TROPICAL AGRICULTURAL SCIENCE

Diversity, Composition, Taxa Biomarkers, and Functional Genes of Fish Gut Microbes in Peat Swamp Forests and its Converted Areas in North Selangor, Malaysia

Hamidu Saadu[1,2], Jumria Sutra[1], Amalia Mohd Hashim[3], Ahmad Ismail[1], Syaizwan Zahmir Zulkifli[1] and Mohammad Noor Azmai Amal[1,4]*

[1]Department of Biology, Faculty of Science, Universiti Putra Malaysia,
43400 UPM Serdang, Selangor, Malaysia
[2]Department of Biology, School of Science, Federal College of Education Katsina, PMB 2041,
Katsina, Katsina State, Nigeria
[3]Department of Microbiology, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia,
43400 UPM Serdang, Selangor, Malaysia
[4]Laboratory of Aquatic Animal Health and Therapeutics, Institute of Bioscience, Universiti Putra Malaysia,
43400 UPM Serdang, Selangor, Malaysia

ABSTRACT

The aquatic organisms in peat swamp forests are under threat due to habitat degradation resulting from human activities. This study determines the fish gut microbes' diversity, composition, taxa biomarkers, and functional genes in peat swamp forests and its converted areas in North Selangor, Malaysia. Three undisturbed and disturbed areas nearby the peat swamp forests were selected. First, the 16S amplicon metagenomic analysis was conducted to assess the composition and diversity of bacterial communities in fish gut contents from both areas. Then, Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt) and Linear discriminant analysis Effect Size (LEfSe) were used to predict disease/pathogen related functional genes. This study revealed Proteobacteria, Firmicutes, Bacteroidetes, Fusobacteria, and Actinobacteria as the predominant phyla in both studied areas. In contrast, bacterial community profiles of disturbed and undisturbed areas were slightly dissimilar. Metagenome predictions revealed that genes are related to metabolism, environmental information processing, genetic information processing, cellular processes, human diseases, and

organismal systems. Further investigation revealed six potential biomarker genes, including chronic myeloid leukaemia in an undisturbed area, *Vibrio cholerae* infection, bladder cancer, pathogenic *Escherichia coli* infection, *Staphylococcus aureus* infection, and pertussis in disturbed areas. This study revealed that the fish gut microbiome could be used as an indicator in comparing the undisturbed and disturbed ecosystems.

Keywords: Composition, diversity, fish gut microbes, functional genes, peat swamp forests, taxa biomarkers

INTRODUCTION

Tropical peat swamp forest is one of the few biospheres considered an extreme ecosystem, usually characterised by low pH, anaerobic or low oxygen, acidic, and oligotrophic conditions (Spahni et al., 2013). There are about 882,000 hectares (ha) of remaining peat swamp forest (PSF) in Malaysia, where approximately 39% of the forests are located in Peninsular Malaysia (Posa et al., 2011). Specifically, North Selangor Peat Swamp Forest (NSPSF), covers an area of 81,304 ha and is the second-largest contiguous PSF in Peninsular Malaysia (Selangor State Forestry Department [SSFD], 2014).

Peat swamp is home to many aquatic and terrestrial macro and microorganisms (Too et al., 2018). These macro and microorganisms interact with each other in a complex and dynamic way. Microorganisms play significant roles in maintaining the equilibrium and functioning of the peat swamp ecosystem and facilitate the formation of peat, sequestration of carbon,

and nutrient recycles (Chávez-Romero et al., 2016). Moreover, microbes maintain the health of the ecosystem, and host organisms residing in it (R. Zhang et al., 2020; T. Zhang et al., 2016).

A total of 198 peat swamp-associated fish species have been recorded in Malaysia. From this number, a total of 114 species from 23 families, representing about 40% of the known fish fauna in Peninsular Malaysia, were recorded from NSPSF alone (Sule et al., 2016). Fish in peat swamp is an essential part of the food web as it is entirely dependent on peat swamp habitats (Posa et al., 2011). The fish gut environment has a crucial role in maintaining metabolic homeostasis in the fish (Li et al., 2020). Many bacteria are considered desirable to the host and display certain beneficial biological activities to promote the healthy growth of the host (Banerjee et al., 2000; X. Wang et al., 2019). Gut bacteria also trigger the development of the immune system, nutrient assimilation, vitamin biosynthesis and energy metabolism, and growth (Guivier et al., 2020).

The next-generation sequencing (NGS) methods allowed us to investigate the entire complement of organisms inhabiting a particular environment. The availability of bioinformatics tools, such as Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt) to predict the functions of 16S rRNA gene sequences, is desirable for microbial ecologists. It allows them to study the metagenomes of complex microbial communities with reasonable precision and

confidence at a high taxonomic resolution, while being able to construct robust hypotheses for further works (Langille et al., 2013; Sedlar et al., 2017). In this study, the diversity, composition, potential taxa biomarkers, and functional genes of fish gut microbes found in the extreme environment of peat swamp forests and its converted areas in North Selangor, Malaysia were explored. The fish gut microbiome was hypothesised to be used as an indicator in assessing and comparing the pristine peat swamp forest and its converted areas.

MATERIALS AND METHODS

Study Area

This study was conducted in two main areas, namely, in the undisturbed North Selangor Peat Swamp Forest (NSPSF) (undisturbed forest: UF), Malaysia and adjacent disturbed areas surrounding the NSPSF (disturbed forest: DF). The undisturbed and disturbed areas are already classified by previous studies (Sule et al., 2019). Three sampling sites in NSPSF were selected, which were in Sungai Karang Peat Swamp Forest site 1 (UF1; GPS coordinate at 3°30'00.1"N 101°12'08.7"E), Raja Musa Peat Swamp Forest (UF2; 3°29'30.7"N 101°21'51.1"E), and Sungai Karang Peat Swamp Forest site 2 (UF3; 3°41'47.9"N 101°11'03.7"E). Whereas, the sites in the adjacent disturbed area comprises of paddy field (DF1; 3°29'25.8"N 101°11'23.2"E), fire forest (DF2; 3°28'00.6"N 101°26'28.9"E), and oil palm plantation (DF3; 3°42'30.9"N 101°11'04.0"E) areas (Figure 1).

In general, for the undisturbed area, the UF1 site having a high water level and tall trees that almost covering the sampling site,
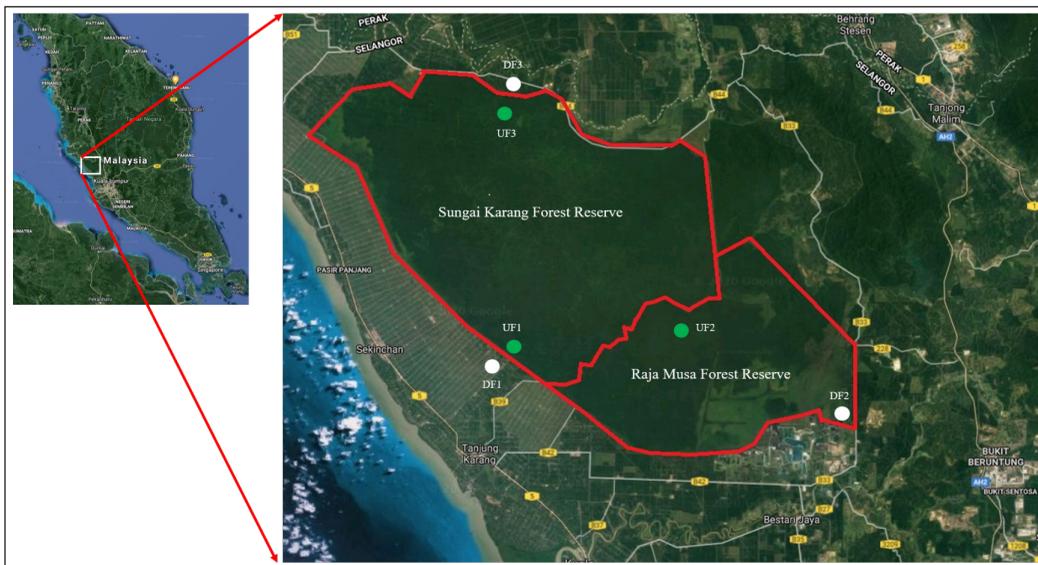

Figure 1. The map of North Selangor Peat Swamp Forest, Selangor, Malaysia. The sampling sites of UF1: Sungai Karang Peat Swamp Forest site 1; UF2: Raja Musa Peat Swamp Forest; and UF3: Sungai Karang Peat Swamp Forest site 2 (undisturbed area; green dots), while adjacent to North Selangor Peat Swamp Forest are DF1: Paddy field; DF2: Forest fire; and DF3: Oil palm plantation (disturbed area; white dots)

sometimes with more than 1 m of depth. UF2 was the second site of the undisturbed area. The forest floor is covered by water. The water depth is averagely about 1 m in height and clearer than the UF1 site. On the other hand, the UF3 site has a high water level but is generally lower than the UF1 and UF2.

However, for the disturbed area, they are said to be anthropogenically altered. The three sites were typically artificial freshwater swamps following conversion of the peat swamp. The entire areas within the sites were majorly exposed to direct sunlight with shrubs and few trees at the bank. These sites were subjected to changes from the peat swamp forest, as evident in the peat characteristics retained in the soil and water. The DF1 site is a mixture of residential buildings and other human activities, such as the main road. The DF2 site shows few signs of irrigated canals and experienced forest fire several years ago, as young trees and plantations can be observed. On the other hand, the DF3 site was covered by oil palm plantations.

Fish Sampling

The fish samples were randomly collected in December 2018 using a combination of gill, cast, and scoop nets in each sampling sites. The weight and length of each fish were measured using an electronic weighing balance and a vernier calliper, respectively. The collected fishes were then placed in a 30 cm long × 60 cm wide × 50 cm high plastic container, and they were euthanised with an overdose of 250 mg/L of MS-222

(Sigma-Aldrich, United States of America). The abdomen was opened at the ventral midline by inserting a fine scalpel blade into the anus of the fish. At the same time, the incision was extended anteriorly, and the gastrointestinal contents (GI) were removed under aseptic conditions. The GI samples were then placed in sterile 50 mL falcon tubes containing RNALater solution (Thermo Fisher Scientific, United States of America) stored at -20°C. The fish were sampled, handled, and sacrificed according to the methods approved by Institutional Animal Care and Use Committee, Universiti Putra Malaysia. All methods were carried under relevant guidelines and regulations. The permit to access and collect samples was obtained from Selangor State Forestry Department, Selangor [Ref. no: P.H.D. 600-5/2/6(19)], while none of the sampled fish species used in this study was considered endangered and protected by the government of Malaysia.

DNA Extraction, PCR, and Sequencing Library Preparation

FavorPrep™ Stool DNA Isolation Mini Kit (Favorgen, Taiwan) was used for DNA extraction of fish gut contents. The quantity and purity of the extracted DNA were then measured and checked using NanoDrop™ 2000 UV-Vis Spectrophotometer (Thermo Fisher Scientific, United States of America) and 1% agarose gel electrophoresis. For further analyses, the concentration of DNA was determined at > 40 µg; meanwhile, the purity of the excellent quality DNA will have an A260/280 ratio of 1.8 – 2.0. The

extracted DNA of the fish gut contents was stored in a freezer at -80ºC until further analyses.

The PCR mixture and the thermal cycling were carried out in duplicates according to Q. Wang et al. (2017) for each extracted DNA. The reactions in a final volume of 25 µL were prepared to contain the following 2.5 µL of 5 ng/µL DNA extract, 5 µL of 0.2 µM 16S (V3-V4) forward primer of 5' TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG CCT ACG GGN GGC WGC AG 3' and 0.2 µM of 5 µL reverse primer of 5' GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA GGA CTA CHV GGG TAT CTA ATC C 3' and 12.5 µL of master mix (DreamTaq Green PCR Master Mix (2×), Thermo Fisher Scientific, United States of America) containing 1.0 µL of 0.2 µM *Taq* polymerase, 2.5 µL of 1 µM buffer, 0.5 µL of 200 µM dNTP, 1.0 µL of 2.0 mM MgCl$_2$, and 7.5 µL of sterilised distilled water. The PCR reaction was carried out in Mastercycler® Nexus Gradient (Eppendorf®, Germany) with an initial denaturation at 95ºC for 3 min, 30 cycles of denaturation at 95ºC for 30 sec, annealing at 55ºC for 30 sec, elongation at 72ºC for 30 sec, extension at 72ºC for 5 min and finally, hold at 10ºC. In addition, gel electrophoresis was carried out on DNA samples.

A total of six DNA samples (three replicates of pooled DNA samples × two undisturbed and disturbed areas) were sent for DNA sequencing on the Illumina Miseq (United States of America) sequencing platform in Novogene Biological Information Technology Co. (China) through Apical Scientific Sdn. Bhd., Malaysia, which results in 250 bp paired-end reads.

Analysis of 16S rRNA Amplicon Sequencing

The raw MiSeq sequencing results in fastq. format was pre-processed and analysed using Quantitative Insights into Microbial Ecology (QIIME ver. 1.9.1 64-bit) pipeline (Caporaso et al., 2010). All sequences were trimmed to remove the primer and barcode sequences using fast length adjustment of short reads (FLASH) (v2.0) (Magoč & Salzberg, 2011) and merged through paired-end read merger (PEAR) (J. Zhang et al., 2014). The reads in fastq. file were subjected to quality filtering at Phred Quality Score of q = 20 and p = 70 using fastq_quality_filter under fastx_toolkit (http://hannonlab.cshl. edu/fastx toolkit/). Chimeric sequences were screened using UCHIME, against the RDP_GOLD (v9) database and were removed from the downstream processing (Haas et al., 2011). The short reads may lead to an unspecific match that could interfere with accuracy. Therefore, to increase the accuracy of the calling reads shorter than 100 bp or longer than 600 bp were removed along with the low-quality bases (Q < 33). The fastq was converted to fasta file by fastq_to_fasta in FastX-toolkit in QIIME. Operational taxonomic unit (OTU) was selected at least 97% similarity threshold using the pick_otus.py script with the usearch_ref method against the Greengenes database through closed references OTU picking. OTU table was

constructed containing the abundance and taxonomic assignments of all OTUs.

Diversity Analysis and Top Percentage Contribution of Taxa

Alpha diversity indexes were calculated to explain the species richness and diversity in each sample. At the same time, rarefactions curves were plotted to determine the adequacy of sequencing depth (Udayangani et al., 2017). A total of 674,795 ($\geq$ 500 bp) clean reads were used to get the best quality in calculating the observed species, Chao1, seChao1, ACE, se.ACE, Shannon, Simpson, InvSimpson, and Fisher.

Beta diversity using Bray-Curtis distance was also determined (Bray & Curtis, 1957). The permutation-based multivariate analysis of variance (PERMANOVA) and analysis of similarities (ANOSIM) was used to test the taxonomic dispersion homogeneity across the areas with 2D estimated stress. The computation of Bray-Curtis, PERMANOVA, and ANOSIM was conducted using PAleontological STatistics (PAST) software (v3.11) (Hammer et al., 2001). The OTU table of the disturbed and undisturbed samples was used to generate non-metric multidimensional scaling (NMDS) in PAST3. Principal component analysis (PCA) was used to visualise similarities or dissimilarities of data based on phylogenetic or distance metrics (Jonsson et al., 2016).

Similarity percentage (SIMPER) was carried out using PAST3 software (v3.11) to determine the top percentage contribution of bacterial taxa from fish gut contents at family and genus levels in disturbed and undisturbed areas. The SIMPER results were visualised in the form of heatmaps, and extended error bars using MicrobiomeAnalyst (https://www.microbiomeanalyst.ca/MicrobiomeAnalyst/faces/docs/Contact.xhtml), Sparse Correlations for Compositional data (SparCC) network of OTU abundance at the class level of the fish gut contents in disturbed and undisturbed areas were constructed.

Predictions of Health-Related Functional Genes

The Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt) (Langille et al., 2013) in the Galaxy server was used to impute the prevalence of genes encoding selected bacterial pathogen/diseases, antibiotic resistance genes, degradation, and other toxic related molecules against Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthology database. Similarly, the presence of specific metabolic and digestive genes whose products are likely to contribute to the fish survival and well-being was determined using the Linear discriminant analysis Effect Size (LEfSe), which measured both biological relevance and statistical significance (Segata et al., 2011). For this analysis, the alpha parameter significance threshold for the Kruskal-Wallis among class was set to 0.01, and the logarithmic linear discriminant analysis (LDA) score cut-off was set to 2.0. This analysis was performed on the Galaxy server.

Statistical Analysis and Data Availability

Shapiro-Wilk's test was conducted to test for normality and homoscedasticity (Levene's test). Similarly, statistical analysis was conducted using Statistical Package for the Social Sciences (SPSS) version 26.0 (SPSS®, United States of America), and statistical significance was determined at $p < 0.05$. The SPSS v26.0 was also used to carry out one-way ANOVA with Tukey's post-hoc test ($p < 0.05$ was considered significant) to compare the differences of alpha diversity indexes among the three disturbed and undisturbed areas. Similarly, ANOVA with Tukey's post-hoc test ($p < 0.05$ was considered significant) was carried out to determine the level of taxa (phylum, class, order, family, and genus) of statistically significant difference among the samples. All the fastq. files for 16S rRNA gene sequence data were also submitted to *National Center for Biotechnology Information* (NCBI) Sequence Read Archive (SRA) database, under the following accession numbers SAMN13909365 (DF1), SAMN13909366 (DF2), SAMN13909367 (DF3), SAMN13909368 (UF1), SAMN13909369 (UF2), and SAMN13909370 (UF3).

RESULTS

Fish Samples

This study collected a total of nine fish species. UF2 site recorded the highest fish samples with 12, while the lowest was in UF3 and DF1 sites with five fish samples. Details of the fish species, number, length and weight, are provided in Appendix 1.

Rarefaction Curves of Bacterial Communities in Undisturbed and Disturbed Areas

The rarefaction and associated diversity analysis showed some variation in bacterial diversity and taxon richness of 16S rRNA genes between undisturbed and disturbed areas. The number of new OTUs found in undisturbed and disturbed areas increased as the number of sequences increased (Figure 2). The plot was rarefied at a minimum library size (850 sequence reads). All samples reached a plateau which indicated that the maximum number of OTU had been identified.

Alpha Diversity of Bacterial Communities in Fish Gut Contents

The 16S rRNA gene sequences of bacterial communities in fish gut contents from undisturbed and disturbed areas were analysed for alpha diversity. In general, species richness was higher in the undisturbed area, in which Observed OTU, Chao1, se.Chao1, ACE, se.ACE, and Fisher had greater values than disturbed areas. On the other hand, diversity indexes such as Shannon, Simpson, and InvSimpson indexes were lower in the undisturbed area (Table 1). However, there was no significant difference ($p > 0.05$) in bacterial diversity indexes and richness among the samples.

Beta Diversity of Bacterial Community among Sites

Bray-Curtis similarity scores were inferred from the taxonomic data generated by QIIME in PAST v3.11 and consequently
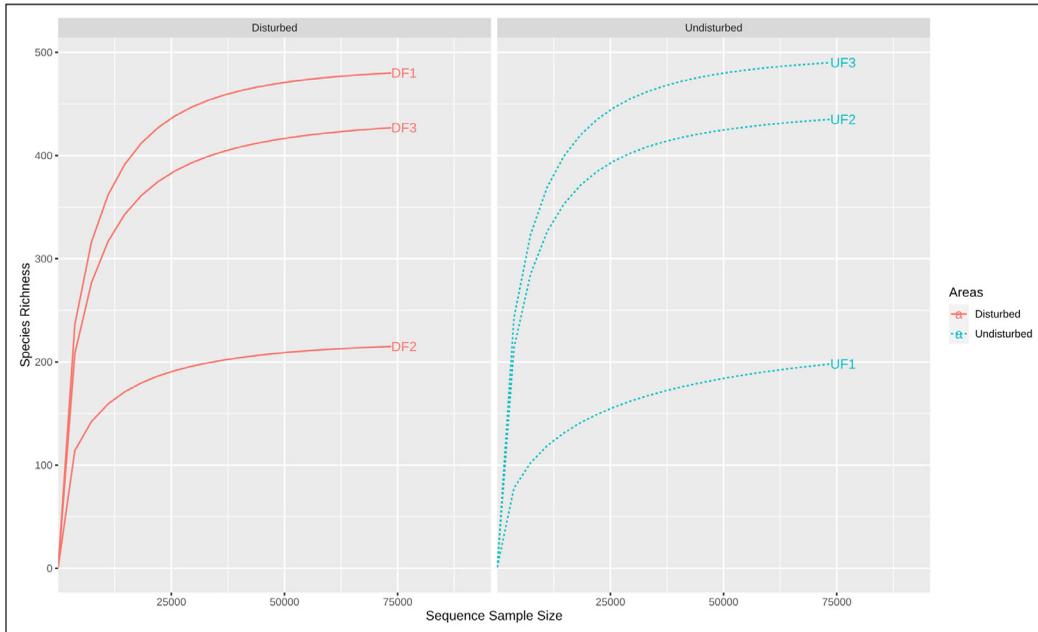
Figure 2. The rarefaction curves show the increase in bacterial operational taxonomic unit's (OTU) and the number of randomly sampled sequences of the bacteria in fish gut contents from undisturbed and disturbed areas. The OTU's detected per sample as a function of the sequencing effort. The plot was rarefied at minimum library size (850 sequence reads)
Note. DF1: Paddy field; DF2: Forest fire; DF3: Oil palm plantation; UF1: Sungai Karang Peat Swamp Forest site 1; UF2: Raja Musa Peat Swamp Forest; UF3: Sungai Karang Peat Swamp Forest site 2

reduced to a two-dimensional space using NMDS to estimate the structural similarity between samples. Although they were slightly separated based on samples between disturbed and undisturbed areas (Figure 3a), it was statistically insignificant ($p >$ 0.05) with a stress value of 0.1796. Figure 3b shows the number of commonly shared (core) and unique OTUs in disturbed and undisturbed samples.

Bacterial Community Structures in Fish Gut Contents

The structure of the bacterial community, particularly at phylum, class, order, family, and genus levels, are described in Appendix 2. There were about 19 - 25 phyla detected

across the samples. DF2 site recorded the lowest, while UF3 site showed the highest number of phyla. A similar pattern was obtained at the genera level, having 165 and 239 genera in DF2 and UF3 areas, respectively. There is no significant difference ($p >$ 0.05) in fish gut bacterial community structures between undisturbed and disturbed areas across taxa (phylum, class, order, family, and genus) (Appendix 3).
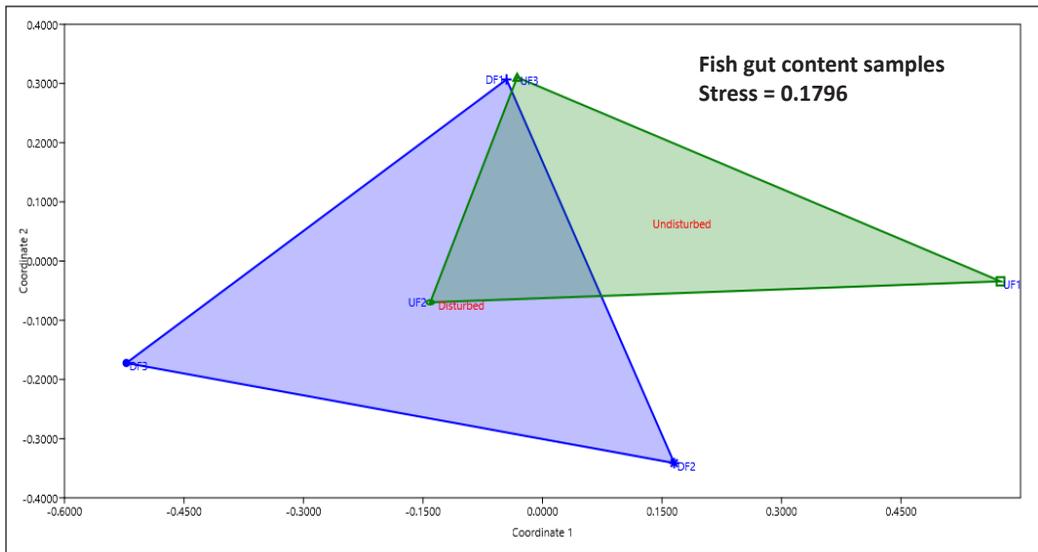
Composition of Bacterial Communities in Fish Gut Contents

A total of six samples of 16S rRNA gene libraries were generated, representing the fish gut contents from the three undisturbed

Table 1
Alpha diversity metrics of bacterial communities in fish gut contents collected from undisturbed and disturbed areas

Site	Observed OTU	Chao1	se.Chao1	ACE	se.ACE	Shannon	Simpson	InvSimpson	Fisher
DF1	1380	1596.47	30.19	1662.06	19.41	3.41	0.89	9.64	246.99
DF2	765	822.50	14.27	825.37	13.14	3.17	0.88	8.77	121.53
DF3	1554	1959.61	47.87	1995.24	22.16	2.43	0.59	2.47	285.51
UF1	1265	1599.94	42.53	1680.64	21.18	2.07	0.54	2.20	2.22
UF2	1596	1853.38	34.80	1863.69	19.94	2.77	0.63	2.74	294.99
UF3	1509	1846.59	41.96	1890.60	21.40	3.54	0.90	10.63	275.43
DF	1233.00 ± 414.53	1459.53 ± 580.79	30.78 ± 16.80	1494.23 ± 602.72	18.24 ± 4.62	3.01 ± 0.51	0.79 ± 0.17	6.97 ± 3.91	218.01 ± 85.74
UF	1456.67 ± 171.59	1766.64 ± 144.40	39.76 ± 4.31	1811.65 ± 114.25	20.85 ± 0.78	2.80 ± 0.74	0.70 ± 0.19	5.19 ± 4.72	264.22 ± 37.66

Note. DF1: Paddy field; DF2: Forest fire; DF3: Oil palm plantation; UF1: Sungai Karang Peat Swamp Forest site 1; UF2: Raja Musa Peat Swamp Forest; UF3: Sungai Karang Peat Swamp Forest site 2; DF: Disturbed forest area; UF: Undisturbed forest area

Pertanika J. Trop. Agric. Sci. 44 (3): 617 - 641 (2021)

625

(a)
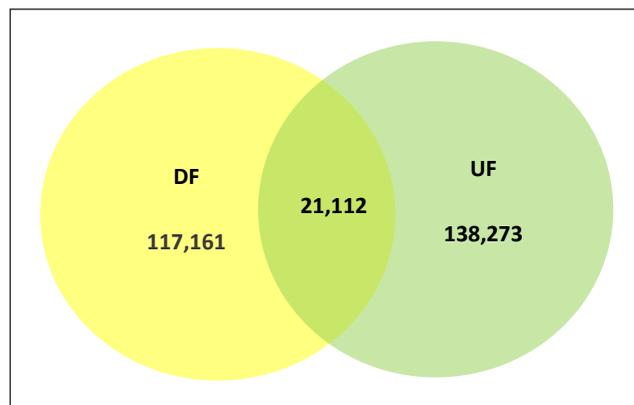
(b)

Figure 3. a) Comparison of NMDS ordination of bacterial communities in fish gut contents collected from undisturbed and disturbed areas with ANOSIM at R = 0.2222. b) Venn diagram summarising the number of core and unique OTUs in disturbed and undisturbed samples
Note. DF1: Paddy field; DF2: Forest fire; DF3: Oil palm plantation; UF1: Sungai Karang Peat Swamp Forest site 1; UF2: Raja Musa Peat Swamp Forest; UF3: Sungai Karang Peat Swamp Forest site 2; DF: Disturbed forest area; UF: Undisturbed forest area

and disturbed areas. The Proteobacteria (36.69%) was the predominant phylum in undisturbed area samples and across all the samples. However, Firmicutes (26.95%) and Bacteroides (4.36%) were the members of the phyla that were higher in disturbed area samples (Figure 4a).

Enterobacteriaceae (42.64%) is the most common family across all samples and is higher in an undisturbed area at the family level. In comparison, additional families present include Ruminococcaceae (10.62%) and Erysipelotrichaceae (9.27%). Whereas, Clostridiaceae (14.63%), Rikenellaceae

(10.94%), and Fusobacteriaceae (7.02%) were also present in disturbed area samples (Figure 4b).

The comparison at the genus level revealed that *Clostridium*, *Providencia*, and *Cetobacterium* representing 28.73%, 22.13%, and 19.14%, respectively were the most dominant genera, particularly in disturbed area samples. On the other hand, the top members recorded in undisturbed area samples include *Lactococcus* (16.36%), *Bacteroides* (12.07%), *Desulfovibrio* (11.30%), and *Candidatus* Arthromitus (6.97%) (Figure 4c).

Percentage Contribution of Bacterial Taxa

The percentage contribution of bacterial taxa was determined at the family level, and overall average dissimilarities showed 46.08 (Appendix 4). The results revealed that Enterobacteriaceae (77.90%) and Streptococcaceae (6.70%) recorded a higher percentage contribution in an undisturbed area, particularly in the UF1. On the other hand, the disturbed area recorded a higher percentage contribution of families Clostridiaceae (19.20%) and Peptostreptococcaceae (10.30%) in the DF2 area, while Fusobacteriaceae (6.15%) in the DF3 area. Comparison of disturbed and undisturbed sample areas showed Clostridiaceae, Fusobacteriaceae, Erysipelotrichaceae, and Peptostreptococcaceae were family members with a higher percentage contribution in the disturbed area. Conversely, Enterobacteriaceae, Ruminococcaceae, Streptococcaceae, and Desulfovibrionaceae families were more prevalent in undisturbed areas (Figure 5a).

Figure 5b shows the top 15 contributors at the genus level. The high percentage contributors of taxa in undisturbed and disturbed areas were *Providencia* (15.10%) and *Weissella* (0.98%), respectively. Whereas, *Desulfovibrio* (5.09%) and *Methylosium* (1.44%) were recorded in the UF1 area. In contrast, the disturbed area samples showed *Lactococcus* (6.57%), *Enterobacter* (1.61%), *Turicibacter* (1.12%), *Pseudomonas* (0.64%), and *Epulopiscium* (0.39%) were mainly higher in the DF1 area. At the same time, *Bacteroides* and *Candidatus* Arthromitus have the highest percentage contribution in the DF3 area, represented by 5.33% and 2.67%, respectively. Thus, the results indicated relatively higher *Providencia*, *Clostridium*, *Cetobacterium*, *Desulfovibrio*, and *Weissella* in undisturbed than in disturbed area samples. However, the genera of *Lactococcus*, *Enterobacter*, *Turicibacter*, *Bacteroides*, *Candidatus* Arthromitus, and *Methylosium* were predominant in the disturbed area.

Co-occurrence Networks between Different Fish Gut Bacterial Classes

The SparCC was used to elucidate the networks of interaction among the classes of fish gut microbiota. The bacterial abundance in disturbed and undisturbed areas showed significant associations among the taxonomic clades in a single giant cluster with one main taxa, Deferribacteres in the
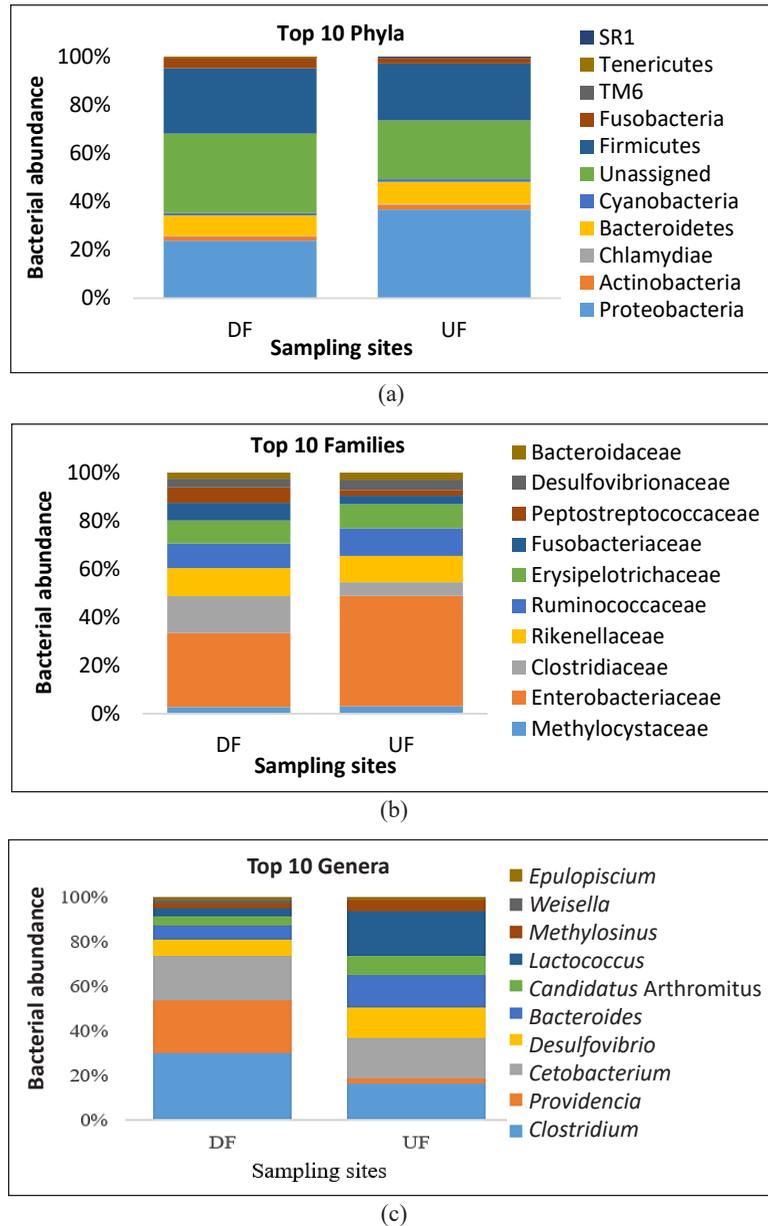
(a)

(b)

(c)

Figure 4. Comparison of relative abundances of top 10 bacterial communities in gastrointestinal contents collected from disturbed and undisturbed areas: (a) Phyla; (b) Family; and (c) genus level
Note. DF: Disturbed forest area; UF: Undisturbed forest area

centre of the cluster, having positive or negative correlations to the other four sub-clusters (Figure 6). The class Deferribacteres linked with the first sub-cluster through

Chlamydia and second sub-cluster via Erysipelotrichia, while Bacteroidia bonded the third sub-cluster by class SJA_4. All the networks depicted either positive or
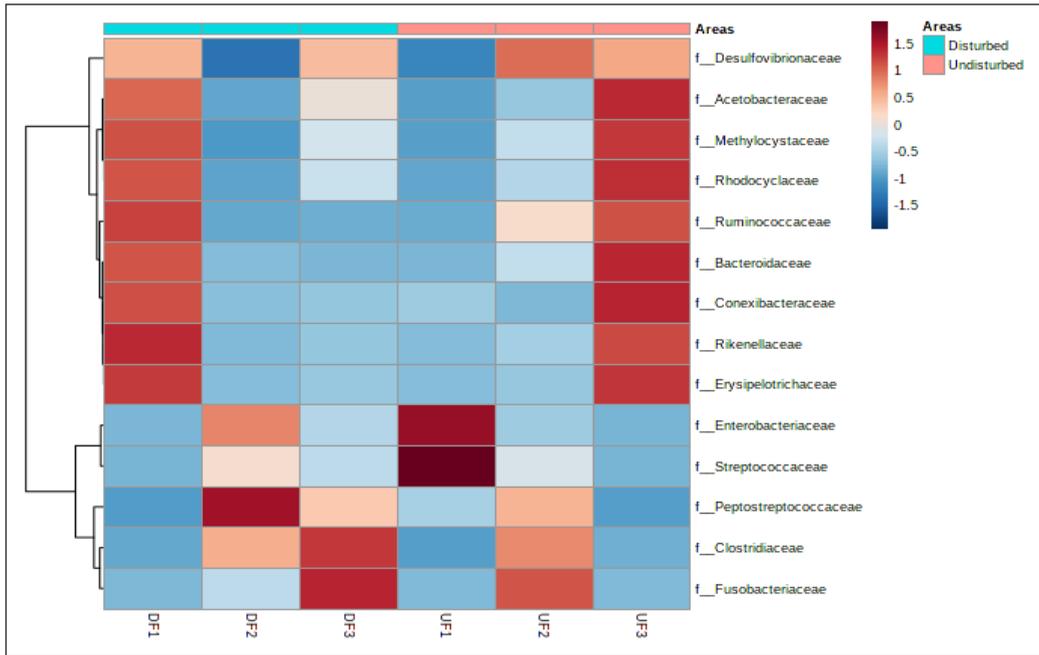
(a)
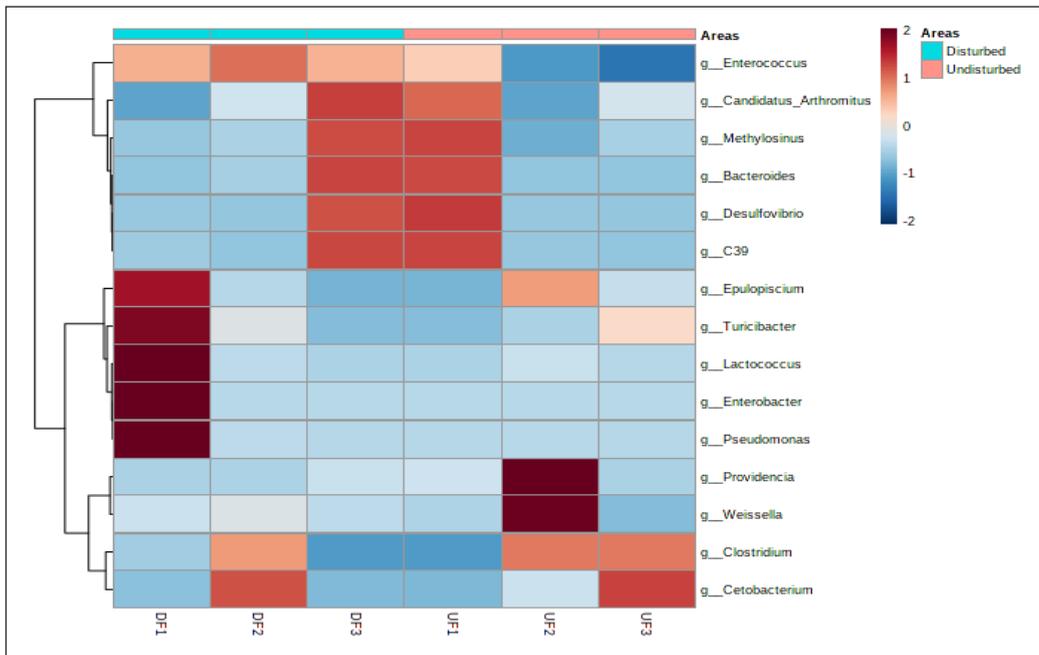
(b)

Figure 5. Heatmaps of top 15 high percentage contribution of bacterial abundance at a) family and b) genus level across samples of fish gut contents collected from disturbed and undisturbed areas. ANOSIM showed no significant difference ($p > 0.05$) in fish gut contents collected from disturbed and undisturbed areas
Note. DF1: Paddy field; DF2: Forest fire; DF3: Oil palm plantation; UF1: Sungai Karang Peat Swamp Forest site 1; UF2: Raja Musa Peat Swamp Forest; UF3: Sungai Karang Peat Swamp Forest site 2

negative associations. The first sub-cluster, Acidobacteria, showed a strong and positive association between Nostocophycideae and Ocillatoriophycideae, prevalent in the disturbed area. The Flavobacteriia recorded higher abundance in the undisturbed area and correlated positively with Acidobacteria. However, *Chlamydia* showed a negative and strong correlation with the three classes.

The second sub-cluster linked to Deferribacteres was prevalent in the disturbed area with a positive and strong association with Betaproteobacteria and Bacteroidia.

On the other hand, Deltaproteobacteria were higher in the undisturbed area, negatively and strongly correlated with Fusobacteria. Deferribacteres displayed positive associations with Erysipelotrichi, *Chlamydia*, Betaproteobacteria, ABS_6, and Bacteroidia, forming the third sub-cluster. Finally, the fourth sub-cluster revealed Thermoplasmata exclusively found in the disturbed area, having negative and strong correlations with Methanobacteria and SJA_4. In contrast, class SJA_4 was correlated positively with Methanobacteria and Bacteroidia.
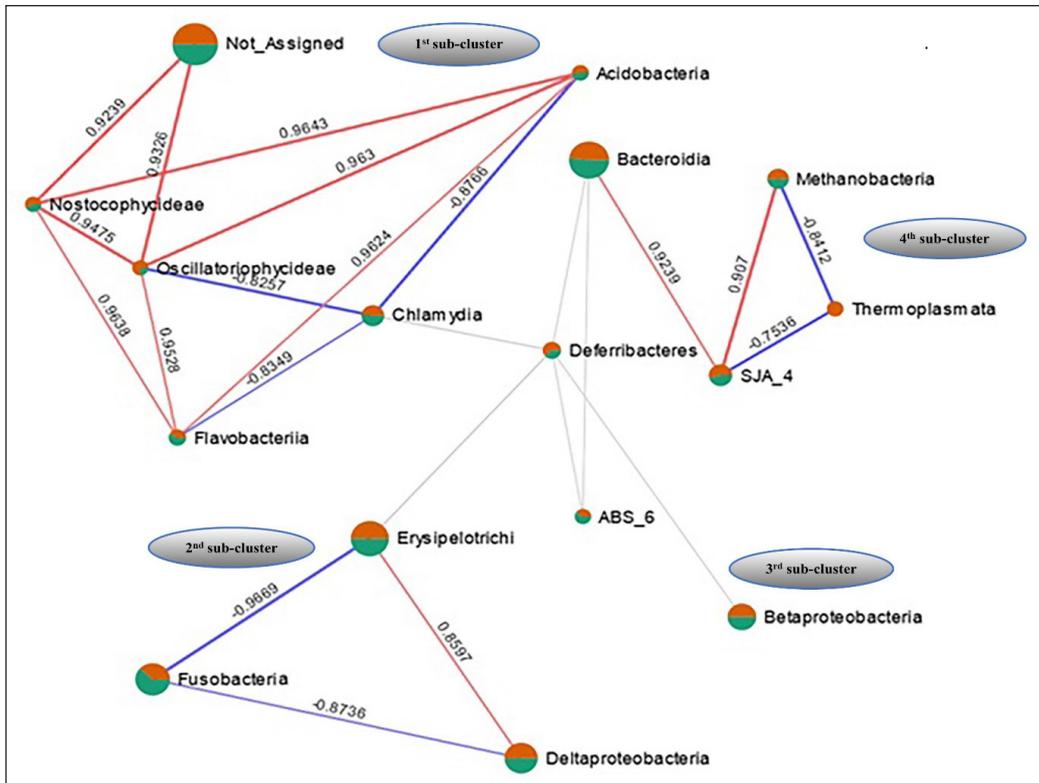

Figure 6. SparCC network of the mean OTU abundance at a class level from fish gut contents collected from disturbed and undisturbed areas at 200 permutations. The significant correlations exhibited between the classes of bacteria at *p* value = 0.05 and *r* = 0.7 thresholds. The taxonomic clades are represented in blue circles. The fish gut samples collected from the disturbed area are represented by orange colour. In contrast, fish gut contents collected from the undisturbed area were labelled green. Additionally, the thicker line indicates a more positive or negative association between or among taxa. In contrast, the thinner lines showed a lesser association

Functional and Biomarker Genes Associated with Fish Health

PICRUSt analysis predicted eight primary functional genes in KEGG pathways (KEGG level 1). Metabolism was comparatively higher in undisturbed area samples. Others include environmental information processing, genetic information processing, unclassified, cellular processes, human diseases, organismal systems, and none recorded the least functional genes across all samples (Figure 7a).

Health-related functional genes from PICRUSt prediction were sorted, and LEfSe was used to find health-related biomarkers. The abundance of the biomarker genes in fish gut samples from undisturbed and disturbed areas was compared. LEfSe identified six biomarker genes associated with diseases, antibiotics, and toxins in the samples. The biomarkers include chronic myeloid leukaemia, *Vibrio cholerae* infection, bladder cancer, pathogenic *Escherichia coli* infection, *Staphylococcus aureus* infection,
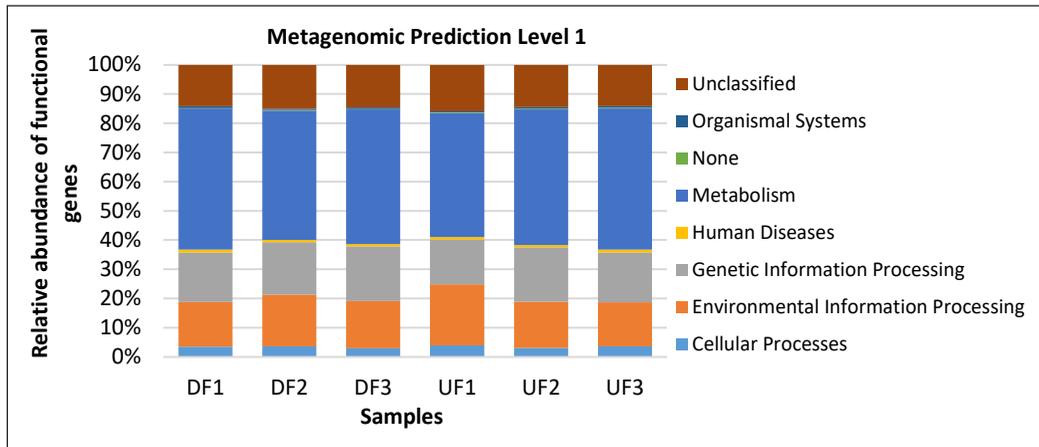
(a)
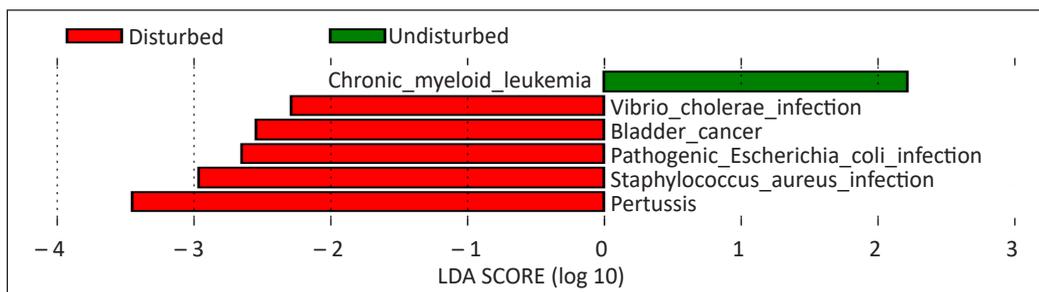
(b)

Figure 7. (a) Relative abundance of predicted KEGG Orthologs functional profiles (KEGG level 1) of fish gut bacterial community was collected from disturbed and undisturbed areas. The results were based on log2 (+1 transformed values). In addition, the functional prediction between groups was tested using bootstrap Mann-Witney U-test with cut off at $p < 0.05$; (b) LEfSe of pathogen or disease-related functional biomarker genes at KEGG level 3 in disturbed and undisturbed samples of fish gut contents
Note. DF1: Paddy field; DF2: Forest fire; DF3: Oil palm plantation; UF1: Sungai Karang Peat Swamp Forest site 1; UF2: Raja Musa Peat Swamp Forest; UF3: Sungai Karang Peat Swamp Forest site 2

and pertussis. Except for chronic myeloid leukaemia, the rest were mainly associated with disturbed area samples. Interestingly, pertussis and *S. aureus* infection were over-represented in the disturbed area (Figure 7b).

DISCUSSION

In this study, following our fish sampling activities, we could not find a uniform fish species representing a fair comparison between the undisturbed and disturbed areas and within the undisturbed and disturbed areas themselves. However, all fish used in this study were primarily omnivorous, which feed on various food of plant and animal (insects) origin. Thus, it will rule out the bias in the context of fish feeding behaviour. Moreover, samples from each site were pooled to represent the fish gut microbiome data from each area.

Alpha diversity measures of the bacterial community in the fish gut contents were rich and diverse. In the peat forest, relatively higher bacterial diversity in undisturbed soil was also reported (Sun et al., 2014). By implication, the comparative low bacterial diversity in the disturbed area might be affected by effluents produced by anthropogenic activity, which appeared higher in the disturbed area. The rarefaction curve began to reach a plateau pattern, suggesting decreasing frequency of discovery of new taxa with increasing sequencing depth. The number of species becomes saturated, regardless of increased sampling frequency. The rarefaction curves indicate that the sampling sufficiently

captures the species in the study area. Unique taxa recorded in both disturbed and undisturbed areas indicated that the bacterial communities are most likely reflect by the environmental condition surrounding the host, which in this study is the anthropogenic factors (Nayak, 2010; Nolorbe-Payahua et al., 2020).

The tropical peat swamp forests are nutrient and oxygen-deficient (Aw et al., 2016), displaying unique systems with core microbial assemblages, such as Proteobacteria, Bacteroides, Actinobacteria, Cyanobacteria, Chlamydiae, and some other phyla. Fish intestines particularly harbour large and diverse populations of bacteria (Givens et al., 2015). Similar bacterial phyla and families were also predominant in NSPSF (Too et al., 2018). Proteobacteria can withstand polluted and extreme environments as in the peat swamp forests (Ward et al., 2009). Members of the proteobacterial methanotrophs, such as the genera *Methylomonas*, *Methylobacter*, *Methylococcus* (type I), *Methylosinus*, and *Methylocystis* (type II), oxidise atmospheric and subsurface methane before it can be released into the atmosphere (Nguyen et al., 2018). This process explains the prevalence of the phylum Proteobacteria in the fish gut from undisturbed area samples.

This study found a change in bacterial diversity in fish microflora inhabiting a disturbed forest compared to an undisturbed forest area geographically distant from any possible source of human activities and pollutants. Enterobacteriaceae appeared to be dominant in undisturbed than disturbed

area samples. Other prevalent members in the disturbed area included Clostridiaceae, Fusobacteriaceae, Erysipelotrichaceae, and Methylocystaceae. The results showed differences between undisturbed and disturbed areas on the fish gut microbiome composition, confirming the previous finding that the surrounding habitat influences the microbial flora in the fish intestinal tract (Sule et al., 2019). However, the composition of the fish gut bacterial community varied depending on habitat (Yukgehnaish et al., 2020). Enterobacteriaceae is a sizeable Gram-negative group with more than 50 genera and over 200 species (Don et al., 2005). Members of this family are frequently found in the normal microbiota of the fish gut (Egerton et al., 2018; Oliveira et al., 2017). Therefore, it might be among the important resident microbiota having a symbiotic relation to its host (Yukgehnaish et al., 2020). However, the family members vary, from beneficial or harmless to several pathogenic members such as *Salmonella*, *Escherichia*, *Yersinia*, *Klebsiella*, *Shigella*, *Serratia*, and *Proteus* (Don et al., 2005). Moreover, the higher abundance of Enterobacteriaceae in the fish gut sample of the undisturbed area could be due to its important role as a core microbiota in the fish gut. In contrast, Enterobacteriaceae in the disturbed area has low abundance compared to the undisturbed area. At the same time, Clostridiaceae and Fusobacteriaceae were found higher abundance in disturbed compared to the undisturbed area. Hence, the impacts of anthropogenic activities and the by-product carried out in that area can alter the core microbiome of fish

species. This finding agrees with the other previous studies that were suggesting that environmental waste causes changes in the gut microflora of aquatic species. The fish species exposed to anthropic contaminants within their ecosystem lead to changes in microbial composition (Evariste et al., 2019; Giang et al., 2018). In the present study, we found bacterial groups in fish guts that could represent potential biomarkers of environmental contamination by anthropogenic activities. Furthermore, genus *Clostridium* was observed more abundant in disturbed compared to undisturbed areas. *Clostridium* were considered one of the markers for persistent faecal contamination in stream water (Mushi, 2018). Hence, the presence of *Clostridium* in fish gut suggests that the fish's environment was disturbed by anthropogenic influences.

Several functional genes related to health consequences in fish were deduced from the predicted PICRUSt results. The health-related genes were sorted, and differentially abundant features in undisturbed and disturbed areas were analysed using LEfSe. Bladder cancer might be associated with oxidative DNA damage, which plays an important role in the pathogenesis of some human diseases, including cancer (Szymańska & Długosz, 2017). Fusobacteriacea and *Acinetobacter* were among the taxa with the maximum percentage contribution in the undisturbed fish gut samples. Moreover, *Fusobacterium nucleatum* and some members of the Ruminococcaceae family (Bučević et al., 2018) and *Acinetobacter* (Mai et al., 2019)

were detected in patients with bladder cancer. Thus, genera *Fusobacterium* and *Acinetobacter* could be a potential protumorigenic pathogen.

Other detected functions are chronic myeloid leukaemia and *Staphylococcus aureus* infection. Similarly, chronic myeloid leukaemia is associated with increased and unregulated growth of myeloid cells in the bone marrow and blood (Nowell, 2007). On the other hand, the pathogenic *Escherichia coli* infection is responsible for enteropathogenic and enterohaemorrhagic *E. coli*. It is causing gastroenteritis and bloody diarrhoea in infants, posing a major threat to human health, especially in the developing world (Humphreys et al., 2016). *Staphylococcus aureus* infection is linked with humans and animals' infection. The presence of staphylococci in fish indicating contamination has occurred. The situation is worrisome as it is a harmful human pathogen causing nosocomial infections. Fish consumption contaminated with *S. aureus* could lead to food poisoning (Hammad et al., 2012). *Vibrio cholerae* pathogenic cycle is connected to infective stages of *V. cholerae*, a cosmopolitan bacterium that inhabits a vast range of environments and is associated with many aquatic organisms. *Vibrio cholerae* possesses a diverse metabolism characterised by rapid inter-regulation of energy-producing pathways to adapt between environments (Bueno et al., 2020). The bacterium is found in water contaminated with human faeces/sewage. It is considered a significant source of pathogenic bacteria responsible for cholerae in humans (Pandey & Mishra, 2020). Lastly, pertussis is associated with a respiratory infectious disease caused by a pathogen, *Bordetella pertussis*. The disease is characterized by paroxysmal cough, inspiratory wheezing, and vomiting (Beier & Gross, 2006).

CONCLUSION

The results revealed a high diversity of microbes in fish gut contents collected from undisturbed and disturbed areas. However, the species richness was relatively lower in disturbed areas. The partial overlap between the undisturbed and disturbed area samples in NMDS analysis suggests high similarity in the composition of the fish gut bacterial community. There could be a possible movement of fishes from disturbed to undisturbed areas, or *vice versa* during high water levels in raining season. Moreover, omnivorous feeding behaviours of all the collected fishes also might be the factor. Rarefaction analysis indicates sufficient sequencing depth. Overall, fish gut bacterial communities in the disturbed area showed a lower total absolute count than the undisturbed area. This study revealed that the fish gut microbiome could be used as an indicator in accessing and comparing the healthy and disturbed ecosystems.

ACKNOWLEDGEMENTS

REFERENCES

Aw, Y. K., Ong, K. S., Lee, L. H., Cheow, Y. L., Yule, C. M., & Lee, S. M. (2016). Newly isolated *Paenibacillus tyrfis* sp. nov. from Malaysian tropical peat swamp soil with broad spectrum antimicrobial activity. *Frontiers in Microbiology*, *7*, 219. https://doi.org/10.3389/fmicb.2016.00219

Banerjee, S., Azad, S. A., Vikineswary, S., Selvaraj, O. S., & Mukherjee, T. K. (2000). Phototrophic bacteria as fish feed supplement. *Asian-Australian Journal of Animal Science*, *13*(7), 991-994. https://doi.org/10.5713/ajas.2000.991

Beier, D., & Gross, R. (2006). Regulation of bacterial virulence by two-component systems. *Current Opinion in Microbiology*, *9*(2), 143-152. https://doi.org/10.1016/j.mib.2006.01.005

Bray, J. R., & Curtis, J. T. (1957). An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs*, *27*(4), 326-349. https://doi.org/10.2307/1942268

Bučević, P. V., Šitum, M., Chow, C. E. T., Chan, L. S., Roje, B., & Terzić, J. (2018). The urinary microbiome associated with bladder cancer. *Scientific Reports*, *8*(1), 12157. https://doi.org/10.1038/s41598-018-29054-w

Bueno, E., Pinedo, V., & Cava, F. (2020). Adaptation of *Vibrio cholerae* to hypoxic environments. *Frontiers in Microbiology*, *11*, 739. https://doi.org/10.3389/fmicb.2020.00739

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., & Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, *7*(5), 335-336. https://doi.org/10.1038/nmeth.f.303

Chávez-Romero, Y., Navarro-Noya, Y. E., Reynoso-Martínez, S. C., Sarria-Guzmán, Y., Govaerts, B., Verhulst, N., & Luna-Guido, M. (2016). 16s metagenomics reveals changes in the soil bacterial community driven by soil organic C, N-fertilizer and tillage-crop residue management. *Soil and Tillage Research*, *159*, 1-8. https://doi.org/10.1016/j.still.2016.01.007

Don, J. B., Noel, R. K., & James, T. S. (2005). Part B: The Gammaproteobacteria. In *Bergey's manual of systematic bacteriology* (2nd ed., pp. 1107-1109). Springer.

Egerton, S., Culloty, S., Whooley, J., Stanton, C., & Ross, R. P. (2018). The gut microbiota of marine fish. *Frontiers in Microbiology*, *9*(873), 1-17. https://doi.org/10.3389/fmicb.2018.00873

Evariste, L., Barret, M., Mottier, A., Mouchet, F., Gauthier, L., & Pinelli, E. (2019). Gut microbiota of aquatic organisms: A key endpoint for ecotoxicological studies. *Environmental Pollution*, *248*, 989-999. https://doi.org/10.1016/j.envpol.2019.02.101

Giang, P. T., Sakalli, S., Fedorova, G., Tilami, S. K., Bakal, T., Najmanova, L., & Grabic, R. (2018). Biomarker response, health indicators, and intestinal microbiome composition in wild brown trout (*Salmo trutta* m. *fario* L.) exposed to a sewage treatment plant effluent-dominated stream. *Science of the Total Environment*, *625*, 1494-1509. https://doi.org/10.1016/j.scitotenv.2018.01.020

Givens, C., Ransom, B., Bano, N., & Hollibaugh, J. (2015). Comparison of the gut microbiomes of 12 bony fish and 3 shark species. *Marine Ecology Progress Series*, *518*, 209–223. https://doi.org/10.3354/meps11034

Guivier, E., Pech, N., Chappaz, R., & Gilles, A. (2020). Microbiota associated with the skin, gills, and gut of the fish *Parachondrostoma toxostoma* from the Rhône basin. *Freshwater Biology*, *63*(3), 446-459. https://doi.org/10.1111/fwb.13437

Haas, B. J., Gevers, D., Earl, A. M., Feldgarden, M., Ward, D. V., & Giannoukos, G. (2011). Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons.

Pertanika J. Trop. Agric. Sci. 44 (3): 617 - 641 (2021)

635

Genome Research, *21*, 494-504. https://doi.org/10.1101/gr.112730.110

Hammad, A. M., Watanabe, W., Fujii, T., & Shimamoto, T. (2012). Occurrence and characteristics of methicillin-resistant and -susceptible *Staphylococcus aureus* and methicillin-resistant coagulase-negative staphylococci from Japanese retail ready-to-eat raw fish. *International Journal of Food Microbiology*, *156*(3), 286–289.

Hammer, Ø., Harper, D. A., & Ryan, P. D. (2001). PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica*, *4*(1), 1-9.

Humphreys, D., Singh, V., & Koronakis, V. (2016). Inhibition of WAVE regulatory complex activation by a bacterial virulence effector counteracts pathogen phagocytosis. *Cell Reports*, *17*(3), 697–707. https://doi.org/10.1016/j.celrep.2016.09.039

Jonsson, V., Österlund, T., Nerman, O., & Kristiansson, E. (2016). Statistical evaluation of methods for identification of differentially abundant genes in comparative metagenomics. *BMC Genomics*, *17*(1), 1-12. https://doi.org/10.1186/s12864-016-2386-y

Langille, M. G. I., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A., & Huttenhower, C. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature Biotechnology*, *31*(9), 814-821. https://doi.org/10.1038/nbt.2676

Li, Z., Tian, J., Lai, Y., Lee, C.-H., Cai, Z., & Yu, C.-F. (2020). Puffer fish gut microbiota studies revealed unique bacterial co-occurrence patterns and new insights on tetrodotoxin producers. *Marine Drugs*, *18*(5), 278. https://doi.org/10.3390/md18050278

Magoč, T., & Salzberg, S. L. (2011). FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, *27*(21), 2957-2963. https://doi.org/10.1093/bioinformatics/btr507

Mai, G., Chen, L., Li, R., Liu, Q., Zhang, H., & Ma, Y. (2019). Common core bacterial biomarkers of bladder cancer based on multiple datasets. *BioMed Research International*, *2019*, 4824909. https://doi.org/10.1155/2019/4824909

Mushi, D. (2018). *Clostridium perfringens* identifies source of pollution and reference streams in a tropical highland environment. *Journal of Water and Health*, *16*(4), 501-507. https://doi.org/10.2166/wh.2018.192

Nayak, S. K. (2010). Role of gastrointestinal microbiota in fish. *Aquaculture Research*, *41*(11), 1553-1573. https://doi.org/10.1111/j.1365-2109.2010.02546.x

Nguyen, N. L., Yu, W. J., Gwak, J. H., Kim, S. J., Park, S. J., Herbold, C. W., & Rhee, S. K. (2018). Genomic insights into the acid adaptation of novel methanotrophs enriched from acidic forest soils. *Frontiers in Microbiology*, *9*, 1982. https://doi.org/10.3389/fmicb.2018.01982

Nolorbe-Payahua, C. D., de Freitas, A. S., Roesch, L. F. W., & Zanette, J. (2020). Environmental contamination alters the intestinal microbial community of the livebearer killifish *Phalloceros caudimaculatus*. *Heliyon*, *6*(6), e04190. https://doi.org/10.1016/j.heliyon.2020.e04190

Nowell, P. C. (2007). Discovery of the Philadelphia chromosome: A personal perspective. *Journal of Clinical Investigation*, *117*(8), 2033-2035. https://doi.org/10.1172/JCI31771

Oliveira, R. V., Oliveira, M. C., & Pelli, A. (2017). Disease infection by Enterobacteriaceae family in fishes: A review. *Journal of Microbiology and Experimentation*, *4*(5), 128. https://doi.org/10.15406/jmen.2017.04.00128

Pandey, V. K., & Mishra, P. K. (2020). Nanoconjugates for detection of waterborne bacterial pathogens. *Waterborne Pathogens*, *2020*, 363-384. https://doi.org/10.1016/B978-0-12-818783-8.00018-9

Posa, M. R. C., Wijedasa, L. S., & Corlett, R. T. (2011). Biodiversity and conservation of tropical peat swamp forests. *BioScience*, *61*(1), 49-57. https://doi.org/10.1525/bio.2011.61.1.10

Sedlar, K., Kupkova, K., & Provaznik, I. (2017). Bioinformatics strategies for taxonomy independent binning and visualization of sequences in shotgun metagenomics. *Computational and Structural Biotechnology Journal*, *15*, 48-55. https://doi.org/10.1016/j.csbj.2016.11.005

Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., & Huttenhower, C. (2011). Metagenomic biomarker discovery and explanation. *Genome Biology*, *12*(6), 1-60. https://doi.org/10.1186/gb-2011-12-6-r60

Selangor State Forestry Department. (2014). *Integrated management plan for North Selangor Peat Swamp Forest 2014-2023*. Retrieved from file:///C:/Users/user/Downloads/201511_imp-nspsf-vol-1_web-03%20(2).pdf

Spahni, R., Joos, F., Stocker, B. D., Steinacher, M., & Yu, Z. C. (2013). Transient simulations of the carbon and nitrogen dynamics in northern peat lands: From the last glacial maximum to the 21st century. *Climate of the Past*, *9*(3), 1287-1308. https://doi.org/10.5194/cp-9-1287-2013

Sule, H. A., Ismail, A., & Amal, M. N. A. (2016). A review of the ichthyofauna of Malaysian peat swamp forest. *Pertanika Journal of Tropical Agricultural Science*, *39*(4), 421-458.

Sule, H. A., Ismail, A., Amal, M. N. A., Zulkifli, S. Z., & Roseli, M. A. M. (2019). Associations between the presence of bacteria and the physico-chemical parameters of water in peat swamp forest, paddy field and oil palm plantation in north Selangor, Malaysia. *Pertanika Journal of Tropical Agricultural Science*, *42*(1), 185-207.

Sun, H., Terhonen, E., Koskinen, K., Paulin, L., Kasanen, R., & Asiegbu, F. O. (2014). Bacterial diversity and community structure along different

peat soils in boreal forest. *Applied Soil Ecology*, *74*(2014), 37-45. https://doi.org/10.1016/j.apsoil.2013.09.010

Szymańska, B., & Długosz, A. (2017). The role of the BLCA-4 nuclear matrix protein in bladder cancer. *Postepy Hig Med Dosw*, *2017*(71), 681-691. https://doi.org/ 10.5604/01.3001.0010.3847

Too, C., Keller, A., Sickel, W., Lee, S., & Yule, C. M. (2018). Microbial community structure in a Malaysian tropical peat swamp forest: The influence of tree species and depth. *Frontiers in Microbiology*, *9*, 2859. https://doi.org/10.3389/fmicb.2018.02859

Udayangani, R. M. C., Dananjaya, S. H. S., Fronte, B., Kim, C. H., Lee, J., & De Zoysa, M. (2017). Feeding of nano scale oats β-glucan enhances the host resistance against *Edwardsiella tarda* and protective immune modulation in zebrafish larvae. *Fish and Shellfish Immunology*, *60*, 72-77. https://doi.org/10.1016/j.fsi.2016.11.035

Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2017). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, *73*(16), 5261-5267. https://doi.org/10.1128/AEM.00062-07

Wang, X., Li, Q., Sui, J., Zhang, J., Liu, Z., Du, J. and Liu, X. (2019). Isolation and characterization of antagonistic bacteria *Paenibacillus jamilae* HS-26 and their effects on plant growth. *BioMed Research International*, *2019*, 3638926. https://doi.org/10.1155/2019/3638926

Ward, N. L., Challacombe, J. F., Janssen, P. H., Henrissat, B., Coutinho, P. M., Wu, M., & Kuske, C. R. (2009). Three genomes from the Phylum Acidobacteria provide insight into the lifestyles of these microorganisms in soils. *Applied and Environmental Microbiology*, *75*(7), 2046-2056. https://doi.org/10.1128/AEM.02294-08

Yukgehnaish, K., Kumar, P., Sivachandran, P., Marimuthu, K., Arshad, A., Paray, B. A.,

& Arockiaraj, J. (2020). Gut microbiota metagenomics in aquaculture: Factors influencing gut microbiome and its physiological role in fish. *Reviews in Aquaculture*, *12*(3), 1903-1927. https://doi.org/10.1111/raq.12416

Zhang, J., Wang, X., Huo, D., Li, W., Hu, Q., Xu, C., & Li, C. (2014). Metagenomic approach reveals microbial diversity and predictive microbial metabolic pathways in Yucha, a traditional Li fermented food. *Scientific Reports*, *6*(1), 32524. https://doi.org/10.1038/srep32524

Zhang, R., Liu, L.-L., Wang, X.-W., Guo, C.-Y., & Zhu, H. (2020). Dietary tea polyphenols induce changes in immune response and intestinal microbiota in Koi carp, *Cyprinus carpio*. *Aquaculture*, *516*, 734636. https://doi.org/10.1016/j.aquaculture.2019.734636

Zhang, T., Yang, C., Qu, H., Xia, Y., Wang, Y., Li, A. D., & Liu, R. (2016). Discovery of new cellulases from the metagenome by a metagenomics-guided strategy. *Biotechnology for Biofuels*, *9*(1), 138. https://doi.org/10.1186/s13068-016-0557-3

APPENDICES

Appendix 1
The collected fish species in this study

Sites	Fish species	Length (cm)	Weight (g)
UF1	*Pristolepis fasciata*	4.5	4.3
	Anabas testudineus	3.5	0.5
	Anabas testudineus	4.4	4.3
	Anabas testudineus	2.6	0.3
	Trichopsis vittata	6.7	2.0
	Pristolepis fasciata	4.0	0.5
UF2	*Mystus singaringan*	11.5	9.0
	Hampala macrolepidota	8.0	5.0
	Trichopsis vittata	11.5	10.0
	Rasbora dusonensis	12.0	16.0
	Trichopsis vittata	7.0	2.0
	Pristolepis fasciata	5.0	2.0
	Rasbora dusonensis	10.5	9.0
	Trichopsis vittata	6.2	2.0
	Hemibagrus capitulum	6.0	1.0
	Pristolepis fasciata	6.0	5.0
	Trichopsis vittata	4.5	1.0
	Trichopsis vittata	6.7	2.0
UF3	*Helostoma temmincki*	4.5	4.3
	Pristolepis fasciata	5.6	4.0
	Anabas testudineus	5.8	4.0
	Anabas testudineus	5.7	3.0
	Anabas testudineus	6.3	5.0
DF1	*Pristolepis fasciata*	7.0	5.0
	Trichopodus trichopterus	7.0	4.0
	Trichopodus trichopterus	6.0	3.0
	Trichopodus trichopterus	5.5	2.0
	Pristolepis fasciata	5.0	4.0
DF2	*Pristolepis fasciata*	6.5	4.0
	Anabas testudineus	4.5	1.0
	Pristolepis fasciata	6.7	6.0
	Anabas testudineus	5.0	7.5
	Pristolepis fasciata	6.5	5.0
	Anabas testudineus	5.0	7.0
	Anabas testudineus	5.5	8.0
	Anabas testudineus	5.0	7.0
	Anabas testudineus	6.5	5.0
DF3	*Trichopodus trichopterus*	2.5	0.2
	Trichopodus trichopterus	3.4	0.4
	Trichopodus trichopterus	2.1	0.2
	Trichopodus trichopterus	2.4	0.2
	Trichopodus trichopterus	2.3	0.2
	Trichopodus trichopterus	3.5	1.0

Note. UF1: Sungai Karang Peat Swamp Forest site 1; UF2: Raja Musa Peat Swamp Forest; UF3: Sungai Karang Peat Swamp Forest site 2; DF1: Paddy field; DF2: Forest fire; DF3: Oil palm plantation
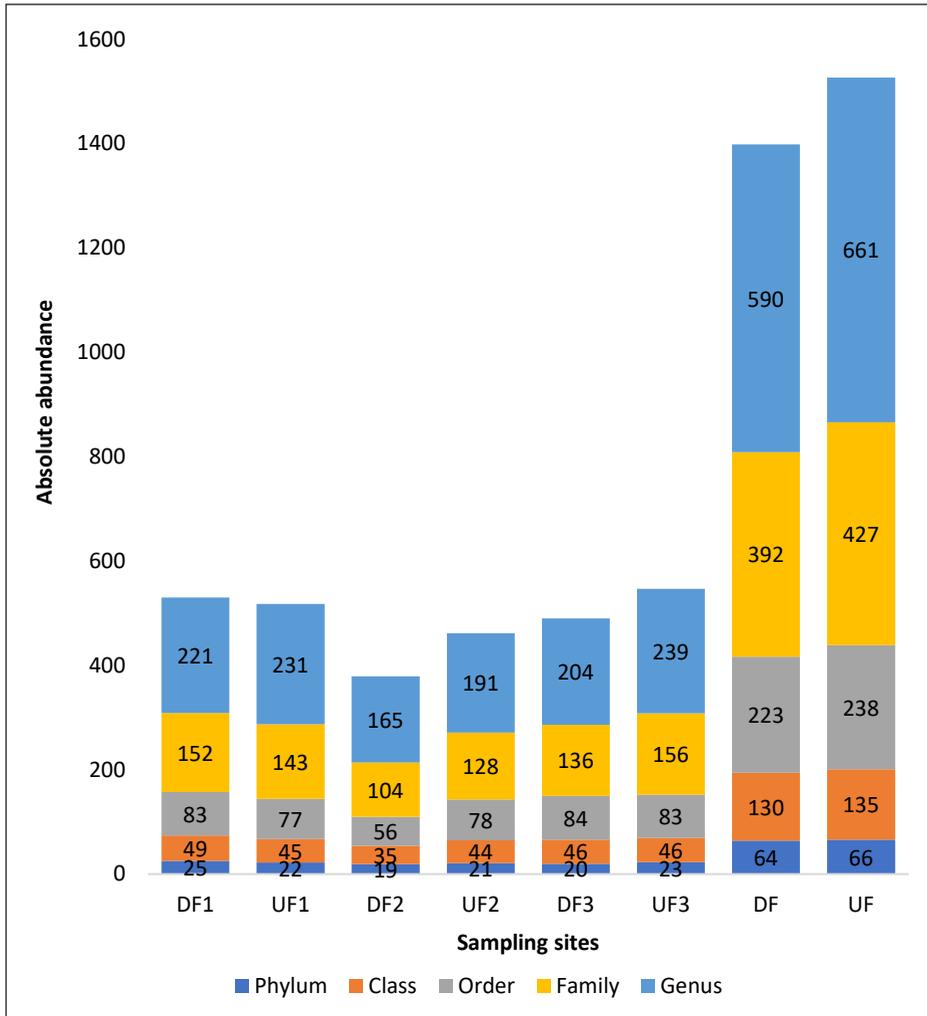
Appendix 2. Bacterial community structures of fish gut contents collected from disturbed and undisturbed area across phylum, class, order, family, and genus displayed in the absolute abundance
Note. DF1: Paddy field; DF2: Forest fire; DF3: Oil palm plantation; UF1: Sungai Karang Peat Swamp Forest site 1; UF2: Raja Musa Peat Swamp Forest; UF3: Sungai Karang Peat Swamp Forest site 2; DF: Disturbed forest area; UF: Undisturbed forest area

Appendix 3
One way-ANOSIM of fish gut bacterial community structures between undisturbed and disturbed area across taxa (phylum, class, order, family and genus)

Permutation N:	9999
Mean rank within:	8.5
Mean rank between:	7.667
R:	-0.1111
p (same):	0.9038

Note. N: Size of the population; R: Size of each sampling group; *p*: *p*-value

Appendix 4
Top 15 % contribution of family level in gastrointestinal contents collected from disturbed and undisturbed areas

Taxon	Av. dissim	Contrib. %	Percentage contribution of taxa					
			DF1	DF2	DF3	UF1	UF2	UF3
Unassigned	19.53	27.03	1.33	2.89	77.00	2.45	70.10	1.22
f_Enterobacteriaceae	17.54	24.28	2.83	52.00	2.76	77.90	2.24	2.69
f_Rikenellaceae	5.913	8.184	23.10	0.00	0.22	0.07	0.54	20.60
f_Ruminococcaceae	5.266	7.288	20.10	0.30	0.15	0.63	2.39	19.10
f_Erysipelotrichaceae	5.075	7.025	19.00	0.04	0.22	0.03	0.23	19.10
f_Clostridiaceae	3.198	4.427	2.89	19.20	5.53	1.93	5.10	3.48
f_Peptostreptococcaceae	1.832	2.535	0.46	10.30	1.12	2.38	1.44	0.52
f_Fusobacteriaceae	1.744	2.414	0.24	5.62	6.15	0.39	6.14	0.30
f_Desulfovibrionaceae	1.466	2.029	5.64	0.11	1.12	0.65	1.66	5.80
f_Bacteroidaceae	1.407	1.948	4.88	0.15	0.00	0.00	0.28	5.49
f_Methylocystaceae	1.339	1.853	5.00	0.12	0.41	0.26	0.41	5.30
f_Streptococcaceae	1.317	1.823	0.13	2.31	0.23	6.70	0.39	0.16
f_Rhodocyclaceae	0.5333	0.7382	1.99	0.09	0.14	0.12	0.13	2.17
f_Acetobacteraceae	0.3631	0.5026	1.30	0.12	0.14	0.08	0.07	1.51
f_Conexibacteraceae	0.2751	0.3807	0.97	0.07	0.02	0.14	0.01	1.08

Note. Overall average dissimilarity = 46.08. DF1: Paddy field; DF2: Forest fire; DF3: Oil palm plantation; UF1: Sungai Karang Peat Swamp Forest site 1; UF2: Raja Musa Peat Swamp Forest; UF3: Sungai Karang Peat Swamp Forest site 2; Av. dissim: Average dissimilarity; Contrib.: Taxa contribution