

Host genomic influence on bacterial composition in the switchgrass rhizosphere

Jeremy Sutherland^{1,2,3}  | Terrence Bell^{1,2,4}  | Ryan V. Trexler^{3,5}  | John E. Carlson^{2,5} |
Jesse R. Lasky³ 

¹Department of Plant Pathology and Environmental Microbiology, The Pennsylvania State University, University Park, Pennsylvania, USA

²Intercollege Graduate Degree Program in Bioinformatics and Genomics, The Pennsylvania State University, University Park, Pennsylvania, USA

³Department of Biology, The Pennsylvania State University, University Park, Pennsylvania, USA

⁴Department of Ecosystem Science and Management, The Pennsylvania State University, University Park, Pennsylvania, USA

⁵Intercollege Graduate Degree Program in Ecology, The Pennsylvania State University, University Park, Pennsylvania, USA

Correspondence

Jeremy Sutherland, Department of Plant Pathology and Environmental Microbiology, The Pennsylvania State University, University Park, PA, USA.
Email: jus1990@psu.edu

Funding information

Foundation for the National Institutes of Health, Grant/Award Number: 1R35GM138300-01; National Institutes of Health, Grant/Award Number: 1R35GM138300-01; USDA; U.S. Department of Agriculture, Grant/Award Number: 2012-68005-19703; National Institute of Food and Agriculture, Grant/Award Number: PENO4532, 1000326 and 2019-67009-29006; Cornell University

Handling Editor: Jeremy B Yoder

Abstract

Host genetic variation can shape the diversity and composition of associated microbiomes, which may reciprocally influence host traits and performance. While the genetic basis of phenotypic diversity of plant populations in nature has been studied, comparatively little research has investigated the genetics of host effects on their associated microbiomes. Switchgrass (*Panicum virgatum*) is a highly outcrossing, perennial, grass species with substantial locally adaptive diversity across its native North American range. Here, we compared 383 switchgrass accessions in a common garden to determine the host genotypic influence on rhizosphere bacterial composition. We hypothesized that the composition and diversity of rhizosphere bacterial assemblages would differentiate due to genotypic differences between hosts (potentially due to root phenotypes and associated life history variation). We observed higher alpha diversity of bacteria associated with upland ecotypes and tetraploids, compared to lowland ecotypes and octoploids, respectively. Alpha diversity correlated negatively with flowering time and plant height, indicating that bacterial composition varies along switchgrass life history axes. Narrow-sense heritability (h^2) of the relative abundance of 21 core bacterial families was observed. Overall compositional differences among tetraploids, due to genetic variation, supports widespread genotypic influence on the rhizosphere microbiome. Tetraploids were only considered due to complexities associated with the octoploid genomes. Lastly, a genome-wide association study identified 1861 single-nucleotide polymorphisms associated with 110 families and genes containing them related to potential regulatory functions. Our findings suggest that switchgrass genomic and life-history variation influences bacterial composition in the rhizosphere, potentially due to host adaptation to local environments.

KEYWORDS

heritability, host genomic influence, microbiome, rhizosphere, switchgrass

1 | INTRODUCTION

Macroscopic organisms host diverse microbiomes, which often differ substantially among individuals (Smith et al., 2015; Trivedi et al., 2020; Wilson et al., 2020). Some of the proximate causes of this variation in microbiome composition arise from differences between host environments and traits, the latter of which can be divided into genetic and plastic components. Host genetic differences can affect a wide range of traits that influence associated microbiome composition, including traits that determine the supply of resources used by microbes or traits that control environmental variables affecting microbial fitness in other ways (e.g., pH) (Lebeis et al., 2015; Sasse et al., 2018). These host genetic influences on microbiomes may arise from different evolutionary processes. For example, host traits that affect microbiome composition may evolve neutrally or in response to nonmicrobial components of the local environment (Zeng et al., 2015). Alternatively, host trait variation could have evolved as a direct response to beneficial interactions with microbes in the environment (Gould et al., 2018). In the latter case, hosts could either influence the abundance of microbes that are simply present in the environment, and functionally important to the host, or hosts could form species or strain-specific symbiotic relationships where the host depends on and influences the abundance of those taxa for reproductive success (Parker et al., 2006).

Because of environmental and dispersal limitations, the same microbes may not be present across the entire range of a host, although functionally similar ones might be. It has been shown that reciprocal interactions between host immune systems and host-associated microorganisms can lead to preferential recruitment of beneficial microbial types and repel pathogenic types, to varying degrees (Hacquard et al., 2017; Jones et al., 2019). This suggests that hosts might (directly or indirectly) select for functional microbial traits from certain lineages. Several studies have observed the presence of “core microbiota” – persistent members of the microbiome that appear across a large portion of a host's population or species (Risely, 2020), but it is often unclear whether these microorganisms are functionally important or commensal to the host. Yet, because certain microbial traits are often phylogenetically conserved and host selection of microbes can occur in related microbial lineages (Lemanceau et al., 2017; Martiny et al., 2015), a core microbiome might be preserved for higher taxonomic levels but undetectable for lower ones, at the same occupancy frequency threshold. Relatedly, many functional microbial strains that have consequences for plants are often rare, such as plant pathogens, and abundance-defined cores miss important members that are not dominant in the community. The study of diverse host genotypes and phenotypes in common environments could reveal the genetic and host physiological drivers of variation in associated microbiomes and, potentially, types of microbes that are functionally important to the host.

The demonstrated influence of host trait variation on microbial assemblages (Hestrin et al., 2021; Jones et al., 2019; Wagner et al., 2016) suggests that microbiome composition can be treated

as an extension of host phenotype, one that may be under selective pressure by both the host and the environment (Bordenstein & Theis, 2015; Hunter, 2018; Moran & Sloan, 2015; Wagner et al., 2016; Whitham et al., 2003). Genome-wide association studies (GWAS) have shown that the relationship between hosts and specific microbial taxa can be linked to single-nucleotide polymorphisms (SNPs) (Beilsmith et al., 2019). However, at the microbiome level, it becomes challenging to directly link host genotype and microbial composition, due to the high diversity and dimensionality of microbiomes (Wray et al., 2013). Common garden experiments and genetic mapping offer a window into host genetic impacts on host-associated microbial composition. The resolution of association mapping can be improved by exploiting high levels of genomic diversity within and between populations (Holland, 2007).

Switchgrass (*Panicum virgatum*) is a C4 perennial grass species that has been extensively studied as a bioenergy crop since 2005 (Bouton, 2007; McLaughlin & Kszos, 2005; Parrish & Fike, 2005; Sanderson et al., 2006). Unlike first generation biofuels, switchgrass can be grown on marginal land with minimal inputs while also providing ecological services (e.g., wildlife habitat and soil erosion mitigation) (Werling et al., 2014). In addition to its ecological and agricultural value, switchgrass provides an ideal system for studying the influence of host genotypic diversity on rhizosphere bacterial composition (i.e., bacteria living on or in close proximity to the root system of the plant) because of the high degree of genotypic diversity in the species.

Switchgrass is primarily an outcrossing species with high genetic diversity within populations, showing evidence of inbreeding depression, which increases heterozygosity in the genome (Sharma et al., 2012). The species is comprised of two locally-adapted ecotypes, upland and lowland, based on differences in phenotype, physiology, and habitat (Das et al., 2004). Separate migration events fostered by interglacial periods, in addition to gene duplication through polyploidization, have also played a major role in shaping the genomic diversity of switchgrass (Casler et al., 2015). Lowland switchgrass is primarily allotetraploid ($4x: n = 36$) and exhibits disomic inheritance – chromosome inheritance that follows diploid association patterns. Upland accessions tend to vary in their ploidy level ($4x: n = 36$, $6x: n = 54$, $8x: n = 72$) and exhibit widespread aneuploidy between and within localized populations (Costich et al., 2010). Additionally, populations of switchgrass are adapted to local climate conditions and linkage mapping suggests a polygenic basis to this local adaptation (Lowry et al., 2019). Here, we studied genotypes largely from the geographically Northeastern, upper Southern, and Midwestern United States that can be genetically clustered into five groups (roughly North, South, East, West, and Northeast, with some geographic overlap) (Figure 1, Figure S1). The relationship between genetic cluster, ploidy, and ecotype is largely consistent across genotypes in this study. However, the South genetic cluster is less homogenous (Figure 1). Similar relationships for these genotypes were previously defined using genotype-by-sequencing (GBS) data (Lu et al., 2013). We have opted to retain

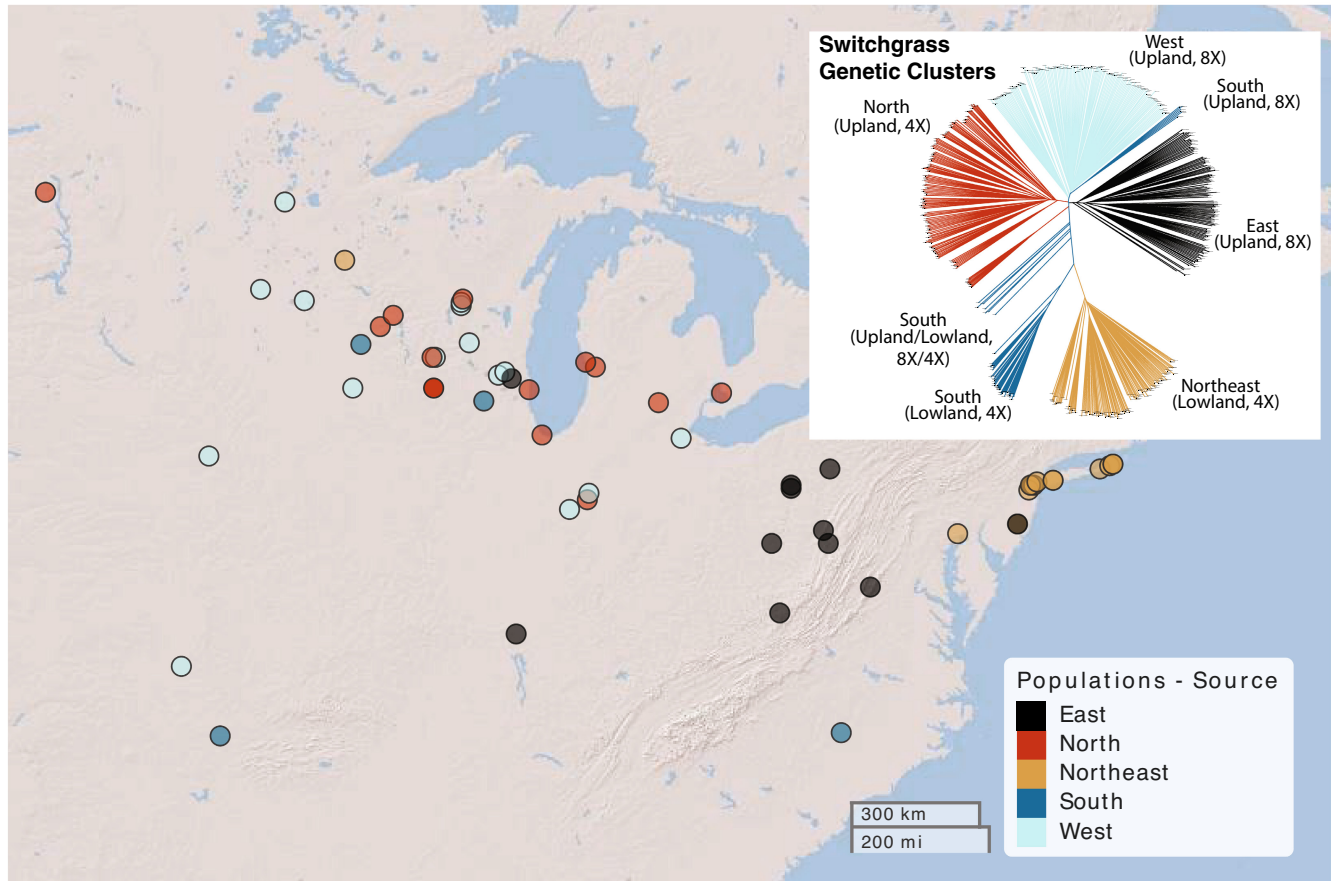


FIGURE 1 Accession map and phylogeny. Accession source population locations and phylogenetic analysis with switchgrass genetic clusters coloured as North (red), South (dark blue), Northeast (beige), West (light blue), and East (black)

the previous cluster names for continuity between studies despite some differences in inferred relatedness, particularly with certain genotypes in the South genetic cluster. Genetic variation in switchgrass has been linked to measures of plant fitness and performance, including biomass production (Lowry et al., 2019) and soil carbon inputs (Adkins et al., 2016; Stewart et al., 2017). While substantial research has explored the switchgrass microbiome, key gaps remain, including the functional role of nonfungal microorganisms and host mechanisms that underly switchgrass–microbiome interactions (Hestrin et al., 2021).

Vascular plants influence microbes in the rhizosphere through rhizodeposits (Dennis et al., 2010) and carbon-based root exudates (Compant et al., 2010; Hu et al., 2018; Yu & Hochholdinger, 2018). Root exudates provide signalling and substrate compounds for beneficial microbial recruitment (Carvalhais et al., 2013; Lakshmanan et al., 2012), and antibiotic compounds that defensively modulate the microbiome composition against pathogenic taxa (Baetz & Martinoia, 2014; Lebeis et al., 2015). When available, root exudates potentially represent a larger portion of carbon in the soil compared to other carbon inputs from the environment (e.g., leaf litter) (Heijboer et al., 2018). Switchgrass can benefit from a range of microbes that are influenced by root exudates (Hestrin et al., 2021).

In this study, we analysed bacterial 16S rRNA gene sequences from total DNA extracted from rhizosphere soil samples for 383 switchgrass accessions from 63 local populations planted within a common garden in Ithaca, NY, USA. Our aim was to better understand the influence of switchgrass genotype on the composition of bacteria in the rhizosphere. We hypothesized that overall bacterial composition and diversity within the growing site would differentiate due to host genetic variation along the three dimensions of switchgrass genetic diversity (ecotype, ploidy level, and genetic cluster), as is true for multiple fitness-related traits (Lovell et al., 2021). Separately, bacteria experience higher rates of gene loss, horizontal gene transfer, and shorter generation times compared to their hosts. Therefore, microbial specialization on a given host can be evolutionarily fast compared to changes in the plant community (Bever et al., 2012). Given that the evolutionary response rate is mismatched between switchgrass genotypes and their microbiomes, we sought to determine the lowest taxonomic level for which a potential host influence could be detected. We show that switchgrass genotype influences bacteria in the rhizosphere, at least, at the family level. Our findings demonstrate how plant host genotype can influence host-associated microbiomes, the taxonomic level at which the microbiome is influenced, and the host genetic architecture involved.

2 | 2 MATERIALS AND METHODS

2.1 | Switchgrass germplasm data

The original switchgrass association panel consisted of 3–10 clonally propagated individuals each from 66 discrete switchgrass local populations grown from seed in the greenhouse of the USDA-ARS Dairy Forage Research Centre in Madison WI in 2007 (Lu et al., 2013). The accessions were sourced from much of the natural range of switchgrass in the eastern and midwestern United States. Previously, exome capture sequencing was used to characterize variation in the genic regions of the v1.1 switchgrass reference genome (<https://phytozome-next.jgi.doe.gov/>) for the Northern Switchgrass Association Panel used in this study ($n = 537$) (Evans et al., 2018). For this study, the HapMapv2 matrix was obtained from the Dryad Digital Repository (10.5061/dryad.mp6cp) in 2019. Several modifications were made to the original data set. The data set was modified in python (Van Rossum & Drake, 2009) to reflect heterozygote positions more conservatively than the original, unfiltered, data set, such that heterozygote positions with read count ratios >0.75 or <0.25 were called homozygous relative to the appropriate allele (e.g., A(20)/T(1) = A). SNPs with a minor-allele frequency (MAF) below 0.05 were excluded, resulting in 103,776 SNPs for all genotypes studied here ($n = 362$) spanning 18 chromosomes: Chr01–09 (subgenomes a and b) of the v1.1 *P. virgatum* reference genome (<https://phytozome-next.jgi.doe.gov/>) and fifteen unanchored scaffolds (Evans et al., 2018). The unanchored scaffolds of Evans et al. (2018) were renamed to the corresponding contig names within the v1.1 reference genome using python. A phylogenetic tree was generated in Tassel v5 (Bradbury et al., 2007) to illustrate the relationship between the five genetic clusters.

We compared bacterial composition to published data on several important switchgrass functional and life history traits, using data from 2009, 2010, and 2011 field seasons at the common garden in Ithaca, NY (genetic variation in these traits is largely consistent from year to year) (Lipka et al., 2014). The 2011 measurements for anthesis date and full plant height were used to identify correlations with rhizosphere alpha diversity present in 2016. Plant height and anthesis date are major determinants of biomass yield since later flowering allows for an extended period of vegetative growth and increased biomass production. These traits are also key life history traits in plant evolution (Anderson, Condron, et al., 2011; Anderson, Willis, Mitchell-Olds, 2011) and have been linked to climate adaptation in switchgrass (Lovell et al., 2021; Lowry et al., 2015).

2.2 | Rhizosphere sampling of switchgrass accessions

Young plants were deposited in the common garden in Ithaca, NY, USA in 2008. In 2016, plugs were collected for transplanting from 525 cloned accessions, representing all but two (perished) original populations. Prior to replanting the plugs, soil samples were

collected into a polythene bag from each of the switchgrass accessions by vigorously shaking off 15 to 50ml of soil attached to the roots (amounts of soil varied because the number of tillers and size of root structure varied widely among genotypes based on variation in growth among plants at the common garden site). The bags with soil and fine roots were initially stored at 4°C while being processed - each soil sample was thoroughly mixed into a uniform distribution and then transferred to 50ml conical cryogenic tubes and stored at -80°C prior to soil DNA extraction.

2.3 | Soil DNA extractions

To determine the composition of bacteria established in the switchgrass rhizosphere, we selected soil samples from 383 of the 525 cloned switchgrass accessions, based on a phylogenetically determined set of clones from 63 of the local populations in the association panel. Total DNA was extracted from ~300mg of soil using the NucleoSpin Soil 96 kit (Macherey-Nagel). Lysis was performed using the buffer SL with Enhancer SX, and was performed on the FastPrep 24 homogenizer (MP Biomedicals) at 4.0 m/s for 30s.

2.4 | Amplicon sequencing

Briefly, amplicons targeting the V3–V4 region of the 16S rRNA gene were generated from all 383 samples using universal bacterial primers 515F (5'-GTGYCAGCMGCCGCGGTAA-3') and 806R (5'-GGACTACNVGGGTWTCTAAT-3') (Apprill et al., 2015; Parada et al., 2016) with overhangs for attaching barcodes and standard Illumina overhang adaptors in a second PCR step (full protocol outlined in the supplemental methods of Trexler & Bell, 2019). Sequencing was performed on an Illumina MiSeq using the 2 × 250cycle v2 kit. Raw reads are available through NCBI SRA under project number PRJNA689762.

2.5 | Amplicon sequence analysis pipeline

The raw 16S rRNA gene sequences were analysed using an adapted version of the dada2 pipeline (Callahan et al., 2016). dada2 was used to process the raw sequences into Amplicon sequence variants (ASVs). Reads were truncated above 240 bp and below 160 bp. Sequences with any undetermined bases ($\text{maxN} = 0$) after truncation were discarded. Reads were then truncated at the first instance of a quality score <2 . After truncation, reads that matched against the phiX Genome were discarded. Reads with higher than two "expected errors" were also discarded. Expected errors are calculated from the nominal definition of the quality score: $\text{EE} = \sum(10^{-Q/10})$. Filtered sequences were used to determine the error rate using the dada2 function: learnErrors(). The filtered sequences were then processed with the core sample inference algorithm, dada(), incorporating the learned error rates. Forward and reverse sequence reads were then

merged. A sequence table was made and chimeras were removed using the “consensus” method. Taxonomic assignments for the ASVs were defined against the SILVA 138 ribosomal RNA gene database (Quast et al., 2013) using DECIPHER v2.14.0 (Wright, 2016). The sequence variant table and taxonomy table were exported for downstream processing in phyloseq (McMurdie & Holmes, 2013).

In phyloseq, ASVs designated “NA” at the phylum level, non-bacterial entries at the domain level, and those associated to chloroplasts and mitochondria were removed from the taxonomy file before further processing. Finally, we also removed samples with no ASVs remaining after pruning. Rarefying samples has been shown to effectively reduce false discovery rates when there are large differences between the average sample library size (Weiss et al., 2017). Our data was characteristic of this scenario. Therefore, we rarefied samples to 1000 sequences, to retain as many samples as possible while also maintaining a sufficient sampling depth to detect the most prevalent taxa comprising the switchgrass core microbiome (see below).

2.6 | Alpha diversity analysis

Using the rarefied ASV data set, we generated alpha diversity metrics using phyloseq (McMurdie & Holmes, 2013). Statistical significance was calculated using the vegan 2.5.6 package (Oksanen et al., 2019), and other core functions in R version 3.6.3 (Ihaka & Gentleman, 1996). The Chao1 and Shannon diversity indices were calculated using the estimate_richness() function in phyloseq and then used to test for differences in alpha diversity between genotypes differing in ecotype, ploidy level, and genetic cluster groups using an analysis of variance (ANOVA) (R version 3.6.3). Correlations between diversity metrics, anthesis date and full plant height were generated using the Pearson correlation test.

2.7 | Beta diversity and core microbiome analysis

Bacterial diversity data rarely conform to the assumptions of MANOVA-like procedures, largely due to inflated zero values among rare taxa that skew the distribution, thus nonparametric methods based on permutation tests are preferred (Anderson, 2001). A nonparametric, one-way analysis of variance (NPMANOVA) is also tolerant of nonindependent observations (e.g., microbe-microbe interactions). We therefore calculated an NPMANOVA using the adonis2 function in vegan to test for significant differences between the ecotype, ploidy level, and genetic cluster groups on the overall rhizosphere microbiome composition using pairwise Bray-Curtis distances at the ASV level.

The “core microbiome” can be defined as microbial taxa that associate with a host above a particular occupancy frequency threshold (i.e., their prevalence). Biological justifications for such thresholds tend to be subjective, often ranging between 30 and 95% (Risely, 2020). However, core taxa prevalence within a population

could be a reflection of an environmental response by the host due to local adaptation, resulting in higher intergroup prevalence of certain “core taxa” relative to the whole population. Microbial abundance, as opposed to prevalence, may not reflect the influence they exert on the host, or vice versa (e.g., potentially pathogenic taxa being influenced by the host). Thus, we determined two core microbiomes: A high-fidelity core (prevalence $\geq 90\%$), to determine taxa that reliably associate with the vast majority of our panel, and a low-fidelity core (prevalence $\geq 10\%$), to study host genetic variation effects on bacterial composition. Taxa present within our rarefied data set were initially agglomerated using the tax_glom() function at all taxonomic levels in phyloseq. Sample counts were then transformed to relative abundances. A filter was applied to subset phylogenetic groups in abundance $>2\%$. Each core microbiome was determined using the microbiome package in R (Lahti et al., 2017). The core() function in the microbiome package (detection = 0, prevalence = 0.9 and 0.1, respectively) was applied at each taxonomic level to determine the core taxa among all genotypes.

2.8 | Heritability estimates

Heritability is an estimate that describes the proportion of phenotypic variance that is due to genetic variance. Narrow-sense SNP-based heritability (h^2) was estimated using the sommer package in R as: σ_A^2 / σ_p^2 , where σ_A^2 is the additive variance and σ_p^2 is the total phenotypic variance (additive + dominance) (Covarrubias-Pazarán, 2016). The core function of the sommer package is the mmer function that fits multivariate linear mixed models. The modified exome capture SNP data set, described above, was used to generate a kinship matrix used in mmer to estimate narrow-sense heritability for each of bacterial families' abundance. The sommer package includes a tolerance parameter for the matrix inverse when singularities are encountered in the estimation procedure. Here, inversion tolerance parameters were adjusted to 10 to avoid model singularity.

2.9 | Redundancy analysis (RDA) and variance partitioning

To determine how whole microbiome variation changed across host genomic variation, much of which is due to population structure, we implemented redundancy analysis (RDA). RDA models variation in a set of response variables as a function of a set of explanatory variables. We used the first 10 principal components of switchgrass SNPs as explanatory variables that describe population structure in switchgrass (Brown et al., 2018) and log-scaled ASV counts for each switchgrass genotype as the response variables. We performed the RDA to explain the relative proportion of total genotypic variance on the overall microbiome composition. The rda() function in vegan (Oksanen et al., 2019) was used to perform the RDA for the tetraploid accessions. For host genomic analyses such as this RDA, we focused on tetraploids as they (1) represent both upland and

lowland ecotypes, (2) exhibit disomic inheritance unlike octoploids, and (3) exhibit more intergroup alpha diversity variation compared to octoploids.

ASVs beyond (+/-) 1.58 times the interquartile range (IQR) were considered outliers for the first two RDA canonical axes; these ASVs with the strongest loadings are those most representative of microbiome-wide turnover across switchgrass hosts of different genomic background. ASVs that were labelled as outliers were then assigned taxonomy using the phyloseq method described above, and compared to assigned families in the original ASV data set. ASVs in the full microbiome data set ($n = 31,181$) belonged to 268 families. Outlier ASVs ($n = 9215$) belonged to 188 families. ~70% of the families in the original data set match the outlier data set (i.e., 188 of 268 families are represented by ASVs in both data sets).

To dissect how genomic variation combined with ecotype, ploidy, and phenotypes to explain bacterial compositional turnover we also implemented variance partitioning of the RDA. Variance partitioning allows one to estimate the portions of compositional turnover explained by multiple sets of factors (Peres-Neto et al., 2006), which here were the first 10 principal components of SNPs, ecotype, geography-of-origin (latitude, longitude, and their squared values), and our two focal phenotypes (anthesis date and height). We implemented two versions of variance partitioning with RDA: one on ASV counts and the other on family abundances, to test whether family-level composition was differently associated with these factors than was ASV level composition.

2.10 | Genome-wide association study (GWAS)

To identify specific loci in the switchgrass genome linked to variation in rhizosphere bacterial composition, a genome-wide association study (GWAS) was conducted using the statgenGWAS package in R (van Rossum et al., 2020). This fast single trait GWAS method was developed by Biometris, following the method described in Kang et al. (2010). The modified SNP data set described above was once again used for this analysis.

Family-level abundance counts data (1) were imported from phyloseq using the rarified phyloseq object and a chromosome positional map (2) was generated from the SNP matrix (3). These three components were used to construct the R object used by statgenGWAS. Initial results from the heritability estimates and RDA indicated that certain core families of bacteria were influenced by switchgrass genotype. Therefore, to determine associated genes, a single trait GWAS, where the trait was the abundance of each of the families, was completed using a generalized least squares (GLS) method for estimating the marker effects and corresponding p -values. Our GWAS included random effects correlated according to the kinship matrix, which was calculated with the VanRaden method (VanRaden, 2008). A p -value threshold is required for this method to minimize the false discovery rate (FDR) following the algorithm proposed by Brzyski et al. (2017). The method limits the number of SNPs in statistical linkage with each other passing FDR control.

Significant SNPs were identified using a 0.01 threshold for the false-discovery rate (FDR). Linkage disequilibrium (LD) in switchgrass decays over kilobase length. Following Grabowski et al. (2017), we used a window of 25 kb to identify genes potentially linked to significant SNPs. Thus, in this study, we defined SNPs within 25 kb as being "within LD of significant SNPs". For each trait, statgenGWAS outputs significant SNPs and SNPs within the defined cutoff. A full description of the method can be found in van Rossum et al. (2020).

2.11 | GO enrichment analysis

Significant SNPs associated with bacterial family abundance and positions within 25 kb of significant SNPs were then used to identify associated genes and gene ontologies. Here, we focused on three families: Xanthobacteraceae, Sphingomonadaceae, and Micromonosporaceae. The Xanthobacteraceae and Sphingomonadaceae families were chosen because they represent the high-fidelity core microbiome. Micromonosporaceae was chosen due to its high relative heritability within our panel and its importance in agriculture applications (Trujillo et al., 2014). *Phytomine* was used to identify gene IDs for each family using the *P. virgatum* v1.1 reference genome at <https://phytozome-next.jgi.doe.gov/>. Gene IDs for corresponding SNPs were then used to perform a singular enrichment analysis (SEA) against the gene ID (version 4) background for *P. virgatum* using the agrigo v2 web program (Du et al., 2010). Default parameters were used.

3 | 3 RESULTS

3.1 | 3.1 Amplicon sequence analysis

For 383 initial rhizosphere soil samples, 96,902 amplicon sequences variants (ASVs) were obtained following initial quality filtering and sequence processing in dada2 (Callahan et al., 2016). After sample pruning (removing nonbacteria from the taxonomy table) and rarefaction (1000 sequences/sample) in phyloseq, 365 samples containing 31,181 ASVs remained. Counts for unique phylogenetic classifications are as follows: Domain: 1, Phylum: 26, Class: 59, Order: 142, Family: 268, Genus: 510.

3.2 | 3.2 Alpha diversity analysis

We detected significant differences in alpha diversity (Chao1 and Shannon diversity) between groups. Upland ecotypes, which are adapted to regions at higher latitudes and those geographically closer to our growing site, exhibited higher rhizosphere diversity compared to lowland ecotypes (Figure 2a, Upland-Lowland: $p_{\text{adj}}[\text{Shannon}] = 1.58e^{-2}$). Tetraploids, which are present in both upland and lowland ecotype groups, also exhibited higher diversity overall, compared to octoploids, which are exclusively upland

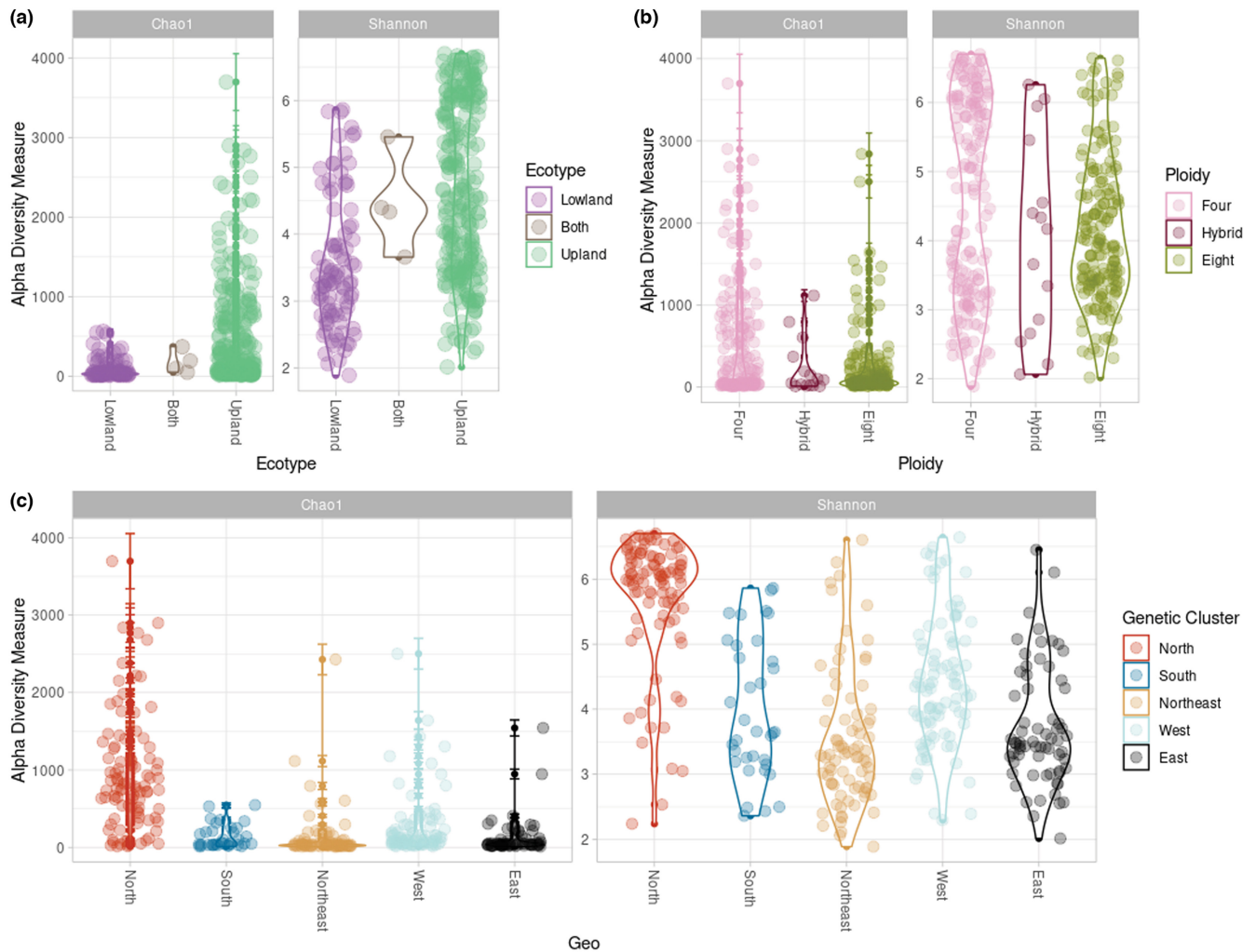


FIGURE 2 Alpha-diversity, measured by Chao1 and Shannon diversity index, plotted for (a) switchgrass ecotypes with Lowland (purple), Both (brown), and Upland (green); (b) ploidy levels with Tetraploid – Four (pink), Hybrid (maroon), and Octoploid-Eight (olive); and (c) genetic clusters with North (red), South (dark blue), Northeast (beige), West (light blue), and East (black)

(Figures 2b and 4–8x: p_{adj} [Shannon] = <.001). Finally, the North genetic cluster, primarily comprised of upland tetraploids, exhibited the highest rhizosphere diversity among all of the genetic clusters and there was a significant difference between the East and West genetic clusters (Figure 2c, North-East: p_{adj} (Shannon) = <.001, North-West: p_{adj} (Shannon) = <.001, North-South: p_{adj} (Shannon) = <.001, East-West: p_{adj} (Shannon) = $5.92e^{-2}$). All diversity measurements for accessions can be found in Table S1.

Overall, we found that Chao1 and Shannon diversity was negatively correlated with anthesis date (Chao1: $r = -0.17$, p -value = .0014; Shannon: $r = -0.18$, p -value = <.001) and full plant height (Chao1: $r = -0.18$, p -value = .0012; Shannon: $r = -0.13$, p -value = .017) (Figures 3a,e). We also considered these trait-diversity associations stratified by ecotype, ploidy, and genetic cluster (Figures 3b–h). In general, the negative associations between alpha diversity and anthesis date, and height were significant while positive associations were generally not. Shannon diversity negatively correlated with plant height for Upland ecotypes ($r = -0.25$, p -value = <.001),

octoploids ($r = -0.23$, p -value = .0045), hybrids ($r = -0.73$, p -value = .016), the North genetic cluster ($r = -0.27$, p -value = .011), the West genetic cluster ($r = -0.24$, p -value = .032), and marginally the Northeast genetic cluster ($r = -0.22$, p -value = .061) (Figures 3b–d). Likewise, Shannon diversity negatively correlated with anthesis date for Upland ecotypes ($r = -0.19$, p -value = .002), tetraploids ($r = -0.23$, p -value = .0025), hybrids ($r = -0.94$, p -value = <.001), and the Northeast genetic cluster ($r = -0.45$, p -value = <.001) (Figures 3f–h). A subset ($n = 7$) of ecotype/ploidy/genetic cluster combinations showed positive trait correlations with Shannon diversity, but these were not significant.

3.3 | Beta diversity analysis

We performed separate multivariate analyses of variance on a Bray-Curtis distance matrix using the NPMANOVA test. There were significant differences in overall bacterial composition between ecotypes ($p < .001$), ploidy levels ($p < .001$), and genetic

clusters ($p < .001$), indicating differences in overall composition associated with switchgrass genetics and life-history. Intergroup pairwise Bray-Curtis distances can be found in Figure S2.

Although all taxonomic levels were tested, the family level was identified as the lowest taxonomic rank with the ability to retain a high-fidelity core (Figure S3). The high-fidelity core microbiome ($> = 90\%$ of samples) contained two families: Xanthobacteraceae

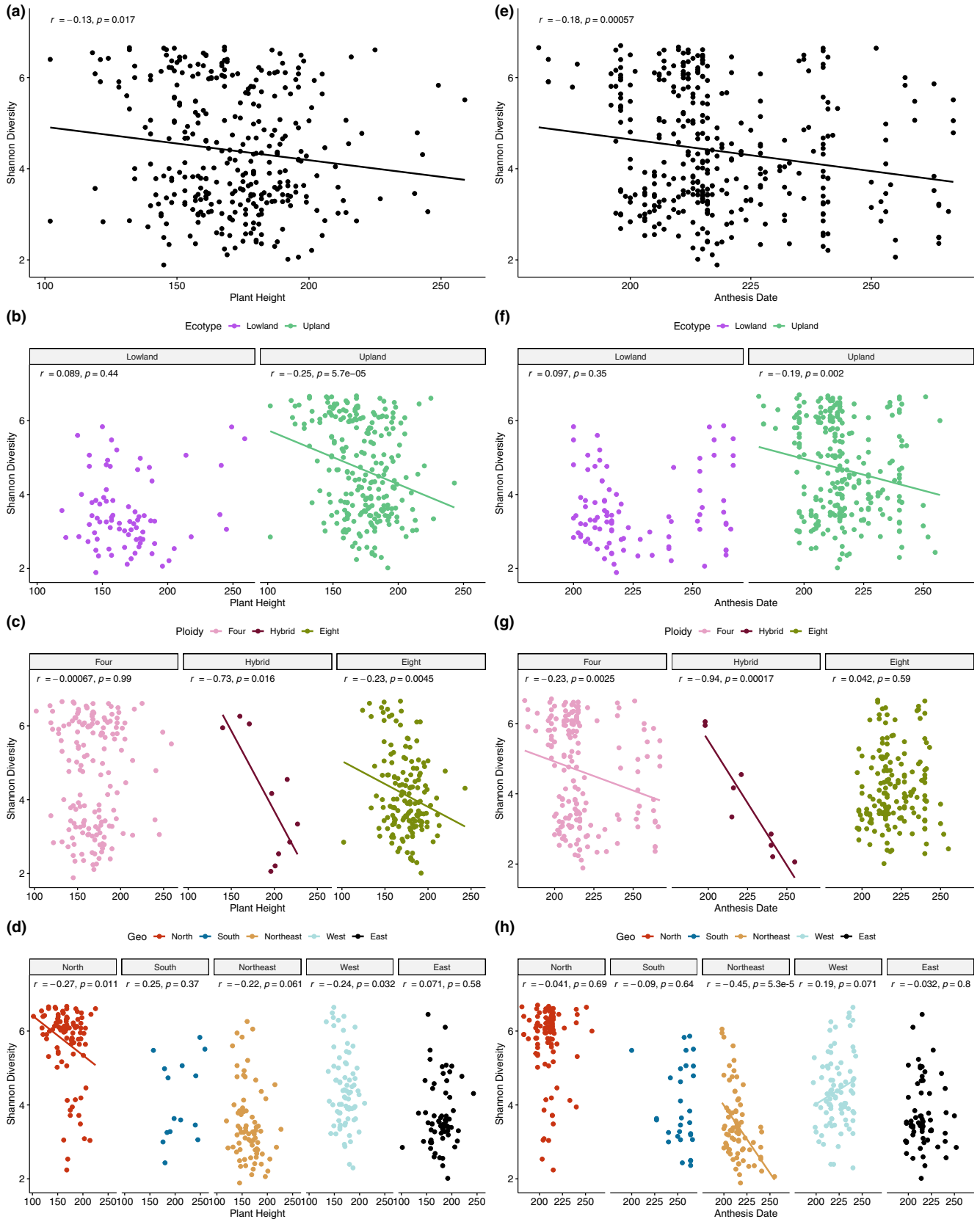


FIGURE 3 Relationships between Shannon diversity and plant height. (a) Genotype: plotted for all switchgrass genotypes. (b) Ecotype: plotted for switchgrass ecotypes with Lowland (purple), Both (brown), and Upland (green). (c) Ploidy: plotted for switchgrass ploidy levels with Tetraploid - Four (pink), Hybrid (maroon), and Octoploid - Eight (olive). (d) Genetic cluster: plotted for switchgrass genetic clusters with North (red), South (dark blue), Northeast (beige), West (light blue), and East (black). Relationships between Shannon diversity and anthesis date. (d) Genotype: plotted for all switchgrass genotypes. (e) Ecotype: plotted for switchgrass ecotypes with Lowland (purple), Both (brown), and Upland (green). (f) Ploidy: plotted for switchgrass ploidy levels with Tetraploid - Four (pink), Hybrid (maroon), and Octoploid - Eight (olive). (g) Genetic cluster: plotted for switchgrass genetic clusters with North (red), South (dark blue), Northeast (beige), West (light blue), and East (black)

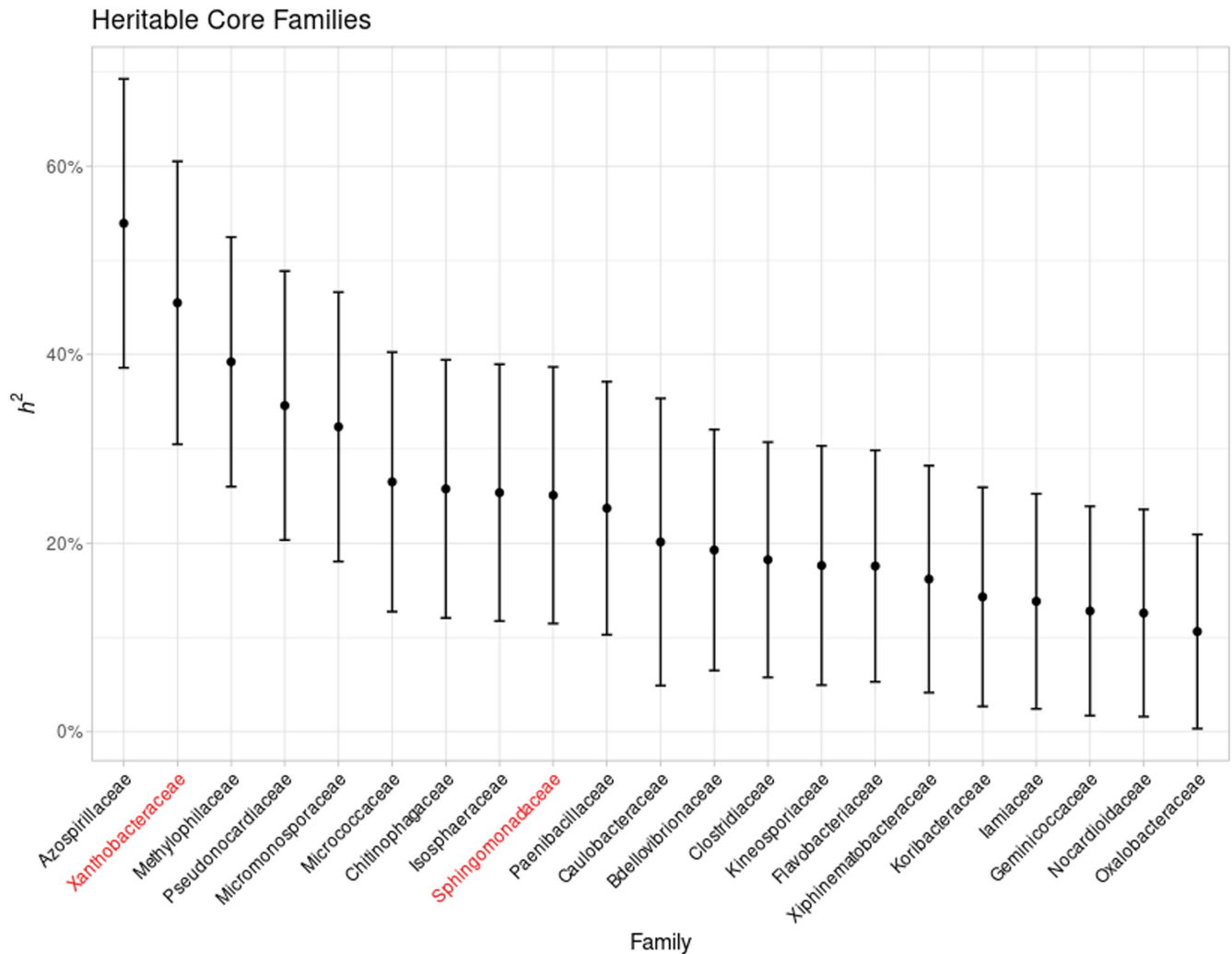


FIGURE 4 Heritability estimates for bacterial families in the switchgrass rhizosphere. For 110 bacterial families that were observed in more than 10% of switchgrass samples, narrow-sense heritability (h^2) was calculated using a mixed linear model. 21 families had heritability estimates with standard error that did not include zero. Families observed in more than 90% of our samples are represented in red

and Spingomonadaceae. The low-fidelity core microbiome (>10% of samples) contained 110 families (Table S2).

3.4 | Host genomic influence

Of the 110 families in the low-fidelity core microbiome, twenty-one showed narrow-sense SNP-based heritability (h^2) significantly greater than zero (Figure 4). The standard errors of narrow-sense

heritability estimates are very large and there is considerable uncertainty in the estimate overall (Furlotte et al., 2014). Therefore, we consider estimates with standard errors that do not intersect zero as showing evidence of heritability (i.e., heritable). Heritability estimates for low-fidelity core taxa in the switchgrass rhizosphere ranged from 0.106 to 0.539, with a mean of 0.241. Heritability estimates presented here for bacterial lineages have a range similar to that of corn and sorghum (Deng et al., 2021). The heritability estimates indicate that, in part, switchgrass genotype influences the

variability in the relative abundance of certain families within our growing site. Heritability estimates for all bacterial families can be found in Table S2.

Next, we performed a redundancy analysis (RDA) to characterize how the whole rhizosphere bacterial assemblage changes in association with host genomic background (Figure 5, Figure S4). The RDA results indicate that genotypic variability among tetraploids (i.e., first 10 PCs of SNPs) had modest influences on bacterial ASV composition ($R^2 = 0.063$, $R^2_{\text{adjusted}} = 0.007$). Switchgrass SNP PC1, which is associated with separation between the North and the Northeast and South genetic clusters, loaded most strongly to the first primary canonical axis (RDA1, PC1 = 0.99, the next strongest loaded PC was PC10, with a 0.097 loading). Switchgrass SNP PC7, which is associated the variance in the North genetic cluster, loaded most strongly to the second canonical axis (RDA2, PC7 = 0.899). The proportion of explained variance in bacterial ASV composition was fairly low along these individual canonical axes (RDA1 = 0.9%, RDA2 = 0.7%), which was unsurprising given the large diversity in the bacterial composition. Still, the RDA

identified bacterial ASV compositional turnover in association with switchgrass population structure, as shown by the turnover between the North and the Northeast and South switchgrass genetic clusters (Figure 5).

Outlier ASVs were determined for each of the first two canonical RDA axes ($n = 9215$). These outlier ASVs are most representative of the variation in bacterial composition among the tetraploid genotypes. We then identified the taxonomic ranks for the outlier ASVs. Our results show that Xanthobacteraceae and Sphingomonadaceae represent the high-fidelity core for both the outlier ASV data set and the original ASV data set, signifying host genotypic influence might be strongest on these families. When we consider the low-fidelity core ($n = 110$ families), 70.9% of families in the RDA outlier ASV core microbiome are in original core microbiome (Table S2). When we consider just the heritable families, 80.95% of heritable families in the outlier ASV core microbiome are in the original core microbiome (Table S2).

When we applied variance partitioning to estimate the portion of ASV-level compositional turnover explained by SNP PCs, ecotype, and

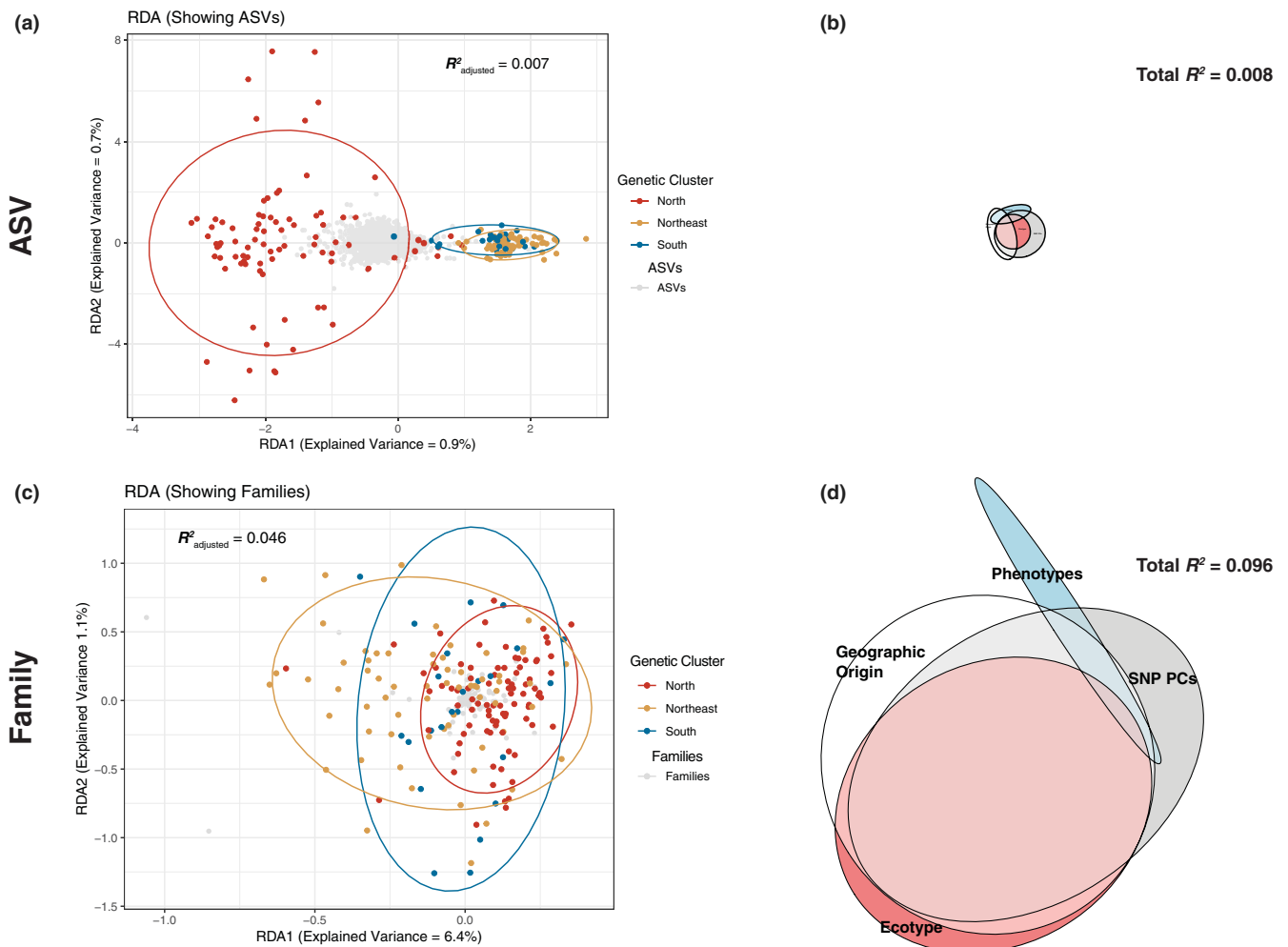


FIGURE 5 Redundancy analyses and variance partitioning. Results of RDA analysis showing the North (red), Northeast (beige), and South (blue) host genetic clusters with corresponding amplicon sequence variants (grey). (a) and bacterial families (b). Results of variance partitioning analysis showing Ecotype (red), SNP PCs (grey), Phenotypes (blue), and Geographic Origin (white) for ASVs (c) and bacterial families (d)

geographic origin (i.e., latitude and longitude coordinates), and the two phenotypes, we found that the total R^2_{adjusted} was low, approximately 1% of ASV composition. However, variance partitioning of family-level turnover found these four factors explained a larger portion, with $R^2_{\text{adjusted}} = 0.096$ (Figure 5). Variance partitioning at the family level found that the largest portion of bacterial compositional variation was explained by the collinear portion of SNP PCs, ecotype, and geographic origin ($R^2_{\text{adjusted}} = 0.0866$), suggesting that geographic population structure associated with ecotypic variation explains a large portion of turnover in bacterial family composition. At the family level, however, we lose the separation among genetic clusters in the RDA. This suggests that amplicon sequence level diversity between groups is higher than at the family level. This line of reasoning is further supported by the ASV based alpha diversity analysis where the distinction of the North genetic cluster is characterized by many unique ASVs (Figure 2).

3.5 | Genome-wide association study

A primary limitation of heritability estimates and the RDA was that we could not account for which SNPs have the greatest effect on compositional variation. To overcome this, a genome-wide association study (GWAS) was conducted to identify specific loci associated with the relative abundance of families in the core microbiome. In total, 1861 SNPs were found to be associated with the abundance of the 110 tested core families. 878 SNPs were found to be associated with the abundance of the 21 heritable core families. Furthermore, 36 SNPs were found to be associated with the abundance of the two families (Xanthobacteraceae and Sphingomonadaceae) that represent the high-fidelity core microbiome (Figure 6, Figure S5, Table S2).

3.6 | GO enrichment analysis

Despite the general difficulty identifying specific molecular functions or pathways in non-model plant species, resources for gene ontology (GO) still remain useful for biological inference (Tian et al., 2017). Therefore, we completed a GO enrichment analysis for genes harbouring significant SNPs and positions within 25 kb of significant SNPs to determine possible cellular processes (P), cellular components (C), and molecular functions (F) associated with three core family abundances (Table S3). Primarily, we identified two types of cellular processes: metabolic and stimuli response processes. Broadly, we identified several molecular functions relating to ATPase activity, hydrolase activity, and pyrophosphatase activity (Figure S6, Table S3). Lastly, genes harbouring SNPs associated with Sphingomonadaceae abundance, in particular, were enriched for GO terms associated with cellular membrane components, metallopeptidase and metalloendopeptidase activity, and cellular processes that relate to metabolism and response to stimuli (Table S3).

4 | DISCUSSION

Macroscopic organisms often exhibit striking variation among individuals in the microbial assemblages they host, but the genetic basis and ecological relevance of this variation is often unclear. Common garden experiments with diverse genotypes offer a powerful window into dissecting the ecological genomics of host influences on their microbiomes. Previous studies have demonstrated that genomic differences among switchgrass genotypes and populations may, in part, have led to much of the adaptive phenotypic differences observed between switchgrass populations (Adkins et al., 2016; Lowry et al., 2019). We add to these observations by showing a similar phenomenon within a less visible portion of the switchgrass phenotype. We found that switchgrass genotype influences the rhizosphere microbiome in a variety of ways relative to bacterial alpha and beta diversity as a result of variability within the switchgrass genome.

We demonstrated that the alpha diversity (Chao1 and Shannon diversity) of rhizosphere bacteria within a common garden was significantly differentiated between switchgrass ecotypes, ploidy levels, and genetic clusters. These groupings with distinct bacterial assemblages are associated with life history variation and locally-adapted traits in this species. Furthermore, alpha diversity within each group typically had no correlation with, or was negatively correlated with, anthesis date and plant height (Figure 3). Flowering time is an important adaptation to local growing conditions and has major fitness consequences for the plant (Grabowski et al., 2017). In general, in common garden experiments, upland ecotypes have an anthesis date 2–4 weeks earlier than lowland ecotypes (Schwartz & Amasino, 2013). As plants move through different developmental stages the diversity and availability of metabolites influencing the rhizosphere microbiome shifts (Wagner, 2021). Rhizosphere sampling in this study occurred in May 2016. Therefore, it is reasonable to expect genotypes at different stages of development at the time of sampling to exert different selective pressure on the microbiome composition. Future research examining temporal exudate availability in the rhizosphere of switchgrass could expand on this hypothesis.

It is already widely hypothesized that hosts mediate their microbiomes relative to environmental conditions, resource availability, and microbial function, not necessarily always a specific species or strain (Lemanceau et al., 2017; Martiny et al., 2015). Given the disparity in the speed of adaptation (due to differences in generation time), it is unlikely that switchgrass has mechanisms to distinguish between closely related bacterial taxa unless doing so provided a substantial fitness advantage (Yin et al., 2021). Thus, some basis for comparing taxa between host groupings becomes important when determining the taxonomic level at which host influence is detectable. A "core" microbiome can be defined in a variety of ways, and may not represent microbes that have the most important functional impacts on a system (Bell & Bell, 2021). For example, commensal microorganisms remain widespread throughout host populations (Zeng et al., 2015). Still, the consistent occurrence of

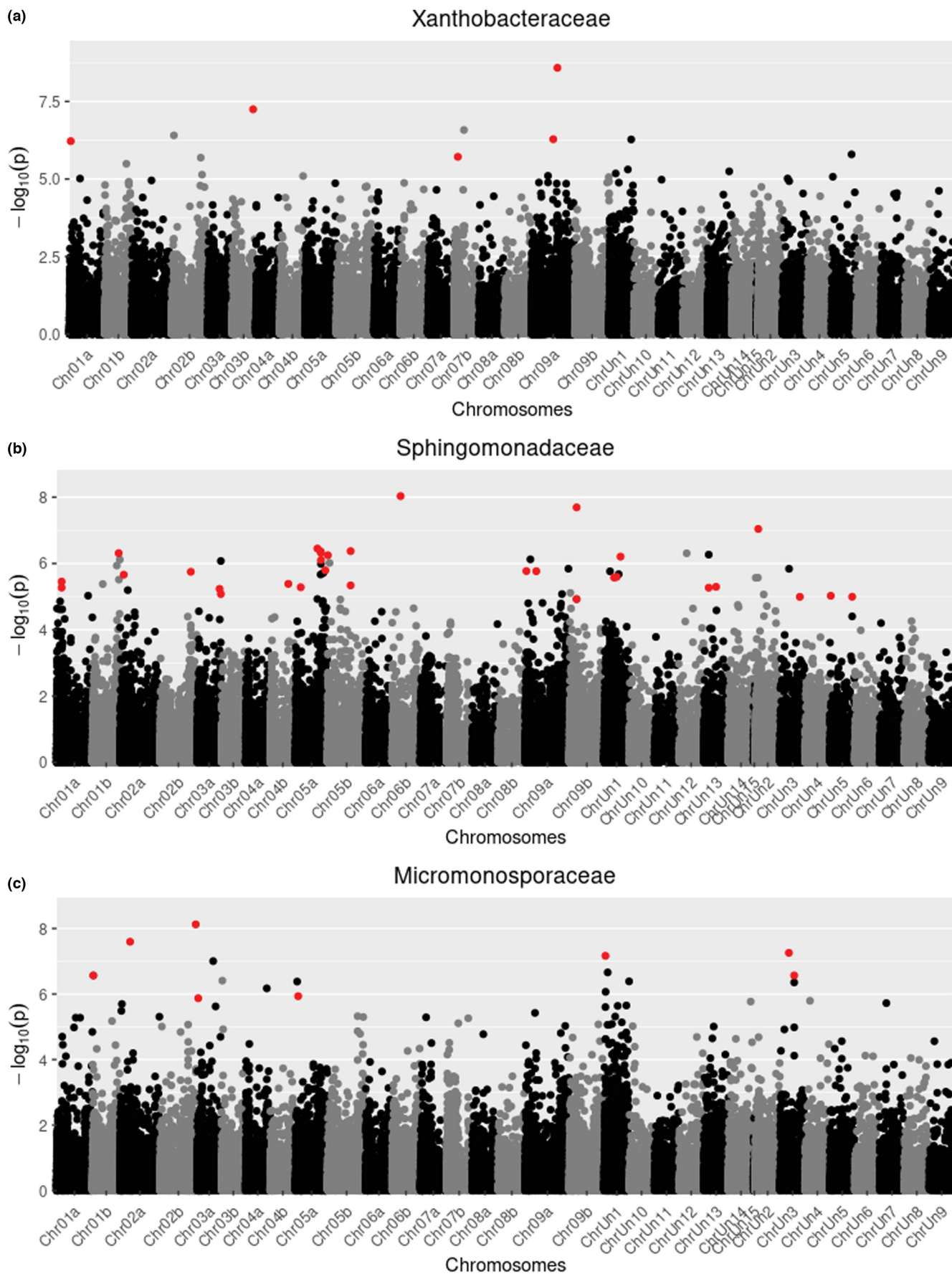


FIGURE 6 Genome-wide association study (GWAS). Manhattan plot presenting significant SNPs (red) associated with Xanthobacteraceae rhizosphere abundance (a), Sphingomonadaceae rhizosphere abundance (b), and micromonosporaceae rhizosphere abundance (c)

microbial taxa provides an opportunity to assess relationships between host genetics and microbial composition. Here, we applied the “common core” concept (Risely, 2020) to identify bacterial taxa that were consistently associated with diverse switchgrass genotypes in a common garden. Our perspective for emphasizing prevalence was that (1) we did not know which microbes would benefit or harm the host, and (2) that we did not know whether the relative abundance of a microbe would reflect its importance to the host. For instance, rare microbes within the context of the microbiome can have large impacts on plant biomass and nutritional quality (Hol et al., 2010). As such, the core microbiome provides both a basis for measuring bacterial prevalence within the common garden as well as a means to measure host genomic influence at the lowest possible taxonomic rank.

For example, Singer, Bonnette, Woyke, et al. (2019) observed a large core endophytic switchgrass microbiome dominated by root-colonizing bacterial genera such as *Streptomyces*, *Pseudomonas*, and *Bradyrhizobium*, while rhizosphere diversity was more variable between their two sampling sites. Likewise, Singer, Bonnette, Woyke, et al. (2019) observed different core bacterial classes associated with upland and lowland switchgrass ecotypes, with each ecotype preferentially enriched for Alphaproteobacteria and Actinobacteria, respectively. For our study, out of 268 families, only two associated with the majority ($\geq 90\%$) of our rhizosphere samples (Xanthobacteraceae and Sphingomonadaceae) and less than half ($n = 110$) associated with more than 10% of our samples. Xanthobacteraceae and Sphingomonadaceae are families in the Alphaproteobacteria class. We were unable to detect any genera or species associated with our genotypes at the 90% occupancy threshold. This indicates that a large swath of the switchgrass rhizosphere microbiome assembly is stochastic and only certain lineages were able to consistently associate with our genotypes in the common garden.

The abundance of 21 of the 110 families were determined to show evidence of heritability. After reasonably establishing heritability of select core families, the RDA allowed us to identify how genome-wide variation in switchgrass influenced the rhizosphere microbiome composition. RDA methods used in microbial ecology typically explain the variance in microbiome composition due to environmental variables, such as temperature or pH (Paliy & Shankar, 2016). However, here, we modelled microbial compositional data as a function of host genomic data. As such, the RDA in this study represents the compositional variance of the microbiome due to genetic variation in the hosts. We then extended the analysis to explore outlier ASVs along each of the first two RDA axes, because those ASVs represent taxa associated with host genomic influence on overall turnover in bacterial composition. Once clustered at the family level, we determined that seventeen out of the 21 heritable families were in the outlier ASV data set. Additionally, the results of the variance partitioning suggest that geographic population structure associated with ecotypic variation explains a larger portion of turnover in bacterial family composition compared to ASV composition. Thus, we determined that genotypic variability in switchgrass

sampled from across its natural range influences the composition of the rhizosphere bacterial microbiome, particularly for tetraploids.

The ecophysiology of host–microbe interactions in our study is unclear, requiring detailed investigation, but there may be hints in the function of switchgrass genes that are strongly associated with core family abundances. We conducted a GWAS to identify potential loci that strongly associate with core family abundances. The outcrossing nature of switchgrass and its rapid decay in LD can help pinpoint causal loci with GWAS. Several considerations had to be made though, in order to reliably trust the significance of the resulting QTLs. First, we considered only tetraploids for the same reasons as the RDA, that they (1) represent both upland and lowland ecotypes, (2) exhibit disomic inheritance unlike octoploids, increasing the effect size of additive effects, and (3) exhibit the most intergroup alpha diversity variation compared to octoploids or hybrids. We used exome capture SNPs (Evans et al. (2018), which are more effective in their ability to tag causal polymorphisms and parts of adjacent nonexonic DNA, compared to GBS SNPs that are often far from any genes (Kaur et al., 2017). The identification of genes containing significant SNPs could provide insights into potential host–microbe interaction pathways. To explore this relationship further, we focused on three core families: Xanthobacteraceae, Sphingomonadaceae, and Micromonosporaceae.

We identified hundreds of switchgrass genes containing SNPs (including those within 25kb of significant SNPs) associated with Micromonosporaceae ($n = 315$), Xanthobacteraceae ($n = 578$), and Sphingomonadaceae ($n = 718$) abundance in the rhizosphere. Many species within the Micromonosporaceae family degrade chitin, cellulose, lignin, and pectin, and play an important role in the turnover of organic plant material. Moreover, many strains produce useful secondary metabolites and enzymes used by plants (Trujillo et al., 2014). Members of Xanthobacteraceae have been experimentally identified as nitrate reducers (Anderson, Condron, et al., 2011, Anderson, Willis, Mitchell-Olds, 2011) and therefore their regulation could be important to host metabolic activity. Our GO enrichment analysis indicates that genes associated with Micromonosporaceae and Xanthobacteraceae abundance are involved in hydrolase, pyrophosphatase, and ATPase activity suggesting that the enrichment for this set of genes plays an important role in overall plant fitness. Functions involving ATP hydrolysis as an energy source either typically catalyse a reaction or drive membrane transport against a concentration gradient (Berg et al., 2002). In the context of our GWAS, this could suggest possible root exudate activity whereby switchgrass genotypes differentially transport exudates against a lower soil compound gradient influencing core family abundances. Perturbing these regions in future studies could reveal the underlying pathways and mechanisms. Members of the Sphingomonadaceae family are known to be antagonistic against plant pathogens and induce plant growth promotion (Glaeser & Kämpfer, 2014). GO terms for genes associated with this family are enriched for metallopeptidase and metalloendopeptidase activity, membrane cellular components, and processes involved in stimuli response. Altogether, this suggests that the activity of these genes is linked to environmental stimuli,

possibly initiated by Sphingomonadaceae, and interacting with plant cell membranes. Again, given the context of our GWAS, we suspect that the genes associated with these GO terms could interact with Sphingomonadaceae populations in the soil. Future work examining these families could make these connections to plant function more robust.

In conclusion, bacterial composition in the switchgrass rhizosphere is influenced by host genotypic variability. Rhizosphere diversity shows major differences among switchgrass ecotypes, ploidy levels, genetic clusters, and life-history axes. In particular, the North switchgrass genetic cluster was most distinct at our common garden site. Two high-fidelity core families, Xanthobacteraceae and Sphingomonadaceae, were observed across the majority of our switchgrass genotypes. Twenty-one bacterial families were found to exhibit evidence for heritability. Though the explained variance was low, a redundancy analysis revealed widespread genotypic influence on the composition of the rhizosphere bacterial microbiome and outlier ASVs along each primary axis belonged largely to our defined core microbiome. Finally, we identified 1861 SNPs associated with the abundance of core microbiome families and discussed potential pathway mechanisms that could influence a subset of those families that we expected could influence host fitness. A GO enrichment analysis determined several functional and cellular processes for which host genotypic influence could be expressed, notably by hydrolase activity. This study should function as a foundation on which further exploration can be achieved investigating the relationship between switchgrass genotype and its associated rhizosphere microbiome in different environmental contexts or under artificial genetic perturbation.

AUTHOR CONTRIBUTIONS

All authors designed the research. R.T. provided technical lab assistance. J.S. performed research, analyzed data, and wrote the paper. All authors edited multiple drafts.

ACKNOWLEDGEMENTS

We thank Wanyan Wang for assistance in collecting and cataloguing the soil samples from the GWAS panel and selecting samples for metagenome analysis. We thank Maureen Mailander for conducting the initial rhizosphere soil DNA extractions. We are also grateful to our entire project team (investigators Marvin Hall, Stacy Bonos, Julie Hanson, Don Viands, and their students and staff for the collection of the GWAS family plants from the common garden site at Ithaca, NY). We greatly appreciate the opportunity provided by Dr Michael Casler for allowing us to collect samples of all of the genotypes in this range-wide collection of switchgrass genotypes and check cultivars at of Cornell University for this project. This research was supported by the National Institute of Food and Agriculture, U.S. Department of Agriculture, under award no. 2019-67009-29006 to JEC, TBH, and JRL. Seed support was provided by the Northeast Woody/warm-season BIOmass consortium, funded by USDA-AFRI grant no. 2012-68005-19703. Programme support to JEC was provided through the USDA National Institute of Food and

Agriculture Federal Appropriations under Project PEN04532 and Accession no. 1000326. JRL was also supported by NIH award 1R35GM138300-01.

CONFLICTS OF INTEREST

The authors have no conflicts of interest to declare.

DATA AVAILABILITY STATEMENT

Raw amplicon sequence reads are deposited in the NCBI SRA (BioProject PRJNA689762). Switchgrass exome capture reads are deposited in the NCBI SRA (BioProject PRJNA280418). Rscripts and python scripts are available at https://github.com/jeremysutherland/host_genomic_influence_on_bacterial_composition_in_the_switchgrass_rhizosphere

OPEN RESEARCH BADGES



This article has earned an Open Data Badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at https://github.com/jeremysutherland/host_genomic_influence_on_bacterial_composition_in_the_switchgrass_rhizosphere

ORCID

Jeremy Sutherland <https://orcid.org/0000-0003-3535-0302>

Terrence Bell <https://orcid.org/0000-0003-3603-7270>

Ryan V. Trexler <https://orcid.org/0000-0002-8321-6127>

Jesse R. Lasky <https://orcid.org/0000-0001-7688-5296>

REFERENCES

- Adkins, J., Jastrow, J. D., Morris, G. P., Six, J., & de Graaff, M. A. (2016). Effects of switchgrass cultivars and intraspecific differences in root structure on soil carbon inputs and accumulation. *Geoderma*, 262, 147–154. <https://doi.org/10.1016/j.geoderma.2015.08.019>
- Anderson, C. R., Condrón, L. M., Clough, T. J., Fiers, M., Stewart, A., Hill, R. A., & Sherlock, R. R. (2011). Biochar induced soil microbial community change: Implications for biogeochemical cycling of carbon, nitrogen and phosphorus. *Pedobiologia*, 54(5–6), 309–320. <https://doi.org/10.1016/j.pedobi.2011.07.005>
- Anderson, J. T., Willis, J. H., & Mitchell-Olds, T. (2011). Evolutionary genetics of plant adaptation. *Trends in Genetics*, 27(7), 258–266. <https://doi.org/10.1016/J.TIG.2011.04.001>
- Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, 26(1), 32–46. <https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x>
- Apprill, A., McNally, S., Parsons, R., & Weber, L. (2015). Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquatic Microbial Ecology*, 75, 129–137. <https://doi.org/10.3354/ame01753>
- Baetz, U., & Martinoia, E. (2014). Root exudates: The hidden part of plant defense. *Trends in Plant Science*, 19, 90–98. <https://doi.org/10.1016/j.tplants.2013.11.006>
- Beilsmith, K., Thoen, M. P. M., Brachi, B., Gloss, A. D., Khan, M. H., & Bergelson, J. (2019). Genome-wide association studies on the phyllosphere microbiome: Embracing complexity in host-microbe interactions. *The Plant Journal*, 97(1), 164–181. <https://doi.org/10.1111/tbj.14170>

- Bell, T. H., & Bell, T. (2021). Many roads to bacterial generalism. *FEMS Microbiology Ecology*, 97. <https://doi.org/10.1093/femsec/fiaa240>
- Berg, J. M., Tymoczko, J. L., & Stryer, L. (2002). A family of membrane proteins uses ATP hydrolysis to pump ions across membranes. <https://www.ncbi.nlm.nih.gov/books/NBK22464/>
- Bever, J. D., Platt, T. G., & Morton, E. R. (2012). Microbial population and community dynamics on plant roots and their feedbacks on plant communities. *Annual Review of Microbiology*, 66, 265–283. <https://doi.org/10.1146/annurev-micro-092611-150107>
- Bordenstein, S. R., & Theis, K. R. (2015). Host biology in light of the microbiome: Ten principles of holobionts and hologenomes. *PLoS Biology*, 13(8), e1002226. <https://doi.org/10.1371/journal.pbio.1002226>
- Bouton, J. H. (2007). Molecular breeding of switchgrass for use as a biofuel crop. *Current Opinion in Genetics and Development*, 17, 553–558. <https://doi.org/10.1016/j.gde.2007.08.012>
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23(19), 2633–2635. <https://doi.org/10.1093/BIOINFORMATICS/BTM308>
- Brown, B. C., Bray, N. L., & Pachter, L. (2018). Expression reflects population structure. *PLoS Genetics*, 14(12), e1007841. <https://doi.org/10.1371/journal.pgen.1007841>
- Brzyski, D., Peterson, C. B., Sobczyk, P., Candès, E. J., Bogdan, M., & Sabatti, C. (2017). Controlling the rate of GWAS false discoveries. *Genetics*, 205, 61–75. <https://doi.org/10.1534/genetics.116.193987>
- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High resolution sample inference from Illumina amplicon data. *Nature Methods*, 13, 581–583. <https://doi.org/10.1038/nmeth.3869>
- Carvalhais, L. C., Dennis, P. G., Fan, B., Fedoseyenko, D., Kierul, K., Becker, A., von Wiren, N., & Borriss, R. (2013). Linking plant nutritional status to plant-microbe interactions. *PLoS One*, 8(7), e68555. <https://doi.org/10.1371/journal.pone.0068555>
- Casler, M. D., Vogel, K. P., & Harrison, M. (2015). Switchgrass germplasm resources. *Crop Science*, 55(6), 2463–2478. <https://doi.org/10.2135/cropsci2015.02.0076>
- Compant, S., Clément, C., & Sessitsch, A. (2010). Plant growth-promoting bacteria in the rhizo- and endosphere of plants: Their role, colonization, mechanisms involved and prospects for utilization. *Soil Biology and Biochemistry*, 42, 669–678. <https://doi.org/10.1016/j.soilbio.2009.11.024>
- Costich, D. E., Friebe, B., Sheehan, M. J., Casler, M. D., & Buckler, E. S. (2010). Genome-size Variation in Switchgrass (*Panicum virgatum*): Flow Cytometry and Cytology Reveal Rampant Aneuploidy. *The Plant Genome*, 3(3), 130–141. <https://doi.org/10.3835/plantgenom2010.04.0010>
- Covarrubias-Pazarán, G. (2016). Genome-Assisted prediction of quantitative traits using the r package sommer. *PLoS One*, 11, e0156744. <https://doi.org/10.1371/journal.pone.0156744>
- Das, M. K., Fuentes, R. G., & Taliaferro, C. M. (2004). Genetic variability and trait relationships in switchgrass. *Crop Science*, 44(2), 443–448. <https://doi.org/10.2135/cropsci2004.4430>
- Deng, S., Caddell, D. F., Xu, G., Dahlen, L., Washington, L., Yang, J., & Coleman-Derr, D. (2021). Genome wide association study reveals plant loci controlling heritability of the rhizosphere microbiome. *ISME Journal*, 15(11), 3181–3194. <https://doi.org/10.1038/s41396-021-00993-z>
- Dennis, P. G., Miller, A. J., & Hirsch, P. R. (2010). Are root exudates more important than other sources of rhizodeposits in structuring rhizosphere bacterial communities? *FEMS Microbiology Ecology*, 72, 313–327. <https://doi.org/10.1111/j.1574-6941.2010.00860.x>
- Du, Z., Zhou, X., Ling, Y., Zhang, Z., & Su, Z. (2010). agriGO: A GO analysis toolkit for the agricultural community. *Nucleic Acids Research*, 38, W64–W70. <https://doi.org/10.1093/nar/gkq310>
- Evans, J., Sanciangco, M. D., Lau, K. H., Crisovan, E., Barry, K., Daum, C., Hundley, H., Jenkins, J., Kennedy, M., Kunde-Ramamoorthy, G., Vaillancourt, B., Acharya, A., Schmutz, J., Saha, M., Kaeppler, S. M., Brummer, E. C., Casler, M. D., & Buell, C. R. (2018). Extensive genetic diversity is present within north American switchgrass germplasm. *The Plant Genome*, 11(1), 1–16. <https://doi.org/10.3835/plantgenome2017.06.0055>
- Furlotte, N. A., Heckerman, D., & Lippert, C. (2014). Quantifying the uncertainty in heritability. *Journal of Human Genetics*, 59, 269–275. <https://doi.org/10.1038/jhg.2014.15>
- Glaeser, S. P., & Kämpfer, P. (2014). The family Sphingomonadaceae. In *The prokaryotes: alphaproteobacteria and betaproteobacteria* (Vol. 9783642301971, pp. 641–707). Springer. https://doi.org/10.1007/978-3-642-30197-1_302
- Gould, A. L., Zhang, V., Lamberti, L., Jones, E. W., Obadia, B., Korasidis, N., Gavryushkin, A., Carlson, J. M., Beerenwinkel, N., & Ludington, W. B. (2018). Microbiome interactions shape host fitness. *Proceedings of the National Academy of Sciences of the United States of America*, 115(51), E11951–E11960. <https://doi.org/10.1073/pnas.1809349115>
- Grabowski, P. P., Evans, J., Daum, C., Deshpande, S., Barry, K. W., Kennedy, M., Ramstein, G., Kaeppler, S. M., Buell, C. R., Jiang, Y., & Casler, M. D. (2017). Genome-wide associations with flowering time in switchgrass using exome-capture sequencing data. *New Phytologist*, 213(1), 154–169. <https://doi.org/10.1111/nph.14101>
- Hacquard, S., Spaepen, S., Garrido-Oter, R., & Schulze-Lefert, P. (2017). Interplay between Innate immunity and the plant microbiota. *Annual Review of Phytopathology*, 55, 565–589. <https://doi.org/10.1146/annurev-phyto-080516-035623>
- Heijboer, A., de Ruiter, P. C., Bodelier, P. L. E., & Kowalchuk, G. A. (2018). Modulation of litter decomposition by the soil microbial food web under influence of land use change. *Frontiers in Microbiology*, 9, 2860. <https://doi.org/10.3389/fmicb.2018.02860>
- Hestrin, R., Lee, M. R., Whitaker, B. K., & Pett-Ridge, J. (2021). The switchgrass microbiome: A review of structure, function, and taxonomic distribution. *Phytobiomes Journal*, 5, 14–28. <https://doi.org/10.1094/pbiomes-04-20-0029-fi>
- Hol, W. H. G., de Boer, W., Termorshuizen, A. J., Meyer, K. M., Schneider, J. H. M., van Dam, N. M., van Veen, J., & van der Putten, W. (2010). Reduction of rare soil microbes modifies plant-herbivore interactions. *Ecology Letters*, 13(3), 292–301. <https://doi.org/10.1111/J.1461-0248.2009.01424.X>
- Holland, J. B. (2007). Genetic architecture of complex traits in plants. *Current Opinion in Plant Biology*, 10, 156–161. <https://doi.org/10.1016/j.pbi.2007.01.003>
- Hu, L., Robert, C. A. M., Cadot, S., Zhang, X., Ye, M., Li, B., Manzo, D., Chervet, N., Steinger, T., MGA, v. d. H., Schlaeppli, K., & Erb, M. (2018). Root exudate metabolites drive plant-soil feedbacks on growth and defense by shaping the rhizosphere microbiota. *Nature Communications*, 9(1), 1–13. <https://doi.org/10.1038/s41467-018-05122-7>
- Hunter, P. (2018). The revival of the extended phenotype. *EMBO Reports*, 19(7), 1–4. <https://doi.org/10.15252/embr.201846477>
- Ihaka, R., & Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5(3), 299–314. <https://doi.org/10.1080/10618600.1996.10474713>
- Jones, P., García, B. J., Furches, A., Tuskan, G. A., & Jacobson, D. (2019). Plant host-associated mechanisms for microbial selection. *Frontiers in Plant Science*, 10, 862. <https://doi.org/10.3389/fpls.2019.00862>
- Kang, H. M., Sul, J. H., Service, S. K., Zaitlen, N. A., Kong, S. Y., Freimer, N. B., Sabatti, C., & Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics*, 42(4), 348–354. <https://doi.org/10.1038/ng.548>
- Kaur, P., & Gaikwad, K. (2017). From genomes to gene-omes: Exome sequencing concept and applications in crop improvement. *Frontiers in Plant Science*, 8, 2164. <https://doi.org/10.3389/fpls.2017.02164>

- Lahti, L., Shetty, S., & Blake, T. (2017). Tools for microbiome analysis in R. *Microbiome Package Version 0.99*.
- Lakshmanan, V., Kitto, S. L., Caplan, J. L., Hsueh, Y. H., Kearns, D. B., Wu, Y. S., & Bais, H. P. (2012). Microbe-associated molecular patterns-triggered root responses mediate beneficial rhizobacterial recruitment in arabidopsis. *Plant Physiology*, *160*(3), 1642–1661. <https://doi.org/10.1104/pp.112.200386>
- 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., & Dangl, J. L. (2015). Salicylic acid modulates colonization of the root microbiome by specific bacterial taxa. *Science*, *349*(6250), 860–864. <https://doi.org/10.1126/science.aaa8764>
- Lemanceau, P., Blouin, M., Muller, D., & Moënne-Loccoz, Y. (2017). Let the core microbiota be functional. *Trends in Plant Science*, *22*, 583–595. <https://doi.org/10.1016/j.tplants.2017.04.008>
- Lipka, A. E., Lu, F., Cherney, J. H., Buckler, E. S., Casler, M. D., & Costich, D. E. (2014). Accelerating the switchgrass (*panicum virgatum* L.) breeding cycle using genomic selection approaches. *PLoS One*, *9*(11), e112227. <https://doi.org/10.1371/journal.pone.0112227>
- Lovell, J. T., MacQueen, A. H., Mamidi, S., Bonnette, J., Jenkins, J., Napier, J. D., Sreedasyam, A., Healey, A., Session, A., Shu, S., Barry, K., Bonos, S., Boston, L., Daum, C., Deshpande, S., Ewing, A., Grabowski, P. P., Haque, T., Harrison, M., ... Schmutz, J. (2021). Genomic mechanisms of climate adaptation in polyploid bioenergy switchgrass. *Nature*, *590*(7846), 438–444. <https://doi.org/10.1038/s41586-020-03127-1>
- Lowry, D. B., Behrman, K. D., Grabowski, P., Morris, G. P., Kiniry, J. R., & Juenger, T. E. (2015). Adaptations between ecotypes and along environmental gradients in *panicum virgatum*. *The American Naturalist*, *183*(5), 682–692. <https://doi.org/10.1086/675760>
- Lowry, D. B., Lovell, J. T., Zhang, L., Bonnette, J., Fay, P. A., Mitchell, R. B., Lloyd-Reilley, J., Boe, A. R., Wu, Y., Rouquette, F. M., Jr., Wynia, R. L., Weng, X., Behrman, K. D., Healey, A., Barry, K., Lipzen, A., Bauer, D., Sharma, A., Jenkins, J., ... Juenger, T. E. (2019). QTL × environment interactions underlie adaptive divergence in switchgrass across a large latitudinal gradient. *Proceedings of the National Academy of Sciences of the United States of America*, *116*, 12933–12941. <https://doi.org/10.1073/pnas.1821543116>
- Lu, F., Lipka, A. E., Glaubitz, J., Elshire, R., Cherney, J. H., Casler, M. D., Buckler, E. S., & Costich, D. E. (2013). Switchgrass genomic diversity, ploidy, and evolution: Novel Insights from a network-based SNP discovery protocol. *PLoS Genetics*, *9*, e1003215. <https://doi.org/10.1371/journal.pgen.1003215>
- Martiny, J. B. H., Jones, S. E., Lennon, J. T., & Martiny, A. C. (2015). Microbiomes in light of traits: A phylogenetic perspective. *350 Science*, *350*. <https://doi.org/10.1126/science.aac9323>
- McLaughlin, S. B., & Kszos, L. A. (2005). Development of switchgrass (*Panicum virgatum*) as a bioenergy feedstock in the United States. *Biomass and Bioenergy*, *28*(6), 515–535. <https://doi.org/10.1016/j.biombioe.2004.05.006>
- McMurdie, P. J., & Holmes, S. (2013). phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*, *8*(4), e61217. <https://doi.org/10.1371/journal.pone.0061217>
- Moran, N. A., & Sloan, D. B. (2015). The hologenome concept: Helpful or hollow? *PLoS Biology*, *13*(12), e1002311. <https://doi.org/10.1371/journal.pbio.1002311>
- Oksanen, J., Simpson, G., Blanchet, F., Kindt, R., Legendre, P., Minchin, P., O'Hara, R., Solymos, P., Stevens, M., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., De Caceres, M., Durand, S., ... Weedon, J. (2022). vegan: Community Ecology Package. R package version 2.5-5. <https://CRAN.R-project.org/package=vegan>
- Paliy, O., & Shankar, V. (2016). Application of multivariate statistical techniques in microbial ecology. *Molecular Ecology*, *25*, 1032–1057. <https://doi.org/10.1111/mec.13536>
- Parada, A. E., Needham, D. M., & Fuhrman, J. A. (2016). Every base matters: Assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environmental Microbiology*, *18*, 1403–1414. <https://doi.org/10.1111/1462-2920.13023>
- Parker, M. A., Malek, W., & Parker, I. M. (2006). Growth of an invasive legume is symbiont limited in newly occupied habitats. *Diversity Distributions*, *12*(5), 563–571. <https://doi.org/10.1111/j.1366-9516.2006.00255.x>
- Parrish, D. J., & Fike, J. H. (2005). The biology and agronomy of switchgrass for biofuels. *Critical Reviews in Plant Sciences*, *24*, 423–459. <https://doi.org/10.1080/07352680500316433>
- Peres-Neto, P. R., Legendre, P., Dray, P., & Borcard, D. (2006). Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology*, *87*(10), 2614–2625. <https://doi.org/10.1890/0012-9658>
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, *41*, D590–D596. <https://doi.org/10.1093/nar/gks1219>
- Risely, A. (2020). Applying the core microbiome to understand host-microbe systems. *Journal of Animal Ecology*, *89*, 1549–1558. <https://doi.org/10.1111/1365-2656.13229>
- Rossum, B. van, Kruijer, W., Eeuwijk, F. van, & Boer, M. (2020). Package “statgenGWAS” Title Genome Wide Association Studies. doi: <https://doi.org/10.1038/ng.548>
- Sanderson, M. A., Adler, P. R., Boateng, A. A., Casler, M. D., & Sarath, G. (2006). Switchgrass as a biofuels feedstock in the USA. *Canadian Journal of Plant Science*, *86*(5), 1315–1325. <https://doi.org/10.4141/p06-136>
- Sasse, J., Martinoia, E., & Northen, T. (2018). Feed your friends: Do plant exudates shape the root microbiome? *Trends in Plant Science*, *23*, 25–41. <https://doi.org/10.1016/j.tplants.2017.09.003>
- Schwartz, C., & Amasino, R. (2013). Nitrogen recycling and flowering time in perennial bioenergy crops. *Frontiers in Plant Science*, *4*, 76. <https://doi.org/10.3389/FPLS.2013.00076>
- Sharma, M. K., Sharma, R., Cao, P., Jenkins, J., Bartley, L. E., Qualls, M., Grimwood, J., Schmutz, J., Rokhsar, D., & Ronald, P. C. (2012). A genome-wide survey of switchgrass genome structure and organization. *PLoS One*, *7*(4), e33892. <https://doi.org/10.1371/journal.pone.0033892>
- Singer, E., Bonnette, J., Woyke, T., & Juenger, T. E. (2019). Conservation of endophyte bacterial community structure across two *panicum* grass species. *Frontiers in Microbiology*, *10*, 2181. <https://doi.org/10.3389/FMICB.2019.02181>
- Smith, C. C. R., Snowberg, L. K., Gregory Caporaso, J., Knight, R., & Bolnick, D. I. (2015). Dietary input of microbes and host genetic variation shape among-population differences in stickleback gut microbiota. *ISME Journal*, *9*(11), 2515–2526. <https://doi.org/10.1038/ismej.2015.64>
- Stewart, C. E., Roosendaal, D., Deneff, K., Pruessner, E., Comas, L. H., Sarath, G., Jin, V. L., Schmer, M., & Soundararajan, M. (2017). Seasonal switchgrass ecotype contributions to soil organic carbon, deep soil microbial community composition and rhizodeposit uptake during an extreme drought. *Soil Biology and Biochemistry*, *112*, 191–203. <https://doi.org/10.1016/j.soilbio.2017.04.021>
- Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., Xu, W., & Su, Z. (2017). AgriGO v2.0: A GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Research*, *45*, W122–W129. <https://doi.org/10.1093/nar/gkx382>
- Trexler, R. V., & Bell, T. H. (2019). Testing sustained soil-to-soil contact as an approach for limiting the abiotic influence of source soils during experimental microbiome transfer. *FEMS Microbiology Letters*, *366*, fnz228. <https://doi.org/10.1093/femsle/fnz228>

- Trivedi, P., Leach, J. E., Tringe, S. G., Sa, T., & Singh, B. K. (2020). Plant-microbiome interactions: from community assembly to plant health. *Nature Reviews Microbiology*, 18, 607–621. <https://doi.org/10.1038/s41579-020-0412-1>
- Trujillo, M. E., Hong, K., & Genilloud, O. (2014). The family micro-monosporaceae. In *The prokaryotes: Actinobacteria* (pp. 499–569). Springer. https://doi.org/10.1007/978-3-642-30138-4_196
- Van Rossum, G., & Drake, F. L. (2009). *Python 3 reference manual*. CreateSpace.
- VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, 91(11), 4414–4423. <https://doi.org/10.3168/jds.2007-0980>
- Wagner, M. R. (2021). Prioritizing host phenotype to understand microbiome heritability in plants. *New Phytologist*, 232, 502–509. <https://doi.org/10.1111/NPH.17622>
- Wagner, M. R., Lundberg, D. S., Del Rio, T. G., Tringe, S. G., Dangl, J. L., & Mitchell-Olds, T. (2016). Host genotype and age shape the leaf and root microbiomes of a wild perennial plant. *Nature Communications*, 7, 12151. <https://doi.org/10.1038/ncomms12151>
- Weiss, S., Xu, Z. Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., Lozupone, C., Zaneveld, J. R., Vázquez-Baeza, Y., Birmingham, A., Hyde, E. R., & Knight, R. (2017). Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome*, 5(1), 27. <https://doi.org/10.1186/s40168-017-0237-y>
- Werling, B. P., Dickson, T. L., Isaacs, R., Gaines, H., Gratton, C., Gross, K. L., Liere, H., Malmstrom, C. M., Meehan, T. D., Ruan, L., Robertson, B. A., Robertson, G. P., Schmidt, T. M., Schrottenboer, A. C., Teal, T. K., Wilson, J. K., & Landis, D. A. (2014). Perennial grasslands enhance biodiversity and multiple ecosystem services in bioenergy landscapes. *Proceedings of the National Academy of Sciences of the United States of America*, 111(4), 1652–1657. <https://doi.org/10.1073/pnas.1309492111>
- Whitham, T. G., Young, W. P., Martinsen, G. D., Gehring, C. A., Schweitzer, J. A., Shuster, S. M., Wimp, G. M., Fischer, D. G., Bailey, J. K., Lindroth, R. L., Woolbright, S., & Kuske, C. R. (2003). Community and ecosystem genetics: A consequence of the extended phenotype. *Ecology*, 84, 559–573. [https://doi.org/10.1890/0012-9658\(2003\)084\[0559:CAEGAC\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2003)084[0559:CAEGAC]2.0.CO;2)
- Wilson, A. S., Koller, K. R., Ramaboli, M. C., Nesengani, L. T., Ocvirk, S., Chen, C., Flanagan, C. A., Sapp, F. R., Merritt, Z. T., Bhatti, F., Thomas, T. K., & O'Keefe, S. J. D. (2020). Diet and the human gut microbiome: An international review. *Digestive Diseases and Sciences*, 65, 723–740. <https://doi.org/10.1007/s10620-020-06112-w>
- Wray, N. R., Yang, J., Hayes, B. J., Price, A. L., Goddard, M. E., & Visscher, P. M. (2013). Pitfalls of predicting complex traits from SNPs. *Nature Reviews Genetics*, 14, 507–515. <https://doi.org/10.1038/nrg3457>
- Wright, E. S. (2016). Using DECIPHER v2.0 to Analyze Big Biological Sequence Data in R.
- Yin, C., Casa Vargas, J. M., Schlatter, D. C., Hagerty, C. H., Hulbert, S. H., & Paulitz, T. C. (2021). Rhizosphere community selection reveals bacteria associated with reduced root disease. *Microbiome*, 9(1), 1–18. <https://doi.org/10.1186/s40168-020-00997-5>
- Yu, P., & Hochholdinger, F. (2018). The role of host genetic signatures on root-microbe interactions in the rhizosphere and endosphere. *Frontiers in Plant Science*, 871. <https://doi.org/10.3389/fpls.2018.01896>
- Zeng, Q., Sukumaran, J., Wu, S., & Rodrigo, A. (2015). Neutral Models of Microbiome Evolution. *PLoS Computational Biology*, 11(7), e1004365. <https://doi.org/10.1371/journal.pcbi.1004365>

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Sutherland, J., Bell, T., Trexler, R. V., Carlson, J. E., & Lasky, J. R. (2022). Host genomic influence on bacterial composition in the switchgrass rhizosphere. *Molecular Ecology*, 00, 1–17. <https://doi.org/10.1111/mec.16549>