Whole-genome duplication and host genotype affect rhizosphere microbial communities in *Arabidopsis thaliana*

Julian C. B. Ponsford [a], Charley J. Hubbard [a,b], Joshua G. Harrison[a], Lois Maignien[c,d] C. Alex Buerkle [a,b], Cynthia Weinig[a,b,e]#

a. Department of Botany, University of Wyoming, Laramie, WY, USA
b. Program in Ecology, University of Wyoming, Laramie, WY, USA
c. Marine Biological Laboratory, Josephine Bay Paul Center, Woods Hole, MA, USA
d. Laboratory of Microbiology of Extreme Environments, UMR 6197, Institut Européen de la Mer, Université de Bretagne Occidentale, Plouzane, France
e. Department of Molecular Biology, University of Wyoming, Laramie, WY, USA

Corresponding author: Cynthia Weinig, cweinig@uwyo.edu

Keywords: *Arabidopsis thaliana*, whole genome duplication, multinomial modeling, plant microbe interactions

Funding: National Science Foundation grant: IOS-1444571 to LM and CW & EPS-16557226 to CW and CAB.

Abstract

The composition of complex microbial communities found in association with plants is influenced in part by host genotype. Yet, the salient genetic architecture is often unknown. Genome duplication events are common in the evolutionary history of plants, influence many important diverse plant traits, and may affect associated microbial communities. Using genotypes with experimentally induced whole genome duplication (WGD), we tested the effect of WGD on rhizosphere bacterial communities in *Arabidopsis thaliana*. Specifically, we performed 16s rRNA amplicon sequencing to characterize differences between microbiomes associated with specific host genotypes (Columbia *vs*. Landsberg) and ploidy levels (diploid *vs*. tetraploid). We modeled abundances of individual bacterial taxa by utilizing a hierarchical Bayesian framework, based on the Dirichlet and multinomial distributions. We found that host genotype and host ploidy level affected rhizosphere community composition. Further, the microbiome of the tetraploid Columbia genotype differed from that of other host genotypes. We then tested to what extent microbiomes derived from a given host genotype or ploidy level affected plant performance by inoculating sterile seedlings of each genotype with microbial communities harvested from a prior generation, using a full factorial design. We found a negative effect of the tetraploid Columbia microbiome on growth of all four plant genotypes. The findings suggest that while both host genotype and ploidy affect microbial community assembly, bacterial communities found in association with only some host genotypes may affect growth of subsequent plant generations.

Introduction

Plants and microbes have a tight feedback cycle. Changes in microbial community composition can affect plant performance and the genetics of the plant host can alter microbial community membership. (Chaparro *et al.*, 2013; Schlaeppi & Bulgarelli, 2015; Van Nuland *et al.*, 2017). The traits of bacterial taxa in plant-associated communities can have both functional and fitness effects on their hosts (Talbot *et al.*, 2014). Rhizosphere bacteria, particularly, affect many aspects of plant performance, such as increasing access to nutrients (Richardson & Simpson, 2011), relieving abiotic and biotic stress (Mendes *et al.*, 2011), and promoting growth (Bulgarelli *et al.*, 2013). Differences in the abundance of one taxon in the rhizosphere microbiome or broader can affect host plant performance (Henning *et al.*, 2016). For instance, Korir *et al.* (2017) found that increased abundance of *Bacillus megaterium* led to enhanced nitrogen access and growth of *Phaseolus vulgaris* in field conditions (Korir *et al.*, 2017). More widespread changes in the abundance of taxa in complex rhizosphere bacterial communities can also affect plant performance (Zolla *et al.*, 2013; Wagner *et al.*, 2014). Specifically, in *Arabidopsis thaliana*, differences in microbial community composition attributable to past or nearby plant communities strongly influenced host growth (Sugiyama *et al.*, 2013; Hubbard *et al.*, 2017).

Conversely, a plant host's genetic background can affect the composition of complex microbial communities consisting of thousands of taxa (Lundberg *et al.*, 2012; Marques *et al.*, 2014; Wagner *et al.*, 2016). For instance, over a range of environmental conditions, host genotype in maize explains on average ~19% of the variance in abundance of root microbial taxa (Peiffer *et al.*, 2013). Among host genotypes, allelic variation segregating at loci with diverse functions could potentially contribute to differences in rhizosphere community composition. The

contribution of specific host-plant loci and pathways has been demonstrated experimentally using genetic knock-outs or transgenic overexpression (Berendsen *et al.*, 2012; Agler *et al.*, 2016). Genetic manipulation to shift the plant circadian clock by ± 4 hours explained ~22% of the variation in rhizosphere bacterial communities among experimental *Arabidopsis* lines (Hubbard *et al.*, 2017). However, the number of studies characterizing causal genetic factors is small, and the extent to which specific host genetic and genomic features alter the assembly and function of microbial communities remains largely unknown, despite the demonstrably important effect of the rhizosphere microbiome on host plant performance.

Plants influence rhizosphere microbial communities via root exudation of simple and complex carbon molecules (Walker, 2003; Jones *et al.*, 2009; Dennis *et al.*, 2010). For simple symbioses consisting of one plant and one microbial taxon, such as between legumes and *Rhizobia*, the host genetic mechanisms that underlie these interactions are increasingly well characterized (Sanchez-Contreras *et al.*, 2007; Udvardi & Poole, 2013; Bravo *et al.*, 2016). How more complex genetic mechanisms affect associated root microbial communities is less clear.

Whole-genome duplication (WGD) is one genetic feature of host plants that could potentially influence microbial community composition via root exudation. WGD can lead to changes in cell size, life-history, and physiology, which may influence numerous other phenotypes (Stebbins, 1971; Garbutt & Bazzaz, 1983). WGD also occurs spontaneously in natural populations and is ubiquitous in the evolutionary history of plants and in the domestication history of many crop species (Ramsey & Ramsey, 2014).

In laboratory settings, colchicine is used to induce autopolyploidization and create stable lines of tetraploids in *Arabidopsis thaliana*. This mutation theoretically produces no genetic

changes besides those associated with genome duplication (Breuer *et al.*, 2007). Therefore, the comparison of rhizosphere microbiomes between *A. thaliana* tetraploid genotypes developed from inbred diploid lines provides a basis for testing the effects of host ploidy level on microbial community composition, without the confounding effects of allelic variation or variable fixed heterozygosity between tetraploids and diploid progenitors. Tetraploidy can induce changes in core metabolic pathways such as the tricarboxylic acid cycle (TCA), malate and citrate concentrations, and potassium uptake in *A. thaliana* (Chao *et al.*, 2013; Vergara *et al.*, 2016). These phenotypic shifts may alter root exudate profiles or other phenotypes that consequentially change microbial colonization and community composition.

In the current study, we were interested in characterizing the effects of host WGD on rhizosphere bacterial communities and testing how changes in bacterial community composition affected plant performance. We hypothesize that WGD will affect bacterial communities due to changes in host physiology induced by polyploidization. Further, if tetraploid plants assemble disparate communities compared to diploid hosts, then we predict that differences in these communities could affect plant performance.

Material and Methods

Plant material and growth conditions

To test the effects of whole genome duplication and plant genotype on rhizosphere bacterial community composition and plant host performance, we selected two *A. thaliana* diploid genotypes: Columbia (Col-2x) and Landsberg *erecta* (Ler-2x) and their tetraploid counterparts (Col-4x CS922178) and (Ler-4x CS3900). Tetraploid genotypes were multiple generations removed from initial colchicine treatment, and therefore can be assumed to be

mutationally stable (Hollister *et al.*, 2012). In our first experiment, each genotype was planted into sterilized potting mix and inoculated with a microbial community from the Catsburg region of Durham, North Carolina, USA (36.0622^{0}N, -78.8496^{0}W), a site with a history of *A. thaliana* growth. The microbiomes associated with each host plant genotype were characterized by 16s rRNA gene amplicon sequencing (experiment one) and tested for effects on growth of a second generation of plants, in experiment two.

In both experiments, seeds were surface sterilized using a solution of 15% bleach, 0.1% Tween, and autoclave sterilized reverse osmosis purified water (RO H_2O). Seeds were thoroughly rinsed in RO H_2O to remove any remaining detergents. Seeds were placed in 1.5mL tubes containing 1mL RO H_2O and stored at 4°C for seven days and then placed on greenhouse benches under natural 14 hr photoperiods to induce synchronous germination (Kerwin *et al.*, 2015). On the day root radicles were observed, seeds were transferred to 2-inch diameter net pots filled with a mixture of autoclaved Redi-Earth Potting Mix (Sungro Horticulture, Agawam, MA, USA) and 2ml of liquid inoculate (described below). To ensure sterility, potting mix was autoclaved on a wet cycle for 60 min at 121°C, allowed to rest for at least one hour and autoclaved for another 60 min. No microbes could be cultured on tryptic soy agar media using serial dilutions of autoclaved soil as inoculum. To create our inoculants for the first experiment, 60 g of the Catsburg soil was mixed with 540ml of autoclaved RO H_2O and sieved through 1000µm, 212µm and 45µm sterile sieves to remove nematodes that could potentially affect plant performance (van de Voorde *et al.*, 2012). The Catsburg site not only has a well-documented occurrence of naturalized *A. thaliana* populations but also has a soil pH approximately identical to the potting substrate, which minimized potential selection by the common soil matrix on

microbial community composition (Mauricio, 1998; Fierer & Jackson, 2006; Lauber *et al.*, 2009).

For the second experiment, we used plant-conditioned soil collected from each host genotype to create inoculate. In all experiments, after ten days of growth, seedlings were thinned to one plant per pot. All experiments were performed at the Williams Conservatory or Agricultural Experiment Station at the University of Wyoming.

Experimental design

Experiment 1: Effects of host plant genotype and ploidy on microbial community

To measure the influence of host ploidy on rhizosphere bacterial community composition, 20 replicates of Col-2x, Col-4x, Ler-2x, and Ler-4x were planted in a fully randomized four block design. To avoid confounding the effects of ploidy with developmental stage, eight rhizosphere samples from each genotype were collected on day 24, by which time all plants displayed visible elongation of the primary inflorescence from the apical meristem (N = 32). Unharvested plants were allowed to grow until senescence to quantify additional phenotypes. To account for variation in microbial communities due to greenhouse conditions and intrinsic variation in soil independent of plant host effects, we collected soil from empty pots containing only growth substrate that were potted simultaneously with our experiment (N = 7).

Experiment 2: Microbial effects on plant performance

To investigate the functional consequences of differences in bacterial community composition that were attributable to host plant genotype, we used a fully factorial design with four host plant genotypes grown in four host-influenced microbiomes (harvested from the host genotypes in expt. 1). 20 replicates of each plant genotype were grown in sterilized potting mix inoculated with one of the four rhizosphere microbiomes from plant-conditioned soil in a fully randomized block design. Plants were checked daily for bolting and flowering. Upon the first plant bolting, a subset of plants was collected for above- and belowground biomass (N =398) to measure plant growth and resource allocation.

DNA extraction and sequencing

Rhizosphere samples were collected from the root surface as described in (Bulgarelli *et al*., 2012) by first uprooting plants, manually agitating plants, removing loosely adhering soil particles, subsequently placing roots with closely adhering soil in a buffer of sterile PBS with 0.1% Silwet in a 15 mL tube, and vortexing for 10 min on maximum speed. Rhizosphere samples were centrifuged at 3714 rcf, and a total soil mass of no more than 250 mg was transferred from each tube into a sterile Qiagen PowerSoil (Qiagen, Valencia, CA) bead tube. DNA was extracted following manufacturer's instructions. Extracted DNA was shipped to the Marine Biological Laboratories (Woods Hole, MA, USA) for amplification and sequencing of the V4-V5 (primers 518f and 926r) region of the 16s rRNA gene (Huse *et al*., 2014) on an Illumina MiSeq (Illumina, San Diego, CA, USA).

Sequence and data analysis

We used the R package *dada2* (ver. 1.5.8) to filter and trim reads based on quality, estimate the error rate using 1 000 000 reads, dereplicate reads, infer amplicon sequence variants (ASV), merge paired end reads, remove chimeras, and assign unique sequences to taxa using the Silva 16s database (ver. 128) (Quast *et al.*, 2013; Callahan *et al.*, 2016). Next, we ascertained the abundance of operational taxa by quantifying unique sequences and amplicon sequence variants using the default settings in the *microbiome* R package (Shetty *et al.*, 2017). We used *adonis* (Permutational Multivariate Analysis of Variance Using Distance Matrices) on Jaccard (presence-absence) and Bray-Curtis (abundance) dissimilarities between community samples to characterize the effects of genotype (Col-2x & Col-4x vs. Ler-2x & Ler-4x) and ploidy (Col-2x & Ler-2x vs. Col-4x & Ler-4x), and calculated Shannon's diversity index to characterize intra- and inter-genotype variation using the *vegan* and *phyloseq* R packages (Dixon, 2003; McMurdie & Holmes, 2013; R Development Core Team, 2018). Raw sequences were uploaded to the NIH NCBI Short Read Archive under project ID PRJNA474006.

Dissimilarity analyses identify coarse-grained changes in community composition across treatments, which arise from the combined ffects of all community members. Yet, significant differences in relative abundance may exist when comparing individual taxa between treatments that are not discernible when aggregating effects of total communities. To test for differences in the relative abundances of individual taxa between groups (contrasts of host ploidy and host genotype), we used a hierarchical Bayesian modelling approach reliant upon the Dirichlet and multinomial distributions (Holmes *et al.*, 2012; Harrison *et al.*, 2019). The Dirichlet-multinomial model (DMM) explicitly accounts for the compositional nature of community sampling (whether in sequencing data or any finite number of taxa observations from an assemblage), readily allows information-sharing among replicates within a category and obtains estimates of the relative

abundance of each microbial taxon, while propagating uncertainty in those estimates. We estimated differences in the relative abundance of each microbial taxon between experimental treatments through comparison of parameter estimates as per Harrison et al. 2019.

Briefly, DMM estimates the multinomial parameters describing the relative abundances of each taxon in a replicate, denoted as a vector of parameters (***p***). These multinomial parameters were informed by a Dirichlet distribution, with parameters characterizing the expected frequencies of each taxon ($\boldsymbol{\pi}^* \Theta$), where $\boldsymbol{\pi}$ is a vector describing the expected relative abundance of each taxon in the sampling group and Θ is an intensity parameter that models among-replicate variation (Fordyce *et al.*, 2011; Harrison *et al.*, 2019). To quantify differences in the relative abundance of taxa between groups, we calculated the posterior probability distribution for the difference in π_i parameters between groups for some taxon *i*. The probability of an effect of treatment on a focal taxon can be determined by the location of zero in this distribution of differences. Following convention, if 95% or more of the distribution did not overlap zero, then there was little evidence that the relative abundance of the focal taxon differed between treatment groups.

The DMM offers several advantages over existing analytical methodologies. First, parameters were estimated while propagating uncertainty, thus avoiding cumbersome multiple comparison correction, and precluding the use of p-values. Second, information was shared among replicates and treatment groups. Third, rarefaction is unnecessary because analyses are performed on taxon frequencies, which are estimated using information from all replicates within a sampling group. Fourth, the Dirichlet and multinomial distributions have interdependent parameters that reflect the compositional nature of sequencing data, thus we model the

composition as a whole as opposed to estimating the relative abundance of each taxon separately, as in DESeq2 (Love *et al.*, 2014).

DMM was specified in the Stan probabilistic programming language and implemented through the rstan v2.18.2 package in the R statistical computing environment v3.6.0 (Carpenter *et al.*, 2017; Stan Development Team, 2018); full code can be found in the Supplementary Material. For each of four chains, the sampler was run for 1500 steps as a burn-in period and was followed by an additional 1000 steps (a total of 4000 samples were drawn from the posterior distributions of focal parameters [$\boldsymbol{\pi}$ from each sampling group]). The Gelman-Rubin statistic was computed to measure convergence among chains (Besag *et al.*, 1989). Separate models were constructed for the following comparisons: all diploid versus all tetraploid hosts, Columbia versus Landsberg *erecta*, and Columbia tetraploid versus all other treatments (pooled).

For those taxa that differed credibly among our treatment groups, we used NCBI BLAST to confirm 16s rRNA identifications generated using the SILVA database (Altschul *et al.*, 1990; Johnson *et al.*, 2008). We report species level identification in cases where reads align 100% with entries in the 16s ribosomal RNA sequences database, and report bacterial family if genus level resolution is not corroborated by SILVA (Agarwala *et al.*, 2018).

To assess effects on plant performance in experiment 2, we used fixed effect, three-way ANOVAs, where inoculum, genotype, and block were the explanatory variables. For purposes of data representation, residuals were plotted after statistically accounting for the effect of genotype and block. Finally, we used planned comparisons to contrast plant performance between plants grown in the Col-2x versus Col-4x inoculums and Ler-2x versus Ler-4x inoculums.

Results

Sequencing results

After quality filtering, sample inference and chimera removal in *dada2* as well as removal of bacterial taxa occurring in plant-less pots and removal of 16s rRNA sequences from chloroplasts, mitochondria, and Eukaryotes, 1 647 741 processed reads remained from an initial 3 282 410 raw reads. We retained 36 033 reads to 143 254 reads per plant rhizosphere (**Supplemental Figure 1)**. In total, there were 2689 microbial taxa (ASVs) present.

Genetic background (Col versus Ler) affected both the presence (P = 0.021), and abundance of taxa (P = 0.024), and explained 5.8% and 7.2%, of the variance between groups, respectively (**Figure 1)**. In the first experiment, the bacterial composition of our plant-less blanks differed significantly from rhizosphere communities of plants, based on PERMANOVA analyses of Jaccard and Bray-Curtis dissimilarities (P < 0.01).

Ploidy level did not significantly affect whole rhizosphere community composition based on either Jaccard and Bray-Curtis dissimilarities of the entire community (PERMANOVA, Table 1), whether we considered all taxa (Supplemental Figure 1; PERMANOVA, P = 0.52 & P = 0.456), or only the 742 most abundant taxa, which were represented by > 100 reads across all replicates. However, because we observed differential growth effects of the Col-4x vs. all other rhizospheres, we performed a post hoc comparison of the rhizosphere microbiome of the Col-4x genotype to that of other sampling groups including Col-2x. We found that Col-4x differed significantly from all other rhizospheres based on both Jaccard and Bray-Curtis dissimilarities (Table 1; PERMANOVA, P = 0.018 & P = 0.013).

Shannon's diversity index revealed no significant differences between genotype (P = 0.911), between ploidy levels (P = 0.565) or between ploidy level nested within genotype (Landsberg: P = 0.768, Columbia: P = 0.564), based on Welch's two sample t-test. The average diversity across all samples was 4.59 (+/-0.043, n = 29).

We utilized DMM to identify individual taxa with significantly different relative abundance due to host genotype and host ploidy level. This model allowed us to quantify differences in the abundance of individual taxa among treatments that might not be evident in whole community dissimilarity analyses. Based on this approach we identified 25 taxa that were significantly enriched in both Ler genotypes and 29 taxa that were more abundant in the Col genotypes (**Figure 2a)**. A taxon of genus *Pedobacter* was the most highly enriched in Ler associated bacterial communities with an 8-fold increase. In Col bacterial communities, the most enriched taxon was *Pseudomonas* at a 4.8-fold increase in abundance. DMM also identified 17 taxa that were significantly enriched in all diploid genotypes. Of these taxa, the most enriched taxon was a member of the genus *Bacillus*, enriched 3.1-fold relative to tetraploid communities. Tetraploid communities' most enriched member was *Mucilaginibacter*, which was present at 4-fold greater abundance compared to diploid rhizosphere bacterial community abundance. 23 taxa were significantly enriched across both tetraploid genotypes (**Figure 2b)**. Our model identified 16 taxa as higher abundance in the Columbia tetraploid rhizosphere when compared to 30 taxa enriched in all other bacterial microbiomes. In Col-4x a bacterium identified as *Sphingomonas* had the highest change in abundance at 6-fold. Conversely, a taxon of genus *Pedobacter* was enriched in all other communities. This bacterium appeared at 5.8-fold greater frequency in these communities than Col-4x. Across all comparisons, members of the phylum Proteobacteria were more commonly differentially abundant than any other phylum.

Effects of rhizosphere microbes on plant performance

When bacterial communities from plant-conditioned soil were harvested following experiment one, and used as inoculant in experiment 2, plants grown in soils inoculated with Col-4x rhizosphere had significantly lower aboveground and belowground biomass compared to plants grown in soils inoculated with Col-2x rhizosphere ($P < 0.001$ & $P = 0.012$) or rhizospheres influenced by either of the two Ler genotypes (**Figure 3ab**; $P < 0.001$ & $P < 0.001$). The soil microbiome had no effect on phenological characteristics or fruit number (**Supplemental Table 2**).

Discussion

Whole genome duplication is estimated to have occurred across 30-70% of the angiosperm phylogeny over its evolutionary history and is a common genetic feature of many economically important plants (Alix *et al.*, 2017). We found that host genotype and WGD influenced numerous individual taxa within the rhizosphere microbiome. Despite this, dissimilarity analyses identified no differences in overall community composition attributable to WGD. Furthermore, host-genotype specific effects on rhizosphere microbiomes influenced plant growth. In contrast to prior studies (Bai *et al.*, 2015), our experiments also utilized a soil with a history of *Arabidopsis* occurrence as inoculant. This should in principle allow genotypes to assemble root microbiota from existing and potentially coevolved plant-proximal microbes.

Previous studies examining the root microbiota found in association with common lab accessions of *A. thaliana* have found a link between genotype and rhizosphere bacteria (Jones *et*

al., 2009; Lundberg *et al.*, 2012). These studies report a < 10% shift in composition between the Ler and Col, which we corroborate here (**Figure 1**). As noted above, we did not find a significant effect of ploidy on bacterial community composition based on dissimilarity analysis but did identify individual taxa influenced by genome duplication (**Figure 2a**). This provides support for the hypothesis that these specific taxa are directly associated with genome duplication, although the causal physiological mechanisms remain unclear. Previous studies suggest that carbon root exudates strongly affect the presence-absence and abundance of microbial taxa in the rhizosphere (Chaparro *et al.*, 2013; Staley *et al.*, 2017), and these exudates can be influenced by genome duplication (Vergara *et al.*, 2016). Consequently, shifts in root exudates, or other metabolites produced by the plant, in response to WGD could be partially responsible for the association between ploidy and the relative abundance of microbial taxa that we observed (Edger *et al.*, 2015; Lebeis *et al.*, 2015). We conclude, that while host ploidy level affects the abundance of some taxa, it does not lead to broad-scale changes in the rhizosphere microbiome (**Figure 1**).

We found that differences in the rhizosphere bacterial microbiome lead to differences in host performance [**Figure 3**: 23, 61, 62]. Here, plants grown in soils inoculated with microbial rhizosphere communities harvested from Col-4x plants had reduced vigor compared to plants grown in soils inoculated with microbial communities from all other genotypes (**Figure 3**). Although 16 ASVs were identified as significantly enriched in Col-4x root associated communities, the nature of 16s rRNA amplicon data prevents us from obtaining the species level resolution necessary to make physiological conclusions about these taxa (Callahan *et al.*, 2017). Therefore, further study is required to elucidate the effects of these bacteria on *Arabidopsis*' growth. These taxa may be especially impactful in the context of mixed populations of diploid

and tetraploid individuals, where stressful conditions could be compounded or mitigated by host-associated microbes (Segraves, 2017).

Conclusions

In sum, we have shown that whole genome duplication influences rhizosphere bacterial community composition by affecting the relative abundance of certain bacterial taxa. In some instances, plant ploidy level affects microbial function, as indicated by enhanced above and belowground plant biomass in the Col-2x versus Col-4x genotypes. Multiple bacterial taxa are affected by host ploidy level, irrespective of host genetic background. The ASVs enriched (or reduced) in the rhizosphere of tetraploid hosts could help explain the advantages tetraploids appear to have in stressful environments. Given the prevalence of WGD among plants in both natural and agricultural systems, our results highlight a novel mechanism by which plant evolutionary history influences the root-associated microbiome, thereby affecting plant phenotype.

Acknowledgments

Computing was performed in the Teton Computing Environment at the Advanced Research Computing Center, University of Wyoming, Laramie (https://doi.org/10.15786/M2FY47). We would like to thank Lindsay Leverett for collecting soil, and Meredith Pratt and Ryan Pendleton for help with experimental and greenhouse maintenance.

Conflict of Interests

The authors report no conflicts of interest.

References

Agarwala R, Barrett T, Beck J *et al.*, 2018. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research* **46**, D8–D13.

Agler MT, Ruhe J, Kroll S *et al.*, 2016. Microbial Hub Taxa Link Host and Abiotic Factors to Plant Microbiome Variation. (MK Waldor, Ed,). *PLoS biology* **14**, e1002352.

Alix K, Gérard PR, Schwarzacher T, Heslop-Harrison JSP, 2017. Polyploidy and interspecific hybridization: Partners for adaptation, speciation and evolution in plants. *Annals of Botany* **120**, 183–194.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ, 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403–410.

Badri D V., Zolla G, Bakker MG, Manter DK, Vivanco JM, 2013. Potential impact of soil microbiomes on the leaf metabolome and on herbivore feeding behavior. *New Phytologist* **198**, 264–273.

Bai Y, Müller DB, Srinivas G *et al.*, 2015. Functional overlap of the Arabidopsis leaf and root microbiota. *Nature* **528**, 364–369.

Berendsen RL, Pieterse CMJ, Bakker PAHM, 2012. The rhizosphere microbiome and plant health. *Trends in Plant Science* **17**, 478–486.

Besag J, Green P, Higdon D, Mengersen K, 1989. Institute of Mathematical Statistics is

collaborating with JSTOR to digitize, preserve, and extend access to Statistical Science. ® www.jstor.org. *Statistics* **10**, 409–435.

Bravo A, York T, Pumplin N, Mueller LA, Harrison MJ, 2016. Genes conserved for arbuscular mycorrhizal symbiosis identified through phylogenomics. *Nature Plants* **2**, 15208.

Breuer C, Stacey NJ, West CE *et al.*, 2007. BIN4, a Novel Component of the Plant DNA Topoisomerase VI Complex, Is Required for Endoreduplication in Arabidopsis. *THE PLANT CELL ONLINE* **19**, 3655–3668.

Bulgarelli D, Rott M, Schlaeppi K *et al.*, 2012. Revealing structure and assembly cues for Arabidopsis root-inhabiting bacterial microbiota. *Nature* **488**, 91–5.

Bulgarelli D, Schlaeppi K, Spaepen S, van Themaat EVL, Schulze-Lefert P, 2013. Structure and Functions of the Bacterial Microbiota of Plants. *Annual Review of Plant Biology* **64**, 807–838.

Callahan BJ, McMurdie PJ, Holmes SP, 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME journal* **11**, 2639–2643.

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP, 2016. DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods* **13**, 581–583.

Carpenter B, Gelman A, Hoffman MD *et al.*, 2017. Stan: A Probabilistic Programming Language. *Journal of Statistical Software* **76**, 1–32.

Chao D-Y, Dilkes B, Luo H *et al.*, 2013. Polyploids Exhibit Higher Potassium Uptake and Salinity Tolerance in Arabidopsis. *Science* **341**, 658–659.

Chaparro JM, Badri D V, Vivanco JM, 2013. Rhizosphere microbiome assemblage is affected by plant development. *The ISME Journal* **8**, 790–803.

Dennis PG, Miller AJ, Hirsch PR, 2010. Are root exudates more important than other sources of rhizodeposits in structuring rhizosphere bacterial communities? *FEMS Microbiology Ecology* **72**, 313–327.

Dixon P, 2003. VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science* **14**, 927–930.

Edger PP, Heidel-Fischer HM, Bekaert M *et al.*, 2015. The butterfly plant arms-race escalated by gene and genome duplications. *Proceedings of the National Academy of Sciences* **112**, 8362–8366.

Fierer N, Jackson RB, 2006. The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences* **103**, 626–631.

Fitzpatrick CR, Copeland J, Wang PW, Guttman DS, Kotanen PM, Johnson MTJ, 2018. Assembly and ecological function of the root microbiome across angiosperm plant species. *Proceedings of the National Academy of Sciences of the United States of America* **115**, E1157–E1165.

Fordyce JA, Gompert Z, Forister ML, Nice CC, 2011. A Hierarchical Bayesian Approach to Ecological Count Data: A Flexible Tool for Ecologists (E Scalas, Ed,). *PLoS ONE* **6**, e26785.

Garbutt K, Bazzaz FA, 1983. Leaf demography, flower production and biomass of diploid and tetraploid populations of phlox-drummondii hook on a soil-moisture gradient. *New Phytologist* **93**, 129–141.

Harrison JG, Calder WJ, Shastry V, Buerkle CA, 2019. Dirichlet-multinomial modelling outperforms alternatives for analysis of microbiome and other ecological count data. *bioRxiv*.

Henning JA, Weston DJ, Pelletier DA, Timm CM, Jawdy SS, Classen AT, 2016. Root bacterial endophytes alter plant phenotype, but not physiology. *PeerJ* **4**, e2606.

Hollister JD, Arnold BJ, Svedin E, Xue KS, Dilkes BP, Bomblies K, 2012. Genetic Adaptation Associated with Genome-Doubling in Autotetraploid Arabidopsis arenosa (R Mauricio, Ed,). *PLoS Genetics* **8**, e1003093.

Holmes I, Harris K, Quince C, 2012. Dirichlet multinomial mixtures: Generative models for microbial metagenomics. *PLoS ONE* **7**.

Hubbard CJ, Brock MT, van Diepen LT, Maignien LL, Ewers BE, Weinig C, 2017. The plant circadian clock influences rhizosphere community structure and function. *The ISME Journal* **doi**, in press.

Huse SM, Young VB, Morrison HG *et al.*, 2014. Comparison of brush and biopsy sampling methods of the ileal pouch for assessment of mucosa-associated microbiota of human subjects. *Microbiome* **2**, 5.

Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL, 2008. NCBI BLAST: a better web interface. *Nucleic acids research* **36**.

Jones DL, Nguyen C, Finlay RD, 2009. Carbon flow in the rhizosphere: Carbon trading at the soil-root interface. *Plant and Soil* **321**, 5–33.

Kerwin R, Feusier J, Corwin J *et al.*, 2015. Natural genetic variation in *Arabidopsis thaliana* defense metabolism genes modulates field fitness. *eLife* **2015**, 1–28.

Korir H, Mungai NW, Thuita M, Hamba Y, Masso C, 2017. Co-inoculation Effect of Rhizobia and Plant Growth Promoting Rhizobacteria on Common Bean Growth in a Low Phosphorus Soil. *Frontiers in Plant Science* **08**, 141.

Lauber CL, Hamady M, Knight R, Fierer N, 2009. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Applied and Environmental Microbiology* **75**, 5111–5120.

Lebeis SL, Paredes SH, Lundberg DS *et al.*, 2015. Salicylic acid modulates colonization of the root microbiome by specific bacterial taxa. *Science* **349**, 860–864.

Love MI, Huber W, Anders S, 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550.

Lundberg DS, Lebeis SL, Paredes SH *et al.*, 2012. Defining the core Arabidopsis thaliana root microbiome. *Nature* **488**, 86–90.

Marques JM, da Silva TF, Vollu RE *et al.*, 2014. Plant age and genotype affect the bacterial community composition in the tuber rhizosphere of field-grown sweet potato plants. *FEMS Microbiology Ecology* **88**, 424–435.

Mauricio R, 1998. Costs of Resistance to Natural Enemies in Field Populations of the Annual

Plant Arabidopsis thaliana. *The American Naturalist* **151**, 20–28.

McMurdie PJ, Holmes S, 2013. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data (M Watson, Ed,). *PLoS ONE* **8**, e61217.

Mendes R, Kruijt M, de Bruijn I *et al.*, 2011. Deciphering the Rhizosphere Microbiome for Disease-Suppressive Bacteria. *Science* **332**, 1097–1100.

Van Nuland ME, Bailey JK, Schweitzer JA, 2017. Divergent plant–soil feedbacks could alter future elevation ranges and ecosystem dynamics. *Nature Ecology & Evolution* **1**, 0150.

Peiffer JA, Spor A, Koren O *et al.*, 2013. Diversity and heritability of the maize rhizosphere microbiome under field conditions. *Proceedings of the National Academy of Sciences* **110**, 6548–6553.

Quast C, Pruesse E, Yilmaz P *et al.*, 2013. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research* **41**, D590–D596.

R Development Core Team, 2018. R: a language and environment for statistical computing. [3.5].

Ramsey J, Ramsey TS, 2014. Ecological studies of polyploidy in the 100 years following its discovery. *Philosophical Transactions of the Royal Society B: Biological Sciences* **369**, 20130352–20130352.

Richardson AE, Simpson RJ, 2011. Soil Microorganisms Mediating Phosphorus Availability Update on Microbial Phosphorus. *Plant Physiology* **156**, 989–996.

Sanchez-Contreras M, Bauer WD, Gao M, Robinson JB, Allan Downie J, 2007. Quorum-sensing

regulation in rhizobia and its role in symbiotic interactions with legumes. *Philosophical Transactions of the Royal Society B: Biological Sciences* **362**, 1149–1163.

Schlaeppi K, Bulgarelli D, 2015. The Plant Microbiome at Work. *Molecular Plant-Microbe Interactions MPMI* **212**, 212–217.

Segraves KA, 2017. The effects of genome duplications in a community context. *New Phytologist* **215**, 57–69.

Shetty SA, Hugenholtz F, Lahti L, Smidt H, de Vos WM, 2017. Intestinal microbiome landscaping: Insight in community assemblage and implications for microbial modulation strategies. *FEMS Microbiology Reviews* **41**, 182–199.

Staley C, Ferrieri AP, Tfaily MM *et al.*, 2017. Diurnal cycling of rhizosphere bacterial communities is associated with shifts in carbon metabolism. *Microbiome* **5**, 65.

Stan Development Team, 2018. rstan: the R interface to Stan v.

Stebbins GL, 1971. *Chromosomal Evolution in Higher Plants*. Edward Arnold Ltd., London.

Sugiyama A, Bakker MG, Badri D V., Manter DK, Vivanco JM, 2013. Relationships between Arabidopsis genotype-specific biomass accumulation and associated soil microbial communities. *Botany* **91**, 123–126.

Talbot JM, Bruns TD, Taylor JW *et al.*, 2014. Endemism and functional convergence across the North American soil mycobiome. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 6341–6.

Udvardi M, Poole PS, 2013. Transport and Metabolism in Legume-Rhizobia Symbioses. *Annual*

Review of Plant Biology **64**, 781–805.

Vergara F, Kikuchi J, Breuer C, 2016. Artificial Autopolyploidization Modifies the Tricarboxylic Acid Cycle and GABA Shunt in Arabidopsis thaliana Col-0. *Scientific reports* **6**, 26515.

van de Voorde TFJ, van der Putten WH, Bezemer TM, 2012. Soil inoculation method determines the strength of plant-soil interactions. *Soil Biology and Biochemistry* **55**, 1–6.

Wagner MR, Lundberg DS, Coleman-Derr D, Tringe SG, Dangl JL, Mitchell-Olds T, 2014. Natural soil microbes alter flowering phenology and the intensity of selection on flowering time in a wild Arabidopsis relative (J Klironomos, Ed,). *Ecology Letters* **17**, 717–726.

Wagner MR, Lundberg DS, Del Rio TG, Tringe SG, Dangl JL, Mitchell-Olds T, 2016. Host genotype and age shape the leaf and root microbiomes of a wild perennial plant. *Nat Commun* **7**, 1–15.

Walker TS, 2003. Root Exudation and Rhizosphere Biology. *PLANT PHYSIOLOGY* **132**, 44–51.

Zolla G, Badri D V., Bakker MG, Manter DK, Vivanco JM, 2013. Soil microbiomes vary in their ability to confer drought tolerance to Arabidopsis. *Applied Soil Ecology* **68**, 1–9.

Table 1: Effects of plant (vs. soil only), plant genotype, and ploidy level on microbial community composition as estimated from PERMANOVA.

Comparison	Jaccard	Bray Curtis
Planted v. Unplanted	P = 0.004**	P = 0.001**
Landsberg v. Columbia	P = 0.021*	P = 0.016*
Diploid v. Tetraploid	P = 0.530	P = 0.449
Col-2x v. Col-4x	P = 0.165	P = 0.158
Ler-2x v. Ler-4x	P = 0.978	P = 0.959
Col-4x v. All Others	P = 0.018*	P = 0.013*

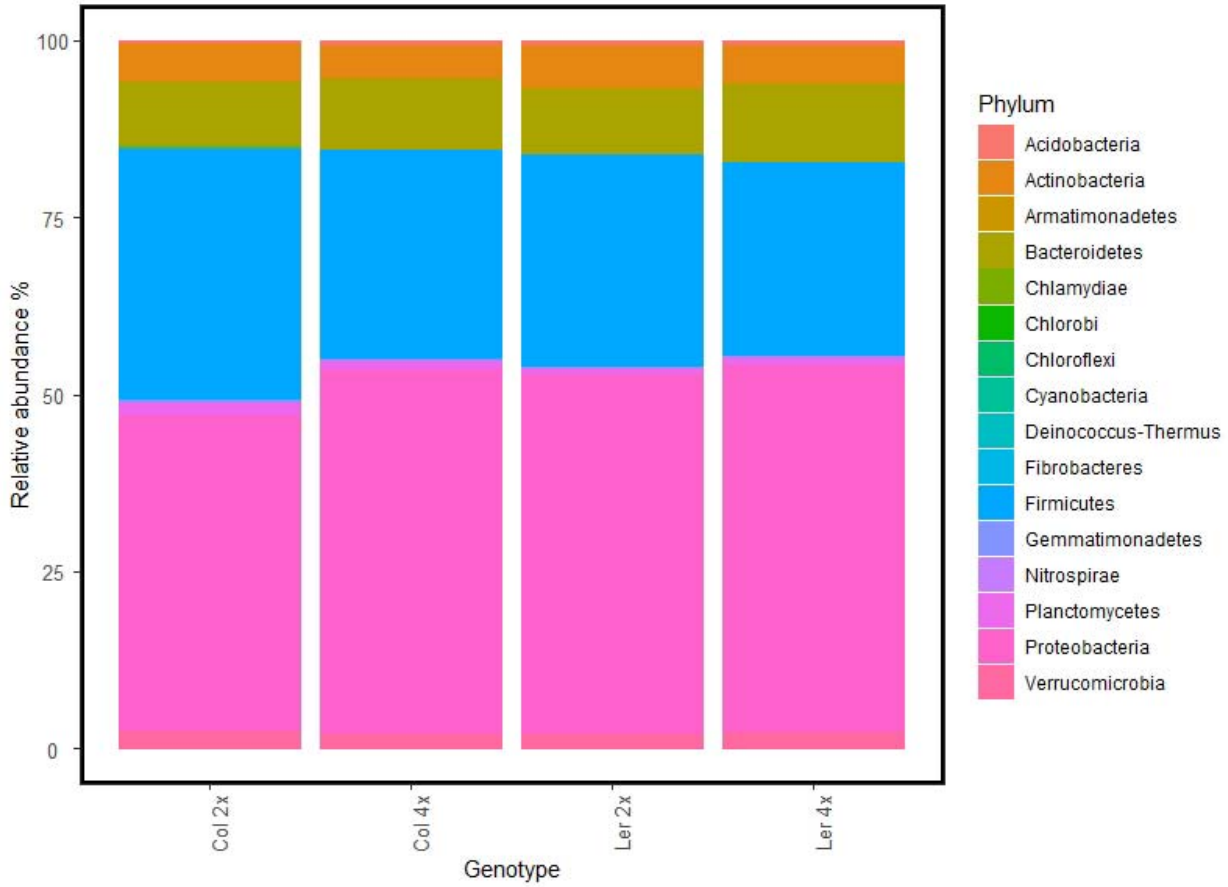


Figure 1: Percent relative abundance as estimated from read counts separated by each genotype. On average, bacterial communities were dominated by Proteobacteria (49.9%), Firmicutes (30.3%), Bacteroidetes (9.8%) and Actinobacteria (5.1%). The remaining twelve phyla made up less than 5% of each community.

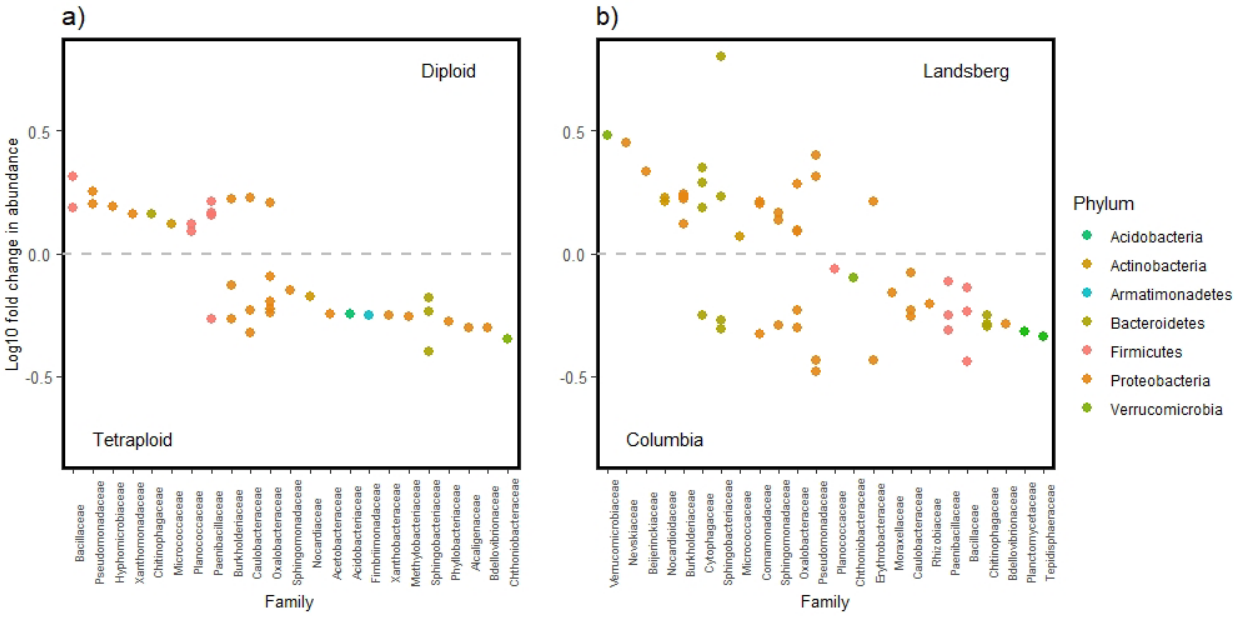


Figure 2: Genotype and ploidy differentially affects rhizosphere bacterial community abundances between the Col and Ler genotypes. **a)** Families identified as more abundant in the Col rhizosphere are shown above grey line, and those more abundant in the Ler rhizosphere appear below grey line. **b)** Families identified as more abundant in the diploid rhizosphere are shown above grey line, and those more abundant in the rhizosphere of tetraploids appear below grey line. Log10 fold change estimates were calculated from multinomial estimates of the difference in parameter $\boldsymbol{\pi}$ (relative taxon abundance), between treatments. Points represent individual genera within families.

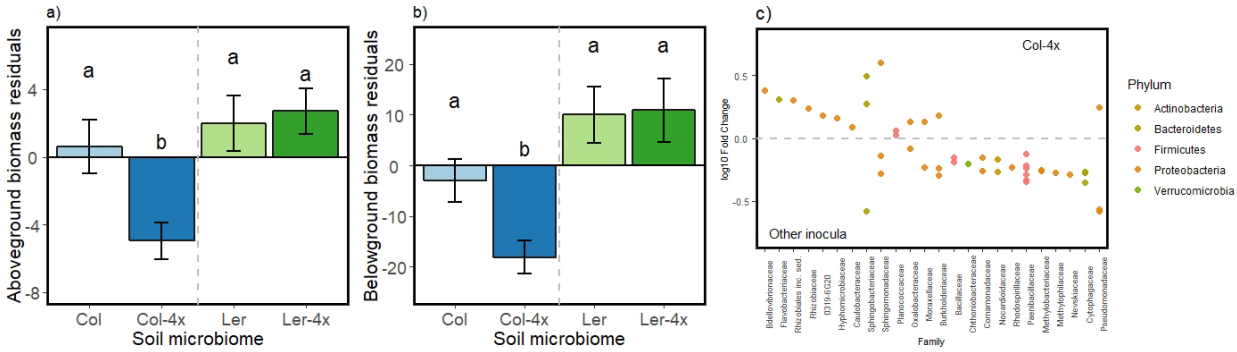


Figure 3: **a,b)** Residuals of observed growth differences in above and belowground biomass for plants grown in soils inoculated with microbiomes shaped by each genotype, with effects of block removed. All genotypes inoculated with Columbia tetraploid microbiome had significantly reduced above and below ground biomass. **c)** Families identified as more abundant in the Columbia tetraploid rhizosphere are shown above grey line. Those more abundant in the rhizosphere of all other inocula appear below grey line. Log10 fold change estimates were calculated from multinomial estimates of the difference in parameter $\boldsymbol{\pi}$ (relative taxon abundance), between treatments. Points represent individual genera within each family.

Supplemental Tables and Figures

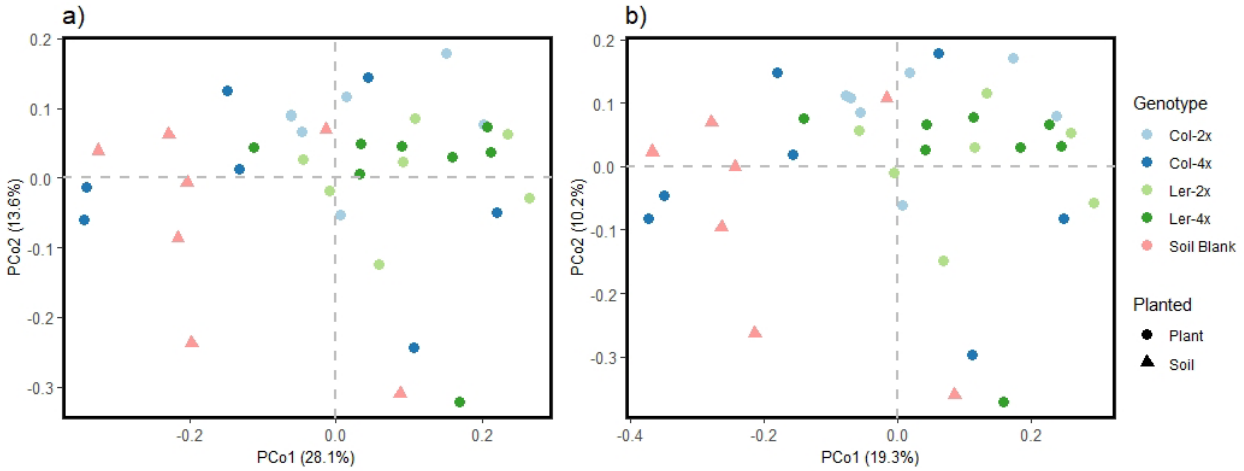
Supplemental Table 1: Number of reads at each step of sequence data processing

Processing Step	Number of Reads
Initial	3 282 410
Filtering and Trimming, Sample Inference, Chimera Removal	1 985 490

Removal of Chloroplast, Mitochondrial, Eukaryotic DNA and Plant-less blanks	1 647 741

Supplemental Table 2: Effect of inoculum shaped by plant genotype on plant performance

Trait	P - Value
Aboveground biomass	< 0.001
Belowground biomass	< 0.001
Root : Shoot	**ns**
Days to germination	**ns**
Days to flowering	**ns**
Seed Mass	**ns**



Supplemental Figure 1:a) Principal coordinate analysis of Jaccard dissimilarities (n=29). Rhizosphere community composition does not differ significantly by host plant ploidy level (P =

0.53) but does differ between genetic backgrounds (Col vs. Ler backgrounds) ($P = 0.021$). **b)** Principal coordinate analysis of Bray-Curtis dissimilarities ($n = 29$). Rhizosphere community abundance does not significantly differ between host ploidy levels ($P = 0.45$) but does differ between genetic backgrounds (Col vs. Ler backgrounds) ($P = 0.016$). Both dissimilarity analyses show significant divergence from unplanted soil ($P < 0.01$)