Heritable associations with microbial communities are essential for necrotrophic pathogen resistance

Cloe S. Pogoda^{*1}, Stephan Reinert^{*1}, Zahirul I. Talukder³, Ziv Attia¹, Jason A. Corwin¹,
Kennedy L. Money³, Erin C.E. Collier-zans¹, William Underwood², Thomas J. Gulya², C.
Alisha Quandt¹, Nolan C. Kane¹, Brent S. Hulke²⁺

¹ University of Colorado, Ecology and Evolutionary Biology Department, 1900 Pleasant Street,
334 UCB, Boulder, CO 80309-0334, USA

² USDA-ARS Sunflower and Plant Biology Research Unit, Edward T Schafer Agricultural
Research Center, 1616 Albrecht Blvd. N., Fargo, ND 58102-2765, USA

³ North Dakota State University, Department of Plant Sciences, NDSU Dept. 7670, Fargo, ND
58108-6050, USA

^{*} Authors contributed equally

⁺Corresponding author: brent.hulke@usda.gov, +1 701 239 1321, ORCID: 0000-0001-5380-0827

Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

Abstract

Host-microbe interactions are increasingly recognized as important drivers of organismal health, growth, longevity, and community-scale ecological processes. However, less is known about how genetic variation affects hosts' associated microbiomes and downstream phenotypes. We demonstrate that sunflower (*Helianthus annuus*) harbors substantial, heritable variation in microbial communities under field conditions. We show that microbial communities explain up to 77.5% of the heritable variation in resistance to root infection caused by the necrotrophic pathogen *Sclerotinia sclerotiorum*, and that plants grown in sterilized soil showed almost complete elimination of pathogen resistance. Association mapping revealed 69 genetic locations related to microbial abundance and *Sclerotinia* resistance. Although the genetic architecture is complex and quantitative, we have, in large part, elucidated previously unexplained genetic variation for resistance to this pathogen. This suggests new targets for plant breeding and demonstrates the potential for heritable microbial associations to play important roles in defense in natural and human-altered environments.

Introduction

Plant-microbial interactions have long been understood to be important by ecologists ^{1,2}, but the advent of agroecology and DNA methods for profiling these communities has stimulated renewed interest among scientists and plant breeders. A very important site of these interactions is the soil-root interface within the soil rhizosphere ^{3,4}. Previous research has indicated that certain microbial components of the rhizosphere can be protective to the plant and enhance tolerance and resistance to biotic and abiotic stresses, benefit overall biodiversity, and increase nutrient uptake and availability ^{5,6}. While research has suggested that the microbial community associated with each plant can be affected by species and even genotype,⁷ many previous reports have also suggested that differences in soil abiotic characteristics have a larger impact on rhizosphere microbial community composition than host natural genetic variation ^{8–10}. Thus, a better understanding of how plant genetic diversity can shape the soil rhizosphere microbiome will affect how the microbiome is used to enhance important plant traits and soil health. Particularly intriguing are suggestions that the microbiome may play an important role in plant pathogen resistance².

Cultivated sunflower (*Helianthus annuus* L.) offers an excellent system to study how plant genotype may influence plant-microbial interactions and associated disease resistance. Sunflowers are the fourth most valuable annual oil crop, with 480 million tons grown worldwide ¹¹, used as a high value vegetable oil for human consumption. It is also a naturally cross-pollinated species with excellent diversity panels and associated genomic resources. One serious pathogen of sunflower, *Sclerotinia sclerotiorum*, is a necrotroph known to infect most herbaceous, dicotyledonous plants in temperate climates, including important crops such as soy, canola, lettuce, lentils, peas, and beans ¹². *Sclerotinia*-induced basal stalk rot and wilt is the result of myceliogenic infection of the roots, which is unique to sunflowers and related species ⁹. Diseased plant symptoms include sudden wilting of the plant during seed filling and basal stalk lesions, often leading to visible signs of the pathogen including white mycelia and black sclerotized bodies, and resulting in yield loss due to plant lodging and premature death ^{13,14}. Disease management in sunflower has only been successful by selective breeding

for heritable resistance to the pathogen ^{14,15}so a better understanding of genetics underlying resistance is essential.

Despite its clear importance, little is known about the mechanisms through which host-plant resistance is conferred. Previous studies have not identified obvious plant traits associated with resistance. With the rhizosphere being the site of infection for *Sclerotinia* basal stalk rot, we focus here on mechanisms that specifically include rhizosphere microbial community dynamics. In so doing, we identify how the non-*Sclerotinia* microbial communities affect *Sclerotinia* resistance. Together, these analyses suggest that microbial partners explain previously unexplained heritable variation in *Sclerotinia* resistance. By exploring the tripartite interaction of plant genotype, rhizosphere composition, and *Sclerotinia* disease, we provide new insight into plant defense against this cosmopolitan pathogen, which in turn will provide the potential for new methods of control that include enhancement of the host-plant genetic contributions to soil community health together with probiotic applications or improved soil microbial functioning.

Methods

***Helianthus annuus* Plant Material**

Plant materials used in all field experiments were from the Open Pollinated Variety diversity panel of sunflower, which contains 260 USDA plant introductions (PIs), of which ten were released lines from the USDA sunflower breeding program (a list is available in ^{14,16}. Of these, 95 were chosen randomly for further analysis of rhizosphere composition and another environment of inoculated basal stalk rot testing at the Carrington Research and Extension Center, Carrington, ND, USA, in 2017. Six of the released sunflower lines from the USDA sunflower breeding program were included in the subset.

For the greenhouse experiment, 40 accessions were further chosen from the subset of 95 lines (Table S1). These 40 were divided equally between highly susceptible or highly resistant to *Sclerotinia* basal stalk rot in the field.

The sunflower lines associated with each aspect of our work are listed in Table S1.

***Helianthus annuus* DNA Extraction and Sequencing**

Lyophilized leaf material for the 260 PIs were ground using tungsten carbide bearings in a Qiagen 96-well plate shaker. A total of 10 plants per PI were co-processed for each sample. Genomic DNA (gDNA) was extracted from leaf tissues using a Qiagen DNeasy 96 plant kit. The manufacturer's protocol was modified to include the addition of 10 mM SMB (sodium metabisulfite) to the initial lysis buffer, a 45 minute incubation at 65 °C for the ground material in lysis buffer, a 100% ethanol wash before final drying of the membrane prior to elution, and DNA storage in an elution buffer that contained 10mM DTT (dithiothreitol), which have all been shown to improve DNA concentration and purity^{17,18}. Extracted samples were stored at -20 °C prior to library preparation.

Genomic libraries were prepared following standard protocols using Nextera® XT DNA library prep kits (Illumina®) and were barcoded with the Nextera® adaptors i5 and i7. Pools that passed QC were processed for an average coverage depth of 5x, 151 bp, paired-end HiSeq® reads at the NovoGene sequencing facility in Sacramento, CA.

***Helianthus annuus* Variant Calling**

Demultiplexed data was downloaded directly from NovoGene's servers. FASTQ data were trimmed using Trimmomatic version 0.38 with the following parameters: NexteraPE-PE.fa:2:30:10 LEADING:3 TRAILING:3 MINLEN:100, with NexteraPE-PE.fa containing the standard set of Nextera adapters to be trimmed from reads. Resulting FASTQ files were aligned to the most up-to-date assembly available as of writing, HA412-HO.v2.fasta (L. Rieseberg, pers. comm., 2019). Variant calling was performed using samtools *mpileup* (samtools mpileup -l -uf) and bcftools *call* (bcftools call -v -m). A single variant call file (vcf) table was created and then pruned to contain the individuals specific to each environment/trait to remove poor quality genotypic and phenotypic data from downstream analysis.

The vcf table was filtered for single copy sites based on depth (sites with depth between 200-1000 were retained). Additionally, filtering was done to select sites with a minimum quality score of 100 (minQ=100), minor allele frequency of 0.1 (maf=0.1) and max missingness value of 0.75 (max-missing=0.75). Missing data were imputed using BEAGLE version 5.0 with default settings retained¹⁹. Plink version 1.9 with the

parameters: - - indep 100 5 10, corresponding to a window size of 100 SNPs, step size of 5 SNPs, and a VIF threshold of 10 (equivalent to an R^2 value of 0.9), were used to select markers with low redundancy for use in subsequent steps²⁰.

***Sclerotinia* Disease Incidence / Resistance Phenotyping**

A previously studied, high quality, multi-environment *Sclerotinia* basal stalk rot data set was re-analyzed in our study. Details of the experiments are available in Talukder et al. (2014)¹⁴. Briefly, the Open Pollinated Variety diversity panel was grown in four environments in the USA (Crookston, MN, 2008; Crookston, MN, 2009; Davenport, ND, 2008; and Grandin, ND, 2009). Field plots included about 20 plants and were arranged in a randomized complete block design with two replications per environment and 13 nested sets-in-reps incomplete blocks per replication, to effectively control error due to uncontrollable field trends²¹. Plots were side-dressed with dried millet seed infested with *Sclerotinia* isolate NEB-274 and were rated at maturity as percentage of plants with disease symptoms in each plot. These datasets are referred to by location name and year when analyzed individually, and as the multi-environment dataset when analyzed based on multi-environment mean values.

Rhizosphere sample collection

The analysis of rhizosphere composition was conducted in one environment (Carrington, ND) in 2017, in a field with diverse crop history and multiple years of inoculated *Sclerotinia* trials. This site was chosen because the likelihood that soil microbial populations antagonistic to *Sclerotinia* existed there was high, based on crop history and previous use as a *Sclerotinia* disease trial site. Approximately 20 plants of each PI were grown in single row, 6.1 m plots, spaced 0.76 m apart, using a Case IH 1200 series planter fitted with Almaco belt cones. The experiment consisted of a case-control design with nested randomized complete block design, such that two replications on one-half of the field area were inoculated with dried millet seed infested with *S. sclerotiorum* isolate NEB-274, and another two replications were mock-inoculated with autoclaved, dried millet only. At late bud stage, two plants from each mock-inoculated plot were destructively sampled for rhizosphere composition analysis.

Soil samples were collected by extracting the root ball with a shovel (approximately 15 cm wide and 15 cm deep) and removing the loose dirt from the roots. The remaining soil adhered to the roots and was separated from the roots by shaking it into a metal pan. Roots and non-soil matter were removed, and the rhizosphere-associated soil further sized using a 2 mm aperture sieve (Seedburo). Total volume of rhizosphere soil collected per sunflower accession was approximately 50 ml. Samples were stored at -20 °C prior to DNA extraction. All plots were rated near maturity according to Talukder et al. (2014) for basal stalk rot percent incidence ¹⁴.

Rhizosphere DNA and Bioinformatic Processing

DNA from rhizosphere-associated soil was extracted using the standard protocol for the Qiagen DNeasy PowerSoil HTP 96 kit. The phylogenetically informative prokaryotic 16S were amplified using the barcoded 515 f/926r and ITS1F/ITS2 primer pairs, respectively. Sample amplicons were pooled and sequenced on an Illumina MiSeq instrument using paired end sequencing for 251 bp at the BioFrontiers sequencing center at the University of Colorado, Boulder.

Raw reads were processed using a previously described uSearch bioinformatic pipeline ²². Briefly, primer dimers and adapter sequences were removed using the ‘cutadapt’ software and filtered for high quality reads using USEARCH v10 ‘fastq_filter’ function ^{22,23}. Paired-end reads were merged using USEARCH v10 ‘fastq_mergepairs’ function and samples were demultiplexed using a custom in-laboratory python script. The sequences were de-replicated and denoised using the ‘derep_fulllength’ and ‘unoise3’ algorithms, respectively. Raw reads were mapped back to the unique denoised sequence database and an operational taxonomic unit (OTU) count table was created using custom python scripts. Putative phylogeny of unique denoised sequences was determined using the RDP classifier against the Greengenes phylogenetic database v13.5; for 16S sequences ²⁴. Any sequences that mapped to eukaryotic chloroplast or mitochondrial sequences were removed from the analysis.

Greenhouse validation experiment

Native soil was obtained in late 2018 from the same field as the original rhizosphere study near Carrington, ND. The soil was placed in plastic boxes and transported to the USDA-ARS in Fargo, ND. The soil inside the boxes was broken up with a shovel and large plant debris and rocks were removed. Half of the soil was marked for autoclaving, with the remaining half not autoclaved. Autoclaved soil was treated twice for 45 minutes on a wet cycle at 121 °C, with agitation between cycles. To measure the effectiveness of the autoclave sterilization, *Geobacillus stearothermophilus* bioindicators (Crosstex, Little Falls, NJ) were placed in the middle of the bags of soil. The vials were removed during agitation and placed back in the center of the soil for the second round of autoclaving. After the two cycles of autoclaving were complete, twenty bioindicators and two control indicators were activated and incubated for 24 hours at 60 °C, as prescribed by the manufacturer. The treatment bioindicators were compared to the controls and the results showed that two of the twenty were not sterilized. The soil then underwent a third 121 °C wet cycle for 30 minutes, with five bioindicators placed randomly in the soil. These vials were evaluated after the 24 h incubation period and indicated that the soil had been effectively sterilized.

The experimental design was based on the procedure of Underwood et al. (2020)²⁵. All plants were grown from seed in the greenhouse. Initially, 36 plants of each of the 40 accessions were planted in each of the native and the autoclaved soil, in cell flat trays (TO Plastics 730654C, Clearwater, MN, USA). Plants grown in autoclaved soil and native soil were physically separated on different benches. These plants were allowed to grow approximately 5 weeks after planting with up to twice-a-day watering to avoid water stress.

After 5 weeks, the plants were removed from the early growth trays. They were transplanted into new trays in a split-plot design with three replications nested within the main plot treatment of autoclaved vs. native soil, to prevent crossover inoculation across soil treatments within a replication. Each replication was completely randomized, containing 12 plants each of the 40 accessions, and placed in fifteen 32-cell trays with 2.5 mL (or about ~0.38 g) of *S. sclerotiorum* inoculated millet (isolate NEB-274) at the bottom of each plant. Development of millet inoculum has been previously described¹⁴. Each replicate was placed on a separate bench in the same greenhouse.

After inoculation and randomization, for the next 28 days, each plant was evaluated daily for symptoms of *Sclerotinia* basal stalk rot. Plants were evaluated for days-to-death by noting meristematic necrosis, typified by browning of the apical meristem / emerging leaves, and in most cases, permanent terminal wilt. If a plant exhibited either or both of these symptoms for two consecutive days, it was noted as dead on the first day and removed from the tray the second day.

Statistical Analysis

Disease incidence / resistance

Percent incidence data for all basal stalk rot experiments were analyzed using PROC MIXED of SAS v. 9.4 (SAS Institute, 2013), with all effects random except environment, which remained fixed²⁶. Phenotypes were inverted in some analyses for scaling purposes, and the inverted phenotypes were called percent resistance. Variance components were used to compute broad-sense heritability on an entry-mean basis, using the formula $H^2 = V_g / (V_g + V_{ge}/l + V_e/r)$ for multiple environments and $H^2 = V_g / (V_g + V_e/r)$ for single environment analyses, where V_g , V_{ge} , and V_e are the accession, accession-by-environment, and error variances, respectively; r is the number of replications per environment; and l is the number of environments²⁷.

OTU abundances

We filtered our dataset for all samples with less than 5,000 OTU counts and OTUs that occurred in less than 25% of samples were removed, as their inclusion could skew downstream statistical analysis. The resulting filtered OTU table was normalized using the trimmed M-means algorithm using the 'edgeR' package within the R statistical environment^{28,29}. PERMANOVAs of the normalized pseudo-count matrix were obtained from `adonis()` in the 'vegan' package of R using the following model:

$Y_{rg} = R_r + G_g + \epsilon_{rg}$, where Y_{rg} is the normalized pseudo-count, R_r is the field replicate, G_g is the sunflower accession, and ϵ_{rg} is the error effect³⁰.

The effect of plant accession on each individual OTU was also obtained with a negative binomial generalized linear model using the same model for each individual OTU and the resulting p-values were FDR corrected for multiple comparisons. Broad-sense heritability across the sunflower accessions for each OTU was defined as $H^2 =$

$V_g/(V_g + V_\epsilon/r)$. Model-corrected means for each OTU and sunflower accession were determined from the nb.glm using the LSmeans function within the 'doBy' package.

Network analysis and association with disease

Bacterial OTUs that may serve as a protective function against *Sclerotinia*-induced stalk rot were first extracted using network analysis. Briefly, model-corrected means for each sunflower accession and OTU were correlated with the model-corrected means of basal stalk rot from the Carrington 2017 inoculated study. OTUs were chosen that were most negatively correlated with incidence of stalk rot with an FDR corrected p-value < 0.0001. This analysis resulted in a total of 42 bacterial taxa negatively associated with incidence of *Sclerotinia* infection. NMDS plots were generated using R-package *vegan* in R version 3.6.

Each of the 42 bacterial taxa that were negatively correlated with *Sclerotinia* incidence were then tested individually against resistance phenotypic data using a general linear model ('anova' function in the 'lm' package of R)²⁹. The model p-values were corrected using the Bonferroni method, and OTUs significant at the corrected-p < 0.05 threshold were tested for pairwise interaction in a series of models that consisted of all individually significant OTUs as main effects plus one interaction effect of each pair of significant OTUs. Resulting p-values were again Bonferroni-corrected by the number of interactions tested. R^2_P values were extracted from the regressions.

The percentage of the heritable portion of the disease resistance phenotype explained by the heritable portion of the contribution of rhizosphere was calculated as $R^2_G = R^2_P \times H^2_{SR} \times H^2_{OTU}$, where: R^2_G is the percentage of the heritable portion of the stalk rot disease phenotypic variation explained by the heritable portion of variation in a single OTU, R^2_P is the percentage of the phenotypic variation in stalk rot disease in a given environment explained by the phenotypic variation in each OTU, as above, H^2_{SR} is the heritability of the stalk rot disease phenotype in a given environment or multi-environment mean, as calculated above, and H^2_{OTU} is the heritability of the OTU phenotype, as calculated above.

Greenhouse validation experiment

Analysis of the greenhouse experiment was conducted using PROC MIXED of SAS v. 9.4²⁶. Data were modelled as follows: $Y_{ijkl} = S_i + R[S]_{ij} + G_k + SG_{ik} + \varepsilon_{ijk} + w_{ijkl}$, where Y_{ijkl} is the number of days to death, post-inoculation, S_i is the effect of i^{th} soil type (native vs. autoclaved), $R[S]_{ij}$ is the effect of the j^{th} replication (1, 2, 3) nested within the i^{th} soil type, G_k is the effect of the k^{th} sunflower accession (1 to 40), SG_{ik} is the interaction of the i^{th} soil type with the k^{th} sunflower accession, ε_{ijk} is the residual model error, and w_{ijkl} is the within treatment effect of the l^{th} plant (1 to 12) nested within the i^{th} soil type, j^{th} replication, and k^{th} sunflower accession.

Least square means were calculated for each level of the soil-by-sunflower-accession interaction, using the LSMEANS option. Tukey's multiple comparison method was used to compare the least square means. Least square means of days to death for each accession, separately for native and autoclaved soil, were then tested individually against the 42 OTUs using a general linear model ('anova' function in the 'lm' package of R), in the same way as the field data²⁹. The percentage of the heritable portion of the days-to-death phenotype explained by the heritable portion of the contribution of rhizosphere was also calculated in the same manner as the field stalk rot evaluations.

Association mapping

Sunflower quantitative trait nucleotides (QTN) associated with each OTU and disease phenotype were identified using the R-package *GAPIT* and specifically the *FarmCPU* function³¹. Least square means of abundance for each OTU, disease incidence of SR at each field environment (including greenhouse experiments) and for the multi-environment mean were the dependent variables used in the analyses. The first principle component, with a maximum number of iterations equal to 10 was used to detect associations between effective markers and traits. Extreme outlier genotypes were culled from the dataset based on visual observation of the principle component graph (Supplemental Figure S5). In order to identify QTNs of interest, we calculated the q-value (estimation of false discovery rate) using the R-package *qvalue* and selected sites above a threshold q-value = 0.5³². These QTNs were marked as potentially interesting drivers of their respective traits and environments.

Linkage Disequilibrium

To understand the extent of linkage disequilibrium (LD) in the vicinity of significant markers, we utilized two separate methods. First local LD was assessed chromosome-wide (for all 17 chromosomes) to identify the extent of the LD blocks due mainly to physical linkage and other population genetic forces for each of the significant QTNs. For that, we pruned the variant call file (vcf; genotypic data) of each chromosome to include every hundredth marker as well as the QTNs of interest (described above in the association mapping section) associated with each chromosome. The resulting pruned vcf table was converted into plink formatted files using vcfutils. Next, plink version 1.9 was utilized to determine the R^2 LD matrix²⁰. This matrix was further filtered to only include the pairwise comparisons for each of the significant QTNs to all other markers on the same chromosome. The resulting chromosome-wide data were graphed using *ggplot2* in R version 3.6.0²⁹. These data were parsed to determine the size of the LD block surrounding each QTN (Table S2). Second, to find longer range LD among significant QTN within and among chromosomes, due to selection forces on unlinked loci, we used Haploview Desktop version 4.2 with default settings to examine LD among the significant QTNs of all field sites and significant OTUs³³.

Results

Identification of Rhizosphere Taxa-of-interest and Associated Heritability Values

16S marker profiling of the sunflower rhizosphere microbiome identified 890 unique bacterial and archaeal OTUs in the collected sunflower samples. Of these 890, statistical microbial diversity analysis revealed 42 taxa which are negatively correlated with disease incidence but are relatively unstructured among the 890 OTUs identified (Figure 1) and are from a variety of different phyla (Table S3). Due to their putative importance to inhibiting *Sclerotinia* establishment on the host, we focused on these 42 OTUs for further study.

Figure 1: NMDS plot of the 890 (grey dots) operational taxonomic units identified in the soil sampled from Carrington, ND, in 2017. Forty-two taxa (black triangles) that are negatively correlated with *Sclerotinia* basal stalk rot disease resistance are highlighted in black. Black polygon indicates the boundaries of the 42 taxa.

Individual regression analysis of the 42 OTUs associated with disease resistance identified four bacterial taxa which showed a statistically significant effect on basal stalk rot resistance in Carrington (adjacent, *Sclerotinia*-inoculated field plots in the same environment as the rhizosphere assessment; Figure S1, Table S4). These OTUs were identified as OTU77: Acidobacteria, OTU293: Verrucomicrobia bacteria, OTU750: Rhizomicrobium bacteria, and OTU908: Actinobacteria. When OTU abundance in Carrington 2017 was regressed on the multi-environment data set of Talukder et al., 2014¹⁴, OTU750 showed a highly significant response (Bonferroniized $p = 0.003$). When the multi-environment data from the Talukder study was reduced to individual environments, OTU293, OTU750, and OTU908 were significant in at least two of the four environments. This result shows the conservation of these relationships over a 250 km distance among several field sites and with different crop history. The broad sense heritability values for the four significant OTUs were centrally distributed among the values for all 890 individuals initially identified (Figure 2).

Figure 2: Heritability plot of the 890 bacterial operational taxonomic units sampled in 2017 from the Carrington, ND, field. Forty-two taxa that were negatively correlated with *Sclerotinia* disease resistance fall within the bars highlighted in red. The heritability values of the four statistically significant OTUs are given in the figure legend (top right).

Variation for Basal Stalk Rot Resistance Explained by Rhizosphere

In order to better understand the role of genetics in the stalk rot phenotype, we calculated the broad-sense heritability of disease resistance for each environment studied. The broad-sense heritability estimate for *Sclerotinia* resistance in the Carrington 2017 field environment was high (0.76), while for the multi-environment mean of the Talukder et al., 2014 study, the estimate was 0.70¹⁴. Individually, those four environments were as follows: Crookston in 2008, 0.57; Crookston in 2009, 0.47; Davenport in 2008, 0.39; and Grandin in 2009, 0.35.

We then correlated the rhizosphere community present at the Carrington 2017 environment with the disease resistance phenotypes in all of our evaluations (Figure 3 & S2). We note that the only portion of phenotypic variation that could co-vary across environments was the heritable portion. Table 1 shows the percentage of the heritable portion of the phenotypic variance in stalk rot resistance explained by the heritable variance for each OTU. The percent heritable variation that was explained by each of the four OTUs in each separate field environment ranged between 11.0% for OTU77 in Grandin in 2009 and 77.5% for OTU293 in Davenport in 2008.

Table 1: Percent heritable *Sclerotinia* basal stalk rot resistance variation explained by each operational taxonomic unit in field and greenhouse environments.

Taxon	Individual evaluation sites					Carrington ^a		
	Crookston08	Crookston09	Davenport08	Grandin09	Multi ^b	Field	Autoclaved	Native
	<hr/> ----- R^2_G ----- <hr/>							
OTU77	40.9	14.2	42.2	11.0	31.1	61.1	10.6	12.9
OTU293	67.7	17.6	77.5	23.5	28.9	37.9	57.5	33.3
OTU750	57.2	17.8	66.5	28.1	45.7	38.7	33.6	37.1
OTU908	73.3	66.5	41.7	16.6	31.8	58.5	67.0	56.4

^a Field: *Sclerotinia* field trial adjacent to rhizosphere collection experiment; Autoclaved: Greenhouse study using autoclaved Carrington soil; Native: Greenhouse study using non-autoclaved Carrington soil.

^b Multi: Mean of *Sclerotinia* basal stalk rot disease incidence across four field sites (Crookston08, Crookston09, Davenport08, Grandin09).

To validate these findings, we used a greenhouse test for basal stalk rot resistance screening that has been previously verified to replicate field resistance ²⁵. Forty sunflower accessions from the 95 that were sampled for rhizosphere composition in Carrington were re-evaluated in replicated greenhouse trials under both native soil and autoclaved soil conditions, with high correspondence between the field resistance and greenhouse native soil experiments (Figure 3; Table S5). Results showed that days to death from stalk rot under native soil for each sunflower accession was between 2 to 19 days longer for native soil compared to autoclaved (mean = 9 d, median = 8 d, $p < 0.0001$). The distribution in days to death in autoclaved soil was considerably narrower than under native soil, which resulted in a significant soil-by-accession effect ($p <$

0.0001). The most resistant accessions in the Carrington 2017 test survived, on average, 10 d longer in native soil than autoclaved (median = 10 d) in the greenhouse test, while the least resistant from the field test averaged 7 d (median = 6 d) longer. Despite being autoclaved, a small but significant effect of OTU on resistance occurred in that treatment, which could be due to environmental contaminants that increased over the 5 week plant growth interval before inoculation or increase from some very low level of microbial inoculum that survived autoclaving.

Figure 3: Correlation between disease resistance and operational taxonomic unit (OTU) abundance for the top four OTUs that were significant in multiple field environments. Data points are colored to indicate the OTU they represent as indicated by the figure legend. The x-axis is relative OTU abundance (count in thousands) in Carrington, 2017, for each panel. Y-axis is disease resistance calculated as 1 minus [disease incidence] (disease incidence: plants showing wilt or lesions) in panels A and B and days-to-death in panels C and D. Regression bands indicate a 95% confidence interval. **A)** Carrington field environment **B)** Multi-environment mean (combined data from four separate sites: Crookston08, Crookston09, Davenport08 and Grandin09) **C)** Greenhouse native soil **D)** Greenhouse autoclaved soil

Genetic architecture of the rhizosphere - *Sclerotinia* resistance association

We examined the *Sclerotinia* and rhizosphere datasets for significant QTNs using standard GWAS techniques detailed above. Due to moderate heritabilities of both traits and quantitative inheritance, multiple genomic loci of variable effect size were associated with each OTU and disease screening environment. Markers that fell above a q-value threshold of 0.5 were identified for further analysis, which resulted in 69 QTNs total for all traits and environments (Figure 4, 5, S3). For each QTN, local LD tended to decay dramatically in most cases. There were notable exceptions, including chromosomes 3, 6, 7, 8, 9, 10, 13, 14, 16 and 17, which had LD at significant QTN extending beyond 1 Mbp (Table S2). Chromosome 9 had the most QTNs at 7, while chromosome 5 had the fewest at 1 QTN.

Figure 4: Genome-wide Manhattan plot for the abundance of four operational taxonomic units (OTUs) that are significant in multiple environments. Chromosomes are labeled and represented in alternating greyscale. The marker sites above a q-value of 0.5 are highlighted in red for each plot. **A)** OTU77 **B)** OTU293 **C)** OTU750 **D)** OTU908

Figure 5: Genome-wide Manhattan plot for stalk rot resistance in four experiments. Chromosomes are labeled and represented in alternating greyscale. The marker sites above a q-value of 0.5 are highlighted in red for each plot. **A)** Carrington 2017 **B)** Multi-environment (mean of Crookston08, Crookston09, Davenport08 and Grandin09) **C)** Greenhouse native soil **D)** Greenhouse autoclaved soil

In order to identify marker-trait QTNs in high LD due to historical selection forces on the diversity panel, we calculated LD for all significant sites and focused solely on those sites which were in LD at $R^2 \geq 0.6$. Even after previous culling of collinear QTN due to local LD, there was still evidence of extensive interchromosomal LD, including 20 instances of QTN for both *Sclerotinia* and OTU phenotypes in high LD (Table S2, Figure S4). Since many of the open pollinated varieties on the panel diverged ca. 500 years ago with the diaspora of sunflower from North America overseas, it is likely that a bottleneck, followed by a population expansion, has influenced our diversity panel at loci of functional importance under cultivation or natural environments. Logically, this would include *Sclerotinia* resistance, as the disease occurs worldwide. In this scenario functional loci within each haplotype may have been maintained in high LD while neutral, intervening nucleotides underwent mutation and drift. Additionally, high interchromosomal LD was also observed among many pairs of QTN, which is almost certainly due to selection, given negligible population structure in the panel (Figure S5). Given the complexity of this LD, we subjected the LD between QTN to a neighbor-joining tree decomposition to determine networks of loci working in tandem for our phenotypes of interest (Figure S6). It is apparent that multiple, physically unlinked loci

have been selected in tandem across the genome for the purpose of increasing abundance of our four OTUs of interest (and likely others), reducing *Sclerotinia* stalk rot disease (a known target of sunflower breeding since domestication), and perhaps other, yet unknown purposes.

Discussion

Using a population of cultivated sunflower genotypes that were grown in multiple environments and evaluated for resistance to *Sclerotinia sclerotiorum* myceliogenic root infection, also known as basal stalk rot, we show a strong link between host-plant resistance and heritable associations with the rhizosphere microbial community. We hypothesize that specific bacterial OTUs increase resistance to this plant pathogen. It has been shown that healthy rhizosphere communities can contribute to not only overall plant vigor but ability to resist biotic stressors^{35–37}. By connecting the disease resistance of specific sunflower genotypes with bacterial strains present in a healthy rhizosphere, we achieve greater understanding in how to fight this destructive and detrimental pathogen.

In our study, we sampled soil from one field site (Carrington) which had a rich history of *Sclerotinia* disease studies on several different crops. We identified 890 distinct bacterial OTUs in the soil sampled. Among these 890 individuals, we further identified 42 OTUs that were significantly and negatively associated with *Sclerotinia sclerotium* disease incidence. Regression analyses identified four of these 42, individually, as being statistically significant associations. While we did not directly measure the soil communities in the other field environments we analyzed (Crookston08, Crookston09, Devenport08, Grandin09, autoclaved greenhouse soil, native greenhouse soil), we observed the same pattern of correlation in all of the field environments for some or all of these four OTUs (Figure 3 & S2), suggesting that plant genetics associated with microbial abundance and disease resistance in Carrington was also associated with disease resistance in these other environments.

The heritability values of the four OTUs identified were centrally distributed, indicating that these bacteria are truly significant and not spuriously correlated due to

statistical bias. We identified the four most significant OTUs as bacteria that are commonly present in healthy rhizospheres from a wide range of environmental conditions³⁸. The ubiquity of these species in healthy plants' microbiomes suggests that these taxa may be of general importance and not specific to the confines of our study, though further validation will be necessary to support this. Nevertheless, it supports the hypothesis of a healthy bacterial community increasing resistance to biotic stress in the common sunflower^{38–44}.

Our field results were validated with a greenhouse experiment where we grew a subset of the plant genotypes in both autoclaved (sterilized) soil and unsterilized soil directly from the Carrington field. Greenhouse experiments in unsterilized (native) field soil showed similar results to field conditions, but in sterilized soil resistant plant genotypes performed much worse in the *Sclerotinia* assay, to the extent that nearly all resistance was eliminated, thus adding to the data that a healthy rhizosphere plays an important role in a plant's ability to resist infection^{45–47}.

Quantitative trait nucleotides (QTN) in the host plant contributing to OTU abundance and *Sclerotinia* disease incidence formed LD clusters inter-chromosomally. This provides a solid basis to assert that some of the same genetic loci that affect OTU abundance in sunflower are also resulting in disease resistance. This could be either by a pleiotropic effect (the underlying genes provide a product that enhances certain OTUs but is detrimental to *Sclerotinia*) or because of a causative effect of genes enhancing OTUs and the OTUs, in turn, inhibiting *Sclerotinia*. Our greenhouse work, already discussed, suggests the latter is more likely. Additionally, LD blocks identified between significant QTNs on chromosomes 3, 4, 9, 11, 14 and 17 are supported by previous studies which found *Sclerotinia* resistance loci on these chromosomes^{48–50}. Given the diversity panel is composed of cultivated sunflower lines at various levels of improvement and different geographic origins, high LD among QTN is likely the result of artificial and natural selection for OTU abundance together with *Sclerotinia* disease resistance. This is bolstered by the negligible signals of population structure (Figure S5). While we reason that pleiotropy is occurring, epistasis and local clustering of genes with important additive effects into haplotypes could also be important.

Conclusion

The microbial taxa analyzed here hold promise as probiotics that could reduce the risk of *Sclerotinia* disease or as indicator organisms for improvement of agricultural soil health. More broadly, our demonstration that plant genotype shapes abundance of numerous taxa in the soil microbial community indicates a high potential for selection to act on this genetic variation, enabling evolution of more favorable plant-microbe interactions. If these bacterial associations provide benefits or costs to plant fitness, by provisioning disease resistance or other phenotypes, natural and artificial selection would be expected to shape microbial associations, either in a plant breeding context or in natural ecosystems. The heritable microbiome variation in this cultivated sunflower diversity panel suggests the potential for tradeoffs that are leveraged differently in each selection environment, but this panel also provides a valuable system with which to investigate the ecology, genetics and physiology of microbial interactions, with numerous applications as well as importance for basic science.

References

1. Waksman, S. A. & Starkey, R. L. The Soil and the Microbe. *Soil Sci.* **32**, (1931).
2. Smith, K. P., Handelsman, J. & Goodman, R. M. Genetic basis in plants for interactions with disease-suppressive bacteria. *Proc. Natl. Acad. Sci.* **96**, 4786–4790 (1999).
3. Schmidt, E. L. Initiation of Plant Root-Microbe Interactions. *Annu. Rev. Microbiol.* **33**, 355–376 (1979).
4. Schirawski, J. & Perlin, M. H. Plant–microbe interaction 2017—the good, the bad and the diverse. *International Journal of Molecular Sciences* (2018) doi:10.3390/ijms19051374.
5. Lugtenberg, B. J. J., Chin-A-Woeng, T. F. C. & Bloemberg, G. V. Microbe-plant interactions: Principles and mechanisms. *Antonie van Leeuwenhoek, Int. J. Gen. Mol. Microbiol.* (2002) doi:10.1023/A:1020596903142.
6. Morrissey, J. P., Dow, J. M., Mark, G. L. & O’Gara, F. Are microbes at the root of a solution to world food production? Rational exploitation of interactions between microbes and plants can help to transform agriculture. *EMBO Rep.* (2004) doi:10.1038/sj.embor.7400263.
7. Berg, G. & Smalla, K. Plant species and soil type cooperatively shape the structure and function of microbial communities in the rhizosphere. *FEMS Microbiol. Ecol.* **68**, 1–13 (2009).
8. Leff, J. W. *et al.* Predicting the structure of soil communities from plant community taxonomy, phylogeny, and traits. *ISME J.* (2018) doi:10.1038/s41396-018-0089-x.
9. Fierer, N. Embracing the unknown: Disentangling the complexities of the soil microbiome. *Nature Reviews Microbiology* (2017) doi:10.1038/nrmicro.2017.87.
10. Walters, W. A. *et al.* Large-scale replicated field study of maize rhizosphere identifies heritable microbes. *Proc. Natl. Acad. Sci. U. S. A.* (2018) doi:10.1073/pnas.1800918115.
11. Faostat. 2019.
12. Bolton, M. D., Thomma, B. P. H. J. & Nelson, B. D. *Sclerotinia sclerotiorum* (Lib.) de Bary: Biology and molecular traits of a cosmopolitan pathogen. *Molecular Plant Pathology* (2006) doi:10.1111/j.1364-3703.2005.00316.x.
13. Huang, H. C., Dueck, J., Huang, H. C. & Dueck, J. Wilt of sunflower from infection by mycelial-germinating sclerotia of *Sclerotinia sclerotiorum*. *Can. J. Plant Pathol.* **2**, 47–52 (1980).
14. Talukder, Z. I., Hulke, B. S., Marek, L. F. & Gulya, T. J. Sources of resistance to sunflower diseases in a global collection of domesticated USDA plant introductions. *Crop Sci.* **54**, 694–705 (2014).
15. Masirevic, S. & Gulya, T. J. *Sclerotinia* and *Phomopsis* - two devastating

- sunflower pathogens. *F. Crop. Res.* (1992) doi:10.1016/0378-4290(92)90004-S.
16. USDA. Germplasm Resources Information Network - (GRIN) [Online Database]. *Taxonomy for Plants* (1998).
17. Gao, Q.-M. *et al.* Genetic Architecture of Capitate Glandular Trichome Density in Florets of Domesticated Sunflower (*Helianthus annuus* L.). *Front. Plant Sci.* **8**, 1–15 (2018).
18. Pogoda, C. S., Keepers, K. G., Lendemer, J. C., Kane, N. C. & Tripp, E. A. Reductions in complexity of mitochondrial genomes in lichen-forming fungi shed light on genome architecture of obligate symbioses. *Mol. Ecol.* **27**, 1155–1169 (2018).
19. Browning, B. L., Zhou, Y. & Browning, S. R. A One-Penny Imputed Genome from Next-Generation Reference Panels. *Am. J. Hum. Genet.* **103**, 338–348 (2018).
20. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* (2007) doi:10.1086/519795.
21. Hallauer, A. R., Carena, M. J. & Miranda Filho, J. B. *Quantitative genetics in maize breeding*. (Springer, 1988).
22. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* (2010) doi:10.1093/bioinformatics/btq461.
23. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* (2011) doi:10.14806/ej.17.1.200.
24. McDonald, D. *et al.* An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* (2012) doi:10.1038/ismej.2011.139.
25. Underwood, W., Misar, C. G., Talukder, Z. I., Hulke, B. S. & Markell, S. A greenhouse method to evaluate sunflower quantitative resistance to basal stalk rot caused by *Sclerotinia sclerotiorum*. *Plant Dis.*, in press (2020).
26. SAS Institute. SAS system for Windows. (2013).
27. Nyquist, W. E. Estimation of Heritability and Prediction of Selection Response in Plant Populations. *CRC. Crit. Rev. Plant Sci.* (1991) doi:10.1080/07352689109382313.
28. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* (2009) doi:10.1093/bioinformatics/btp616.
29. R Core Team. R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing, Vienna, Austria* (2013) doi:http://www.R-project.org/.
30. Dixon, P. VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science* (2003) doi:10.1111/j.1654-1103.2003.tb02228.x.

31. Liu X, Huang M, Fan B, Buckler ES, Z. Z. Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome- Wide Association Studies. *PLoS Genet.* **12**, 1–24 (2016).
32. Dabney, A., Storey, J. D., & Warnes, G. R. qvalue: Q-value estimation for false discovery rate control. *R Packag. version 1*, (2010).
33. Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
34. JB, E. Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome* **4**, 250–255 (2011).
35. Bonkowski, M., Cheng, W., Griffiths, B. S., Alpehi, J. & Scheu, S. Microbial-faunal interactions in the rhizosphere and effects on plant growth. *European Journal of Soil Biology* (2000) doi:10.1016/S1164-5563(00)01059-1.
36. Singh, M. *et al.* Complementarity among plant growth promoting traits in rhizospheric bacterial communities promotes plant growth. *Sci. Rep.* (2015) doi:10.1038/srep15500.
37. Marasco, R. *et al.* *Salicornia strobilacea* (synonym of *Halocnemum strobilaceum*) grown under different tidal regimes selects rhizosphere bacteria capable of promoting plant growth. *Front. Microbiol.* (2016) doi:10.3389/fmicb.2016.01286.
38. Da Rocha, U. N., Plugge, C. M., George, I., Van Elsas, J. D. & Van Overbeek, L. S. The rhizosphere selects for particular groups of Acidobacteria and Verrucomicrobia. *PLoS One* (2013) doi:10.1371/journal.pone.0082443.
39. Lee, S. H., Ka, J. O. & Cho, J. C. Members of the phylum Acidobacteria are dominant and metabolically active in rhizosphere soil. *FEMS Microbiol. Lett.* (2008) doi:10.1111/j.1574-6968.2008.01232.x.
40. Nunes Da Rocha, U., Van Overbeek, L. & Van Elsas, J. D. Exploration of hitherto-uncultured bacteria from the rhizosphere. *FEMS Microbiology Ecology* (2009) doi:10.1111/j.1574-6941.2009.00702.x.
41. Zhao, K. *et al.* The rhizospheres of traditional medicinal plants in Panxi, China, host a diverse selection of actinobacteria with antimicrobial properties. *Appl. Microbiol. Biotechnol.* (2012) doi:10.1007/s00253-011-3862-6.
42. Arunachalam Palaniyandi, S., Yang, S. H., Damodharan, K. & Suh, J. W. Genetic and functional characterization of culturable plant-beneficial actinobacteria associated with yam rhizosphere. *J. Basic Microbiol.* (2013) doi:10.1002/jobm.201200531.
43. Pii, Y. *et al.* Microbial interactions in the rhizosphere: beneficial influences of plant growth-promoting rhizobacteria on nutrient acquisition process. A review. *Biology and Fertility of Soils* (2015) doi:10.1007/s00374-015-0996-1.
44. Kalam, S., Das, S. N., Basu, A. & Podile, A. R. Population densities of indigenous Acidobacteria change in the presence of plant growth promoting rhizobacteria

- (PGPR) in rhizosphere. *J. Basic Microbiol.* (2017) doi:10.1002/jobm.201600588.
45. Ahkami, A. H., Allen White, R., Handakumbura, P. P. & Jansson, C. Rhizosphere engineering: Enhancing sustainable plant ecosystem productivity. *Rhizosphere* (2017) doi:10.1016/j.rhisph.2017.04.012.
 46. Majeed, A., Muhammad, Z. & Ahmad, H. Plant growth promoting bacteria: role in soil improvement, abiotic and biotic stress management of crops. *Plant Cell Reports* (2018) doi:10.1007/s00299-018-2341-2.
 47. Crawford, K. M. *et al.* When and where plant-soil feedback may promote plant coexistence: a meta-analysis. *Ecology Letters* (2019) doi:10.1111/ele.13278.
 48. Bert, P. F. *et al.* Comparative genetic analysis of quantitative traits in sunflower (*Helianthus annuus* L.) 1. QTL involved in resistance to *Sclerotinia sclerotiorum* and *Diaporthe helianthi*. *Theor. Appl. Genet.* (2002) doi:10.1007/s00122-002-1004-3.
 49. Micic, Z. *et al.* QTL mapping of *Sclerotinia* midstalk-rot resistance in sunflower. *Theor. Appl. Genet.* **109**, 1474–1484 (2004).
 50. Talukder, Z. I., Seiler, G. J., Song, Q., Ma, G. & Qi, L. SNP discovery and QTL mapping of *Sclerotinia* basal stalk rot resistance in sunflower using genotyping-by-sequencing. *Plant Genome* (2016) doi:10.3835/plantgenome2016.03.0035.

Acknowledgements

The authors would like to thank the staff of CHS Sunflower, Croplan Genetics, and the NDSU Carrington Research and Extension Center for providing space for our field experiments, with special thanks to the Carrington Research and Extension Center for soil for the greenhouse experiment. The assistance of multiple students and technicians in conducting this work is also acknowledged, particularly Christopher Misar, Brady Koehler, and Michael Grove. Noah Fierer and Jessica Henley provided important advice and assistance. This work was supported by USDA-ARS CRIS projects 5442-21220-024, 3060-21000-039, and 3060-21000-043; and the USDA-ARS National *Sclerotinia* Initiative. This research was also supported by BARD, the United States - Israel Binational Agricultural Research and Development Fund, Vaadia-BARD Postdoctoral Fellowship Award No.FI-577-2018.

Author Contribution Statement

BSH, NCK, and TJG designed the study and obtained funding; ZIT, WU, and TJG managed field studies and collected *Sclerotinia* field resistance data; SR and BSH collected rhizosphere samples; KLM and BSH managed greenhouse studies and collected greenhouse data; ZIT and BSH analyzed field and greenhouse sunflower disease data; JAC, CSP, and CAQ analyzed rhizosphere samples and associated data; ECEC and CSP sequenced sunflower lines and developed the genomic marker dataset; CSP and SR conducted genome-wide association studies; CSP, SR, ZA, BSH, NCK, and CAQ initially drafted the manuscript; and the manuscript was revised and all aspects approved by all authors.

Competing Interests

We declare that none of the authors have competing financial or non-financial interests as defined by Nature Research.

NMDS2

▲ Sclero Resistance

NMDS1

LOD

Chromosome

LOD

Chromosome