Supporting Online Material for

Soil nutrient levels are associated with suppression of banana *Fusarium* wilt disease

Yifan Cao[1], Yannan Ou[1], Nana Lv[1], Chengyuan Tao[1], Hongjun Liu[1], Rong Li[1, 2], Zongzhuan Shen[1, 2*], Qirong Shen[1]

Affiliations

[1] Jiangsu Provincial Key Lab for Organic Solid Waste Utilization, National Engineering Research Center for Organic-based Fertilizers, Jiangsu Collaborative Innovation Center for Solid Organic Waste Resource Utilization, Nanjing Agricultural University, Nanjing, 210095, PR China.

[2] The Sanya Institute of Nanjing Agricultural University, Sanya, Hainan Province, China.

[*]Corresponding author:

Zongzhuan Shen, College of Resources and Environmental Sciences, Nanjing Agricultural University, 210095, Nanjing, China. E-mail: shenzongz@njau.edu.cn. Tel: (86)02584395521. Fax: (86)02584396291.

Supplemental material and methods

Material and Methods

1 Collection and definition of the disease suppressive (S) and conducive (C) soils

In 2016, we identified paired banana orchards on Hainan Island, China, characterized by their suppressive (19.477605 N, 109.423841 E) or conducive (19.491065 N, 109.392632 E) nature to banana Fusarium wilt. This region, a notable banana production area, experiences a tropical monsoon climate with extended sunshine and significant annual precipitation. The orchards, under long-term monocropping, cultivated the susceptible *Musa acuminata* Cavendish cv. Brazil variety. It is important to note that the pesticide management and irrigation practices at the suppressive (S) and conducive (C) sites were roughly similar, as per farm records.

For the purpose of our study, disease-suppressive soil was defined as having a maximum disease incidence of 15%, a threshold deemed acceptable by local farmers due to the minimal economic impact at this rate. Conversely, disease-conducive soil was characterized by severe Fusarium wilt incidence, exceeding 50%.

The identification of Fusarium wilt was based on the observation of typical wilt symptoms (Dita et al., 2018). We designated five representative subplots (50m × 40m) at each site for soil sampling and disease incidence assessment post-harvest. Fusarium wilt incidence was calculated as: (number of infected plants) / (total number of banana plants) × 100%. In each subplot, five banana trees without wilt symptoms were selected for soil sampling, following a previously established method (Shen et al., 2018). A total of twenty soil samples, encompassing both bulk and rhizosphere soils from S and C sites, were collected for further analysis. Plant residues were removed from the bulk soil samples, a portion of which was air-dried for soil chemical property determinations. The remaining bulk soil and all rhizosphere soil samples were preserved at -80°C for subsequent microbial analysis.

2 Assay of soil chemical properties

The analysis of soil chemical properties was comprehensive, including soil pH, electrical conductivity (EC), content of total carbon (TC), total nitrogen (TN), ammonium nitrogen (NH_4^+ -N), nitrate nitrogen (NO_3^- -N), total phosphorus (TP), available phosphorus (AP), total potassium (TK), available potassium (AK), calcium (Ca), magnesium (Mg), copper (Cu), zinc (Zn), manganese (Mn), silicon (Si), iron (Fe) and molybdenum (Mo). These properties were determined following established methodologies (Bao, 2000). Additionally, various nutrient stoichiometry ratios such as carbon-nitrogen

ratio C/N, carbon-phosphorus ratio C/P, carbon-potassium ratio C/K, nitrogen-phosphorus ratio N/P, nitrogen-potassium ratio N/K and phosphorus-potassium ratio P/K were calculated based on the proportions of their total contents.

3 Amplicon sequencing for soil bacterial and fungal community

For the analysis of the soil bacterial and fungal communities, genomic DNA of all soil samples was extracted by using the DNeasy® PowerSoil® Kit (QIAGEN GmbH, Germany), following the manufacturer's guidelines.

Bacterial and fungal sequencing libraries were prepared in accordance with established protocols (Kozich et al., 2013). The V4 region of the bacterial 16S rRNA genes and the internal transcribed spacer 1 region (ITS1) of fungi were amplified using specific primers (515F/806R for bacteria and ITS1F/ITS2 for fungi). The quality and concentration of the amplicons were assessed using an Agilent 2100 Bioanalyzer Instrument (Agilent Technologies Co. Ltd, USA) and a KAPA Library Quantification Kit (KapaBiosystems, USA). All constructed libraries were sequenced using the Illumina HiSeq 2000 platform at the Novogene Bioinformatics Institute (Beijing, China).

The resulting raw DNA sequences were processed in USEARCH (v. 9.1.13) (Edgar, 2016). Unoise2 algorithm was employed in denoising to identify all correct biological sequences in the reads. The classification of representative sequences for each operational taxonomic unit (OTU) was performed against the silva_16s_v123 or UNITE 8.2 databases, using the 'sintax' command with a bootstrap confidence threshold of 0.6.

4 Metagenomic sequencing of soil microbial community and functional analysis

For a comprehensive analysis of the soil microbial community and its functional aspects, metagenomic sequencing was conducted on all soil samples. Genomic DNA, extracted from 5 g of each soil sample using the PowerMax® Soil DNA Isolation kit (MoBio Laboratories Inc., USA), was prepared for sequencing according to the Illumina Paired-End Prep kit protocol. DNA fragmentation was achieved using Covaris M220 (Gene Company Limited, China), resulting in an average fragment size of approximately 300 bp. This was followed by the construction of paired-end libraries using TruSeqTM DNA Sample Prep Kits (Illumina, San Diego, CA, USA), with adapters containing sequencing primer hybridization sites ligated to the fragment ends. Shotgun sequencing was conducted on an Illumina HiSeq 4000 platform (Illumina Inc., San Diego, CA, USA) at Majorbio.

The initial processing of sequencing data involved adapter sequence removal using SeqPrep

software (https://github.com/jstjohn/SeqPrep), and read trimming with the library sickle (https://github.com/najoshi/sickle). Trimming was based on quality, with any portion of a read dropping below a mean quality score of 20 being discarded. Reads shorter than 50 bp or containing ambiguous bases were also eliminated. Subsequent de novo assembly of the filtered sequences was executed using SOAP software (http://soap.genomics.org.cn, V. 1.06), and k-mer values for the main splicing parameter were set between 39-47. Open reading frames (ORFs) of the assembled contigs were predicted using MetaGene software. The predicted gene sequences were clustered using CD-HIT software (v. 4.6.4), selecting the longest gene sequence from each cluster to create a nonredundant gene set.

The high-quality reads from each sample were aligned to this nonredundant gene set (95% identity) using SOAPaligner software, building a comprehensive metagenome reference gene set for in-depth analysis. This reference gene set underwent a BLASTP search against the NCBI NR database for taxonomic assignment. Annotations were also acquired by blasting the gene set against the eggnog database for Clusters of Orthologous Groups (COG) identification (v4.0) and against the Kyoto Encyclopedia of Genes and Genomes (KEGG) database for pathway information. Additionally, contigs exceeding 5 kb were analyzed using antiSMASH (V. 5.1.2) with default parameters, to evaluate secondary metabolism capabilities, following previously described methods (Blin et al., 2019).

5 Statistical analyses

All statistical analyses were conducted using R software (version 4.0.4). The "Metafor" package was used to evaluate the difference of soil chemical properties between disease-suppressive (S) and disease-conducive (C) soil. The natural log of the response ratio (RR) of all soil chemical properties was calculated between S and C soils according to this equation: $RR = \ln (X_s / X_c)$. The variance of RR was calculated using the equation: $vRR = SD_c^2 / (n_c * X_c^2) + SD_s^2 / (n_s * X_s^2)$ (Hedges et al., 1999). The X_c, n_c, and SD_c represent the mean, replicate number and standard deviation of soil chemical properties in C soil, and X_s, n_s, SD_s represent those indexes in S soil respectively. A soil chemical property was deemed significantly higher in S soil compared to C soil if its RR was greater than 0, and its 95% confidence interval did not overlap with zero.

In order to determine the influence of soil chemical properties on the Fusarium wilt disease incidence, which differed significantly between S and C soils, hierarchical linear regression was considered appropriate. A linear mixed-effects model was applied with soil chemical properties as fixed factors and disease type (S or C soils) as a random factor. This was achieved by the "lme4" and “lmerTest”

packages. Additionally, the Boruta algorithm, an all-relevant feature selection wrapper capable of working with Random Forest (Kursa and Rudnicki, 2010), was used to identify the important soil chemical properties influencing functional genes of disease suppression. This algorithm ranks factors according to their importance, categorizing them as “Confirmed” (green factors with strong influence), “Tentative” (yellow factors with moderate influence), or “Rejected” (red factors with minimal influence).

Statistical difference analyses for single factors between two or multiple groups were conducted using non-parametric tests: the Wilcoxon rank sum test and Kruskal Wallis test, respectively, implemented with “stats” and “agricolae” packages. Spearman correlation analyses between paired parameters were performed using the “cor.test” command.

Microbial community analysis was conducted using the “vegan” package. Nonmetric Multidimensional Scaling (NMDS) of soil microbial community was calculated based on Bray-Curtis distance, and PERMANOVA tests was used to detect differences among treatments. Shannon diversity was estimated from a rarefied OTU table. The “ggtree” package was used for phylogenetic tree generation. Graphical representations were produced using “ggplot2”, “ggbeeswarm”, “RColorBrewer”, “corrplot”, and “pheatmap” packages, while “dplyr”, “reshape2”, and “xlsx” assisted in data processing.

Reference

Dita, M., Barquero, M., Heck, D., Mizubuti, E.S., Staver, C.P., 2018. Fusarium wilt of banana: current knowledge on epidemiology and research needs toward sustainable disease management. Frontiers in plant science 9, 1468.

Shen, Z., Penton, C.R., Lv, N., Xue, C., Yuan, X., Ruan, Y., Li, R., Shen, Q., 2018. Banana Fusarium wilt disease incidence is influenced by shifts of soil microbial communities under different monoculture spans. Microbial ecology 75, 739-750.

Bao, S.D., 2000. Soil and agricultural chemistry analysis. China Agriculture Press, Beijing.

Edgar, R.C., 2016. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. BioRxiv, 081257.

Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., Schloss, P.D., 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Applied and environmental microbiology 79, 5112-5120.

Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S.Y., Medema, M.H., Weber, T., 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic acids research 47, W81-W87.

Hedges, L.V., Gurevitch, J., Curtis, P.S., 1999. The meta-analysis of response ratios in experimental ecology. Ecology 80, 1150-1156.

Kursa, M.B., Rudnicki, W.R., 2010. Feature selection with the Boruta package. Journal of statistical software 36, 1-13.

Supplemental tables

Table S1 Soil chemical properties of disease suppressive (S) and conducive (C) soils

	C1	C2	C3	C4	C5	S1	S2	S3	S4	S5
DI	64	68	72	64	60	13	10	9	7	10
pH	5.73	5.01	4.99	4.98	5.78	7.12	6.96	7.2	6.96	6.75
EC	222	99	107	108	206	192	165	178	223	144
AP	44	15	19	22	102	131	143	142	217	146
AK	158	72	93	92	151	126	91	109	116	85
TP	0.86	0.7	0.38	0.5	0.79	2	1.88	2.51	2.45	1.99
TK	8.03	5	5.57	5.28	8.77	9.23	7.28	8.03	8.26	7.91
NH_4-N	0.34	0.41	0.43	0.36	0.4	0.35	0.38	0.34	0.4	0.42
NO_3-N	0.31	0.11	0.1	0.2	0.26	0.43	0.26	0.76	0.59	0.37
TN	0.13	0.13	0.13	0.13	0.16	0.18	0.197	0.23	0.25	0.19
TC	1.54	1.42	1.34	1.25	1.61	1.67	1.93	2.2	2.32	1.81
Zn	2.63	3.19	3.08	2.81	5.13	5.28	5.28	8.3	10.05	5.48
Mn	79	95.3	102	83.8	91.5	40.2	32.8	45.6	60	49.2
Fe	11.6	21.9	21	19.9	15.6	12.6	13.5	11.5	19.4	19.8
Cu	1.02	1.35	1.28	1.23	1.1	2.9	2.47	3.56	2.47	2.16
Mo	0.28	0.26	0.27	0.258	0.31	0.3	0.16	0.2	0.17	0.22
Si	177	104.6	142.7	130.8	171.5	321.3	236	203.8	224.9	200.5
Ca	2.9	2.4	1.4	2.1	3.3	7.8	7.8	10.5	11.6	5.5
Mg	2.9	3	2.4	2.9	3	2	1.9	2.1	2.1	2
C/N	11.68	11.3	10.4	9.73	10.29	9.44	9.77	9.74	9.36	9.65
N/P	0.151	0.186	0.342	0.26	0.203	0.09	0.105	0.092	0.102	0.095
N/K	0.016	0.026	0.023	0.025	0.018	0.02	0.027	0.029	0.03	0.024
P/K	0.107	0.14	0.068	0.095	0.09	0.217	0.258	0.313	0.297	0.252
C/P	1.791	2.029	3.526	2.5	2.038	0.835	1.027	0.876	0.947	0.91
C/K	0.192	0.284	0.241	0.237	0.184	0.181	0.265	0.274	0.281	0.229

Table S2 Detailed results of the response ratio between disease suppressive and conducive soils

label	RR	lci	uci	*p*
C/P	-0.9503	-1.2149	-0.6857	<.0001
N/P	-0.8582	-1.1512	-0.5651	<.0001
C/N	-0.1074	-0.1747	-0.0402	0.0017
C/K	0.0787	-0.1372	0.2946	0.4751
N/K	0.1778	-0.0461	0.4018	0.1197
P/K	0.9824	0.7195	1.2453	<.0001
Mn	-0.6843	-0.899	-0.4696	<.0001
Mg	-0.3407	-0.4263	-0.2552	<.0001
Mo	-0.2718	-0.5138	-0.0299	0.0276
Fe	-0.1586	-0.4663	0.1491	0.3124
AK	-0.0714	-0.4032	0.2604	0.6732
NH_4^-N	-0.0261	-0.1402	0.088	0.6537
EC	0.1953	-0.1887	0.5792	0.3189
TK	0.2206	-0.0249	0.4662	0.0782
pH	0.2783	0.2059	0.3507	<.0001
TC	0.3271	0.1782	0.4759	<.0001
TN	0.4316	0.281	0.5821	<.0001
Si	0.4904	0.2344	0.7463	0.0002
Zn	0.714	0.3309	1.0971	0.0003
Cu	0.8187	0.6177	1.0197	<.0001
NO_3^-N	0.8998	0.3562	1.4434	0.0012
TP	1.2098	0.9128	1.5069	<.0001
Ca	1.2726	0.9108	1.6344	<.0001
AP	1.3497	0.5399	2.1596	0.0011

Table S3 Detailed results of the hierarchical multiple regression between soil chemical properties and disease incidence

lable	regression coefficient	*p*
pH	-0.22	0.0825
EC	-0.07661	0.0696
AP	-0.1815	0.0171
AK	-0.06047	0.145
TP	-0.2546	0.0734
TK	-0.0727	0.145
NH_4^-N	0.03044	0.492
NO_3^-N	-0.08232	0.173
TN	-0.1474	0.0722
TC	-0.1204	0.086
Zn	-0.09447	0.133
Mn	0.06997	0.567
Fe	0.03408	0.463
Cu	0.04172	0.688
Mo	0.0216	0.714
Si	0.001647	0.982
Ca	-0.1283	0.159
Mg	-0.215	0.0275
C/N	0.02189	0.735
N/P	0.1075	0.126
N/K	0.008766	0.863
P/K	-0.1406	0.278
C/P	0.1428	0.0618
C/K	0.01747	0.702

Supplemental figures

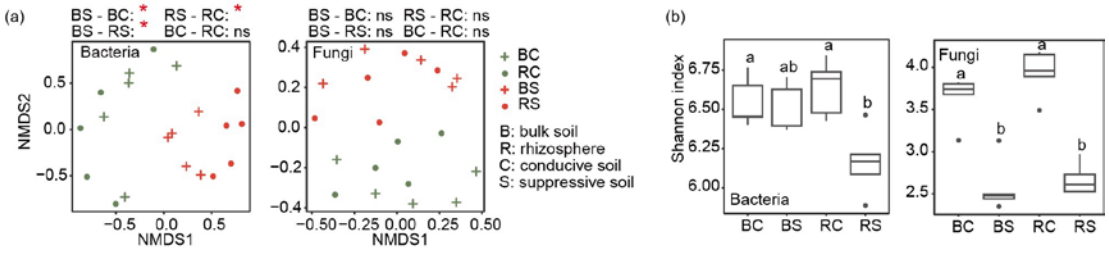

Fig S1 Diversity of soil microbial community in soils suppressive or conducive to Fusarium wilt (a) Non-metric Multidimensional Scaling (NMDS) analysis illustrated the Bray-Curtis distance in bacterial and fungal communities across different soil types. Statistical significance of the differences was assessed using PERMANOVA test. **(b)** Boxplot depicted the Shannon diversity index in different soil types. Significance testing was conducted using the Kruskal-Wallis test. An asterisk indicates $p < 0.1$.

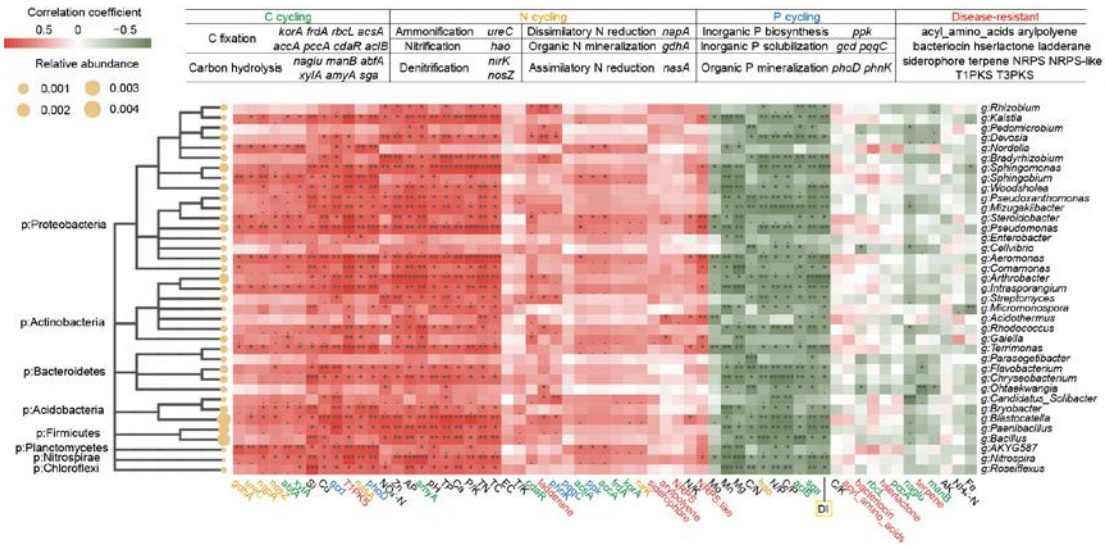

Fig S2 Relationship between soil microbial composition and microbial function genes and soil chemical properties A phylogenetic tree represented the microbial OTUs that were enriched in suppressive soil compared to conducive soil ($p < 0.1$) in rhizosphere soil. They were presented at genus level. Accompanying this was a heatmap delineating the Spearman correlations between these microbial groups and disease incidence, all the soil chemical properties and microbial functional genes. A dot, an asterisk, two asterisks and three asterisks indicate $p < 0.1$, $p < 0.05$, $p < 0.01$ and $p < 0.001$, respectively.