Comparative metagenomics analysis of the rhizosphere microbiota influence on *Radix Angelica sinensis* in different growth soil environments in China

Hui YAN[1]*# , Lei ZHU[1]#, Yingjun WANG[1], Sen ZHANG[1], Pei LIU[1], Tina T. X. DONG[2], Qinan WU[1], Jin-Ao DUAN[1]*

Abstract

We studied the correlation between the soil microbial status and the crop yield or medicinal properties of *Angelica s.*. Soil from different counties of China were collected and used in this study. Readfq, SOAP denovo, MetaGeneMark, and DIAMOND Resistance Gene Identifier (RGI) software were used in the analysis. A total of 131,345 genes of high quality were obtained, with the number of genes in all samples ranging from 483-42,545 and the average length ranging from 354.88-523.47. The acidobacteria in the non-sterilized group were significantly increased by comparing with the sterilized group, while actinobacteria in the non-sterilized group were significantly reduced by comparing with the sterilized group. There was a strong correlation between the two groups of the planting group and the two groups of the non-planting group, while the correlation between the planting and non-planting group was very weak. There were significant differences in the soil microbial community structure between Min county and Yunnan, as well as significant differences between the dominant genera of Min county and Yunnan. In summary, our study showed that the soil microbial flora in different areas affected the characteristics of *Angelica s.* roots, while sterilized soil was not conducive to the growth of *Angelica s.*

Keywords: metagenomics; *Radix Angelica sinensis*; rhizosphere microbiota; soil environment; traditional Chinese medicine.

Practical Application: Study on the Chinese medicine *Radix Angelica sinensis* in depth and find the way to improve the environment for this nutrition's development.

1 Introduction

Radix Angelica sinensis (*Angelica s.*) is a perennial herb in China. The root of *Angelica s.* is terete and branched with a predominantly fleshy fibrous root and full-bodied aroma. *Angelica s.* needs low temperatures, long sunshine, and an appropriate cold cool climate. *Angelica s.* is suitable for cultivation in sandy loam with good drainage and rich humus, and mainly produced in southeast Min county, Gansu Province, with large output and good quality, followed by Yunnan, Sichuan, Shaanxi, Hubei, and other provinces (Wei et al., 2016). *Angelica s.* has significant effects on sepsis (Wang et al., 2006), promotes the proliferation of bone cells (Yang et al., 2002), maintains stem cell capacity (Mu et al., 2017), reduces diabetic kidney injury (Sui et al., 2019), has anti-inflammatory activity (Lv et al., 2018), and has an anti-tumor effect (Yan et al. 2014).

The characteristics of growing soil may greatly affect the yield and character of *Angelica s.*, and the effects of soil microbes on plants are also important (Miki, 2012; Kandasamy et al., 2019; Stolze et al., 2015). Metagenomics avoids the isolation and culture of microorganisms in samples and provides a way to study microorganisms that cannot be isolated and cultured. Metagenomics reflects the composition and interaction of microorganisms in samples more realistically and studies the metabolic pathways and gene functions at the molecular level (Garrido-Cardenas & Manzano-Agugliaro, 2017; Mackelprang et al., 2017; Bengtsson-Palme et al., 2017).

With the rapid development of sequencing and information technologies in recent years, next-generation sequencing (NGS) applied to metagenomics can quickly and accurately obtain abundant biological data, and thus has become an important method to study microbial diversity and community characteristics (Tringe et al., 2005; Luo et al., 2014).

In the current study, *Angelica s.* seedlings were cultivated in sterilized and non-sterilized soils from different producing areas (Min county, Gansu Province/Weixi, and Yunnan Province), and comparative macrogenomic analysis was performed to determine the main microbial clusters that may affect the yield and traits of *Angelica s.*

2 Methods

2.1 Collection of *Angelica s.* plant samples

Based on preliminary research involving two experiments, Min county, Gansu province and Weixi, Yunnan province, were selected for long-term cooperation.

Soil (brown red soil, pH 6.2) from Shili town, Min county, Gansu province (gray brown soil, pH 7.8) and Taicheng town, Weixi county, Yunnan province (brown red soil, pH 6.2) were obtained and transported to Shili town, Min county, Gansu province (E 103° 59° 10, N 34° 25° 28, 2288 m). The seedlings were cultivated by Xhucai traditional Chinese medicine planting professional cooperative in Xizhai town, Min county, and transplanted in the traditional planting period. Without the use of chemical fertilizers, plant hormones and other inputs, *Angelica s.* was harvested 1 year after planting.

2.2 Collection of rhizosphere soil samples

The seeds and seedlings came from a *Angelica S.* seedling base, which was the planting instruction station of traditional Chinese medicinal materials in Min county. Each pot of rhizosphere soil (root peripherals) was 30-50 g, and 500 g of root peripherals were obtained from 4 blank samples, and 500 g of root peripherals were obtained from 4 sample groups (Gansu, Yunnan, sterilized and non-sterilized), and stored in a refrigerator at -20°C.

2.3 Comparative macrogenomics database construction

To further clarify the soil microorganism composition in different areas, soil representing different conditions were collected for comparative macrogenomics analysis. Experiments were set up in eight treatment groups, as follows: Min county, Gansu province, non-sterilized sterilized and *Angelica s.* not planted, blank control group (MTWK); Min county, Gansu province, sterilized soil and *Angelica S.* not planted, blank control group (MTMK); Min county, Gansu province, non-sterilized soil and *Angelica s.* planted group (MTWMJ); Min county, Gansu province, sterilized soil and *Angelica s.* planted group (MTMJ); Weixi county, Yunnan province non-sterilized soil and *Angelica s.* not planted, blank control group (YTWK); Weixi county, Yunnan province, sterilized soil and *Angelica s.* not planted blank control group (YTMK); Weixi county, Yunnan province, non-sterilized soil and *Angelica s.* planted group (YTWMJ); and Weixi county, Yunnan, sterilized soil and *Angelica s.* planted group (YTMJ).

2.4 Sequencing results pre-treatment

Pre-processing the raw data obtained from the Illumina HiSeq sequencing platform using Readfq (V8; GitHub Inc., 2020a) was performed to acquire the clean data for subsequent analysis. The specific processing steps were as follows: a) removing the reads which contained low-quality bases (default quality threshold value ≤ 38) above a certain portion (default length = 40 bp); b) removing the reads in which the N base had reached a certain percentage (default length = 10 bp); and c) removing reads which shared the overlap above a specific portion with Adapter (default length = 15 bp). Considering the possibility that host pollution may exist in samples, clean reads were aligned to the host database by Bowtie 2.2.4 software (Johns Hopkins University, 2020) to filter the reads that are of host origin; the parameters were as follows: --end-to-end; --sensitive; -I 200; and -X 400.

2.5 Metagenome assembly

Clean reads were assembled by SOAP denovo software (V2.04), the parameters were as follows: -d 1; -M 3; -R; -u; -F; and -K 55. MEGAHIT software (v1.0.4-beta) was used to assemble the clean reads and the parameters were -presets meta-large(-- min-count 2, --k-min 27, --k-max 87, --k-step 10), then the assembled Scaftigs were interrupted from N to scaffolds. Clean reads from all samples were aligned to each scaffold by Bowtie 2.2.4 software to acquire the reads which were not aligned and the parameters were as follows: --end-to-end; --sensitive; -I 200; and -X 400. The reads which were not aligned to the scaffolds were combined, then assembled by SOAP denovo (V2.04) with the parameters which were the same as the single sample assembly; the assembled Scaftigs were interrupted from N to scaffolds. The fragments which were shorter than 500 bp in all Scaftigs were filtered out for statistical analysis.

2.6 Gene prediction and abundance analysis

The ORFs were predicted by MetaGeneMark (V2.10; Georgia Institute of Technology, 2020) software of Scaftigs (≥ 500 bp), which were assembled from both single and mixed samples. The ORFs shorter than 100nt were filtered out. CD-HIT software (V4.5.8) was adopted to do ORF clustering and obtain the unique initial gene catalogue. The parameters option was -c 0.95, -G 0, -aS 0.9, -g 1, and -d 0. The clean reads of each sample were mapped to the initial gene catalogue using Bowtie 2.2.4 to obtain the number of mapped reads of each gene catalogue and the parameters were as follows: --end-to-end; --sensitive; -I 200; and -X 400. Based on the number of mapped reads and the length of the genes, statistical analysis of the abundance profile of each gene in each sample was presented. The basic information statistic, core-pan gene analysis, correlation analysis of samples, and Venn analysis of number of genes were all based on the abundance profile.

2.7 Taxonomy prediction

DIAMOND software (V0.9.9; GitHub Inc., 2020b) was used to align the Unigenes to the sequences of bacteria, fungi, archaea, and viruses which were all extracted from the NR database (version 20161115; National Center for Biotechnology Information, 2020b). The LCA algorithm was used to do the species annotation with MEGAN software. The table containing the number of genes and the abundance information of each sample in each taxonomy hierarchy (kingdom, phylum, class, order, family, genus, and species) were obtained based on the LCA annotation result and the gene abundance table. The abundance of a species in one sample was equal to the total number of the gene abundance profile of the species. Krona analysis, the exhibition of generation situation of relative abundance, the exhibition of abundance cluster heat map, PCA (R ade4 package, version 2.15.3), and NMDS (R vegan package, version 2.15.3) decrease-dimension analysis were based on the abundance table of each taxonomic

hierarchy. Hypothetical testing between groups is presented by Anosim (R vegan package, version 2.15.3). Metastats and LEfSe analysis were used to identify the different species between groups. Permutation testing was used in Metastats analysis for each taxonomy and the P value was obtained, then the Benjamini and Hochberg False Discovery Rate was used to correct the P value and acquire the q value. LEfSe analysis was conducted by LEfSe software (the default LDA score = 3).

2.8 Common functional database annotations

DIAMOND software (V0.9.9) was used to align Unigenes to the functional database. The parameter setting was "blastp, -e 1e-5." The functional database consisted of the KEGG database (version 201609), eggNOG database (version 4.5), and CAZy database (version 20150704). The relative abundance of each functional hierarchy was equal to the sum of relative abundance annotated to the functional level. The gene number table of each sample in each taxonomy hierarchy was obtained from the function annotation result and gene abundance table. PCA and NMDS analyses of the general relative abundance were performed with the default parameter. The Anosim analysis between groups (inside), which was based on the functional abundance, is presented. Comparative analysis of metabolic pathways, and the Metastat and LEfSe analyses between groups were performed.

2.9 Resistance gene annotation

Resistance Gene Identifier (RGI) software was used to align the Unigenes to CARD database. The parameter settings were "blastp, -e 1e-30." The relative abundance of Antibiotic Resistance Ontology (ARO) was counted from the alignment. Bar charts and heatmap were used to show the difference between groups.

3 Results and discussion

3.1 Character differences of *Angelica s.* under different planting conditions

The results showed that *Angelica s.* growing in sterilized soil had weak roots and fewer branches. The root system of *Angelica s.* grown in non-sterilized soil was developed, with more and stronger branches. The roots and branches of *Angelica s.* cultivated in Yunnan soil were small and weak, and not as developed as Min county in Gansu province. Even if the Yunnan soil was sterilized, the cultivated *Angelica s.* roots and branches were still small and weak, and not as developed as those in Min country, Gansu province (Figure 1). This finding suggested that there may be other factors besides soil microorganisms influencing crop traits.

3.2 Comparative macrogenomics database construction and data processing

A total of 131,345 genes of high quality were obtained with the number of genes in all samples ranging from 483-42,545 and the average length ranging from 354.88-523.47. The number of genes in the Yunnan blank sterilization group was 75,098.

The not planted blank control group in Yunnan had the fewest number (483). The average gene length of Yunnan soil was longer than Min country. The GC content of all samples ranged from 41.79-67.55%, and were all different, indicating that the soil microbial composition of all samples were different. There were clear differences between the number and length of the soil gene catalogue in Yunnan and Gansu (Table 1). Compared with the non-sterilized group, more genes were assembled in the sterilized group. The number of genes in the Yunnan sterilization group was more than the Min county sterilization group, while the number of genes in the Min county non-sterilization group was more than the Yunnan non-sterilization group. Compared with the implant group, there were more genes in the non-implant group, except for the non-sterilized control group in Yunnan.

Thus, the sequence length distribution of each group in Min county was relatively stable, while each group in Yunnan was relatively unstable, mainly manifested by the short sequence length of the YTWK group. The GC content distribution of all groups was relatively discrete, which verifies the reliability of the experimental results. The mean GC content of the MTWJ, YTWJ, MTMJ, YTMJ, MTWK, YTWK, MTMK, and YTMK groups was 62.47%, 56.17%, 56.45%, 33.81%, 34.40%, 42.28%, 37.29%, and 32.13%, respectively.

There were significant differences in the soil species composition between Yunnan and Gansu. Compared with the non-sterilized soil planting group in Yunnan, the GC content in the non-sterilized soil planting group in Min county was significantly higher. The same results were found with respect to Cyanobacteria in the soil sterilization groups of Min county and Yunnan province (Figure 2).

The microbial diversity of Min county soil was greater, whether or not the soil was sterilized. The soil GC content in the planting group was significantly higher than the non-planting group, and the GC content in the non-sterilization group was higher than the sterilization group.

3.3 Sample correlation analysis based on gene number

There was a strong correlation between the two groups in the planting group and the two groups in the non-planting group, while the correlation between the planting and non-planting groups was very weak. There was a strong correlation between the MTWMJ and YTWMJ groups, the MTWMJ and MTMJ groups, the YTWK and MTWK groups, the YTMJ and MTMJ groups, the YTWK and YTMK groups, and the YTWMJ and YTMJ groups (Figure 3). It is worth noting that the correlation between the MTWMJ and YTWMJ groups was the strongest, followed by the MTWMJ and MTMJ groups, and the YTWK and MTWK groups. These results indicated that the planting of *Angelica s.* had a great influence on the community composition of rhizosphere soil microorganisms.

3.4 Species relative abundance profile

The differences in species composition between all groups were analyzed at the phylum and genus levels (Figure 4). At the phylum level, the root microflora of *Angelica* in all types of soil were dominated by Proteobacteria (relative abundance = 86.92%),

Figure 1. Soil origin and soil microbial status affect the root character of *Angelica sinensis*. (A) Photographs of *Angelica s.* in sterilized soil of Gansu (a1) and non-sterilized soil of Gansu (a2); and (B) in sterilized soil of Yunnan (b1) and non-sterilized soil of Yunnan (b2); (C) Photographs of *Angelica s.* in soil of Yunnan (c1) and soil of Gansu (c2); (D) in sterilized soil of Yunnan (d1) and sterilized soil of Gansu (d2).

Table 1. Statistical table of basic gene catalogue information.

Sample ID	Gene Number	Gene Length (bp)	Gene Length/ Genome(%)	Gene Average Length (bp)	GC Content in Gene Region (%)
MTMJ	12245	4541025	70.88%	370.85	49.38%
MTMK	37262	15536694	81.98%	416.96	60.46%
MTWK	5352	2154663	78.62%	402.59	62.07%
MTWMJ	1388	492567	64.64%	354.88	41.79%
YTMJ	31433	16454151	85.64%	523.47	66.31%
YTMK	42545	19179549	86.20%	450.81	67.55%
YTWK	637	399560	69.61%	627.25	56.95%
YTWMJ	483	194469	51.55%	402.63	50.80%

Description: gene number represents the number of genes in the gene catalogue. Gene length (bp): Gene length/genome (%) indicates the percentage of total gene length in the genome. Gene average length (bp) represents the average gene length. GC content in the gene region (%) represents the overall GC content value of genes within the predicted gene catalogue.

Actinobacteria (6.43%), and Bacteroidetes (2.79%), accounting for 96.14% of the total flora. Cyanobacteria in the MTMJ (3.22%) and MTWK groups (12.92%) were significantly enriched. In the YTWK group (4.43%), Acidobacteria was significantly enriched. Actinobacteria were significantly enriched in the YTMJ (10.47%) and YTMK groups (16.85%; $p<0.05$). Compared with the non-sterilized group in Min county, the Yunnan non-sterilized group, Bacteroidetes, and Acidobacteria were significantly increased, while Actinobacteria was significantly decreased. In contrast, compared with the sterilization group in Min county, the Yunnan sterilization group had fewer Bacteroidetes, while Actinobacteria and Acidobacteria were increased. In addition, compared with the sterilized group, Acidobacteria and Actinobacteria in the non-sterilized group

A

B
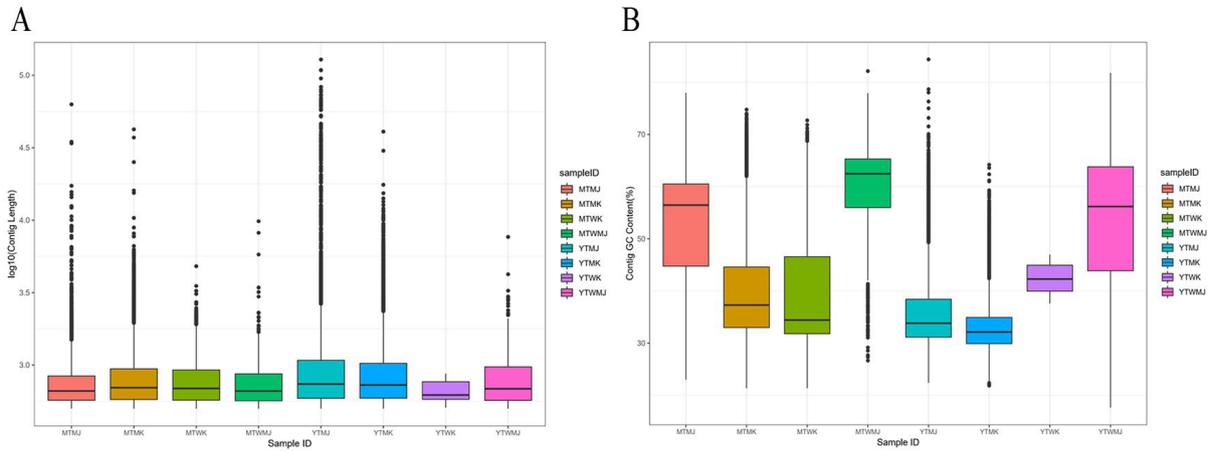
Figure 2. The length distribution of the assembly sequence. Contig length boxplot (A) and contig GC boxplot (B).
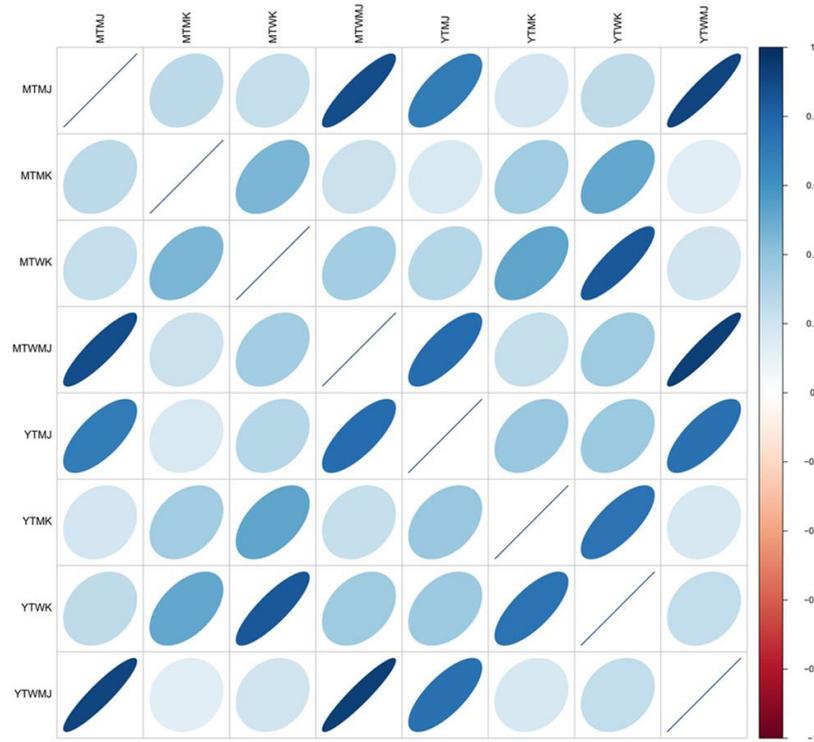
Figure 3. The heat map of correlation coefficient between samples. Different colors represent the level of Spearman correlation coefficient. The relationship between correlation coefficient and color is illustrated in the legend on the right. The darker the color, the greater the absolute value of the correlation coefficient between samples. The left deviation of the ellipse indicates that the correlation coefficient is positive and the right deviation is negative. The flatter the ellipse, the greater the absolute value of the correlation coefficient.

were significantly increased and decreased, respectively. There was no significant difference between the planting and non-planting groups.

At the genus level, 61.24% of the top 10 genera in all groups were found. *Variovorax* was significantly enriched in the MTMJ (10.98%) and MTWMJ groups (13.40%). The short wave monomonas in the MTMK (25.45%) and YTWK groups (13.90%) were significantly enriched by *Brevundimonas*. For the YTMJ (35.52%) and YTMK groups (19.65%), *Rhodanobacter* bacteria were significantly enriched (p<0.05). In the Yunnan

planting group, compared with other Min county planting groups, *Variovorax* and *Sphingopyxis* were significantly decreased, while *Mesorhizobium* was significantly increased. It is noteworthy that, except for the YTWMJ group, the abundance of *Pseudomonas* of soil samples in the Yunnan group was significantly reduced. For Min country soil, *Rhodopseudomonas* in the non-sterilized group was more than the sterilized group and *Caulobacter* was less than the sterilized group. For Yunnan soil, *Rhodopseudomonas* and *Mesorhizobium* of *Rhodopseudomonas* and *Mesorhizobium*
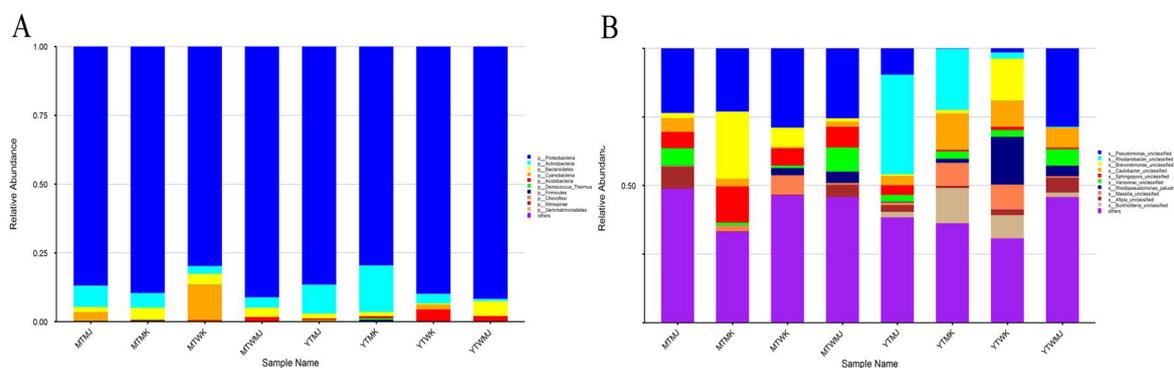
Figure 4. Column chart of relative abundance of species at the phylum and genus levels (samples). (A) Column diagram of relative abundance of the gate level; (B) Column diagram of horizontal relative abundance; the horizontal axis represents the sample name, and the vertical axis shows the relative proportion of species annotated to a certain type. The species categories corresponding to each color block are shown in the legend on the right.

of mesorobacter were more common in the non-sterilized group than the sterilized group, while *Rhodanobacter obacter* was less. Compared with the planting group, there were fewer V*ariovora* and *Afipia,* and more *Brevundimonas* and *Massilia* in the non-planting group.

3.5 Cluster analysis of the number and relative abundance of annotated genes

There were significant differences in the soil microbial community structure between Min county and Yunnan, as well as significant differences between dominant genera Min county and Yunnan (Figure 5). Specifically, *Pseudomonas*, *Sphingopyxis*, Agrobacterium bacteria (*rhizobia*) and *Pedobacter* (*Sphingomonas*) of *Pseudomonas* were significantly higher in Min county soil than Yunnan soil. In contrast, *Caulobacter*, *Asticcacaulis*, *Mesorhizobium*, and *Burkholderia* belonged to the genus *Stibaciaceae* and both belonged to the family. *Stibaciaceae*, *Mesorhizobium*, *Mesorhizobium*, and *Burkholderia*, which were significantly higher in Yunnan soils than in Min county soils.

3.6 GO annotation

The Gene Ontology [GO] (Figure 6) is an internationally standardized classification system of gene functions that provides a controlled vocabulary of dynamically updated information about the properties of genes and their products in organisms. GO has three ontologies, which describe the molecular function, cellular component, and biological process of genes.

3.7 COG annotation

The orthologous database (Cluster of Orthologous Groups of proteins [COG]) is a commonly used protein annotation database (National Center for Biotechnology Information, 2020a; Figure 7). By means of pin-like clustering of genes of different races, genes can be divided into various direct homologous clusters so that the function of unknown genes can be annotated by the known genes in the same cluster. This method can predict the biological function of the unknown "open reading framework"

by searching for the direct homologues of the unknown genes at the genomic level. BLAST software default parameters were used to compare gene sequences with the COG database, and the comparison results were filtered to remove the alignment sequences with an e_value greater than 1e-5. In addition, the best alignment sequence among multiple sequences of the same sequence was selected for statistical analysis of the comparison results.

3.8 Unifrac analyses

Analysis of the distance between the samples was performed by the weighted unifrac (only considering the abundance of species) and unweighted unifrac distances (and thinking about the species abundance and species evolution relationship) to measure these samples of different coefficient. The value was smaller, and the species diversity and differences among the samples were smaller. The abundance of species is plotted in Figure 8.

3.9 Discussion

Angelica s., as a medicinal resource with great developmental value, is one of the common Chinese traditional medicines to invigorate the human body, with a long history and clinical basis. *Angelica s.* is a perennial herb, with a sweet taste and the effect of invigorating blood, activating blood circulation, regulating menstruation, relieving pain, and embellishing bowel (Bengtsson-Palme et al., 2017).

The root system development and yield of *Angelica s.* produced in different regions were different, as were the content of trace elements and the medicinal effect. This finding was closely related to the species and population of soil microorganism. Microorganisms are the most active, diverse, and abundant components in soil. The microorganisms have unique functions in the soil ecosystem and play an important role in the earth's material circulation, energy conversion, environment, and health (Cicatelli et al., 2019).

With the advances in DNA sequencing technology and the continuous improvement of bioinformatics analysis in
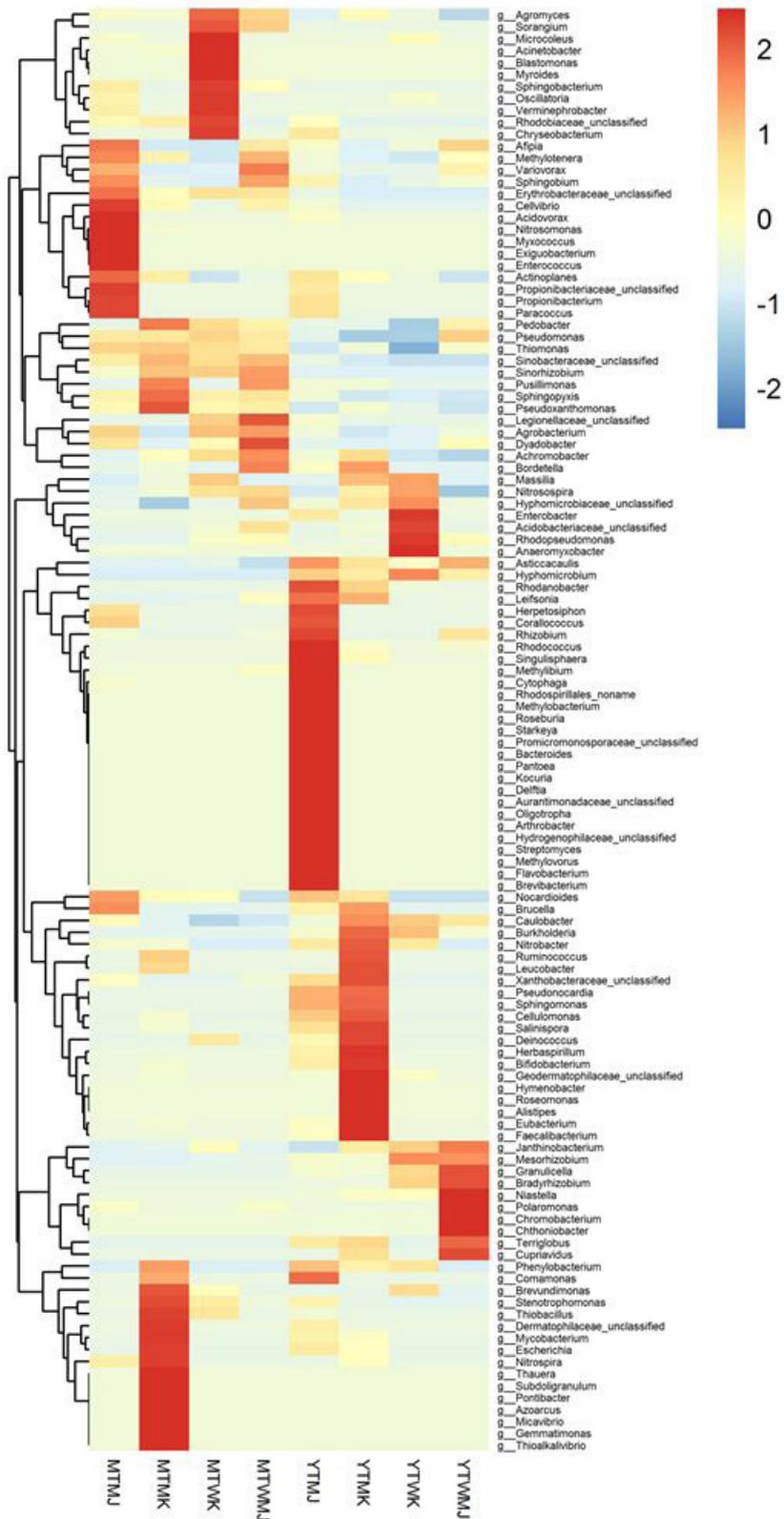
Figure 5. Clustering heat map of number and abundance of generic horizontal genes. a) Unigenes notes the number of the statistical heat map. The horizontal axis is the sample name. The vertical axis is the species information. Different colors represent the height of the Unigenes number. b) Cluster heat map of relative abundance of the genus level (transverse sample information; longitudinal is species information). The clustering tree on the left of the figure is the species clustering tree. The value corresponding to the intermediate heat map is the z-value obtained after standardization treatment for the relative abundance of the species in each row. Thus, the z-value of a sample in a classification is the value obtained by dividing the difference between the relative abundance of samples in that classification and the average relative abundance of all samples in that classification by the standard deviation of all samples in that classification.
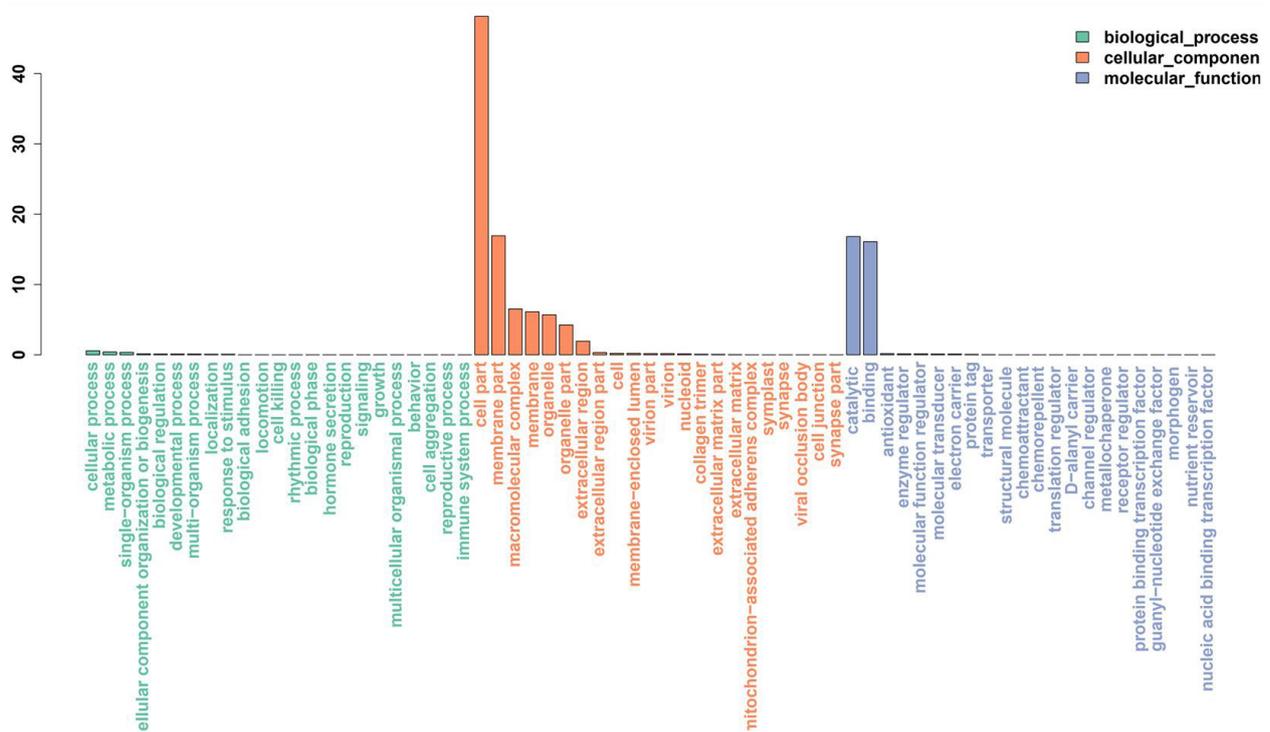
Figure 6. GO annotation graphic. The horizontal coordinate represents the GO term, and the vertical coordinate represents the proportion corresponding to each term. Three different colors were used to represent the three different classes, in the order from highest-to-lowest of each class.
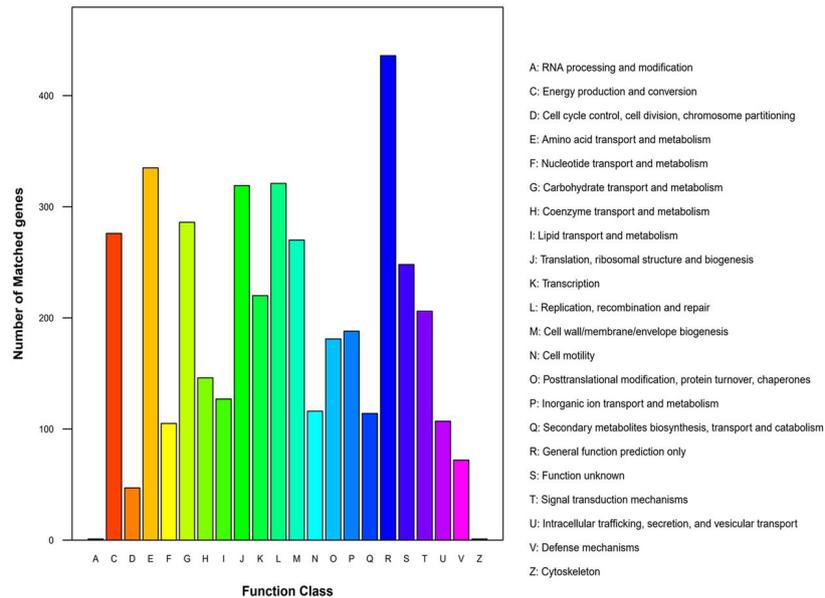
Figure 7. COG annotation graphic.

recent years, metagenomics is considered to be a new way to study the genomics of microbial communities (Ngara & Zhang, 2018). Metagenomics takes the total genome of microorganisms in environmental samples, and selects and analyzes the functional genes to study the structure, diversity, and evolutionary relationship between the microbial community and the environment (Tringe et al., 2005). Therefore, we chose comparative metagenomics to study the interaction between

Angelica s. and the rhizosphere microorganisms in the planting environment.

The root characteristics of *Angelica s.* were compared, and cultivated in sterilized or non-sterilized soil in different areas. The local soil in Gansu province was shown to significantly promote the root growth of local *Angelica s.*, while the soil in other regions of Yunnan did not produce relatively robust and developed roots.
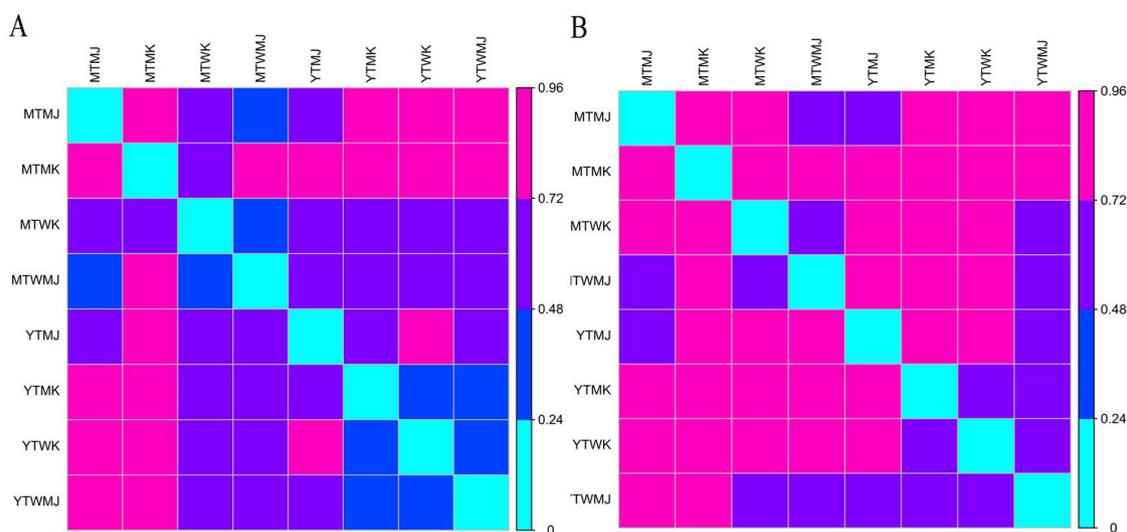
Figure 8. Wang ghted unifrac matrix (A) and unweighted unifrac matrix (B).

This phenomenon is probably related to the rhizosphere microbiome in the soil. Through comparative metagenomics analysis, we found that this phenomenon was highly correlated with rhizosphere microorganisms. Different soil microbial flora could affect the developmental degree of *Angelica s.* root, and *Angelica s.* could also affect the distribution of microbiota around the root.

After planting *Ange-lica s.* in sterilized or non-sterilized soil, the distribution of the rhizosphere microbial flora changed. *Angelica s.* and soil microbial flora were mutually regulated. Indeed, some strains of bacteria thrived and reproduced better around the roots of *Angelica s.*

We also found that *Pseudomonas* were the dominant species in Gansu soil, while the species and genera of bacteria in Yunnan soil were distributed evenly with less abundant. In addition, *Rhodopseudomonas palustris*, which promotes plant growth, was significantly higher in the non-sterilized group. After planting *Angelica s.*, the proportion of *Pseudomonas* in Yunnan soil was significantly increased. *Pseudomonas* can kill insects, improve plant nutrition, promote the transformation and decomposition of soil substances, produce plant growth regulatory substances, and induce systemic resistance (Kuzmanović et al., 2018; Prabhukarthikeyan et al., 2018). The study suggested that a high proportion of *Pseudomonas* in the soil of Gansu province plays an important role in promoting the branches and growth of *Angelica s.* root system.

Shown as in the study, rhizosphere microbiota and plant roots are interdependent and interact with each other. Different rhizosphere microbiota participate in the regulation of plant growth through the transformation of soil substances. When farming, we can tend to add the corresponding beneficial strains, which can significantly promote the root growth of *Angelica s.* and affect the medicinal properties.

4 Conclusion

In this study, local soil in Gansu province with abundant *Pseudomonas* promotes the root growth of local *Angelica S.*, while the soil in Yunnan with less abundant *Pseudomonas* cannot produce relatively robust and developed roots. After planting *Angelica s.*, the proportion of *Pseudomonas* in Yunnan soil was increased significantly. A high proportion of *Pseudomonas* in the soil of Gansu province plays an important role in promoting the branches and growth of the *Angelica s.* root system. The study showed that the microbial environment of soil in different regions is different, and the soil microbial flora in different areas affected the characteristics of Angelica s. roots, while sterilized soil was not conducive to the growth of *Angelica s.*

Acknowledgements

References

Bengtsson-Palme, J., Larsson, D. G. J., & Kristiansson, E. (2017). Using metagenomics to investigate human and environmental resistomes. *The Journal of Antimicrobial Chemotherapy*, 72(10), 2690-2703. http://dx.doi.org/10.1093/jac/dkx199. PMid:28673041.

Cicatelli, A., Ferrol, N., Rozpadek, P., & Castiglione, S. (2019). Effects of plant-microbiome interactions on phyto- and bio-remediation capacity. *Frontiers of Plant Science*, 10, 533. http://dx.doi.org/10.3389/fpls.2019.00533. PMid:31105729.

Garrido-Cardenas, J. A., & Manzano-Agugliaro, F. (2017). The metagenomics worldwide research. *Current Genetics*, 63(5), 819-829. http://dx.doi.org/10.1007/s00294-017-0693-8. PMid:28401295.

Georgia Institute of Technology. (2020). *GeneMark.* Retrieved from http://topaz.gatech.edu/GeneMark/

GitHub Inc. (2020a). *Readfq.* Retrieved from https://github.com/cjfields/readfq

GitHub Inc. (2020b). *Diamond.* Retrieved from https://github.com/bbuchfink/diamond/

Johns Hopkins University. (2020). *Bowtie 2.2.4.* Retrieved from http://bowtie-bio.sourceforge.net/bowtie2/index.shtml

Kandasamy, S., Liu, E. Y. R., Patterson, G., Saldias, S., Ali, S., & Lazarovits, G. (2019). Introducing key microbes from high productive soil transforms native soil microbial community of low productive soil. *MicrobiologyOpen*, 8(10), e895. http://dx.doi.org/10.1002/mbo3.895. PMid:31250991.

Kuzmanović, N., Eltlbany, N., Ding, G., Baklawa, M., Min, L., Wei, L., & Smalla, K. (2018). Analysis of the genome sequence of plant beneficial strain *Pseudomonas sp* RU47. *Journal of Biotechnology*, 281, 183-192. http://dx.doi.org/10.1016/j.jbiotec.2018.07.023. PMid:30031092.

Luo, C. W., Rodriguez-R, L. M., Johnston, E. R., Wu, L. Y., Cheng, L., Xue, K., Tu, Q. C., Deng, Y., He, Z. L., Shi, J. Z., Yuan, M. M., Sherry, R. A., Li, D. J., Luo, Y. Q., Schuur, E. A. G., Chain, P., Tiedje, J. M., Zhou, J. Z., & Konstantinidis, K. T. (2014). Soil microbial community responses to a decade of warming as revealed by comparative metagenomics. *Applied and Environmental Microbiology*, 80(5), 1777-1786. http://dx.doi.org/10.1128/AEM.03712-13. PMid:24375144.

Lv, J. L., Zhang, L. B., & Guo, L. M. (2018). Phthalide dimers from *Angelica sinensis* and their COX-2 inhibition activity. *Fitoterapia*, 129, 102-107. http://dx.doi.org/10.1016/j.fitote.2018.06.016. PMid:29935258.

Mackelprang, R., Burkert, A., Haw, M., Mahendrarajah, T., Conaway, C. H., Douglas, T. A., & Waldrop, M. P. (2017). Microbial survival strategies in ancient permafrost: insights from metagenomics. *The ISME Journal*, 11(10), 2305-2318. http://dx.doi.org/10.1038/ismej.2017.93. PMid:28696425.

Miki, T. (2012). Microbe-mediated plant-soil feedback and its roles in a changing world. *Ecological Research*, 27(3), 509-520. http://dx.doi.org/10.1007/s11284-012-0937-5.

Mu, X. Y., Zhang, Y. Y., Li, J., Xia, J. Y., Chen, X. B., Jing, P. W., Song, X. Y., Wang, L., & Wang, Y. P. (2017). *Angelica sinensis* polysaccharide prevents hematopoietic stem cells senescence in D-galactose-induced aging mouse model. *Stem Cells International*, 2017, 3508907. http://dx.doi.org/10.1155/2017/3508907. PMid:28491095.

National Center for Biotechnology Information – NCBI. (2020a). *COG links.* Retrieved from http://www.ncbi.nlm.nih.gov/COG

National Center for Biotechnology Information – NCBI. (2020b). Retrieved from https://www.ncbi.nlm.nih.gov/

Ngara, T. R., & Zhang, H. J. (2018). Recent advances in function-based metagenomic screening. *Genom Proteom Bioinf*, 16(6), 405-415. http://dx.doi.org/10.1016/j.gpb.2018.01.002. PMid:30597257.

Prabhukarthikeyan, S. R., Keerthana, U., & Raguchander, T. (2018). Antibiotic-producing *Pseudomonas fluorescens* mediates rhizome rot disease resistance and promotes plant growth in turmeric plants. *Microbiological Research*, 210, 65-73. http://dx.doi.org/10.1016/j.micres.2018.03.009. PMid:29625661.

Stolze, Y., Zakrzewski, M., Maus, I., Eikmeyer, F., Jaenicke, S., Rottmann, N., Siebner, C., Pühler, A., & Schlüter, A. (2015). Comparative metagenomics of biogas-producing microbial communities from production-scale biogas plants operating under wet or dry fermentation conditions. *Biotechnology for Biofuels*, 8(1), 14. http://dx.doi.org/10.1186/s13068-014-0193-8. PMid:25688290.

Sui, Y., Liu, W. J., Tian, W., Li, X. Q., & Cao, W. (2019). A branched arabinoglucan from *Angelica sinensis* ameliorates diabetic renal damage in rats. *Phytotherapy Research*, 33(3), 818-831. http://dx.doi.org/10.1002/ptr.6275. PMid:30672023.

Tringe, S. G., von Mering, C., Kobayashi, A., Salamov, A. A., Chen, K., Chang, H. W., Podar, M., Short, J. M., Mathur, E. J., Detter, J. C., Bork, P., Hugenholtz, P., & Rubin, E. M. (2005). Comparative metagenomics of microbial communities. *Science*, 308(5721), 554-557. http://dx.doi.org/10.1126/science.1107851. PMid:15845853.

Wang, H., Li, W., Li, J., Ochani, M., Ashok, M., Yang, H., Tracey, K. J., Wang, P., & Sama, A. E. (2006). Therapeutic potential of a popular Chinese herb, *Angelica sinensis*, in experimental sepsis. *Shock*, 25(Suppl. 1), 25. http://dx.doi.org/10.1097/00024382-200606001-00075.

Wei, W. L., Zeng, R., Gu, C. M., Qu, Y., & Huang, L. F. (2016). *Angelica sinensis* in in China: a review of botanical profile, ethnopharmacology, phytochemistry and chemical analysis. *Journal of Ethnopharmacology*, 190, 116-141. http://dx.doi.org/10.1016/j.jep.2016.05.023. PMid:27211015.

Yan, D. Q., Liu, Y. Q., Li, Y. D., Li, D., Cheng, X. L., & Wu, Z. W. (2014). Tumor inhibition effects and mechanisms of angelica s.inensis and sophorae flavescentis ait decoction combined with cisplatin in xenograft mice. *Asian Pacific Journal of Cancer Prevention*, 15(11), 4609-4615. http://dx.doi.org/10.7314/APJCP.2014.15.11.4609. PMid:24969893.

Yang, Q., Populo, S. M., Zhang, J. Y., Yang, G. G., & Kodama, H. (2002). Effect of *Angelica sinensis* on the proliferation of human bone cells. *Clinica Chimica Acta*, 324(1-2), 89-97. http://dx.doi.org/10.1016/S0009-8981(02)00210-3. PMid:12204429.