

Current Biology

Genetic determinants of switchgrass-root-associated microbiota in field sites spanning its natural range

Highlights

- A large proportion of root-associated bacteria show heritable variation
- Genotypes growing in their native habitat show affinity to the local microbiota
- Microbiome acquisition is a genetically complex trait varying by environment
- Root microbiome variation correlates with variation in host immunity phenotypes

Authors

Joseph A. Edwards,
Usha Bishnoi Saran,
Jason Bonnette, ..., Felix B. Fritsch,
David B. Lowry, Thomas E. Juenger

Correspondence

j_edwards@utexas.edu (J.A.E.),
tjuenger@austin.utexas.edu (T.E.J.)

In brief

Plant roots are a hotspot for colonization by soil bacteria that can greatly impact host fitness. Edwards et al. investigate the role of host plant genotype in determining the composition of root microbiota in switchgrass, a North American native perennial species. They find a large proportion of the root microbiota is under host genetic control.

Article

Genetic determinants of switchgrass-root-associated microbiota in field sites spanning its natural range

Joseph A. Edwards,^{1,6,*} Usha Bishnoi Saran,¹ Jason Bonnette,¹ Alice MacQueen,¹ Jun Yin,¹ Tu uyen Nguyen,¹ Jeremy Schmutz,^{2,3} Jane Grimwood,² Len A. Pennacchio,³ Chris Daum,³ Tijana Glavina del Rio,³ Felix B. Frittschi,⁴ David B. Lowry,⁵ and Thomas E. Juenger^{1,*}

¹Department of Integrative Biology, University of Texas, Austin, 2415 Speedway, Austin, TX 78712, USA

²Genome Sequencing Center, HudsonAlpha Institute for Biotechnology, 601 Genome Way Northwest, Huntsville, AL 35806, USA

³Joint Genome Institute, Lawrence Berkeley National Laboratory, 91R183 1 Cyclotron Road, Berkeley, CA 94720, USA

⁴Department of Plant Science and Technology, University of Missouri, Agriculture Bldg, 52, Columbia, MO 65201, USA

⁵Department of Plant Biology, Michigan State University, 612 Wilson Road, Rm 166, East Lansing, MI 48824, USA

⁶Lead contact

*Correspondence: j_edwards@utexas.edu (J.A.E.), tjuenger@austin.utexas.edu (T.E.J.)

<https://doi.org/10.1016/j.cub.2023.03.078>

SUMMARY

A fundamental goal in plant microbiome research is to determine the relative impacts of host and environmental effects on root microbiota composition, particularly how host genotype impacts bacterial community composition. Most studies characterizing the effect of plant genotype on root microbiota undersample host genetic diversity and grow plants outside of their native ranges, making the associations between host and microbes difficult to interpret. Here, we characterized the root microbiota of a large diversity panel of switchgrass, a North American native C4 bioenergy crop, in three field locations spanning its native range. Our data, composed of 1,961 samples, suggest that field location is the primary determinant of microbiome composition; however, substantial heritable variation is widespread across bacterial taxa, especially those in the Sphingomonadaceae family. Despite diverse compositions, relatively few highly prevalent taxa make up the majority of the switchgrass root microbiota, a large fraction of which is shared across sites. Local genotypes preferentially recruit/filter for local microbes, supporting the idea of affinity between local plants and their microbiota. Using genome-wide association, we identified loci impacting the abundance of >400 microbial strains and found an enrichment of genes involved in immune responses, signaling pathways, and secondary metabolism. We found loci associated with over half of the core microbiota (i.e., microbes in >80% of samples), regardless of field location. Finally, we show a genetic relationship between a basal plant immunity pathway and relative abundances of root microbiota. This study brings us closer to harnessing and manipulating beneficial microbial associations via host genetics.

INTRODUCTION

Root-associated microbes are known to boost host plant productivity and fitness through increasing nutrient accessibility,¹ manipulating plant growth and development pathways,² and antagonizing pathogen colonization.³ Recent insight into the composition, ecology, and functional importance of the plant microbiome has greatly increased interest in the potential to harness root microbiota to sustainably increase crop resilience and yield. Microbial inoculants have historically been discussed to achieve this goal, but more recent calls for using plant breeding to enrich beneficial bacteria from the existing soil microbiota have begun to emerge. A roadblock hampering this effort is a lack of understanding about which microbes can respond to breeding practices, whether breeding can instill consistent effects on microbial assemblages across differing environments, and which genes and pathways from the host can be adjusted to modify microbiomes.

Plant root bacterial microbiomes are derived from soil-borne communities, for which membership is largely driven by environmental factors such as soil type, geography and climate,^{4,5} land use history,⁶ and seasonal variation.^{7–9} The host plant exerts additional influence over its microbiota, resulting in filtered subsets of soil microbiota often composed of consistently enriched microbial taxa on and inside root tissue (summarized by the two-step selection model¹⁰). Given that microbiota can impact plant health,^{6,11} especially under varying environmental conditions,^{12–15} it follows that the filtering process may be under selection and lead to microbe-mediated local adaptation.¹⁶

Heritable variation is required for a trait to respond to selection. Indeed, several studies suggest that rhizosphere and root microbiota vary by host genotype.^{17–22} These studies indicate the possibility of enriching for beneficial microbial associations through breeding. A challenge, however, is that most of these studies include only a few host genotypes and/or grow host plants outside of their native ranges, making the role of host

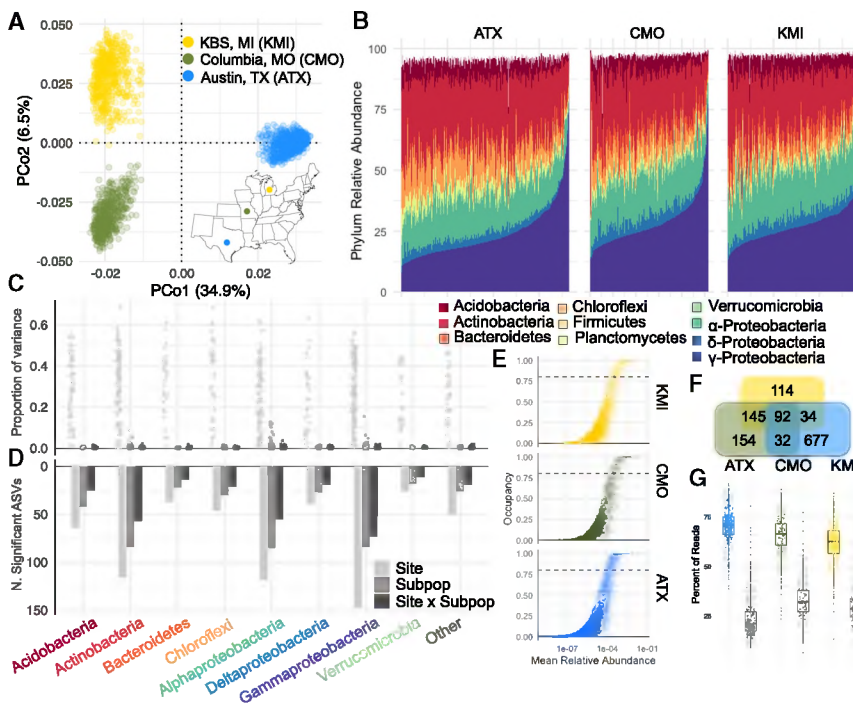


Figure 1. Field site is the primary determinant of switchgrass root microbiota composition

(A) Principal coordinate analysis based on Bray-Curtis dissimilarities. Inset: map of field locations, colors match those in the figure legend. (B) Relative abundance of phyla and Proteobacterial classes in every sample at each site. (C) Effect sizes for site, host subpopulation, and subpopulation $\times$ site for ASVs in dataset, broken down by phylum/class. (D) Number of ASVs with significant contrasts from the models displayed (C). (E) Prevalence/abundance curves for each field site. Each point represents a single ASV and the black dashed line is the 80% prevalence threshold used to call core taxa. (F) Venn diagram displaying overlaps of core microbiota from each site. (G) Fraction of reads belonging to the core microbiota (colored boxes) and the study-wide core microbiota (92 overlapping microbes from F, gray boxes). See also [Figures S1 and S2](#) and [Data S1](#).

genetics in root-microbe interactions difficult to interpret. Furthermore, given our relatively recent understanding that features of the microbiome are heritable,^{23–25} genomic loci underlying root-associated microbiome composition are still largely uncharacterized. There are notable exceptions, however, Deng et al. used the Sorghum Association Panel to uncover loci impacting rhizosphere community composition.²⁶ Bergelson et al. performed GWAS on *Arabidopsis* root (and leaf) microbiome community metrics, including richness and principal coordinates based upon community dissimilarity.²⁷ Uncovering the effects of host genetics on microbiomes across multiple native environments remains incomplete, but these studies provide exciting avenues to leverage host genetics to enrich for beneficial properties of the microbiome.

Switchgrass (*Panicum virgatum*) is a wild C4 perennial prairie grass native to North America and has been promoted as a potential biofuel crop due to its biomass yield potential when grown in marginal soil. Its biological features and environmental and economic impact have made switchgrass a popular model to investigate root-associated microbiota assembly.^{28,29} Sutherland et al. used a panel of switchgrass genotypes grown in a single location in the northeast United States to uncover the role of host genotype on rhizosphere bacterial assemblages.³⁰ This study used GWAS to uncover putative loci affecting the abundance of several bacterial families in the rhizosphere and found gene ontology (GO) enrichments for diverse sets of functions. Still, relatively little is known about how host genetics drive tightly adhering or endophytic root-associated bacterial communities.

In this study we addressed the following questions. (1) What bacteria are prominent members of the switchgrass-root-associated microbiome when plants are grown across their natural range? (2) How does the effect of host genotype compare with

that of the environment when determining the composition of root-associated bacterial microbiota? (3) Which microbial lineages show heritable variation in roots and is heritability consistent across field sites? (4) Which host genomic loci impact the abundance of root-associated bacteria? (5) Does microbial abundance show patterns of association with variation in host immune response? Answering these questions will bring us closer to harnessing and manipulating beneficial microbial associations via host genetics.

RESULTS

Field site is a primary determinant of switchgrass root microbiota composition

We used a diversity panel of fully resequenced switchgrass (*Panicum virgatum*, see [STAR Methods](#)) natural accessions that were clonally replicated and grown in field sites at Austin, TX, Columbia, MO; and Kellogg Biological Research Station, MI (from here on referred to as ATX, CMO, and KMI, respectively; [Figure 1A](#), map inset) to uncover the role of environmental variation and host genetics in shaping root microbiota composition. The field sites are geographically and climatically distinct,^{31,32} with soils that differ in physical and chemical properties ([Data S1A](#)). These plants had been established for 2 years and show signatures of local adaptation including differential survivorship and biomass accumulation across gardens as well as genetic loci associated with environmental variables.^{33,34} We first investigated the effect of field site on root bacterial microbiota. Principal coordinate analysis (PCA) of root and soil microbiomes revealed three site-specific clusters ([Figures 1A and S1](#)) and the significance of this observation was confirmed by perMANOVA ($R^2 = 0.51$, $p < 0.001$). Although the communities showed large differences between field sites at the amplicon sequence variant

(ASV) level, we found that phylum level relative abundances were largely consistent between sites (Figure 1B). Actinobacteria and Proteobacteria (namely alpha and gamma Proteobacteria) were the dominant phyla associated with switchgrass roots at every site, which is consistent with most other terrestrial plant microbiota studies.^{17–19,21}

A recent population genomic study of switchgrass found that tetraploid switchgrass can be broadly classified into three genetic subpopulations: Gulf, Midwest, and Atlantic.³³ The ranges for these subpopulations are largely distinct, with Gulf occupying habitats in the southern US, Atlantic occupying the Atlantic coast, and Midwest spread across northern latitudes. We compared the effect of field site, host subpopulation, and their interaction using linear models run on bacteria present in $\geq 50\%$ of the samples study-wide. The effect of field site was much larger than the secondary effects of host subpopulation and subpopulation $\times$ site interactions (Figure 1C). We then compared the variance explained by site between bacterial phyla/classes to better understand how experimental factors impact broader taxonomic groupings. Effect sizes were largely consistent between phyla and Proteobacteria classes, except for Chloroflexi and Actinobacteria, which showed larger effect sizes than Deltaproteobacteria ($p < 0.05$, Tukey's post hoc test; Figure 1D). The large influence of field site on ASV relative abundance was also visible in the number of ASVs, which exhibited significant differences in relative abundance across field sites (Figure 1D).

We next evaluated the relationship between ASV prevalence (i.e., the proportion of samples for which a given ASV is detected) and mean relative abundance at each site (Figure 1E). Our study used an atypically high depth of sequencing ($>250,000$ reads per library on average; Figure S1C), which gave us greater confidence in assessing presence/absence of microbes in samples. We found that ASVs with greater relative abundances were also present in a higher proportion of root microbiomes. We next defined site-specific core microbiota; to be consistent with other studies, we used a threshold of 80% prevalence.¹⁷ ATX had the most ASVs passing this prevalence threshold (Figures 1F and S1D); we expected this, because we sequenced ATX samples at greater sequencing depths than the other two sites (Figure S2; see STAR Methods). Still, we found that each site hosted overlapping core microbiota: for all three sites, an overlap of 92 core microbes was found (from here on referred to as the study-wide core). CMO and KMI shared the most ASVs. The site-specific core microbiota typically comprised $\sim 60\%$ – 70% of the total microbial population (Figure 1G, colored boxplots; Figure S3B) within each respective site, while the study-wide core microbiota made up $\sim 25\%$ of the total population (Figure 1G, gray boxplots; Figure S3B). We compared the relative abundance of the study-wide core microbiota in roots versus soil. In each site, over half of the ASVs showed significant enrichment in the roots compared with soil, while a smaller set of ASVs were enriched in soil (Figure S2A). The root-enriched ASVs largely overlapped between sites, with 46 ASVs showing significant root-enrichment at every site (Figure S2B). Thus, though field site acts as the primary determinant of switchgrass-root-associated microbiota composition, large proportions of switchgrass root bacterial microbiome are shared between locations.

Evidence of affinity between host genotypes and local microbiota

Our analyses revealed that host subpopulation and subpopulation by location interactions are determinants of microbiota composition (Figures 1C and 1D). Because the three switchgrass subpopulations are largely constrained to distinct geographic regions (Figure 2A), we hypothesized that plants grown in gardens within their subpopulation's native range would show affinity for the microbes that persist and are abundant within these ranges. If this were true, then we would expect, at each site, that more ASVs would show preferential colonization of individuals in the subpopulation grown in its native range than in the other two subpopulations. To test this, we used linear models to analyze the abundance of ASVs within each site and contrasted the abundances between the different subpopulations. We defined a specific association as occurring if the relative abundance of an ASV was significantly greater in one subpopulation compared with the other two. Gulf plants in their native ATX site had the most specific associations, while Midwest plants enriched the most ASVs in native CMO and KMI sites (Figure 2B; Data S1B), supporting the notion that subpopulations enrich more microbes in their native habitats. Furthermore, we found the ASVs with subpopulation-specific associations tended to have significantly greater prevalence (Figure 2C), but only for subpopulations growing within their native range. We compared the relative abundance of subpopulation-enriched ASVs with their relative abundance in the soil. In general, we found that microbes showing enrichment in genetic subpopulations had greater abundance in roots (Figure S2C). There were notable exceptions to this trend: Gulf-specific microbes in ATX tended to show greater relative abundance in soil than roots (46 soil-enriched vs. 21). In the KMI site, 7 Midwest-enriched microbes were enriched in soil vs. 16 in roots. These comparisons suggest that there is preferential sorting of local microbiota onto locally adapted plant genotypes, especially for highly prevalent microbes.

Switchgrass root microbiota show widespread heritable variation and genotype by environment interactions

We next used a kinship matrix denoting finer genetic relationships among individuals of the diversity panel to model how host genetic variation contributes to variation in microbe abundance. We used a suite of linear mixed effects models to partition additive genetic variance in microbial abundance (V_A) using the host population genetic relationship matrix and tested how V_A differs across the three environments ($V_{G \times E}$) with a compound symmetry model. Because microbiomes can be defined and analyzed at various taxonomic levels by aggregating counts at nodes of the bacterial phylogenetic tree, we tested the effect of host genotype on the relative abundance of taxa at various taxonomic levels. Across each taxonomic level, both $V_{G \times E}$ and V_A significantly explained variation in microbial abundance (Figure 3A; Data S1C). For microbial features within the top 10th percentile for V_A and $V_{G \times E}$, we found generally increasing estimates for V_A and decreasing estimates for $V_{G \times E}$ from phylum to ASV (Figure 3B). We next asked whether taxonomic groupings of microbes at the ASV level were more likely to be under the influence of host genetics. Significant, non-zero V_A and $V_{G \times E}$ were widespread across the microbial phylogeny, however, specific orders were overrepresented in the data (Figure 3C). Each tested

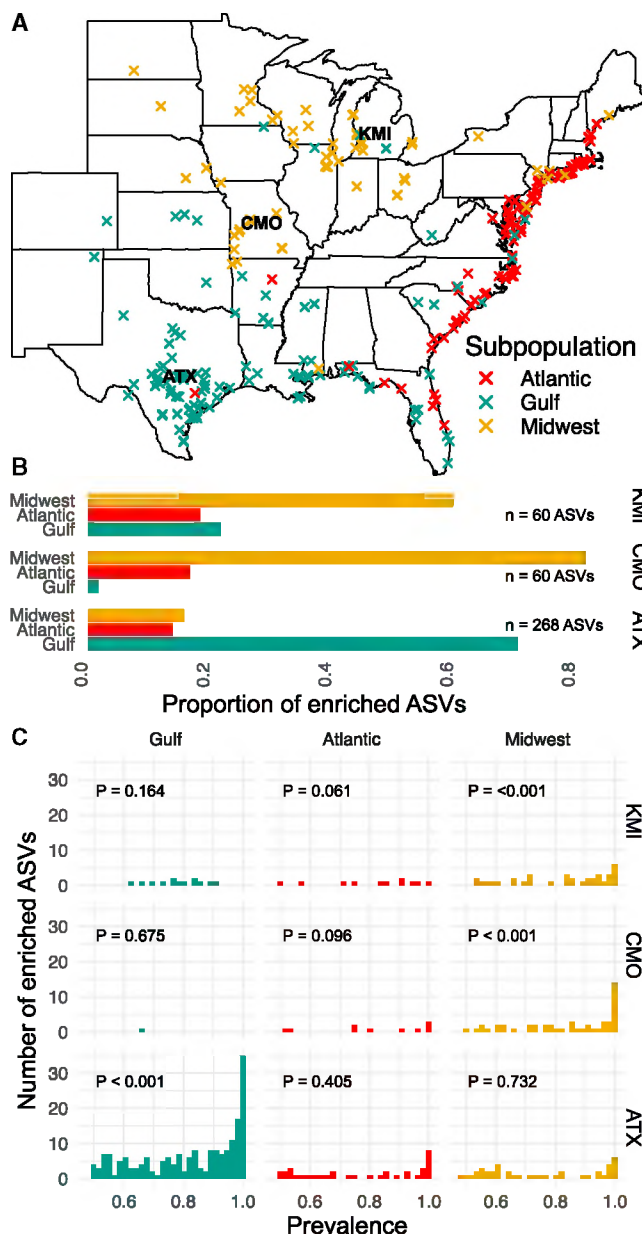


Figure 2. Plants show evidence of affinity to local bacterial strains
(A) Map depicting locations where individuals within the population were collected. Colors represent their subpopulation memberships. Field sites are depicted with their three-letter abbreviations. ATX, Austin, TX; CMO, Columbia, MO; KMI, KBS, MI.
(B) Proportion of ASVs showing specific enrichments in one subpopulation compared with the other two by site.
(C) Histograms of prevalence showing specific enrichments by subpopulation and site. p values represent the significance of the mean prevalence being greater than that of the background distribution. This was calculated by randomly drawing the number of enriched ASVs from the background distribution and asking how often we saw a mean prevalence greater than that of the focal set. See also [Figure S2](#) and [Data S1B](#).

ASV within the orders Sphingomonadales, subgroup 6 (Acidobacteria), Gammaproteobacteria Incertae Sedis displayed significant non-zero V_A or $V_{G \times E}$. In general, across microbial taxa,

$V_{G \times E}$ was greater than V_A (Figure 3D). The prominence of $G \times E$ suggested that the magnitude of V_A differed between locations. To better understand the contribution of V_A within each site, we fit an unstructured model to ASVs that allowed for site-specific V_A and as many unique covariances as site combinations. We applied these models to ASVs with prevalence >80% in at least two field sites (Figure 3E), finding similar trends to the compound symmetry model (Figure S3A). When analyzing the study-wide core microbiota, we found 95 instances of significant site-specific V_A spread across 64 unique ASVs (Data S1C). CMO had the most ASVs displaying significant V_A ($n = 38$) while KMI had the least ($n = 24$). We also tested whether there was a genetic association between the abundance of an ASV across multiple sites by focusing on the genetic covariance of root-associated microbial traits across sites. Genetic covariances were mainly positive (Figure S3B), and site comparison had a significant effect on covariance strength ($p = 0.005$, ANOVA). CMO/KMI covariances were significantly greater than those from ATX/KMI (adjusted $p = 0.006$, Tukey's post hoc test), but not ATX/CMO ($p > 0.05$, Tukey's post hoc test). We tested for ASVs that showed significant genetic covariance between sites and found 78 total significant estimates spread across 59 unique ASVs. Like the aggregate genetic covariance distributions, we found the most cases of significant genetic covariance between CMO/KMI, while CMO/ATX and KMO/ATX had equal instances of significant estimates (Figure 3C). Together, these results indicate that the host genetics play a significant role in modulating an extensive phylogenetic diversity of root-associated microbiota.

GWAS reveals microbiota assembly is a complex trait with extensive pleiotropy

We next asked whether host genomic regions responsible for heritable variation in associated bacteria could be localized using a GWAS framework. We first performed GWAS on community composition using the first three principal coordinates for each site (Figure S4). Significant associations between SNPs and community composition were detected for each site. To better understand how host allelic variation influences individual microbes, we extended our analysis to perform GWAS on each ASV $\times$ site combination, resulting in 1,019 independent analyses. We found 1,153 SNPs associated with 459 ASV $\times$ site combinations. Most ASVs with significant SNP associations were from the ATX site (253 ASVs), while CMO and KMI had similar numbers of ASVs with associated SNPs (101 and 105 ASVs, respectively). Taxa with associated SNPs were diverse, but no bacterial orders were overrepresented (Figures 4A–4C). Most ASVs with associated SNPs were specific to field sites; however, of the 179 ASVs that were tested in multiple sites, 50 showed associations across multiple field sites, with 9 showing associations across all three sites (Figure S5D). In line with our heritability analysis, bacteria within Sphingomonadales featured prominently among ASVs, with GWAS hits across multiple sites: 7 of the 10 ASVs within this family showed hits across 2 or more sites, and 2 *Sphingobium* ASVs had at least one significantly associated SNP at all three sites (Figure S5D).

We next asked whether any host genomic loci affected multiple microbial taxa (i.e., had statistically pleiotropic effects on microbiota, from here on referred to as pleiotropic loci) by compiling the 0.5% tail of 25 kb genomic bins into a quantitative

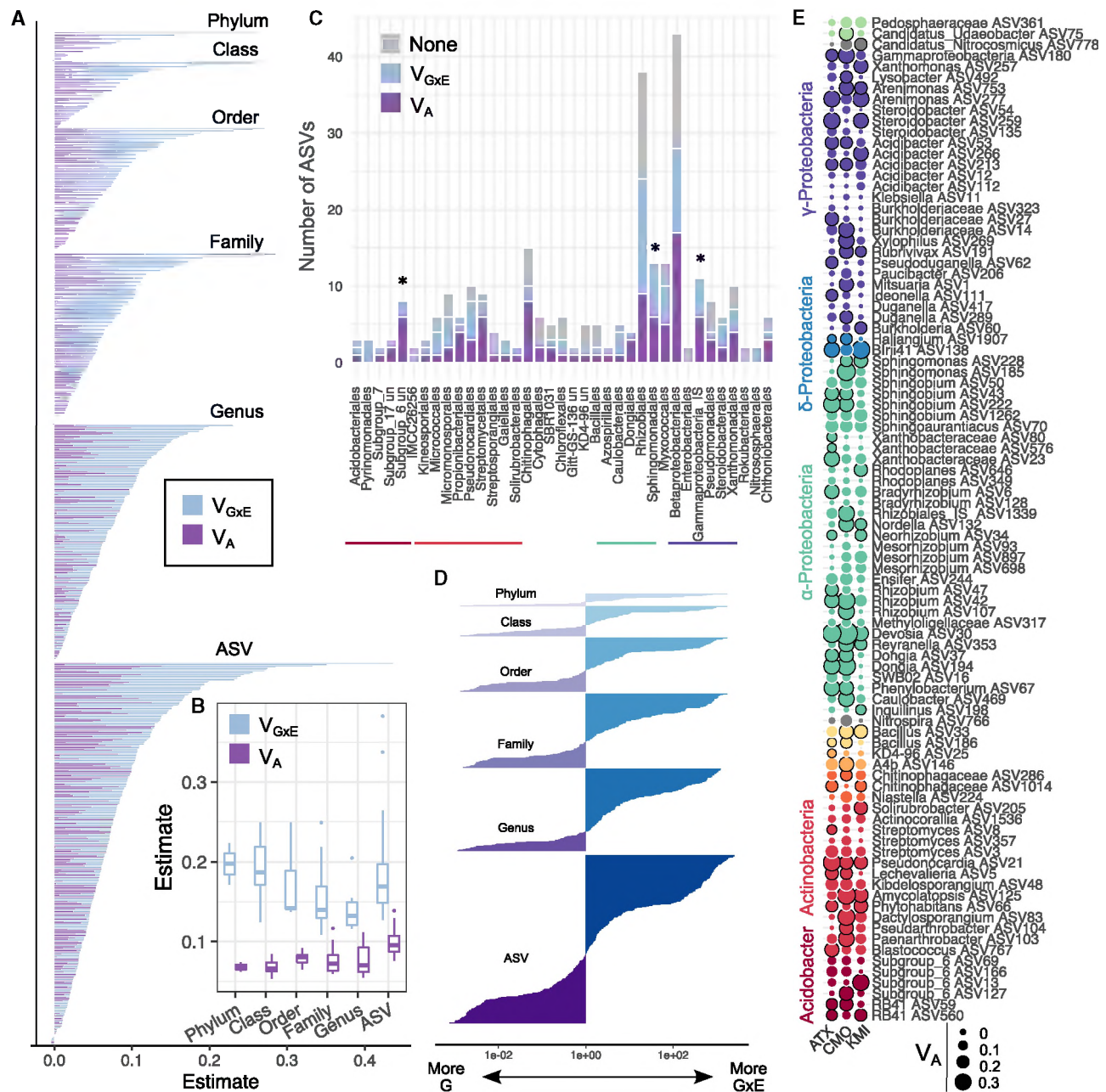


Figure 3. Switchgrass root microbiota show widespread heritability, which is influenced by field site differences

(A) Variance components for aggregated abundances of different taxonomic levels and for ASVs. V_A is the additive genetic variance while $V_{G \times E}$ is the variance attributable to genotype by environment interactions.
(B) The relationship between genetic variance components and microbial taxonomic rank.
(C) The number of ASVs showing either significant GxE, V_A , or no association to host genotype.
(D) Comparison of the magnitude of V_A vs. GxE is presented as the log fold change in the ratio of V_A to GxE for measured units within each taxonomic level.
(E) V_A estimates for the study-wide core microbiota. The size of the circles indicates the magnitude of estimated V_A and dark perimeters of the circles indicate a significant association (FDR < 0.1).
See also Figure S3 and Data S1C.

trait locus (QTL) $\times$ ASV matrix for each site (Figure 4; see STAR Methods). We first investigated the most commonly observed 25 kb genomic bins for each site by selecting the top 5 loci associated with the most ASVs within each site (ATX = 38–45 ASVs;

CMO = 18–23 ASVs; KMI = 19–25 ASVs, Data S1D). Two pleiotropic loci overlapped with loci detected from our initial GWAS on community metrics (Figure S4; CMO:Chr01N and ATX:Chr02K), indicating that while some pleiotropic loci account

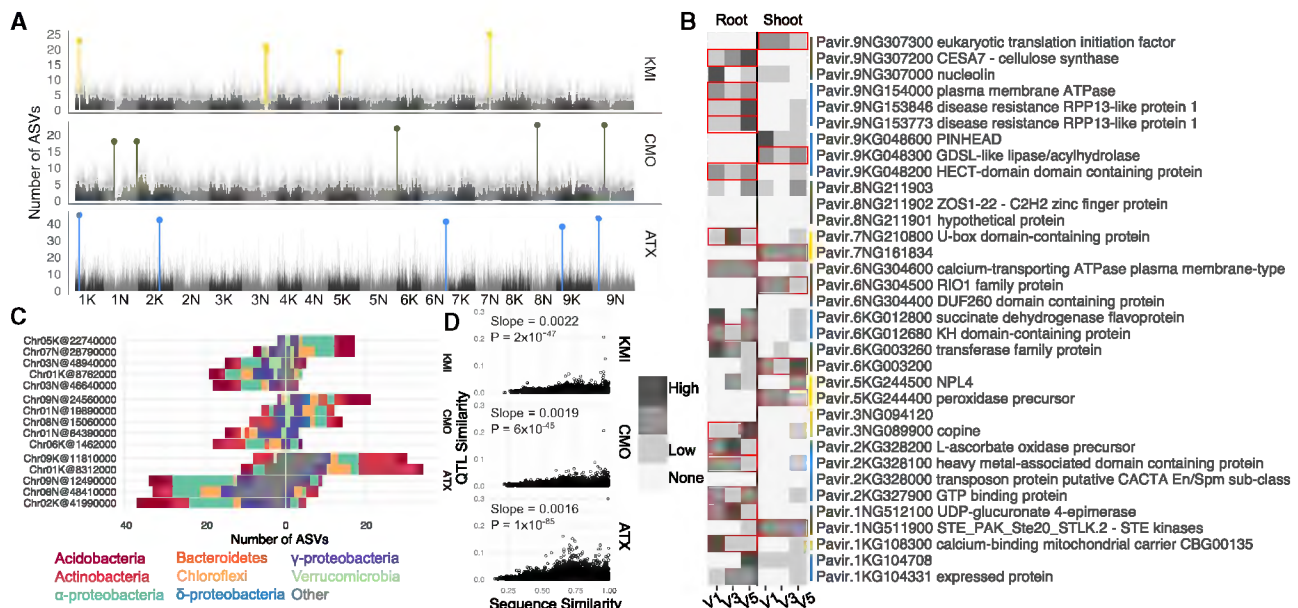


Figure 4. Pleiotropic loci influencing root microbiota

(A) Number of ASVs detected in the 0.5% tails of the ASV x site GWAS p value distributions. The top 5 most frequently observed genomic bins for each site are highlighted in site-specific colors.

(B) Candidate genes underlying the pleiotropic loci and their expression pattern in switchgrass roots and shoots. V1–V3 represent phenological stages of the plant, and red boxes around expression values represent genes differentially expressed between roots and shoots (FDR < 0.05).

(C) Taxonomic breakdown of ASVs affected by putatively pleiotropic loci.

(D) Comparison of QTL similarity (1 - Jaccard dissimilarity) and ASV sequence similarity.

See also [Figures S4–S6](#) and [Data S1](#).

for larger trends in community composition, most identify variation not seen along the first three axes of community composition.

To better characterize the candidate genes underlying these loci, we next compiled expression patterns for genes within these intervals ([Figure 4B](#)). Most loci contained genes displaying higher expression patterns in switchgrass roots than shoots, implicating promising candidate genes affecting multiple microbiota members. These included several proteins involved in calcium signaling, immunity, and secondary cell-wall biosynthesis. The microbes associated with pleiotropic loci were taxonomically diverse, with multiple bacterial phyla affected by each locus. In general, the additive effects of the QTL were largely consistent in sign across the different ASVs. This observation was also reflected in the taxa being affected by the loci: several loci show patterns where the relative abundances of Actinobacteria, Chloroflexi, or Alphaproteobacteria ASVs had consistent effect signs ([Figure 4C](#)). This observation led us to the hypothesis that there may be an association between the QTL landscape and phylogenetic relationship for pairs of microbes. We found a positive and significant association between the sequence similarity of ASVs and their associated QTL ([Figure 4D](#)). This association differed weakly but significantly between sites, with ATX showing a weaker correlation than CMO or KMI ($p = 0.06$ and 0.0015 , respectively). Each site had a closely related ASV pair, which stood out in terms of shared QTLs. These included two *Sphingobium* ASVs in ATX, *Bacillus* in CMO, and *Acidibacter* in KMI. Together, these results indicate that host genomic

variation can have pleiotropic effects on microbiota and that the abundances of related microbes are more likely to be affected by the same host loci.

The discovered pleiotropic loci included several promising candidate genes, but to have a more robust understanding of the functional categories influencing switchgrass-root-associated microbiota, we performed GO enrichments for annotated genes underlying the ASV x QTL matrix ([Figure S6](#)). We found that 789 of the ASV x site combinations displayed at least one significant GO enrichment. The most commonly observed GO term enrichments showed overlapping as well as contrasting patterns between sites ([Figure S6](#)). For example, the terms “response to biotic stimulus,” “response to auxin,” “negative regulation of growth,” and “sucrose biosynthesis” were observed in multiple ASVs across every site, while “defense response,” “prophenate biosynthetic process,” and “carbohydrate binding” showed more site-specific patterns. These results indicate that variation in host molecular pathways can influence the abundance of microbiota members and that some pathways are putatively dependent on environmental conditions.

To better understand the contribution of loci independent of field site, we subsetting our scans to ASVs in the study-wide core microbiota, joining p values generated during GWAS for a single ASV across each field site using Fisher’s method, a practice commonly used in meta-analyses to identify statistical tests with repeatable signals across multiple trials. A total of 239 SNPs passed a p value threshold of 5×10^{-8} , revealing that 44 out of

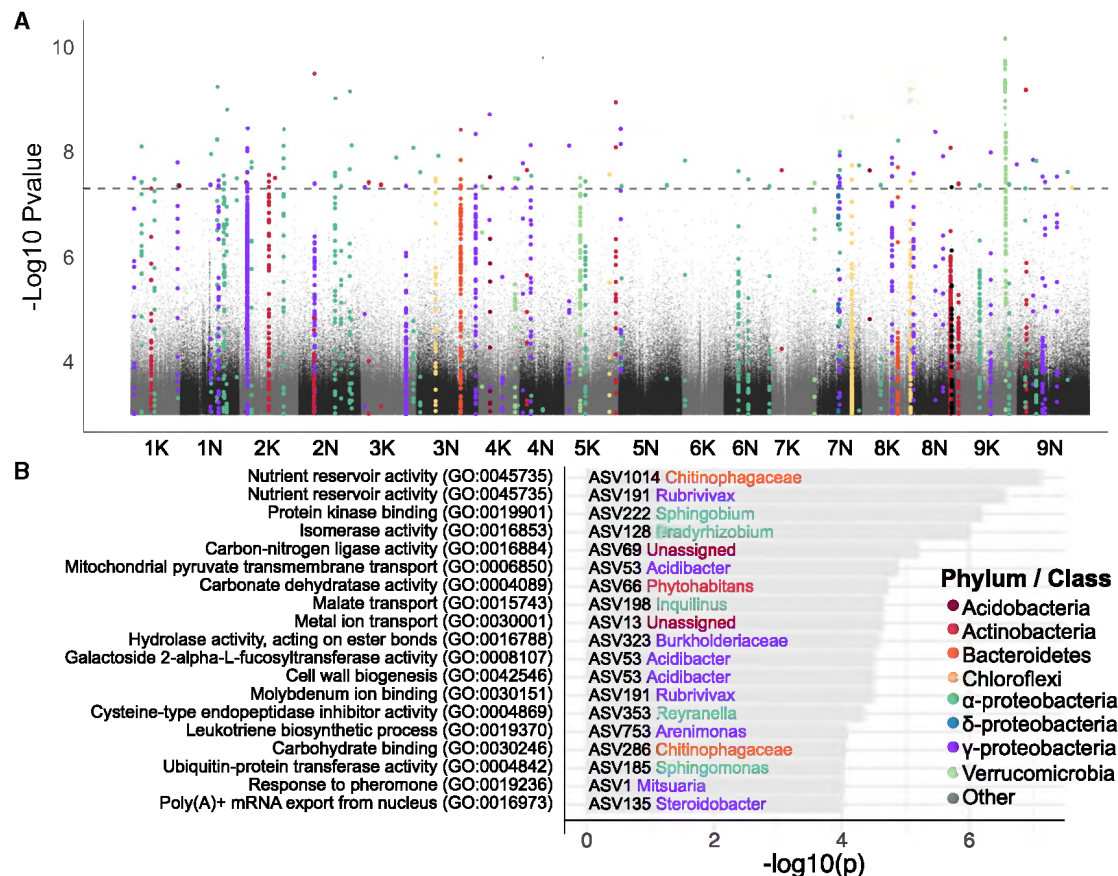


Figure 5. GWAS reveals loci associated with study-wide core switchgrass root microbiota

(A) Manhattan plot showing the association between SNPs and abundances of study-wide core ASVs. p values are derived from combining p values using Fisher's method. Peaks are colored by the phylum/class of the ASV.

(B) The most strongly enriched gene ontology (GO) terms within the core ASV GWAS tails.

See also [Data S1](#).

92 study-wide core ASVs had a significant association (Figures 5A and S5D; [Data S1E](#)). More than half of the ASVs with significant associations (23/44) showed significant GWAS hits across multiple sites (Figures 5A and S5D). Some ASVs with combined p values passing this genome-wide threshold did not display any significant associations in the ASV × site GWAS analyses. For example, ASV6, a highly abundant *Bradyrhizobium* strain, displayed two significant peaks when p values were combined that were not present during the initial site by ASV GWAS (Figure S5D). We calculated the amount of variance explained by significant loci using a multi-QTL model. Our results indicate that SNPs explained 1.2%–24% of the observed phenotypic variance, with a mean of 5.5% ([Data S1F](#)). These results indicate that leveraging multi-site GWAS by combining p values can identify loci impacting the study-wide core microbiota and that the variance explained through SNPs on phenotypic variation is in line with the results reported in an earlier study in *Arabidopsis*.³⁵

We explored the functional enrichments of combined p value GWAS scans from the study-wide core microbiota (Figure 5B). We identified 76 distinct GO terms enriched across 48 ASVs, some of which have *a priori* implications in microbiome

assembly. For example, malate transport and cell-wall biogenesis were among the most frequently enriched terms. Malate is a prominent root exudate involved in shaping rhizospheric microbiome composition,³⁶ and cell walls form physical barriers as well as energy sources for microbes.³⁷ This analysis revealed that although observations of loci associated with the abundance of various microbes is environmentally dependent, some loci can be implicated across multiple environments and the processes by which the host plant modulates core microbiota are diverse.

Pattern-triggered immunity responses genetically co-vary with root-associated microbiome composition

Plants surveil their biotic environment through perception of microbial associated molecular patterns (MAMPs), eliciting the activation of the pattern-triggered immunity (PTI) pathway. We hypothesized that loci responsible for the observed variation in PTI may overlap with host genetic variation controlling microbial abundance. To test this hypothesis, we measure reactive oxygen species (ROS) burst in response to the elicitor flg22 in the switchgrass diversity panel. Flg22 elicited a range of ROS burst profiles in the population, while mock treated samples did not display the

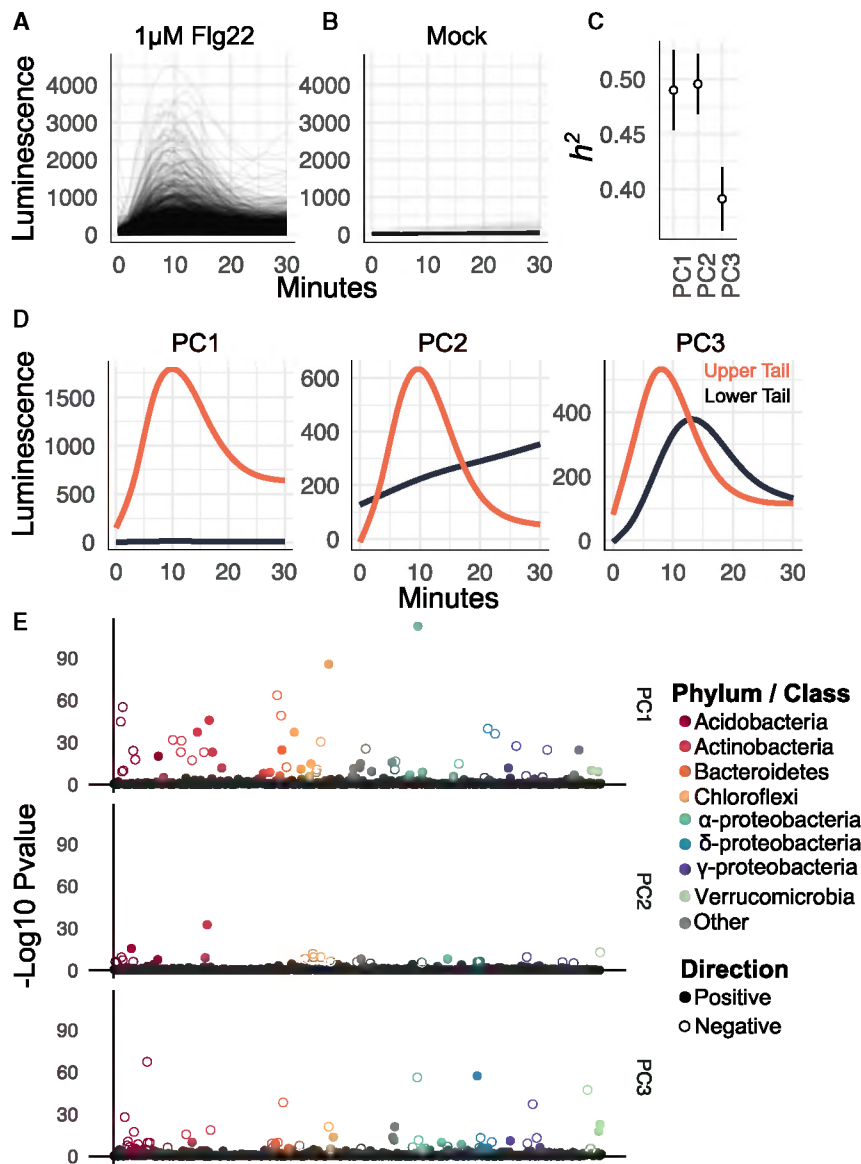


Figure 6. ASV abundances co-vary with pattern-triggered immune responses

(A) Response curves for the switchgrass population planted at the ATX site for treatment with 1 μ M Flg22. (B) Response curves for mock inoculated plants. (C) Narrow-sense heritability estimates for the three PC axes of PTI response variation. Bars represent standard error estimates. (D) The 5% and 95% percent tails of the first three PC axes of PTI response variation. (E) Microbial Manhattan plot displaying the p values for the covariances between ASV relative abundance and the PC axes of PTI variation. Colored circles represent ASVs passing a Bonferroni threshold of 0.05.

ASVs, while PC1 had a similar amount of positive and negative covariances (Figure 6D). These results indicate that bacterial microbiota show positive and negative genetic correlations with PTI responsiveness and that associations between these traits are not phylogenetically limited.

DISCUSSION

Here, we have used natural switchgrass accessions growing in common gardens spanning its native range to evaluate the contribution of environment and host genotype on root-associated bacterial assemblages. A similar study using a separate switchgrass population at a single site also found a significant effect of host genotype on rhizosphere microbiome assembly.³⁰ While our studies analyzed the microbiomes of different root compartments, there was notable overlap in results. For example, both studies identified microbes within Sphingomonadaceae as heritable members of the switchgrass microbiome.

A key finding of our study was that relative abundances of bacteria were strongly influenced by the interaction of host genetic variation and field site (Figures 2 and 3). Further, we found that there were affinities between genotypes growing in their home environments and the local microbiota (Figure 2B). Interestingly, microbes with specific enrichments to local genotypes consistently had higher prevalence than expected (Figure 2C). A potential explanation is that genotypes grown in their subpopulation's range, as opposed to genotypes grown outside of their subpopulation's range, are more in sync with their native climates, photoperiods, and soil properties. This, in turn, may reduce host stress³⁸ and culminate in the acquisition of consistent microbiota. Alternatively, these results could be explained by a co-evolutionary framework, where the host and microbes impose reciprocal selection that leads to evolutionary change in both the host and the microbiome.²⁴ However, given the stochastic dispersal of soil microbes,³⁹ a more likely explanation is one-sided evolution, where the local microbe population

typical response curve of treated plants (Figure 6A). We converted the time series into principal components to better understand the different modes of variation displayed across treated samples. The tails of the PC axes were informative of the type of variation observed in the population (Figure 6B): PC1 best explained the magnitude of response; PC2 separated plants with acute vs. gradual responses; and PC3 showed a timing difference of peak ROS burst. All three axes showed significant narrow-sense heritability (h^2), ranging from 0.48 to 0.38 (Figure 6C). These results indicate that switchgrass genotypes significantly vary in their response to the PTI elicitor flg22.

We next calculated the genetic covariances for the PTI PC axes against the relative abundance of ATX-specific core microbes. We found significant genetic covariances across each PTI axis: in total 126/739 ASVs showed significant genetic covariances with PTI axes (Bonferroni $p < 0.05$, Figure 6D). PTI PC1 had the most associations and PC2 had the least. PTI PCs 2 and 3 predominantly had negative covariances with

imposes selection and evolution on the host population, rather than the host imposing selection on the microbial populations.^{24,40} The elevated prevalence of enriched microbes may equate to more chances for interaction and acts to exert stronger selection on hosts (Figure 2C). Another display of GxE was that ASVs rarely showed heritable variation across every site. While GxE for microbial community composition is often complex in these types of studies, the fundamental “disease triangle” framework from the plant pathology field⁴¹ is useful when considering host-microbe associations, regardless of pathogenesis. This theory dictates that, for disease to occur, a susceptible host genotype, virulent pathogen, and favorable environmental condition must co-exist. Each of the three points of the triangle can be explored further to explain GxE in root microbiota assemblages. We discuss these three points in the context of our study below.

First, environmental variation occurs in biotic and abiotic components, which are not mutually exclusive. Our results indicate that the environment greatly influences the composition of root microbiota (Figure 1A). The three field sites differ in their management and soil chemical and physical properties (DataS1A), factors likely contributing to soil microbiome variation.⁶ The CMO and KMI sites are converted prairie and forest, respectively, and cultivate crops either agriculturally or experimentally. The ATX field site is located on a campus with no known history of agricultural cultivation. Furthermore, climate patterns differ between the sites, CMO and KMI having more similar patterns.³² Differing conditions may promote growth of certain taxa, which may ultimately influence the abundance of other microbes.

The microbial component of the disease triangle states that a virulent form of the pathogen must be present to infect a host and initiate disease. Implicit to this point is that genetic variation exists for microbes in addition to hosts. We could not examine genetic variation of individual ASVs in our study because detection and abundance of taxa was based on a single gene, which is insufficient to explore bacterial strain level variation. ASVs in a site are under selective pressure by the local environment.⁴² Therefore, an ASV detected at one site can have distinct polymorphisms with adaptive consequences compared with the same ASV at a different field site. Even within sites, ASVs can be composed of multiple microbial lineages,⁴³ some of which convey distinct phenotypes to the host.⁴⁴ Genetic polymorphisms within an ASV group may preclude the microbe from falling under the genetic influence of the host, explaining why we detect significant heritability for the same ASV in some sites but not others. Nevertheless, we identified ASVs where combined p values generated from site-specific GWAS helped to uncover loci consistently associated with their abundance. This was the case for half of the study-wide core microbiome, suggesting that the modulation of ASVs through shared mechanisms across field sites is relatively common, yet may not have effects passing a threshold in single ASV × site GWAS. A potential method to study GxE with host-associated microbiomes is through the construction of synthetic communities, which offers an ecologically relevant, yet controlled system for plants and microbes to interact while experiencing an experimental environment change.⁴⁵

Finally, the third point of the disease triangle stipulates that a host plant must be susceptible to infection for pathogenesis to

occur. In our case, this equates to host accessions being compatible for colonization by the local ASV. Susceptibility or compatibility is likely dependent upon both biotic and abiotic environmental conditions. That is, habitat variation and microbial community variation between sites may activate or repress the expression of the allelic variants responsible for regulation of microbial colonization. For example, increased temperature attenuates effector-triggered immunity in *Arabidopsis*, increasing susceptibility to *Pseudomonas syringae*.⁴⁶ Xin et al. demonstrate that elevated humidity can greatly influence the pathogenesis of *Pseudomonas syringae*, but in a host-genotype-dependent manner.¹³ In addition, given that the microbiomes vary substantially between sites, the biotic component of the environment may contribute to expression differences between allelic variants, thus leading to differential enrichment of metabolic, immunity, and developmental pathways.^{25,47} One fascinating possibility is that microbes that subvert plant immunity may ultimately serve as keystone taxa^{48–50} by dampening the immune response, allowing other microbiota to side-step the host immune system. Given that the biotic environment largely varies between sites, contrasting keystone taxa may exert alternative effects on different genotypes.

Genetic architecture of host-microbiome interactions in roots

We identified regions of the host genome associated with the abundance of study-wide core taxa. In addition, our results indicate that associated SNPs passing a genome-wide threshold are rarely shared across multiple ASVs, yet the tails of GWAS p value distributions contain commonly associated loci. This suggests that loci with the largest effects on any ASV's abundance are specific to that microbe, while loci with smaller effects are shared between ASVs. Together, these results indicate that microbiome assembly is a complex trait, given that the microbiome constitutes a consortium of interdependent bacteria, that many significant loci with small effect sizes were identified associated with these microbes' abundances (Figure 5A; Data S1F), and that many GO term enrichments were uncovered associated with these loci (Figures 5B and S6). The latter of these two observations were also reported by Sutherland et al.³⁰ These results suggest that many genes and processes contribute relatively small effects to influence the relative abundance for various ASVs.

A difficulty in presenting these data is their complexity and the plethora of uncovered candidate genes putatively involved in microbiota assembly. We therefore focused on loci impacting the most members of the microbiome (i.e., pleiotropic loci; Figure 4). Several compelling candidate genes were identified among the commonly associated loci, which showed enriched expression in roots. Among these were a cellulose synthase subunit, whose ortholog in *Arabidopsis* is involved in secondary cell-wall synthesis and has been reported to influence resistance to soil-borne bacterial pathogens in a defense-hormone-independent manner.⁵¹ We also identified two root-expressed candidate nucleotide-binding leucine rich repeat proteins (NLRs) showing associations to multiple ASVs. NLRs are important sensors involved in effector-triggered immunity and have been implicated in affecting both the sorghum and barley rhizosphere microbiota.^{26,52} Given the diversity of NLR genes within plant

species (switchgrass has >1,500 annotated NLR genes) and the presence/absence variation between individuals within species,⁵³ an open question is how the repertoire of NLR genes shapes root-associated microbiota.

An association between PTI and root microbiota composition

Several of our analyses implicated physical and immune defenses as modulators of microbiome composition. In our study, we investigated the role of plant genotype in explaining PTI variation using the elicitor flg22. Although flg22 is one of many known elicitors, it serves as a good proxy for PTI given that pattern recognition receptors share similar co-receptors, which funnel into similar pathways,⁵⁴ and downstream transcriptional responses show strong overlaps.⁵⁵ Much like a recent study in *Arabidopsis* that used seedling root growth inhibition as a proxy for PTI sensitivity, our results revealed strong heritable variation in PTI response within our population.⁵⁶ Further, our analysis revealed a link between the abundance of the ATX core microbiota and modes of PTI variation within our switchgrass diversity panel. Particularly strong associations, both negative and positive, were observed between the first axis of PTI variation (ROS burst magnitude) and a phylogenetically broad set of root-associated microbes (Figure 6D). PTI canonically inhibits the entry of perceived pathogens,⁵⁷ but our results suggest that it may also gate or limit the proliferation of commensal bacteria and their interactors, at least for ASVs with negative genetic covariances. This result is in line with previous studies showing that the attenuation of PTI can lead to altered microbiota composition and even dysbiosis.⁵⁸ Similarly, *Arabidopsis* plants with altered defense hormone production host atypical root microbiota, indicating that immune signaling is an important modulator of microbiota assembly.⁵⁹ On the other hand, we found ASVs with strong positive genetic covariance with PTI. These ASVs may (1) stimulate PTI sensitivity, such as in the case of induced systemic resistance; (2) escape the effects of PTI; or (3) benefit from the exclusion of PTI-sensitive microbes. Deciphering the role and mechanisms of the host immune system in regulating microbiota assembly processes and how the assembly of microbiota in turn modulates the host immune system is an active area of investigation, with implications for the design of plant probiotics.⁶⁰

Leveraging the microbiota via manipulation of host genetics to favor desirable outcomes on plant fitness or yield is a goal that is currently unrealized. By characterizing which microbes are responsive to plant genotype and the potential loci involved in host-microbiome interactions, the insights from this study may be of use for configuring associations between plants and microbes in the field.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
 - Lead contact
 - Materials availability
 - Data and code availability

● EXPERIMENTAL MODEL AND SUBJECT DETAILS

- Plants

● METHOD DETAILS

- Root sample collection and processing
- DNA extraction
- Library preparation and sequencing
- Sequence processing and ASV calling
- Beta diversity measurements
- Modeling site and subpopulation effects on ASVs
- Core microbiome considerations
- Which taxonomic level is appropriate for calculating heritability of bacteria
- Genetic variance component analyses
- Microbial genome-wide associations
- Detection of pleiotropic loci affecting multiple microbes
- Gene ontology enrichments
- Gene expression analysis
- Pattern-triggered immunity assays
- Genetic covariances of PTI axes and bacterial abundances

● QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cub.2023.03.078>.

ACKNOWLEDGMENTS

This research was supported by the Office of Science (BER), US Department of Energy, grant no. DE-SC0014156 and DE-SC0021126. The work (proposal: <https://doi.org/10.46936/10.25585/60000507>) conducted by the US Department of Energy Joint Genome Institute (<https://ror.org/04xm1d337>), a DOE Office of Science User Facility, is supported by the Office of Science of The US Department of Energy operated under contract no. DE-AC02-05CH11231. This work was supported in part by the Great Lakes Bioenergy Research Center, US Department of Energy, Office of Science, Office of Biological and Environmental Research under award number DE-SC0018409. Support for this research was provided by the National Science Foundation Long-term Ecological Research Program (DEB 1832042) at the Kellogg Biological Station and by Michigan State University AgBioResearch. J.A.E. acknowledges the support of the USDA National Institute of Food and Agriculture postdoctoral fellowship (grant no. 2019-67012-2971/project accession no. 1019437). In addition, we would like to acknowledge Allison Hutt, Nick Ryan, and Lisa Vormwald for their help in collecting samples.

AUTHOR CONTRIBUTIONS

J.A.E., T.E.J., F.B.F., and D.B.L. conceptualized the project; J.A.E., U.B.S., J.B., J.Y., and T.u.N. performed experiments; J.A.E., A.M., and T.E.J. analyzed data; J.G., T.G.d.R., and C.D. performed sequencing; J.S. and L.A.P. facilitated sequencing of samples; F.B.F., D.B.L., and T.E.J. oversaw experimentation in the common garden sites; J.A.E. and T.E.J. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: August 23, 2022

Revised: February 3, 2023

Accepted: March 27, 2023

Published: April 19, 2023

REFERENCES

- Hiltner, L. (1903). Studien über die Bakterienflora des Ackerbodens: mit Besonderer Berücksichtigung Ihres Verhaltens nach Einer Behandlung mit Schwefelkohlenstoff und nach Brache (Parey).
- Beattie, G.A. (2015). Microbiomes: curating communities from plants. *Nature* 528, 340–341.
- Carrión, V.J., Perez-Jaramillo, J., Cordovez, V., Tracanna, V., de Hollander, M., Ruiz-Buck, D., Mendes, L.W., van Ijcken, W.F.J., Gomez-Exposito, R., Elsayed, S.S., et al. (2019). Pathogen-induced activation of disease-suppressive functions in the endophytic root microbiome. *Science* 366, 606–612.
- Jiao, S., Xu, Y., Zhang, J., Hao, X., and Lu, Y. (2019). Core microbiota in agricultural soils and their potential associations with nutrient cycling. *mSystems* 4, e00313–e00318. <https://doi.org/10.1128/mSystems.00313-18>.
- Santos-Medellín, C., Liechty, Z., Edwards, J., Nguyen, B., Huang, B., Weimer, B.C., and Sundaresan, V. (2021). Prolonged drought imparts lasting compositional changes to the rice root microbiome. *Nat. Plants* 7, 1065–1077.
- Edwards, J., Santos-Medellín, C., Nguyen, B., Kilmer, J., Liechty, Z., Veliz, E., Ni, J., Phillips, G., and Sundaresan, V. (2019). Soil domestication by rice cultivation results in plant-soil feedback through shifts in soil microbiota. *Genome Biol.* 20, 221.
- Edwards, J.A., Santos-Medellín, C.M., Liechty, Z.S., Nguyen, B., Lurie, E., Eason, S., Phillips, G., and Sundaresan, V. (2018). Compositional shifts in root-associated bacterial and archaeal microbiota track the plant life cycle in field-grown rice. *PLoS Biol.* 16, e2003862.
- Zhang, J., Zhang, N., Liu, Y.X., Zhang, X., Hu, B., Qin, Y., Xu, H., Wang, H., Guo, X., Qian, J., et al. (2018). Root microbiota shift in rice correlates with resident time in the field and developmental stage. *Sci. China Life Sci.* 61, 613–621.
- Xu, L., Naylor, D., Dong, Z., Simmons, T., Pierroz, G., Hixson, K.K., Kim, Y.M., Zink, E.M., Engbrecht, K.M., Wang, Y., et al. (2018). Drought delays development of the sorghum root microbiome and enriches for monoderm bacteria. *Proc. Natl. Acad. Sci. USA* 115, E4284–E4293.
- Bulgarelli, D., Schlaeppi, K., Spaepen, S., Ver Loren van Themaat, E., and Schulze-Lefert, P. (2013). Structure and functions of the bacterial microbiota of plants. *Annu. Rev. Plant Biol.* 64, 807–838.
- Vogel, C.M., Potthoff, D.B., Schäfer, M., Barandun, N., and Vorholt, J.A. (2021). Protective role of the *Arabidopsis* leaf microbiota against a bacterial pathogen. *Nat. Microbiol.* 6, 1537–1548.
- Hu, L., Robert, C.A.M., Cadot, S., Zhang, X., Ye, M., Li, B., Manzo, D., Chervet, N., Steinger, T., van der Heijden, M.G.A., et al. (2018). Root exudate metabolites drive plant-soil feedbacks on growth and defense by shaping the rhizosphere microbiota. *Nat. Commun.* 9, 2738.
- Xin, X.F., Nomura, K., Aung, K., Velásquez, A.C., Yao, J., Boutrot, F., Chang, J.H., Zipfel, C., and He, S.Y. (2016). Bacteria establish an aqueous living space in plants crucial for virulence. *Nature* 539, 524–529.
- Voges, M.J.E.E.E., Bai, Y., Schulze-Lefert, P., and Sattely, E.S. (2019). Plant-derived coumarins shape the composition of an *Arabidopsis* synthetic root microbiome. *Proc. Natl. Acad. Sci. USA* 116, 12558–12565.
- Mueller, U.G., Juenger, T.E., Kardish, M.R., Carlson, A.L., Burns, K.M., Edwards, J.A., Smith, C.C., Fang, C.C., and Des Marais, D.L. (2021). Artificial selection on microbiomes to breed microbiomes that confer salt tolerance to plants. *mSystems* 6, e0112521.
- Petipas, R.H., Geber, M.A., and Lau, J.A. (2021). Microbe-mediated adaptation in plants. *Ecol. Lett.* 24, 1302–1317.
- Walters, W.A., Jin, Z., Youngblut, N., Wallace, J.G., Sutter, J., Zhang, W., González-Peña, A., Peiffer, J., Koren, O., Shi, Q., et al. (2018). Large-scale replicated field study of maize rhizosphere identifies heritable microbes. *Proc. Natl. Acad. Sci. USA* 115, 7368–7373.
- Peiffer, J.A., Spor, A., Koren, O., Jin, Z., Tringe, S.G., Dangl, J.L., Buckler, E.S., and Ley, R.E. (2013). Diversity and heritability of the maize rhizosphere microbiome under field conditions. *Proc. Natl. Acad. Sci. USA* 110, 6548–6553.
- Wagner, M.R., Lundberg, D.S., del Rio, T.G., Tringe, S.G., Dangl, J.L., and Mitchell-Olds, T. (2016). Host genotype and age shape the leaf and root microbiomes of a wild perennial plant. *Nat. Commun.* 7, 1–15.
- Edwards, J., Johnson, C., Santos-Medellín, C., Lurie, E., Podishetty, N.K., Bhatnagar, S., Eisen, J.A., and Sundaresan, V. (2015). Structure, variation, and assembly of the root-associated microbiomes of rice. *Proc. Natl. Acad. Sci. USA* 112, E911–E920.
- Lundberg, D.S., Lebeis, S.L., Paredes, S.H., Yourstone, S., Gehring, J., Malfatti, S., Tremblay, J., Engelbrektson, A., Kunin, V., del Rio, T.G., et al. (2012). Defining the core *Arabidopsis thaliana* root microbiome. *Nature* 488, 86–90.
- Wagner, M.R., Roberts, J.H., Balint-Kurti, P., and Holland, J.B. (2020). Heterosis of leaf and rhizosphere microbiomes in field-grown maize. *New Phytol.* 228, 1055–1069.
- Wagner, M.R. (2021). Prioritizing host phenotype to understand microbiome heritability in plants. *New Phytol.* 232, 502–509. <https://doi.org/10.1111/nph.17622>.
- Koskella, B., and Bergelson, J. (2020). The study of host-microbiome (co) evolution across levels of selection. *Philos. Trans. R. Soc. Lond. B* 375, 20190604.
- Beilsmith, K., Thoen, M.P.M., Brachi, B., Gloss, A.D., Khan, M.H., and Bergelson, J. (2019). Genome-wide association studies on the phyllosphere microbiome: embracing complexity in host-microbe interactions. *Plant J.* 97, 164–181.
- Deng, S., Caddell, D.F., Xu, G., Dahlen, L., Washington, L., Yang, J., and Coleman-Derr, D. (2021). Genome wide association study reveals plant loci controlling heritability of the rhizosphere microbiome. *ISME J.* 15, 3181–3194. <https://doi.org/10.1038/s41396-021-00993-z>.
- Bergelson, J., Mittelstrass, J., and Horton, M.W. (2019). Characterizing both bacteria and fungi improves understanding of the *Arabidopsis* root microbiome. *Sci. Rep.* 9, 24.
- Singer, E., Bonnette, J., Kenaley, S.C., Woyke, T., and Juenger, T.E. (2019). Plant compartment and genetic variation drive microbiome composition in switchgrass roots. *Environ. Microbiol. Rep.* 11, 185–195.
- Ulbrich, T.C., Friesen, M.L., Roley, S.S., Tiemann, L.K., and Evans, S.E. (2021). Intraspecific variability in root traits and edaphic conditions influence soil microbiomes across 12 switchgrass cultivars. *Phytobiomes J.* 5, 108–120.
- Sutherland, J., Bell, T., Trexler, R.V., Carlson, J.E., and Lasky, J.R. (2022). Host genomic influence on bacterial composition in the switchgrass rhizosphere. *Mol. Ecol.* 31, 3934–3950.
- Zhang, L., MacQueen, A., Bonnette, J., Fritsch, F.B., Lowry, D.B., and Juenger, T.E. (2021). QTL × environment interactions underlie ionome divergence in switchgrass. *G3 (Bethesda)* 11, jkab144.
- Lowry, D.B., Lovell, J.T., Zhang, L., Bonnette, J., Fay, P.A., Mitchell, R.B., Lloyd-Reilly, J., Boe, A.R., Wu, Y., Rouquette, F.M., et al. (2019). QTL × environment interactions underlie adaptive divergence in switchgrass across a large latitudinal gradient. *Proc. Natl. Acad. Sci. USA* 116, 12933–12941.
- Lovell, J.T., MacQueen, A.H., Mamidi, S., Bonnette, J., Jenkins, J., Napier, J.D., Sreedasyam, A., Healey, A., Session, A., Shu, S., et al. (2021). Genomic mechanisms of climate adaptation in polyploid bioenergy switchgrass. *Nature* 590, 438–444. <https://doi.org/10.1038/s41586-020-03127-1>.
- MacQueen, A.H., Zhang, L., Bonnette, J., Boe, A.R., Fay, P.A., Fritsch, F.B., Lowry, D.B., Mitchell, R.B., Rouquette, F.M., Wu, Y., et al. (2021). Mapping of genotype-by-environment interactions in phenology identifies two cues for flowering in switchgrass (*Panicum virgatum*). Preprint at bioRxiv. <https://doi.org/10.1101/2021.08.19.456975>.
- Horton, M.W., Bodenhausen, N., Beilsmith, K., Meng, D., Muegge, B.D., Subramanian, S., Vetter, M.M., Vilhjálmsson, B.J., Nordborg, M.,

- Gordon, J.I., and Bergelson, J. (2014). Genome-wide association study of *Arabidopsis thaliana* leaf microbial community. *Nat. Commun.* **5**, 5320.
36. Kawasaki, A., Dennis, P.G., Forstner, C., Raghavendra, A.K.H., Mathesius, U., Richardson, A.E., Delhaize, E., Gilliam, M., Watt, M., and Ryan, P.R. (2021). Manipulating exudate composition from root apices shapes the microbiome throughout the root system. *Plant Physiol.* **187**, 2279–2295.
37. Veach, A.M., Yip, D., Engle, N.L., Yang, Z.K., Bible, A., Morrell-Falvey, J., Tschaplinski, T.J., Kalluri, U.C., and Schadt, C.W. (2018). Modification of plant cell wall chemistry impacts metabolome and microbiome composition in *Populus PdKOR1* RNAi plants. *Plant Soil* **429**, 349–361.
38. VanWallendael, A., Soltani, A., Emery, N.C., Peixoto, M.M., Olsen, J., and Lowry, D.B. (2019). A molecular view of plant local adaptation: incorporating stress-response networks. *Annu. Rev. Plant Biol.* **70**, 559–583.
39. Richter-Heitmann, T., Hofner, B., Krah, F.S., Sikorski, J., Wüst, P.K., Bunk, B., Huang, S., Regan, K.M., Berner, D., Boeddinghaus, R.S., et al. (2020). Stochastic dispersal rather than deterministic selection explains the spatio-temporal distribution of soil bacteria in a temperate grassland. *Front. Microbiol.* **11**, 1391.
40. Henry, L.P., Bruijning, M., Forsberg, S.K.G., and Ayroles, J.F. (2021). The microbiome extends host evolutionary potential. *Nat. Commun.* **12**, 5141.
41. Francl, L.J. (2001). The disease triangle: a plant pathological paradigm revisited. *The Plant Health Instructor* **10**. <https://doi.org/10.1094/PHI-T-2001-0517-01>.
42. Crits-Christoph, A., Olm, M.R., Diamond, S., Bouma-Gregson, K., and Banfield, J.F. (2020). Soil bacterial populations are shaped by recombination and gene-specific selection across a grassland meadow. *ISME J.* **14**, 1834–1846.
43. Lundberg, D.S., de Pedro Jové, R., Ayutthaya, P.P.N., Karasov, T.L., Shalev, O., Poersch, K., Ding, W., Bollmann-Giolai, A., Bezrukov, I., and Weigel, D. (2021). Contrasting patterns of microbial dominance in the *Arabidopsis thaliana* phyllosphere. *Proc. Natl. Acad. Sci. USA* **119**, e2211881119. <https://doi.org/10.1101/2021.04.06.438366>.
44. Karasov, T.L., Almario, J., Friedemann, C., Ding, W., Giolai, M., Heavens, D., Kersten, S., Lundberg, D.S., Neumann, M., Regalado, J., et al. (2018). *Arabidopsis thaliana* and *Pseudomonas* pathogens exhibit stable associations over evolutionary timescales. *Cell Host Microbe* **24**, 168.e4–179.e4.
45. Vorholt, J.A., Vogel, C., Carlström, C.I., and Müller, D.B. (2017). Establishing causality: opportunities of synthetic communities for plant microbiome research. *Cell Host Microbe* **22**, 142–155.
46. Menna, A., Nguyen, D., Guttman, D.S., and Desveaux, D. (2015). Elevated temperature differentially influences effector-triggered immunity outputs in *Arabidopsis*. *Front. Plant Sci.* **6**, 995.
47. Santos-Medellín, C., Edwards, J., Nguyen, B., and Sundaresan, V. (2022). Acquisition of a complex root microbiome reshapes the transcriptomes of rice plants. *New Phytol.* **235**, 2008–2021.
48. Teixeira, P.J.P.L., Colaianni, N.R., Law, T.F., Conway, J.M., Gilbert, S., Li, H., Salas-González, I., Panda, D., Del Risco, N.M., Finkel, O.M., et al. (2021). Specific modulation of the root immune system by a community of commensal bacteria. *Proc. Natl. Acad. Sci. USA* **118**, e2100678118. <https://doi.org/10.1073/pnas.2100678118>.
49. Niu, B., Paulson, J.N., Zheng, X., and Kolter, R. (2017). Simplified and representative bacterial community of maize roots. *Proc. Natl. Acad. Sci. USA* **114**, E2450–E2459.
50. Ma, K.W., Niu, Y., Jia, Y., Ordon, J., Copeland, C., Emonet, A., Geldner, N., Guan, R., Stölze, S.C., Nakagami, H., et al. (2021). Coordination of microbe-host homeostasis by crosstalk with plant innate immunity. *Nat. Plants* **7**, 814–825.
51. Hernández-Bianco, C., Feng, D.X., Hu, J., Sánchez-Vallet, A., Deslandes, L., Llorente, F., Berrocal-Lobo, M., Keller, H., Barlet, X., Sánchez-Rodríguez, C., et al. (2007). Impairment of cellulose synthases required for *Arabidopsis* secondary cell wall formation enhances disease resistance. *Plant Cell* **19**, 890–903.
52. Escudero-Martínez, C., Coulter, M., Alegria Terrazas, R., Foito, A., Kapadia, R., Pietrangeli, L., Mavrić, M., Sharma, R., Aprile, A., Morris, J., et al. (2022). Identifying plant genes shaping microbiota composition in the barley rhizosphere. *Nat. Commun.* **13**, 3443.
53. Van de Weyer, A.L., Monteiro, F., Furzer, O.J., Nishimura, M.T., Cevik, V., Witek, K., Jones, J.D.G., Dangel, J.L., Weigel, D., and Bemm, F. (2019). A species-wide inventory of NLR genes and alleles in *Arabidopsis thaliana*. *Cell* **178**, 1260–1272.e14.
54. Couto, D., and Zipfel, C. (2016). Regulation of pattern recognition receptor signalling in plants. *Nat. Rev. Immunol.* **16**, 537–552.
55. Björnson, M., Pimprikar, P., Nürnberger, T., and Zipfel, C. (2021). The transcriptional landscape of *Arabidopsis thaliana* pattern-triggered immunity. *Nat. Plants* **7**, 579–586.
56. Vetter, M., Karasov, T.L., and Bergelson, J. (2016). Differentiation between MAMP triggered defenses in *Arabidopsis thaliana*. *PLoS Genet.* **12**, e1006068.
57. Zipfel, C., Robatzek, S., Navarro, L., Oakeley, E.J., Jones, J.D.G., Felix, G., and Boer, T. (2004). Bacterial disease resistance in *Arabidopsis* through flagellin perception. *Nature* **428**, 764–767.
58. Chen, T., Nomura, K., Wang, X., Sohrabi, R., Xu, J., Yao, L., Paasch, B.C., Ma, L., Kremer, J., Cheng, Y., et al. (2020). A plant genetic network for preventing dysbiosis in the phyllosphere. *Nature* **580**, 653–657.
59. Lebeis, S.L., Paredes, S.H., Lundberg, D.S., Breakfield, N., Gehring, J., McDonald, M., Malfatti, S., Glavina del Rio, T., Jones, C.D., Tringe, S.G., et al. (2015). PLANT MICROBIOME. Salicylic acid modulates colonization of the root microbiome by specific bacterial taxa. *Science* **349**, 860–864.
60. Teixeira, P.J.P., Colaianni, N.R., Fitzpatrick, C.R., and Dangel, J.L. (2019). Beyond pathogens: microbiota interactions with the plant immune system. *Curr. Opin. Microbiol.* **49**, 7–17.
61. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12.
62. Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., and Holmes, S.P. (2016). DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583.
63. McCaw, Z.R., Lane, J.M., Saxena, R., Redline, S., and Lin, X. (2020). Operating characteristics of the rank-based inverse normal transformation for quantitative trait analysis in genome-wide association studies. *Biometrics* **76**, 1262–1272.
64. Searle, S.R., Speed, F.M., and Milliken, G.A. (1980). Population marginal means in the linear model: an alternative to least squares means. *Am. Stat.* **34**, 216–221.
65. Covarrubias-Pazaran, G. (2016). Genome-assisted prediction of quantitative traits using the R package sommer. *PLoS One* **11**, e0156744.
66. Oksanen, J., Kindt, R., Legendre, P., O'Hara, B., Stevens, M.H.H., and Oksanen, M.J.; MASS Suggests (2007). The vegan package: community ecology package **10**, 631–637.
67. Privé, F., Aschard, H., Ziyatdinov, A., and Blum, M.G.B. (2018). Efficient analysis of large-scale genome-wide data with two R packages: bigstatsr and bigsnpr. *Bioinformatics* **34**, 2781–2787.
68. Dewey, M. (2022). metap: meta-analysis of significance values. <https://cran.r-project.org/web/packages/metap/index.html>.
69. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842.
70. Federer, W.T., and Crossa, J. (2012). I4 screening experimental designs for quantitative trait loci, association mapping, genotype-by-environment interaction, and other investigations. *Front. Physiol.* **3**, 156.
71. Moehring, J., Williams, E.R., and Piepho, H.P. (2014). Efficiency of augmented p-rep designs in multi-environment trials. *Theor. Appl. Genet.* **127**, 1049–1060.
72. Bollmann-Giolai, A., Giolai, M., Heavens, D., Macaulay, I., Malone, J., and Clark, M.D. (2020). A low-cost pipeline for soil microbiome profiling. *MicrobiologyOpen* **9**, e1133.
73. Parada, A.E., Needham, D.M., and Fuhrman, J.A. (2016). Every base matters: assessing small subunit rRNA primers for marine microbiomes with

- p mock communities, time series and global field samples.
- Environ. Microbiol.*
- 18**
- , 1403–1414.
74. Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glöckner, F.O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596.
 75. Shade, A., and Stopnisek, N. (2019). Abundance-occupancy distributions to prioritize plant core microbiome membership. *Curr. Opin. Microbiol.* **49**, 50–58.
 76. Shade, A., and Handelsman, J. (2012). Beyond the Venn diagram: the hunt for a core microbiome. *Environ. Microbiol.* **14**, 4–12. <https://doi.org/10.1111/j.1462-2920.2011.02585.x>.
 77. Neu, A.T., Allen, E.E., and Roy, K. (2021). Defining and quantifying the core microbiome: challenges and prospects. *Proc. Natl. Acad. Sci. USA* **118**, <https://doi.org/10.1073/pnas.2104429118>.
 78. Sutherland, J., Bell, T., Trexler, R.V., Carlson, J.E., and Lasky, J.R. (2021). Host genomic influence on bacterial composition in the switchgrass rhizosphere. Preprint at bioRxiv. <https://doi.org/10.1101/2021.09.01.458593>.
 79. Melnyk, R.A., Hossain, S.S., and Haney, C.H. (2019). Convergent gain and loss of genomic islands drive lifestyle changes in plant-associated pseudomonas. *ISME J.* **13**, 1575–1588.
 80. Levy, A., Salas Gonzalez, I., Mittelviefhaus, M., Clingenpeel, S., Herrera Paredes, S., Miao, J., Wang, K., Devescovi, G., Stillman, K., Monteiro, F., et al. (2017). Genomic features of bacterial adaptation to plants. *Nat. Genet.* **50**, 138–150.
 81. Bai, Y., Müller, D.B., Srinivas, G., Garrido-Oter, R., Potthoff, E., Rott, M., Dombrowski, N., Münch, P.C., Spaepen, S., Remus-Emsermann, M., et al. (2015). Functional overlap of the Arabidopsis leaf and root microbiota. *Nature* **528**, 364–369.
 82. Kurilshikov, A., Medina-Gomez, C., Bacigalupe, R., Radjabzadeh, D., Wang, J., Demirkan, A., Le Roy, C.I., Raygoza Garay, J.A., Finnicum, C.T., Liu, X., et al. (2021). Large-scale association analyses identify host factors influencing human gut microbiome composition. *Nat. Genet.* **53**, 156–165.
 83. Lopera-Maya, E.A., Kurilshikov, A., van der Graaf, A., Hu, S., Andreu-Sánchez, S., Chen, L., Vila, A.V., Gacesa, R., Sinha, T., Collij, V., et al. (2022). Effect of host genetics on the gut microbiome in 7,738 participants of the Dutch Microbiome Project. *Nat. Genet.* **54**, 143–151.
 84. Brachi, B., Filiault, D., Whitehurst, H., Darne, P., Le Gars, P., Le Mentec, M., Morton, T.C., Kerdaffer, E., Rabanal, F., Anastasio, A., et al. (2022). Plant genetic effects on microbial hubs impact host fitness in repeated field trials. *Proc. Natl. Acad. Sci. USA* **119**, e2201285119.
 85. Grabowski, P.P., Evans, J., Daum, C., Deshpande, S., Barry, K.W., Kennedy, M., Ramstein, G., Kaeppler, S.M., Buell, C.R., Jiang, Y., and Casler, M.D. (2017). Genome-wide associations with flowering time in switchgrass using exome-capture sequencing data. *New Phytol.* **213**, 154–169.
 86. Sreedasyam, A., Plott, C., Hossain, M.S., Lovell, J.T., Grimwood, J., Jenkins, J.W., Daum, C., Barry, K., Carlson, J., Shu, S., et al. (2022). JGI Plant Gene Atlas: an updateable transcriptome resource to improve structural annotations and functional gene descriptions across the plant kingdom. Preprint at bioRxiv. <https://doi.org/10.1101/2022.09.30.510380>.
 87. Samira, R., Zhang, X., Kimball, J., Cui, Y., Stacey, G., and Balint-Kurti, P. (2019). Quantifying MAMP-induced production of reactive oxygen species in sorghum and maize. *Bio Protoc.* **9**, 9. <https://doi.org/10.21769/BioProtoc.3304>.
 88. Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
Fig22	EZBiolab	Cat# cp7201
L-012	Wako Chemicals	Cat# 120-04891
Horseradish Peroxidase, type VI-A	Sigma	Cat #P6782
Deposited data		
Raw sequencing reads for root associated microbiota	This paper	Bioproject: PRJNA822373
Raw sequencing reads for root associated microbiota	This paper	Bioproject: PRJNA919067
Code used to analyze data	This paper	GitHub: https://github.com/bulksoil/VirgatumMicrobiomeGWAS
Large data files	This paper	Database: https://figshare.com/projects/Data_for_first_switchgrass_microbiome_GWAS_study/162628
Experimental models: Organisms/strains		
Switchgrass Natural Diversity Panel	The Juenger Laboratory	Lovell et al. ³³
Oligonucleotides		
DNA Primers	Integrated DNA Technologies	https://github.com/bulksoil/VirgatumMicrobiomeGWAS
Software and algorithms		
Cutadapt	Martin ⁶¹	https://cutadapt.readthedocs.io/en/stable/index.html
Dada2	Callahan et al. ⁶²	https://www.bioconductor.org/packages/release/bioc/html/dada2.html
RNOmni	McCaw et al. ⁶³	https://cran.r-project.org/package=RNOmni
Emmeans	Searle et al. ⁶⁴	https://cran.r-project.org/package=emmeans
Sommer	Covarrubias-Pazaran ⁶⁵	https://cran.r-project.org/web/packages/sommer/index.html
Vegan	Oksanen et al. ⁶⁶	
Bigsnp	Privé et al. ⁶⁷	https://privefl.github.io/bigsnp/
switchgrassGWAS	Lovell et al. ³³	https://github.com/Alice-MacQueen/switchgrassGWAS
Metap	Dewey ⁶⁸	https://cran.r-project.org/web/packages/metap/index.html
BEDtools	Quinlan and Hall ⁶⁹	https://github.com/arq5x/bedtools2

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by lead contacts Joseph Edwards (j_edwards@utexas.edu) and Thomas Juenger (tjuenger@austin.utexas.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Raw sequencing data have been deposited at NCBI SRA and are publicly available as of the date of publication. Accession numbers are listed in the [key resources table](#), under Bioproject PRJNA822373 and PRJNA919067
- All original code has been deposited at Github and is publicly available as of the date of publication. A link to the repository is listed in the [key resources table](#), <https://github.com/bulksoil/VirgatumMicrobiomeGWAS>.
- Any additional information required to analyze the data reported in this paper is available from the [lead contact](#) upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Plants

The *Panicum virgatum* natural variation panel used in this study was originally described in Lovell et al.³³ Details on field plot locations, geographic origin of accessions, subpopulation and ecotype classification, and genetic relatedness including SNP calls can be found in Lovell et al. Briefly, the diversity population was established by collecting seeds and rhizomes from natural as well as common garden resources and transported to Austin, TX where the accessions were clonally propagated. Switchgrass is an outcrossing perennial plant, hence individuals in the planting populations are clonally propagated ramets and it is not possible to raise identical plants from seed. The genomes for individuals within the population were resequenced and aligned to the reference genome AP13 to identify SNPs. Initial growth of plants and seedlings occurred in a mixture of Promix peat-based potting soil and calcined clay (Turface). Rhizome propagules were transplanted into 5-gallon pots containing finely ground pine-bark mulch and nutrients were supplied through slow-release fertilizer (14-14-14, Osmocote). Final propagation of the accessions occurred in 2018 where ramets were grown in 1-gallon pots containing pine-bark mulch. In May to June 2018 the ramets were transplanted into the common gardens. Briefly, the fields were covered with weed cloth and the layout was arrayed in a honey-comb design with minimum interplant distance of 1.56 m. Holes were cut into the weed cloth and the soil was excavated using a spade shovel. The plants were placed into the holes, surrounded by soil, and hand watered. The lowland cultivar ‘Blackwell’ was planted around the edge of the field sites to account for border effects. We used an augmented unreplicated design for each common garden utilizing the AP13 genotype (the genome reference) as a highly replicated check. This is a common experimental design for multi-location field trials and in the early stages of many breeding programs^{70,71} and focuses on planting single representative genotypes of a diversity panel at many locations. It’s an especially effective design in the case of GWAS and studies of genotype-by-environment interaction, where the focus centers on contrasting alternative alleles as the unit of analysis rather the specific comparisons of individual genotypes. We sampled 729 samples root samples from ATX, 514 root samples from CMO, and 581 root samples from KMI. Included in the analyses were also 48 bulk soil samples from each location.

METHOD DETAILS

Root sample collection and processing

Samples were collected in the summer of 2019. Samples from ATX were collected in June, 2019 while CMO and KMI samples were collected in early August of 2019. The gap in sample collection timing between the sites was intentionally set to account for phenological differences in AP13, the reference genome accession, between locations. The size of our plantings as well as various characteristics of switchgrass plants presented several challenges during sampling. Given that microbiomes can be dynamic, and can potentially respond to weather events, sampling of the fields had to occur within one day. Our plantings are large, and a team of samplers was employed to quickly collect root samples. A 1-inch diameter punch core was used for sample collection. Briefly, the corer was placed at the edge of the crown and rotated to be tangential to the crown. This allowed us to avoid the original potting soil directly underneath the crown where the original transplantation occurred and minimized the chance of capturing legacy microbiota from the pre-transplanted roots. The corers were pushed 10-15 cm below the surface at a 45-degree angle. The soil-bound roots were extracted from the instrument using a scoopula and placed into a plastic baggie. Between samples, the corer was cleaned of remaining soil using a paper towel, but no effort was made to sterilize the instrument between samples as ethanol cannot remove DNA and bleaching / washing the instruments was not feasible for conducting the sampling in a reasonable timeframe. Roots were encased by surrounding soil in the core; therefore the risk of cross contamination was negligible. After a row was completed, the sampler returned to a workstation and the baggies were organized and placed into a cooler with ice packs or wet ice. Bulk soil samples were collected on the same day by collecting soil cores between plants in the field. The samples were placed into plastic baggies and stored on wet ice.

The samples were processed the next day. Living roots from the baggies were picked using ethanol and flame sterilized forceps. Two or three 1-inch pieces of roots were placed into a 2 mL tube with 1 mL sterile PBS. Typical root samples contained both transport roots with attached absorptive roots. The roots were vortexed in PBS for 10 seconds then sterilely transferred to a new, clean tube with 1 mL PBS. The roots were again vortexed to remove soil adhering to the surface and the resulting dirty PBS was discarded. This process was repeated until the PBS solution was clear and no soil remained in the tube. The roots in the tubes were then frozen and stored at -80 degrees until DNA extraction took place.

DNA extraction

DNA was extracted from samples using a procedure similar to Bollman-Giolai et al.⁷² Briefly, root samples are ground to a fine powder with two sterile steel beads in a 2 mL tube using a GenoGrinder for 30s at 1750 rpm. For soil samples, the soils in the baggies were homogenized by squeezing and shaking the bags, then 0.25 g of soil was placed into a tube using a flame sterilized spatula. After grinding roots (soils were not ground), 0.25 g of garnet particles (Lysing Matrix A, BioSpec) were decanted into the tube and 540 μ L of Buffer I (181 mM NaPO₄, 121 mM Guanidinium Thiocyanate) was pipetted into each tube. The samples were briefly vortexed, and 60 μ L of buffer II (150 mM NaCl, 4% SDS, 500 mM Tris pH 8) was added. The samples were then placed into the Genogrinder for 2 min at 1500 RPM to grind / lyse. The tubes were centrifuged at 10,000 g for 1 min to palette debris. The supernatant (500 μ L) was transferred to a deepwell (1mL) 96-well plate and 250 μ L of Buffer III (133 mM Ammonium Acetate) was added to the

samples and vortexed to precipitate SDS and proteins. The plates were placed in 4 degrees for 5 min, then centrifuged at 4000 g. The supernatant (500 uL) was transferred to a new plate and 120 uL of Buffer IV (120 mM Aluminum Ammonium Sulfate Dodecahydrate) was added to precipitate fulvic and humic acids, typical PCR inhibitors from plant and soil samples. The samples were put at 4 degree for 5 min, then centrifuged for 2 min at 4000 g. After this step, the supernatant can be frozen /stored or directly used for the next SPRI bead purification step. For the SPRI cleanup, 300 uL of the supernatant is mixed with 240 uL of SPRI beads in a deepwell 96-well plate and incubated for 5 min. The plates were then placed on a magnet, allowed to clear, and the supernatant was discarded. The beads were then washed twice with 80% ethanol and allowed to dry for 5 min. DNA was then eluted using 50 uL of water and transferred to a 96 well plate for storage at -20.

Library preparation and sequencing

We amplified the V4 region of 16S rRNA gene to survey microbial membership and relative abundance in the samples. We used a two-step strategy, where V4 regions were first amplified using modified primers published by Parada et al.⁷³ The primers were modified to add nextera sequencing primer annealing sites to the amplicons. The resulting PCRs were checked for amplification on a gel and cleaned using SPRI beads. The second round of PCR added barcodes and flow cell annealing adapters to the amplicons. Our barcoding strategy adds 12 bp Golay barcodes to both ends of the amplicon. The libraries were purified again using SPRI beads and quantified using Qubit high sensitivity assays. The amplicons were normalized for concentration by pooling samples at different volumes depending on their concentrations. The resulting pools were then concentrated using SPRI beads and run on a 2% agarose gel. The appropriate band was cut from the gel and purified (Nucleospin) and sent for sequencing.

Sequencing occurred at multiple centers. Our first two library pools contained the ATX samples and were sent to both the HudsonAlpha Genomic Sequencing Facility and to the Joint Genome Institute (JGI). Therefore, these samples were sequenced twice and the reads attributable to corresponding samples were pooled. This explains why ATX samples had such deep sequencing. The library pools for CMO and KMI were sequenced at JGI. All sequencing was performed using Illumina NovaSeq configured with the SP flowcell which is capable of 250 x 250 bp paired end read lengths.

Sequence processing and ASV calling

Resulting reads were demultiplexed, if needed, using the demultiplex Python software (<https://demultiplex.readthedocs.io/en/latest/index.html>). Reads were trimmed to remove adapter sequences using cutadapt.⁶¹ ASVs were called using the dada2 R software package.⁶² The forward reads were trimmed to 240 bp while the reverse reads were trimmed to 230 bp. A maximum of 1 expected error was allowed for both the forward and reverse reads during the filtering process of the DADA2 pipeline. ASVs identified as chimeras were discarded from the ASV table along with ASVs less than 248 bp and greater than 256 bp. A taxonomy was assigned to each ASV sequence using DADA2's assignTaxonomy() function using the Silva version 138.1 reference database.⁷⁴ ASVs with taxonomies assigned to mitochondria or chloroplast were discarded as host contamination and therefore removed from the analysis. Samples with less than 10,000 reads were removed from the analysis. The count data was converted to relative abundance on a per-mille scale by dividing the raw count by the library total count and multiplying by 1000.

Beta diversity measurements

Bray-Curtis dissimilarities were calculated using the *vegdist* function from the Vegan R package⁶⁶ on log₂ transformed ASV relative abundances. Log₂ transformation brings the count data closer to a normal distribution which better suites the ordination algorithms. Principal coordinate analysis was done using the *capscale* function from the Vegan package. Permanova was conducted using the *adonis* function.

Modeling site and subpopulation effects on ASVs

We used a linear modeling framework to model the effect of field site, genetic subpopulation, and subpopulation by site effects on microbes. To be included in the analysis, an ASV must have been present in >= 50% of the total samples included in the study. For every ASV a linear model was run with the following structure

$$\text{lm}(\text{ASV_abundance}_i \sim \log_{10}(\text{depth}) + \text{Site} + \text{Subpopulation} + \text{Site} : \text{Subpopulation})$$

Where *ASV_abundance_i* is the vector of rank-based inverse normal transformation for the *i*th ASV. This transformation was performed using the function *RankNorm()* from the R package *RNOmni*.⁶³ Sequencing depth was accounted for by including the log₁₀(depth) term in the model. Site represents the vector of field locations and Subpopulation represents the switchgrass genetic population of the host. Site:Subpopulation is the term capturing interaction effects between these two factors. Rank-based inverse normal transformations were performed to coax ASV relative abundances into a normal distribution to better fit the underlying assumptions of the model. Variance partitioning of the terms was performed by running the function *Anova()* from the Car package on individual models and percent variance was calculated by dividing a factor's sum of squares by the total sum of squares. Contrasts across model variables were calculated using the *emmeans* package.⁶⁴

Core microbiome considerations

There are various methods which exist for calling a core microbiome,^{75–77} but the scope and scale at which to define a core microbiome is currently unknown, especially across location, populations, and temporal scales. In this study, we use a prevalence

threshold of 80% to define microbes belonging to the core microbiomes at each location. The overlap of these core microbiota between sites were then termed the study-wide core microbiome. Our reasoning for using the 80% prevalence threshold was two-fold. One is that other studies have used this cutoff, and therefore there is a precedent in this area of research. Secondly, the analyses we perform are sensitive to data distributions and statistical power. Removing samples where the focal ASV was not detected may imbalance data in such a way to make GxE impossible to accurately calculate, if the prevalences are different between the sites. We have included an analysis to show how the size of various core microbiomes change in relation to adjusting the prevalence threshold (Figures S1A and S1B). The list of study-wide and site-specific core microbiota members can be found at <https://github.com/bulksoil/VirgatumMicrobiomeGWAS>.

Which taxonomic level is appropriate for calculating heritability of bacteria

We find that heritable variation of microbiota members can be observed across every taxonomic level. Several studies have calculated heritability of rhizosphere or root associated bacteria.^{17–19,78} Typically, the analysis is conducted at the OTU or ASV level (i.e. the taxonomic level with the highest resolution for metabarcoding). In the case of Sutherland et al., the authors found significant heritable variation for aggregated counts of bacterial families in the switchgrass rhizosphere, but found little evidence for the effect of host population structure at the ASV level.³⁰ This begs the question: which taxonomic level is appropriate for calculating heritability of host-associated bacteria? Our results indicate that, while individual ASVs displayed the greatest association to host genetic variation, relatively high V_A can be observed even at the bacterial order and family level. This observation lends some support to the idea that plants do not select for particular microbes (i.e. specific ASVs), but rather for microbes with particular functional attributes.^{10,25} In some cases, it may be that functional attributes impacting colonization of the host diverge across closely related microbes,⁷⁹ therefore the ASV level may be most appropriate. In other cases, a functional attribute selected for by the host may be conserved across wider evolutionary distances (i.e. a core genomic feature) allowing for detection of h^2 at higher taxonomic levels.^{80,81} Given the differing conclusions that our study has with Sutherland et al., the unit at which to calculate heritability may depend on the plant compartment sampled, e.g. the proximity of association with the host may be an important determinant for these considerations. Uncovering the appropriate unit for calculating heritable signal in host associated microbial communities will be an important challenge for future studies.

Genetic variance component analyses

Additive genetic variance and GxE variance was first calculated using the compound symmetry model in the R package Sommer.⁶⁵ The compound symmetry structure model assumes constant total variance within each site as well as constant covariance between sites. This is the simplest model structure and was selected as the first step in our analysis because the model returns components for additive genetic variance and genotype by environment variance. To be included in the analysis, a feature must have been detected in $\geq 80\%$ of the samples. The full model was run with the following structure.

```
Full_model <- mmer(rst ~ Site + log 10(depth), random = ~ vs(PLANT_ID, Gu = K)
+ vs(Site : PLANT_ID, Gu = EK), rcov = ~ units, data = x2, tolparinv = 1e - 01, verbose = T)
```

rst is the vector of rank-based inverse normal transformed ASV relative abundance (or aggregated relative abundance if classification is above ASV). Rank-based inverse normal transformations were applied to the counts within each site for each ASV and resulted in a constant overall variance, fulfilling this assumption of the compound symmetry structure. In this model Site and sequencing depth were fit as fixed effects. PLANT_ID is the plant accession name and K is the kinship matrix with pairwise relationships between individuals in the population based upon SNP data. Site is the field location and 'vs(Site:PLANT_ID, Gu=EK)' captures the variance of GxE in the model, where EK is a list of site-specific kinship matrices. Reduced models were constructed to test the contribution of $V_{G \times E}$ and V_A to the models. They were encoded as follows

```
reduced_1 <- mmer(rst ~ Site + log 10(depth), random = ~ vs(PLANT_ID, Gu = K), rcov =
~ units, data = x2, tolparinv = 1e - 01, verbose = T)
```

Notably, this model lacks the GxE term 'vs(Site:PLANT_ID, Gu=EK)'. This model was compared to the full model using a likelihood ratio test to examine whether GxE influenced the abundance of the tested ASV. To test for the effect of host genotype, we compared reduced_1 to the below model.

```
reduced_2 <- mmer(rst ~ Site + log 10(depth), rcov = ~ units, data = x2, tolparinv = 1e - 01, verbose = T)
```

This model lacks the effect of genotype altogether, thus comparing reduced_2 to reduced_1 using a likelihood ratio test examining whether host genotype contributes to the observed variance of the tested ASV. To make a call on whether GxE or V_A influenced microbial abundances, we first asked if GxE showed an adjusted P-value < 0.1 . If so, our analysis stopped and we flagged the tested ASV as showing significant GxE. If not, then we tested whether V_A had an effect with an adjusted P-value < 0.1 . If so, we made a call that the ASV is affected by host additive genetic variance. If not, we inferred that the ASV was not affected by host genotype.

We next used the unstructured model in the Sommer package to ask about additive genetic variance within each site. The unstructured model allows for unequal additive genetic variances within sites as well as unequal covariances between sites. This allowed us to ask about the influence of host genotype within sites and whether the influence of host genotype is consistent across multiple sites.

Multiple testing was accounted for through correction by the Benjamini-Hochberg approach, and a significant contribution of either parameter was determined at FDR < 0.1.

Microbial genome-wide associations

To perform GWAS on bacterial community composition, we first performed independent PCoA for each field site. The first three PCs from the ordination of each field site were used as dependent variables in the GWAS scans (see below). We performed GWAS for microbes found in >80% of the samples within each site. For this analysis, where we were performing quantitative models, we removed samples where the focal ASV was not detected and the relative abundance were transformed as previously mentioned using the rank-based inverse normal transformation. Genome wide association analysis were completed using the SwitchgrassGWAS R package (<https://github.com/Alice-MacQueen/switchgrassGWAS>).³³ This package is a wrapper around bigsnpr⁶⁷ package which, for each SNP, fits a simple linear model testing for an additive effect and controls for population structure by incorporating a series of PCs as fixed effects. SwitchgrassGWAS dynamically chooses the number of genetic PCs to include as covariates in the model to control for population structure and reduce genomic inflation. The SNP matrix used in the analysis was dense, composed of over 25 million SNPs with a minor allele frequency > 5% generated from the *Panicum virgatum* V5 genome. GWAS results were examined using a genome-wide significance threshold of 5×10^{-8} to identify SNPs associated with the abundance of various microbes, a common cutoff used in microbiome GWAS studies where many phenotypes are analyzed together.^{82–84} The gene content near SNPs passing a threshold of 5×10^{-8} was generated using BEDTools window⁶⁹ on the *P. virgatum* v5.1 genome annotation with a window size of 50 kb.

For the study-wide core microbiota, i.e. microbes detected in $\geq 80\%$ of the samples in each field site, the P-values for the GWAS scans of each microbe were combined using Fisher's Method from the R package 'metap'⁶⁸. Phenotypic variance attributable to SNPs was calculated using a multi-QTL model on ASVs whose GWAS scans had SNPs passing the significance threshold. Leading SNPs were identified for each significant peak, i.e. the SNP with the lowest P-value within a 25kb window containing SNPs passing the genome-wide significance threshold ($P < 5 \times 10^{-8}$). The variance explained by the allelic variation at each of these loci was calculated using ANOVA using the following parameters with the base R aov() function.

$$\text{aov}(\text{rst} \sim \text{SNP}_x + \text{SNP} \dots + \text{SNP}_x : \text{Site} + \text{SNP} \dots : \text{Site} + \text{PC1} + \text{PC2} + \text{PC3} + \text{PC4} + \text{PC5} + \text{PC5})$$

In this strategy we controlled for population structure using the first six principal components of the kinship matrix (i.e PC1 through PC5). We fit terms for both SNPs and an interaction term between SNP and the different field sites. We report the total variance as a summation of the variance attributable to SNPs and SNPs by location interactions.

Detection of pleiotropic loci affecting multiple microbes

To identify regions of the host genome putatively influencing the abundance of multiple microbes we divided the genome into 25 kb bins, consistent with average linkage equilibrium decays suggested in other switchgrass studies.⁸⁵ For each microbe, this resulted in 43,402 bins. We next calculated the minimum P-value of the SNPs within each bin for each microbe and retained the top 0.5% of bins with the lowest P-values (217 bins which we refer to as QTL bins). The resulting QTL bins were then compiled into a presence / absence matrix and we present 5 bins from each site showing association to the most ASVs for further analysis. We tested the likelihood of observing the number of overlapping loci in our data by using a permutation framework. In our QTL x ASV matrix, the ASVs were the rows and QTL were the columns. We randomized the QTLs for each ASV in the matrix and counted the maximum number of overlaps, stratifying by field location. This was performed 1000 times to develop a null distribution. All of our top 5 pleiotropic loci had $p < 0.001$. We chose to only analyze the top 5 loci for each site for presentability but include the other loci passing this significance threshold in the supplemental tables.

Gene ontology enrichments

We identified the gene content of the QTL matrix composed above using bedtools window,⁶⁹ then extracted the Gene Ontology categories for each gene within each 25 Kb genomic bin. Enrichment was calculated against the background genome GO counts using a hypergeometric test and P-values were corrected for multiple tests using the Benjamini-Hochberg procedure.

Gene expression analysis

The expression values for gene underlying putative pleiotropic loci were extracted from the *Panicum virgatum* gene expression atlas which can be found on Phytozome 13. The FPKM values for *P. virgatum* gene expression across tissues and environments were gathered from the JGI Gene Atlas Project.⁸⁶ Differential expression between root and shoot tissue was performed using the following linear model on FPKM values.

$$\text{lm}(\log 2(\text{expression}) \sim \text{Tissue})$$

The resulting P-values for the term 'Tissue' were corrected using the Benjamini-Hochberg procedure and significance was called at adjusted P-value < 0.05.

Pattern-triggered immunity assays

We used a protocol of Samira et al.⁸⁷ to study plant immune responses to flg22 using leaf tissue collected from the ATX field site plants in the spring of 2020. Leaf disks (3 mm) were punched from the leaves on location in the field and immediately placed in 2 mL of sterile DI water in a 48 well plate and covered with aluminum foil. The plates were gently shaken for 2 hours, then the disks were transferred to white, opaque 96 well plates in 50 uL of sterile DI water, wrapped in aluminum foil, and left overnight. The next day, the disks were treated with 50 uL of Flg22 elicitor cocktail (10ug/mL horseradish peroxidase, 34 ug/mL L-012, and 1 uM Flg22). The plates were read over a time series on a SpectraMax M3 plate reader. Negative control plates with a randomly selected group of genotypes were mock treated (10ug/mL horseradish peroxidase, 34 ug/mL L-012, water). Each genotype was read in triplicate. To analyze the data, we log transformed the relative luminescence units of the time series and reduced the dimensionality using PCA with the `princomp()` command from base R.

Genetic covariances of PTI axes and bacterial abundances

We estimated genetic covariances between the first three PTI PCA axes and ATX root microbe relative abundances using the R package `Sommer`. We used the following mixed effects model.

```
covar_mod <- mmer(cbind(ASV_abund, PTI_PC) ~ 1, random = ~ vs(PLANT_ID, Gu = K), data = data, tolparinv = 1e - 1)
```

The terms for `ASV_abund` and `PTI_PC` changed depending on the focal ASV and focal PTI PC axis. Covariance estimates and standard errors for the estimates were gathered using the following command.

```
covar <- vpredict(covar_mod, covar ~ V2 / sqrt(V1 * V3))
```

P-values for observing the covariance estimate or larger (in magnitude) were calculated as $p = 2 * \text{pnorm}(\text{estimate} / \text{standard_error}, \text{lower.tail} = \text{FALSE})$

QUANTIFICATION AND STATISTICAL ANALYSIS

The R programming environment (version 4.2.1) was used for data analysis and visualization. All statistical tests were performed in R. Between group contrasts in linear models were conducted with the packages `emmeans` (Searle et al.⁶⁴). ASV relative abundance were transformed using rank-based inverse normal transformations with the function `RankNorm` from the R package `RNOmni` (1.0.1 McCaw et al.⁶³). Meta-analysis of multiple GWAS was performed using the `sumlog` function from the `metap` package (1.8, Dewey⁶⁸). Genetic variance components were estimated using the `mmer` function from the `sommer` package (4.2.0.1 Covarrubias-Pazaran⁶⁵). Principal coordinates analysis was performed using the function `capscale` from the `Vegan` packages (2.6.4 Oksanen et al.⁶⁶). Principal components analysis was performed using the `prcomp` function from the `stats` package. Permutational MANOVA was performed using the `adonis2` function from the `Vegan` package (2.6.4 Oksanen et al.⁶⁶). GWAS was performed using the `switchgrassGWAS` R package which uses tools from the `bigsnpr` package (1.11.6 Privé et al.⁶⁷) from Where mentioned, false discovery rate was controlled for with multiple testing correction by the Benjamini and Hochberg method using the `p.adjust` function from the `stats` package (Benjamini and Hochberg⁸⁸).