COMPARATIVE METAGENOMIC PROFILING OF SOIL MICROBIOMES IN LOW AND HIGH YIELDING COCOA TREES ASSOCIATED WITH FARM MANAGEMENT PRACTICES

Rahman M. Z. A[1]*, Norasekin T.[1], Ishak Z.[2] , Lea J.[1], Roslina M.S.[1], and Fisal A.[2]
[1] Malaysian Cocoa Board, Centre for Cocoa Biotechnology Research, Commercial Zone 1, Norowot Road, South KKIP, 88460, Kota Kinabalu, Sabah, Malaysia.
[2] Malaysian Cocoa Board, Cocoa Innovative and Technology Centre, Lot 12621 Nilai Industrial Area,71800 Nilai, Negeri Sembilan, Malaysia..
**Corresponding author email: hilmi@koko.gov.my.*

Malaysian Cocoa J. 15 (2): 7-12 (2023)

ABSTRACT – *Soil microbial communities or microbiomes, comprise a large portion of life's genetic diversity, acting as important regulators of cocoa trees productivity and growth. In this study, comparative metagenomics was performed in order to determine the relationship between the microbial diversity associated with yield and farm management practices. The findings indicate that farm management practices have a significant effect in shaping the structure of microbial communities. The properly managed cocoa farms have higher microbial diversity compared to abandoned farms. Moreover, in this study, we have demonstrated that cocoa trees with higher yield will have higher microbial diversity in contrast with low yield trees. Agricultural management practice is one of the important factors in determining the microbial community structure and proper farm management not only affect the production and yield but will promote diversification of favourable microbial community for enhancing crop productivity and global ecosystem health.*

Key words: Metagenomics, soil microbiome, farm management.

INTRODUCTION

Theobroma cacao L. also known as cacao tree is the most significant crop for countries such as Ghana, Ivory Coast, Indonesia, and Malaysia. This commodity contributes significantly to the economic growth of cocoa-producing country's agricultural sector and it remains an important global commodity due to its fundamental role for the chocolate production and as enabler in the production of many foods (Vega & Kwik-Uribe, 2012), as well as providing raw materials for the cosmetics and pharmaceutical industries.

Worldwide, about 95% of cocoa is grown on smallholder farms ranging from 2-5 hectares (Anga, 2016), and employing about 5–6 million cocoa farmers from Asia, Africa, Oceania, and Latin America (Kozicka *et al*., 2018). However, bean production around the world has been declining, and Malaysia is not an exception. A number of factors, including the drop in cocoa prices, cocoa area abandonment and conversion of the plantation due to low yield have contributed to these effects (Arshad & Ibragimov, 2015).

In the early stages of cocoa plantation, productivity of the beans reached more than one tonne per hectare as the crop enjoyed the soil fertility of first-time cocoa planting in new areas. However, despite the extensive use of chemical fertilizers and pesticides, the fertility of the soil declined over time as a result of the abandonment of plantation. Soil microbial communities and plant-microbe interactions (phyllosphere, endosphere and rhizosphere) have important roles in biogeochemical cycling, and maintenance of plant health and productivity, yet remain poorly understood (Turner *et al*., 2013). The composition and diversity of these soil microbial communities are important indicators of the soil fertility (Wang *et al*., 2020). In the past, culture method was used to study the microbes from their own environments. The main constraint of the traditional studies is the method only focused on culturable bacteria which is estimated only 0.1% to 1%. Other limitations include necessity to grow as co-culture with other isolates, the growth condition with range of temperature tolerance and pH (Stewart, 2012). Uncovering these microbes' adaptations and advantageous traits is crucial for gaining a thorough knowledge of their contribution and function. Advances in molecular biology have led to the development of omics techniques, which recently gained prominence in the diversity and abundance study of microbes (Alawiye and Babalola, 2019; Brader *et al*., 2017a), such as metagenomics, metatranscriptomic and metaproteomic.

Metagenomics is a broad discipline facilitating the taxonomic and physiological information of species collected from their true habitat. It is also

defined as the evaluation and collective set of genomes of a mixed microbial community in a niche (Petrosina *et al.*, 2009). The technique allowed exploration of taxonomic communities and help linking of associated functions without cultivation of microorganisms (Schloss and Handelsman, 2003).

In this study, metagenomic approach was used to investigate the correlation between potential yield associated farm management practices and microbial community to understand their association and interaction with the host. This could help to improve soil fertility by reducing the use of chemical fertilizers and farm management thus increase the cocoa yield.

MATERIALS AND METHODS

Study Area and Sample Collection

The research plots were selected based on farm management conditions that are divided into three cocoa farm conditions which are properly managed, neglected and abandoned farm (Figure 1). The criteria for properly managed farm are defined as the farm that implementing good agricultural practice (GAP) and have active agriculture activity. Neglected farm is defined as the farm that is not managed properly, not implementing GAP and have active agriculture activity. Whereas, abandoned farm is defined as the farm that have been totally abandoned and agriculture activity was not active. The study was conducted in Sabah, Malaysia and specifically located in Tuaran, Kota Marudu, Kudat and Madai districts. Rhizosphere's soil samples were collected from 18 identified locations as listed in Table 1. In each sampling area, three soil sample replicates were collected from each high and low yielding cocoa tree identified for each sampling sites. For this study, high yielding tree was defined a tree that can produce more that 60 cocoa pods per year. For each soil sampling experiments, 5 metadata parameters were recorded (coordinate, altitude, pH, soil texture and yield).

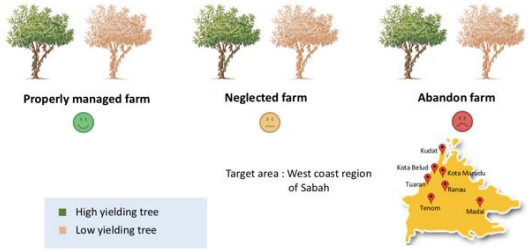

Figure 1: Rhizosphere soil sampling criteria according to farm management condition.

Table 1: Sampling Location and Farm Management Condition

Sampling Location No.	*District*	*Region*	*Farm Condition*
1	Kota Marudu	Goshen	Partially Neglected
2	Kota Marudu	Goshen	Partially Neglected
3	Kota Marudu	Goshen	Neglected
4	Kota Marudu	Goshen	Neglected
5	Ranau	Togis	Properly Managed
6	Ranau	Kibbas	Properly Managed
7	Ranau	Kibbas	Properly Managed
8	Tuaran	Tuaran	Abandoned
9	Kudat	Pinawantai	Partially Neglected
10	Kudat	Indarsun Darat	Neglected
11	Kudat	Batu 6	Properly Managed
12	Kudat	Perapat	Properly Managed
13	Madai	PPPK Madai	Properly Managed
14	Tenom	Baru Jumpa	Properly Managed
15	Tenom	Baru Jumpa	Properly Managed
16	Tenom	Baru Jumpa	Properly Managed
17	Keningau	Bunga Raya	Neglected
18	Kota Kinabalu	PPBK KKIP	Partially Neglected

DNA Extraction

DNA extraction for all soil samples were performed using MoBio PowerSoil kit (Mo-Bio, USA) as per the manufacturer's instructions. The DNA quantitation and quality assessment were performed using NanoDrop 2000c spectophotometer and Qubit dsDNA BR assay. The sample was also run on 1% agarose gel to determine the integrity of genomic DNA.

Library Preparation and Targeted Sequencing

The 16S (V1-V3 region) regions were chosen for the targeted sequencing. The primers used for 16S

regions amplification are listed in Table 2. The following steps of library preparation and sequencing were performed by Malaysian Genomics Resource Centre Berhad, Kuala Lumpur, Malaysia. Paired-end sequencing runs (2 X 300bp) were performed on an Illumina MiSeq platform with v3 kit chemistry (Illumina, San Diego, CA, United States).

Table 2: Primer sequence used for targeted sequencing.

Target name	*Primer Sequence*
16S: 27F - 519R (V1-V3)	Forward primer AGAGTTTGATCMTGGCTCAG Reverse Primer GWATTACCGCGGCKGCTG

Metagenomic Analysis

Paired-ends reads were assembled by aligning the forward and reverse reads using Fastq-join. Low quality data sets (at Phred < Q19) were trimmed and screened to remove chimeras (sequences generated due to the PCR amplification of multiple templates or parent sequences). Sequences were clustered followed by the chimera filtered. It consists of picking Operational Taxonomic Units (OTUs) based on sequence similarity within the reads, and picking a representative sequence from each OTU. It also assigns taxonomic identities using reference databases, aligns the OTU sequences, and constructs an OTU table, representing the abundance of each OTU in each sample. Using QIIME taxonomy which was assigned using Greengenes database5 (Version 13_8, Aug 2013) as reference for 16S data sets.

RESULTS AND DISCUSSION

The microbiome profiling between two cocoa populations, which are the high yielding population and low yielding population based on farm conditions (Figure 1) were conducted by using targeted metagenomics sequencing approach based on identified area as listed in Table 1. For each soil sampling experiments, 5 metadata parameters were recorded (coordinate, altitude, pH, soil texture and yield). The recorded metadata for each sample were listed in Table 3. The samples were collected in various location as different location will have different composition of bacterial communities and the microbial profiling carried out in this study will help us determine the association between location and farm management practice in shaping the communities.

A total of 24 collected soil samples were subjected to DNA extraction using PowerSoil DNA Isolation Kit (Mo-Bio, USA). The extracted DNA samples were subjected to assessment to check for its quality, quantity and integrity. The DNA quality and quantity were measured using NanoDrop and Qubit dsDNA BR assay as summarized in Table 4. The sample was also run on 1% agarose gel to determine the integrity of genomic DNA (Figure 2). Approximately 1 µg of high-quality DNA (as measured by Qubit) is required for library preparation.

Table 3: Metadata collected for soil rhizosphere

Sample Name	*Location*	*Coordinate*	*Altitude (m)*	*pH*	*Soil type / Texture % (sand:silt:clay)*	*Yield (H / L)*
KIP-1	Kota Kinabalu	6.0935N 116.1838E	13.13	7.0	0:0:100	H
KIP-2	Kota Kinabalu	6.0940N 116.1839E	14.42	<3	47:53:0	L
TRN-1	Tuaran	6.15N 116.216667E	24.10	5.0	31:60:9	L
KM-1	Kota Marudu	6.466667N 116.75E	37.18	6.0	15:70:15	H
KM-2	Kota Marudu	6.45N 116.783333E	38.74	6.5	73:22:5	L
KM-3	Kota Marudu	6.4512N 116.7550E	86.37	6.0	41:45:14	L
KM-4	Kota Marudu	6.4512N 116.7555E	108.08	5.0	34:43:23	L
KDT-1	Kudat	6.7372N 116.7288E	39.58	6.0	48:48:4	H
KDT-2	Kudat	6.7372N 116.7285E	30.34	5.5	49:46:5	L
KDT-3	Kudat	6.7802N 116.6522E	31.80	6.0	26:42:32	H
KDT-4	Kudat	6.9056N 116.8038E	22.54	6.0	42:33:25	H
KDT-5	Kudat	6.9005N 116.8032E	31.89	6.0	37:57:6	L
KDT-6	Kudat	6.8295N 116.7664E	20.41	6.0	52:21:27	H
MDI-1	Madai	4.7828N 117.9673E	225	6.5	28:52:20	H
RNU-1	Ranau	5.9290N 116.6222E	670.52	7.0	63:22:15	L
RNU-2	Ranau	5.9284N 116.6219E	667.22	6.5	32:62:6	L
RNU-3	Ranau	5.9289N 116.6247E	637.58	6.5	22:78:0	H
RNU-4	Ranau	5.9286E 116.6240E	657.86	5.0	0:100:0	H
TNM-1	Tenom	4.9268N 115.8873E	310.59	5.0	50:50:0	H
TNM-2	Tenom	4.9257N 115.8841E	348.37	5.0	85:0:15	H
TNM-3	Tenom	4.9274E 115.8880E	295.83	5.0	58:22:20	H
TNM-4	Tenom	4.9285N 115.8895E	293.70	5.0	0:97:3	H
KNG-1	Keningau	5.4710N 116.2178E	419.07	5.5	18:54:28	H
KNG-2	Keningau	5.4733N 116.2118E	426.85	6.5	43:47:10	H

H=high; L=low

Amplicon Sequencing Experiments (16S region) were performed for 24 DNA samples and > 1.5 million reads (>30,000 reads per sample) were generated. A total of 24 DNA samples (total = 48 targets) were collected from cocoa rhizosphere were sequenced for 16S (24 targets). The generated paired-end reads are grouped in pairs (2 X 300bp), which is identified as the forward read and the other is the reverse read, stored in two FASTQ files, one for each member of the pair. So, having an overlap is a way to merge the two reads of the pair into a single one. These paired-ends reads were assembled by aligning the forward and reverse reads using Fastq-join.

Table 4: QC results for DNA samples.

Sample Barcode	Vol (µl)	NanoDrop				Qubit dsDNA BR	
		Conc (ng/ µl)	Total (µg)	A260 /280	A260 /230	Conc (ng/ µl)	Total (µg)
GSA00516	30	51.3	1.54	1.87	1.72	36.6	1.10
GSA00518	30	77.4	2.32	1.82	1.63	63.4	1.90
GSA00523	30	38.5	1.93	1.88	1.67	31.6	1.58
GSA00523	50	79.2	2.38	1.83	1.74	89.4	2.68
GSA00524	30	62.4	3.62	1.87	1.83	52.6	3.05
GSA00525	58	80.9	2.43	1.82	1.59	60.2	1.81
GSA00529	30	46.2	1.39	1.89	1.59	49.0	1.47
GSA00530	30	98.3	2.95	1.82	1.58	83.0	2.49
GSA00531	30	33.9	1.02	1.86	1.64	28.8	0.86
GSA00533	30	41.4	1.24	1.92	1.61	32.2	0.97
GSA00534	30	64.8	1.94	1.83	1.60	58.0	1.74
GSA00535	30	64.9	1.95	1.82	1.57	58.0	1.74
GSA00536	30	46.4	1.39	1.87	1.61	36.0	1.08
GSA00537	30	57.1	1.71	1.89	1.72	47.2	1.42
GSA00538	30	77.9	2.34	1.85	1.81	43.8	1.31
GSA00646	30	30.9	0.93	1.84	1.53	26.8	0.80
GSA00647	30	44.1	1.32	1.88	1.67	39.2	1.18
GSA00648	30	54.1	1.62	1.93	1.72	49.8	1.49
GSA00649	30	39.5	1.19	1.89	1.69	34.6	1.04
GSA00650	30	41.0	1.23	1.90	1.80	35.2	1.06
GSA00651	30	64.9	1.95	1.89	1.74	61.6	1.85
GSA00652	30	35.6	1.07	1.92	1.80	33.2	1.00
GSA00653	30	38.4	1.15	1.94	1.84	35.6	1.07
GSA00654	30	59.2	1.78	1.96	2.12	67.8	2.03

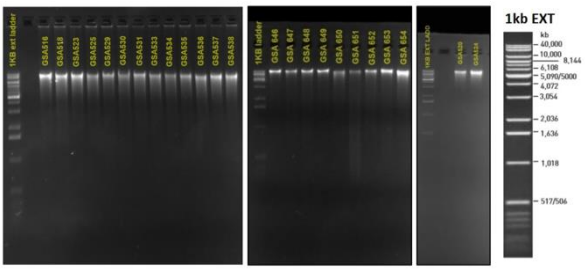

Figure 2: Gel images of DNA samples as run on 1% agarose gel.

Taxonomic analysis at the phylum level detected differences in relative abundance among all three 16S rRNA regions (V1-V3) as shown in Figure 3. There were significant differences in the soil microbial community structure between properly managed farm and abandoned farm soil (Figure 3). Specifically, Proteobacteria and Verrucomicrobia were significantly higher in properly managed farm soil than abandoned farm soil. In contrast, Chloroflexi and Nitrospirae were significantly higher in abandoned farm soils than in properly managed soils. When comparing, the soil microbial community structure between high yielding and low yielding cocoa tree, Bacteroidetes and Planctomycetes were significant higher in high yielding rhizosphere samples compared to low yielding trees. In contrast, Chloroflexi and Saccharibacteria (TM7) phyla were significant higher in low yielding tree rhizosphere as opposed to high yielding tree rhizosphere.

Alpha diversity metrics summarize the structure of an ecological community with respect to its richness (number of taxonomic groups), evenness (distribution of abundances of the groups), or both. Because many perturbations to a community affect the alpha diversity of a community, summarizing and comparing community structure via alpha diversity is a ubiquitous approach to analyzing community surveys (Willis, 2019). As shown in rarefaction plots of observed species for yield (Figure 4), the species richness was slightly higher in high yielding tree rhizosphere compared to low yielding tree rhizosphere suggesting species richness are indirectly affecting crop yield potential. Based on rarefaction plots of observed species for location in Figure 5, there are significant differences in species richness in farming management condition and location, suggesting the farming practice do impart perturbation in microbial diversity and species richness. Therefore, based on our comparative metagenomics study in investigating the effect on low and high yielding cocoa trees associated with farm management practices, we found out different farm management practices directly affect the structure of the microbial community and function of the rhizosphere. Different soil microbial flora could affect the developmental degree of Theobroma cacao root and the distribution of microbiota around the root, thus suggesting rhizosphere microbiota and plant roots are interdependent and interact with each other.

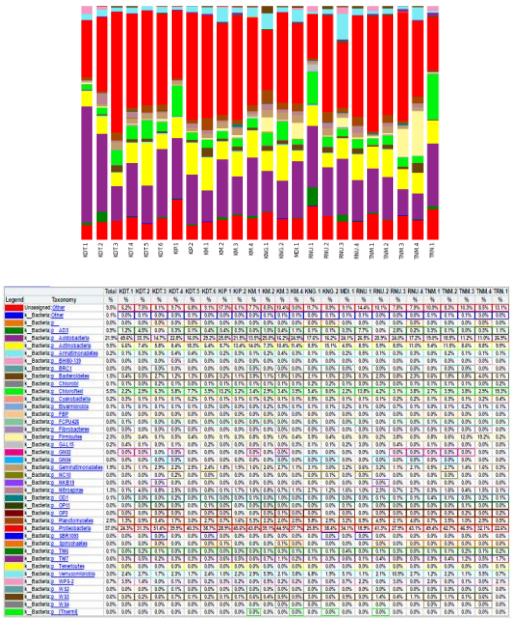

Figure 3: A bar chart showing the relative abundance at phylum level for 16S data sets.

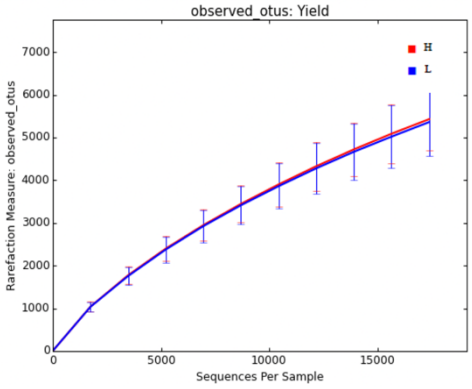

Figure 4: Rarefaction plots of observed species for yield. The vertical axis displays the diversity of the community, while the horizontal axis displays the number of sequences considered in the diversity calculation. Each line on the figure represents the average of all microbial belonging to a group within a category (Yield): here the red line represents all high yielding trees, and blue line represents low yielding trees for 16S data sets.

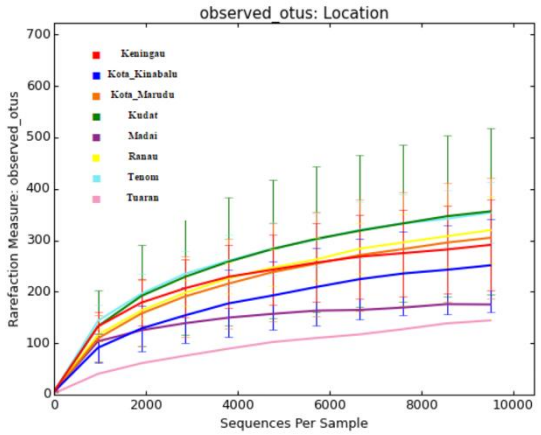

Figure 5: Rarefaction plots of observed species for location. The vertical axis displays the diversity of the community, while the horizontal axis displays the number of sequences considered in the diversity calculation. Each line on the figure represents the average of all microbial belonging to a group within a category (Location): here the red line represents Keningau, blue line represents Kota Kinabalu, orange line represents Kota Marudu, green line represents Kudat, purple line represents Madai, yellow line represents Ranau, cyan represents Tenom, and pink line represents Tuaran for 16S data sets.

CONCLUSION

In this study, we have performed comparative metagenomics in order to determine the relationship between the microbial diversity associated with yield and farm management practices. Farm management practice have significant effect in shaping the structure of microbial communities as based on our data, properly managed cocoa farms have higher microbial diversity compared to abandoned farm. Moreover, in this study we have demonstrated that cocoa trees with higher yield will have higher microbial diversity in contrast with low yield trees. Overall, our results highlighted that agricultural management practice is one of the important factors in determining the microbial community structure and proper farm management not only effect the production and yield but will promote diversification of favourable microbial community for enhancing crop productivity and global ecosystem health.

ACKNOWLEDGEMENT

This work is supported by Developmental Fund (P20001001116013-Collection and Sequencing of Cocoa Genome) from Ministry of Finance under Eleventh Malaysia Plan. Special thanks to the Director General of Malaysian Cocoa Board, Deputy Director General Research and Development and Director of Biotechnology Division for their continuous support and kind approval to publish this research.

REFERENCES

Alawiye, T. T., & Babalola, O. O. (2019). Bacterial diversity and community structure in typical plant rhizosphere. *Diversity*, **11(10)**, 179.

Anga, J. M. (2016). Injecting Innovation to Sustain the Future Supply of Cocoa. *Presentation at the World Cocoa Conference* 2016.

Arshad, F.M., & Ibragimov, A. (2015). Malaysia's cocoa beans decline: a prognosis. International Journal of Agriculture, Forestry and Plantation, Vol. 1 (Sept.)

Brader, G., Compant, S., Vescio, K., Mitter, B., Trognitz, F., Ma, L. J., & Sessitsch, A. (2017). Ecology and genomic insights into plant-pathogenic and plant-nonpathogenic endophytes. *Annual Review of Phytopathology*, **55(1).**

Kozicka, M., Tacconi, F., Horna, D. & Gotor, E. (2018). Forecasting Cocoa Yields for 2050. Bioversity International, Rome, Italy.

Petrosino, J. F., Highlander, S., Luna, R. A., Gibbs, R. A., & Versalovic, J. (2009). Metagenomic pyrosequencing and microbial identification. *Clinical chemistry*, **55(5)**, 856-866.

Schloss, P. D., & Handelsman, J. (2003). Biotechnological prospects from metagenomics. *Current opinion in biotechnology*, **14(3),** 303-310.

Stewart E. J. (2012). Growing Unculturable Bacteria. *Journal of bacteriology*, **194(16),** 4151–4160.

Turner, T. R., Ramakrishnan, K., Walshaw, J., Heavens, D., Alston, M., Swarbreck, D., Osbourn, A., Grant, A., & Poole, P. S. (2013). Comparative metatranscriptomics reveals kingdom level changes in the rhizosphere microbiome of plants. *The ISME journal*, **7(12),** 2248–2258.

Vega, C. & Kwik-Uribe, C. (2012). Theobroma cacao—An Introduction to the Plant, Its Composition, Uses, and Health Benefits.

Wang, C. H., Wu, L., Wang, Z., Alabady, M.S., Parson, D., Molumo, Z. & Fankhauser, S.C. (2020). Characterizing changes in soil microbiome abundance and diversity due to different cover crop techniques. *PLoS ONE* **15(5):** e0232453.

Willis, A. D. (2019). Rarefaction, Alpha Diversity, and Statistics. *Frontier in Microbiology*. (10)2407.