Data in Brief 54 (2024) 110273

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Dataset of core and differentially abundant bacteria in various compartments of farm-cultivated and home-planted chilli plants (*Capsicum frutescens*)

Sze-Looi Song[a,*], Zheng-Xian Lew[b], Hoi-Sen Yong[c], Qi-Hang Lim[b], Rong-Heng Joshua Chai[b]

[a] *Institute for Advanced Studies, Universiti Malaya, Kuala Lumpur, Malaysia*
[b] *China-ASEAN College of Marine Sciences, Xiamen University Malaysia, Sepang, Malaysia*
[c] *Institute of Biological Sciences, Faculty of Science, Universiti Malaya, Kuala Lumpur, Malaysia*

ARTICLE INFO

Article history:
Received 14 January 2024
Revised 18 February 2024
Accepted 23 February 2024
Available online 1 March 2024

Dataset link: Microbial diversity on chilli plants (Original data)

Keywords:
Chilli pepper
Next generation sequencing
Microbiome
16S rRNA gene
Solanaceae

ABSTRACT

Chillies are members of the genus *Capsicum* L. (family Solanaceae). They are native to Central and South America and consist of approximately 35 species [1,2]. Among these, five species (*C. annuum* L., *C. baccatum* L., *C. chinense* Jacq., *C. frutescens* L., and *C. pubescens* Ruiz & Pav.) have been domesticated and are mainly cultivated for consumption as vegetables and spices. Of the domesticated chillies, *C. annuum* is commercially cultivated worldwide, while *C. frutescens* and *C. chinense* are mainly cultivated in American, Asian, and African countries [3]. We compared the diversity of microbiota in various compartments of farm-cultivated (FC) and home-planted (HP) chilli plants (*Capsicum frutescens*). Targeted 16S rRNA gene (V5-V6 region) was sequenced using the Illumina NovaSeq 6000 platform. Proteobacteria, Actinobacteriota, Acidobacteriota, Gemmatimonadota, Bacteroidota, and Firmicutes were present in all compartments of both the FC and HP plants. Proteobacteria (or Pseudomonadota) was the predominant phylum in all the compartments of both HP and FC plants, while Actinobacteriota (or Actinomycetota) was the second most abundant phylum. Most plant compartments (leaves, fruits and roots) exhibited a higher rela-

* Corresponding author.
E-mail address: szelooi@um.edu.my (S.-L. Song).

https://doi.org/10.1016/j.dib.2024.110273

tive abundance of Proteobacteria compared to the soil samples. With few exceptions, the soil compartments (bulk and rhizospheric soils) displayed a higher relative abundance of the phyla Myxococcota, Acidobacteriota, Gemmatimonadota, Bacteroidota, Nitrospirota, Verrucomicrobiota, and Firmicutes than the plant compartments. Diversity indices revealed that the bacterial community in chili plants clustered based on both compartment and cultivation area.

Specifications Table

Subject	Microbiology: Microbiome
Specific subject area	High-throughput amplicon metagenome sequence datasets of the 16S rRNA gene (V5-V6 hypervariable region) from various compartments of chili plants.
Data format	Raw, Analyzed, Filtered
Type of data	Table, Image, Chart, Graph, Figure
Data collection	A total of five sample groups were collected from both farm-cultivated and home-planted chilli plants: bulk soil (BS), rhizopheric soil (RS), root endophytes (Ren), leaves (L), and fruit endophytes (Fen). Total DNA was extracted from the sample and the 16S rRNA gene amplicon (V5-V6 region) was sequenced by the Illumina NovaSeq 6000 System (2 × 150 bp paired-end reads).
Data source location	Samples from the farm-cultivated chili plant were collected from Hoho Farm, Batang Kali, Selangor (3° 26′ 48.5″ N 101° 38′ 52.2″ E). Samples from the home-planted chilli plant were harvested from Ipoh, Perak (4° 38′ 26.9″ N 101° 07′ 45.2″ E). Data are stored at Universiti Malaya, Kuala Lumpur, Malaysia.
Data accessibility	Repository name: GenBank Sequence Read Archive [4] Data identification number: Data are available at the NCBI with BioProject PRJNA1064378 Direct URL to data: https://www.ncbi.nlm.nih.gov/Traces/study/?acc=PRJNA1064378&o=acc_s/3Aa

1. Value of the Data

- The data provide information on the core members and differentially abundant taxa of the bacterial community in various compartments of farm-cultivated and home-planted chilli plants (*Capsicum frutescens*).
- The data are useful for comparative analysis of abundance and core members of the bacterial community among different chili plant species/cultivars.
- The data are useful for culture-dependent technique on the microbiota associated with *Capsicum frutescens*.
- The data are valuable for developing a management program to maintain a healthy soil condition in cultivation areas.

2. Background

Varieties of chillies with fruits of little or no pungency (particularly *C. annuum*) are commonly known as ‘sweet pepper’, while those with strong pungency (such as *C. frutescens* and *C. chinense*) are referred to as ‘chilli pepper’. Many aspects of chilli have been widely studied, including production practices such as breeding and improvement, fruit ripening, pests and

diseases [5–7], as well as genetics, genomics, phylogenetics and systematics [1,2,8-10]. Traditional uses, phytochemistry and pharmacology of bioactive compounds [11], along with quality attributes, flavour compounds, microbial diversity and succession, and metabolome-microbiome interactions of fermented chilli [12,13], have also received considerable attention. Despite the voluminous research, some aspects remain understudied, such as the diversity and roles of naturally occurring microbiota associated with different compartments of the chilli plant. There is perhaps only a single report on the disease-induced changes in microbiome assembly and functional adaptation in *C. annuum* [14]. We report here a fundamental study on the composition and diversity of bacterial communities associated with different compartments of chilli pepper (*C. frutescens*) cultivated under both farm and home conditions.

3. Data Description

3.1. OTUs clustering

A total of 785,289 raw reads were obtained from the 10 samples. During the process, 37,271 reads were removed due to low quality reads and unable to be classified into any known Operational Taxonomic Units (OTUs). The remaining reads were clustered into 16,194 OTUs at 97% similarity threshold. In addition, OTUs classified as chloroplast and mitochondria were filtered out, resulting in 15,974 OTUs being retained for further taxonomy classification. The raw datasets for 16S rRNA gene amplicon sequencing generated for this paper have been deposited in the GenBank Sequence Read Archive (accession number PRJNA1064378).

Rarefaction analysis showed that the observed OTUs of the leaf, fruit and root samples increased and reached a plateau at a sequencing depth of 1700, whereas the bulk soil and rhizospheric soil samples increased sharply but did not reach a plateau (Fig. 1).

3.2. Bacterial community/composition

The 16S rRNA amplicons revealed a different number of bacterial OTUs between the plant (root, leaf and fruit) and soil (bulk and rhizospheric soils) compartments (Table 1). The number of bacterial OTUs at different taxonomic levels in the soil compartments was much higher than in the plant compartments (Table 1).

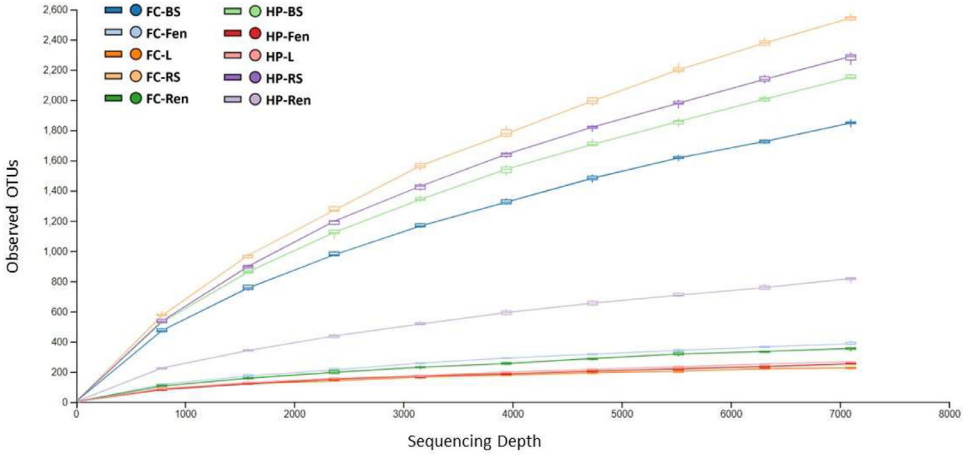

Fig. 1. Alpha rarefaction curve of bacterial community in different compartments of chilli plant (*Capsicum frutescens*).

Table 1
Number of bacterial Operational Taxonomic Units (OTUs) detected in the field-cultivated (FC) and home-planted chilli plants (*Capsicum frutescens*) and the associated soil.

OTU	FC-BS	FC-Fen	FC-L	FC-Ren	FC-RS	HP-BS	HP-Fen	HP-L	HP-Ren	HP-RS
Phylum	9	9	6	7	9	9	8	7	9	9
Class	27	5	4	5	26	27	5	4	8	24
Order	63	14	7	17	70	71	9	12	33	69
Family	75	17	12	19	97	92	15	16	44	94
Genus	16	7	8	8	16	21	7	6	13	17

Table 2
Relative abundance (%) of bacterial phyla in 5 different compartments of home-planted and farm-cultivated chilli plants. FC: farm-cultivated; HP: home-planted; Fen: fruit endophyte; L: leaf; Ren: root endophyte; BS: bulk soil; RS: rhizospheric soil.

Phylum	FC-BS	FC-Fen	FC-L	FC-Ren	FC-RS	HP-BS	HP-Fen	HP-L	HP-Ren	HP-RS
Proteobacteria	46.03	78.03	93.86	67.44	38.29	56.69	95.32	92.97	44.72	55.73
Actinobacteriota	12.06	21.01	5.89	19.66	32.39	7.61	4.20	6.44	49.38	22.72
Myxococcota	9.82	0.04	0.00	0.02	6.64	8.40	0.01	0.03	0.84	5.35
Acidobacteriota	9.19	0.02	0.01	0.01	3.76	6.73	0.02	0.01	0.15	3.37
Gemmatimonadota	8.49	0.01	0.02	0.01	5.74	3.91	0.01	0.01	0.08	1.77
Bacteroidota	4.30	0.08	0.02	12.42	5.06	9.49	0.02	0.01	3.48	5.47
Nitrospirota	3.74	0.01	0.00	0.00	1.55	3.91	0.01	0.00	0.02	2.42
Verrucomicrobiota	3.40	0.01	0.00	0.00	0.95	2.30	0.00	0.00	0.08	1.10
Firmicutes	2.97	0.77	0.21	0.44	5.61	0.95	0.41	0.52	1.24	2.08

Nine phyla were detected in the soil samples, while the number in the plant compartments ranged from 6 to 9 phyla (Table 1). Proteobacteria, Actinobacteriota, Acidobacteriota, Gemmatimonadota, Bacteroidota, and Firmicutes were present in all the compartments of both the FC and HP plants (Table 2). Three phyla were not detected in some compartments: Mycococcota was not detected in FC-L; Nitrospirota not detected in FC-L, FC-Ren and HP-L; and Verrumicrobiota not detected in FC-L, FC-Ren, HP-Fen and HP-L.

Proteobacteria (or Pseudomonadota) was the predominant phylum present in all compartments of both the HP and FC plants (Table 2; Fig. 2), except HP-Ren with a relative abundance of 44.72% Proteobacteria and 49.38% Actinobacteriota. Actinobacteriota (or Actinomycetota) was the second most abundant phylum – 19.66% in FC-Ren, 32.39% in FC-RS, and 22.72% in HP-RS.

Plant compartments (leaves, fruits and roots) had a higher relative abundance of Proteobacteria compared to the soil samples, except for the root of HP sample (HP-Ren – 44.72%) which had a lower abundance than the bulk soil (HP-BS – 56.69%) and rhizospheric soil (HP-RS – 55.73%). In the 5 compartments of the home-planted chilli plant, HP-Fen had the highest relative abundance (95.32%) of Proteobacteria in the total community. The sample FC-L had the highest relative abundance (93.86%) of Proteobacteria compared to the other 4 samples collected from FC chilli plants.

The root compartments – FC-Ren (19.66%) and HP-Ren (49.38%) – had the highest relative abundance of Actinobacteriota among the plant compartments of both the FC and HP plants. Likewise, the rhizospheric soil had a higher relative abundance of Actinobacteriota than the bulk soil of both the FC and HP plants.

With few exceptions, the soil compartments (bulk and rhizospheric soils) exhibited a higher relative abundance of the phyla Myxococcota, Acidobacteriota, Gemmatimonadota, Bacteroidota, Nitrospirota, Verrucomicrobiota, and Firmicutes compared to the plant compartments. FC-Ren (12.42%) had a higher relative abundance of Bacteroidota than the soil compartments of FC plants; and HP-Ren (1.24%) had a higher relative abundance of Firmicutes than HP-BS (0.95%).

The supplementary table S1 shows bacterial OTUs at different taxonomic levels: Phylum, Class, Order, Family, and Genus. At the genus level, Proteobacteria was represented by 30 genera, Actinobacteriota by 9 genera, Acidobacteriota by 5 genera, Myxococcota by 3 genera, Bac-

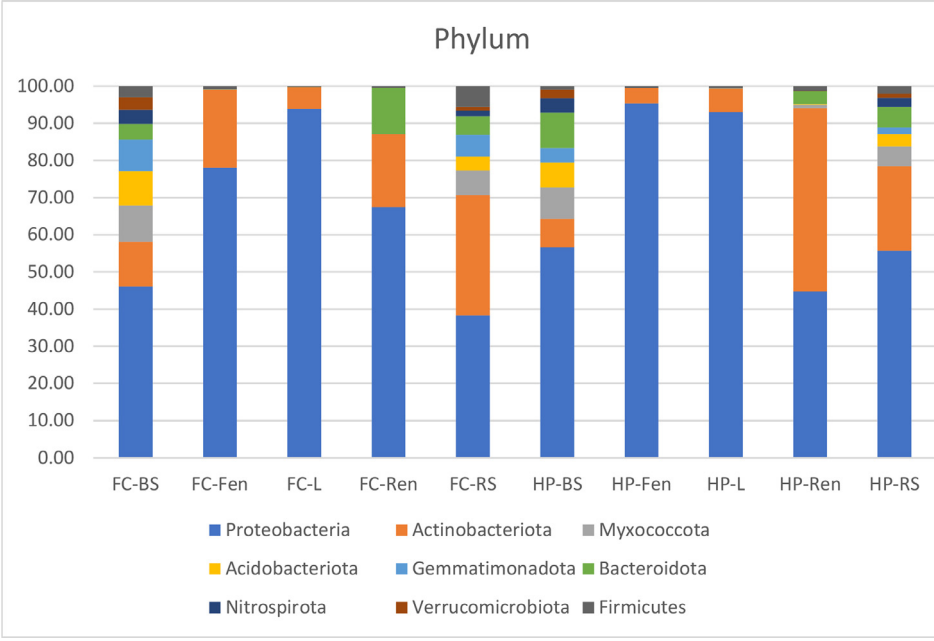

Fig. 2. Relative abundance (%) of bacterial phyla in 5 different compartments of home-planted and farm-cultivated chilli plants. FC: farm-cultivated; HP: home-planted; Fen: fruit endophyte; L: leaf; Ren: root endophyte; BS: bulk soil; RS: rhizospheric soil.

teroidota by 2 genera, and Nitrospirota, Firmicutes and Gemmatimonadota by 1 genus each (Supplementary Table S1). Twenty-nine genera were present in at least one compartment in both the FC and HP samples. Fifteen genera (in 15 families) were not present in FC plants and associated soils, and 9 genera (in 9 families) were not present in HP plants and associated soils. In both the FC and HP plants, 30 genera were not present in all the plant compartments and 22 genera were not present in all the soil compartments.

Halomonas (Proteobacteria) was the predominant genus in the leaf and fruit compartments of both the FC and HP plants – FC-Fen, 47.69%; FC-L, 77.20%; HP-Fen, 55.89%; HP-L, 73.45% (Table 3). It was also the predominant genus in HP-RS (11.15%); it was not present in the other soil compartments. The predominant genera in the other compartments were: MND1 (11.09%) in FC-BS; *Roseateles* (49.93%) in FC-Ren and not present in the other compartments of both the FC and HP plants; *Nocardioides* (12.90%) in FC-RS and only in rhizospheric soils; *Nitrospira* (6.97%) in HP-BS and only in soil compartments; and *Streptomyces* (39.21%) in HP-Ren and only in root endophyte. The bulk and rhizospheric soils had high relative abundance of unidentified genera in both the FC and HP plants – FC-BS (28.41%); FC-RS (36.36%); HP-BS (28.96%); HP-RS (35.03%).

3.3. Alpha diversity

In general, the alpha diversity of the bacterial community (richness and evenness) was not significantly different between the two cultivation (home-planted and farm-cultivated) conditions (Kruskal-Wallis p value of 0.92) (Fig. 3).

The soil compartments (bulk and rhizospheric soils) of the FC and HP chilli plants exhibited significantly higher alpha diversity of the bacterial community (observed OTUs, Shannon diversity, Faith's PD, and Pielou's evenness) compared to the plant compartments (leaf, fruit and root) (Kruskal-Wallis p-value of 0.01) (Fig. 4).

Table 3
Relative abundance of bacterial genera in field-cultivated (FC) and home-planted (HP) chilli plants. BS, bulk soil; Fen, fruit endophyte; L, leaf; Ren, root endophyte; RS, rhizospheric soil.

Genus	FC-BS	FC-Fen	FC-L	FC-Ren	FC-RS	HP-BS	HP-Fen	HP-L	HP-Ren	HP-RS
Halomonas	0.00	47.69	77.20	0.00	0.00	0.00	55.89	73.45	5.51	11.15
Gaiella	1.99	0.00	0.00	0.00	2.49	0.00	0.00	0.00	0.00	6.11
Pedomicrobium	3.21	0.00	0.00	0.00	3.25	2.67	0.00	0.00	0.00	5.12
Nitrospira	5.86	0.00	0.00	0.00	2.87	6.97	0.00	0.00	0.00	4.81
PLTA13	6.36	0.00	0.00	0.00	3.14	3.64	0.00	0.00	0.00	3.79
MND1	11.09	0.00	0.00	0.00	5.35	6.66	0.00	0.00	0.00	3.72
67–14	2.19	0.00	0.00	0.00	4.15	0.00	0.00	0.00	0.00	3.57
SC-I-84	3.10	0.00	0.00	0.00	0.00	2.73	0.00	0.00	0.00	3.05
B1–7BS	0.00	0.00	0.00	0.00	0.00	2.67	0.00	0.00	0.00	2.92
mle1–7	0.00	0.00	0.00	0.00	0.00	4.39	0.00	0.00	0.00	2.83
SWB02	0.00	0.00	0.00	0.00	0.00	2.01	0.00	0.00	0.00	2.78
Blrii41	0.00	0.00	0.00	0.00	0.00	3.93	0.00	0.00	0.00	2.73
Sphingomonas	0.00	3.05	1.22	0.00	3.29	0.00	0.00	0.00	0.00	2.70
Subgroup 10	3.26	0.00	0.00	0.00	0.00	2.14	0.00	0.00	0.00	2.53
Nocardioides	0.00	0.00	0.00	0.00	12.90	0.00	0.00	0.00	0.00	2.49
IMCC26256	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.47
MB-A2–108	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.20
CCD24	3.08	0.00	0.00	0.00	0.00	5.07	0.00	0.00	0.00	0.00
Haliangium	3.82	0.00	0.00	0.00	0.00	2.83	0.00	0.00	0.00	0.00
Pelagibacterium	0.00	9.87	9.37	0.00	0.00	0.00	17.07	14.72	1.32	0.00
Acidibacter	0.00	0.00	0.00	0.00	2.77	2.89	0.00	0.00	0.00	0.00
bacteriap25	6.47	0.00	0.00	0.00	3.23	2.96	0.00	0.00	0.00	0.00
TRA3–20	6.71	0.00	0.00	0.00	3.06	4.77	0.00	0.00	0.00	0.00
Bacillus	3.17	0.00	0.00	0.00	4.88	0.00	0.00	0.00	0.00	0.00
Bryobacter	0.00	0.00	0.00	0.00	2.39	2.69	0.00	0.00	0.00	0.00
Pseudomonas	0.00	0.00	0.00	0.00	0.00	2.09	0.00	0.00	0.00	0.00
Pedosphaeraceae	2.31	0.00	0.00	0.00	0.00	2.33	0.00	0.00	0.00	0.00
Microbacterium	0.00	20.44	3.45	0.00	0.00	0.00	1.67	3.56	0.00	0.00
Subgroup 5	3.12	0.00	0.00	0.00	0.00	2.45	0.00	0.00	0.00	0.00
Streptomyces	0.00	0.00	0.00	20.02	4.14	0.00	0.00	0.00	39.21	0.00
Mycobacterium	0.00	0.00	0.00	0.00	3.07	0.00	0.00	0.00	0.00	0.00
Subgroup 22	5.85	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Bradyrhizobium	0.00	2.79	0.00	0.00	0.00	0.00	1.49	0.00	0.00	0.00
Ensifer	0.00	0.00	0.00	2.20	0.00	0.00	0.00	0.00	0.00	0.00
S0134 terrestrial group	0.00	0.00	0.00	0.00	2.67	0.00	0.00	0.00	0.00	0.00
Steroidobacter	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.14	0.00
Ohtaekwangia	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.27	0.00
Sphingobium	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.05	0.00
Nesterenkonia	0.00	0.00	2.28	0.00	0.00	0.00	1.86	2.64	0.00	0.00
Aureimonas	0.00	0.00	1.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Allorhizobium-Neorhizobium-Pararhizobium-Rhizobium	0.00	0.00	0.00	1.51	0.00	0.00	0.00	0.00	7.57	0.00
Methylophilus	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.62	0.00
Acidovorax	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.35	0.00
*Burkholderia-Caballeronia-*Paraburkholderia	0.00	10.35	1.78	0.00	0.00	0.00	1.69	2.36	1.98	0.00
Afipia	0.00	0.00	0.00	10.99	0.00	0.00	0.00	0.00	1.85	0.00
Bosea	0.00	0.00	0.00	1.60	0.00	0.00	0.00	0.00	1.80	0.00
Pseudoxanthomonas	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.55	0.00
Shinella	0.00	0.00	0.00	1.19	0.00	0.00	0.00	0.00	0.00	0.00
Roseateles	0.00	0.00	0.00	49.93	0.00	0.00	0.00	0.00	0.00	0.00
Methylobacterium-Methylorubrum	0.00	5.83	2.39	0.00	0.00	0.00	18.76	1.77	0.00	0.00
Thauera	0.00	0.00	0.00	0.00	0.00	2.79	0.00	0.00	0.00	0.00
Plasticicumulans	0.00	0.00	0.00	0.00	0.00	2.34	0.00	0.00	0.00	0.00
Chitinophaga	0.00	0.00	0.00	12.56	0.00	0.00	0.00	0.00	0.00	0.00
Other	28.41	0.00	1.06	0.00	36.36	28.96	1.56	1.49	25.78	35.03

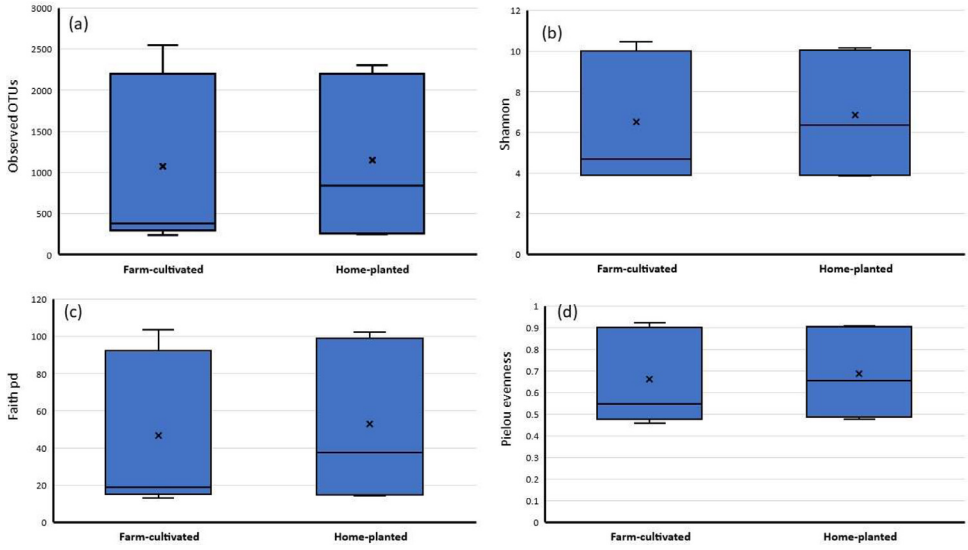

Fig. 3. Alpha diversity tests between home-planted and farm-cultivated chilli plants. (a) Observed OTUs (HP: 1150 ± 991 (244~2302); FC: 1073 ± 1061 (234~2551)); (b) Shannon diversity (HP: 6.83 ± 3.10 (3.83~10.13); FC: 6.48 ± 3.24 (3.85~10.42)); (c) Faith's phylogenetic diversity (HP: 52.64 ± 43.08 (14.09~102.22); 46.60 ± 42.38 (12.99~103.16)); (d) Pielou's evenness (HP: 0.69 ± 0.21 (0.48~0.91); 0.66 ± 0.22 (0.46~0.92)).

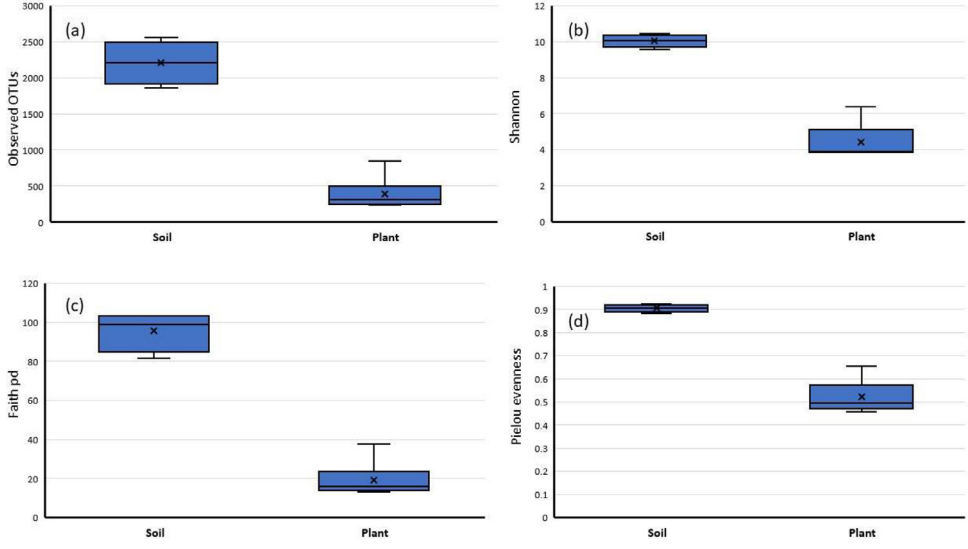

Fig. 4. Alpha diversity tests between plant and soil compartments of FC and HP chilli plants. (a) Observed OTUs (Plant: 462 ± 230 (234~838); Soil: 2201 ± 297 (1851~2551)); (b) Shannon diversity (Plant: 5.30 ± 1.00 (3.83~6.35); Soil: 10.02 ± 0.36 (9.57~10.42)); (c) Faith's phylogenetic diversity (Plant: 22.91 ± 9.12 (12.99~37.25); Soil: 95.40 ± 10.14 (81.21~103.16)); (d) Pielou's evenness (Plant: 0.62 ± 0.07 (0.46~0.65); Soil: 0.90 ± 0.02 (0.88~0.92)).

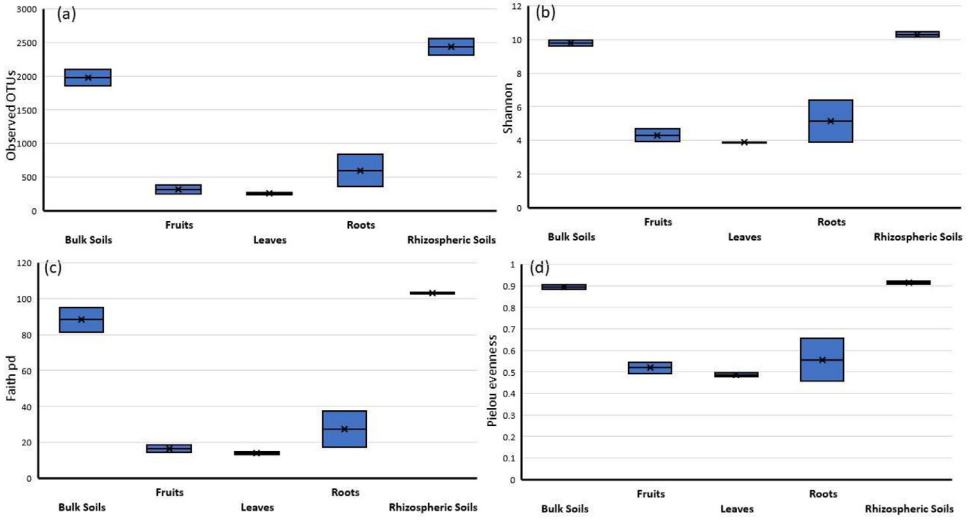

Fig. 5. Alpha diversity tests between different compartments of FC and HP chilli plants. (a) Observed OTUs (Fruits: 311 ± 95 (244∼378); Leaves: 250 ± 22 (234∼265); Roots: 594 ± 246 (349∼838); Rhizospheric soils: 2427 ± 176 (2302∼2551); Bulk soils: 1976 ± 176 (1851∼2100)); (b) Shannon diversity (Fruits: 4.28 ± 0.54 (3.90∼4.66); Leaves: 3.86 ± 0.04 (3.83∼3.89); Roots: 5.10 ± 1.77 (3.85∼6.35); Rhizospheric soils: 10.28 ± 0.21 (10.13∼10.42); Bulk soils: 9.76 ± 0.27 (9.57∼9.95)); (c) Faith's phylogenetic diversity (Fruits: 16.31 ± 3.14 (14.09∼18.53); Leaves: 13.80 ± 1.14 (12.99∼14.61); Roots: 27.17 ± 14.25 (17.10∼37.25); Rhizospheric soils: 102.69 ± 0.66 (102.22∼103.16); Bulk soils: 88.11 ± 9.76 (81.21∼95.02)); (d) Pielou's evenness (Fruits: 0.52 ± 0.04 (0.49∼0.54); Leaves: 0.48 ± 0.01 (0.48∼0.49); Roots: 0.56 ± 0.14 (0.46∼0.65); Rhizospheric soils: 0.91 ± 0.01 (0.91∼0.92); Bulk soils: 0.89 ± 0.01 (0.88∼0.90)).

The bacterial community of the bulk and rhizospheric soils had a significantly higher alpha diversity than that of the plant compartments (leaves, fruits and roots) (Fig. 5). However, the richness and evenness of the bacterial community between leaves, fruits and roots were not statistically significant.

3.4. Beta diversity

Principal Coordinate Analysis (PCoA) showed that the bacterial community from the same compartment of FC and HP chilli plants tend to group closely to each other (Fig. 6). There was a significant difference in the bacterial community among the different compartments with $p = 0.007$ for unweighted UniFrac distance and $p = 0.001$ for weighted UniFrac distance. However, there was no significant difference in the bacterial community between chilli plants cultivated in different conditions ($p = 0.761$ for unweighted UniFrac distance; $p = 0.852$ for weighted UniFrac distance).

4. Experimental Design, Materials and Methods

4.1. Sample collection and DNA extraction

Farm-cultivated chillies were obtained from Hoho Farm, Batang Kali, Selangor (3° 26′ 48.5″ N 101° 38′ 52.2″ E), while home-planted chillies were harvested from a 2-year-old chilli plant grown in a round gardening plastic bag measuring 5 cm in width and filled with organic soil. The home planting practices were conducted in an open field, utilizing fertilizers once every 10

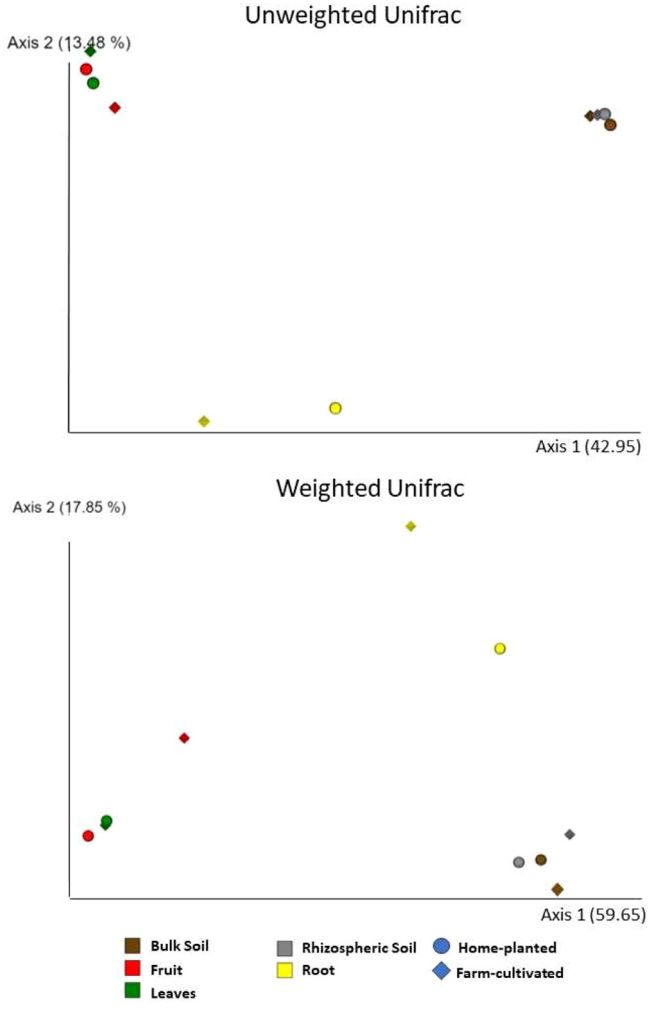

Fig. 6. Beta diversity analysis, based on unweighted UniFrac distance and weighted UniFrac distance, of the bacterial community associated with different compartments of farm-cultivated and home-planted chilli plants.

days and watering once a day in Ipoh, Perak (4° 38′ 26.9″ N 101° 07′ 45.2″ E). Pesticides were only utilized when the plant displayed signs of illness.

Samples of leaves, mature fruits and roots were directly collected from the chilli plants. Bulk soil samples were excavated from a 5 cm depth at the periphery of home-planted chilli, and from the area outside the bucket of farm-cultivated plants, not in direct contact with the roots of chilli plant. All collected samples were stored at −20 °C prior to subsequent DNA extraction.

Rhizospheric soil was collected by spinning the root samples submerged in autoclaved distilled water using vortex mixer at maximum speed for 3–5 min followed by centrifugation at 2800 g for 30 min. The supernatant was discarded, leaving behind the pellet identified as the rhizospheric soil [15]. Subsequently, the processed root samples, along with other plant-based samples, underwent surface sterilization [14].

Plant-based samples were surface sterilized by washing with 70% ethanol and 0.1% of sodium hypochlorite solutions for 1 min each. Next, the samples underwent a 30-second wash with

70% ethanol, followed by a final wash with sterile water. The samples were then crushed and stored at −20 °C before DNA extraction. The flesh and seeds of the chili fruits were separated prior to crushing [14] and only the flesh of the chili samples was used for subsequent processes.

A total of five sample groups were collected from each chili plant: bulk soil (BS), rhizopheric soil (RS), root endophytes (Ren), leaves (L), and fruit endophytes (Fen). Total DNA of plant samples was extracted using i-genomic Plant DNA Extraction Kit (iNtRON Biotechnology, Inc, Korea), according to the manufacturer's protocol with minor modifications of sonicating for at least 1 hour to enhance cell lysis. For soil samples, Mobio Powersoil DNA Isolation Kit (Qiagen, Germany) was used according to the manufacturer's instructions. The quality and quantity of all extracted DNA samples were determined using Nanodrop 2000 spectrophotometer and Qubit® 2.0 fluorometer (Thermo Fisher Scientific™, USA). Samples that had the best quality and quantity were selected from each sample groups for targeted metagenomics sequencing.

4.2. Targeted metagenomics sequencing

The V5-V6 hypervariable region of 16S rRNA chloroplast-excluding bacteria primer was used for library preparation with the following primer set: 799F (Forward), 5′- AAC MGG ATT AGA TAC CCK G −3′; 1115R (Reverse), 5′- AGG GTT GCG CTC GTT G −3′ (Chelius & Triplett, 2001). The reaction solution contained 4 µL of 5 × FastPfu buffer, 2 µL of 2.5 mM dNTPs, 0.8 µL of forward primer, 0.8 µL of reverse primer, 0.4 µL TransStart Fastpfu DNA polymerase, 10 ng of DNA template, and topping up with ddH2O to 20 µL of reaction solution. PCR amplification was performed in ABI GeneAMP® 9700 using the cycling parameters of 95 °C for 5 min, followed by 27 cycles of 95 °C for 30 s, 55 °C for 30 s, and 72 °C for 45 s, and a final extension stage of 72 °C for 10 min. Gel electrophoresis was performed using 2% agarose gel to detect the PCR products obtained from 799F/1115R primer set.

4.3. Bioinformatics and statistical analysis

The demultiplexed paired-end reads were merged and trimmed for low quality reads. The data were then dereplicated and clustered into Operational Taxonomic Units (OTUs) using VSEARCH closed reference clustering method available in QIIME 2 version 2022.2 [16]. The data was clustered under 97% similarity threshold based on the SILVA reference sequences [17] provided in the QIIME 2 data resources page. Taxonomic assignment was performed using pre-trained SILVA-138 99% OTUs full-length sequences classifier [17,18] available in QIIME2 data resources page. Concurrently, reads that were identified as mitochondria or chloroplast were excluded from the results. Alpha rarefaction curve was plotted based on observed OTUs vs sequencing depth to determine a suitable sampling depth that would include as much samples as possible.

The bacterial sequences were rarefied to 7000 reads for both alpha and beta diversity index estimates. Observed OTUs, Shannon diversity, Faith's phylogenetic diversity (PD), and Pielou's evenness index were calculated using QIIME 2 Alpha Diversity plugin. The results were presented in box-plot for comparison of species richness and evenness between different sample types, cultivation areas and compartments. Significance differences were calculated at significance level of $p < 0.05$. Additionally, beta diversity was evaluated by calculating weighted UniFrac and unweighted UniFrac distances. The outcomes were visualized through Principal coordinate analysis (PCoA) plots to determine the degree of similarity in bacterial communities across different compartments as well as different chilli plants. The significance differences were calculated using permutational multivariate analysis of variance (PERMANOVA) method at significance level of $p < 0.05$.

Limitations

The data in this article have certain limitations, primarily in the data collection process. One notable constraint is the sample size, which could limit the precision of data regarding bacterial communities and diversity. Additionally, the absence of specific soil properties from both home-planted and farm-cultivated samples in the dataset might impede a comprehensive understanding of their relationship with bacterial communities.

Ethics Statement

The authors have read and follow the ethical requirements for publication in Data in Brief and confirming that the current work does not involve human subjects, animal experiments, or any data collected from social media platforms.

Data Availability

Microbial diversity on chilli plants (Original data) (GenBank Sequence Read Archive).

CRediT Author Statement

Sze-Looi Song: Conceptualization, Methodology, Writing – original draft; **Zheng-Xian Lew:** Data curation, Software, Writing – original draft; **Hoi-Sen Yong:** Writing – original draft; **Qi-Hang Lim:** Writing – review & editing; **Rong-Heng Joshua Chai:** Writing – review & editing.

Acknowledgements

We thank our institutions for providing various research facilities and other support. This work was funded by the Universiti Malaya (H-5620009 to HSY).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary Materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.dib.2024.110273.

References

[1] C. Carrizo García, M.H. Barfuss, E.M. Sehr, G.E. Barboza, R. Samuel, E.A. Moscone, F. Ehrendorfer, Phylogenetic relationships, diversification and expansion of chili peppers (*Capsicum*, Solanaceae), Ann. Bot. 118 (2016) 35–51, doi:10.1093/aob/mcw079.

[2] F. Liu, J. Zhao, H. Sun, C. Xiong, X. Sun, X. Wang, … X. Zou, Genomes of cultivated and wild *Capsicum* species provide insights into pepper domestication and population differentiation, Nat. Commun. 14 (2023) 5487, doi:10.1038/s41467-023-41251-4.

[3] P. Tripodi, S. Kumar, The *Capsicum* crop: an introduction, in: The *Capsicum* Genome, 2019, pp. 1–8, doi:10.1007/978-3-319-97217-6_1.
[4] National Center for Biotechnology Information (NCBI)[Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988]– [cited 2024 January 14]. Available from: https://www.ncbi.nlm.nih.gov/Traces/study/?acc=PRJNA1064378&o=acc_s/3Aa.
[5] B.Z. Hou, C.L. Li, Y.Y. Han, Y.Y. Shen, Characterization of the hot pepper (*Capsicum frutescens*) fruit ripening regulated by ethylene and ABA, BMC Plant Biol. 18 (2018) 1–12, doi:10.1186/s12870-018-1377-3.
[6] K.R. Karim, M.Y. Rafii, A.B. Misran, M.F.B. Ismail, A.R. Harun, M.M.H. Khan, M.F.N. Chowdhury, Current and prospective strategies in the varietal improvement of Chilli (*Capsicum annuum* L.) specially heterosis breeding, Agronomy 11 (2021) 2217, doi:10.3390/agronomy11112217.
[7] D.N. Lozada, P.W. Bosland, D.W. Barchenger, M. Haghshenas-Jaryani, S. Sanogo, S. Walker, Chile pepper (*Capsicum*) breeding and improvement in the "multi-omics" era, Front. Plant Sci. 13 (2002) 879182, doi:10.3389/fpls.2022.879182.
[8] V. Jaiswal, V. Gahlaut, N. Kumar, N. Ramchiary, Genetics, genomics and breeding of chili pepper *Capsicum frutescens* L. and other *Capsicum* species, in: Advances in Plant Breeding Strategies: Vegetable Crops: Volume 9: Fruits and Young Shoots, 2021, pp. 59–86, doi:10.1007/978-3-030-66961-4_2.
[9] D. Shim, S. Raveendar, J.R. Lee, G.A. Lee, N.Y. Ro, Y.A. Jeon, … J.W. Chung, The complete chloroplast genome of *Capsicum frutescens* (Solanaceae), Appl. Plant Sci. 4 (2016) 1600002, doi:10.3732/apps.1600002.
[10] K. Shiragaki, S. Yokoi, T. Tezuka, Phylogenetic analysis and molecular diversity of *Capsicum* based on rDNA-ITS region, Horticulturae 6 (2020) 87, doi:10.3390/horticulturae6040087.
[11] S.K. Mandal, S.K. Rath, R. Logesh, S.K. Mishra, H.P. Devkota, N. Das, *Capsicum annuum* L. and its bioactive constituents: a critical review of a traditional culinary spice in terms of its modern pharmacological potentials with toxicological issues, Phytother. Res. 37 (2023) 965–1002, doi:10.1002/ptr.7660.
[12] H. Liao, Y. Luo, X. Huang, X. Xia, Dynamics of quality attributes, flavor compounds, and microbial communities during multi-driven-levels chili fermentation: interactions between the metabolome and microbiome, Food Chem. 405 (2023) 134936, doi:10.1016/j.foodchem.2022.134936.
[13] Z. Ye, Z. Shang, S. Zhang, M. Li, X. Zhang, H. Ren, … J. Yi, Dynamic analysis of flavor properties and microbial communities in Chinese pickled chili pepper (*Capsicum frutescens* L.): a typical industrial-scale natural fermentation process, Food Res. Int. 153 (2022) 110952, doi:10.1016/j.foodres.2022.110952.
[14] M. Gao, C. Xiong, C. Gao, C.K. Tsui, M.M. Wang, X. Zhou, … L. Cai, Disease-induced changes in plant microbiome assembly and functional adaptation, Microbiome 9 (2021) 1–18, doi:10.1186/s40168-021-01138-2.
[15] P.R. Hirsch, T.H. Mauchline, Who's who in the plant root microbiome? Nat. Biotechnol. 30 (2012) 961–962, doi:10.1038/nbt.2387.
[16] J.G. Caporaso, J. Kuczynski, J. Stombaugh, K. Bittinger, F.D. Bushman, E.K. Costello, … R. Knight, QIIME allows analysis of high-throughput community sequencing data, Nat. Methods 7 (2010) 335–336, doi:10.1038/nmeth.f.303.
[17] M.S. Robeson, D.R. O'Rourke, B.D. Kaehler, M. Ziemski, M.R. Dillon, J.T. Foster, N.A. Bokulich, RESCRIPt: reproducible sequence taxonomy reference database management, PLoS Comput. Biol. 17 (2021) e1009581, doi:10.1101/2020.10.05.326504.
[18] N.A. Bokulich, B.D. Kaehler, J.R. Rideout, M. Dillon, E. Bolyen, R. Knight, … J.G. Caporaso, Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin, Microbiome 6 (2018) 1–17, doi:10.1186/s40168-018-0470-z.