)** Absolute size of rosettes at 6 dpi. **b)** Percent change in rosette size between 0-6 dpi. **c)** Representative silhouettes at 6 dpi of seedlings grown in 24-well plates. **d)** Rosette size by infection treatment, regardless of *Sphingomonas* pretreatment. * = P < 0.001 (Mann-Whitney U-test with FDR adjustment). **e-g)** Three local *Sphingomonas* isolates, *S. melonis* Fr1, or buffer was sprayed on 2-week old seedlings of Ey15-2 or *eds1*-1, and 4 days later, live *Pseudomonas*, boiled *Pseudomonas,* or buffer was inoculated. **e)** Absolute sizes of rosettes for different *Pseudomonas* treatments, regardless of *Sphingomonas* treatment. * = P < 0.001 (Mann-Whitney U-test with FDR adjustment). **f)** Same data as in **(e)**, but split by *Sphingomonas* pre-treatment. **g)** Bacterial load ratio in plants infected with *Pst* DC3000 as determined by hamPCR. Note that the color legend represents bacteria quantified, not the bacteria used for treatment.

Sphingomonas and *Pseudomonas* strains thrive on multiple plant species

Sphingomonas and *Pseudomonas* are generalists, with no known exclusive hosts of any given strain (*12*). As a first step towards determining the extent of generalization of strains across local plant species, we compared ASVs across plants. We observed that SphASV1 was more abundant overall in the spring collection, but it was consistently detectable on most plant species in both seasons. Plant taxa colonized by SphASV1 were also frequently colonized by multiple additional *Sphingomonas* ASVs to appreciable levels (Figure 4a). In contrast, PseASV1 was relatively more abundant for *A. thaliana, C. hirsuta* and *D. verna* - all from the family Brassicaceae - and less abundant, though still present, on other taxa (Figure 4b). PseASV1 was

especially enriched in the spring, which matches our previous finding of more *Pv-ATUE5* isolates from spring vs. winter collected *A. thaliana* (*18*). The presence of an ASV across plant species is not necessarily an indication that bacterial strains are shared across plants. To test for strain sharing, we compared genome similarity for all PseASV1 and SphASV1 genomes we isolated from *A. thaliana* and four other plant species chosen for individual colony isolation: *T. officinale*, *T. repens*, *C. hirsuta*, and *I. glandulifera* (Figure 4c, 4d). These plants were chosen for isolates due to robust cultivation of both *Pseudomonas and Sphingomonas* from the same glycerol stocks for these species.

While 29 out of the 86 SphASV1 isolates from the 2018 harvest shared at least 99.9% ANI with at least one other isolate, with only one exception these highest-similarity isolates came from the same lysate pool, meaning they could be clonal descendants from the same plant and might not represent a migration event to a new plant. To avoid potential intra-plant clones, we recalculated all pairwise genome similarities between the SphASV1 isolates, ignoring those comparisons within the same lysate pool. We compared similarity within *A. thaliana* isolates (intra-host isolate similarity) and between *A. thaliana* isolates and those from any other plant species (inter-host isolate similarity). A higher intra-host similarity would be evidence of host-specificity, possibly due to unique selective forces within *A. thaliana*, or easier migration between individuals (for example facilitated by insects). We found that the distribution of intra-host ANI values closely matched that of inter-host isolate ANI values (Figure 4e), with the intra-host ANI being only marginally higher (0.9%, Mann-Whitney U test, $p < 0.001$).

PseASV1-associated genomes were more similar to each other than SphASV1-associated genomes (Figure 2a). Every isolate shared at least 97.5% ANI with at least one isolate from a different plant species, and 47 of the 50 isolates shared at least 99.9% ANI with at least one other isolate (Figure 4d) but as with SphASV1, these highest-similarity isolates tended to come from the same plant pool. Three groups of highest-similarity isolates were shared across different plant pools, different species, and even different seasons (Figure 4d). This is consistent with our previous observation that *Pv-ATUE5* strains are common and persistent (*18*), and further demonstrates that other local plant species host these strains. As with SphASV1 isolates, the intra-host ANI values followed a similar distribution to inter-host ANI values, with intra-host isolate ANI values marginally higher (0.4%, Mann-Whitney U-test, $p < 0.001$, Figure 4f).

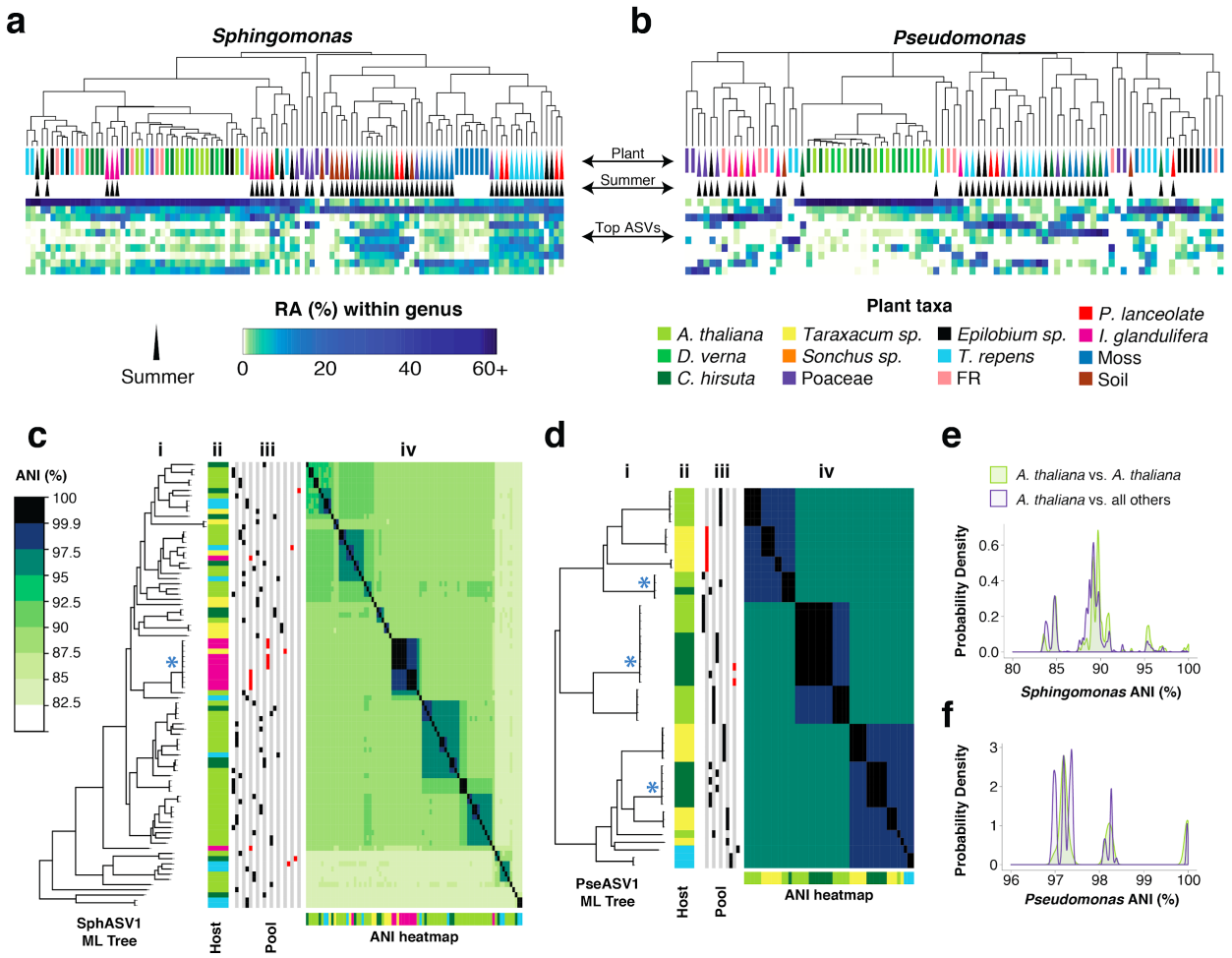

Figure 4. Isolates from other plant species match those isolated from *A. thaliana* | a-b) Plant species from 2018 clustered by Bray-Curtis dissimilarities between relative abundances (RA) for the 10 most abundant 16S rDNA ASVs for each genus. The legend below the figures is shared for both panels. **a)** *Sphingomonas* ASVs, with SphASV1 as the top row in the heatmap **b)** Same as **(a)**, but for *Pseudomonas* ASVs. **c)** Heatmap revealing genome ANI of all pairwise comparisons between SphASV1 isolates. **i:** Isolates arranged by a maximum likelihood (ML) core genome tree. The * represents the position of a group of isolates sharing > 99.9% ANI that derives from different sample pools. **ii**: plant host of origin is indicated by color, corresponding to the plant legend in **(a-b)**. **iii**: Isolates from the same sample pool are indicated by red (summer) or black (spring) marks in the same vertical column. **iv**: heatmap showing pairwise ANI values. **d)** Same as c, but for PseASV1. **e)** Distribution of pairwise intra-host isolate ANI compared to inter-host isolate ANI for SphASV1. **f)** Same as **(e)**, but for PseASV1 isolates.

Bulk culture metagenomics reveals strain sharing across plant species

From our genome-sequenced isolates, it was clear that closely related strains had a regular tendency to colonize diverse hosts. To broaden our survey and extend these observations, we adopted a time and cost-efficient approach to enrich each genus in bulk from plant lysates and sequence the enriched pool as a metagenome. We plated 50 µL of glycerol stock from each lysate, corresponding to ~5 mg of original plant material, on selective *Pseudomonas* or *Sphingomonas* media and grew colonies *en masse* (Figure 5a, Supplementary

Figure 7). This quantity of lysate was chosen empirically for our samples to avoid colonies merging into lawns and allowed visualization of discrete colonies. As a control for lysate viability, we also cultured bacteria from lysates on a non-selective rich LB media. We counted the colonies on each plate after 2 days at room temperature for *Pseudomonas*, or after 7 days for slower-growing *Sphingomonas*, to provide a rough estimate of the bacterial load of each genus (Figure 5b). As expected, rich LB media yielded more colonies than selective media, although counts on LB likely vastly underestimate the total bacteria population in each sample, because a single rich medium does not meet nutritional demands for all bacterial taxa, and it furthermore allows fast-growing colonies to overgrow others before they become visible.

To harvest the bacteria, we flooded the plate with phosphate buffered saline (PBS) and then used a flame-sterilized razor blade to scrape the resulting colonies to a pool, from which we made whole metagenome shotgun libraries (Figure 5a, Methods). This culture-dependent metagenomic strategy (*61*) allowed us to investigate endophytic-enriched bacteria of our genera of interest, while completely avoiding plant DNA. After sequencing the metagenomes using paired 2×150 HiSeq 3000 reads, we mapped the sequences to a comprehensive reference database including all of our local genomes plus selected publicly-available genomes (*45*, *62*) (Methods, Supplementary Figure 8). Because not all colonies on our plates belonged to our targeted genus, we also included in our reference database "decoy" genomes of common plant-associated bacteria to capture those "contaminant" reads and help prevent misclassification to our local strains. A total of 192 metagenomes (119 *Sphingomonas* + 74 = *Pseudomonas*) passed our quality thresholds. For both genera, as also suggested by the amplicon data in (Figure 4a, 4b), there was also a clear shift in dominant bacteria across seasons, including for those plant taxa such as *T. officianale,* Moss, and *C. hirsuta* that were present at both timepoints (Figure 5c, 5d). Finally, it was clear that many of the same successfully-colonizing strains were shared widely across plants. In particular, this could be observed for strains associated with SphASV1 and PseASV1. Interestingly, the relative representation of both SphASV1 and PseASV1 isolates in the summer soil samples was low, suggesting that as the next cohort of *A. thaliana* germinates in the fall, migration of these strains onto the seedlings from nearby plants may be more likely than from soil.

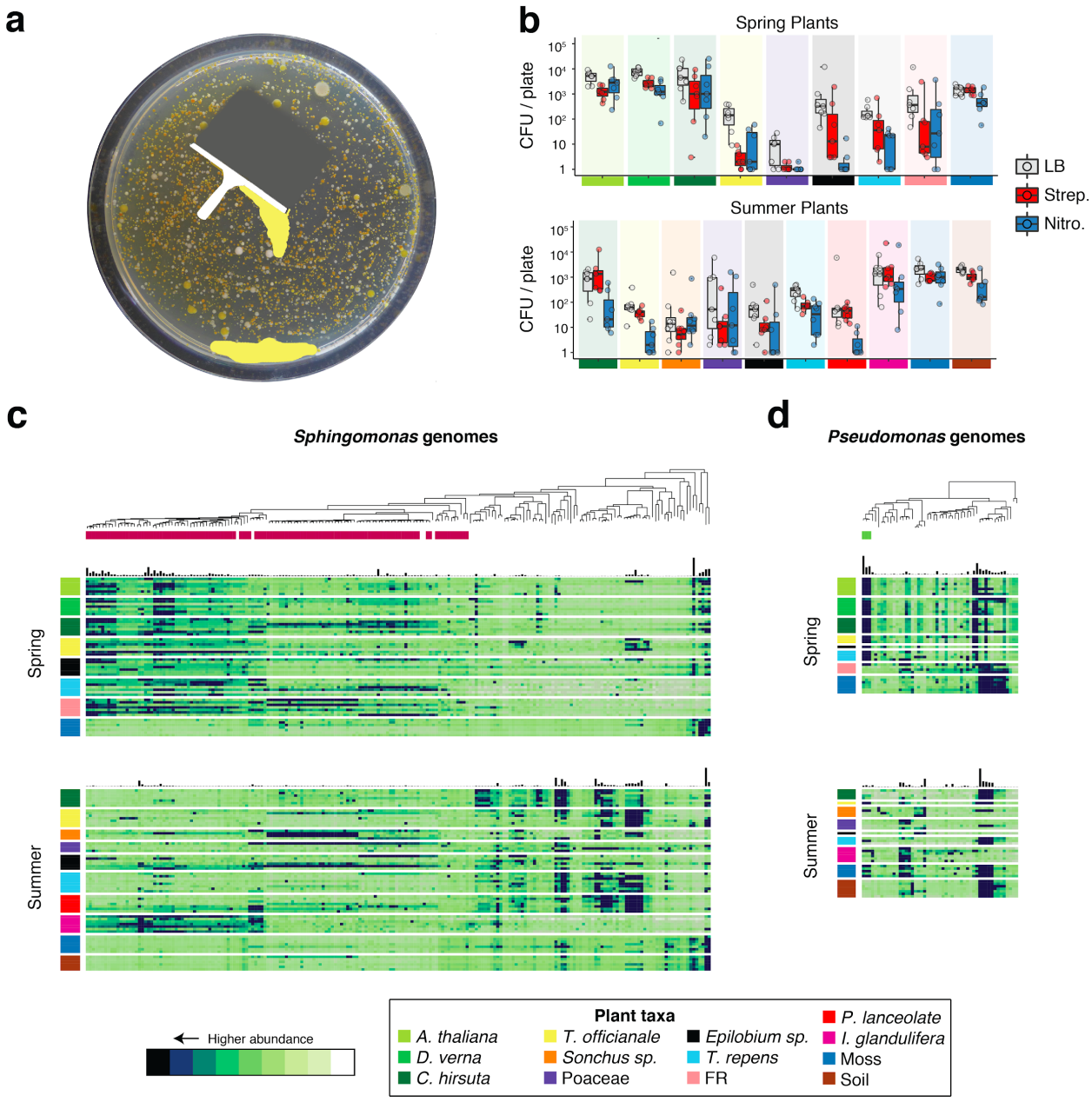

Figure 5: Metagenomics of cultured bacteria reveals strain sharing across plant species | a) Cartoon demonstrating the process of generating and harvesting bulk culture metagenomes. **b)** Colony forming units for each plant species (see color legend, bottom left) on media containing no antibiotic (LB), *Sphingomonas*-selective streptomycin (Strep) media, or *Pseudomonas*-selective nitrofurantoin (Nitro) media. Each plate represents about 4.5 mg fresh weight of original leaf material. **c)** Heatmap showing fourth-root transformed reads from each *Sphingomonas* bulk culture metagenome (rows, colors from plant taxa legend below) that map to a *Sphingomonas* local reference genome (columns), for the spring collection (top) and the summer collection (bottom). The genetic relatedness of the local reference genomes is shown by a maximum likelihood (ML) tree above the heatmap. Those reference genomes associated with SphASV1 are indicated under the ML tree in magenta. Darker colors in the heatmap correspond to genomes to which a greater fraction of reads map. **d)** Same as **c**, showing *Pseudomonas* bulk culture mapping to *Pseudomonas* reference genomes. Reference genomes associated with PseASV1 are indicated under the ML tree in green.

Discussion

Most microbes that live in or on multicellular organisms do not have an obligate relationship with any specific multicellular species, but rather are better adapted to some (hosts) than to others (nonhosts) (*63*, *64*). We sought to use genomic techniques to characterize, at the strain level, the extent of host specialization for the two bacterial genera that are locally most abundant in the phyllosphere of the host plant *A. thaliana*. We found a higher genomic diversity of *Sphingomonas* than for *Pseudomonas*, even among those sharing the same 16S rDNA sequence, highlighting the importance of strain-resolved techniques to address this question in wild ecosystems. To this end we employed direct culturing and sequencing of individual isolates, as well as bulk-culture metagenomics (Figure 5). Although our techniques enable culture of genetically diverse *Sphingomonas* and *Pseudomonas* with low bias (Figure 1f, 1g), bulk culture metagenomics will overrepresent strains that grow faster on the culture media. Therefore, this technique is best thought of as a powerful way to screen for strains of interest in a field site and to qualitatively measure differences in strain membership across plants, rather than to quantitatively characterize the relative strain abundance within a sample. Strain distribution across seasons and other plant species at the same site point to potential explanations for the contrasting colonization patterns of these genera.

Our previous work characterizing PseASV1/*Pv-ATUE5* suggested that much of its standing genetic variation predates *A. thaliana* colonizing SW Germany (*18*), and we had puzzled how ancient variants are apparently able to continue to coinfect the entire population in this pathosystem, in contrast to agricultural systems epidemics are typically monomorphic with a rapid turnover of pathogenic isolates within a few years (*65*, *66*). The present study begins to answer this question. First, PseASV1/*Pv-ATUE5* strains efficiently colonize other local plant species besides *A. thaliana*, especially other Brassicaceae, implying that their performance on these additional hosts may be at least equally, if not more important to their long-term success than their exploitation of *A. thaliana* as a host. The *Pseudomonas* strains we observed on *A. thaliana* also persist through the summer on other hosts when there are no living *A. thaliana* plants. Currently, we do not know how important different hosts are for different subclades of PseASV1/*Pv-ATUE5*, but with a substantial number of host species together with genetic heterogeneity in each host, one can easily see how it is difficult for any one strain to prevail and to come to dominate the entire *Pseudomonas* population. In contrast, in a crop monoculture genetically-identical hosts provide more consistent and uniform challenges for a pathogen, making it more likely for a single strain to become dominant. A second explanation for the lack of a single emergent strain is that the pathogenic nature of PseASV1/*Pv-ATUE5* is strongly context-dependent, and apparently more so than other described *Pseudomonas* pathogens such as *Pst* DC3000. In an agar plate system, disease caused by PseASV1/*Pv-ATUE5* was much more severe than that of *Pst* DC3000, but in potting soil, PseASV1/*Pv-ATUE5* failed to cause disease symptoms and grew more slowly than *Pst* DC3000 in wild-type plants. Consistent with fitness

effects of PseASV1/*Pv-ATUE5* being greatly modulated by the environment, we saw few disease symptoms on field-collected plants at the time of collection, regardless of their *Pseudomonas* load.

Compared to PseASV1/*Pv-ATUE5*, *Sphingomonas* strains that shared the most abundant V3V4 *Sphingomonas* ASV, SphASV1, were more than three times as diverse by nucleotide identity in their core genomes. SphASV1 isolates differed in type 4 and type 6 secretion system presence, anoxygenic photosynthetic ability, plasmid presence, and likely in other unidentified features that are likely to affect host colonization potential or intermicrobial competition. The plasmid count, up to three per strain in our closed genomes, suggests that the *Sphingomonas* genetic tool kit may be highly modular. We observed SphASV1 in diverse 16S rDNA datasets worldwide, including high abundances on *Boechera stricta* in North America (*13*), and while the extent of genome variation among SphASV1 strains in SW Germany makes it difficult to predict genomic features, it will be highly interesting to determine the global diversity in this group of *Sphingomonas* strains.

This study also demonstrated that apart from SphASV1, the genus *Sphingomonas* is not only abundant, but consistently so across plants (Figure 1). It also colonizes diverse plants at our field site (Supplementary Figure 3), a result consistent with reports of high abundances on leaves of diverse plants such as maize (*14*), poplar (*10*, *11*), and others (*10*). The different colonization patterns of the genera *Pseudomonas* and *Sphingomonas* provide clues about their lifestyle strategies. The fact that *Pseudomonas* populations vary greatly in size between plants, occasionally reaching high abundances as homogenous blooms of PseASV1/*Pv-ATUE5* (*18*), suggests that different strains may compete for resources and share the same niche, making stable coexistence within a plant less likely. In contrast, *Sphingomonas* loads varied little across plants, regardless of the load of other bacteria (Figure 1a), and the diversity of *Sphingomonas* ASVs in each plant was also higher. The more balanced coexistence of different *Sphingomonas* strains may mean that each occupies a different niche. To what extent these bacteria inhabit spatially-distinct parts of the leaf or grow as biofilms is unknown; future direct visualisation techniques will help resolve this.

With *Pseudomonas* and *Sphingomonas* among the most abundant genera of leaf bacterial communities, and both genera having epiphytic and endophytic lifestyles (Figure 1c), they may frequently interact. Because it reaches such high abundances on and in leaves, *Sphingomonas* metabolic needs likely substantially reduce substrates of value to other phyllosphere microbes (*33*), a feature that may contribute to the protective effects of various strains against bacterial and fungal pathogens and contribute to their effects on overall community structure (*5*, *31*). However, in at least one known case, a *Sphingomonas* strain secretes an extracellular small molecule that attenuates the virulence of a bacterial pathogen (*21*), indicating that protective mechanisms likely go beyond substrate competition. Although we estimate that in our culture collection only a minority of *Sphingomonas* strains have protective ability against PseASV1/*Pv-ATUE5* or *Pst* DC3000, the high bacterial titres we inoculated and the gnotobiotic

environment in which we observed the strongest effects may limit transferability of our results to the field, especially as the most protective isolate had a 16S rDNA sequence that was not among the more abundant on field plants. However, the high diversity of *Sphingomonas* genomes and the fact that we saw some antagonism suggests there is more to be discovered. The plant protective *S. melonis* Fr1 induces transcriptional changes in *A. thaliana* (*53*), and likely other strains similarly are sensed directly by the plant. The overall implications of *Sphingomonas* populations on leaves will be important to understand, both for their direct impacts on plant evolution, but also for their role in structuring wild microbial communities.

Materials and Methods

Plant collection and lysate glycerol stock preparation

Plants were harvested at Eyach (Starzach) Germany (Supplementary Figure 1). The spring visit occurred on the 20th of April, 2018, and the summer visit occurred on the 14th of September, 2018. For each plant species, we pooled together entire leaves from at least 6 independent plants per sample, and collected 7 such independent samples. After bringing samples back to the lab, we surface-sanitized them in 70% ethanol for 45-60 s (*55*), ground the fresh tissue using a sterile mortar and pestle in a volume of PBS proportional to the sample's fresh weight (7.4 mL PBS per gram of tissue), and mixed the resulting lysate with glycerol to make -80°C freezer stocks with a final glycerol concentration of ~27%, suitable both for direct DNA extraction and culturing.

Isolating individual *Sphingomonas*

Most *Sphingomonas* isolates (Supplementary Table 2) were cultured from glycerol-frozen *A. thaliana* lysates that were originally processed and stored on December 11, 2015 and March 3, 2016 (*18*). Briefly, we processed two leaves each per *A. thaliana* rosette. Each leaf was washed in 75% EtOH for 3-5 s, the ethanol was allowed to evaporate in a sterile hood, and the leaf was ground in 10 mM $MgSO_4$ before being mixed with glycerol to a final concentration of 15%-30%. The glycerol stocks were stored at -80°C. From each sample, 75 mL of lysate was plated on large petri plates (145 x 20 mm, Greiner Bio One International GmbH, Frickenhausen, Germany) containing low sodium Luria broth (LB, 10 gL peptone, 5 g/L yeast extract, 5 g/L NaCl) and 1.5% agar supplemented with 100 mg/mL cycloheximide to suppress fungi (SERVA Electrophoresis GmbH) and 100 mg/mL streptomycin (Thermo Fisher Scientific, Karlsruhe, Germany). Plates were incubated at room temperature for 5 to 10 days, then stored at 4°C until processing. The remaining *Sphingomonas* isolates from pooled *A. thaliana* and other plant species were recovered similarly by plating lysates on selective media (see section "Bulk culture and additional cultured isolates").

Up to 16 colonies per plate were randomly picked and transferred to 24-well plates (CELLSTAR®, Greiner Bio One International GmbH, Frickenhausen, Germany) containing the same selective LB agar, and incubated for 5-10 additional days as necessary to create a small bacterial lawn. All bacteria were cultivated on agar, as we observed that many grew poorly in liquid media. Bacteria were then scraped from the agar with a sterile loop and transferred to a 2 mL screw cap tube (Starstedt, Nümbrecht, Germany) filled with 300 ml of PBS (we later switched to 300 mL of liquid LB media instead of PBS, which improved survival). Three sterile glass balls (Æ 2.85-3.45 mm or Æ 0.25-0.5 mm) (Carl Roth GmbH, Karlsruhe, Germany) were added to each screw cap to help dissociate (but not lyse) clumped bacterial cells, and tubes were shaken in a FastPrep 24™ 5G homogeniser (MP Biomedicals, Eschwege, Germany) at speed 4.0 m/s for 10 seconds. From the mixed cell suspension, 250 mL was stored at -80°C in ~27% glycerol.

Sphingomonas DNA isolation for short read sequencing

The remaining 50 mL was mixed with 150 mL of sterile PBS and transferred to a semi-skirted 96-well PCR microplates (Axygen™) for DNA extraction. Bacterial cell suspensions were lysed by incubation with lysozyme (100 mg/mL final concentration) at 37°C for 30 min followed by incubation with sodium dodecyl sulfate SDS (1.5% final concentration) at 56°C for 1-2 h. Next, ⅓ volume (66 µL) of 5 M potassium acetate (CH_3COOK) was added to each well to precipitate the SDS and other cytosolic components. The 96-well plates were centrifuged at maximum speed for 20 min to pellet the precipitate, and the supernatant transferred to a new 96-well plate for genomic DNA purification using Solid Phase Reversible iImmobilization (SPRI) magnetic beads (*67*). Briefly, home-made SPRI bead mix adapted from (*67*) was thoroughly mixed with our samples at a ratio of 0.6:1 bead to sample ratio and incubated for 15-20 min. The plates were placed on a magnet for 5 min and the supernatant was removed. Following two 80% EtOH washes, the beads were air-dried and resuspended in 50 µL elution buffer (EB, 10 mM Tris, pH 8.0). After an overnight incubation at 4°C, the plates were placed in the magnet and the elution buffer containing the DNA was transferred into a new 96-well PCR plate.

Bulk culture and additional cultured isolates

For *Pseudomonas* and *Sphingomonas* bulk culture analysis, approximately 50 mg of plant lysates collected from Eyach in 2018 was scooped from the frozen glycerol stocks at -80°C, and after thawing, 50 µL of the thawed lysate was pipetted onto 2% agar LB plates (200 mm diameter), using an additional 150 µL of sterile PBS to aid in spreading. The LB medium was supplemented with 100 µg/mL cycloheximide and 100 µg/mL streptomycin for *Sphingomonas* bulk culturing, and 100 µg/mL cycloheximide and 100 µg/mL nitrofurantoin (Sigma-Aldrich,

Steinheim am Albuch, Germany) for *Pseudomonas* bulk culturing (*18*). Plates were incubated at room temperature 5 days for *Sphingomonas* or 1.5-2 days for *Pseudomonas*.

Individual colonies from selected plates were randomly picked to 24-well plates, as described above. To harvest the remaining bacterial colonies in bulk, plates were soaked for 5 min with 4 mL of PBS, and the surface was scraped with a flame-sterilized razor blade to loosen adherent bacteria. The plate was tilted to form a pool of PBS at the lower end, and the scraped bacteria were mixed into the PBS pool by pipetting up and down. An appropriate aliquot of the mixed suspension was then transferred to a 2 mL centrifuge tube, such that the eventual pellet would extend a maximum of 5 mm from the base of the tube (which approaches an upper limit for efficient and uniform DNA extraction using our methods). The amount of bacteria transferred varied depending on the density of colonies on the plates.

DNA extraction from bulk culture bacterial pellets and plant lysates

For bulk culture bacterial pellets, the final bacterial pellets were resuspended in at least 750 µL of DNA lysis buffer containing 10 mM Tris pH 8.0, 10 mM EDTA, 100 mM NaCl, and 1.5% SDS. Especially large pellets (greater than 5 mm from the bottom of the tube) were suspended in proportionally greater volumes of buffer to ensure an efficient lysis, and 750 µL was used for lysis. For DNA extraction from the plant lysates in glycerol, 400 uL of glycerol lysates was mixed with 400 µL of the DNA lysis buffer described in the previous sentence, but with 3% SDS instead of 1.5% SDS to yield a final SDS concentration of 1.5%. The suspensions were pipetted to a screw cap tube containing ~0.5 mL sterile garnet beads (Bio Spec Products Inc., USA) and homogenized in a FastPrep 24™ 5G homogeniser at speed 6.0 m/s for 1 min. The tubes were next centrifuged at 10000 x g for 5 min, and the supernatant (about 600 µL) mixed with 200 µL sterile 5 M potassium acetate in a 1 mL 96-well plate (Ritter Riplate® 43001-0016, Schwabmünchen, Germany) to precipitate the SDS. The plates were spun at 5000 x g for 10 min and the supernatant transferred to a new 1 mL deepwell plate. The resulting supernatant was centrifuged a second time to clear out remaining plant material and precipitate. Finally, 360 µL SPRI beads were added to 600 µL of the supernatant. After mixing and incubating on a 96-Well Magnet Type A (Qiagen, Hilden, Germany) the beads were cleaned with 80% ethanol and DNA was eluted in 100 µL of EB.

16S rDNA V3-V4 amplicon sequencing

The 16S rDNA V3-V4 region was amplified using the 2-step protocol described for V4 amplicons in (*9*), with the exception that the forward PCR primer was 341F (Supplementary Table 1), and due to this different forward primer, the annealing step in the first PCR was done at 55 °C instead of 50 °C in (*9*). In addition, because the amplicons were longer than the V4 amplicons in (*9*), the libraries were sequenced on MiSeq instrument (Illumina) with a V3 2 × 300 bp reagent kit instead of a V2 2 × 250 bp reagent kit. This allowed overlap and assembly

of the forward and reverse reads. The frameshifts built into the primers used in the first PCR made the addition of Illumina PhiX control library to increase sequence diversity unnecessary [27].

Bacterial whole-genome sequencing and bulk culture metagenome sequencing with short reads

DNA from each bacterial isolate, or from each metagenome, was quantified by Quant-iT™ PicoGreen™ dsDNA (Invitrogen™) in a Magellan Infinite 200 PRO plate reader (Tecan Trading AG, Switzerland) and diluted and normalised to 0.5 ng/µl as a prior step to library construction. Bacterial DNA libraries were constructed using an adapted Nextera™ protocol for small volumes (*18*, *68*). In brief, 2.5 ng of DNA was sheared using Nextera Tn5 transposase (Illumina, Inc), and Nextera sequencing adapters were added though 12 cycles of PCR as previously described (*18*). An aliquot of each library was run in an agarose gel for quality control, and the remainder of the library was purified using SPRI beads to remove primers at a ratio 1.5:1 beads to PCR product. The clean DNA was eluted in 40 µL of EB and the concentration of the final product was quantified with PicoGreen. Libraries were pooled in an equimolar ratio. To increase the concentration of the pool to enable further size selection procedures, the pooled library was concentrated by first precipitating the DNA by mixing with it with an equal volume of a solution containing 1 part sodium acetate and 8 parts isopropanol, passing the solution over an EconoSpin Mini Spin Columns (Epoch Life Science Inc.), washing the column twice in 70% ethanol, and eluting in 50 µL EB. The resulting multiplexed, concentrated library molecules were size selected to keep fragments between 350 and 700 bp using a 1.5% cassette in a BluePippin instrument (Sage Science, Inc). After adjusting the concentration of each size-selected pool to 2.5 nM, the DNA was sequenced with 2 ×150 bp paired-end reads on a HiSeq 3000 instrument (Illumina inc, San Diego, USA).

Short-read genome assembly and annotation

Genomes were assembled using SPAdes genome assembler version 3.11.0 correcting for mismatches and short indels (--careful) and k-mer sizes of 21, 33, 55, 77, 99 and 127 (-k) (*69*). Draft genomes were corrected using Pilon version 1.20 with standard parameters (*70*). Annotation of bacterial genomes was accomplished using Prokka version 1.12 using the --compliant parameter (*43*). Coverage and N50 statistics were measured using custom scripts, including the N50.sh script from the GitHub repository of Henk den Bakker (*71*). The completeness of the genome was assessed using BUSCO version 2.0 (*41*) selecting proteobacteria lineage and the gene set (proteins) assessment (-m prot).

Production and assembly of closed Sphingomonas genomes

Assembly of closed *Sphingomonas* genomes required a separate DNA preparation and library production for sequencing on an Oxford Nanopore MinION instrument. First, 12 *Sphingomonas* isolates were each bulked on 15 × 15 cm^2 LB plates with 100 µg / mL streptomycin. After 2-5 days, depending on the growth rate of each strain, the bacterial lawn was harvested, and ~100 mg of bacteria was mixed with 800 µL lysis buffer, and DNA was extracted and purified using the homemade SPRI beads exactly as described above for bulk culture lysates, including the harsh bead beating. We noticed that bead beating did not shear DNA below 10 kb, and fragments greater than approximately 7 kb are sufficiently long to span repeated regions in bacterial genomes (*72*). The harsh bead beating during the lysis step naturally sheared the DNA to approximately 15-20 kb, and the 0.6 : 1.0 SPRI bead cleanup removed most of the smaller fragments, so no additional shearing or size selection was performed. DNA was eluted from the SPRI beads in 400 µL EB. Because some DNA extracts remained discolored or viscous following the SPRI purification, the DNA was further purified by mixing with it with an equal volume of chloroform. The aqueous phase was collected and the HMW DNA was precipitated by mixing it with an equal volume of a solution containing 1 part sodium acetate and 8 parts isopropanol. Precipitated DNA was pelleted by centrifugation at 20,000 x g for 5 minutes. The pellet was washed twice in 70% ethanol and eluted again in 400 µL EB.

Pure DNA was prepared for sequencing on the Nanopore MinION using the Nanopore-recommended protocol entitled "1D Native barcoding genomic DNA" using SQK-LSK109 kit with barcode expansions NBC104 and NBC114 for 24 samples. Briefly, DNA concentration was adjusted to 1 µg of DNA diluted in 49 µL of NFW water (Ambion). From this 48 µL were treated with NEBNext reagents for FFPE and end repair as well as dA-tailing. The end-repaired libraries were cleaned with Agencourt AMPure XP beads at a 1:1 ratio and eluted in 25 µL in a 1.5 mL DNA LoBind® tube (Eppendorf AG, Hamburg, Germany). Native barcodes were attached by ligation, the solution was again cleaned with AMPure beads at a 1:1 ratio, and concentration of all barcoded-added samples were measured using a Qubit (Thermo Fisher) fluorometer and pooled at equimolar ratios for a total of 700 ng. Finally, sequencing adapters were added to the pool by ligation. LFB was selected for washing in the final clean-up with AMPure beads. The library was prepped and loaded into an FLO-MIN106 RevD R9.4.1 flow cell following the manufacturer's instructions and sequencing was run for 24 h.

To assemble closed *Sphingomonas* genomes, Guppy package version 3.0.3 (https://nanoporetech.com/) was first employed to perform initial base calling to produce raw read and quality assessment in the FASTQ format as recommended by (*73*). Next, the samples were demultiplexed using qcat. Draft contigs were then assembled *de novo* by using mini_assemble which is part of the Pomoxis toolkit (version 0.3.6,

https://github.com/nanoporetech/pomoxis) using default parameters for de-novo assembly employing miniasm (*74*) and four rounds of long-reads-based polishing with minimap2 and Racon. Each of the genomic consensus assembly was further improved by additional 4 rounds of long-reads-based polishing with Racon followed by one additional round using medaka_consensus (package version 0.6.5, https://github.com/nanoporetech/medaka). To enhance base corrections, an additional polishing step using high-quality Illumina short reads from the same *Sphingomonas* strains was used. First, the short reads were mapped onto the assembled *Sphingomonas* genome, using Burrows-Wheeler Alignment (BWA) with BWA-MEM (*75*). Next, these mapped reads were used to correct the assembly using Pilon version 1.23 with default parameters (*70*). Assessment of genome completeness and annotations were performed in the same way as short-read draft genomes.

Pan-genome and phylogenetics analysis

The panX pan-genome pipeline (*44*) was used to assign orthology clusters and construct the phylogenetic tree, taking as input the Genbank format files (.gbk) from Prokka (previous section). The following parameters were used for the panX analysis: the divide-and-conquer algorithm (-dmdc), a size of 50 strains per subset to run DIAMOND (-dcs 50) (*76*), and a soft core genome cutoff of 70% that includes all genes present in >70% of the strains as part of the core genome (-cg 0.7). Genomes included in each panX run are indicated in Supplementary Table 2.

Processing bulk culture metagenomic reads

A first quality control of raw sequencing data was performed using Skewer version 0.2.2 (*77*) to trim raw reads and remove highly degenerative reads (-n) or reads shorter than 20 bp (-l 20). For those samples with a total yield of at least 500 Mb, the filtered reads were mapped with BWA-MEM (*75*) using standard parameters against a custom-made reference database (described in the next section). Those reads mapping with a quality score of 30 or higher were output in a BAM file using SAMtools (*78*). Duplicated reads were removed from BAM files using the MarkDuplicates command in Picard tools (http://broadinstitute.github.io/picard/) version 2.0.1 using default parameters. SAMtools -stats and -fastq commands were used to retrieve BWA-MEM mapping statistics and convert the BAM files back to FASTQ files, respectively (Supplementary Figure 8).

Custom local reference database

Metagenome reads from the *Sphingomonas* or *Pseudomonas* bulk cultures were mapped to a three part custom-made reference database of bacterial genomes. First, "Decoy" genomes contained bacteria from other genera, and were used to help classify reads from contaminant bacteria. The collection of Decoy genomes was made by downloading the *A. thaliana*

phyllosphere, root, and soil isolates from (*62*) and removing *Pseudomonas* when mapping *Sphingomonas,* or removing *Sphingomonas* when mapping *Pseudomonas.* Three different genome sets were combined with distinctive headings to allow later recovery of mapped reads:

DECOY: A set of bacterial genomes (~430 genomes) found in *A. thaliana* phyllosphere, roots, and soil surrounding the plants. Published in (*62*). Decoy genomes have the mission to capture contaminant reads.
REFSEQ: A set of bacterial genomes of *Pseudomonas* or *Sphingomonas* found in the NCBI Reference Sequence Database (*45*).
LOCAL: A set of bacterial genomes of interest sequenced in-house. This includes 165 genomes representative of local *Pseudomonas* isolates sequenced by (*18*) and all the single *Sphingomonas* and *Pseudomonas* genomes from this study.

Genome similarity comparisons

Briefly, to generate similarity matrices for *Sphingomonas* and/or *Pseudomonas* genome comparisons using MASH (version 2.1)(*42*), we used the formula 100 × (1-MD), where MD is the MASH distance, to convert MASH distances to similarity scores. To calculate average nucleotide identity (ANI), we used FastANI (*79*) For similarity matrices Heatmaps of similarity scores were illustrated using the function "heatmap.2" in the R package "gplots" (*80*).

Inoculation, infection, and phenotyping in 24-well plates

Plant cultivation: Each well of a 24-well plate (Greiner) was filled with 1.5 mL of 1% agar (Duchefa) containing ½ strength Murashige-Skoog (MS) medium with MES buffer. *A. thaliana* seeds (accession Ey15-2, CS76309) were surface-sterilized by submerging in 70% EtOH with 0.01% Triton X-100 for 1 min, then submerging in 10% household bleach solution for 12 min, and finally washing three times with sterile water. The seeds were then stratified at 4°C for 3 days in water, and then were pipetted onto the agar (1 seed/well). Excess water after distributing seeds was removed by pipetting.

Inoculation and infection: Freshly-planted, ungerminated seeds were inoculated with 4 µL of 10 mM $MgCl_2$ or with *Sphingomonas* sp. suspended in the same amount of buffer to an optical density at 600 nm (OD_{600nm}) of 0.5. The *Sphingomonas* had been previously cultivated for 5 days on selective LB agar plates with 100 mg/mL streptomycin, were scraped from the plates with a sterile loop, and were washed twice by centrifugation and resuspension in $MgCl_2$ to remove residual antibiotics. Inoculated seeds were germinated and seedlings were cultivated for 10 days in growth chambers at 21 °C with 16 hours of light. The seedlings were then challenged with 100 µL of 10 mM $MgCl_2$ or with *Pseudomonas* suspended in the same amount of buffer to OD_{600nm} = 0.01. The bacteria were drip-inoculated by pipette to the center of each rosette. Plates were sealed with Parafilm and returned to the growth chamber for 7 days.

Plant phenotyping: On 0, 2, 4, and 7 days post inoculation, rosettes were imaged in the plates with a custom procedure to eliminate glare. Briefly, an opaque box was filled with a LED light source and covered with a sturdy translucent paper surface to diffuse the light. Each plate was placed on top of the paper in a defined position, and the backlighting allowed imaging from above without removing the lids. Pictures were taken using a LUMIX DMC-TZ71 digital camera (Panasonic Co., Osaka, Japan) without flash. Images were processed similarly to that described in (*81*). Briefly a predefined mask was used to extract each plant in the image, and automatic segmentation based on pixel color was applied to recognize plant leaves from background. The leaf area of each plant was then calculated based on the segmented plant images.

Inoculation, infection, phenotyping, and hamPCR of plants on potting soil

Plant cultivation: *A. thaliana* Ey15-2 seeds were surface-sterilized, stratified for 4 days in sterile water at 4°C, and sowed on potting soil (CL T Topferde; www.einheitserde.de) in 7 cm pots (PL 2832/20, Pöppelmann GmbH, Lohne, Germany). Seedlings were germinated and cultivated under short day (8 hr light) growing conditions at 23°C and 65% relative humidity, illuminated by cool white fluorescent light of 125 to 175 $\mu mol \times m^{-2} \times s^{-1}$.

Inoculation and infection: Two weeks after sowing, seedling leaves were sprayed ad- and ab- axially with *Sphingomonas* strains cultivated identically as for inoculation in 24-well plates as described above, but resuspended in 10 mM $MgCl_2$ buffer to a concentration of OD_{600nm} = 1.0. Plants were also sprayed with heat-killed (boiled) *Sphingomonas* prepared by mixing equal parts of all strains after they had been resuspended at OD_{600nm} = 1.0 in $MgCl_2$ and boiling the resulting solution for 10 minutes. Humidity domes were kept on the flats for 48 hours. On the fourth day following *Sphingomonas* treatment, plants were sprayed with *Pseudomonas* strains also cultivated identically as for 24-well plates, but resuspended in 10 mM $MgCl_2$ buffer to a concentration of OD_{600nm} = 1.0. A heat-killed *Pseudomonas* mix was also prepared, as described above. Prior to spray-inoculating *Pseudomonas,* the surfactant Silwet L-77 added at 0.04% v/v following the protocol in (*54*). Humidity domes were kept on the flats for 72 hours.

Plant phenotyping: Five days (120 h) post inoculation, overhead images of pots arranged in the flat were taken from a height of ~ 1.5 M with a LUMIX DMC-TZ71 digital camera. Because the position of each pot in the flat and the position of each plant in the pot was not fixed in the overhead images, identically-sized squares containing each plant were manually cropped from each image and arranged into a montage for each flat using object alignment functions in Adobe Illustrator. The montages were then processed programmatically as described in the previous section by using a predefined mask to extract each plant in the image and counting green pixels.

Confirmation of viable bacteria and DNA extraction: Following plant phenotyping, whole seedlings were harvested to 2 mL screw-cap tubes (Type I, Sarstedt AG, Nümbrecht, Germany) using flame-sterilized tweezers and scissors, and kept on ice. First, one 5 mm glass

bead (Sigma) and 300 µL of PBS buffer were added to the tubes and the tubes were shaken at speed 4.0 m/s in a FastPrep 24™ 5G homogeniser for 20 s to release viable bacteria from the leaves. From this homogenate, 20 µL was directly plated on both *Sphingomonas* and *Pseudomonas* selective media to confirm viable bacteria were present at the end of the experiment (not shown). Next, 470 µL of DNA lysis buffer containing 3% SDS as described in "DNA extraction from bulk cultures and plant lysates" above was added to the remaining lysate to make a final SDS concentration of 1.88%. Garnet beads (0.5 mL) were added to the lysate and DNA was extracted by bead-beating at speed 6.0 m/s in the FastPrep 24™ 5G, and purified as described earlier.

hamPCR to determine bacterial load and composition the bacterial community: hamPCR (*60*) was performed on seedlings from soil-grown plants using primers for the *A. thaliana GIGANTEA* gene as a host gene and primers for the V4 region of the 16S rDNA (Supplementary Table 1), using primers and cycling conditions recommended in (*60*)

Data availability

All data in this manuscript are in the process of being deposited to the European Nucleotide Archive (ENA) under the project number PRJEB44136. At https://www.ebi.ac.uk/ena

Author Contributions

DSL planned the study. DSL, TLK, OSS, RdP-J, and DW collected samples. DSL, RdP-J, PP, and KP processed plant samples, made sequencing libraries, and analyzed data. AB-G helped troubleshoot the agar plate assays. IB developed algorithms for plant phenotyping from photographs. WD helped with panX analysis. DSL, RdP-J, PP, and DW wrote the manuscript, with input from all authors.

Acknowledgements

We thank Christa Lanz, Manuela Neumann, and Pablo Carbonell for assistance with Nanopore sequencing, Heike Budde for assistance with Illumina sequencing, and Haim Ashkenazy for assistance with the Nanopore genome assembly pipeline. Supported by Human Frontiers Science Program (HFSP) Long-Term Fellowships (LT000565/2015-L, D.S.L.; LT000348/2016-L, T.L.K.), ERC Advanced Grant IMMUNEMESIS (340602), the DFG through SPP Priority Program DECRyPT, and the Max Planck Society (DW).

Competing Interests

The authors declare no competing interests.

References

1. M. L. Avolio, E. J. Forrestel, C. C. Chang, K. J. La Pierre, K. T. Burghardt, M. D. Smith, Demystifying dominant species. *New Phytol.* **223**, 1106–1126 (2019).
2. Y. Zhou, H. Gao, K. A. Mihindukulasuriya, P. S. La Rosa, K. M. Wylie, T. Vishnivetskaya, M. Podar, B. Warner, P. I. Tarr, D. E. Nelson, J. D. Fortenberry, M. J. Holland, S. E. Burr, W. D. Shannon, E. Sodergren, G. M. Weinstock, Biogeography of the ecosystems of the healthy human body. *Genome Biol.* **14**, R1 (2013).
3. S. L. Heaver, E. L. Johnson, R. E. Ley, Sphingolipids in host-microbial interactions. *Curr. Opin. Microbiol.* **43**, 92–99 (2018).
4. A. G. Wexler, A. L. Goodman, An insider's perspective: Bacteroides as a window into the microbiome. *Nature Microbiology*. **2**, 17026 (2017).
5. M. T. Agler, J. Ruhe, S. Kroll, C. Morhenn, S.-T. Kim, D. Weigel, E. M. Kemen, Microbial Hub Taxa Link Host and Abiotic Factors to Plant Microbiome Variation. *PLoS Biol.* **14**, e1002352 (2016).
6. J. A. Vorholt, Microbial life in the phyllosphere. *Nat. Rev. Microbiol.* **10**, 828–840 (2012).
7. T. L. Karasov, M. Neumann, A. Duque-Jaramillo, S. Kersten, I. Bezrukov, B. Schröppel, E. Symeonidi, D. S. Lundberg, J. Regalado, G. Shirsekar, J. Bergelson, D. Weigel, The relationship between microbial biomass and disease in the Arabidopsis thaliana phyllosphere. *bioRxiv* (2019), p. 828814.
8. N. Bodenhausen, M. W. Horton, J. Bergelson, Bacterial communities associated with the leaves and the roots of Arabidopsis thaliana. *PLoS One*. **8**, e56329 (2013).
9. J. Regalado, D. S. Lundberg, O. Deusch, S. Kersten, T. Karasov, K. Poersch, G. Shirsekar, D. Weigel, 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).
10. H. Kim, M. Nishiyama, T. Kunito, K. Senoo, K. Kawahara, K. Murakami, H. Oyaizu, High population of Sphingomonas species on plant surface. *J. Appl. Microbiol.* **85**, 731–736 (1998).
11. J. Massoni, M. Bortfeld-Miller, L. Jardillier, G. Salazar, S. Sunagawa, J. A. Vorholt, Consistent host and organ occupancy of phyllosphere bacteria in a community of wild herbaceous plant species. *ISME J.* (2019), doi:10.1038/s41396-019-0531-8.
12. C. E. Morris, J. R. Lamichhane, I. Nikolić, S. Stanković, B. Moury, The overlapping continuum of host range among strains in the Pseudomonas syringae complex. *Phytopathology Research*. **1**, 4 (2019).
13. M. R. Wagner, D. S. Lundberg, T. G. Del Rio, S. G. Tringe, J. L. Dangl, T. Mitchell-Olds, Host genotype and age shape the leaf and root microbiomes of a wild perennial plant. *Nat. Commun.* **7**, 12151 (2016).
14. J. G. Wallace, K. A. Kremling, L. L. Kovar, E. S. Buckler, Quantitative Genetics of the Maize Leaf Microbiome. *Phytobiomes Journal*. **2**, 208–224 (2018).
15. K. L. Grady, J. W. Sorensen, N. Stopnisek, J. Guittar, A. Shade, Assembly and seasonality of core phyllosphere microbiota on perennial biofuel crops. *Nat. Commun.* **10**, 4135 (2019).
16. C.-J. Dong, L.-L. Wang, Q. Li, Q.-M. Shang, Bacterial communities in the rhizosphere, phyllosphere and endosphere of tomato plants. *PLoS One*. **14**, e0223847 (2019).
17. R. S. C. de Souza, V. K. Okura, J. S. L. Armanhi, B. Jorrín, N. Lozano, M. J. da Silva, M. González-Guerrero, L. M. de Araújo, N. C. Verza, H. C. Bagheri, J. Imperial, P. Arruda, Unlocking the bacterial and fungal communities assemblages of sugarcane microbiome. *Sci. Rep.* **6**, 28774 (2016).
18. T. L. Karasov, J. Almario, C. Friedemann, W. Ding, M. Giolai, D. Heavens, S. Kersten, D. S. Lundberg, M. Neumann, J. Regalado, R. A. Neher, E. Kemen, D. Weigel, Arabidopsis thaliana and Pseudomonas Pathogens Exhibit Stable Associations over Evolutionary Timescales. *Cell Host Microbe*. **24**, 168–179.e4 (2018).
19. H. C. McCann, E. H. A. Rikkerink, F. Bertels, M. Fiers, A. Lu, J. Rees-George, M. T. Andersen, A. P. Gleave, B. Haubold, M. W. Wohlers, D. S. Guttman, P. W. Wang, C. Straub, J. L. Vanneste, P. B. Rainey, M. D. Templeton, Genomic analysis of the Kiwifruit pathogen Pseudomonas syringae pv. actinidiae provides insight into the origins of an emergent plant disease. *PLoS Pathog.* **9**, e1003503 (2013).
20. T. C. Helmann, A. M. Deutschbauer, S. E. Lindow, Genome-wide identification of Pseudomonas syringae genes required for fitness during colonization of the leaf surface and apoplast. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 18900–18910 (2019).
21. H. Matsumoto, X. Fan, Y. Wang, P. Kusstatscher, J. Duan, S. Wu, S. Chen, K. Qiao, Y. Wang, B. Ma, G. Zhu, Y. Hashidoko, G. Berg, T. Cernava, M. Wang, Bacterial seed endophyte shapes disease resistance in rice. *Nat Plants*. **7**, 60–72 (2021).

22. N. C. Snelders, H. Rovenich, G. C. Petti, M. Rocafort, G. C. M. van den Berg, J. A. Vorholt, J. R. Mesters, M. F. Seidl, R. Nijland, B. P. H. J. Thomma, Microbiome manipulation by a soil-borne fungal plant pathogen using effector proteins. *Nat Plants.* **6**, 1365–1374 (2020).
23. C. Vogel, G. Innerebner, J. Zingg, J. Guder, J. A. Vorholt, Forward genetic in planta screen for identification of plant-protective traits of Sphingomonas sp. strain Fr1 against Pseudomonas syringae DC3000. *Appl. Environ. Microbiol.* **78**, 5529–5535 (2012).
24. J. L. W. Thudichum, *A Treatise on the Chemical Constitution of the Brain, etc* (Baillière & Company, 1884).
25. U. Ali, H. Li, X. Wang, L. Guo, Emerging Roles of Sphingolipid Signaling in Plant Response to Biotic and Abiotic Stresses. *Mol. Plant*. **11**, 1328–1343 (2018).
26. E. L. Johnson, S. L. Heaver, J. L. Waters, B. I. Kim, A. Bretin, A. L. Goodman, A. T. Gewirtz, T. S. Worgall, R. E. Ley, Sphingolipids produced by gut bacteria enter host metabolic pathways impacting ceramide levels. *Nat. Commun.* **11**, 2471 (2020).
27. Y. K. Yeoh, P. G. Dennis, C. Paungfoo-Lonhienne, L. Weber, R. Brackin, M. A. Ragan, S. Schmidt, P. Hugenholtz, Evolutionary conservation of a core root microbiome across plant phyla along a tropical soil chronosequence. *Nat. Commun.* **8**, 215 (2017).
28. F. O. Aylward, B. R. McDonald, S. M. Adams, A. Valenzuela, R. A. Schmidt, L. A. Goodwin, T. Woyke, C. R. Currie, G. Suen, M. Poulsen, Comparison of 26 sphingomonad genomes reveals diverse environmental adaptations and biodegradative capabilities. *Appl. Environ. Microbiol.* **79**, 3724–3733 (2013).
29. S. Asaf, M. Numan, A. L. Khan, A. Al-Harrasi, Sphingomonas: from diversity and genomics to functional role in environmental remediation and plant growth. *Crit. Rev. Biotechnol.* **40**, 138–152 (2020).
30. A. L. Khan, M. Waqas, S.-M. Kang, A. Al-Harrasi, J. Hussain, A. Al-Rawahi, S. Al-Khiziri, I. Ullah, L. Ali, H.-Y. Jung, I.-J. Lee, Bacterial endophyte Sphingomonas sp. LK11 produces gibberellins and IAA and promotes tomato plant growth. *J. Microbiol.* **52**, 689–695 (2014).
31. C. I. Carlström, C. M. Field, M. Bortfeld-Miller, B. Müller, S. Sunagawa, J. A. Vorholt, Synthetic microbiota reveal priority effects and keystone strains in the Arabidopsis phyllosphere. *Nat Ecol Evol.* **3**, 1445–1454 (2019).
32. N. C. Snelders, H. Rovenich, G. C. Petti, M. Rocafort, J. A. Vorholt, J. R. Mesters, M. F. Seidl, R. Nijland, B. Thomma, A plant pathogen utilizes effector proteins for microbiome manipulation (2020), p. 2020.01.30.926725.
33. G. Innerebner, C. Knief, J. A. Vorholt, Protection of Arabidopsis thaliana against leaf-pathogenic Pseudomonas syringae by Sphingomonas strains in a controlled model system. *Appl. Environ. Microbiol.* **77**, 3202–3210 (2011).
34. J. Regalado, D. S. Lundberg, O. Deusch, S. Kersten, T. Karasov, K. Poersch, G. Shirsekar, D. Weigel, Combining whole genome shotgun sequencing and rDNA amplicon analyses to improve detection of microbe-microbe interaction networks in plant leaves. *bioRxiv* (2019), p. 823492.
35. S. E. Lindow, M. T. Brandl, Microbiology of the phyllosphere. *Appl. Environ. Microbiol.* **69**, 1875–1883 (2003).
36. C. Vacher, A. Hampe, A. J. Porté, U. Sauer, S. Compant, C. E. Morris, The Phyllosphere: Microbial Jungle at the Plant–Climate Interface. *Annu. Rev. Ecol. Evol. Syst.* **47**, 1–24 (2016).
37. D. S. Lundberg, S. Yourstone, P. Mieczkowski, C. D. Jones, J. L. Dangl, Practical innovations for high-throughput amplicon sequencing. *Nat. Methods*. **10**, 999–1002 (2013).
38. C. R. Fitzpatrick, P. Lu-Irving, J. Copeland, D. S. Guttman, P. W. Wang, D. A. Baltrus, K. M. Dlugosch, M. T. J. Johnson, Chloroplast sequence variation and the efficacy of peptide nucleic acids for blocking host amplification in plant microbiome studies. *Microbiome.* **6**, 144 (2018).
39. C. R. Fitzpatrick, J. Copeland, P. W. Wang, D. S. Guttman, P. M. Kotanen, M. T. J. Johnson, Assembly and ecological function of the root microbiome across angiosperm plant species. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E1157–E1165 (2018).
40. K. Vanbroekhoven, A. Ryngaert, L. Bastiaens, P. Wattiau, M. Vancanneyt, J. Swings, R. De Mot, D. Springael, Streptomycin as a selective agent to facilitate recovery and isolation of introduced and indigenous Sphingomonas from environmental samples. *Environ. Microbiol.* **6**, 1123–1136 (2004).
41. F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov, BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* **31**, 3210–3212 (2015).
42. B. D. Ondov, T. J. Treangen, P. Melsted, A. B. Mallonee, N. H. Bergman, S. Koren, A. M. Phillippy, Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* **17**, 132 (2016).
43. T. Seemann, Prokka: rapid prokaryotic genome annotation. *Bioinformatics.* **30**, 2068–2069 (2014).

44. W. Ding, F. Baumdicker, R. A. Neher, panX: pan-genome analysis and exploration. *Nucleic Acids Res.* **46**, e5 (2018).
45. N. A. O'Leary, M. W. Wright, J. R. Brister, S. Ciufo, D. Haddad, R. McVeigh, B. Rajput, B. Robbertse, B. Smith-White, D. Ako-Adjei, A. Astashyn, A. Badretdin, Y. Bao, O. Blinkova, V. Brover, V. Chetvernin, J. Choi, E. Cox, O. Ermolaeva, C. M. Farrell, T. Goldfarb, T. Gupta, D. Haft, E. Hatcher, W. Hlavina, V. S. Joardar, V. K. Kodali, W. Li, D. Maglott, P. Masterson, K. M. McGarvey, M. R. Murphy, K. O'Neill, S. Pujar, S. H. Rangwala, D. Rausch, L. D. Riddick, C. Schoch, A. Shkeda, S. S. Storz, H. Sun, F. Thibaud-Nissen, I. Tolstoy, R. E. Tully, A. R. Vatsan, C. Wallin, D. Webb, W. Wu, M. J. Landrum, A. Kimchi, T. Tatusova, M. DiCuccio, P. Kitts, T. D. Murphy, K. D. Pruitt, Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–45 (2016).
46. C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T. L. Madden, BLAST+: architecture and applications. *BMC Bioinformatics.* **10**, 421 (2009).
47. N. Atamna-Ismaeel, O. Finkel, F. Glaser, C. von Mering, J. A. Vorholt, M. Koblížek, S. Belkin, O. Béjà, Bacterial anoxygenic photosynthesis on plant leaf surfaces. *Environ. Microbiol. Rep.* **4**, 209–216 (2012).
48. S. Siddaramappa, V. Viswanathan, S. Thiyagarajan, A. Narjala, Genomewide characterisation of the genetic diversity of carotenogenesis in bacteria of the order Sphingomonadales. *Microb Genom.* **4** (2018), doi:10.1099/mgen.0.000172.
49. S. Hanada, Anoxygenic Photosynthesis -A Photochemical Reaction That Does Not Contribute to Oxygen Reproduction. *Microbes Environ.* **31**, 1–3 (2016).
50. O. M. Finkel, T. O. Delmont, A. F. Post, S. Belkin, Metagenomic Signatures of Bacterial Adaptation to Life in the Phyllosphere of a Salt-Secreting Desert Tree. *Appl. Environ. Microbiol.* **82**, 2854–2861 (2016).
51. S. J. Unterholzner, B. Poppenberger, W. Rozhon, Toxin-antitoxin systems: Biology, identification, and application. *Mob. Genet. Elements* . **3**, e26219 (2013).
52. H. Li, Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics.* **34**, 3094–3100 (2018).
53. C. Vogel, N. Bodenhausen, W. Gruissem, J. A. Vorholt, 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).
54. B. Laflamme, M. M. Dillon, A. Martel, R. N. D. Almeida, D. Desveaux, D. S. Guttman, The pan-genome effector-triggered immunity landscape of a host-pathogen interaction. *Science.* **367**, 763–768 (2020).
55. X.-F. Xin, K. Nomura, K. Aung, A. C. Velásquez, J. Yao, F. Boutrot, J. H. Chang, C. Zipfel, S. Y. He, Bacteria establish an aqueous living space in plants crucial for virulence. *Nature.* **539**, 524–529 (2016).
56. R. N. Pruitt, L. Zhang, S. C. Saile, D. Karelina, K. Fröhlich, W.-L. Wan, S. Rao, A. A. Gust, F. Locci, Matthieu H A, Bart P H, J.-M. Zhou, J. L. Dangl, D. Weigel, J. E. Parker, F. El Kasmi, T. Nürnberger, Arabidopsis cell surface LRR immune receptor signaling through the EDS1-PAD4-ADR1 node. *Cold Spring Harbor Laboratory* (2020), p. 2020.11.23.391516.
57. H. Cui, E. Gobbato, B. Kracher, J. Qiu, J. Bautor, J. E. Parker, A core function of EDS1 with PAD4 is to protect the salicylic acid defense sector in Arabidopsis immunity. *New Phytol.* **213**, 1802–1817 (2017).
58. D. D. Bhandari, D. Lapin, B. Kracher, P. von Born, J. Bautor, K. Niefind, J. E. Parker, An EDS1 heterodimer signalling surface enforces timely reprogramming of immunity genes in Arabidopsis. *Nat. Commun.* **10**, 772 (2019).
59. K. Jakob, J. M. Kniskern, J. Bergelson, The role of pectate lyase and the jasmonic acid defense response in Pseudomonas viridiflava virulence. *Mol. Plant. Microbe. Interact.* **20**, 146–158 (2007).
60. D. S. Lundberg, R. de Pedro Jové, P. Pramoj Na Ayutthaya, T. L. Karasov, O. Shalev Skriptchak, K. Poersch, W. Ding, A. Bollman-Giolai, I. Bezrukov, D. Weigel, Contrasting patterns of microbial dominance in the A. thaliana phyllosphere. *bioRxiv*, p. DOI: 10.1101/2021.04.07.438366, , doi: 10.1101/2021.04.07.438366.
61. F. J. Whelan, B. Waddell, S. A. Syed, S. Shekarriz, H. R. Rabin, M. D. Parkins, M. G. Surette, Culture-enriched metagenomic sequencing enables in-depth profiling of the cystic fibrosis lung microbiota. *Nat Microbiol.* **5**, 379–390 (2020).
62. Y. Bai, D. B. Müller, G. Srinivas, R. Garrido-Oter, E. Potthoff, M. Rott, N. Dombrowski, P. C. Münch, S. Spaepen, M. Remus-Emsermann, B. Hüttel, A. C. McHardy, J. A. Vorholt, P. Schulze-Lefert, Functional overlap of the Arabidopsis leaf and root microbiota. *Nature.* **528**, 364–369 (2015).
63. M. Ayliffe, C. K. Sørensen, Plant nonhost resistance: paradigms and new environments. *Curr. Opin. Plant Biol.* **50**, 104–113 (2019).

64. R. Panstruga, M. J. Moscou, What is the Molecular Basis of Nonhost Resistance? *Mol. Plant. Microbe. Interact.* **33**, 1253–1264 (2020).
65. R. Cai, J. Lewis, S. Yan, H. Liu, C. R. Clarke, F. Campanile, N. F. Almeida, D. J. Studholme, M. Lindeberg, D. Schneider, M. Zaccardelli, J. C. Setubal, N. P. Morales-Lizcano, A. Bernal, G. Coaker, C. Baker, C. L. Bender, S. Leman, B. A. Vinatzer, The plant pathogen Pseudomonas syringae pv. tomato is genetically monomorphic and under strong selection to evade tomato immunity. *PLoS Pathog.* **7**, e1002130 (2011).
66. K. Yoshida, V. J. Schuenemann, L. M. Cano, M. Pais, B. Mishra, R. Sharma, C. Lanz, F. N. Martin, S. Kamoun, J. Krause, M. Thines, D. Weigel, H. A. Burbano, The rise and fall of the Phytophthora infestans lineage that triggered the Irish potato famine. *Elife.* **2**, e00731 (2013).
67. N. Rohland, D. Reich, Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Res.* **22**, 939–946 (2012).
68. N. Caruccio, Preparation of next-generation sequencing libraries using Nextera™ technology: simultaneous DNA fragmentation and adaptor tagging by in vitro transposition. *Methods Mol. Biol.* **733**, 241–255 (2011).
69. A. Bankevich, S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, S. I. Nikolenko, S. Pham, A. D. Prjibelski, A. V. Pyshkin, A. V. Sirotkin, N. Vyahhi, G. Tesler, M. A. Alekseyev, P. A. Pevzner, SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
70. B. J. Walker, T. Abeel, T. Shea, M. Priest, A. Abouelliel, S. Sakthikumar, C. A. Cuomo, Q. Zeng, J. Wortman, S. K. Young, A. M. Earl, Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One.* **9**, e112963 (2014).
71. H. den Bakker, *hcdenbakker - Overview* (Github; https://github.com/hcdenbakker).
72. S. Koren, A. M. Phillippy, One chromosome, one contig: complete microbial genomes from long-read sequencing and assembly. *Curr. Opin. Microbiol.* **23**, 110–120 (2015).
73. R. R. Wick, L. M. Judd, K. E. Holt, Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol.* **20**, 129 (2019).
74. H. Li, Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics.* **32**, 2103–2110 (2016).
75. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv [q-bio.GN]* (2013), (available at http://arxiv.org/abs/1303.3997).
76. B. Buchfink, C. Xie, D. H. Huson, Fast and sensitive protein alignment using DIAMOND. *Nat. Methods*. **12**, 59–60 (2015).
77. H. Jiang, R. Lei, S.-W. Ding, S. Zhu, Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics.* **15**, 182 (2014).
78. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* **25**, 2078–2079 (2009).
79. C. Jain, L. M. Rodriguez-R, A. M. Phillippy, K. T. Konstantinidis, S. Aluru, High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* **9**, 5114 (2018).
80. G. R. Warnes, B. Bolker, L. Bonebakker, R. Gentleman, W. Huber, A. Liaw, T. Lumley, M. Maechler, A. Magnusson, S. Moeller, M. Schwartz, B. Venables, gplots: Various R Programming Tools for Plotting Data (2020), (available at https://CRAN.R-project.org/package=gplots).
81. C. A. Barragan, R. Wu, S.-T. Kim, W. Xi, A. Habring, J. Hagmann, A.-L. Van de Weyer, M. Zaidem, W. W. H. Ho, G. Wang, I. Bezrukov, D. Weigel, E. Chae, RPW8/HR repeats control NLR activation in Arabidopsis thaliana. *PLoS Genet.* **15**, e1008313 (2019).