

1 **Leaf bacterial community structure and variation in wild ruderal plants are shaped by the interaction**  
2 **of host species and defense chemistry with environment.**

3 Teresa Mayer<sup>1</sup>, Michael Reichelt<sup>2</sup>, Jonathan Gershenzon<sup>2</sup>, Matthew Agler<sup>1</sup>

4

5 <sup>1</sup>Plant Microbiosis Lab, Department of Microbiology, Friedrich-Schiller University, Jena, Germany

6 <sup>2</sup>Max-Planck Institute for Chemical Ecology, Department of Biochemistry, Jena, Germany

7

8

9 **\* Corresponding author:**

10 Matthew T. Agler

11 Department of Microbiology, Plant Microbiosis Lab

12 Friedrich Schiller University of Jena

13 Neugasse 23

14 07743 Jena, Germany

15 E-mail: [matthew.agler@uni-jena.de](mailto:matthew.agler@uni-jena.de)

16 Tel: +49 (0)3641 9 49980

17

18

19

## 1 **Summary**

2 Variable phenotypes help plants ensure fitness and survival in the face of unpredictable environmental  
3 stresses. Leaf bacteria (bacteriomes) can extend plant phenotypes and are well-known to vary from  
4 one plant to the next, but little is known about controls on this variation. Here, we find in 9 populations  
5 of *Arabidopsis thaliana* that core leaf bacteriomes are largely, but not completely, shared with other  
6 ground-dwelling ruderal plant species. Strength of differentiation between plant species and between  
7 *A. thaliana* populations shifts from year to year, becoming stronger when plants within populations  
8 have more similar leaf bacteriomes (less plant-to-plant variation or stochasticity). Interestingly, across  
9 different populations, plants with shared leaf aliphatic glucosinolate chemotypes exhibited similar  
10 strong year-to-year stochasticity shifts. Therefore, stochasticity of leaf bacteriomes in plant  
11 populations changes in specific ways and might be controlled by plant traits, with important  
12 implications for how plants adapt to complex and shifting environments.

## 1 Introduction

2 Plants are colonized by diverse organisms including bacteria, fungi and protists. These microorganisms  
3 not only use plants as a primary resource but can provide important functions in return. Besides well-  
4 known beneficial fungal-plant relationships, such as arbuscular mycorrhiza, other microorganisms like  
5 bacteria can play critical roles in plant survival. For example, bacteria are required for plants to survive  
6 detrimental effects of normal soil fungi and oomycetes (Duran et al., 2018). On leaves, bacteria can  
7 function for example as barriers to the constant threat posed by pathogen invasion (Berg and Koskella,  
8 2018; Ritpitakphong et al., 2016). Thus, much effort is put into understanding how leaf bacterial  
9 communities (bacteriomes) are shaped to help develop new, sustainable approaches to agriculture  
10 and biodiversity management.

11 Compared to other organs, leaf bacteriomes are still relatively poorly understood. Different plant  
12 species and plant genotypes tend to have overlapping leaf microbiota on the phylum level with  
13 differences at finer levels (Knief et al., 2010; Massoni et al., 2020). Similar to rhizosphere colonizers,  
14 the primary source of the leaf microbiota is thought to be soil (Massoni et al., 2021; Tkacz et al., 2020),  
15 but vertical colonization and deposition from air could also play important roles (García-Suárez et al.,  
16 2017; Maignien et al., 2014). Selection of bacterial plant colonizers from complex inocula is influenced  
17 to some extent by factors like host genetics and host age (Bodenhausen et al., 2014; Wagner et al.,  
18 2016) and much of this selection may happen early in the plant life cycle since early colonizers seem  
19 to be very robust to later invasion (Carlstrom et al., 2019; Leopold and Busby, 2020). In nature,  
20 however, leaf microbiomes are not necessarily completely stable and shifts across the life cycle of  
21 plants are well-documented (Williams et al., 2013). Indeed, leaves are highly dynamic, exposed  
22 environments and both abiotic factors like temperature, humidity and UV (Aydogan et al., 2020) and  
23 biological factors like pathogens can strongly shape communities (Agler et al., 2016). Hosts can also  
24 strongly interact with these factors, resulting in strong interaction effects between plant host factors  
25 and environments on leaf microbiome structures (Wagner et al., 2016).

26 Some threats to plants that can be mitigated with help from the microbiome are likely to be endemic,  
27 specific to the locations and soils where the plants grow (Agler et al., 2016). On the other hand, many  
28 selective pressures on a given plant population are not necessarily constant and in nature can vary in  
29 magnitude and even direction, remaining unpredictable from one year to the next (Postma and Ågren,  
30 2018). Leaf bacteria can help plants to adapt and thrive in complex environments not just by improving  
31 certain plant phenotypes but also by making phenotypes more plastic and variable (Hawkes et al.,  
32 2021; Petipas et al., 2021). Therefore, plants should be expected not only to structure leaf bacteriomes  
33 in specific ways to address certain threats, but also to manage community variation to help ensure  
34 survival in the face of unpredictability. Consistent with this, we observed that *A. thaliana* leaf  
35 bacteriomes planted in three years in a common garden consistently converged toward a specific  
36 community structure, but with significant plant-to-plant variation (Almario et al., 2021). An important  
37 but largely unaddressed question, then, is not just how leaf bacteriomes are shaped into specific  
38 structures, but also how *variation* arises and what factors might control it.

39 Some of these questions about how microorganisms and plants collaborate to adapt to complex,  
40 changing natural environments can be addressed by studying wild plant populations. Persistent and  
41 numerous wild populations of *A. thaliana* can be found throughout Europe (Thiergart et al., 2020). In  
42 much of Europe, *A. thaliana* populations grow together with other diverse ruderal species and *A.*  
43 *thaliana* itself can have a high trait diversity even on small spatial scales (Katz et al., 2021); this diversity  
44 is so far underutilized in microbiome research but can be leveraged to understand how host factors  
45 shape colonization in nature. In this study, we begin to address some of these questions by looking  
46 into 13 distinct *A. thaliana* populations in Jena, Germany, all within a 4km radius. By sampling whole  
47 leaf bacteriomes of both *A. thaliana* and other sympatric plants across multiple years, we asked how

1 they differ between locations and between host species and whether these effects are consistent  
2 across time. We identify stochasticity of leaf bacteriomes as an important aspect defining whether leaf  
3 bacteriomes are distinguishable between host species or between the populations. Interestingly, this  
4 stochasticity changes between populations and over years and is potentially linked to *A. thaliana* leaf  
5 chemotypes.

## 1 **Material & Methods**

### 2 Isolation and characterization of wild *A. thaliana* population lines

3 *Overview.* In spring 2018, we identified 13 wild *A. thaliana* populations in Jena, Germany (**Figure S1A**).  
4 Plants in the populations generally occupy similar niches – growth in sandy soil around or between  
5 cobblestones or sidewalks. We collected leaf samples of nine plants growing naturally in each local  
6 population to extract DNA to infer genotypic heterogeneity using microsatellite length polymorphisms.  
7 Additionally, around five plants were collected from each population, re-potted and grown in the lab  
8 to collect seeds. For isolated lines of seven of the populations, we characterized leaf glucosinolate (GL)  
9 profiles and later confirmed them in a second batch after plants had been selfed to generate mostly  
10 homozygous lines. Additionally, we selected five populations (see above) with completely uniform  
11 microsatellite profiles (NG2 - Neugasse2, PB - Paradiesbahnhof, SW1 - Sandweg1, JT1 - Johannistor1,  
12 Woe - Wöllnitzer) and characterized plant development times in the selfed lines. Seeds of four of the  
13 populations were donated to the NASC seed database and given new names; seeds of the population  
14 PB will be donated as soon as possible. An overview of the data available and NASC accession numbers  
15 for each population is provided in **Table S2**.

16 *Plant DNA extraction of leaf samples.* We rapidly extracted DNA of the nine plants from each local plant  
17 population as well as the control lines Col-0 and Ws-0 (Edwards et al., 1991). We crushed one leaf with  
18 a metal pestle, added 400  $\mu$ L extraction buffer (200 mM Tris HCl pH 8.0, 250 mM NaCl, 25 mM EDTA,  
19 0.5% SDS) and vortexed for 5 s. We centrifuged the samples at 19,000 x g for 1 min, transferred the  
20 supernatant to a fresh tube and added 300  $\mu$ L of isopropanol. Samples were incubated at RT for 2 min.  
21 We centrifuged the samples again and removed the supernatant. After the DNA was air-dried, we  
22 dissolved the pellets in 100  $\mu$ L TE-Buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0).

23 *Microsatellite diversity.* We assessed the intra- and interspecies diversity of the plant population by  
24 the diversity of three microsatellite loci amplified with the primer pairs nag59F/R, nag111F/R,  
25 nag158F/R (**Table S1**) (Bell and Ecker, 1994). We mixed 5-100 ng of template plant DNA in a 50  $\mu$ L  
26 reaction volume, containing 10x Buffer B, 25 mM MgCl<sub>2</sub> and 0.5  $\mu$ L of Taq Polymerase (Biodeal  
27 Handelsvertretung Edelman e.K., Leipzig), as well as 10mM dNTP Mix (Carl Roth GmbH & Co. KG) and  
28 10  $\mu$ M of each primer of the corresponding primer pair. Products were amplified with the following  
29 program: 94°C for 3 min, 40 cycles of 94°C for 15 s, 55°C for 15 s and 72°C for 30 s, followed by a final  
30 elongation of 72°C for 3 min. Samples were stored at 4°C until we separated them on 4% Agarose gels  
31 (Biodeal Handelsvertretung Edelman e.K., Leipzig). We identified the size of the fragments via agarose  
32 gel electrophoresis and created a phylogenetic tree with this data.

33 *Growth rates and flowering times.* We sowed *A. thaliana* seeds (NG2, PB, SW1, Woe, JT1 and Col-0) in  
34 individual pots on soil (made by mixing 4 L Florador Anzucht soil, 2 L Perligran Premium, 25 g Substral  
35 fertilizer and 2 L tap water) and vernalized them for three days at 4°C in the dark before they were set  
36 in the plant growth chamber (PolyKlima, Freising, Germany; growth conditions: 24°C/18°C, 16h/8h  
37 day/night cycle at 75% light intensity). For each *A.thaliana* population we randomly chose three of the  
38 five plants collected in the wild and used their offspring for the assay. Of each plant three individuals  
39 were used, which means that of each plant population we observed nine individuals (n=9). We  
40 checked, watered and took pictures of the plants every other day. After 10 days in the plant incubator,  
41 we pricked three individual plants of each population and moved them to a new pot with fresh soil.  
42 Until the plants grew to senescence we continued to check and water them every other day. We  
43 measured the following data: continuous measurements (rosette radius, length of stem), timepoint  
44 measurements (flower bud visible for first time, first flower opens, first silique appears) and continuous  
45 counting (number of true leaves). Based on these results we assigned growth stages adapted from  
46 (Boyes et al., 2001). Seed germination refers to Principal Growth stage 0 and includes seed imbibition,

1 radicle emergence and hypocotyl and cotyledon emergence. Leaf development (Principal Growth stage  
2 2) and Rosette growth (Principal Growth stage 3) occur simultaneously most of the time (more leaves  
3 develop while the plants also grow), with rosette growth continuing for a while after the last leaves  
4 have developed. The stage Inflorescence emergence (Principal Growth stage 5) begins with the  
5 development of a flowering bud and ends when the first flower developed, which is the begin of the  
6 Flower development (Principal Growth stage 6). We plotted the data as the mean values of nine  
7 individual plant observed of each plant location. The plant population PB needs a second vernalization  
8 period for two weeks at 4°C to develop flowering buds. The local population JT1 did not germinate  
9 consistently in the lab and was not included in this assay.

10 *Leaf glucosinolate analysis.* We measured GL profiles of the plants similarly to previously described  
11 (Kliebenstein et al., 2001). Four 5-week-old plants of each population were harvested by removing  
12 roots and flower stems as close to the rosette as possible. The whole rosettes were transferred into 50  
13 mL tubes, frozen in liquid nitrogen and kept at -80°C until further processing. After freeze-drying for  
14 three days, rosettes were grinded to powder by shaking with metal beads for 2 min. GLs were extracted  
15 from 8-12 mg of the powder by 1 mL of 80% MeOH with 50 µM pOH-BGI (Sinalbin, internal standard)  
16 and thorough vortexing for 30 sec and shaking for 30 min in a paint shaker shaker (Skandex SO-10M,  
17 Fluid Management Europe, The Netherlands) at high speed. The samples were centrifuged for 5 min  
18 at 3,200 rpm and 600 µL of the supernatant was added to a freshly prepared DEAE-Sephadex-filter  
19 plate and allowed to flow through. Samples on the DEAE-Sephadex were washed five times with: (1.)  
20 0.5 mL 80% MeOH, (2./3.) Two times 1 mL ultrapure water, (4.) 0.5 mL 0.02 MES pH 5.2 and (5.) 30 µL  
21 sulfatase (Sulfatase from *Helix pomatia*; Sigma) prepared according to Graser et.al. (Graser et al.,  
22 2000). Samples were kept at room temperature overnight and desulfo-GLs were eluted with 0.5 mL dd  
23 water on the next day. Desulfo-GLs were then analysed by HPLC-UV. Fifty microliters of the desulfo-GL  
24 extract was run on a reversed phase column (Nucleodur Sphinx RP18, 250 x 4.6mm, 5µm, Macherey-  
25 Nagel, Düren, Germany) on an Agilent 1100 HPLC with a water-acetonitrile gradient (1.5% acetonitrile  
26 for 1min, 1.5 to 5% acetonitrile from 1 to 6 min, 5 to 7% acetonitrile from 6 to 8 min, 7 to 21%  
27 acetonitrile from 8 to 18 min, 21 to 29% acetonitrile from 18 to 23 min, followed by a washing cycle;  
28 flow 1.0 mL min<sup>-1</sup>). Detection was performed with a photodiode array detector and peaks were  
29 integrated at 229 nm. We used the following response factors: 3OHP and 4OHB 2.8, all other aliphatic  
30 glucosinolates 2.0, indole glucosinolates 0.5 (Burow et al., 2006) for quantification of individual  
31 glucosinolates. The identity of the peaks was based on a comparison of retention time and UV  
32 absorption spectrum with data obtained for isolated desulfo glucosinolates as described in Brown et.al  
33 (Brown et al., 2003) and by analysis of the desulfoGLs extracts on an LC-ESI-Ion-Trap-mass  
34 spectrometer (Esquire6000, Bruker Daltonics). The following glucosinolates were detected in the  
35 samples: 3-hydroxypropyl glucosinolate (3OHP), 4-hydroxybutyl glucosinolate (4OHB), 3-  
36 methylsulfinylpropyl glucosinolate (3MSOP), 4-methylsulfinylbutyl glucosinolate (4MSOB), 2-propenyl  
37 (allyl), 2-hydroxy-3-butenyl glucosinolate (2OH3But), 3-butenyl glucosinolate (3-Butenyl), 4-pentenyl  
38 glucosinolate (4-Pentenyl), 4-hydroxy-indol-3-ylmethyl glucosinolate (4OHI3M), 4-methylthiobutyl  
39 glucosinolate (4MTB), 7-methylsulfinylheptyl glucosinolate (7MSOH), 8-dmethylsulfinyloctyl  
40 glucosinolate (8MSOO), indol-3-ylmethyl glucosinolate (I3M), 4-methoxy-indol-3-ylmethyl  
41 glucosinolate (4MOI3M), and 1-methoxy-indol-3-ylmethyl glucosinolate (1MOI3M). Results are given  
42 as µmol per g dry weight. Each genotype was tested in four replicates.

43 To construct the GLs diversity tree, we first filtered out samples with a weight less than 1 mg. The rest  
44 of the samples were normalized to the plant weight and combined with the normalized data of a few  
45 genotypes from (Kliebenstein et al., 2001) as references. We combined the datasets of both GL  
46 analyses and constructed the tree in R using the Neighbor Joining method. In addition, we plot the  
47 mean concentrations of each GL.

1 Culture-independent characterization of bacterial diversity in wild plant populations.

2 *Sampling.* We sampled the nine populations in February 2019 and 2020, where each sample is made  
3 up of 3-4 leaves of one plant, collected with sterilized tweezers (6 replicates per population per time  
4 point). Additionally, we sampled six other plants of varying unknown genotypes near each *A. thaliana*  
5 plant. In March of 2019 and 2020 we collected *A. thaliana* and other plants from the five selected  
6 populations NG2, PB, SW1, JT1 and Woe. We washed all plant samples once with autoclaved ultrapure  
7 water to remove dirt and stored them on ice for the remaining sampling time and at -80°C back in the  
8 lab. In March 2020 additionally five samples of the soil at the five locations were taken. The upper layer  
9 of soil (roughly 1 cm) was removed with a sterilized spatula and then the underlaying soil was scooped  
10 in a clean Eppendorf tube. Soil samples were also placed on ice for remaining sampling time and at -  
11 80°C back in the lab until further processing.

12 *DNA extraction of bacterial communities associated with the wild plant populations.* For DNA  
13 extraction we transferred the plant material to a 2 mL screw-cap tube with two metal beads (3 mm)  
14 and 0.2 g of 0.25-0.5 mm glass beads (Carl Roth GmbH & Co. KG) to crush plant and bacterial cells.  
15 Samples were bead-beat two times (30 s at 1,400 rpm on a BioSpec mini bead beater 96) (Biospec  
16 Products, Inc, USA) right after they left the freezer to optimize the crushing process. We then added  
17 200 µL CTAB extraction buffer (100 mM TRIS pH 8.0, 20 mM EDTA pH 8.0, 1.5 M NaCl, 2%  
18 cetyltrimethylammoniumbromid, 1% polyvinylpyrrolidone (MW 40,000) in nuclease-free water) to each  
19 sample and incubated the samples at 37°C for 10 min. Next, the solution was centrifuged at 20,000 x  
20 g at 4°C for 5 min and the supernatants were transferred to a new tube. We precipitated the DNA with  
21 200 µL cold 100% ethanol and centrifuged at 20,000 x g at 4°C for 5 min. We washed the pellet with 70  
22 % ethanol, let it air-dry and resuspended it in 100 µL TRIS-HCl buffer (pH 8). This was followed by a  
23 short clean up with home-made Sera-Mag purification beads pre-activated and stored in a PEG/NaCl  
24 as described in (Rohland and Reich, 2012).

25 *DNA extraction of bacterial communities associated with soil samples at the locations of the wild*  
26 *Arabidopsis thaliana* *populations.* Their DNA was extracted with the “DNeasy Power Soil”-Kit from  
27 Quiagen (Quiagen GmbH, Hilden, Germanu). In short, 0.25 g of soil was added to a provided PowerBead  
28 tube and vortexed to mix. 600uL of solution C1 was added and the tube was inverted several times.  
29 The tubes were vortexed for 10 minutes at maximum speed and then centrifuged for 30 sec at  
30 10,000xg. 400-500 µL of supernatant was transferred to a collection tube. 250 µL of solution C2 was  
31 added and samples were vortexed for 5 sec before they were put in the fridge (4°C) for 5 min. Tubes  
32 were centrifuged for 1 min at 10,000xg and the supernatant was transferred to a new collection tube.  
33 200 µL of solution C3 was added, samples were vortexed briefly and again incubated for 5 min at 4°C.  
34 Tubes were centrifuged for 1 min at 10,000xg and 750uL of the supernatant was transferred to a clean  
35 collection tube. 1200 µL of solution C4 was added and samples were vortexed for 5 sec. 675 µL of the  
36 whole solution was loaded onto a MB Spin Colum and centrifuged at 10,000g for 1 min. The flow-  
37 through was discarded and this step was repeated until the whole samples was processed. 500 µL of  
38 solution C5 was added on the column and the sample was centrifuged for 30 sec at 10,000xg. The flow-  
39 through was discarded and the sample was centrifuged again for 1 min at 10,000xg. The column was  
40 transferred to a clean collection tube and 100 µL of solution C6 was added to the filter membrane. The  
41 tubes was centrifuged for 30 sec at 10,000xg and the DNA was used for library preparation.

42 **Library preparation.** We amplified all samples (120 *Arabidopsis* plant samples and 120 “Other” plant  
43 samples from 5 locations and 4 sampling time points, 25 soil samples from 5 locations and 1 sampling  
44 time point and 3 CTAB buffer controls, 3 negative controls, 3 Zymo (positive) controls) in 10 µL  
45 reactions containing 1x Kapa Buffer, 0.3 mM Kapa dNTPs, 0.08 µM of each of the forward primers  
46 (341F-OH and 799R-OH; primers with overhang; **Table S1**), 0.25 µM of each of the blocking oligos  
47 (Blc\_16S\_F5 and Blc\_16S\_R1; **Table S1**) to block amplification of unwanted plant chloroplast DNA, 0.2



1     $\mu$ L Kapa taq DNA polymerase (KAPA Biosystems), 4.85  $\mu$ L nuclease-free water and 2  $\mu$ L Template DNA  
2 (1:4 diluted). The samples were run on a thermocycler block at 95°C for 3 min followed by 5 cycles at  
3 98°C for 20 sec, 55°C for 1 min, 72°C 2 min. The last cycle was followed by a final extension of 72°C for  
4 another 2min. We then cleaned the samples enzymatically using Exonuclease I and Antarctic  
5 phosphatase (New England Biolabs, Inc.) to degrade leftover primers and inactivate nucleotides (0.5uL  
6 each enzyme with 1.22uL Antarctic phosphatase buffer at 37°C for 30 minutes followed by 80°C for 15  
7 min). We used 5uL of the cleaned sample as template for the 2<sup>nd</sup> (extension) PCR. The 20uL reactions  
8 contained as before 1x Kapa Buffer, 0.3 mM Kapa dNTPs and 0.4  $\mu$ L of Kapa taq DNA polymerase. Each  
9 sample received a unique pair of indexed forward and reverse primers in a final concentration of 0.3  
10  $\mu$ M (Mayer et al., 2021). Samples were run on a similar program as in the first PCR with the difference  
11 of the annealing temperature (60°C), a shorter extension time in the cycles (1 min) and for 35 cycles.  
12 We cleaned the products with 0.5x vol. of 2x diluted Sera-Mag purification beads and fluorescently  
13 quantified the gene products using Picogreen (Thermo Fisher Scientific, Inc.). We used a sample with  
14 high fluorescence to prepare a standard curve according to which we calculated the concentrations of  
15 the individual samples. We combined the samples in roughly equimolar concentrations and  
16 concentrated once more using 0.5x vol. of 2x diluted Sera-Mag purification beads. The final library was  
17 quantified with a Qubit (Thermo Fisher Scientific, Inc). The library was denatured and then loaded onto  
18 a MiSeq lane spiked with 10% PhiX genomic DNA to ensure high enough sequence diversity.  
19 Sequencing was performed for 600 cycles to recover 300 bp of information in the forward and reverse  
20 directions with standard Illumina sequencing primers.

21 *Processing amplicon sequencing data.* We split the amplicon sequencing data on indices and trimmed  
22 the adapter sequences from distal read ends using Cutadapt 1.2.1 (Martin, 2011). We then clustered  
23 amplicon sequencing data into amplicon sequencing variants “ASVs” using dada2 (Callahan et al.,  
24 2016). Reads were first filtered based on their quality (`truncLen=c(200,100)`) and then dereplicated to  
25 eliminate exactly redundant reads. Data was denoised using the inferred error rates ASVs called in the  
26 forward and reverse reads. Since not all forward and reverse reads overlapped, we merged forward  
27 and reverse reads by keeping overlapping reads and concatenating the rest. We then removed  
28 detected chimeric sequences and retrieved a sequence table from the merged data. We assigned  
29 taxonomy to the final set of ASVs using the Silva 16S (v 1.32) database (Quast et al., 2013).

30 **Analysis of microbial diversity.** We performed downstream analysis in R with Phyloseq and Vegan  
31 (Dixon, 2003; McMurdie and Holmes, 2013). Positive (Zymo microbial community standard) and  
32 negative (DNA extraction buffer or no template) controls were as expected (**Figure S2**). Host-derived  
33 reads were removed by filtering any ASVs in the order “Chloroplast” and “Rickettsiales” from the 16S  
34 ASV tables. Additionally, taxa were agglomerated at genus level. For most analyses (except the bar  
35 charts based on relative abundance) we filtered the data to only allow samples with more than 100  
36 reads (without the “Chloroplast” and “Rickettsiales” reads) filtering out 3 samples in total. For  
37 barcharts, taxa that were less abundant than 100 times in one sample were removed. For the diversity  
38 analysis (beta diversity, beta dispersion and alpha diversity) samples were additionally rarefied to an  
39 equal number of reads. Beta diversity is based on the Bray-Curtis distance and ordination performed  
40 with principle coordinates analysis. Statistical significance of explaining variables was tested with an  
41 Adonis permutational analysis of variance test. Alpha diversity was inferred using Chao1 and Shannon  
42 metrics. The cores of the individual datasets were defined as taxa being present in at least 40% of the  
43 samples in each subset. Of the resulting core taxa, the relative abundances were extracted and plotted.  
44 Statistical methods used in each analysis are cited in the figures and text.

45 **Data Availability.** All scripts used to generate ASV tables from the raw data, as well as ASV tables,  
46 metadata files and scripts that generate the main figures are publicly available on Figshare:



- 1 <http://doi.org/10.6084/m9.figshare.19352189>. Raw sequencing data is being made publicly available
- 2 on NCBI (BioProject ID: PRJNA815825)
- 3

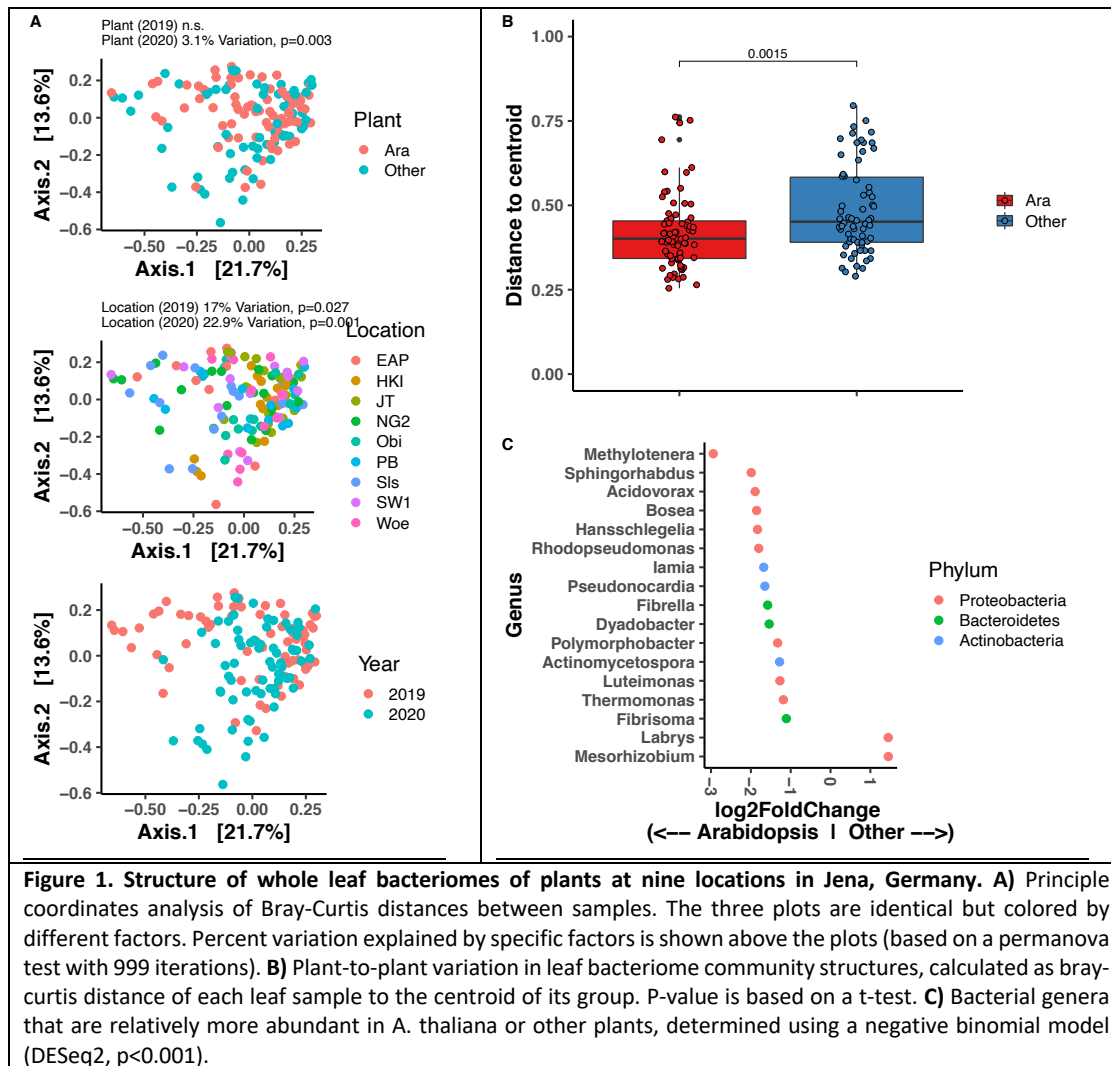
## 1 Results

### 2 Whole leaf bacteriomes depend in part on plant species but strong environmental shaping is similar 3 across species boundaries

4 To study how the assembly of whole leaf bacteriomes are shaped in wild plants, we identified 13  
5 previously unstudied *A. thaliana* populations in Jena, Germany (**Figure S2A and Table S2**). We  
6 characterized nine individual *A. thaliana* plants in each location based on length polymorphisms of  
7 microsatellite loci (**Figure S2B and S2C**). These polymorphisms suggest that most populations are  
8 genetically distinct from one another but have lower diversity within the populations. We sampled  
9 plants in nine of these populations by harvesting leaf tissues of both *Arabidopsis thaliana* and other  
10 sympatric species (other ground-dwelling ruderal plants randomly selected in the immediate vicinity  
11 of the collected *A. thaliana* plants) near the end of February in 2019 and 2020. Each sample is derived  
12 from one plant individual. We used ITS amplicon sequencing to characterize the non-*Arabidopsis* plants  
13 (Figure S3). Most samples consisted of ITS reads from primarily one plant, resulting in identification to  
14 the genus level for most plants. We could not recover ITS reads for some samples and these were not  
15 classified. In total there were 25 genera from 12 plant families, and between 4-9 genera per location.  
16 The collected species were diverse and included mostly herbaceous plants and grasses that were both  
17 annuals and perennials. The exact number of species was not controlled for and varied between  
18 sampling events and locations. Therefore, we simply grouped samples for analysis into *A. thaliana* and  
19 “other” plants, which resulted in comparable numbers of samples in both groups.

20 Using 16S rRNA-based amplicon sequencing, we next analyzed the bacterial diversity in the whole leaf  
21 samples. Overall, principal coordinates analysis of beta diversity (bacteriome community structures)  
22 showed no clean separation of samples based on plant group (*Arabidopsis* vs. other), year or location  
23 (Figure 1A). All three factors significantly correlated to beta diversity, but sampling year strongly  
24 interacted with sampling location and plant group (Table S3). Indeed, within locations, sampling year  
25 was significantly correlated to beta diversity at most sampling sites, indicating strong bacteriome shifts  
26 from one year to the next (Table S3). After splitting data into sampling years, we found that the effects  
27 of both sampling location and plant host changed from year to year. In 2019, location was significantly  
28 correlated to beta diversity (17%,  $p=0.027$ ) and plant identity was not. In 2020, location was correlated  
29 to even more variation (22.9%,  $p=0.001$ ) and plant identity played a small but significant role (3.1%  
30  $p=0.003$ ) (Figure 1A).

31 We reasoned that comparing *A. thaliana* to mixed non-*A. thaliana* leaf bacteriomes might  
32 underestimate the role of plant species. However, if there is a consistent species effect, the leaf  
33 bacteriome *A. thaliana* plants should be more similar to one another than the mixed other plant  
34 samples are to one another. To test this, we calculated the Bray-Curtis distance between *A. thaliana*  
35 samples and between other samples (distance to the group centroids). Indeed, leaf bacteriomes in  
36 between individual *A. thaliana* plants were more similar than between other random plants (Figure  
37 1B). This trend was robust within sampling years and within locations (Figure S4). In line with this  
38 observation, we were able to detect more taxa enriched in *A. thaliana* than in other plants (Figure 1C).  
39 Together, leaf bacteriomes in diverse ground-dwelling ruderals depend on the population location, but  
40 the effect of both the location and the plant species varies from year to year.



**Figure 1. Structure of whole leaf bacteriomes of plants at nine locations in Jena, Germany. A)** Principle coordinates analysis of Bray-Curtis distances between samples. The three plots are identical but colored by different factors. Percent variation explained by specific factors is shown above the plots (based on a permanova test with 999 iterations). **B)** Plant-to-plant variation in leaf bacteriome community structures, calculated as Bray-Curtis distance of each leaf sample to the centroid of its group. P-value is based on a t-test. **C)** Bacterial genera that are relatively more abundant in *A. thaliana* or other plants, determined using a negative binomial model (DESeq2,  $p < 0.001$ ).

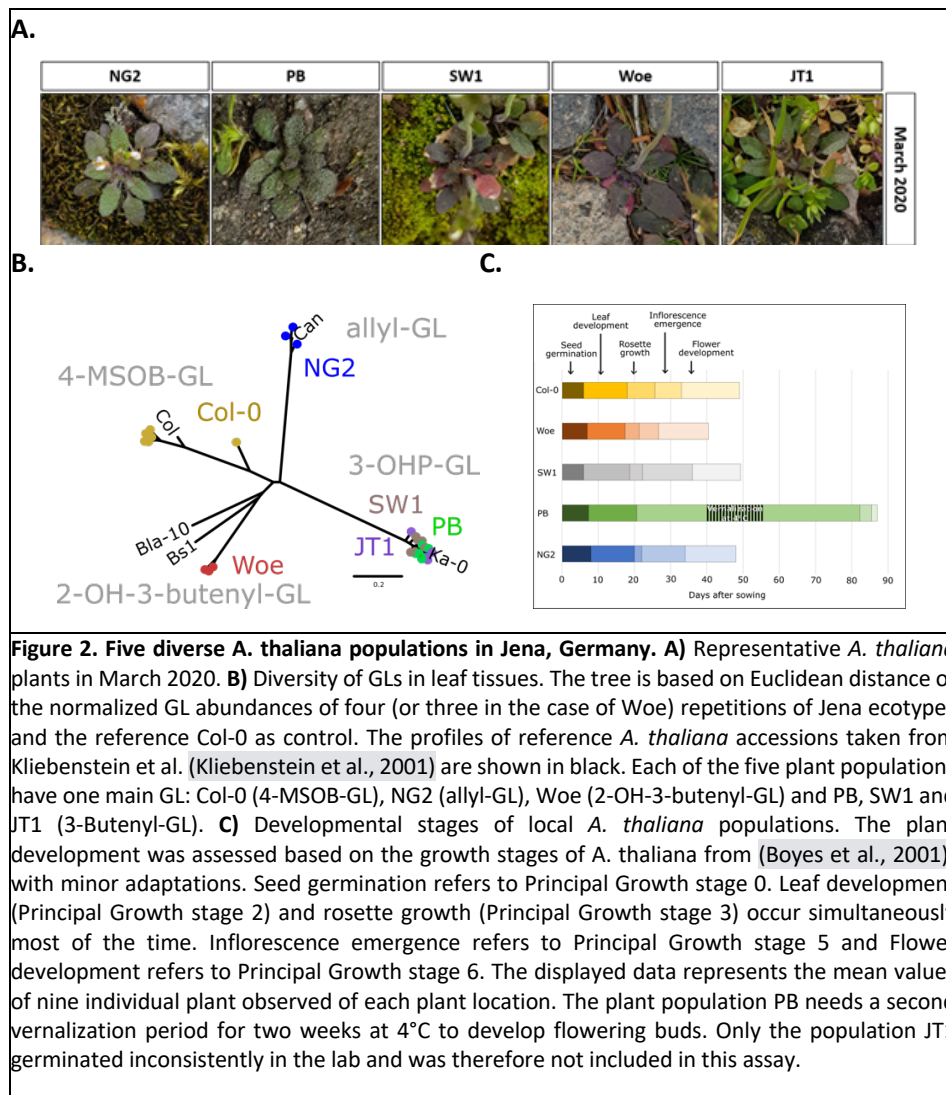
1

2 A core plant whole-leaf bacteriome common among plant species is clearly differentiated from soil

3 To better understand the origins and distribution of whole leaf bacterial colonizers, we focused on the  
 4 most common bacterial colonizers in plants at five selected locations (NG2 - Neugasse2, PB -  
 5 Paradiesbahnhof, SW1 - Sandweg1, JT1 - Johannistor1, Woe - Wöllnitzer; **Figure 2A**). The *A. thaliana*  
 6 plants in these populations have high inter-population diversity as measured by microsatellite loci  
 7 (**Figure S2**). This is supported by both the characteristic profiles of 15 glucosinolates (GLs) in the leaves  
 8 (**Figure 2B, Figure S5**) and their development times (**Figure 2C**). In these populations the whole leaf  
 9 bacteriome was characterized in leaf samples collected in both February and March of 2019 and 2020.  
 10 The bacteriomes were dominated by the phyla *Proteobacteria* (Max. 77% in Woe and min. 63% in  
 11 SW1), followed by *Bacteroidetes* (Max. 34% in SW1 and min. 12% in Woe; 2020) and *Actinobacteria*  
 12 (Max. 15% in JT1 and min. 3% in SW1) (**Figure S6, Table S4**).

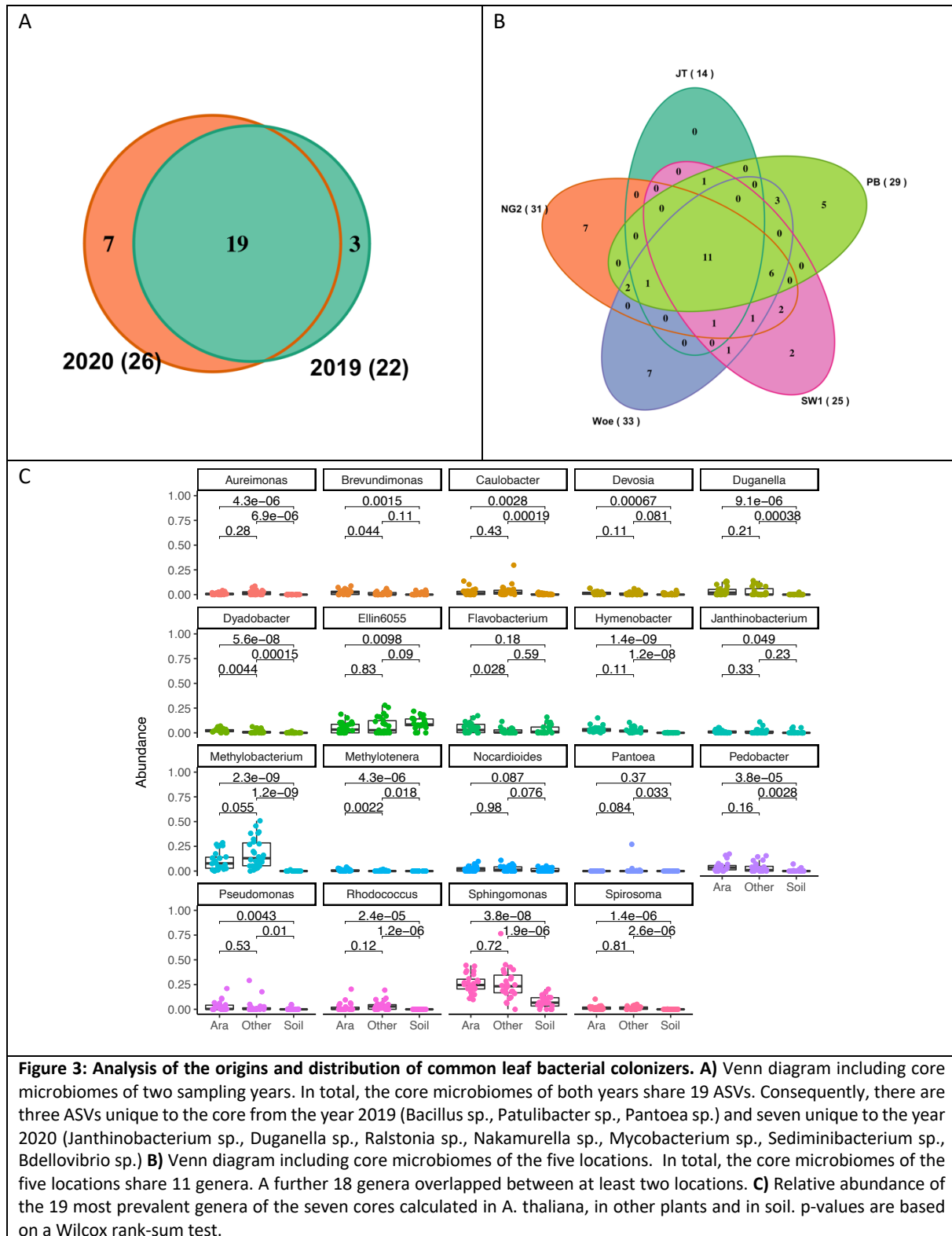
13

14



1

2 To focus on consistent bacterial colonizers, we defined “cores” that appeared in >40% of plant samples,  
 3 even if they were low relative abundance. We calculated seven cores, one from each location (including  
 4 both years) and one from each of 2019 or 2020 (including all locations). The overall cores were  
 5 consistent from 2019 to 2020, with 22 genera in 2019 and 26 in 2020 (Figure 3A and Table S5). 19 of  
 6 the core genera overlapped between the years, including the most prevalent genera *Sphingomonas*  
 7 *sp.*, *Methylobacterium sp.* *Ellin6055 sp.* and *Hymenobacter sp.* The cores at NG2 consisted of 31 genera,  
 8 PB 29 genera, JT1 14 genera, SW1 25 genera and Woe 33 genera (Figure 3B and Table S5). 11 genera  
 9 shared at every location were also in the 19 core genera overlapping between years. A further 18  
 10 genera overlapped between at least two locations and 21 genera were unique to a particular site.  
 11 (Figure 3B and Table S5). Thus, the most common taxa overall were common at all locations and in  
 12 each year, while a few taxa were more common at certain locations and times.



1

2 To broadly understand recruitment of these consistent colonizers, we combined the ten most  
3 prevalent genera of each year and site core (70 genera total), which considering overlaps resulted in  
4 19 genera. This group thus includes the most prevalent taxa overall (the 11 common to all cores) plus  
5 taxa that were prevalent in only some locations or years (all were in at least two cores). We compared  
6 the relative abundance of these taxa between *A. thaliana*, other plants and soil in March 2020, when  
7 soil samples from the five different locations were collected at the same time as plant samples. Not  
8 surprisingly, the soil samples are overall more diverse on phylum level than the plant samples, including

1 seven phyla instead of six (**Figure S7 and Table S6**). There are clear differences between the abundance  
2 of the core taxa in the soil compared to the plants. Most (15 of 19) core taxa were detected in soil  
3 samples at least once, but many significantly differed in abundance between locations (**Figure 3C and**  
4 **Figure S8**). All of the genera except *Flavobacterium* sp., *Pantoea* sp. and *Ellin6055* sp. were enriched in  
5 *A. thaliana* compared to soil. *Ellin6055* sp. was most consistently found high abundance in soil samples  
6 and was the only taxa significantly depleted in leaves (**Figure 3C**). Interestingly, although non-*A.*  
7 *thaliana* plants had a mix of species, 13 of 19 core taxa were enriched compared to soil at similar  
8 abundances as in *A. thaliana*, suggesting overall similar recruitment patterns. Only *Methylobacterium*  
9 sp. and *Pantoea* sp. were higher abundance in other plants, while *Brevundimonas* sp., *Dyadobacter* sp.,  
10 *Flavobacterium* sp. and *Methylothera* sp. were higher abundance in *A. thaliana* (**Figure 3C**). Together,  
11 the most common leaf bacteriome taxa are largely shared among diverse ground-dwelling ruderal  
12 plants and are highly enriched compared to the soil they are in contact with.

13 Stronger differentiation of whole-leaf bacteriomes in *A. thaliana* populations correlates to lower  
14 variability.

15 Next, we looked at just *A. thaliana* leaf bacteriomes at the five local locations to gain deeper insight  
16 into formation of location-specific bacteriomes (**Figure 4A**). Overall, leaf bacterial community structure  
17 in the *A. thaliana* bacterial communities correlated to location (9.3%,  $p=0.001$ ), sampling year (5.7%,  
18  $p=0.001$ ), and sampling month (2%,  $p=0.019$ ). However, the factors interacted significantly, indicating  
19 that location effects are not stable over time (**Table S7**). In all locations, sampling year had a much  
20 stronger effect than sampling month, which had little or no effect (**Table S8**). Accordingly, from 2019  
21 to 2020, many taxa became enriched (**Table S9**), corresponding to an increase in alpha diversity  
22 (significant in three locations) (**Figure 4B**). Although leaf bacteriomes in 2020 were more diverse, they  
23 were also more defined by the sampling location: While in 2019 neither sampling location nor month  
24 significantly correlated to bacteriome structure, in 2020 location correlated to almost 20% of beta  
25 diversity ( $p=0.001$ ) (**Figure 4A and Table S10**) with significant differences in alpha diversity (**Table S11**).  
26 Similarly, most bacterial genera enriched between locations that could be detected were found in 2020  
27 (**Table S12 and Figure S9**). A more nuanced view of the increase in location specificity can be observed  
28 by comparing how similar leaf bacteriomes were to one another within locations vs. between locations.  
29 In 2020, samples within locations were more similar than between locations at 4 of 5 locations vs. 2 of  
30 5 in 2019 (based on Bray-Curtis distances, **Figure 4C**). This was caused by a major decrease in variability  
31 in 2020 at locations JT1, NG2 and PB (**Figure S11**, all  $p < 5 \times 10^{-5}$ ) and an increase only at SW1 (**Figure**  
32 **S11**,  $p=0.008$ ). Thus, while leaf bacteria at most locations were more specific in 2020, this was not  
33 universal.

34 Importantly, the trends in plant-to-plant variation and leaf alpha diversity observed in 2020 were  
35 distinct from local soils: While soil bacterial communities at different locations were significantly  
36 different (explaining 25.7% of variation, permanova  $p=0.003$ ), in contrast to leaves, soil alpha diversity  
37 did not significantly differ between locations and distance between samples was lowest in NG2 and  
38 highest in JT1 (**Figure S12 and Table S13**).

39 Taken together, leaf bacteriomes in *A. thaliana* populations are more location- and host-specific (less  
40 variable within groups) in some years than others (see also **Figure 1**) with the direction of observed  
41 trends depending on the population.

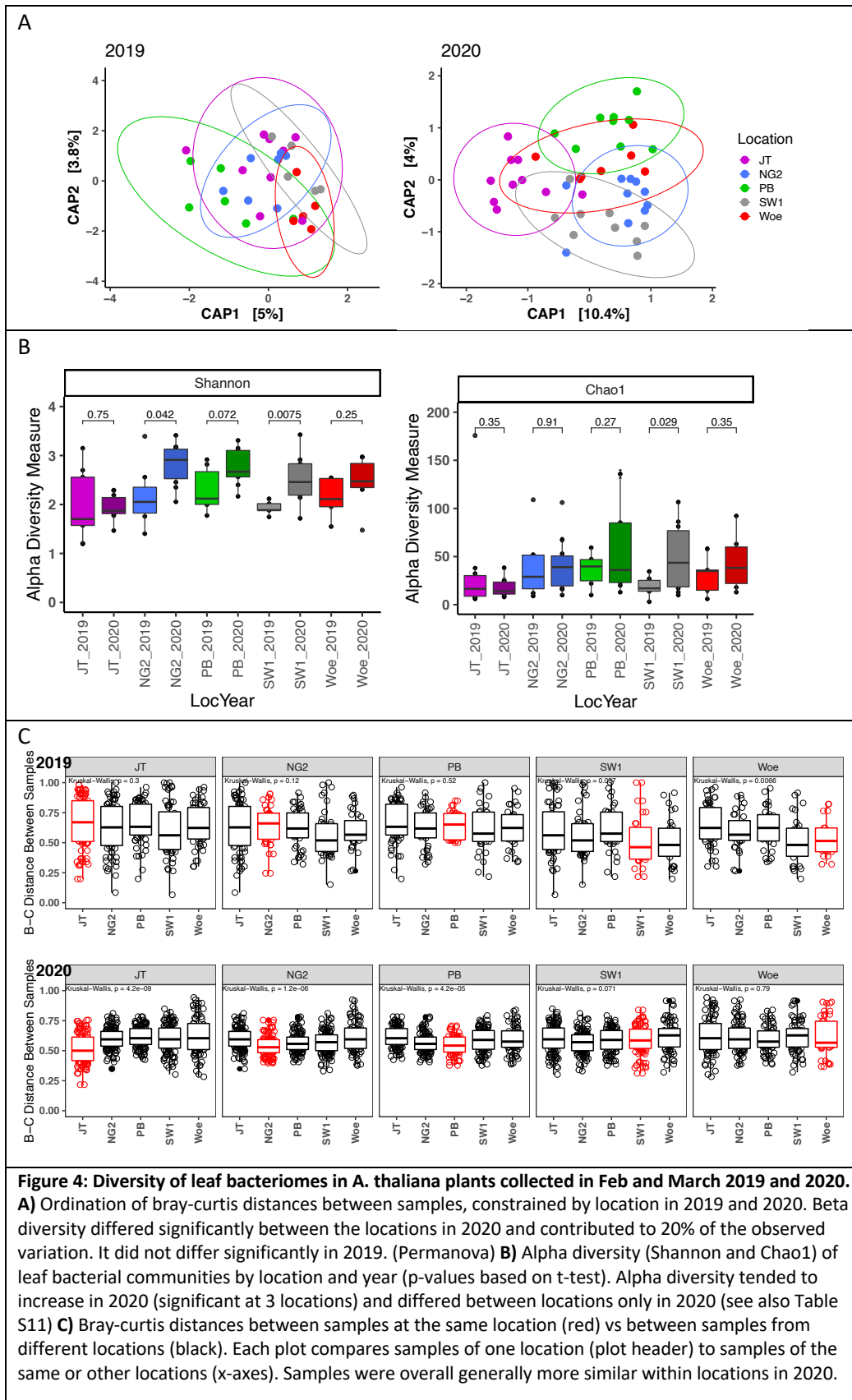
42

43

44



1  
2



3

1 Leaf chemotype strongly correlates to level of leaf bacteriome variability.

2 Since locations (environments) are in this case fully nested with *A. thaliana* ecotypes, we cannot  
3 distinguish whether different patterns in leaf bacteriome recruitment could depend on host genotypes.  
4 Therefore, we turned to leaf traits. At seven locations where we characterized *A. thaliana* whole leaf  
5 bacteriomes, we also measured indole and aliphatic GL profiles (**Figure S5**). While indole GLs were  
6 similar across ecotypes, the main aliphatic GL profiles were distinctly different. In Europe, seven  
7 different aliphatic glucosinolate chemotypes are known whose composition is determined by only  
8 three genetic loci (Katz et al., 2021). Plants of different aliphatic chemotypes usually have one main  
9 aliphatic GL with very little overlap of individual compounds, which was also true for our plant  
10 populations (**Figure S5**). We observed three of seven chemotypes: allyl-GL at NG2, 3-OH-Propyl-GL  
11 (hereafter 3OHP) at PB, SW1 and JT1 and 2-OH-3-butenyl-GL (hereafter 2OH3But) in Woe, SLS and EAP.  
12 Allyl-GL and 3OHP-GL are common in Germany, whereas 2OH3But-GL is more common in other parts  
13 of Europe (Katz et al., 2021). Since aliphatic glucosinolates have been shown to play important roles in  
14 bacterial colonization (Fan et al., 2011), we checked whether we could detect effects of leaf chemistry  
15 independent of sampling locations using the three 3OHP and three 2OH3But populations (all sampled  
16 in February 2019 and 2020).

17 As expected, whole leaf bacterial community structures overall significantly correlated to year (6.8%,  
18  $p=0.001$ ), with significant interactions between year and chemotype (5%,  $p=0.002$ ). Therefore, we  
19 divided the data into years and found that chemotype only significantly correlated to bacterial  
20 community structures in 2020 (**Figure 5A and 5B**, 9.3%,  $p=0.003$ ). A DESeq2 differential abundance  
21 analysis suggested that in 2019, 12 genera were significantly enriched ( $p<0.05$ ) between chemotypes  
22 and in 2020 that one genus was enriched. Of those, only *Bacillus* sp. in 2019 and *Shingopyxis* sp. in  
23 2020 (both enriched in 2OH3But), were significantly enriched when evaluated using a stringent  
24 Wilcoxon rank-sum test (Figure S13).

25 The ordinations did appear to show that the bacteriomes of one or the other chemotype were more  
26 variable in each year. Thus, we measured the similarity of bacterial communities in leaves sharing a  
27 chemotype and in leaves with different chemotypes. Since location has strong effects, to make a fair  
28 comparison, we excluded comparisons between samples at the same location. Indeed, in 2019,  
29 2OH3But leaf bacteriomes were significantly more similar to one another than they were to 3OHP leaf  
30 bacteriomes ( $p=1.5 \times 10^{-7}$ ), while in 2020 this was true of 3OHP bacteriomes ( $p=7.3 \times 10^{-7}$ ) (**Figure 5C and**  
31 **5D**). To our surprise, in each year, leaf bacteriomes in the other chemotype (3OHP in 2019 and  
32 2OH3But in 2020) were significantly *less* similar to one another than to plants in populations with a  
33 different chemotype ( $p=9.8 \times 10^{-5}$  and 0.00045, respectively) (**Figure 5C and 5D**). Therefore, our data  
34 suggests a possible role for leaf chemotypes in constraining leaf community structures *and/or* in  
35 promoting beta diversity of leaf community structures. Importantly, the outcome is apparently  
36 dependent on an unknown factor that changed between growing seasons.

37

38

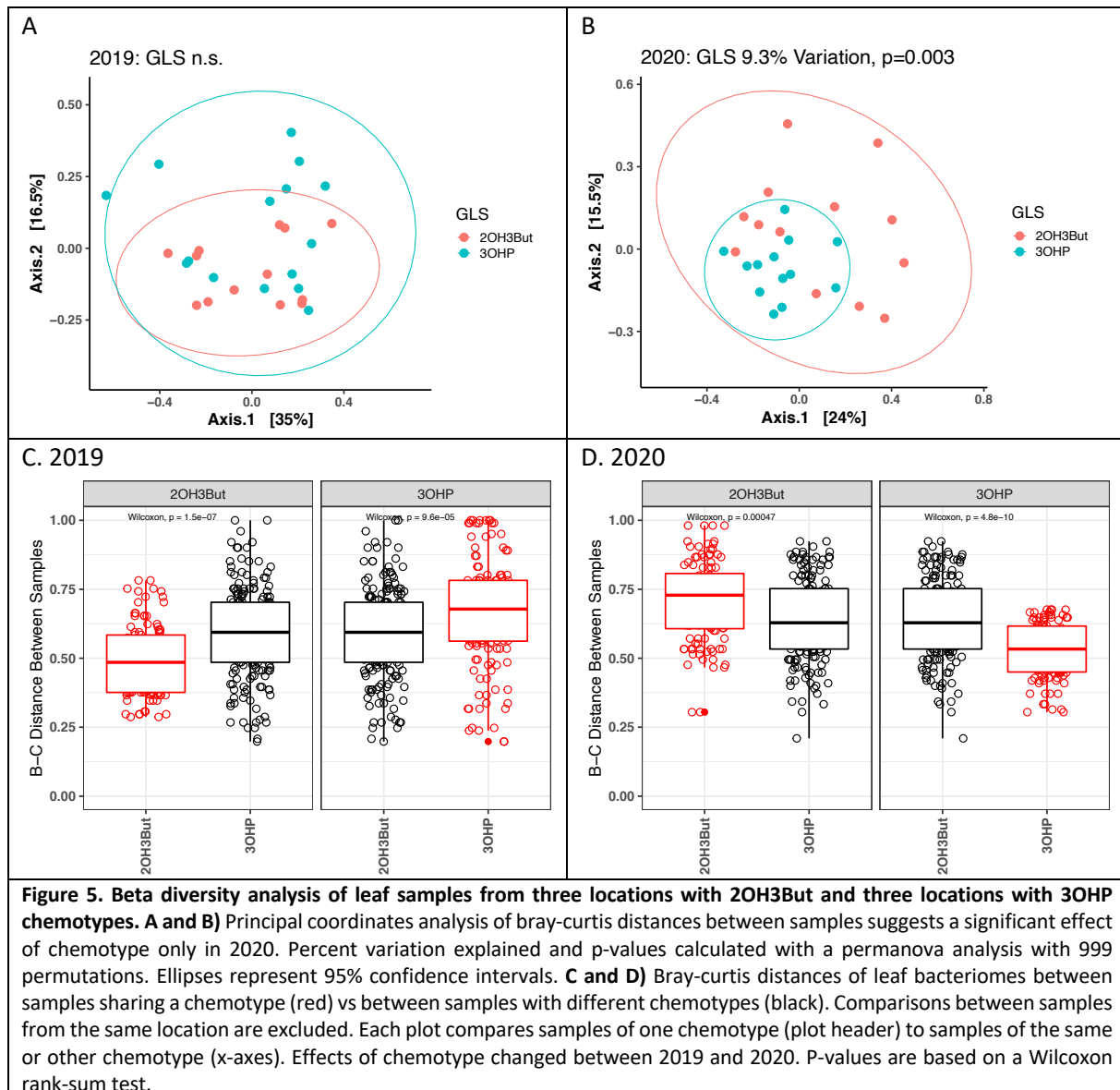
39

40

41

42

43



**Figure 5. Beta diversity analysis of leaf samples from three locations with 20H3But and three locations with 3OHP chemotypes. A and B)** Principal coordinates analysis of bray-curtis distances between samples suggests a significant effect of chemotype only in 2020. Percent variation explained and p-values calculated with a permanova analysis with 999 permutations. Ellipses represent 95% confidence intervals. **C and D)** Bray-curtis distances of leaf bacteriomes between samples sharing a chemotype (red) vs between samples with different chemotypes (black). Comparisons between samples from the same location are excluded. Each plot compares samples of one chemotype (plot header) to samples of the same or other chemotype (x-axes). Effects of chemotype changed between 2019 and 2020. P-values are based on a Wilcoxon rank-sum test.

1

2

## 1 Discussion

2 Microbial communities associated with the leaves of plants are diverse and composed of bacteria,  
3 fungi, oomycetes and other eukaryotes (Mayer et al., 2021) that play key roles in host health (Liu et  
4 al., 2020). Therefore, understanding host controls on leaf colonization processes and the role the  
5 microbiome plays in the local adaptation of plants will be important in developing strategies to shape  
6 plant microbiomes in beneficial ways. One approach to do this is to use panels of diverse plants in  
7 garden experiments, which can provide valuable insights into colonization under controlled conditions  
8 (Horton et al., 2014) or to use reciprocal planting experiments to gain insights under partially controlled  
9 conditions (Wagner et al., 2016). While powerful, such experiments likely cannot fully capture natural  
10 dynamics since wild plants over time build up required associations in response to local context,  
11 including disease threats and polyculture with other plants (Santhanam et al., 2015). Therefore, wild  
12 plant populations can still serve as an important tool for studying the assembly of plant microbiomes  
13 (Agler et al., 2016). To address the challenge of understanding how host factors shape microbiomes,  
14 *Arabidopsis thaliana* is a valuable tool due to persistent wild populations, usually co-colonized by other  
15 plants, and diverse physical and chemical phenotypes, sometimes across small geographic scales (Katz  
16 et al., 2021; Seren et al., 2017; Weigel, 2012).

17 Soil is a primary source of inoculum for both roots and leaves of plants (Massoni et al., 2021; Tkacz et  
18 al., 2020) and probably continuously inoculates ground-dwelling ruderal plants due to their contact  
19 with the ground and splashing during rain events. Soil bacteria are extremely diverse (Ramirez et al.,  
20 2014) but we observed similar to previous reports that leaves of *A. thaliana* tend to be colonized mainly  
21 by bacteria belonging to a few phyla, including *Proteobacteria*, *Bacteroides* and *Actinobacteria*  
22 (Bodenhausen et al., 2013) Consistent with previous work (Wagner et al., 2016), most bacteria that  
23 consistently colonized leaves were also observed in soil samples but in differential abundance. Many  
24 taxa were extremely enriched compared to soil, such as *Sphingomonas* sp., *Hymenobacter* sp. and  
25 *Rhodococcus* sp., underscoring strong plant selection and their fitness as leaf colonizers. Other taxa  
26 like the common soil taxa *Ellin6055* sp., while still common in leaves, was selected against in the leaf  
27 bacteriome. Interestingly, we found that all common *A. thaliana* colonizers were also found on other  
28 plants, in most cases at similar abundances. Previous observations similarly showed in three  
29 herbaceous plant species that 98% of identifiable taxa were shared with similar prevalence across hosts  
30 (Massoni et al., 2020). Thus, most wild ground-dwelling ruderal plants in these populations appear to  
31 have conserved colonization dynamics leading to clearly distinguished bacteriomes from the soil but  
32 similar bacteriomes in different plant species and one location.

33 We did observe some differences between hosts, in *Flavobacterium*, *Brevundimonas* sp.,  
34 *Methylobacterium* sp. and *Methylothenera* sp., the latter two of which are facultative methanotrophs  
35 that can grow on C1 compounds (Kalyuzhnaya et al., 2012). Together with previous results observing  
36 similarly that *Methylobacterium* sp. community composition is significantly affected by plant host  
37 species (Knief et al., 2010), C1 metabolism seems to be an important factor underlying differentiation  
38 of multiple prevalent leaf colonizers. This could be important, since at least *Methylobacterium* sp. can  
39 provide plant-beneficial effects (Grossi et al., 2020). *Flavobacterium* sp. was previously shown in lab  
40 colonization experiments to be differentially abundant between plant species rhizosphere  
41 microbiomes (Wippel et al., 2021), but the reasons for their differentiation in leaves will need more  
42 research. Recently, the same strains of *Pseudomonas* sp. and *Sphingomonas* sp. were shown to thrive  
43 in the wild in leaves of both *A. thaliana* and other plant hosts (Lundberg et al., 2021). Such approaches  
44 that can distinguish bacteria at finer levels are needed because they can lead to better understanding  
45 of plant recruitment strategies in nature.

46 Besides effect of host plant species, we observed a strong effect of sampling location. Differences  
47 between populations in *A. thaliana* have been observed before (Agler et al., 2016) and many factors

1 could contribute. For example, NG2, PB and JT1 are near the city center of Jena, and an urban/rural  
2 divide has previously been suggested to influence tree-leaf bacterial communities (Laforest-Lapointe  
3 et al., 2017). Whatever the cause of inter-location effects, they were inconsistent: we observed an  
4 increased effect of location from 2019 to 2020 in both *A. thaliana* (~11% to ~20%) and in all plants  
5 (from 17% to 23% of variation). Similarly, the effect of both host plant species and leaf glucosinolate  
6 chemistry on leaf bacterial communities also increased in 2020. A shift in 2020 could also be observed  
7 within locations where sampling year correlated to 11.7% to 15.7% of observed variation. Importantly,  
8 we observed not just a shift in strength of explanatory variables, but also that this was due to changes  
9 in plant-to-plant bacteriome variation. Phyllosphere microbiota are known to have temporal dynamics,  
10 most likely due to environmental conditions and nutrient availability (Copeland et al., 2015). The winter  
11 of 2019-2020 (the growing season of *A. thaliana*) was abnormally warm and dry with no days below  
12 freezing and frequent temperature spikes up to 15°C (data for the Jena Sternwarte weather station,  
13 DWD climate data center, [https://opendata.dwd.de/climate\\_environment/CDC/](https://opendata.dwd.de/climate_environment/CDC/)). Thus, we  
14 hypothesize that climate could be to blame for the effects in 2020.

15 Temperature could affect leaf bacteriome assembly in multiple ways. In our ecotypes, development  
16 times under lab conditions were highly divergent. In an unusually warm winter, development could  
17 become magnified because of missing temperature cues. The absence of winter temperature spikes  
18 was previously shown to be important in regulation of flowering in *A. thaliana* (Hepworth et al., 2018)  
19 and could regulate other processes like temperature-dependent phyllosphere microbial community  
20 assembly (Aydogan et al., 2020). Additionally, since we sampled individual plants, we could observe  
21 that the shift in 2020 to more differentiated leaf bacterial communities was caused by a strong  
22 decrease in the variation in leaf bacteriomes at several locations and an increase only in one location.  
23 Our results suggest that this population-specific behavior might in part be explained by leaf  
24 chemotypes. If so, it is reasonable to expect temperature to play a role, since production of  
25 glucosinolates (GLs) themselves are regulated by temperature in *A. thaliana* (Kissen et al., 2016).  
26 Additionally, upon leaf damage GLs encounter myrosinase enzymes and break down to a range of  
27 potentially toxic products (Halkier and Gershenzon, 2006). For example, sulforaphane, the  
28 isothiocyanate product of 4-MSOB GLs, regulates bacterial colonization by inhibiting type three  
29 secretion system expression (Wang et al., 2020). The exact hydrolysis products of each chemotype are  
30 different and further depend on temperature (Hanschen et al., 2017), which likely comes with  
31 differences in volatility and biological activity. Thus, GLs could reasonably account for temperature-  
32 dependent bacteriome differences.

33 An open question then is would plant ecotypes regulate plant-to-plant variation in leaf microbiomes  
34 or are colonization phenotypes just a passive effect of environment and for example plant chemistry?  
35 Plants in the wild must deal with unpredictable threats and therefore, phenotypic variation is critical  
36 survival and adaptability. Leaf microorganisms play important roles in defining plant traits and thus can  
37 extend plant phenotypes (Hawkes et al., 2021). Therefore, increased plant-to-plant leaf microbiome  
38 variation could increase chances of survival. This would not come without risk, since essentially  
39 allowing or even driving random colonization can be detrimental. Thus, a regulated system that auto-  
40 tunes according to environment might be a good strategy to balance adaptability and risks. At any rate,  
41 our results show that factors shaping leaf bacteriomes exert effects that probably depend on  
42 environmental conditions, but more work is needed to understand whether and how this affects plant  
43 health. Gaining this understanding, however, could help design better approaches to harness  
44 microorganisms for sustainability and should be an important focus of ongoing research.

#### 45 **Acknowledgements**

46 TM and MTA were supported by the Carl Zeiss Stiftung via the Jena School for Microbial  
47 Communication and the Friedrich Schiller University of Jena. They were also funded by the Deutsche

1 Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy –  
2 EXC 2051 – Project-ID 390713860. MR and JG were supported by the Max Planck Society.

#### 4 **Author Contributions**

5 MTA conceptualized the project. TM performed plant sampling and other analyses and generated  
6 sequencing data. MR analyzed the glucosinolate composition of leaf samples. TM designed and wrote  
7 the scripts to process data and generate figures with help from MTA. TM and MTA wrote the  
8 manuscript and all authors edited and approved it.

#### 10 **Bibliography**

- 11  
12 Agler, M.T., Ruhe, J., Kroll, S., Morhenn, C., Kim, S.-T., Weigel, D., Kemen, E.M., 2016. Microbial Hub  
13 Taxa Link Host and Abiotic Factors to Plant Microbiome Variation. *PLoS Biol* 14, e1002352.  
14 <https://doi.org/10.1371/journal.pbio.1002352>
- 15 Almario, J., Mahmudi, M., Kroll, S., Agler, M., Placzek, A., Mari, A., Kemen, E., 2021. Life on leaves :  
16 uncovering temporal dynamics in *Arabidopsis*' leaf microbiota (BioRxiv preprint).  
17 <https://doi.org/10.1101/2021.07.06.450897>
- 18 Aydogan, E.L., Budich, O., Hardt, M., Choi, Y.H., Jansen-Willems, A.B., Moser, G., Müller, C., Kämpfer,  
19 P., Glaeser, S.P., 2020. Global warming shifts the composition of the abundant bacterial  
20 phyllosphere microbiota as indicated by a cultivation-dependent and -independent study of  
21 the grassland phyllosphere of a long-term warming field experiment. *FEMS Microbiology*  
22 *Ecology* 96, fiae087. <https://doi.org/10.1093/femsec/fiae087>
- 23 Bell, C.J., Ecker, J.R., 1994. Assignment of 30 Microsatellite Loci to the Linkage Map of *Arabidopsis*.  
24 *Genomics* 19, 137–144. <https://doi.org/10.1006/geno.1994.1023>
- 25 Berg, M., Koskella, B., 2018. Nutrient- and dose-dependent microbiome-mediated protection against  
26 a plant pathogen. *Current Biology: CB* 28, 2487-2492.e3.  
27 <https://doi.org/10.1016/j.cub.2018.05.085>
- 28 Bodenhausen, N., Bortfeld-Miller, M., Ackermann, M., Vorholt, J.A., 2014. A synthetic community  
29 approach reveals plant genotypes affecting the phyllosphere microbiota. *PLoS Genet* 10,  
30 e1004283. <https://doi.org/10.1371/journal.pgen.1004283>
- 31 Bodenhausen, N., Horton, M.W., Bergelson, J., 2013. Bacterial communities associated with the  
32 leaves and the roots of *Arabidopsis thaliana*. *PloS One* 8, e56329.  
33 <https://doi.org/10.1371/journal.pone.0056329>
- 34 Boyes, D.C., Zayed, A.M., Ascenzi, R., McCaskill, A.J., Hoffman, N.E., Davis, K.R., Görlach, J., 2001.  
35 Growth Stage–Based Phenotypic Analysis of *Arabidopsis*: A Model for High Throughput  
36 Functional Genomics in Plants. *Plant Cell* 13, 1499–1510.  
37 <https://doi.org/10.1105/TPC.010011>
- 38 Brown, P.D., Tokuhisa, J.G., Reichelt, M., Gershenzon, J., 2003. Variation of glucosinolate  
39 accumulation among different organs and developmental stages of *Arabidopsis thaliana*.  
40 *Phytochemistry* 62, 471–481. [https://doi.org/10.1016/S0031-9422\(02\)00549-6](https://doi.org/10.1016/S0031-9422(02)00549-6)
- 41 Burow, M., Müller, R., Gershenzon, J., Wittstock, U., 2006. Altered Glucosinolate Hydrolysis in  
42 Genetically Engineered *Arabidopsis thaliana* and its Influence on the Larval Development of  
43 *Spodoptera littoralis*. *J Chem Ecol* 32, 2333–2349. [https://doi.org/10.1007/s10886-006-9149-](https://doi.org/10.1007/s10886-006-9149-1)  
44 1
- 45 Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J., Holmes, S.P., 2016. DADA2: High-  
46 resolution sample inference from Illumina amplicon data. *Nat Meth* 13, 581–3.  
47 <https://doi.org/10.1038/nmeth.3869>
- 48 Carlström, C.I., Field, C.M., Bortfeld-Miller, M., Muller, B., Sunagawa, S., Vorholt, J.A., 2019. Synthetic  
49 microbiota reveal priority effects and keystone strains in the *Arabidopsis* phyllosphere. *Nat*  
50 *Ecol Evol* 3, 1445–1454. <https://doi.org/10.1038/s41559-019-0994-z>



- 1 Copeland, J.K., Yuan, L., Layeghifard, M., Wang, P.W., Guttman, D.S., 2015. Seasonal Community  
2 Succession of the Phyllosphere Microbiome. *MPMI* 28, 274–285.  
3 <https://doi.org/10.1094/MPMI-10-14-0331-FI>
- 4 Dixon, P., 2003. VEGAN, a package of R functions for community ecology. *J Veg Science* 14, 927–930.  
5 <https://doi.org/10.1111/j.1654-1103.2003.tb02228.x>
- 6 Duran, P., Thiergart, T., Garrido-Oter, R., Agler, M., Kemen, E., Schulze-Lefert, P., Hacquard, S., 2018.  
7 Microbial Interkingdom Interactions in Roots Promote *Arabidopsis* Survival. *Cell* 175, 973–  
8 983.e14. <https://doi.org/10.1016/j.cell.2018.10.020>
- 9 Edwards, K., Johnstone, C., Thompson, C., 1991. A simple and rapid method for the preparation of  
10 plant genomic DNA for PCR analysis. *Nucl Acids Res* 19, 1349–1349.  
11 <https://doi.org/10.1093/nar/19.6.1349>
- 12 Fan, J., Crooks, C., Creissen, G., Hill, L., Fairhurst, S., Doerner, P., Lamb, C., 2011. *Pseudomonas* sax  
13 genes overcome aliphatic isothiocyanate-mediated non-host resistance in *Arabidopsis*.  
14 *Science* 331, 1185–8. <https://doi.org/10.1126/science.1199707>
- 15 García-Suárez, R., Verduzco-Rosas, L.A., Del Rincón-Castro, M.C., Délano-Frier, J.P., Ibarra, J.E., 2017.  
16 Translocation of *Bacillus thuringiensis* in *Phaseolus vulgaris* tissues and vertical transmission  
17 in *Arabidopsis thaliana*. *J Appl Microbiol* 122, 1092–1100. <https://doi.org/10.1111/jam.13407>
- 18 Graser, G., Schneider, B., Oldham, N.J., Gershenzon, J., 2000. The Methionine Chain Elongation  
19 Pathway in the Biosynthesis of Glucosinolates in *Eruca sativa* (Brassicaceae). *Arch Biochem*  
20 *Biophys* 378, 411–419. <https://doi.org/10.1006/abbi.2000.1812>
- 21 Grossi, C.E.M., Fantino, E., Serral, F., Zawoznik, M.S., Fernandez Do Porto, D.A., Ulloa, R.M., 2020.  
22 *Methylobacterium* sp. 2A Is a Plant Growth-Promoting Rhizobacteria That Has the Potential  
23 to Improve Potato Crop Yield Under Adverse Conditions. *Front Plant Sci* 11, 71.  
24 <https://doi.org/10.3389/fpls.2020.00071>
- 25 Halkier, B.A., Gershenzon, J., 2006. Biology and biochemistry of glucosinolates. *Annu Rev Plant Biol*  
26 57, 303–33. <https://doi.org/10.1146/annurev.arplant.57.032905.105228>
- 27 Hanschen, F.S., Klopsch, R., Oliviero, T., Schreiner, M., Verkerk, R., Dekker, M., 2017. Optimizing  
28 isothiocyanate formation during enzymatic glucosinolate breakdown by adjusting pH value,  
29 temperature and dilution in Brassica vegetables and *Arabidopsis thaliana*. *Sci Rep* 7, 40807.  
30 <https://doi.org/10.1038/srep40807>
- 31 Hawkes, C.V., Kjølner, R., Raaijmakers, J.M., Riber, L., Christensen, S., Rasmussen, S., Christensen, J.H.,  
32 Dahl, A.B., Westergaard, J.C., Nielsen, M., Brown-Guedira, G., Hestbjerg Hansen, L., 2021.  
33 Extension of Plant Phenotypes by the Foliar Microbiome. *Annu Rev Plant Biol* 72, 823–846.  
34 <https://doi.org/10.1146/annurev-arplant-080620-114342>
- 35 Hepworth, J., Antoniou-Kourounioti, R.L., Bloomer, R.H., Selga, C., Berggren, K., Cox, D., Collier Harris,  
36 B.R., Irwin, J.A., Holm, S., Säll, T., Howard, M., Dean, C., 2018. Absence of warmth permits  
37 epigenetic memory of winter in *Arabidopsis*. *Nat Commun* 9, 639.  
38 <https://doi.org/10.1038/s41467-018-03065-7>
- 39 Horton, M.W., Bodenhausen, N., Beilsmith, K., Meng, D., Muegge, B.D., Subramanian, S., Vetter,  
40 M.M., Vilhjalmsson, B.J., Nordborg, M., Gordon, J.I., Bergelson, J., 2014. Genome-wide  
41 association study of *Arabidopsis thaliana* leaf microbial community. *Nat Comm* 5, 5320.  
42 <https://doi.org/10.1038/ncomms6320>
- 43 Kalyuzhnaya, M.G., Beck, D.A.C., Vorobev, A., Smalley, N., Kunkel, D.D., Lidstrom, M.E.,  
44 Chistoserdova, L., 2012. Novel methylotrophic isolates from lake sediment, description of  
45 *Methylothena versatilis* sp. nov. and emended description of the genus *Methylothena*. *Int J*  
46 *System Evol Microb* 62, 106–111. <https://doi.org/10.1099/ijs.0.029165-0>
- 47 Katz, E., Li, J.-J., Jaegle, B., Ashkenazy, H., Abrahams, S.R., Bagaza, C., Holden, S., Pires, C.J.,  
48 Angelovici, R., Kliebenstein, D.J., 2021. Genetic variation, environment and demography  
49 intersect to shape *Arabidopsis* defense metabolite variation across Europe. *eLife* 10, e67784.  
50 <https://doi.org/10.7554/eLife.67784>
- 51 Kissen, R., Eberl, F., Winge, P., Uleberg, E., Martinussen, I., Bones, A.M., 2016. Effect of growth  
52 temperature on glucosinolate profiles in *Arabidopsis thaliana* accessions. *Phytochemistry*  
53 130, 106–118. <https://doi.org/10.1016/j.phytochem.2016.06.003>

- 1 Kliebenstein, D.J., Kroymann, J., Brown, P., Figuth, A., Pedersen, D., Gershenzon, J., Mitchell-Olds, T.,  
2 2001. Genetic control of natural variation in *Arabidopsis* glucosinolate accumulation. *Plant*  
3 *Phys* 126, 811–25.
- 4 Knief, C., Rasmann, A., Frances, L., Alonso-Blanco, C., Vorholt, J.A., 2010. Site and plant species are  
5 important determinants of the *Methylobacterium* community composition in the plant  
6 phyllosphere. *ISME J* 4, 719–728. <https://doi.org/10.1038/ismej.2010.9>
- 7 Laforest-Lapointe, I., Messier, C., Kembel, S.W., 2017. Tree Leaf Bacterial Community Structure and  
8 Diversity Differ along a Gradient of Urban Intensity. *mSystems* 2, e00087-17.  
9 <https://doi.org/10.1128/mSystems.00087-17>
- 10 Leopold, D.R., Busby, P.E., 2020. Host Genotype and Colonist Arrival Order Jointly Govern Plant  
11 Microbiome Composition and Function. *Current Biology* 30, 3260-3266.e5.  
12 <https://doi.org/10.1016/j.cub.2020.06.011>
- 13 Liu, H., Brettell, L.E., Singh, B., 2020. Linking the Phyllosphere Microbiome to Plant Health. *Trends*  
14 *Plant Sci* 25, 841–844. <https://doi.org/10.1016/j.tplants.2020.06.003>
- 15 Lundberg, D.S., de Pedro Jové, R., Na Ayutthaya, P.P., Karasov, T.L., Shalev, O., Poersch, K., Ding, W.,  
16 Bollmann-Giolai, A., Bezrukov, I., Weigel, D., 2021. Contrasting patterns of microbial  
17 dominance in the *Arabidopsis thaliana* phyllosphere (BioRxiv preprint).  
18 <https://doi.org/10.1101/2021.04.06.438366>
- 19 Maignien, L., DeForce, E.A., Chafee, M.E., Eren, A.M., Simmons, S.L., 2014. Ecological Succession and  
20 Stochastic Variation in the Assembly of *Arabidopsis thaliana* Phyllosphere Communities.  
21 *mBio* 5. <https://doi.org/10.1128/mBio.00682-13>
- 22 Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads.  
23 *EMBnet. J* 17, 10–12.
- 24 Massoni, J., Bortfeld-Miller, M., Jardillier, L., Salazar, G., Sunagawa, S., Vorholt, J.A., 2020. Consistent  
25 host and organ occupancy of phyllosphere bacteria in a community of wild herbaceous plant  
26 species. *ISME J* 14, 245–258. <https://doi.org/10.1038/s41396-019-0531-8>
- 27 Massoni, J., Bortfeld-Miller, M., Widmer, A., Vorholt, J.A., 2021. Capacity of soil bacteria to reach the  
28 phyllosphere and convergence of floral communities despite soil microbiota variation. *Proc*  
29 *Natl Acad Sci* 118. <https://doi.org/10.1073/pnas.2100150118>
- 30 Mayer, T., Mari, A., Almario, J., Murillo-Roos, M., Syed M Abdullah, H., Dombrowski, N., Hacquard, S.,  
31 Kemen, E.M., Agler, M.T., 2021. Obtaining deeper insights into microbiome diversity using a  
32 simple method to block host and nontargets in amplicon sequencing. *Mol Ecol Resour*  
33 <https://doi.org/10.1111/1755-0998.13408>
- 34 McMurdie, P.J., Holmes, S., 2013. phyloseq: an R package for reproducible interactive analysis and  
35 graphics of microbiome census data. *PLoS One* 8, e61217.  
36 <https://doi.org/10.1371/journal.pone.0061217>
- 37 Petipas, R.H., Geber, M.A., Lau, J.A., 2021. Microbe-mediated adaptation in plants. *Ecol Lett* 24,  
38 1302–1317. <https://doi.org/10.1111/ele.13755>
- 39 Postma, F.M., Ågren, J., 2018. Among-year variation in selection during early life stages and the  
40 genetic basis of fitness in *Arabidopsis thaliana*. *Mol Ecol* 27, 2498–2511.  
41 <https://doi.org/10.1111/mec.14697>
- 42 Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O., 2013.  
43 The SILVA ribosomal RNA gene database project: improved data processing and web-based  
44 tools. *Nucl Acids Res* 41, D590-596. <https://doi.org/10.1093/nar/gks1219>
- 45 Ramirez, K.S., Leff, J.W., Barberan, A., Bates, S.T., Betley, J., Crowther, T.W., Kelly, E.F., Oldfield, E.E.,  
46 Shaw, E.A., Steenbock, C., Bradford, M.A., Wall, D.H., Fierer, N., 2014. Biogeographic patterns  
47 in below-ground diversity in New York City's Central Park are similar to those observed  
48 globally. *Proc Biol Sci* 281. <https://doi.org/10.1098/rspb.2014.1988>
- 49 Ritpitakphong, U., Falquet, L., Vimoltust, A., Berger, A., Metraux, J.P., L'Haridon, F., 2016. The  
50 microbiome of the leaf surface of *Arabidopsis* protects against a fungal pathogen. *The New*  
51 *Phyt* 210, 1033–43. <https://doi.org/10.1111/nph.13808>

- 1 Rohland, N., Reich, D., 2012. Cost-effective, high-throughput DNA sequencing libraries for  
2 multiplexed target capture. *Genome Res* 22, 939–946.  
3 <https://doi.org/10.1101/gr.128124.111>
- 4 Santhanam, R., Luu, V.T., Weinhold, A., Goldberg, J., Oh, Y., Baldwin, I.T., 2015. Native root-  
5 associated bacteria rescue a plant from a sudden-wilt disease that emerged during  
6 continuous cropping. *Proc Natl Acad Sci* 112, E5013-5020.  
7 <https://doi.org/10.1073/pnas.1505765112>
- 8 Seren, Ü., Grimm, D., Fitz, J., Weigel, D., Nordborg, M., Borgwardt, K., Korte, A., 2017. AraPheno: a  
9 public database for *Arabidopsis thaliana* phenotypes. *Nucl Acids Res* 45, D1054–D1059.  
10 <https://doi.org/10.1093/nar/gkw986>
- 11 Thiergart, T., Durán, P., Ellis, T., Vannier, N., Garrido-Oter, R., Kemen, E., Roux, F., Alonso-Blanco, C.,  
12 Ågren, J., Schulze-Lefert, P., Hacquard, S., 2020. Root microbiota assembly and adaptive  
13 differentiation among European *Arabidopsis* populations. *Nat Ecol Evol* 4, 122–131.  
14 <https://doi.org/10.1038/s41559-019-1063-3>
- 15 Tkacz, A., Bestion, E., Bo, Z., Hortalá, M., Poole, P.S., 2020. Influence of Plant Fraction, Soil, and Plant  
16 Species on Microbiota: A Multikingdom Comparison. *mBio* 11.  
17 <https://doi.org/10.1128/mBio.02785-19>
- 18 Wagner, M.R., Lundberg, D.S., Del Rio, T.G., Tringe, S.G., Dangl, J.L., Mitchell-Olds, T., 2016. Host  
19 genotype and age shape the leaf and root microbiomes of a wild perennial plant. *Nat*  
20 *Commun* 7, 12151. <https://doi.org/10.1038/ncomms12151>
- 21 Wang, W., Yang, J., Zhang, J., Liu, Y.-X., Tian, C., Qu, B., Gao, C., Xin, P., Cheng, S., Zhang, W., Miao, P.,  
22 Li, L., Zhang, X., Chu, J., Zuo, J., Li, J., Bai, Y., Lei, X., Zhou, J.-M., 2020. An *Arabidopsis*  
23 Secondary Metabolite Directly Targets Expression of the Bacterial Type III Secretion System  
24 to Inhibit Bacterial Virulence. *Cell Host Microbe* 27, 601-613.e7.  
25 <https://doi.org/10.1016/j.chom.2020.03.004>
- 26 Weigel, D., 2012. Natural Variation in *Arabidopsis*: From Molecular Genetics to Ecological Genomics.  
27 *Plant Phys* 158, 2–22. <https://doi.org/10.1104/pp.111.189845>
- 28 Williams, T.R., Moyne, A.-L., Harris, L.J., Marco, M.L., 2013. Season, Irrigation, Leaf Age, and  
29 *Escherichia coli* Inoculation Influence the Bacterial Diversity in the Lettuce Phyllosphere. *PLoS*  
30 *One* 8, e68642. <https://doi.org/10.1371/journal.pone.0068642>
- 31 Wippel, K., Tao, K., Niu, Y., Zgadzaj, R., Kiel, N., Guan, R., Dahms, E., Zhang, P., Jensen, D.B.,  
32 Logemann, E., Radutoiu, S., Schulze-Lefert, P., Garrido-Oter, R., 2021. Host preference and  
33 invasiveness of commensal bacteria in the *Lotus* and *Arabidopsis* root microbiota. *Nat*  
34 *Microbiol* 6, 1150–1162. <https://doi.org/10.1038/s41564-021-00941-9>
- 35