

## Mining endophytic microbiome information from plant and animal transcriptome data

Guomin Han<sup>1,2,\*</sup>, Xianjin Wang<sup>1,2</sup>, Guiping Qiu<sup>1,2</sup>

<sup>1</sup> School of Life Sciences, Anhui Agricultural University, Hefei, 230036, China

<sup>2</sup> National Engineering Laboratory of Crop Stress Resistance Breeding, Anhui Agricultural University, Hefei, 230036, China

\* Author for correspondence: G. H. ([guominhan@ahau.edu.cn](mailto:guominhan@ahau.edu.cn))

### Abstract

Endophytic microorganisms play important physiological functions in plants and animals. In this paper, we developed a method to obtain endophytic microbiome information directly by analyzing transcriptome sequencing data files of plants and animals. Compared with the use of amplicon analysis or whole-genome sequencing of animal and plant tissues to analyze microbial composition information, this method can obtain endophytic microbiome information in addition to obtaining gene expression information of host plants and animals.

### Keywords

New method; Mining; Endophytic microbiome; Plant and animal; Transcriptome data

### Introduction

Microorganisms are almost ubiquitous, such as rivers, lakes, seas, soils, air, body surfaces, and insides of humans, animals, and plants. Many microorganisms, including bacteria, archaea, and fungi, can live in the tissues of plants. Endophytes are microorganisms that are present in various tissues of the host plant in a symbiotic or beneficial manner or without causing any detrimental effects (Ali, et al., 2021; Kaul, et al., 2016). A few endophytes can enhance the host plant tolerance to environmental abiotic stresses, such as thermotolerance (Marquez, et al., 2007; Shekhawat, et al., 2021), while others can promote plant growth through the production of phytohormones, solubilization of phosphorus and potassium, biological fixing of nitrogen, inhibition of ethylene biosynthesis (Fadiji, et al., 2021; Lata, et al., 2019). In addition, many endophytes can protect plants from microbial pathogens by the production of ammonia, hydrogen cyanide, and siderophores (Ali, et al., 2021; Carrion, et al., 2019; Lata, et al., 2019). Pathogenic fungi and symbiotic fungi can also penetrate plant cells for absorption or exchange of nutrients (Han, et al., 2019; Han, et al., 2020).

Current endophytic microorganisms detection methods mainly including culture-dependent and culture-independent methods (Dissanayake, et al., 2018; Tian, et al., 2019). Several culture-independent methods, such as Real-time PCR, In situ hybridization FISH, microarray, metabarcoding approaches, whole plant genome DNA sequencing have been used (Aslam, et al., 2017; Dissanayake, et al., 2018; Fadiji, et al., 2021; Fadiji, et al., 2021), and the mystery of a large number of plant endophytes was revealed (Aslam, et al., 2017; Matsumoto, et al., 2021; Trivedi, et al., 2020). The use of whole plant tissues for DNA extraction for amplifying 16S rDNA, fungal ITS area, or amplifying specific target genes and then sequencing with the second or third-generation sequencing technology is the often-used method. Another next-generation sequencing method is the direct sequencing of the whole genomic DNA from plant tissues and analysis using microbiome analysis software to generate endophytic microbiome information.

In this paper, a new method for mining endophytic microbiome information from plant transcriptome data was described. Briefly, after the removal of adaptor and low-quality bases, the clean data were mapped to the host plant genome or all cDNA files to generate unaligned files. The unaligned files were further analyzed by using a microbiome analysis pipeline to obtain plant endophytic microbiome information.

## Materials and Methods

### Endophytic microbiome mining method

The Workflow diagram is shown in Figure 1. The animal or plant transcriptome raw data were assessed via fastqc, and then cleaned via trimmomatic with the default parameter. The clean data were mapped to the host genome by using hisat2 with the default parameter. Use its output unaligned sequence parameter "--un-conc" to obtain an unaligned data file, which contains not only a part of the host animal and plant sequences but also expressed microbial gene fragments. Use microbiome analysis pipelines, such as Kraken2, bracken and Pavian (Breitwieser and Salzberg, 2020; Lu, et al., 2017; Wood, et al., 2019), to analyze the unaligned data files to generate endophytic microbiome information.

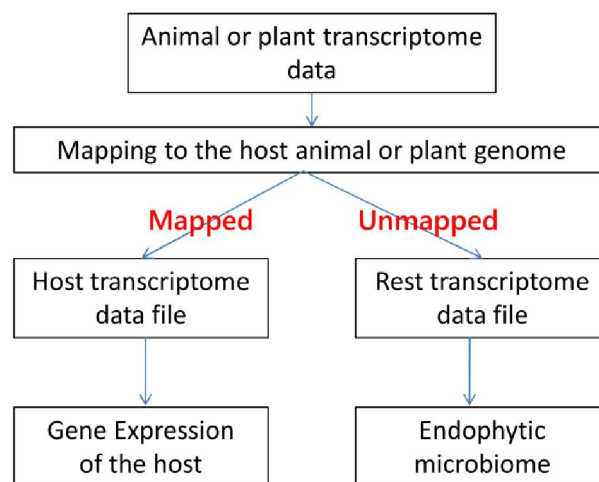


Figure 1 Workflow diagram

### Data sets

We selected the RNA-seq data of two plants inoculated or infected by microorganisms.

#### (1) Maize inoculated with arbuscular mycorrhiza (AM) fungus data sets

In our previous study, the maize (*Zea mays*) roots inoculated with AM fungus *Rhizophagus irregularis* DAOM-197198 (previously known as *Glomus intraradices*) were compared with the control without fungal inoculation (Han, et al., 2019). The maize raw RNA-Seq data (Bioproject accession: PRJNA553580) were used in this study. The maize inbred line B73 genome was download from Ensembl Plants.

#### (2) Common bean infected by *Xanthomonas* data sets

In the study by Foucher et al. (2020), a common bean (*Phaseolus vulgaris* L.) susceptible genotype (JaloEEP558) was infected by *Xanthomonas phaseoli* pv. *phaseoli* strain CFBP6546R 48 h after inoculation (Foucher, et al., 2020). The raw RNA-Seq data of Common bean (SRA accession: SRP273448) were used. The Common bean genome was download from Ensembl Plants.

## Results and discussion

### Generation the unaligned files from the host RNA-seq data

Approximately 18-28% of maize RNA-seq data cannot be mapping to the maize genome (Table S1), while about 5-11% of common bean RNA-seq data cannot mapping to the common bean genome (Table S2). The unmapping rates from maize inoculated with AM fungus are higher than that of the control, similar results were also observed in the common bean RNA-seq data (Figure 2).

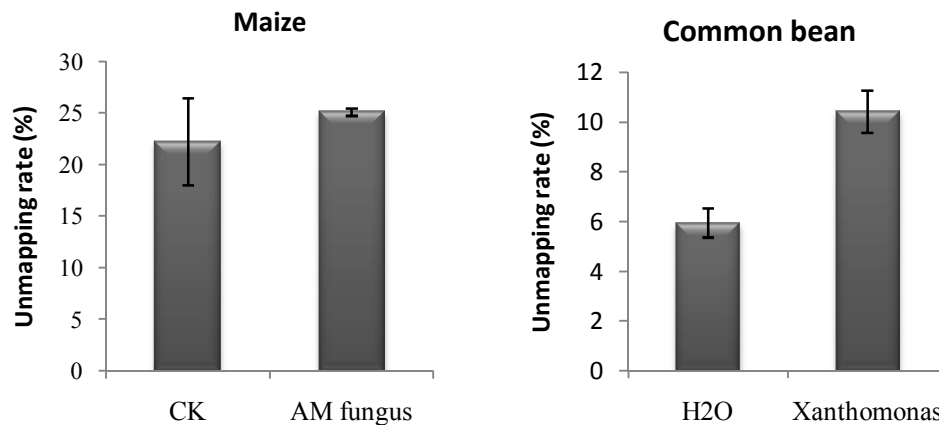


Figure 2 The unmapping rate of RNA-Seq data to each host genome.

### Endophytic microbiome information

As an example, in addition to maize root transcriptome data, the transcriptome data also included fungi, bacteria, archaea, and viruses, and similar results were found in the common bean transcriptome data (Figure 3). Further analysis can be done to obtain the content relationships between various types of microorganisms (Figure 4). The microorganisms in the maize root inoculated with AM fungi were mostly fungi, whereas microorganisms in common beans were dominated by bacteria (Figure 4).

In our previous study, the maize seedling was inoculated with *R. irregularis* DAOM-197198 to investigate the regulatory network responsive to AM fungi colonization in maize roots. The sankey diagram clearly shows that a large number of *R. irregularis* can be detected in the maize inoculated with AM fungus while the control was not detected (Figure 3). Similarly, *X. phaseoli* can be detected in the sample inoculated with *X. phaseoli* while the control was not detected (Figure 3). Species *R. irregularis* is the most significant biomarker in the maize roots inoculated with AM fungus *R. irregularis* (Figure 5), while Genus *Xanthomonas* is one of the most significant biomarkers in the common bean sample inoculated with *X. phaseoli* pv. *phaseoli* strain CFBP6546R (Figure 6). The evidence shows that the method of mining endophytic microbiome information from plant transcriptome data is reliable, and this method is also applicable to animal transcriptomes.

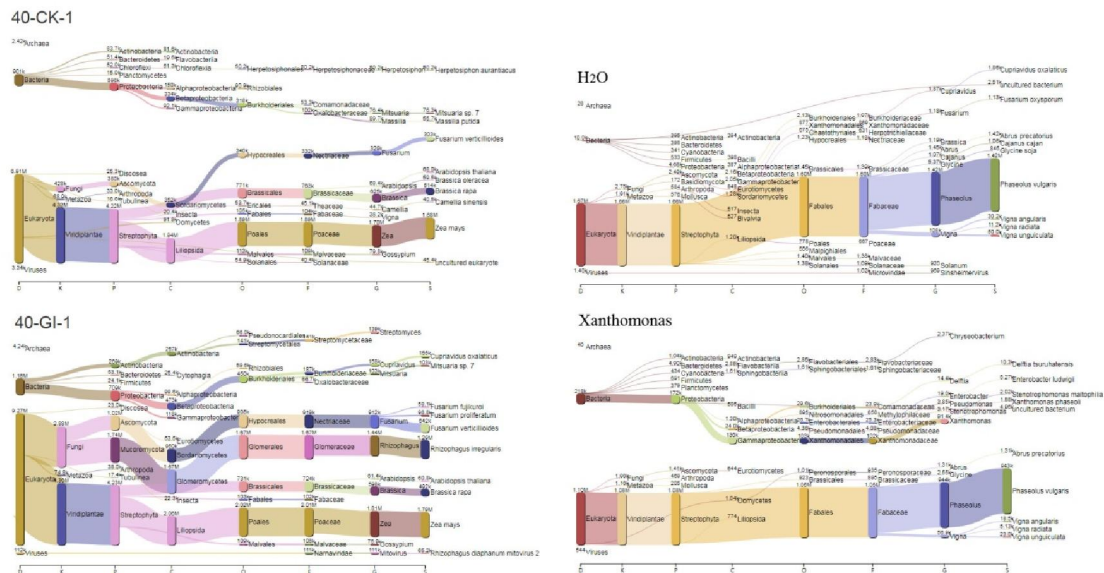


Figure 3 Species composition generated from unmapped files of the maize RNA-seq data and the common bean RNA-seq data, respectively.

Note: 40-CK-1, one of the control samples of maize; 40-GI-1, one of the maize roots inoculated with *R. irregularis* DAOM-197198; H<sub>2</sub>O, one of the control samples of common bean; Xanthomonas, one of the common bean sample inoculated with *X. phaseoli* pv. *phaseoli* strain CFBP6546R

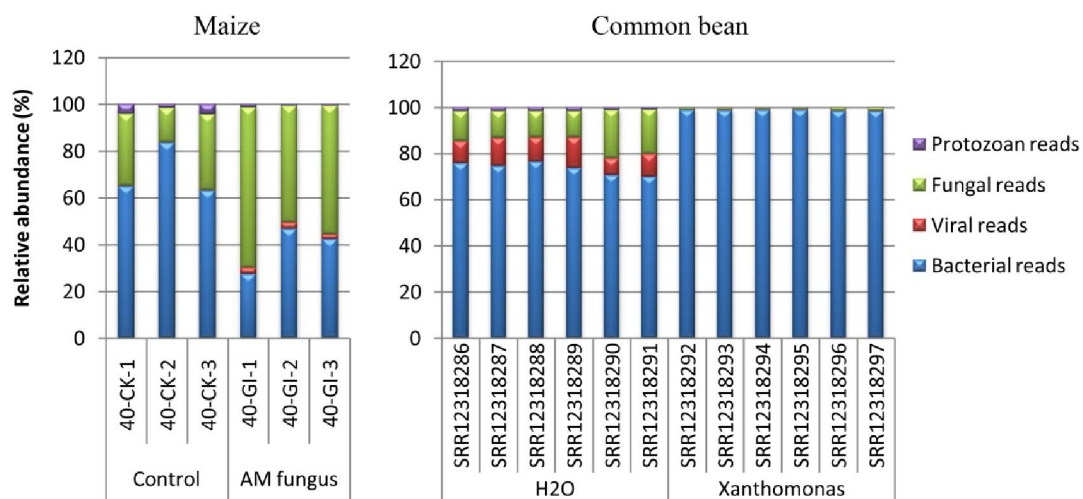


Figure 4 Relative abundance of different microorganisms generated from the maize RNA-seq data and the common bean RNA-seq data, respectively.

Note: Ck, the control sample of maize; AM fungus, the maize roots inoculated with *R. irregularis* DAOM-197198; H<sub>2</sub>O, the control sample of common bean; Xanthomonas, the common bean sample inoculated with *X. phaseoli* pv. *phaseoli* strain CFBP6546R

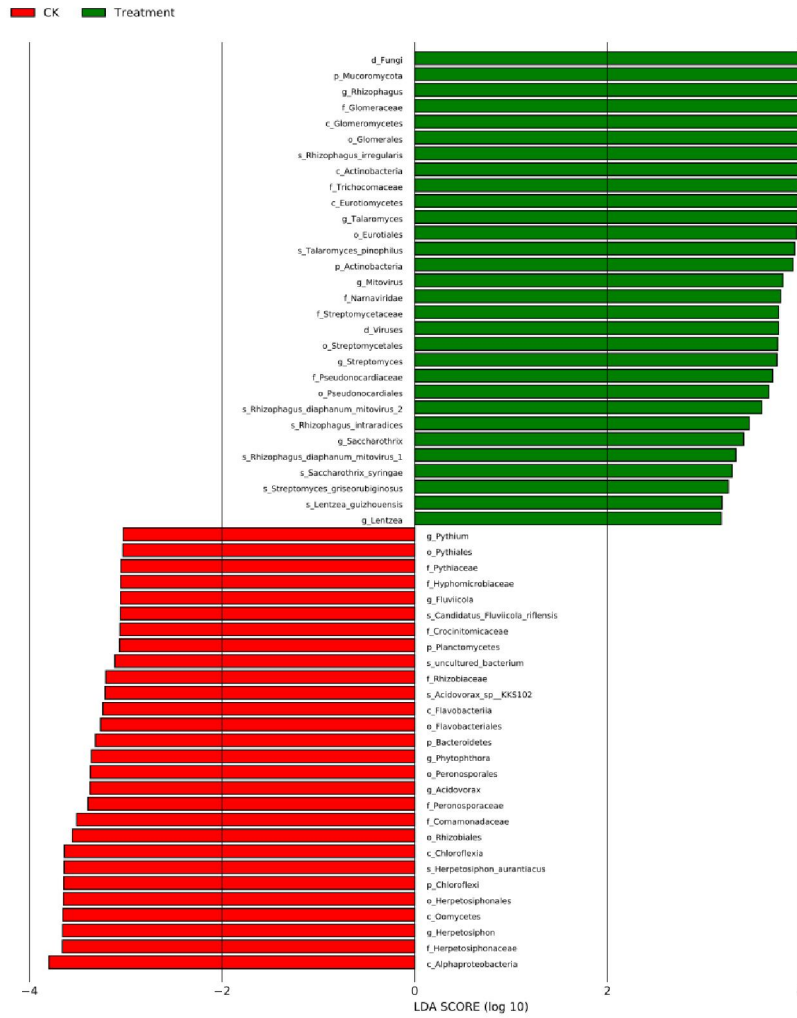


Figure 5 Taxons with significant difference between treatments in maize.

Note: Ck, the control sample of maize; Treatment, the maize roots inoculated with *R. irregularis* DAOM-197198

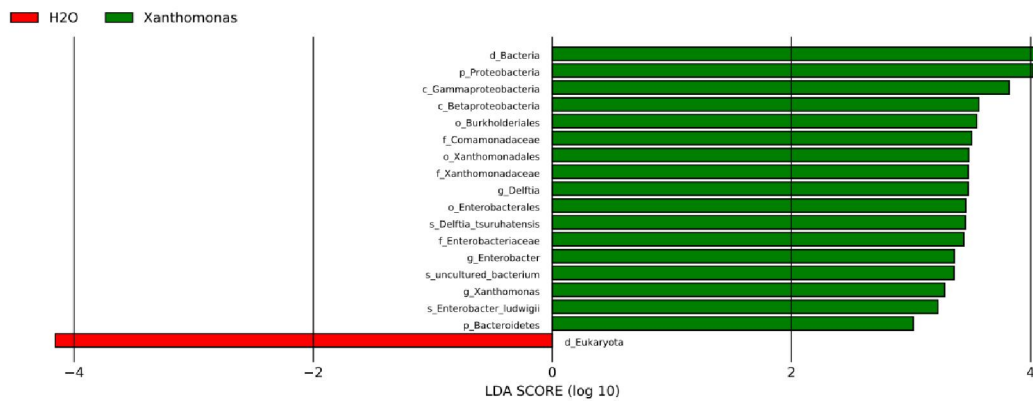


Figure 6 Taxons with significant difference between treatments in common bean.

Note: H2O, the control sample of common bean; Xanthomonas, the common bean sample inoculated with *X. phaseoli* pv. *phaseoli* strain CFBP6546R

## Conclusions

We developed a method of mining endophytic microbiome information from plant or animal transcriptome data, which is reliable and useful in investigating microbiome information while investigating the gene expression information of the host plant or animal.

## References

- Ali, M., *et al.* Maize endophytic microbial-communities revealed by removing PCR and 16S rRNA sequencing and their synthetic applications to suppress maize banded leaf and sheath blight. *Microbiol Res* 2021;242.
- Aslam, S., *et al.* Recent advances in molecular techniques for the identification of phytopathogenic fungi - a mini review. *J Plant Interact* 2017;12(1):493-504.
- Breitwieser, F.P. and Salzberg, S.L. Pavian: interactive analysis of metagenomics data for microbiome studies and pathogen identification. *Bioinformatics* 2020;36(4):1303-1304.
- Carrion, V.J., *et al.* Pathogen-induced activation of disease-suppressive functions in the endophytic root microbiome. *Science* 2019;366(6465):606-612.
- Dissanayake, A.J., *et al.* Direct comparison of culture-dependent and culture-independent molecular approaches reveal the diversity of fungal endophytic communities in stems of grapevine (*Vitis vinifera*). *Fungal Divers* 2018;90(1):85-107.
- Fadji, A.E., Ayangbenro, A.S. and Babalola, O.O. Shotgun metagenomics reveals the functional diversity of root-associated endophytic microbiomes in maize plant. *Current Plant Biology* 2021;25:100195.
- Fadji, A.E., Ayangbenro, A.S. and Babalola, O.O. Unveiling the putative functional genes present in root-associated endophytic microbiome from maize plant using the shotgun approach. *J Appl Genet* 2021;62(2):339-351.
- Foucher, J., *et al.* Common bean resistance to *Xanthomonas* is associated with upregulation of the salicylic acid pathway and downregulation of photosynthesis. *BMC Genomics* 2020;21(1):566.
- Han, G.M., *et al.* Identification of Long Non-Coding RNAs and the Regulatory Network Responsive to Arbuscular Mycorrhizal Fungi Colonization in Maize Roots. *Int J Mol Sci* 2019;20(18).
- Han, G.M., *et al.* Genome-wide association study leads to novel genetic insights into resistance to *Aspergillus flavus* in maize kernels. *Bmc Plant Biol* 2020;20(1).
- Kaul, S., Sharma, T. and Dhar, M.K. "Omics" Tools for Better Understanding the Plant-Endophyte Interactions. *Front Plant Sci* 2016;7.
- Lata, R.K., Divjot, K. and Nath, Y.A. Endophytic microbiomes: biodiversity, ecological significance and biotechnological applications. *Research Journal of Biotechnology Vol* 2019;14:10.
- Lu, J., *et al.* Bracken: estimating species abundance in metagenomics data. *PeerJ Computer Science* 2017;3.
- Marquez, L.M., *et al.* A virus in a fungus in a plant: Three-way symbiosis required for thermal tolerance. *Science* 2007;315(5811):513-515.
- Matsumoto, H., *et al.* Bacterial seed endophyte shapes disease resistance in rice. *Nat Plants* 2021;7(1).
- Shekhawat, K., *et al.* Root endophyte induced plant thermotolerance by constitutive chromatin modification at heat stress memory gene loci. *EMBO Rep* 2021;22(3):e51049.
- Tian, X.L., *et al.* Infection of *Plasmodiophora brassicae* changes the fungal endophyte community of tumorous stem mustard roots as revealed by high-throughput sequencing and culture-dependent methods. *Plos One* 2019;14(6).

Trivedi, P., *et al.* Plant-microbiome interactions: from community assembly to plant health. *Nat Rev Microbiol* 2020;18(11):607-621.

Wood, D.E., Lu, J. and Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol* 2019;20(1):257.

Table S1 The cleaning and mapping information of the maize RNA-seq data

Treatment	Sample id	Input Read Pairs	Both Surviving Pairs	Aligned concordantly 0 times	Aligned concordantly exactly 1 time	Aligned concordantly >1 times
CK	40-CK-1	43293767	37896934	24.24%	68.02%	7.74%
CK	40-CK-2	53402529	47289618	18.25%	74.14%	7.60%
CK	40-CK-3	55730952	50485352	24.19%	68.27%	7.54%
AM fungus	40-GI-1	51811515	45108797	28.26%	64.73%	7.00%
AM fungus	40-GI-2	57804926	49615744	23.77%	69.48%	6.75%
AM fungus	40-GI-3	53450977	46803974	23.27%	69.84%	6.89%

Note: Ck, the control sample of maize; Treatment, the maize roots inoculated with *R. irregularis* DAOM-197198

Table S2 The cleaning and mapping information of the common bean RNA-seq data

Treatment	Sample id	Input Read Pairs	Both Surviving reads	Aligned concordantly 0 times	Aligned concordantly exactly 1 time	>Aligned concordantly >1 times
H2O	SRR12318286	29080436	28032538	5.65%	90.78%	3.57%
H2O	SRR12318287	30814752	29503109	5.94%	90.53%	3.53%
H2O	SRR12318288	30404194	29090030	5.27%	90.76%	3.96%
H2O	SRR12318289	31867409	30251534	5.61%	90.47%	3.91%
H2O	SRR12318290	26394779	25461066	6.51%	90.68%	2.80%
H2O	SRR12318291	27198114	26091099	6.81%	90.39%	2.80%
Xanthomonas	SRR12318292	15056487	14362639	10.14%	86.96%	2.89%
Xanthomonas	SRR12318293	14962446	14202256	10.72%	86.45%	2.83%
Xanthomonas	SRR12318294	18387872	17490043	11.04%	85.69%	3.27%
Xanthomonas	SRR12318295	19056924	18003970	11.64%	85.17%	3.19%
Xanthomonas	SRR12318296	16698533	16043346	9.28%	86.48%	4.24%
Xanthomonas	SRR12318297	16563339	15850911	9.86%	85.96%	4.18%

Note: H2O, the control sample of common bean; Xanthomonas, the common bean sample inoculated with *X. phaseoli* pv. *phaseoli* strain CFBP6546R