Supplementary files

The relative importance of soil moisture in predicting bacterial wilt disease occurrence

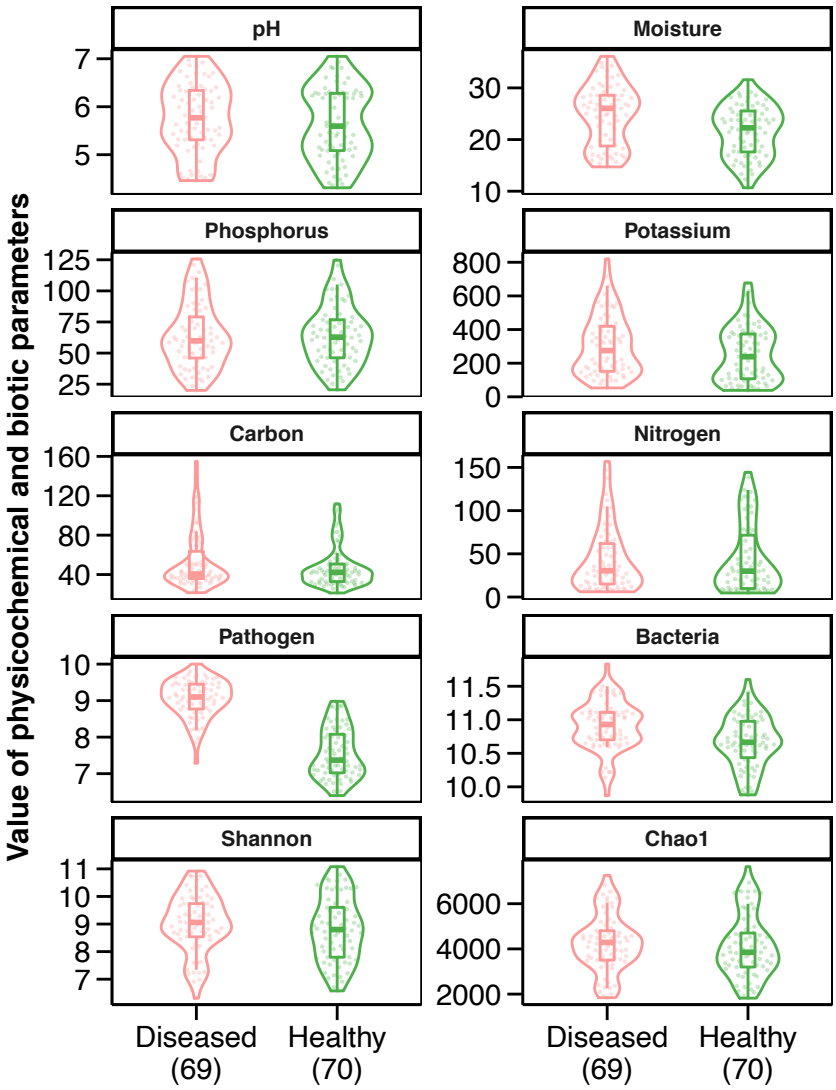
Gaofei Jiang, Ningqi Wang, Yaoyu Zhang, Zhen Wang, Yuling Zhang, Jiabao Yu, Yong Zhang, Zhong Wei, Yangchun Xu, Stefan Geisen, Ville-Petri Friman, Qirong Shen

The supplementary information contains six files. The table file includes statistical information about multiple regression analysis, and figure files provide further detail on the variation of physicochemical properties and bacterial communities in tomato rhizosphere and how it was linked to pathogen density and plant healthy in tomato rhizosphere microbiomes.

Supplementary Table 1. ANOVA table summarizing the relative importance of abiotic physicochemical and biotic soil properties in predicting pathogen densities in tomato rhizosphere samples.

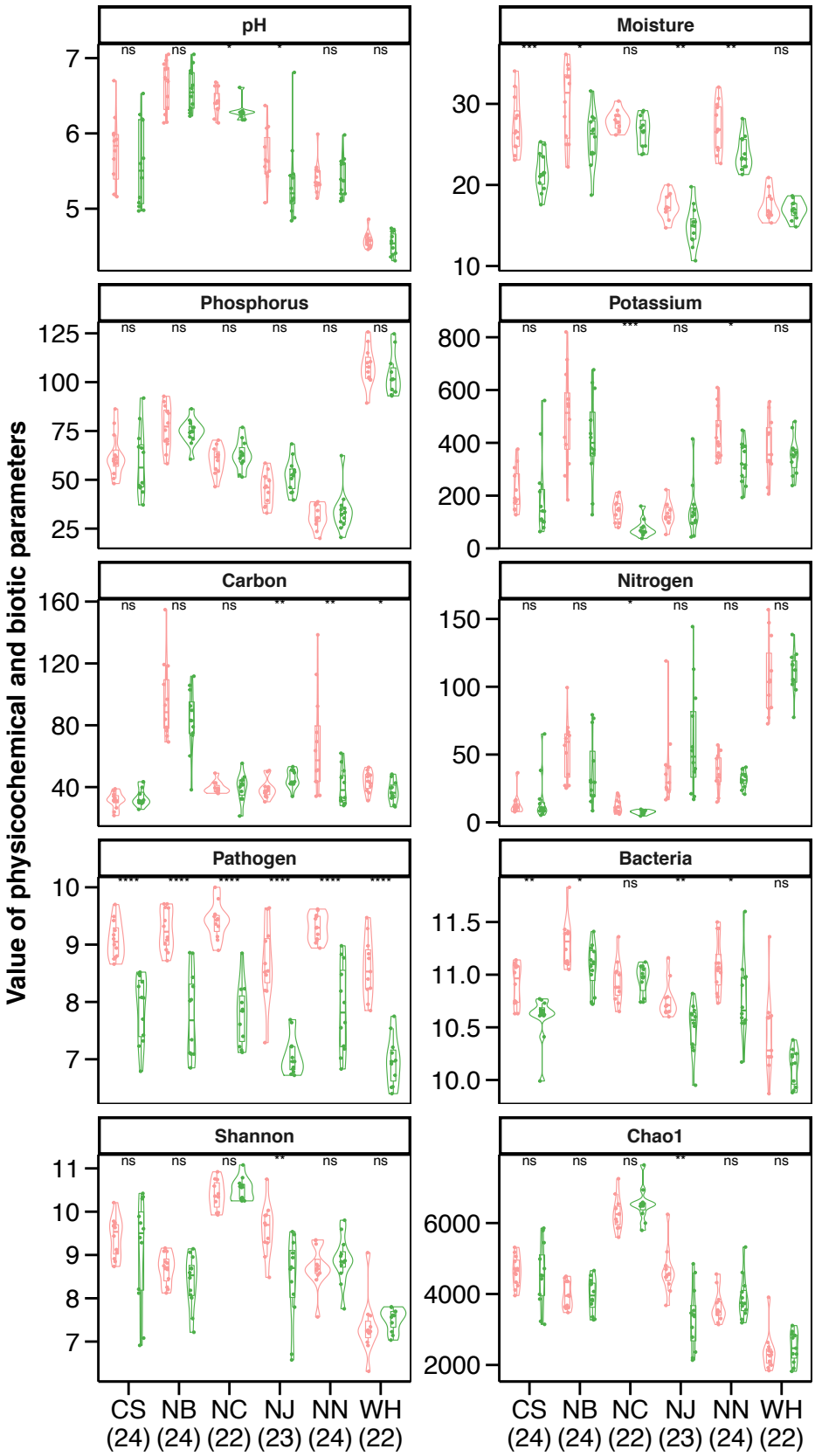
Predictor variable	Df	Sum Square	Mean Square	*F*-value	*P*-value	Relative weight
pH	1	10.29	10.29	18.68	<0.0001↓	14.59%
Moisture	1	28.45	28.45	51.66	<0.0001↑	40.36%
Phosphorus	1	0.14	0.14	0.25	0.6170↑	0.20%
Potassium	1	0.14	0.14	0.26	0.6141↑	0.20%
Carbon	1	0.07	0.07	0.12	0.7310↓	0.09%
Nitrogen	1	0.03	0.03	0.05	0.8271↓	0.04%
Bacterial abundance	1	16.05	16.05	29.15	<0.0001↑	22.77%
Shannon	1	1.78	1.78	3.24	0.0744↑	2.53%
Chao1	1	1.33	1.33	2.42	0.1226↓	1.89%
Bray-Curtis	1	0.11	0.11	0.21	0.6499↑	0.16%
Residuals	128	70.50	0.55			
Model Summary			AIC: 324.09; $F_{10,128} = 10.6$, $R^2 = 0.45$, $P < 0.0001$			

Note: The significant effects ($P < 0.05$) are shown in red colour and the 'up' and 'down' arrows denote for positive and negative effects, respectively, based on multiple regression model.

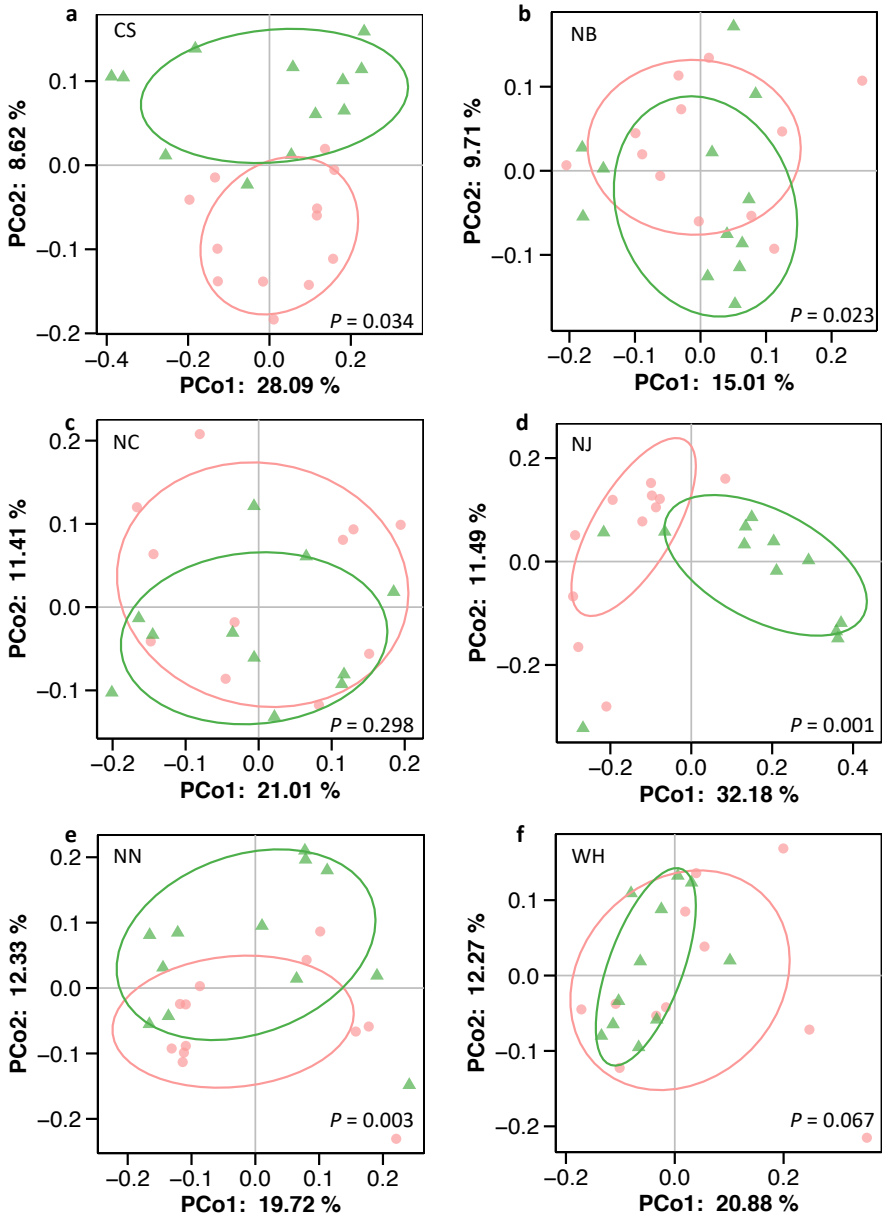


Supplementary Figure 1. Differences in abiotic physicochemical and biotic soil properties between diseased and healthy plants.

All data is pooled over sampling locations and the numbers in the parentheses denote the sample size in each group. The 'ns' denotes for non-significant difference ($P > 0.05$) and stars (*, **, *** and ****) show significant differences at levels $P < 0.05$, $P < 0.01$, $P < 0.001$ and $P < 0.0001$, respectively. Pathogen and total bacterial abundances are abbreviated as 'Pathogen' and 'Bacteria', respectively. Each violin plot shows the distribution of 69 and 70 rhizosphere soil samples in diseased and healthy plants, respectively.

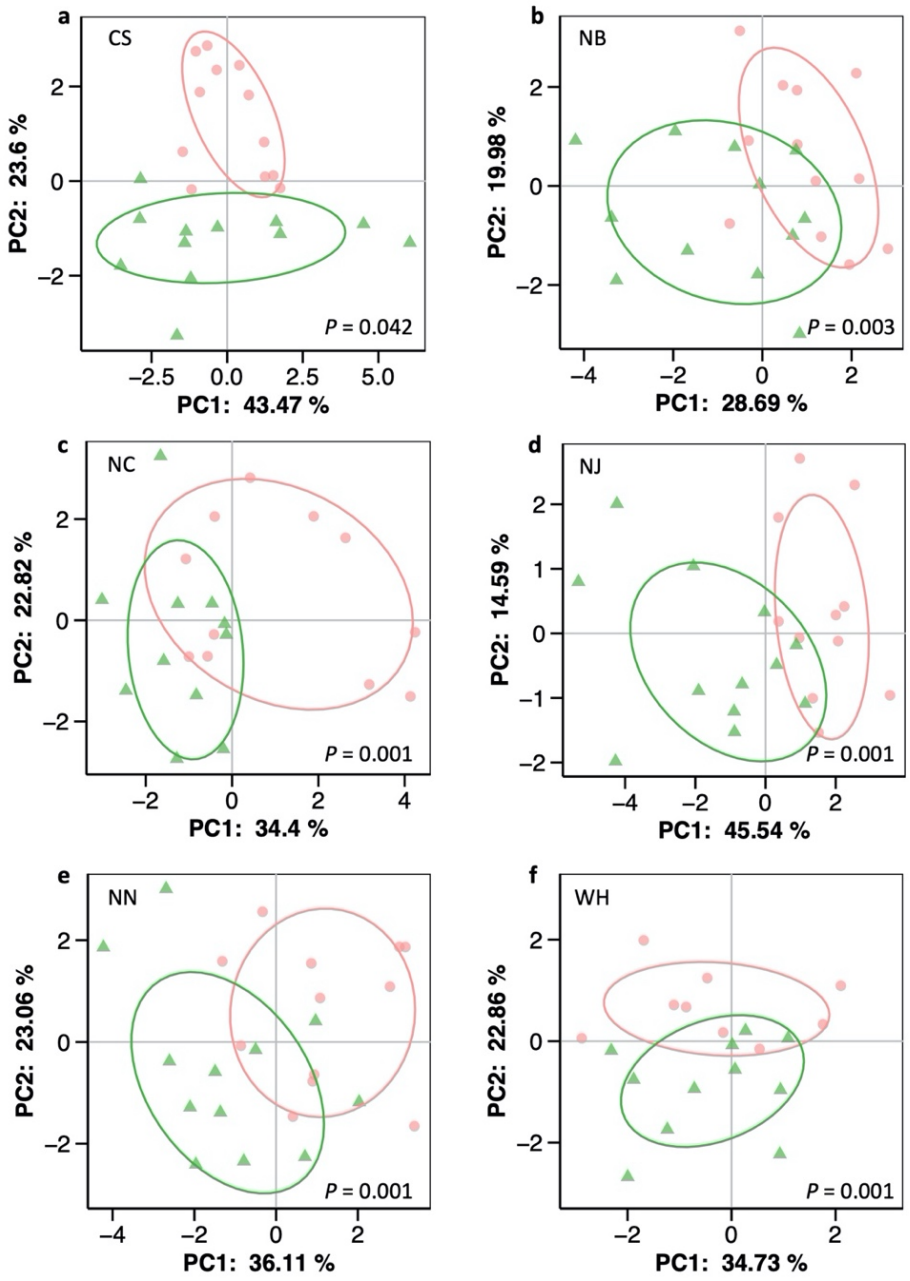

Value of physicochemical and biotic parameters
pH
Moisture
Phosphorus
Potassium
Carbon
Nitrogen
Pathogen
Bacteria
Shannon
Chao1
CS (24)
NB (24)
NC (22)
NJ (23)
NN (24)
WH (22)

Supplementary Figure 2. Differences in abiotic physicochemical and biotic soil properties of diseased and healthy plants in different sampling locations (provinces). Numbers in the parentheses show sample size in each group. The 'ns' denotes for non-significant difference ($P > 0.05$) and stars (*, **, *** and ****) show significant differences at levels $P < 0.05$, $P < 0.01$, $P < 0.001$ and $P < 0.0001$, respectively. Pathogen and total bacterial abundances are abbreviated as 'Pathogen' and 'Bacteria', respectively. Each violin plot shows the distribution of rhizosphere soils in each province. Sampling locations are abbreviated as follows: CS = Changsha, NB = Ningbo, NC = Nanchang, NJ = Nanjing, NN = Nanning and WH = Wuhan.



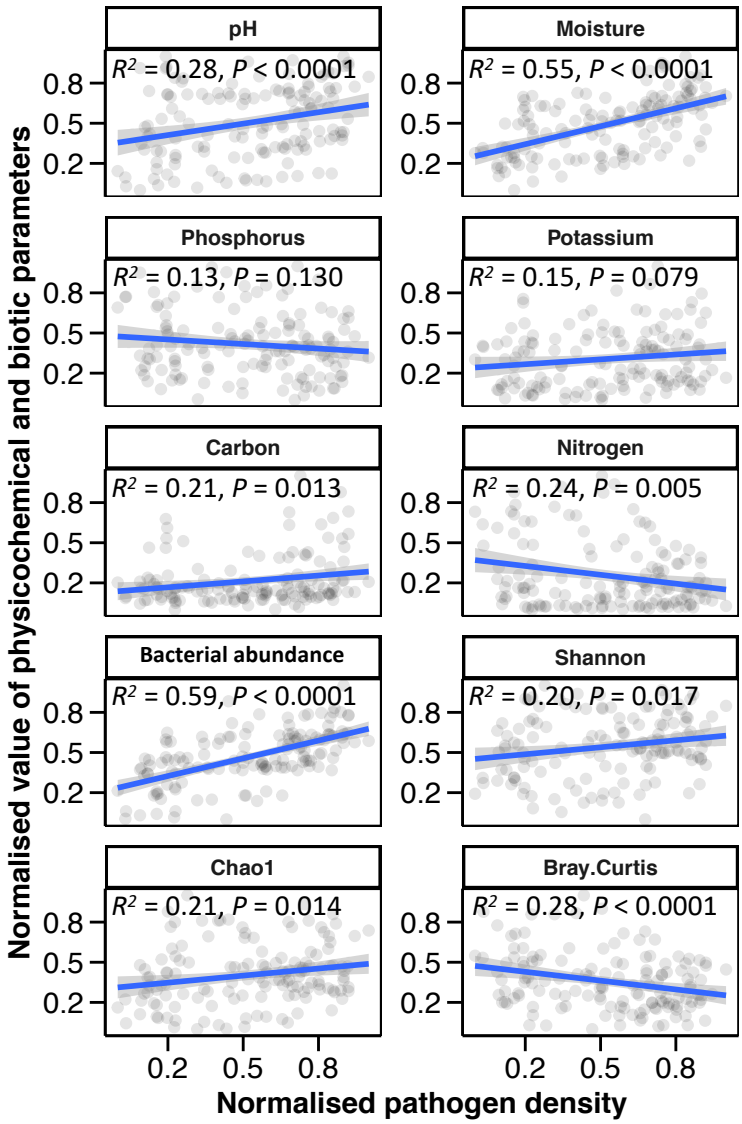
Supplementary Figure 3. Differences in bacterial community composition (PCoA) between healthy (green) and diseased (red) rhizosphere soil samples in different sampling locations (a-f).

PERMANOVA was used to identify microbial composition difference between the rhizosphere soil of diseased and healthy plants in each location based on Bray–Curtis distance matrices. *P*-values are indicated in each panel. Green triangles and red circles denote healthy and diseased plants, respectively. Sampling locations are abbreviated as follows: CS = Changsha, NB = Ningbo, NC = Nanchang, NJ = Nanjing, NN = Nanning and WH = Wuhan.



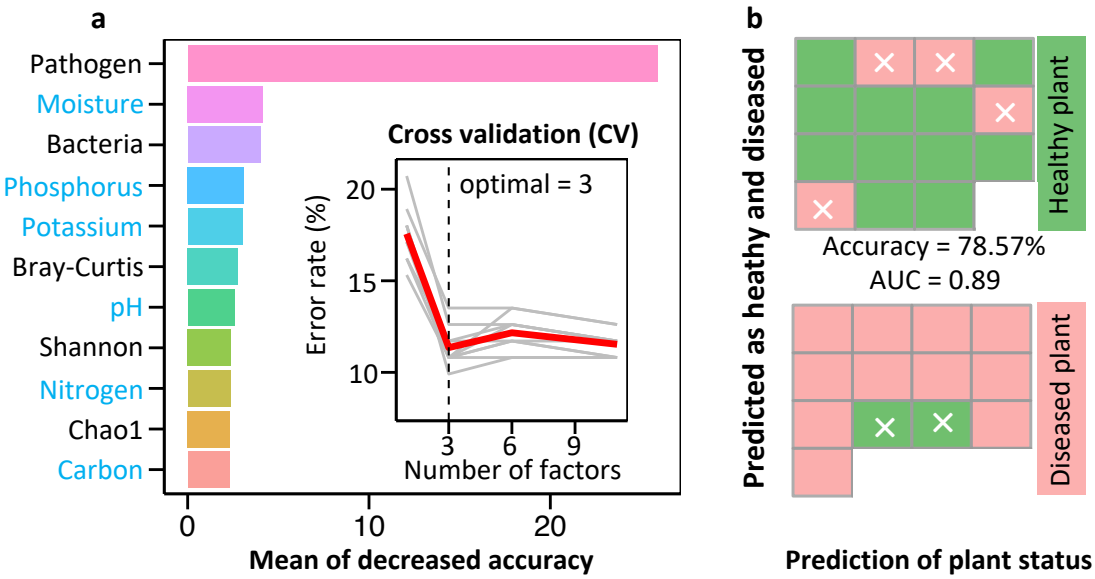
Supplementary Figure 4. Differences in physicochemical and biotic parameters (PCA) between healthy (green) and diseased (red) rhizosphere soil samples in different sampling locations (a-f).

PERMANOVA was used to for identify the environmental difference between the rhizosphere soil of diseased and healthy plants in each location based on Euclidean distance matrices. *P*-values are indicated in each panel. Green triangles and red circles denote healthy and diseased plants, respectively. Sampling locations are abbreviated as follows: CS = Changsha, NB = Ningbo, NC = Nanchang, NJ = Nanjing, NN = Nanning and WH = Wuhan.



Supplementary Figure 5. Linear correlations between the normalised pathogen densities and abiotic physicochemical and biotic soil properties (averaged over heathy and diseased plant samples).

Blue lines indices the linear fitting of curves, while R^2 indicates the Spearman correlation coefficient of the linear regression and *P*-values the significance of each correlation.



Supplementary Figure 6. Comparing the relative importance of all soil parameters in predicting plant health status.

(**a**) Relative importance rank of overall abiotic physicochemical (blue) and biotic (black) soil properties and ten-fold cross-validation of random forest model (inset in **a**) based on the training set (80% of randomly selected rhizosphere samples). Pathogen and total bacterial abundances are abbreviated as 'Pathogen' and 'Bacteria', respectively. (**b**) Validation of random-forest model with a test set (20% of remaining samples) predicting plant disease outcomes based on soil properties: green and red filled cells denote for correct predictions, while filled cells with white crosses denote for false predictions. The overall model gained an average accuracy (78.6%) in classifying plant status, with 84.6% accuracy for diseased and 73.3% accuracy for health plants (AUC = 0.89).