Supplementary Methods for Maignien et al.

Plant germination and growth

Seeds of Arabidopsis thaliana Col-0 and gl-1 (Lehle Seeds) were surface-sterilized with a 2 minute wash in 70% EtOH, a 20 minute wash in 50% bleach, and rinsed 10X in sterile MilliQ H_2O. Phytoagar plates for seed germination were prepared using 1x standard MS medium with 3% sucrose and 0.7% agar, pH 5.7. Seeds were placed on phytoagar plates in a laminar flow hood, and plates were incubated at 4 C in the dark for 72 hours. Plates were then transferred to a growth chamber at 22 C and incubated for 12 days on a 16 hour light/8 hour dark cycle with a PAR level of ~ 160 µmol $m^{-2}s^{-1}$ generated by four 6500K T5 fluorescent bulbs. Hypocotyl and cotyledon emergence occurred after 6 days of incubation for a majority of seeds. After 12 days incubation, most plants had 2-4 rosette leaves and were ready for transfer to soil.

PMP Arabidopsis medium soil (Lehle Seeds) was sterilized as described (Wolf, Dao, Scott, & Lavy, 1989), moistened with autoclaved ddH_2O, and placed in EtOH and UV-treated 6-welled plastic planting inserts. During transfer of seedlings to soil, we used elbow length laboratory gloves, face masks, and repeated sterilization of tools with EtOH to minimize possible contamination from human associated microbes. Plants were watered every three days using 1.5 L sterile MilliQ water placed into each tray and drained after 10-15 minutes to prevent algal growth. We fertilized plants with a sterile 200 ppm solution of 20-10-20 Jack's General Purpose Fertilizer (JR Peters) once a week starting on day 27. Visual observations of soil indicated that the soil surface remained free of algal contamination throughout the experiment.

Data on temperature, RH, and PAR was collected at 1-minute intervals throughout the entire experiment using HOBO sensors (Onset Computer) (10/18/11-12/13/2011; Figure S1). Temperatures were held mostly constant throughout the experiment at 20 C, with the exception of periodic increases due to AC failure during days 21-27. Maximum daytime light levels ranged from 36.2 to 686.2 μmol $m^{-2}s^{-1}$. Day length ranged from ~10 hours at the start of the experiment to ~8 hours at the end.

Sampling airborne microbes

We captured airborne microbes using standard microscope slides coated with the biological adhesive CellTak (BD Biosciences) at a density of 3.5 μg/cm^2. Slides were sterilized with 100% EtOH prior to coating and dried in a laminar flow hood. We constructed small platforms to hold slides at the level of plant leaves. Slides were left in place during each sampling interval (~5 days), collected, and replaced at the next sampling event. In preliminary experiments, fluorescent microscopy using DAPI staining indicated the presence of intact cells on slides exposed to air for ~1 week periods (data not shown). Additionally, control experiments with live E. coli added to slides indicated that cells remained intact for ~1 week (data not shown). Slides were incubated with trypsin to remove CellTak and detach microbes; the wash solution was collected on sterile filters and DNA was extracted using the Biostic Bacteremia DNA extraction kit (MoBio Labs). Due to the low biomass present in air, we used rigorous sterile technique and extensive negative controls to exclude contaminant microbes (see Results).

Experimental CellTak slides collected from the greenhouse were placed in sterile 50 ml eppendorf tubes in the greenhouse and transported on ice to the laboratory.

Each slide was incubated in 30 ml of sterile 0.5% trypsin (in NaCl/EDTA; Invitrogen) in a rotary shaker/incubator at 37 C, 200 rpm for 2 hours. In the laminar flow hood, slides were scraped with sterile, disposable cell scrapers and washed with TE to remove and collect CellTak from the surface. Slides were then stained with a solution of 0.1% w/v Coomassie blue to verify removal of CellTak. The trypsin/buffer/CellTak solution was filtered onto sterile 0.22 micron filters using 25 mm Swinnex cartridges. Each filter was placed in a bead tube provided in the Biostic Bacteremia DNA extraction kit (MoBio Labs) and kept at -80 C until extraction following kit protocols.

Phyllosphere sampling

Three whole Col-0, three whole gl-1 plants, and six individual leaves from one Col-0 and one gl-1 plant were sampled at each of 11 timepoints (gl-1 and individual leaf data not shown). Phyllosphere microbes were washed from leaves using gentle bath sonication in a solution of 0.2% Silwet in TE (10 mM Tris, 1mM EDTA, pH 7). Silwet is a surfactant originally used in Arabidopsis transformation (1) that penetrates into the apoplast and is effective in removing bacteria from leaves (2, 3) Each plant was removed from soil and roots and flowering stems were carefully trimmed off. Rosettes and individual leaves were separately placed into 50 ml Falcon tubes containing 15-30 mL Silwet solution (depending on rosette size), and bath sonicated for 10 minutes. After sonication, each rosette was removed and retained for drying and subsequent weighing. Rosettes were dried at 70 C overnight and weighed on a high precision scale. We converted weights to area through imaging 12 wet leaves, calculating surface area with ImageJ, and subsequently drying and weighing the imaged leaves. There was a strong linear relationship between area and weight

(R^2=0.95). Additionally, we sampled ~0.25g of soil surrounding each sampled plant and stored at -80 C (data not shown).

In the laboratory, Silwet solutions containing phyllosphere microbes were vortexed at high speed for ~20 seconds to break up biofilms. Solutions were then pre-filtered through 5 micron filters to remove plant material and captured on 0.22 micron filters using sterile Swinnex cartridges. Filters were placed in MoBio Biostic Bacteremia kit bead tubes and stored at -80 C until extraction. Extraction procedures followed kit protocols with slight modifications. First, tubes with buffer CB1 were incubated at 70 C in a rotating hybridization oven. Second, final elution was performed by incubating columns with 25 ul of solution CB5 for 5 min, centrifuging at 10,000 rpm, then repeating with another 25 ul of solution CB5 to yield 50 ul of extracted DNA. DNA was quantified using PicoGreen on a Turner fluorometer.

Quantitative PCR

A SYBR green qPCR assay was developed for quantification of bacterial 16S rRNA copy number among rosette leaf washes using a modified 16S rRNA primer that limits chloroplast 16S amplification (4). We reverse complemented this primer to amplify the v6 region and increased the degeneracy (783F: 5'-GBCCTAATCTATGGGVCCATC-3'). A 10-fold serially diluted, 5 point standard curve (range $3x10^3$ – $3x10^7$) was generated with plasmids containing 16S rRNA 783F/1046R inserts from 9 microbial species of varying GC content. Plasmids were generated with the CloneJet PCR cloning kit (Fermentas, Glen Burnie, MD) and NEB competent cells (Ipswich, MA) and purified with the MoBio Mini Plasmid Prep Kit. Purified plasmids were linearized with XbaI restriction enzyme (NEB) at 37ºC for 4

hours and purified with the MoBio UltraClean PCR Clean-up Kit. Linearized plasmids were quantified with a Nanodrop spectrophotometer, pooled in equimolar amounts and stored in single use aliquots at -80ºC. Aliquots were quantified a second time with PicoGreen immediately prior to use, converted into units of copy number per microliter and serially diluted for the standard curve.

The standard curve, environmental samples and NTCs were run in triplicate using 1X Kapa SYBR FAST ABI Prism qPCR master mix, 0.2 µM of each primer and 2ul of template DNA in a 20ul total volume. Cycling was performed on a StepOnePlus Real-Time PCR System (Life Technologies) software version 2.2 using the following 3-step PCR cycling conditions: 95ºC for 3min, 40 cycles of 95ºC for 3sec, 58ºC for 20sec, 72ºC for 10sec followed by a standard melt curve.

No template control (NTC) amplification of approximately 300 copies from trace bacterial DNA in qPCR reagents requires that low biomass samples <3,000 copies be omitted from the assay to allow a 10-fold difference between sample C_q values and NTCs. The average C_q for triplicate NTCs was 33.67 cycles with a standard deviation of 0.15, as expected for general microbial primers, which can amplify trace contamination in reaction reagents. A single melt curve peak was observed at approximately 85ºC for all samples and standards.

Amplification and 454 sequencing of the v4v6 region of 16S rRNA

The v4v6 variable region of 16S rRNA was amplified in triplicate with 454 fusion primers containing adapters and barcodes with bacterial primer sequences 518F (5'-CCAGCAGCYGCGGTAAN-3') and 1046R (5'-CGACRRCCATGCANCACCT-3') as previously described (5). Primer sequences flank the v4v6 hypervariable region. Reaction mixes included 13 units of Platinum Taq High Fidelity, 1X HiFi PCR Buffer,

3.7mM $MgSO_4$ (Life Technologies, Grand Island, NY), 200μM PurePeak dNTPs (Thermo Fisher, Waltham, MA) and 0.2μm of each forward and reverse primer in a 125μl reaction volume. 25μl was removed for a negative water control. 4μl of gDNA was added to the remaining 100ul and distributed among 3 separate PCR tubes (34.7ul each). Cycling conditions included initial denaturation at 94ºC for 3min, followed by 30 cycles of 94ºC for 30sec, 60ºC for 45sec and 72ºC for 1min with a final 2-min extension at 72ºC. Products were pooled after cycling, visualized on a Bioanalyzer 2100 using a DNA 1000 LabChip (Agilent, Palo Alto, CA) and purified with Agencourt AMPure XP beads (Beckman Coulter, Danvers, MA) with a 0.75x ratio of beads:DNA using 70% ethanol washes. Purified amplicons were eluted in variable volumes of 10mM Tris-Cl, pH 8.0 depending on Bioanalyzer amplicon concentration. Purified products were run a second time on a DNA 1000 LabChip to ensure removal of excess primer. Final products were quantified with Quant-it PicoGreen (Invitrogen) using the TBS-380 Mini-Fluorometer (Turner Biosystems) prior to equimolar library pooling for emPCR.

Bioinformatic analyses

Processing and filtering of v4v6 pyrosequencing reads was carried out using a standard MBL pipeline (5, 6). Lower quality reads with ambiguous bases (N's), mismatch to the proximal primer, or average Q score < 30 were discarded. Chimeras detected by UChime (7) using either the "de novo" or "reference database" methods were also removed from the dataset. Remaining higher quality reads were classified with GAST (6) based on the SILVA 105 database (8, 9).

Per-base quality declined toward the end of reads (v4 region). We thus used a custom python script to trim off all reads downstream of the conserved sequence:

G[T,G]AG[A,G]GT[A,G][A,G]AAT (“anchor350”), located between the V5 and V4 regions. Subsequent analyses were carried out on this high-quality subset of sequences containing only the v5v6 hypervariable region. These sequences were clustered into operational taxonomic units (OTUs) using a 97% similarity cutoff with the Usearch 6 software (10): sequences were denoised by clustering at 99% similarity, then representative sequences were further clustered at 97% similarity. Representative sequences from the 97% clusters were used to define final OTU representative sequences. Individual reads of the high-quality v5v6 dataset were then aggregated around these OTU representatives using a 97% similarity cutoff with the Usearch software in “reference database” mode.

A Sample-OTU abundance matrix was computed based on the results of this clustering step and passed to QIIME software (11) for calculation of alpha-diversity and beta-diversity indices and principal coordinate analysis. For alpha-diversity in each sample, the non-parametric Chao1 index was calculated as the average of Chao1 values resulting from 100 independent rarefactions. The sensitivity of this index to sub-sampling was tested. The matrix was also passed to the Catchall Software (12) (via the Mothur interface (13)) to compute parametric richness in each sample. For calculation of beta-diversity indices, singleton and doubleton OTUs were removed and the abundance matrix was subsampled to the number of reads present in the smallest library (895 reads).The *vegan* package in R (14) was used for PERMANOVA tests on distance matrices. The robustness of beta-diversity indices (binary Jaccard, Morisita-Horn, weighted and unweighted UNIFRAC) to sub-sampling was also confirmed with a jackknife analysis (100 iterations). The Sample-OTU abundance matrix was also used to determine biomarker OTUs and taxonomic lineages with the LEfSE software (15). We sought to identify biomarker OTUs

separating day 60+ air and day 60+ plant samples as well as tray 1 versus tray 2 plant samples.

We used oligotyping (16) (oligotyping.org) to identify significant sub-OTU level variation. Trimmed reads were aligned using PyNAST (17) to the Greengenes (18) multiple sequence alignment template (10/6/2010 release). We generated separate alignments for each of the 20 most abundant genera. Following the initial identification of highly variable nucleotide positions in these alignments using Shannon entropy, we let the oligotyping pipeline generate short, unique oligomers ("oligotypes") for each read by merging nucleotides from manually selected variable positions to identify closely related but distinct groups of reads in each genus. To reduce noise, we required that the most abundant unique sequence in each oligotype be present at a minimum abundance of 20. Finally, we analyzed the abundance patterns of these oligotypes across sample sets.

References Cited

1. **Clough SJ, Bent AF.** 1998. Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. The Plant journal : for cell and molecular biology **16:**735-743.
2. **Delmotte N, Knief C, Chaffron S, Innerebner G, Roschitzki B, Schlapbach R, von Mering C, Vorholt JA.** 2009. Community proteogenomics reveals insights into the physiology of phyllosphere bacteria. Proceedings of the National Academy of Sciences of the United States of America **106:**16428-16433.
3. **Tornero P, Dangl JL.** 2001. A high-throughput method for quantifying growth of phytopathogenic bacteria in Arabidopsis thaliana. The Plant journal **28:**475-481.
4. **Sakai M, Matsuka A, Komura T, Kanazawa S.** 2004. Application of a new PCR primer for terminal restriction fragment length polymorphism analysis of the bacterial communities in plant roots. Journal of microbiological methods **59:**81-89.
5. **Filkins LM, Hampton TH, Gifford aH, Gross MJ, Hogan Da, Sogin ML, Morrison HG, Paster BJ, O'Toole Ga.** 2012. The Prevalence Of Streptococci And Increased Polymicrobial Diversity Associated With Cystic Fibrosis Patient Stability. Journal of bacteriology.
6. **Huse SM, Dethlefsen L, Huber JA, Mark Welch D, Relman DA, Sogin ML.** 2008. Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. PLoS genetics **4:**e1000255.
7. **Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R.** 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics (Oxford, England) **27:**2194-2200.
8. **Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner FO.** 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic acids research **41:**D590-596.
9. **Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glockner FO.** 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic acids research **35:**7188-7196.
10. **Edgar RC.** 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics **26:**2460-2461.
11. **Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R.** 2010. QIIME allows analysis of high-throughput community sequencing data. Nature methods **7:**335-336.
12. **Bunge J.** 2011. Estimating the number of species with catchall. Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing**:**121-130.
13. **Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B,**

Thallinger GG, Van Horn DJ, Weber CF. 2009. Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. Applied and environmental microbiology **75:**7537-7541.

14. **Oksanen J.** 2011. vegan: Community Ecology Package. R package version 1.17-9. *In* Blanchet FG (ed.).
15. **Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, Huttenhower C.** 2011. Metagenomic biomarker discovery and explanation. Genome biology **12:**R60.
16. **Eren AM, Zozaya M, Taylor CM, Dowd SE, Martin DH, Ferris MJ.** 2011. Exploring the diversity of Gardnerella vaginalis in the genitourinary tract microbiota of monogamous couples through subtle nucleotide variation. PloS one **6:**e26732.
17. **Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R.** 2010. PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics (Oxford, England) **26:**266-267.
18. **McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, Andersen GL, Knight R, Hugenholtz P.** 2012. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. The ISME journal **6:**610-618.