Tropical mangrove swamp metagenome reveals unusual abundance of ecologically important microbes

Zahidah Ismail[1], Choon-Kook Sam[2], Wai-Fong Yin[1] and Kok-Gan Chan[1,*]

[1]Division of Genetics and Molecular Biology, Institute of Biological Sciences, Faculty of Science, University of Malaya, Kuala Lumpur, Malaysia
[2]Natural Sciences and Science Education AG, National Institute of Education, Nanyang Technological University, 1 Nanyang Walk, Singapore 637616

The tropical ecosystem is extremely rich in flora and fauna, but the tropical mangrove soil microbial diversity remains illusive. To address this knowledge gap, we characterized the metagenome of a Malaysian mangrove soil sample and its microbial ecological roles via next-generation sequencing (NGS). Shotgun NGS data analysis revealed high diversity of ecologically important microbes from bacteria and archaea domains. Also, an unusually high number of archaea was detected together with high abundance of Deltaproteobacteria. In a functional study by the SEED classification in MEGAN, virulence factor genes were abundantly present, implying that the mangrove soil is a potential reservoir of pathogens.

Keywords: Mangrove, metagenomics, pathogens, soil.

MANGROVE forests are usually located at the tropical and subtropical latitudes. They are present at the transition of land and sea, which makes them susceptible to tidal change and salinity. Mangrove soils generally comprise of soft, muddy and anaerobic sediment with thin top layer of aerobic sediment. They also function as heavy-metal sink[1], acting as a natural sink and filtration system. Mangrove swamps are the habitat for a diverse variety of fauna, especially juvenile fishes, and they also act as breeding and nursery grounds for these aquatic animals.

Microorganisms in the mangrove habitat play an important role in maintaining the productivity, conservation and nutrients of this ecosystem. Microorganisms are involved in biogeochemical cycles that supply nutrients to plants and animals[2,3]. Mangroves are rich in organic matter, but usually lack phosphorus and nitrogen[4–7]. Their activities of microorganisms are high because they are efficient in recycling the nutrients contained therein. Microorganisms are directly involved in nitrogen fixation, phosphate solubility, photosynthesis, sulphate reduction and production of other substances.

The mangrove environment is highly susceptible to anthropogenic effects such as pollution, deforestation and human activities. These could change the dynamic mangrove ecosystem, which in turn affects the mangrove community and disturbs the microorganism community that maintains the productivity and conservation of the mangroves.

This study aimed to investigate the metagenome of a mangrove soil sample and its ecological role through metabolic reconstruction. We used the Illumina HiSeq 2000 platform to carry out shotgun metagenome next-generation sequencing (NGS). This method avoids bias of PCR amplification as in the case of amplicon sequencing and enables parallel study on both the taxonomic and functional diversities. We hypothesized the abundance and diversity of microbes and their functional attributes to be similar to those of previous studies[8,9] .

Materials and methods

Sampling was done on a soil sample obtained in the east coast of Peninsular Malaysia, namely Rantau Abang (RA) (04°54.189′N, 103°22.208′E). No specific permissions were required for the chosen locations and such research activities. Our work also did not involve endangered or protected species. The top 5–20 cm of soil was collected and stored at –20°C until processing. A portion of the soil sample was sent for biochemical analyses of its pH, carbon : nitrogen ratio, and contents of phosphorus, sulphur, and heavy metals like arsenic, cadmium, lead and mercury, as described previously[10].

DNA extraction was carried out according to the protocol as described previously[11], with modifications. Traces of plant materials were removed from the soil prior to extraction. Briefly, 5 g of soil was added with 13.5 ml of DNA extraction buffer (Tris-HCl 100 mM, pH 8; EDTA 100 mM, pH 8; Na_2HPO_4 100 mM, pH 7.8; NaCl 1.5 M and CTAB 1% w/v), 100 µl of proteinase K (10 mg/µl), and 200 µl of lysozyme (10 mg/µl). The mixture was incubated horizontally at 37°C with orbital shaking (225 rpm). After 30 min, 0.5 ml of SDS (20% w/v) was added and the mixture further incubated in a 65°C water bath for 2 h with gentle mixing by inverting the tube at

*For correspondence. (e-mail: kokgan@um.edu.my)

15 min intervals. The supernatant was collected by centrifugation at 6000 *g* for 10 min. The pellet was suspended in 4.5 ml of DNA extraction buffer and 0.5 ml of SDS (20% w/v), and vortexed for 10 s followed by incubation at 65°C for 10 min. The supernatant was then collected by centrifugation and pooled with the supernatant collected previously. Equal volume of chloroform : isoamyl alcohol (24 : 1, v/v) was added to the pooled supernatant and the mixture was gently mixed by inversion. The aqueous phase was transferred to a clean, sterile tube after centrifugation at 6000 *g* for 10 min. The chloroform : isoamyl alcohol step was repeated once. For DNA precipitation, 0.6 volume of cold isopropanol was added and the resultant mixture incubated at –20°C for 30 min. DNA was collected by centrifugation at 16,000 *g* for 20 min, followed by washing with 70% (v/v) ethanol and kept at –20°C for 15 min. Ethanol was removed by centrifugation at top speed in a table-top centrifuge for 10 min and the pellet was air-dried aseptically. The DNA pellet was then dissolved in elution buffer (Roche High Pure PCR Product Purification Kit).

The soil metagenomic DNA was further purified by gel elution in a 3% (w/v) low-melting-temperature agarose electrophoresis. The metagenomic DNA was mixed with 80% (v/v) glycerol and 6× loading dye, and the mixture was then loaded into a well. Electrophoresis was carried out at 15 V for 16–20 h. DNA was excised from the gel with a sterile blade and recovered using the Qiagen Gel Extraction Kit (Venlo, The Netherlands). DNA concentration and purity were determined using Qubit and Nanodrop 2000c respectively (Thermo Fischer Scientific, Waltham, MA, USA). The purified DNA was then subject to NGS using Illumina HiSeq 2000.

For taxonomic analysis, the metagenomic nucleotide sequences obtained were trimmed using CLC Bio Genomic Workbench 5.5.2 (Aarhus, Denmark) at 50-nucleotide length to remove short, low-quality reads. The trimmed data were then blasted against the NCBI Microbial database (dated 22 January 2013) using Blastall 2.2.25 (NCBI) at the expected value of 1×10^{-20}.

For functional gene study, the trimmed nucleotide sequences were assembled using *de novo* assembly in CLC Bio Genomic Workbench at the minimum contig length of 400 nucleotides. The assembled data were extracted at coverage of ≥10%. Gene prediction was performed on the extracted sequences using Prodigal 2.60 (ref. 12), and each predicted gene was annotated using RAPsearch 2.09 (against the NCBI NR database; dated 22 January 2013)[13,14]. Data obtained for both taxonomic and functional distributions were analysed in MEGAN 4.70.4 (ref. 15). Taxonomic analysis was done according to the percentage identity filter to get the best sequence match. In MEGAN, functional analysis was accomplished with the SEED classification[16].

The data of the present study are available as NCBI database accession number SRR748204.

Results

Biochemical analyses

Table 1 shows the biochemical properties of the RA soil sample. The pH for the sample was recorded as 5.1.

Metagenomic library analysis

The RA sample metagenome library shows the higher number of contigs and the longer read lengths (Table 2).

Microbial taxonomic distribution

A total of 98% of the reads from RA sample was assigned to the domains level by MEGAN and they excluded the 'No hits' reads category. Majority of the assigned reads from the RA samples were of the domain bacteria while the remaining were of the domain archaea with 78.52% reads assigned to the former and 21.48% to the latter (Figure 1).

There were 27 phyla hits from the domain bacteria for the RA metagenome library. The phylum Proteobacteria dominated other phyla (43.72%) in the RA sample. There were 10 phyla with abundance percentage of more than 1 in the RA sample. In this sample, the phyla detected were Proteobacteria, Acidobacteria (17.68%), Firmicutes

Table 1. Results of biochemical analyses of the soil sample

Parameter	Unit	Rantau Abang (RA) soil sample
Arsenic	mg/kg	ND (<0.5)
Cadmium	mg/kg	0.10
Lead	mg/kg	25.34
Mercury	mg/kg	ND (<0.05)
pH (10% w/w)	–	5.1
Phosphorus	mg/kg	449.70
Sulphur	mg/kg	1374.52
Carbon : Nitrogen	%	5.19:0.31

ND, Not detected.

Table 2. NGS statistics. The reads were generated by Illumina HiSeq 2000. The RA sample showed good quality of reads in terms of number, length and number of contigs generated

Parameters	RA soil sample
Total no. of sequences (bp)	24,227,393,584
Total no. of quality sequences (bp)	20,907,568,942
Average quality read length (bp)	96.81
Total no. of contigs more than 400 bp long	552,541
Total no. of contigs length (bp)	502,888,532
Coverage	41.57
Percentage of hits against NCBI 16S microbial database (%)	0.03
Total CDS assigned to SEED categories (%)	34.71

(13.45), Actinobacteria (4.55%), Nitrospirae (4.22%), Planctomycetes (3.06%), Chloroflexi (2.88%), Verrucomicrobia (2.69%), Spirochaetes (1.70%), Chlamydiae (1.32%) and Bacteroidetes (1.31%) (Figure 2 *a*). In the RA sample, unclassified bacteria were clustered as Caldithrix, Haloplasmatales and some phototrophic bacteria.

Forty-three classes of bacteria were detected in the RA sample (Figure 2 *b*), with the two most abundant classes being Deltaproteobacteria (19.29%) and Alphaproteobacteria (16.89%), followed by Acidobacteria (16.61%) (Figure 3 *b*). The other minor classes included Clostridia (9.32%), Gammaproteobacteria (5.50%), Actinobacteria (4.58%), Nitrospira (4.24%), Bacilli (3.24%), Planctomycetia (3.08%), Spirochaetia (1.71%), Betaproteobacteria (1.69%), Opitutae (1.52%), Chlamydia (1.33%), Verrucomicrobiae (1.19%), Holophagae (1.17%), Anaerolineae (1.06%) and Ktedonobacteria (1.02%).

At the genus level, *Acidobacterium* of the Acidobacteria phylum was the dominant genus in the RA sample. The abundance frequency of this genus was 10.01% (Figure 2 *c*).

Five classes of Proteobacteria, namely Alphaproteobacteria, Betaproteobacteria, Deltaproteobacteria, Gammaproteobacteria and Epsilonproteobacteria were detected in the RA sample. Among these, Deltaproteobacteria (43.88%) was the major class, followed by Alphaproteobacteria (38.43%), Gammaproteobacteria (12.51%), Betaproteobacteria (RA 3.85%) and Epsilonproteobacteria (RA 1.33%) (Figure 3). The segregation of orders within Deltaproteobacteria showed that Syntrophobacterales was the most abundant order in this soil sample (Figure 4). At the genus level of this order, *Syntrophobacter* was the most abundant genus (Figure 5).

The RA sample showed the presence of archaea, but only phyla Crenarchaeota and Euyarchaeota were detected. In the RA sample, Crenarchaeota (63.78%) was present at a higher percentage compared to Euyarchaeota (36.22%) (Figure 6). A total of eight classes of archaea were detected in both soil samples. Among them, Thermoprotei (RA 63.78%) and Methanomicrobia (RA 17.85%) were the two dominant classes. Other minor archaea classes present in the RA sample were Thermococci (5.48%), Methanococci (5.35%), Thermoplasmata (3.52%), Methanobacteria (2.17%), Archaeoglobi (1.36%) and Halobacteria (0.47%).

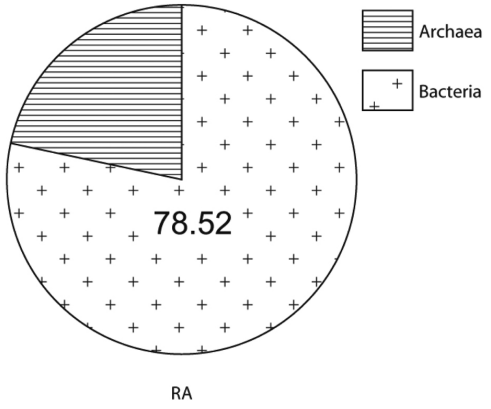

Figure 1. Percentage of assigned reads to domains bacteria and archaea. Majority of the reads were assigned to the domain bacteria by MEGAN. The number of reads detected for archea was significantly higher than previously reported.

The taxonomic diversity for the domains bacteria and archaea was estimated at the genus level using the Shannon–Weaver diversity index, H' in MEGAN; the H' value the RA sample was 7.765.

Metabolic functional analysis via reconstruction of metagenome library

The gene anthology was derived from the SEED classification. Using this approach, the most abundant gene detected in the RA sample was associated with carbohydrate metabolism (12.97%). The second most abundant genes in the RA sample were associated with protein metabolism (9.89%), virulence (9.53%), respiration (8.39%), and amino acids and their derivatives (8.19%) (Figure 7).

Discussion

The application of NGS enables the study of microbial diversity and function in metagenomes without the need of culturing bacteria, thus bypassing the growing of fastidious bacteria which are often unculturable on laboratory media. However, this method depends on the reliability of the NGS data generated.

Most bacteria detected in the RA metagenomes were either anaerobic or facultative anaerobic. They were predominantly from the domain bacteria. However, in the RA metagenome, a significant number of bacteria belonging to the domain archaea were detected, and the percentage of archaeal abundance (21.48) was significantly higher than those reported for other soil metagenomes[17]. Crenarchaeota is known to be present in environments with high sulphur content, and for its ability to utilize sulphur[18]. The high percentage of this phylum detected in the RA sample is consistent with the high concentration of sulphur in the RA sample.

The bacterial diversity detected in the RA sample conforms to the common bacteria present in other types of soil in other geographical locations[19]. In the RA sample, Proteobacteria was the most abundant bacteria, it comprised of five different classes. However, the distribution of Proteobacteria classes in this study differs from those reported for other mangrove habitats[8,20]. In contrast to other mangrove metagenomes reported to date, both the Malaysian mangrove metagenomes possessed Deltaproteobacteria as the dominant class of Proteobacteria. Even though Proteobacteria was the dominant phylum in the RA metagenome, *Acidobacterium* of the Acidobacteria phylum was the most abundant genus in the soil sample.

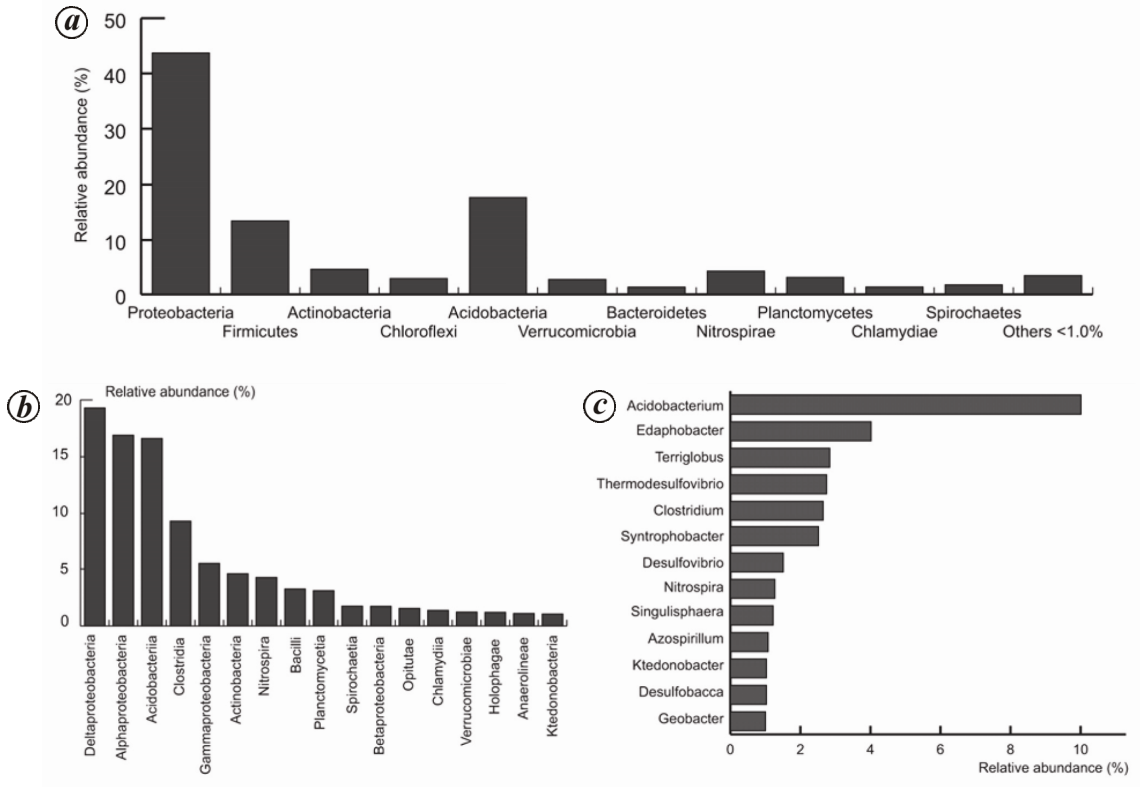

Figure 2. ***a***, Segregation of phyla in the domain bacteria. Proteobacteria was the dominant phylum in the samples with almost half of the reads assigned to it. ***b***, Classes of bacteria present in the Rantau Abang (RA) sample at more than 1%. ***c***, bacterial genera with more than 1% reads. Bacterial genera detected in RA sample in which *Acidobacterium* was the most abundant genus.

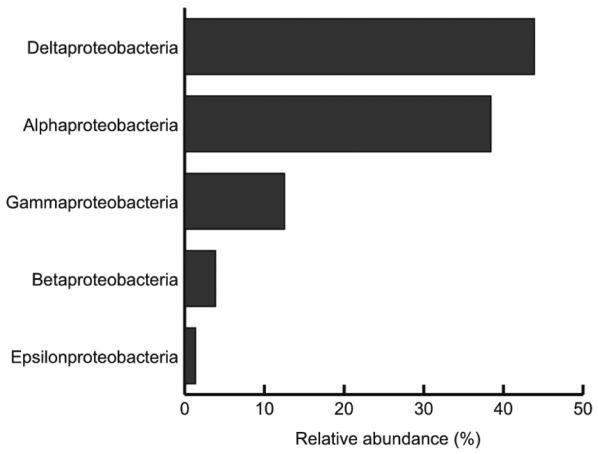

Figure 3. Classes of proteobacteria. Deltaproteobacteria was the most abundant class in RA samples.

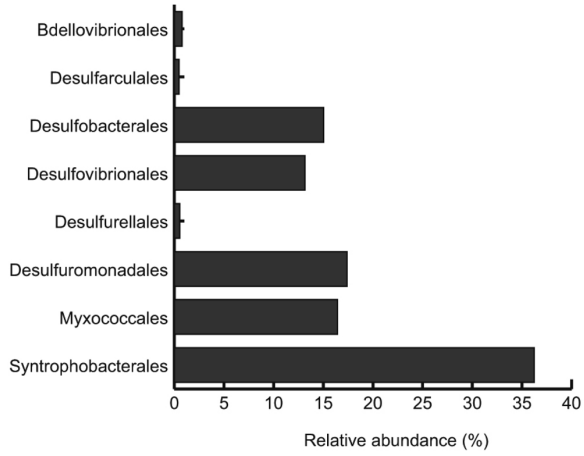

Figure 4. Order level of deltaproteobacteria. Syntrophobacterales were the most abundant order.

The presence of the high frequency of genes associated with carbohydrate metabolism in this soil metagenome analysis is not surprising because these genes are commonly detected in abundance in most studies of soil metagenomes[8,21]. However, the presence of the high frequency of virulence factor genes in the soil metagenomes is unusual, because they are not commonly reported[8,22]. This leads to the speculation that mangrove soil is a potential reservoir of pathogenic bacteria, but further work is necessary to verify this finding.

In the RA soil metagenome, antibiotic and toxic compound resistance genes were also detected frequently. Their abundance may be related to the high percentage of Actinobacteria, which is known to produce a myriad of antibacterial compounds, and Deltaproteobacteria, members of which are known to be resistant to heavy metals and to oxidize heavy metals to their benign forms[23], in the two tropical mangrove soil samples. We also obtained a high hit rate of the stress response gene in the soil metagenomic libraries, suggesting that the tropical

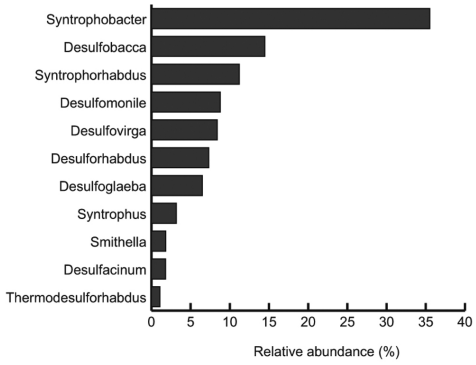

Figure 5. Genus level in Syntrophobacterales. *Syntrophobacter* was the most abundant genus.

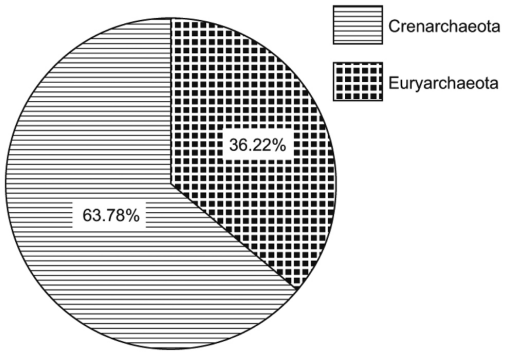

Figure 6. Percentage distribution of archaeal phyla in the sample. RA sample had high percentage of Crenarchaeota.

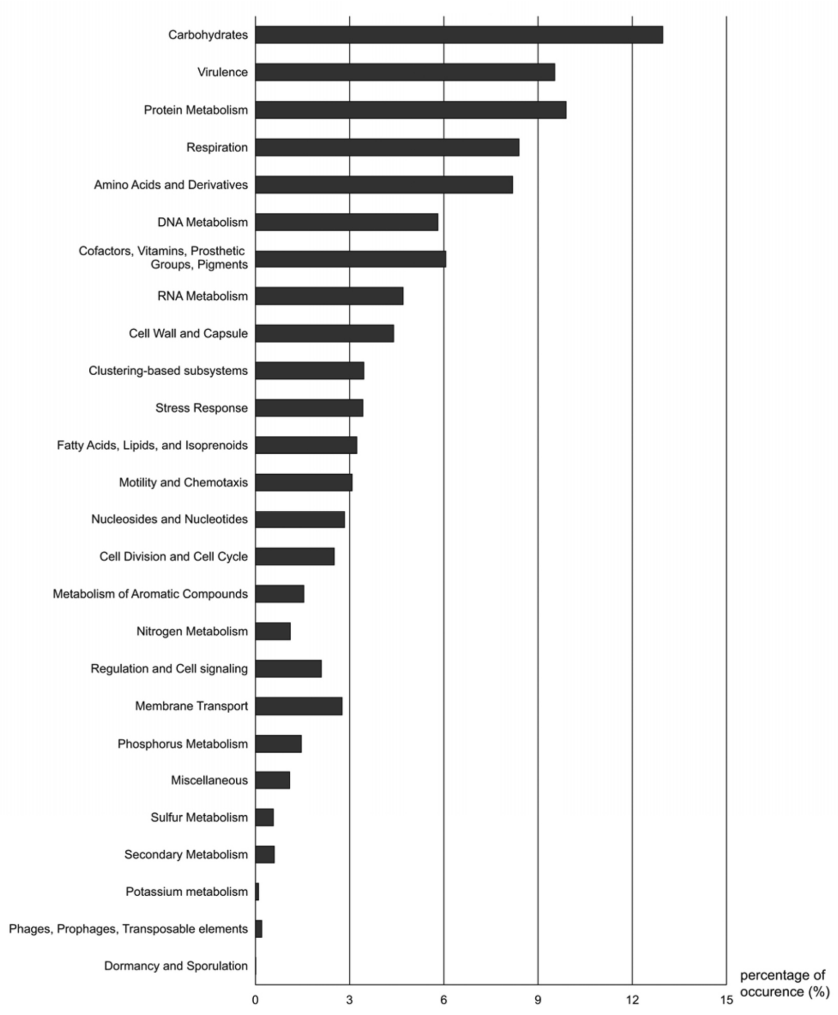

Figure 7. Gene ontology. The SEED classification by MEGAN.

mangrove environment is harsh for microorganisms. This is most probably due to polluted marine waters and high salinity and low aeration available in the muddy mangrove soil.

The biochemical tests showed considerably high amounts of phosphorus and sulphur in the soil sample. However, the gene ontology analysis revealed that the genes associated with the metabolism of compounds containing phosphorus and sulphur were of relatively low abundance, suggesting that these compounds may be mainly involved in redox reactions in electron transport, but not in microbial metabolism. This also implies that

sulphur and phosphorus compounds exist in stable forms as mangroves are sink for inorganic compounds.

Although Actinobacteria and Firmicutes were two of the detected sub-dominant phyla, the present analysis showed low frequency of spore-producing bacteria such as *Bacillus* and *Streptomyces*. This may explain why the genes for sporulation and dormancy were not detected in the soil metagenomic DNA.

Conclusion

This study demonstrates the high level of microbial diversity in mangrove swamps compared to the limited vegetation that is able to survive in this environment. The high abundance of members of Deltaproteobacteria, and heavy metal and toxic compound resistance genes indicates that microorganisms have the potential for bioremediation of heavy metals. The differences in the distribution of microorganisms compared with previous studies on mangrove soils are most likely due to the different geographical locations. To the best of our knowledge, there have been no earlier studies on microbial diversity of mangrove soil in Malaysia, using the NGS metagenomic approach. More mangrove soil samples collected from different locations in Malaysia need to be analysed by this approach before a more defined conclusion on the microbiome of mangrove soils and its functional genes can be reached.

1. Tama, N. F. Y. and Wong, Y. S., Retention and distribution of heavy metals in mangrove soils receiving wastewater. *Environ. Pollut.*, 1996, **94**(3), 283–291.
2. Holguin, G., Vazquez, P. and Bashan, Y., The role of sediment microorganisms in the productivity, conservation, and rehabilitation of mangrove ecosystems: an overview. *Biol. Fertil. Soils*, 2001, **33**, 265–278.
3. Sahoo, K. and Dhal, N. K., Potential microbial diversity in mangrove ecosystems: a review. *Indian J. Mar. Sci.*, 2009, **38**(2), 249–256.
4. Sengupta, A. and Chaudhuri, S., Ecology of heterotrophic denitrogen fixation in the rhizosphere of mangrove plant community at the Ganges river estuary in India. *Oecologia*, 1991, **87**, 560–564.
5. Holguin, G., Guzman, M. A. and Bashan, Y., Two new nitrogen-fixing bacteria from the rhizosphere of mangrove trees: their isolation, identification and *in vitro* interaction with rhizosphere *Staphylococcus* sp. *FEMS Microbiol. Ecol.*, 1992, **101**, 207–216.
6. Alongi, D. M., Christoffersen, P. and Tirendi, F., The influence of forest type on microbial nutrient relationships in tropical mangrove sediments. *J. Exp. Mar. Biol. Ecol.*, 1993, **171**(2), 201–223.
7. Vazquez, P., Holguin, G., Puente, M. E., Lopez-Cortes, A. and Bashan, Y., Phosphate-solubilizing microorganisms associated with the rhizosphere of mangroves in semiarid coastal lagoon. *Biol. Fertil. Soils*, 2000, **30**, 460–468.
8. Andreote, F. D. *et al.*, The microbiome of Brazilian mangrove sediments as revealed by metagenomics. *PLoS ONE*, 2012; doi: 10.1371/journal.pone.0038600.
9. Xu, Z., Hansen, M. A., Hansen, L. H., Jacquiod, S. and Sørensen, S. J., Bioinformatic approaches reveal metagenomic characterization of soil microbial community. *PLoS ONE*, 2014, **9**(4), e93445; doi:10.1371/journal.pone.0093445.
10. Paul, E. A., *Soil Microbiology, Ecology and Biochemistry*, Elsevier Inc., UK, 2007, 3rd edn, pp. 389–430.
11. Zhou, J., Bruns, M. A. and Tiedje, J. M., DNA recovery from soils of diverse composition. *Appl. Environ. Microbiol.*, 1996, **62**(2), 316–322.
12. Hyatt, D., LoCascio, P. F., Hauser, L. J. and Uberbacher, E. C., Gene and translation initiation site prediction in metagenomic sequences. *Bioinformatics*, 2012, **28**(17), 2223–2230.
13. Ye, Y, Choi, J. H. and Tang, H., RAPSearch: a fast protein similarity search tool for short reads. *BMC Bioinfor.*, 2011, doi:10.1186/1471-2105-12-159.
14. Zhao, Y., Tang, H. and Ye, Y., RAPSearch2: a fast and memory-efficient protein similarity search tool for next generation sequencing data. *Bioinformatics*, 2012, **28**(1), 125–126.
15. Huson, D. H., Mitra, S., Weber, N., Ruscheweyh, H. J., Weber, N. and Schuster, S. C., Integrative analysis of environmental sequences using MEGAN4. *Genome Res.*, 2011, **21**, 1552–1560.
16. Disz, T. *et al.*, Accessing the SEED genome databases via web services API: tools for programmers. *BMC Bioinfor.*, 2010; doi:10.1186/1471-2105-11-319.
17. Bates, S. T., Berg-Lyons, D., Caporaso, J. G., Walters, W. A., Knightm R. and Fierer, N., Examining the global distribution of dominant archaeal populations in soil. *ISME J.*, 2011, **5**(5), 908–917.
18. Hou, W. *et al.*, A comprehensive census of microbial diversity in hot springs of Tengchong, Yunnan Province China using 16S RNA gene pyrosequencing. *PLoS ONE*, 2013; doi: 10.1371/journal.pone.0053350.
19. Janssen, P. H., Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. *Appl. Environ. Microbiol.*, 2006, **72**(3), 1719–1728.
20. dos Santos, H. F. *et al.*, Mangrove bacterial diversity and the impact of oil contamination revealed by pyrosequencing: bacterial proxies for oil pollution. *PLoS ONE*, 2011; doi:10.1371/journal.pone.0016943.
21. Jeffries, T. C. *et al*., Substrate type determines metagenomic profiles from diverse chemical habitats. *PLoS ONE*, 2011, **6**(9), e25173; doi:10.1371/journal.pone.0025173.
22. Ghosh, A., Dey, N., Bera, A., Tiwari, A., Sathyaniranjan, K. B., Chakrabarti, K. and Chattopadhyay, D., Culture independent molecular analysis of bacterial communities in the mangrove sediment of Sundarban, India. *Saline Syst.*, 2010, **6**, 1; doi: 10.1186/1746-1448-6-1.
23. Blöthe, M., Akob, D. M., Kostka, J. E., Göschel, K., Drake, H. L. and Küsel, K., pH gradient-induced heterogeneity of Fe(III)-reducing microorganisms in coal mining-associated lake sediments. *Appl. Environ. Microbiol.*, 2008, **74**(4), 1019–1029.

ACKNOWLEDGEMENTS. K.-G.C. thanks the University of Malaya, Malaysia for the High Impact Research Grants (UM.C/625/1/HIR/MOHE/CHAN/01, Grant No. A-000001-50001 and UM-MOHE HIR Grant UM.C/625/1/HIR/MOHE/CHAN/14/1, no. H-50001-A000027).

Received 2 May 2015; revised accepted 12 November 2016

doi: 10.18520/cs/v112/i08/1698-1703