Supplementary Materials

Effect of Microbial Inoculants on Rhizosphere Soil Microbial Community Structure in Melon

Jian Sun [1,#], Meixia Liang [1,#], Mingmei Zhai [1,#], Qi Wang [1], Yuansha He [1], Chunmei Zhao [2], Jianqun Lin [3], Jianming Yang [1], and Zhaobao Wang [1,*]

[1] Energy-rich Compounds Production by Photosynthetic Carbon Fixation Research Center, Shandong Provin-cial Key Laboratory of Microbial Resource Exploration and Innovative Utilization, College of Life Sciences, Qingdao Agricultural University, Qingdao 266109, China.

[2] Biology Public Experimental Teaching Center, Qingdao Agricultural University, Qingdao 266109, China.

[3] State Key Laboratory of Microbial Technology, Shandong University, Qingdao 266200, PR China.

* Address correspondence to Zhaobao Wang

Emails: wangzhaobao123@126.com

[#] These three authors have contributed equally to this work.

Postal address: No. 700 Changcheng Road, Chengyang District, Qingdao 266109, China

Tel & Fax: +86-532-58957375

一、Supplementary Methods

1.1 Statistical Analysis

Alpha diversity indices were calculated using the algorithms implemented in mothur (v1.30.2; https://mothur.org/wiki/calculators/) with R software (version 3.3.1). The permutation test was performed with 999 permutations. PCoA was performed and visualized using the vegan package in R (version 3.3.1). ANOSIM was used to assess differences in microbial community structure among groups, and the corresponding R and P values were calculated. LEfSe analysis was first performed using the nonparametric Kruskal–Wallis (KW) rank-sum test to identify features that differed significantly among treatment groups. Linear discriminant analysis (LDA) was then used to estimate the discriminative effect sizes of these features. RPKM-normalized abundance data were used, and the All-against-all (more strict) strategy was applied for multiple-group comparisons, requiring candidate features to meet the specified significance criteria in all pairwise group comparisons. An LDA score threshold of 2.0 was used to select differential features with a certain level of discriminative effect size, thereby excluding features with relatively weak discriminatory power and potentially limited biological relevance.

1.2 Data Normalization

(1) Alpha diversity was calculated based on read counts, and rarefaction was performed.

$$\mathrm{Gene_i} = \mathrm{R_i}$$

Where R_i represents the abundance of $Gene_i$ in a given sample, defined as the number of reads mapped to $Gene_i$ in that sample.

(2) TPM-based abundance values were used for PCoA, Venn diagram, community composition bar plot, RDA, and correlation heatmap analyses. Specifically, the TPM value of a gene was calculated as its length-normalized abundance divided by the sum of the length-normalized abundances of all genes in the sample, multiplied by one million.

$$\mathrm{TPM_i} = \frac{\left(\frac{\mathrm{R_i}}{\mathrm{L_i}}\right) * 10^6}{\sum_1^x (\mathrm{R_j}/\mathrm{L_j})}$$

Where R_i represents the abundance of Genei in a given sample, defined as the number of reads mapped to Genei in that sample, L_i represents the nucleotide length of Genei, and $\sum_1^x (\mathrm{R_j}/\mathrm{L_j})$ represents the sum of the length-normalized abundances of all genes in that sample.

(3) RPKM-based abundance values were used for LEfSe, COG, KEGG, and CAZy abundance

analyses. RPKM represents the number of reads mapped to a gene per kilobase of gene length per million mapped reads.

$$\mathrm{RPKM_i} = \frac{\mathrm{R_i} * 10^6}{\mathrm{L_i} * \sum_1^x \mathrm{R_j}}$$

Where R_i represents the abundance of Genei in a given sample, defined as the number of reads mapped to $Gene_i$ in that sample, L_i represents the nucleotide length of Genei, and $\sum_1^x R_j$ represents the total number of reads mapped to all genes in that sample.

1.3 Functional Annotation

（1）COG Functional Annotation

Open reading frames (ORFs) were predicted from assembled contigs. Protein sequences of these ORFs were aligned against the COG 2020 database using Diamond v2.0.13 (blastp mode, E-value ≤ $1e^{-05}$). The raw read counts mapped to each ORF were summed to obtain the read abundance for each COG category.

（2）KEGG Functional Annotation

Amino acid sequences of the non-redundant gene set were aligned against the KEGG database using DIAMOND v2.0.13 (blastp mode, E-value ≤ $1e^{-05}$) to obtain KEGG functional annotations for the genes.

（3）Carbohydrate-Active Enzyme Annotation

Amino acid sequences of the non-redundant gene set were aligned against the CAZy database using hmmscan (E-value ≤ $1e^{-05}$). Carbohydrate-active enzyme annotations were assigned to the genes, and the abundances of genes assigned to each carbohydrate-active enzyme category were summed to obtain the abundance of each category.

Table S1. Metagenome sequencing statistics of each sample.

Sample	RawReads	CleanReads	Contigs	Max(bp)	Min(bp)	N50(bp)	N90(bp)
Vck	70644030	70514406	545762	106386	500	975	551
Vck	71820806	71690008	564043	106472	500	978	551
Vck	66311184	66191224	536588	161185	500	1020	555
VT1	69600088	69480150	555416	118437	500	977	551
VT1	72694766	72562450	601557	352294	500	996	553
VT1	68824630	68697050	536981	212065	500	974	551
VT2	68239570	68111890	575738	112176	500	969	554
VT2	73151040	73017354	605147	128078	500	977	554
VT2	69080378	68951692	565958	449574	500	993	553
VT3	73698568	73577388	575212	101163	500	967	549
VT3	72479952	72344860	523448	97790	500	956	549
VT3	67383608	67270616	507535	83068	500	974	550

Table S2. Envfit test results of environmental variables in RDA

	RDA1	RDA2	R^2	P
AK	0.9801	0.1983	0.1983	0.1983
AP	0.8079	0.5893	0.3319	0.1560
OM	-0.2287	0.9735	0.0311	0.8850
NO_3^--N	0.9689	0.2476	0.8430	0.0010
NH_4^+-N	0.9176	-0.3975	0.3584	0.1300
Cl^-	-0.9996	0.0286	0.6055	0.0150