



# Metagenome-Wide Association Study and Machine Learning Prediction of Bulk Soil Microbiome and Crop Productivity

Hao-Xun Chang<sup>1†</sup>, James S. Haudenschild<sup>1,2</sup>, Charles R. Bowen<sup>1,2</sup> and Glen L. Hartman<sup>1,2\*</sup>

<sup>1</sup> Department of Crop Sciences, University of Illinois, Urbana, IL, USA, <sup>2</sup> USDA—Agricultural Research Service, Urbana, IL, USA

## OPEN ACCESS

### Edited by:

Jeanette M. Norton,  
Utah State University, USA

### Reviewed by:

Wei Shi,  
North Carolina State University, USA  
Richard Allen White III,  
Pacific Northwest National Lab, USA  
John Jacob Parnell,  
Novozymes, USA

### \*Correspondence:

Glen L. Hartman  
ghartman@illinois.edu

### † Present Address:

Hao-Xun Chang,  
Department of Plant, Soil, and  
Microbial Sciences, Michigan State  
University, East Lansing, MI, USA

### Specialty section:

This article was submitted to  
Terrestrial Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 08 January 2017

**Accepted:** 13 March 2017

**Published:** 03 April 2017

### Citation:

Chang H-X, Haudenschild JS,  
Bowen CR and Hartman GL (2017)  
Metagenome-Wide Association Study  
and Machine Learning Prediction of  
Bulk Soil Microbiome and Crop  
Productivity. *Front. Microbiol.* 8:519.  
doi: 10.3389/fmicb.2017.00519

Areas within an agricultural field in the same season often differ in crop productivity despite having the same cropping history, crop genotype, and management practices. One hypothesis is that abiotic or biotic factors in the soils differ between areas resulting in these productivity differences. In this study, bulk soil samples collected from a high and a low productivity area from within six agronomic fields in Illinois were quantified for abiotic and biotic characteristics. Extracted DNA from these bulk soil samples were shotgun sequenced. While logistic regression analyses resulted in no significant association between crop productivity and the 26 soil characteristics, principal coordinate analysis and constrained correspondence analysis showed crop productivity explained a major proportion of the taxa variance in the bulk soil microbiome. Metagenome-wide association studies (MWAS) identified more *Bradyrhizodium* and *Gammaproteobacteria* in higher productivity areas and more *Actinobacteria*, *Ascomycota*, *Planctomycetales*, and *Streptophyta* in lower productivity areas. Machine learning using a random forest method successfully predicted productivity based on the microbiome composition with the best accuracy of 0.79 at the order level. Our study showed that crop productivity differences were associated with bulk soil microbiome composition and highlighted several nitrogen utility-related taxa. We demonstrated the merit of MWAS and machine learning for the first time in a plant-microbiome study.

**Keywords:** machine learning, metagenome-wide association study, microbiome, nitrogen fixation, productivity, random forest, rhizobium, soybeans

## INTRODUCTION

The soil microbiome has been a great interest for its potentials in improving plant nutrient utilization and suppressing soil-borne diseases (Müller et al., 2016). While abiotic soil characteristics such as pH, soil types, and trace elements can strongly influence a microbiome composition (Xu et al., 2009; Tkacz and Poole, 2015), biological factors such as plant species or genotypes can also influence a soil microbiome composition, resulting in taxonomic difference between genotypes (Peiffer et al., 2013; Lakshmanan, 2015). Accordingly, a soil microbiome composition could depend on abiotic and biotic factors, and variations in these factors may cause differences in crop productivity (Tkacz and Poole, 2015). Soybean [*Glycine max* (L.) Merr.] is one of the predominant crops grown in rotation with maize in agronomic fields of Illinois in

the USA. Crop productivity differences in areas within a field have been noted by a number of producers, although the field itself may have the same cropping history, the same soybean genotype (cultivar), and the same management practices in a given season. A hypothesis for the crop productivity difference is that some beneficial and/or detrimental abiotic or biotic factors are unequally distributed in the bulk soils among areas in a field. A couple of studies have suggested the link between yield performances and soil microbiome differences for grape and millet (Debenport et al., 2015; Xu et al., 2015). This could also be the case for field crops.

In order to test this hypothesis, quantifications of a variety of abiotic soil characteristics and the taxa in a soil microbiome are needed. Abiotic soil characteristics can be measured by different chemical and physical analyses, but quantification of taxa can be technically challenging because of the complexity of the soil microbiome. Recent advances in metagenomics, which uses the power of next generation sequencing technology, provides for an approach to quantify taxa in the soil microbiome (Simon and Daniel, 2011). Metagenomics allows a direct detection and quantification of DNA sequences and bypasses the necessity to isolate the organisms, which might be rare in proportion and might be fastidious or unable to culture. Moreover, shotgun metagenomics avoids the concern of PCR amplification bias and provides functional annotation through gene enrichment analysis and pathway analysis (Sharpton, 2014). Although there are several technical challenges, such as sampling consistency from environments, DNA integrity and contamination, and bioinformatic difficulties in taxa annotation and quantification, the power of shotgun metagenomics has been demonstrated in several medical studies on finding associations between taxa in a microbiome and human diseases (Le Chatelier et al., 2013; Lakshmanan, 2015; Zhang et al., 2015). One approach to identify the association is using metagenome-wide association study (MWAS), which takes advantages of huge taxa data discovered using metagenomics and applies the concept of genome-wide association study (GWAS) for the association analysis. Instead of using single nucleotide polymorphisms (SNPs) as the explanatory variables, MWAS employs the abundance of a taxa (a metagenomic species or a metagenomic gene cluster) as the explanatory variables (Wang and Jia, 2016), and MWAS has been successfully used for several human diseases such as type 2 diabetes (Karlsson et al., 2013). Another advantage of the huge taxa data from a metagenomics is to use machine learning methods such as the Random Forest (RF) model or Support Vector Machine model, to integrate the abundance of metagenomic species for phenotypic prediction (Soueidan and Nikolski, 2016; Wang and Jia, 2016). Successful integrative studies for human microbiome and its association with human diseases have been demonstrated (Soueidan and Nikolski, 2016; Wang and Jia, 2016), but to our knowledge, the robustness of MWAS and machine learning has not yet been tested or applied on plant or soil metagenomic data.

Our goal in this study was to determine if abiotic or biotic factors associate with high and low crop productivity areas within agronomic fields. The objectives included the association between crop productivity with abiotic soil characteristics, and

crop productivity with the abundance of metagenomic species based on the reference database from shotgun metagenomic analysis. We applied MWAS to find significant associations between taxa and productivity, and adapted machine learning using RF to predict productivity based on soil microbiome composition.

## MATERIALS AND METHODS

### Soil Sampling and Characterization

Soil samples were collected from six agronomic fields in Illinois (Figure S1). Ten soil core (2.5 cm diameter by 13 cm deep) subsamples were collected from each of two areas in each of the six fields. One area in the field was identified to be high in productivity and another area was identified to be low productivity based on the farmer production records. Samples for each bulk were taken from within an area of less than a 100-meter diameter circle. Samples were blended and divided. One part of each sample was frozen at  $-80^{\circ}\text{C}$  and later lyophilized. The other part of each sample was used for CHN analysis (Microanalysis Laboratory, University of Illinois, Urbana, IL, U.S.A.), and were quantified for other 26 characteristics, including one biotic feature: the soybean cyst nematode (SCN) eggs counts, and 25 abiotic characteristics: latitude and longitude of sampling areas, percentage of clay, sand, and silt, 12 elements (B, Ca, Cu, Fe, Mg, Mn, N, P, K, Na, S, and Zn), percent saturation (PS) of five elements (PS.Ca, PS.H, PS.K, PS.Mg, and PS.Na), cation-exchange capacity (CEC), organic matter, and water pH (SGS North America Inc. Rutherford, NJ, U.S.A.). Pairwise Pearson's correlation was performed using R package "psych" version 1.6.6 (Revelle, 2016) in the R environment version 3.3.1 (R Core Team, 2015). The correlation plot was generated using R package "corrplot" version 0.77 with hierarchical clustering Ward.D2 method (Wei and Simko, 2016). Logistic regression was applied to understand the association between the crop productivity and the other 26 soil characteristics. Significance of Pearson's correlation and logistic regression were determined at  $p$ -value of 0.05.

### Shotgun Sequencing and Data Archive

DNA was extracted from 200 mg subsamples of lyophilized, milled (model M20, Ika Works, Wilmington-NC) soil using the FastDNA SPIN Kit for Soil (MP Biomedicals, Solon, OH, U.S.A.) and further purified using the MicroElute DNA Clean-up Kit (Omega Bio-tek, Norcross, GA, U.S.A.). Twelve bulk soil DNA samples were deep shotgun sequenced in pairs through six lanes using Illumina HiSeq2000 (Roy J. Carver Biotechnology Center at the University of Illinois) using TruSeq SDS sequencing kit version 3 according to the manufacturers' protocols. The 12 shotgun sequencing data were deposited in MG-RAST server (Table 1).

### Metagenome Analyses and Metagenome-Wide Association Study (MWAS)

Raw reads were uploaded to the MG-RAST server (Meyer et al., 2008), and quality-controlled reads were analyzed for taxa abundance using the best hit classification to the M5NR

**TABLE 1 | Sampling information and metagenomic statistics of 12 bulk soil samples in Illinois.**

Sample ID	MG-RAST ID	City (IL, USA)	Latitude	Longitude	Productivity*	Raw Reads	Filtered (%)	Post QC Reads	Alpha-diversity (species)
1	4502923.3	Coulterville	38.11880	-89.37656	High	62,087,940	10.40	55,616,919	934.208
2	4502929.3	Coulterville	38.11917	-89.37632	Low	90,120,459	15.30	76,341,732	935.288
3	4502930.3	Greenfield	39.22606	-90.19410	High	124,404,467	8.60	113,747,535	908.248
4	4502931.3	Greenfield	39.22540	-90.19423	Low	93,849,320	4.90	89,256,102	921.719
5	4502932.3	Greenfield	39.22709	-90.19435	High	106,333,248	4.60	101,477,796	889.839
6	4502933.3	Greenfield	39.22728	-90.19357	Low	67,442,497	17.80	55,471,025	946.078
7	4502541.3	Auburn	39.35987	-89.45862	High	78,573,006	17.40	64,893,620	861.629
8	4502539.3	Auburn	39.35907	-89.45862	Low	67,333,977	7.70	62,155,903	897.977
9	4502926.3	Mansfield	40.13227	-88.27992	High	97,576,070	3.70	93,936,766	875.722
10	4502925.3	Mansfield	40.13185	-88.28042	Low	89,341,482	4.80	85,056,308	892.166
11	4502927.3	Urbana	40.04660	-88.13072	High	111,108,508	4.40	106,242,805	919.205
12	4502928.3	Urbana	40.04661	-88.12943	Low	81,221,533	4.40	77,618,380	912.711

\*Distance calculation between paired sampling locations was 46 m (1 vs. 2), 74 m (3 vs. 4), 70 m (5 vs. 6), 89 m (7 vs. 8), 63 m (9 vs. 10), and 11 m (11 vs. 12) based on <http://andrew.hedges.name/experiments/haversine/> and <http://www.movable-type.co.uk/scripts/latlong.html>. Samples were collected: 1–6 = 29 Oct. 2010, 7–8 = 01 Nov. 2010, and 9–12 = 02 Nov. 2010.

database (Wilke et al., 2012) and functional gene abundance using hierarchical classification to the Subsystems. Compared to the default parameters of MG-RAST server, higher stringent parameters were set at a minimum length of 30 nucleotides, a cutoff at 80% of identity, and a cutoff at an  $E$ -value of  $1 \times 10^{-9}$  in this study (Wilke et al., 2016). The first two principal coordinates (PCo1 and PCo2) generated by MG-RAST were also analyzed using Pearson's correlation to the 26 soil characteristics and the logistic regression to the crop productivity. Significance of Pearson's correlation and logistic regression were determined at  $p$ -value of 0.05. The highest-correlated soil characteristics to PCo1 and PCo2 (water pH and productivity) were labeled in the principal coordinate analysis (PCoA) plot generated in MG-RAST using normalized abundance. R package "mvtnorm" version 1.0–5 (Genz et al., 2016) and "ellipse" version 0.3–8 (Murdoch and Chow, 2013) were used to generate 90% confidence intervals for the high and low productivity samples. Constrained correspondence analysis or the canonical correspondence (CCA) with environmental vector fitting was performed using R package "vegan" version 2.4-1 (Oksanen et al., 2016). Since there is a multicollinearity problem among the 26 soil characteristics, the variance inflation factor (VIF) for each variable was estimated using R package "faraway" version 1.0.7 (Faraway, 2016). Because 12 soil samples cannot provide enough degree of freedom for a full model with 26 variables, variables (including water pH, crop productivity, and others with a VIF below 5) were used in vector fitting in the CCA, regarding as the VIF-based model. Akaike's information criterion (AIC) was applied to select useful soil characteristics in the AIC-based model for vector fitting. Permutation tests by marginal effects with 1,000 permutations were applied to estimate the significance of the AIC-based model and the VIF-based model. MWAS was performed to find significant associations between taxa and crop productivity using Wilcoxon rank sum test after filtering taxa with raw abundance below 12 counts across 12 soil samples from MG-RAST (Karlsson et al., 2013; Wang and Jia, 2016). Significant associations were determined

at Benjamini-Hochberg adjusted  $p$ -value or false discovery rate (FDR) at  $\alpha = 0.05$ .

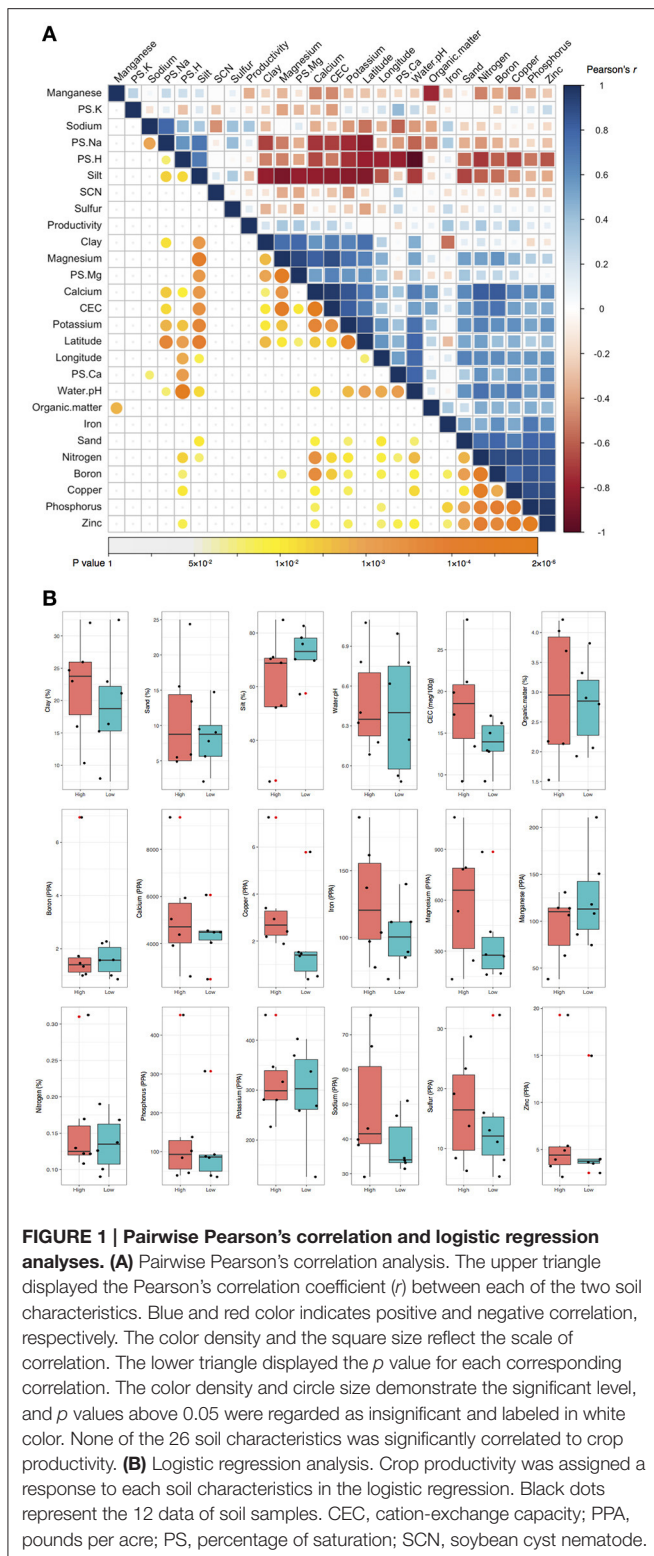
## Machine Learning Using Random Forest (RF)

The RF machine learning was performed in R using "ranger" package (Wright and Ziegler, 2015). A total of 66 possible combinations, which included 10 samples as the training set and the remaining two samples as the testing set ( $C_2^{12}$ ), were iterated in each run. Within each run, the number of trees (num.tree) was set at 500 to build the RF model. The number of variables/taxa that could be selected in each splitting node (mtry) was set in a range one tenth of the maximum taxon number at each taxonomic level (32 for phylum, 66 for class, 134 for order, 175 for family, 149 for genus, and 215 for species), and all of the 10 mtry parameters were run. The importance of each taxon was estimated using the Gini index, and the prediction accuracy was determined by  $\frac{\text{True Positive (TP)} + \text{True Negative (TN)}}{\text{TP} + \text{TN} + \text{False Positive} + \text{False Negative}}$ . A total of 100 runs for each mtry parameter were performed for each taxonomy hierarchy.

## RESULTS

### Soil Characteristics Were Not Associated with Crop Productivity

In the pairwise Pearson's correlation analysis, the 26 soil characteristics formed one negative correlated block and two positive correlated blocks (Figure 1A). In the negative correlated block, water pH significantly correlated with latitude, longitude, and several elements as well as clay and sand soil types. The first positive correlated block included several elements (such as boron, phosphorus, and zinc) and sand soil type. The second positive correlated block contained other elements (such as Ca, Mg, and K) with CEC, latitude, and longitude. A weaker positive correlation between the first and the second positive correlation block was significant as well. The preliminary



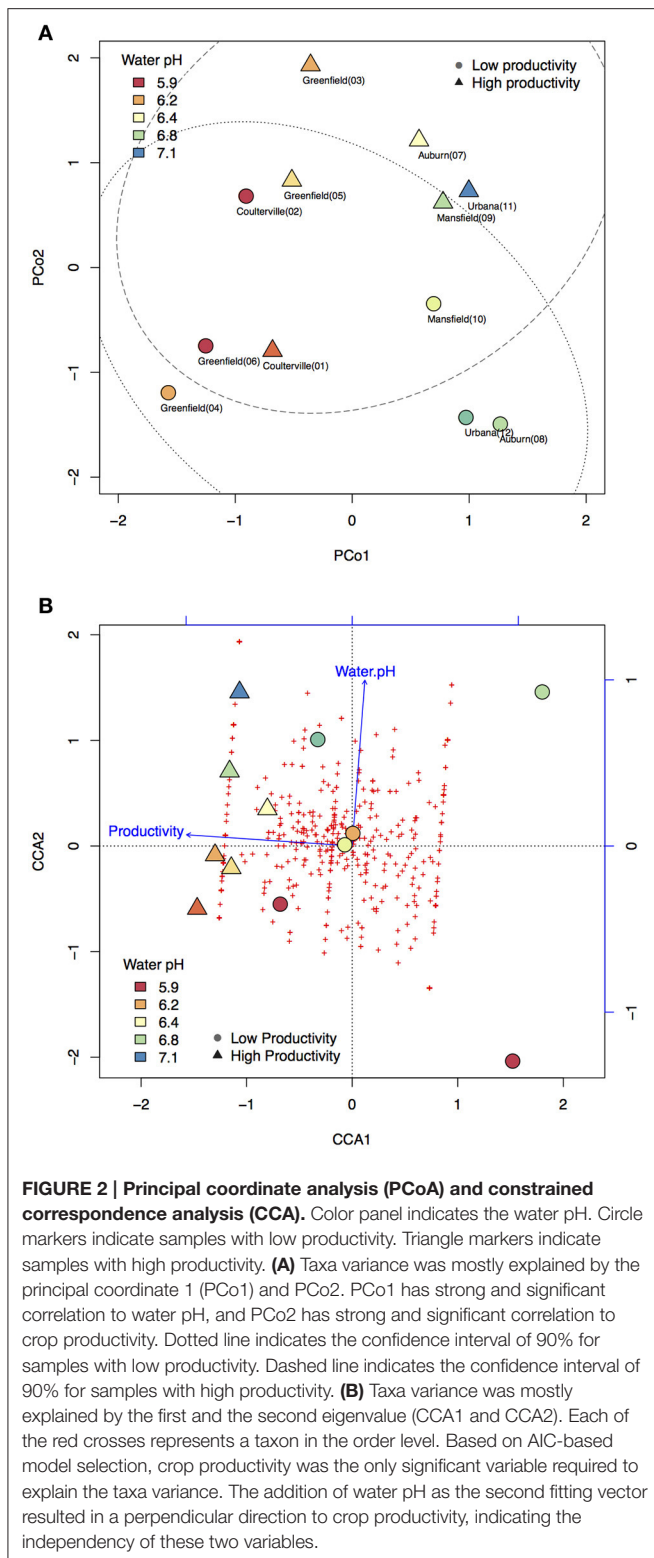
analysis using Person's correlation demonstrated there was no significant association between crop productivity to any of the 26 soil characteristics (Figure 1A). To further confirm the observation, logistic regression was applied to understand the

association between the binary crop productivity and each soil characteristic. The results of logistic regression support the observation of pairwise Pearson's correlation that none of the 26 soil characteristics was significantly associated with crop productivity (Figure 1B). The results together indicate that neither the abundance of SCN, which is the primary soil pathogen for soybean production in Illinois (Niblack and Riggs, 2015), nor the 25 soil characteristics was related to crop productivity difference. Other abiotic and biotic factors might associate with the productivity difference including the composition of bulk soil microbiome.

## Microbiome Composition Significantly Associated with Crop Productivity

To understand if microbiome compositions in bulk soil samples, shotgun sequencing was chosen to profile the soil microbiome. Pearson's correlation analysis indicated the highest variance (PCo1) in microbiome composition displayed a strong correlation to water pH ( $p = 0.0001$ ). PCo1 separated acidic ( $pH < 6.3$ ) and non-acidic ( $pH > 6.3$ ) soil samples into two spaces, accounting for the largest 18% of taxa variance (Figure 2A). On the other hand, PCo2 explained second largest 13% of taxa variance. Moreover, PCo2 was the only factor that displayed a significant correlation to crop productivity ( $p = 0.0126$ ) and the association was further confirmed by logistic regression ( $p = 0.0466$ ). There were three samples (Greenfield 03, Auburn 07, and Urbana11) from high productivity areas and three samples (Greenfield 04, Auburn 08, Urbana 12) from low productivity areas located outside of the 90% confidence intervals (Figure 2A). Our observations in the PCoA were consistent to several publications that described the association between soil pH and microbiome (Rousk et al., 2010; Rascovan et al., 2016). However, the multicollinearity problem (such as the strong correlations between water pH and other characteristics) raises the possibility that water pH and crop productivity might be a confounding factor to PCo1 and PCo2, respectively.

In order to confirm the PCoA results, VIF-based model and AIC-based model were subjected to CCA with permutation supports. In the VIF-based model, six soil characteristics with VIF below 5 were included in the CCA model, including crop productivity (VIF: 1.05), organic matter (VIF: 1.19), PS.K (VIF: 1.26), SCN (VIF: 1.33), sulfur (VIF: 1.26), and water pH (VIF: 1.25) (Figure S2A). Among these six soil characteristics, permutation test identified crop productivity as the only significant explanatory variable ( $p = 0.014$ ). When PCo1 and PCo2 were used as fitting vector in the VIF-based CCA, the result supported that PCo1 is closer to water pH while PCo2 is closer to crop productivity (Figure S2B). On the other hand, AIC-based model selection suggested the crop productivity as the only variable that needs to be included in the CCA to explain the taxa variance (AIC: 93.98,  $p = 0.055$ ). The addition of water pH into the vector fitting CCA resulted in a perpendicular direction to crop productivity (Figure 2B), and when PCo1 and PCo2 were added into the CCA, a consistent result that PCo2 is closer to crop productivity can be observed (Figure S2C). ANOVA model comparison between VIF-based model (six explanatory variables)



and AIC-based model (one explanatory variable) failed to reject the smaller AIC-based CCA model ( $p = 0.709$ ). The results of CCA vector fitting indicated crop productivity is the major factor and explained 16% of taxa variance while water pH explained 8%

of the taxa variance. Both PCoA and CCA demonstrated the taxa variance of microbiome composition was associated with crop productivity, which indicated the possibility that some taxa might vary between high and low productivity areas in the six fields.

## Metagenomic Analyses and MWAS

There were more than 55 million sequencing reads for each sample that passed quality control and MG-RAST estimated an alpha-diversity around 861–935 for the 12 samples (Table 1). Welch independent two sample  $t$ -test suggested no significant difference for the alpha-diversity average from high and low productivity areas ( $p = 0.20$ ). In order to identify what taxa in the bulk soil microbiome differ between the high and low productivity samples, we applied MWAS using Wilcoxon rank sum test. The abundance of a bacterial order *Planctomycetales* and a eukaryotic phylum *Streptophyta* were found significantly higher in low productivity areas. Both phyla can be detected more than three times at different hierarchies in the same taxonomy lineage (Table 2). Other significant taxa repeatedly identified by Wilcoxon rank sum test included a bacterial genus *Bradyrhizodium*, a bacterial class *Gammaproteobacteria*, an unclassified class in the fungal phylum *Ascomycota*, and an unclassified class in the eukaryotic phylum *Streptophyta* (Figure 3). The abundance of *Bradyrhizodium* and *Gammaproteobacteria* were generally higher in high productivity areas, while the abundance of *Ascomycota* was higher in low productivity areas (Figure 3, Table 2). In contrast to the success on MWAS, the association analysis between functional gene abundance and crop productivity failed to find anything significant results (data not shown).

## Productivity Prediction by Using RF Machine Learning

To understand if the microbiome composition in the bulk soils could be informative to predict crop productivity, we adapted RF machine learning and estimated the prediction accuracy at each taxonomic level with 10 different variables/taxa (mtry) included in the RF model. While most of the predictions had low accuracies, we found at the order level with all variables in the model reached the best prediction accuracy at 0.787 (Figure 4A). We further computed the taxon importance assigned by the RF model at the order level, and the results indicated most important taxon was the *Actinomycetiales*, and followed by *Nostocales* and *Rhizobiales* (Figure 4B). While the *Nostocales* in the *Cyanobacteria* phylum was only found to be important by RF machine learning, both *Actinomycetiales* and *Rhizobiales* were identified to be significant in the MWAS (Table 2). In addition to *Rhizobiales*, other taxa such as an unclassified order under the *Gammaproteobacteria* in the *Proteobacteria* phylum was also found to be important by MWAS and RF machine learning (Figure 4B).

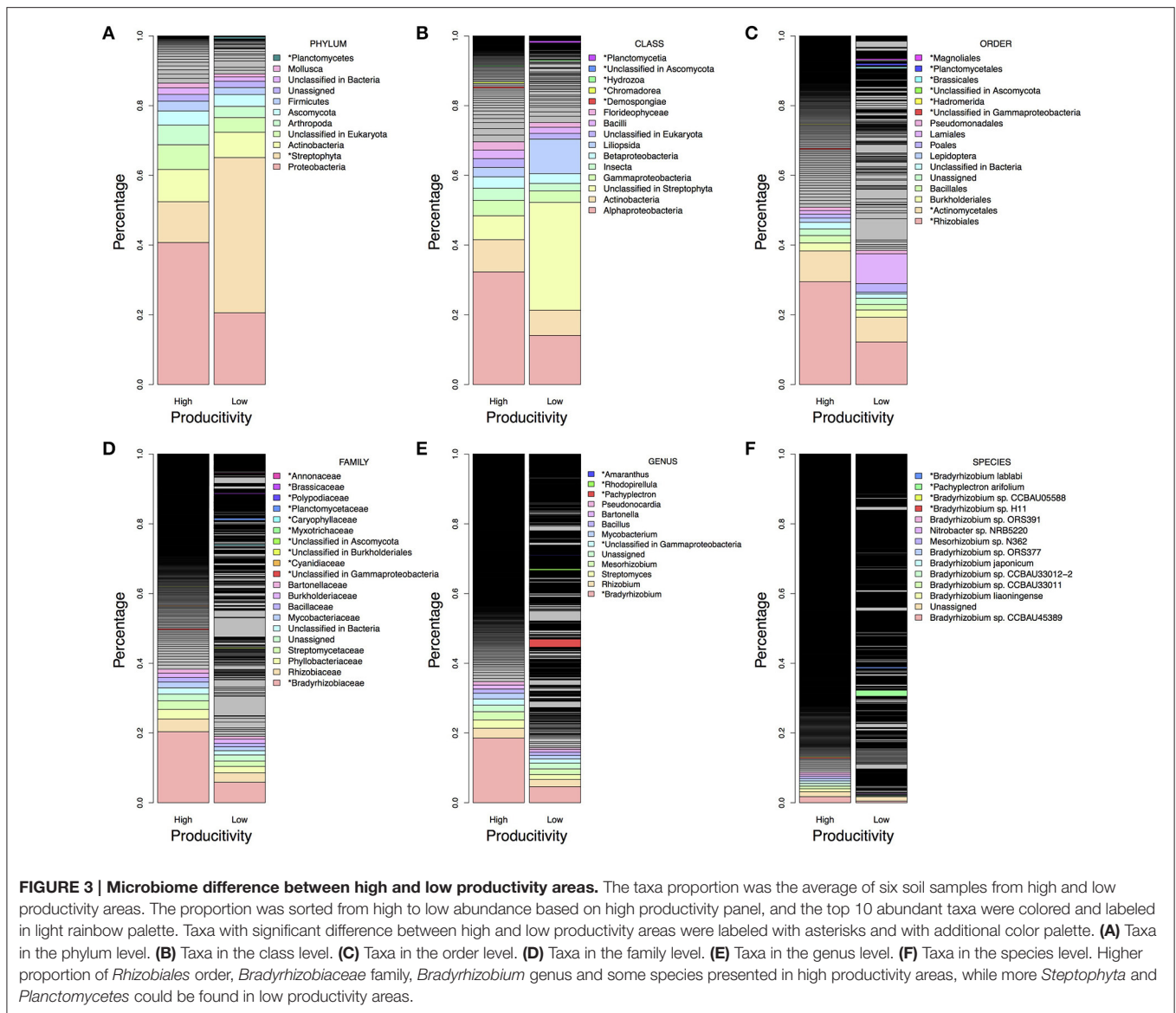
## DISCUSSION

Crop productivity is a quantitative trait determined by a variety of factors (Van Roekel et al., 2015). Abiotic soil characteristics such as water and nitrogen availability are well-known for being

**TABLE 2 | Significant taxon abundance differed between high and low productivity areas using Wilcoxon rank sum test.**

Domain	Phylum	Class	Order	Family	Genus	Species	High Prod#	Low Prod#	FDR*
Bacteria	Actinobacteria	Actinobacteria	Actinomycetales				422	489	4.46E-02
Bacteria	<b>Planctomycetes</b>	<b>Planctomycetia</b>					1	38	2.51E-02
Bacteria	<b>Planctomycetes</b>	<b>Planctomycetia</b>	<b>Planctomycetales</b>				1	38	2.51E-02
Bacteria	<b>Planctomycetes</b>	<b>Planctomycetia</b>	<b>Planctomycetales</b>	Planctomycetaceae			1	38	2.51E-02
Bacteria	<b>Planctomycetes</b>	<b>Planctomycetia</b>	<b>Planctomycetales</b>	Planctomycetaceae	Rhodopirellula		0	31	9.47E-03
Bacteria	<b>Proteobacteria</b>	<b>Alphaproteobacteria</b>	<b>Rhizobiales</b>				1413	844	4.11E-02
Bacteria	<b>Proteobacteria</b>	<b>Alphaproteobacteria</b>	<b>Rhizobiales</b>	<b>Bradyrhizobiaceae</b>			983	410	2.60E-02
Bacteria	<b>Proteobacteria</b>	<b>Alphaproteobacteria</b>	<b>Rhizobiales</b>	<b>Bradyrhizobiaceae</b>	<b>Bradyrhizobium</b>		913	336	2.60E-02
Bacteria	<b>Proteobacteria</b>	<b>Alphaproteobacteria</b>	<b>Rhizobiales</b>	<b>Bradyrhizobiaceae</b>	<b>Bradyrhizobium</b>	<i>Bradyrhizobium lablabi</i>	1	40	2.08E-02
Bacteria	<b>Proteobacteria</b>	<b>Alphaproteobacteria</b>	<b>Rhizobiales</b>	<b>Bradyrhizobiaceae</b>	<b>Bradyrhizobium</b>	<i>Bradyrhizobium sp. COBAU 05588</i>	19	2	3.61E-02
Bacteria	<b>Proteobacteria</b>	<b>Alphaproteobacteria</b>	<b>Rhizobiales</b>	<b>Bradyrhizobiaceae</b>	<b>Bradyrhizobium</b>	<i>Bradyrhizobium sp. HT1</i>	21	4	3.61E-02
Bacteria	<b>Proteobacteria</b>	<b>Alphaproteobacteria</b>	<b>Rhizobiales</b>	<b>Bradyrhizobiaceae</b>	<b>Bradyrhizobium</b>		10	23	1.75E-02
Bacteria	<b>Proteobacteria</b>	Betaproteobacteria	Burkholderiales	Unclassified			19	8	3.98E-02
Bacteria	<b>Proteobacteria</b>	<b>Gammaproteobacteria</b>	<b>Unclassified</b>	Unclassified			19	8	3.98E-02
Bacteria	<b>Proteobacteria</b>	<b>Gammaproteobacteria</b>	<b>Unclassified</b>	Unclassified			19	8	3.98E-02
Bacteria	<b>Proteobacteria</b>	<b>Gammaproteobacteria</b>	<b>Unclassified</b>	Unclassified	Unclassified		19	7	1.53E-02
Eukaryota	<b>Ascomycota</b>	Leotiomycetes	Unclassified	Myxotrichaceae			3	10	2.48E-02
Eukaryota	<b>Ascomycota</b>	<b>Unclassified</b>					3	9	2.39E-02
Eukaryota	<b>Ascomycota</b>	<b>Unclassified</b>					3	9	2.39E-02
Eukaryota	<b>Ascomycota</b>	<b>Unclassified</b>					3	9	2.39E-02
Eukaryota	Cnidaria	Hydrozoa		Unclassified			13	26	4.27E-02
Eukaryota	Nematoda	Chromadorea					21	8	3.47E-02
Eukaryota	Porifera	Demospongiae					23	6	1.87E-02
Eukaryota	Porifera	Demospongiae	Hadromerida				13	1	1.46E-02
Eukaryota	<b>Streptophyta</b>	Lillopsida	Asparagales	Orchidaceae	<i>Pachyplectron</i>		2	168	2.60E-02
Eukaryota	<b>Streptophyta</b>	Lillopsida	Asparagales	Orchidaceae	<i>Pachyplectron</i>		2	168	2.60E-02
Eukaryota	<b>Streptophyta</b>	Polyplodopsida	Polyplodiales	Polyplodaceae			1	11	2.48E-02
Eukaryota	<b>Streptophyta</b>	<b>Unclassified</b>	Brassicales				1	33	2.05E-02
Eukaryota	<b>Streptophyta</b>	<b>Unclassified</b>	Brassicales				0	20	2.84E-02
Eukaryota	<b>Streptophyta</b>	<b>Unclassified</b>	Caryophyllales	Amaranthaceae	<i>Amaranthus</i>		0	13	2.69E-02
Eukaryota	<b>Streptophyta</b>	<b>Unclassified</b>	Caryophyllales	Caryophyllaceae			2	27	1.85E-02
Eukaryota	<b>Streptophyta</b>	<b>Unclassified</b>	Magnoliales				1	26	4.61E-03
Eukaryota	<b>Streptophyta</b>	<b>Unclassified</b>	Magnoliales	Annonaceae			0	14	2.80E-02
Eukaryota	Unclassified	Bangiophyceae	Cyanidiales	Cyanidiaceae			13	1	1.98E-02

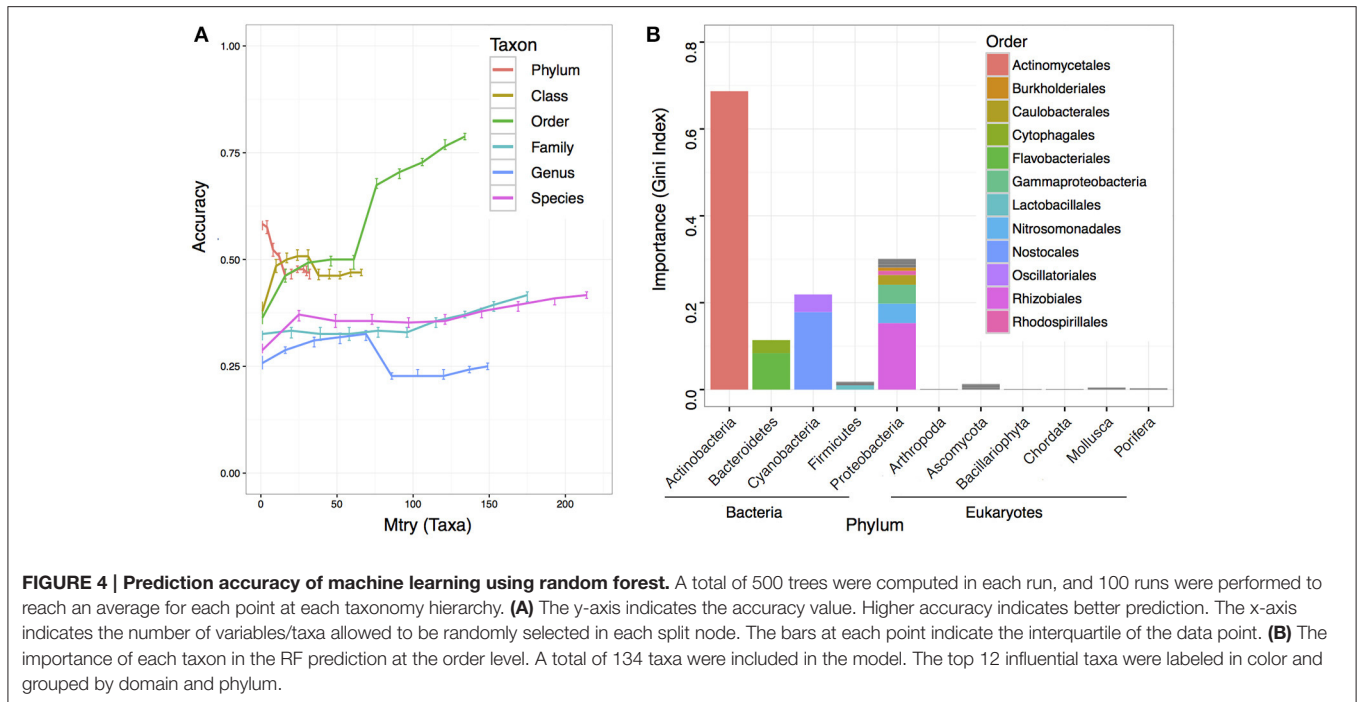
\* FDR, False Discovery Rate.  
 # High Prod: the sum of a taxon abundance in the six soil samples from high productivity areas. Low Prod: the sum of a taxon abundance in the six soil samples from low productivity areas.  
 Bold, taxa detected to be significant for at least three times in MMAS.



productivity-limiting factors (Durán et al., 2014), and weather conditions such as rainfall or temperature may significantly impact on crop productivity. On the other hand, biotic features such as pathogen, pest (Hartman et al., 2015), and beneficial symbiosis such as root nodulation (Tkacz and Poole, 2015) are also involved in yield performance. Moreover, an important biotic feature is the genetics of the crop variety (the genotype), which includes the genetics of the photosynthesis and productivity performance (Dhanapal et al., 2016; Li et al., 2016), the genetics of water and nitrogen utilizing efficiency (Dhanapal et al., 2015; Chen et al., 2016), the genetics of disease and pest resistance (Chang et al., 2016; Reville, 2016), and the genetic influence on structuring the rhizosphere microbiome (Jin et al., 2009; Babujia et al., 2016; de Almeida Lopes et al., 2016). In our study, we discovered additional factors underlying crop productivity when the above factors were identical or similar. Our

experimental design ensured each pair of two areas in the same agronomic field (with no known difference in environmental conditions such as rainfall) received the same management by the farmers (with identical crop genetic variety and agricultural applications such as fertilization) at each of the six locations in Illinois, and because there were known differences in diseases or pests reported in the sampling season between the two areas, we hypothesized that bulk soils with some unevenly distributed abiotic or biotic factors might be a cause of crop productivity difference.

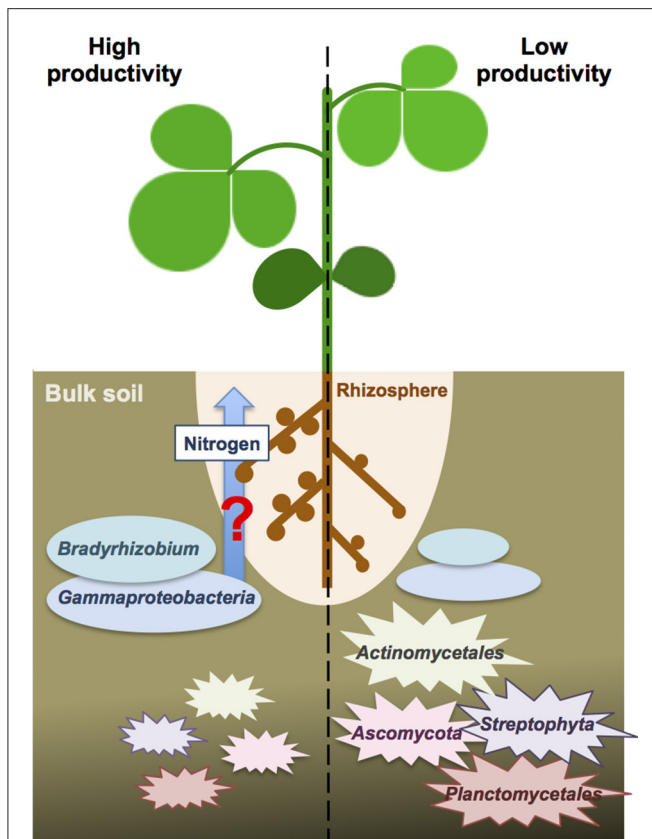
Twenty-six soil characteristics were quantified in the soil samples; however, none of these displayed significant correlations to crop productivity. Nonetheless, other abiotic characteristics such as soil physical compaction, soil moisture, or drainage difference between areas should be considered as potential abiotic factors because the soil system is much more complicated than



the 26 characteristics included in the studies. While the 26 soil characteristics led to no significant result, PCoA and CCA both suggested crop productivity was an important explanatory variable that accounted for the taxa variance of microbiome composition. In other words, microbial difference in the bulk soil samples might associate with crop productivity. To further understand which taxa in the microbiome associated with crop productivity difference, MWAS was applied to dissect the microbiome by individual taxa at different taxonomy levels (from Phylum to Species). Three bacterial taxa (*Bradyrhizodium*, *Gammaproteobacteria*, and *Planctomycetales*) and two eukaryotic taxa (*Ascomycota*, and *Streptophyta*) were found significant for at least three times in the same hierarchical lineage. Interestingly, most of these taxa were related to nitrogen utility one way or another. It was suggested that the abundance of *Proteobacteria* was higher when nitrogen is more available (Fierer et al., 2012), and indeed, we observed higher abundance of *Proteobacteria* in the higher productivity areas (Figure 4). Both *Bradyrhizodium* and *Gammaproteobacteria* belong to *Proteobacteria* phylum, and both taxa related to nodulation. Bacteria in the *Bradyrhizodium* genus are well known for their symbiosis roles with legumes to fix nitrogen and benefit crop productivity (Durán et al., 2014). The interactions between soybean and different *Bradyrhizodium* strains on crop productivity were shown to be significant (Zimmer et al., 2016). The distribution and diversity of *Bradyrhizodium* strains in the U.S.A. were also reported to vary geographically (Shiro et al., 2013). Most bacteria in the *Bradyrhizodium* genus were reported to have nitrogen-fixation genes (Durán et al., 2014), and their ability to fix nitrogen for more than half of soybean N demand was recognized (Salvagiotti et al., 2008).

Although bacteria in the *Gammaproteobacteria* class may not have independent nitrogen fixation ability, some bacteria such as those in the genus *Klebsiella* were able to colonize peanut nodules in the presence of *Bradyrhizodium* species (Ibáñez et al., 2009), and some were assumed to be disease-suppressive or health-promotive (Berendsen et al., 2012). The abundances of beneficial rhizobia (*Bradyrhizodium* and *Gammaproteobacteria*) were generally higher in the high productivity areas (Table 2). On the other hand, the order *Planctomycetales* belongs to a special group of bacteria that contain no peptidoglycan and mainly reproduce by budding. Classification for *Planctomyces* situates the group in between Bacteria and Archaea because some *Planctomyces* have eukaryotic characteristics, such as a membrane-bound nucleoid and the ability to synthesize sterol. Moreover, some *Planctomyces* performs anaerobic oxidation of ammonium to dinitrogen in specialized vesicles called anamoxosomes, which might reduce nitrogen availability in the bulk soils (Fuerst and Sagulenko, 2011). While *Streptophyta* is the phylum of land plants and algae that may directly compete for nitrogen availability with crops (Leliaert et al., 2012; Becker, 2013), *Ascomycota* is the largest fungal phylum that contains diverse soil-borne plant pathogenic fungi, but the phylum also contains many non-pathogens and additional studies that focus on what taxa are beneficial or detrimental for crop productivity are required as direct evidence. The *Actinomycetales* order also contains plant pathogens such as potato scab pathogen, *Streptomyces scabies*. While *Actinomycetales* was detected significantly only at the order level (not significant for *Actinobacteria* phylum and class), the group of bacteria was weighted as the most important taxon in RF machine learning. Although the abundance of *Actinomycetales* was higher in low





**FIGURE 5 | The speculative scheme for the interactions between crop productivity and bulk soil microbiome.** The beneficial taxa of *Bradyrhizodium* and *Gammaproteobacteria* may interact with crops regarding nitrogen fixation and nodule formation to enhance nitrogen availability. On the other hand, higher abundance of *Streptophyta* and *Planctomycetes* in the bulk soil may compete nitrogen with crops and reduce nitrogen availability. Higher abundance of *Actinomycetales* and *Ascomycota* may increase biotic stress to crops.

productivity areas similar to the abundance of *Ascomycota*, *Planctomycetales*, and *Streptophyta* (Table 2), which gives an intuition that *Actinomycetales* may be detrimental to crop health, some studies reported co-inoculation benefits of *Actinomycetes* with *Bradyrhizodium japonicum* that promoted soybean growth (Soe et al., 2012; Nimnoi et al., 2014). As the *Actinomycetales* order still includes too many taxa to be conclusive, additional studies that focus on what taxa in the *Actinomycetales* order are beneficial or detrimental for crop productivity are required as direct evidence.

Because the nitrogen content in the bulk soils was not significantly different between high and low productivity areas (Welch independent two sample *t*-test,  $p = 0.53$ ), we speculated the different crop productivity might associate with nitrogen fixation activities inside the nodules. Higher abundance of *Bradyrhizodium* and *Gammaproteobacteria* may contribute to higher abundance of beneficial rhizobia in the rhizosphere. Indeed, it has been proposed that the rhizosphere microbiome of soybean was specialized from bulk soil microbiome to enhance soybean growth and nutrient utilities (Mendes et al.,

2014). Unfortunately, the association results between functional gene abundance and crop productivity failed to identify any significant result, nor for nitrogen metabolism-related genes. But because DNA-based metagenomic study does not provide direct expression evidence, even if nitrogen metabolism-related genes appeared to be more abundant, a meta-transcriptomic study is still needed to ensure nitrogen metabolism-related genes indeed have differential expression in one condition over another. Advanced studies focus on finding evidence for recruitment of these beneficial rhizobia from the bulk soil into the rhizosphere, and finding proofs for these beneficial rhizobia provide better nitrogen fixation or stimulate more nodules will provide a new insight to the crop productivity difference in a field (Figure 5).

In addition to identify taxa relating to crop productivity difference, our study applied machine learning prediction for the first time using soil microbiome composition. The result demonstrated that microbiome composition indeed could be useful for crop productivity prediction. While the prediction model with a small training set resulted in lower accuracy compare to machine learning prediction in human diseases (generally included 100–300 samples; Pasoli et al., 2016), we expect the accuracy would be improved with a larger sample size. Nonetheless, because soil microbiome could be far more complicated and diverse than human microbiome, limited sequencing depth to detect rare taxa and the reproducibility under the challenge of highly varied environmental factors will be the technical bottlenecks.

## CONCLUSION

We identified four groups of bacteria and two groups of eukaryotes that were significantly associated with crop productivity. The use of a RF model successfully predicted crop productivity at an accuracy of 0.79. With the active progress in metagenome annotation, statistical comparison, and computational power to handle high dimensional data, we expect that MWAS and machine learning will provide a new understanding on how microbial communities interact with crops and deliver direct benefits to agriculture.

## DISCLOSURE

Trade and manufacturers' names are necessary to report factually on available data; however, the USDA neither guarantees nor warrants the standard of the product, and the use of the name by USDA implies no approval of the product to the exclusion of others that may also be suitable.

## AUTHOR CONTRIBUTIONS

HC: Analyzed and interpreted the data and developed the draft of the manuscript; JH: Interpreted the data and helped to write the manuscript; CB: Collected raw materials for experiment and helped to write the manuscript; GH: Coordinated the research and helped to write the manuscript.

## FUNDING

Research reported in this publication was supported by the Illinois Soybean Board and the USDA Agricultural Research Service.

## ACKNOWLEDGMENTS

We thank the grower cooperators for allowing soil sampling of their fields and the USDA Agricultural Research Service for partial funding support.

## REFERENCES

- Babujia, L. C., Silva, A. P., Nakatani, A. S., Cantao, M. E., Vasconcelos, A. T. R., Visentainer, J. V., et al. (2016). Impact of long-term cropping of glyphosate-resistant transgenic soybean [*Glycine max* (L.) Merr.] on soil microbiome. *Transgenic Res.* 25, 425–440. doi: 10.1007/s11248-016-9938-4
- Becker, B. (2013). Snow ball earth and the split of Streptophyta and Chlorophyta. *Trends Plant Sci.* 18, 180–183. doi: 10.1016/j.tplants.2012.09.010
- Berendsen, R. L., Pieterse, C. M., and Bakker, P. A. (2012). The rhizosphere microbiome and plant health. *Trends Plant Sci.* 17, 478–486. doi: 10.1016/j.tplants.2012.04.001
- Chang, H.-X., Lipka, A., Domier, L. L., and Hartman, G. L. (2016). Characterization of disease resistance loci in the USDA soybean germplasm collection using genome-wide associations. *Phytopathology* 106, 1139–1151. doi: 10.1094/PHYTO-01-16-0042-FI
- Chen, W., Yao, Q. M., Patil, G. B., Agarwal, G., Deshmukh, R. K., Lin, L., et al. (2016). Identification and comparative analysis of differential gene expression in soybean leaf tissue under drought and flooding stress revealed by RNA-Seq. *Front. Plant Sci.* 7:1044. doi: 10.3389/fpls.2016.01044
- de Almeida Lopes, K. B., Carpentieri-Pipolo, V., Oro, T. H., Stefani Pagliosa, E., and Degrossi, G. (2016). Culturable endophytic bacterial communities associated with field-grown soybean. *J. Appl. Microbiol.* 120, 740–755. doi: 10.1111/jam.13046
- Debenport, S. J., Assigbetse, K., Bayala, R., Chapuis-Lardy, L., Dick, R. P., and Gardener, B. B. M. (2015). Association of shifting populations in the root zone microbiome of millet with enhanced crop productivity in the Sahel Region (Africa). *Appl. Environ. Microb.* 81, 2841–2851. doi: 10.1128/AEM.04122-14
- Dhanapal, A. P., Ray, J. D., Singh, S. K., Hoyos-Villegas, V., Smith, J. R., Purcell, L. C., et al. (2015). Genome-wide association analysis of diverse soybean genotypes reveals novel markers for nitrogen traits. *Plant Genome* 8. doi: 10.3835/plantgenome2014.11.0086
- Dhanapal, A. P., Ray, J. D., Singh, S. K., Hoyos-Villegas, V., Smith, J. R., Purcell, L. C., et al. (2016). Genome-wide association mapping of soybean chlorophyll traits based on canopy spectral reflectance and leaf extracts. *BMC Plant Biol.* 16:174. doi: 10.1186/s12870-12016-10861-x
- Durán, D., Rey, L., Sanchez-Canizares, C., Jorrián, B., Imperial, J., and Ruiz-Argüeso, T. (2014). “Biodiversity of slow-growing Rhizobia: the genus *Bradyrhizobium*,” in *Beneficial Plant-Microbial Interactions: Ecology and Applications*, eds B. González and J. Gonzalez-López (Boca Raton, FL: CRC Press), 20–46.
- Fierer, N., Lauber, C. L., Ramirez, K. S., Zaneveld, J., Bradford, M. A., and Knight, R. (2012). Comparative metagenomic, phylogenetic and physiological analyses of soil microbial communities across nitrogen gradients. *ISME J.* 6, 1007–1017. doi: 10.1038/ismej.2011.159
- Fuerst, J. A., and Sagulenko, E. (2011). Beyond the bacterium: planctomycetes challenge our concepts of microbial structure and function. *Nat. Rev. Microbiol.* 9, 403–413. doi: 10.1038/nrmicro2578
- Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., et al. (2016). *Mvtnorm: Multivariate Normal and t Distributions*. R package version 1.0-5. Available online at: <http://cran.r-project.org/package=mvtnorm>.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2017.00519/full#supplementary-material>

**Figure S1 | Sampling location in the Illinois state.** A collaborator in each location (two collaborators in the Greensfield) provided two samples from each field, one with high productivity and one with low productivity.

**Figure S2 | Constrained correspondence analysis (CCA).** (A) VIF-based CCA model, which includes six soil characteristics as explanatory variables. (B) VIF-based CCA model with PCo1 and PCo2 included. (C) AIC-based CCA model with PCo1 and PCo2 included.

- Hartman, G. L., Rupe, J. C., Sikora, E. F., Domier, L. L., Davis, J. A., and Steffey, K. L. (2015). *Compendium of Soybean Diseases and Pests, 5th Edn.* St. Paul, MN: APS Press. American Phytopathological Society.
- Ibáñez, F., Angelini, J., Taurian, T., Tonelli, M. L., and Fabra, A. (2009). Endophytic occupation of peanut root nodules by opportunistic Gammaproteobacteria. *Syst. Appl. Microbiol.* 32, 49–55. doi: 10.1016/j.syapm.2008.10.001
- Jin, J., Wang, G. H., Liu, X. B., Liu, J. D., Chen, X. L., and Herbert, S. J. (2009). Temporal and spatial dynamics of bacterial community in the rhizosphere of soybean genotypes grown in a black soil. *Pedosphere* 19, 808–816. doi: 10.1016/S1002-0160(09)60176-4
- Karlsson, F. H., Tremaroli, V., Nookaew, I., Bergstrom, G., Behre, C. J., Fagerberg, B., et al. (2013). Gut metagenome in European women with normal, impaired and diabetic glucose control. *Nature* 498, 99–105. doi: 10.1038/nature12198
- Lakshmanan, V. (2015). Root microbiome assemblage is modulated by plant host factors. *Adv. Bot. Res.* 75, 57–79. doi: 10.1016/bs.abr.2015.09.004
- Le Chatelier, E., Nielsen, T., Qin, J. J., Prifti, E., Hildebrand, F., Falony, G., et al. (2013). Richness of human gut microbiome correlates with metabolic markers. *Nature* 500, 541–549. doi: 10.1038/nature12506
- Leliaert, F., Smith, D. R., Moreau, H., Herron, M. D., Verbruggen, H., Delwiche, C. F., et al. (2012). Phylogeny and molecular evolution of the green algae. *Crit. Rev. Plant Sci.* 31, 1–46. doi: 10.1080/07352689.2011.615705
- Li, H. Y., Yang, Y. M., Zhang, H. Y., Chu, S. S., Zhang, X. G., Yin, D. M., et al. (2016). A genetic relationship between phosphorus efficiency and photosynthetic traits in soybean as revealed by QTL analysis using a high-density genetic map. *Front. Plant Sci.* 7:924. doi: 10.3389/fpls.2016.00924
- Mendes, L. W., Kuramae, E. E., Navarrete, A. A., van Veen, J. A., and Tsai, S. M. (2014). Taxonomical and functional microbial community selection in soybean rhizosphere. *ISME J.* 8, 1577–1587. doi: 10.1038/ismej.2014.17
- Meyer, F., Paarmann, D., D’Souza, M., Olson, R., Glass, E. M., Kubal, M., et al. (2008). The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 9:386. doi: 10.1186/1471-2105-9-386
- Müller, D., Vogel, C., Bai, Y., and Vorholt, J. (2016). The plant microbiota: systems-level insights and perspectives. *Annu. Rev. Genet.* 50, 211–234. doi: 10.1146/annurev-genet-120215-034952
- Murdoch, D., and Chow, E. (2013). *Ellipse: Functions for Drawing Ellipses and Ellipse-Like Confidence Regions*. R package version 0.3-8. Available online at: <https://cran.r-project.org/package=ellipse>
- Niblack, T., and Riggs, R. (2015). “Soybean cyst nematode,” in *Compendium of Soybean Diseases and Pests, 5th Edn.*, eds G. L. Hartman, J. C. Rupe, E. F. Sikora, L. L. Domier, J. A. Davis, and K. L. Steffey (St. Paul, MN: American Phytopathological Society), 100–104.
- Nimnoi, P., Pongsilp, N., and Lumyong, S. (2014). Co-inoculation of soybean (*Glycine max*) with Actinomycetes and Bradyrhizobium japonicum enhances plant growth, nitrogenase activity and plant nutrition. *J. Plant Nutr.* 37, 432–446. doi: 10.1080/01904167.2013.864308
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlenn, D., et al. (2016). *Vegan: Community Ecology Package*. R package version 2.4-1. Available online at: <https://CRAN.R-project.org/package=vegan>
- Pasolli, E., Truong, D., Malik, F., Waldron, L., and Segata, N. (2016). Maching learning meta-analysis fo large metagenomic datasets: tools and biological insight. *PLoS Comput. Biol.* 12:e1004977. doi: 10.1371/journal.pcbi.1004977

- Peiffer, J. A., Spor, A., Koren, O., Jin, Z., Tringe, S. G., Dangl, J. L., et al. (2013). Diversity and heritability of the maize rhizosphere microbiome under field conditions. *Proc. Natl. Acad. Sci. U. S. A.* 110, 6548–6553. doi: 10.1073/pnas.1302837110
- Rascovan, N., Carbonetto, B., Perrig, D., Diaz, M., Canciani, W., Abalo, M., et al. (2016). Integrated analysis of root microbiomes of soybean and wheat from agricultural fields. *Sci. Rep.* 6:28084. doi: 10.1038/srep28084
- R Core Team (2015). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online at: <http://www.r-project.org/>
- Revelle, W. (2016). *Psych: Procedures for Personality and Psychological Research, Version = 1.6.6*. Evanston, IL: Northwestern University. Available online at: <http://cran.r-project.org/package=psych>
- Rousk, J., Baath, E., Brookes, P. C., Lauber, C. L., Lozupone, C., Caporaso, J. G., et al. (2010). Soil bacterial and fungal communities across a pH gradient in an arable soil. *ISME J.* 4, 1340–1351. doi: 10.1038/ismej.2010.58
- Salviaggiotti, F., Cassman, K., Specht, J., Walters, D., Weiss, A., and Dobermann, A. (2008). Nitrogen uptake, fixation and response to fertilizer N in soybeans: a review. *Field Crops Res.* 108, 1–13. doi: 10.1016/j.fcr.2008.03.001
- Sharpton, T. J. (2014). An introduction to the analysis of shotgun metagenomic data. *Front. Plant Sci.* 5:209. doi: 10.3389/fpls.2014.00209
- Shiro, S., Matsuura, S., Saiki, R., Sigua, G. C., Yamamoto, A., Umehara, Y., et al. (2013). Genetic diversity and geographical distribution of indigenous soybean-nodulating Bradyrhizobia in the United States. *Appl. Environ. Microb.* 79, 3610–3618. doi: 10.1128/AEM.00236-13
- Simon, C., and Daniel, R. (2011). Metagenomic analyses: past and future trends. *Appl. Environ. Microb.* 77, 1153–1161. doi: 10.1128/AEM.02345-10
- Soe, K. M., Bhromsiri, A., Karladee, D., and Yamakawa, T. (2012). Effects of endophytic actinomycetes and Bradyrhizobium japonicum strains on growth, nodulation, nitrogen fixation and seed weight of different soybean varieties. *Soil Sci. Plant Nutr.* 58, 319–325. doi: 10.1080/00380768.2012.682044
- Soueidan, H., and Nikolski, M. (2016). Machine learning for metagenomics: methods and tools. arXiv:1510.06621v2 [q-bio.GN]. doi: 10.1515/metgen-2016-0001
- Tkacz, A., and Poole, P. (2015). Role of root microbiota in plant productivity. *J. Exp. Bot.* 66, 2167–2175. doi: 10.1093/jxb/erv157
- Van Roekel, R. J., Purcell, L. C., and Salmeron, M. (2015). Physiological and management factors contributing to soybean potential yield. *Field Crops Res.* 182, 86–97. doi: 10.1016/j.fcr.2015.05.018
- Wang, J., and Jia, H. J. (2016). Metagenome-wide association studies: fine-mining the microbiome. *Nat. Rev. Microbiol.* 14, 508–522. doi: 10.1038/nrmicro.2016.83
- Wei, T., and Simko, V. (2016). *Corrplot: Visualization of a Correlation Matrix*. R package version 0.77. Available online at: <http://cran.r-project.org/package=corrplot>
- Wilke, A., Bischof, J., Gerlach, W., Glass, E., Harrison, T., Keegan, K. P., et al. (2016). The MG-RAST metagenomics database and portal in 2015. *Nucleic Acids Res.* 44, D590–D594. doi: 10.1093/nar/gkv1322
- Wilke, A., Harrison, T., Wilkening, J., Field, D., Glass, E. M., Kyrpides, N., et al. (2012). The M5nr: a novel non-redundant database containing protein sequences and annotations from multiple sources and associated tools. *BMC Bioinformatics* 13:141. doi: 10.1186/1471-2105-13-141
- Wright, M. N., and Ziegler, A. (2015). Ranger: a fast implementation of random forests for high dimensional data in C++ and R. arXiv: 1508.04409 [stat].
- Xu, X. M., Passey, T., Wei, F., Saville, R., and Harrison, R. J. (2015). Amplicon-based metagenomics identified candidate organisms in soils that caused yield decline in strawberry. *Hortic. Res.* 2:15022. doi: 10.1038/hortres.2015.22
- Xu, Y. X., Wang, G. H., Jin, J., Liu, J. J., Zhang, Q. Y., and Liu, X. B. (2009). Bacterial communities in soybean rhizosphere in response to soil type, soybean genotype, and their growth stage. *Soil Biol. Biochem.* 41, 919–925. doi: 10.1016/j.soilbio.2008.10.027
- Zhang, X., Zhang, D. Y., Jia, H. J., Feng, Q., Wang, D. H., Liang, D., et al. (2015). The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. *Nat. Med.* 21, 895–905. doi: 10.1038/nm.3914
- Zimmer, S., Messmer, M., Haase, T., Piepho, H. P., Mindermann, A., Schulz, H., et al. (2016). Effects of soybean variety and Bradyrhizobium strains on yield, protein content and biological nitrogen fixation under cool growing conditions in Germany. *Eur. J. Agron.* 72, 38–46. doi: 10.1016/j.eja.2015.09.008

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Chang, Haudenschild, Bowen and Hartman. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.