MJFAS
Malaysian Journal of Fundamental and Applied Sciences

Salleh and Esa *et al.* | Malaysian Journal of Fundamental and Applied Sciences, Vol. 19 (2023) 1-5

DATASETS ARTICLE

The First Next-Generation Sequencing Metabarcoding Dataset on Faecal Bacterial Diversity from the Southern River Terrapin, *Batagur affinis* ssp.

Mohd Hairul Mohd Salleh[a,b], Yuzine Esa[a,c*]

[a]Department of Aquaculture, Faculty of Agriculture, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia; [b]Royal Malaysian Customs Department, Persiaran Perdana, Presint 2, 62596 Putrajaya, Malaysia; [c] International Institute of Aquaculture and Aquatic Sciences, Universiti Putra Malaysia, Lot 960 Jalan Kemang 6, 71050 Port Dickson, Negeri Sembilan, Malaysia

Abstract The Southern River terrapin, *Batagur affinis* ssp., has the first data on faecal bacterial diversity from next-generation sequencing (NGS). This dataset describes the bacterial diversity of the Southern River terrapin, locally known as Tuntung. *Batagur affinis* spp. are freshwater turtles listed as critically endangered on the International Union for Conservation of Nature (IUCN) Red List since 2000. This is the first dataset on the faecal bacterial diversity of *Batagur affinis* ssp., and the data provided here can be used to comprehensively understand the microbiome's community composition. Seven faeces samples were collected aseptically from captive (*N* = 5) and wild (*N* = 2) adult *B. affinis* ssp. while crossing Peninsular Malaysia's east and west coasts. The data was acquired by metabarcoding using 16S rRNA. The amplicons were further analysed using the SILVA and DADA2 pipelines. The V3-V4 of the 16S rRNA gene region was amplified, and the amplicons were sequenced on the Illumina MiSeq system. In total, 297 bacterial communities' taxonomic profiles (phylum to genus) have been determined. The data for this metagenome can be found in the BioSample Submission Portal as Bio-Project PRJNA767629 and Sequence Read Archive (SRA) accession numbers from SAMN21919713 to SAMN21919722.

Keywords: NGS metabarcoding dataset, Tuntung, bacteria, 16S rRNA, V3-V4 gene region.

***For correspondence:**
yuzine@upm.edu.my

Received: 12 Nov. 2022
Accepted: 28 Dec. 2022

Introduction

The sequencing of the 16S rRNA gene amplicon for microbial sequencing is important in environmental and medicinal studies [1]. Every bacterium (prokaryote) contains the 16S rRNA gene, with a variable evolution rate for each species. Therefore, next-generation sequencing (NGS) of microorganisms can be used to determine the type of bacteria [2]. Also, studies of microbial communities in faeces samples using DNA metabarcoding have been described as a non-invasive, accurate, and time- and cost-efficient way to find host-associated microbial communities that play important roles in hosts' health [3].

Thus, due to the advances in molecular microbial community identification techniques, exploring two *B. affinis* subspecies (ssp.) faeces samples in terms of bacterial community could enhance the understanding of gut microbiome patterns and their potential roles in *B. affinis* ssp. Few examinations have inspected freshwater turtle microbiomes, especially in *B. affinis* ssp. Most investigations focus on sea turtles' microbiomes [4-7]. Many studies have suggested that faecal DNA metabarcoding could be an appealing way to interact with microbial communities [7-10]. Also, this technique has been commonly

used to study the diets of various animals [11, 12]. The symbiotic bacterial community patterns of *B. afiinis* ssp. maybe useful for long-term conservation purposes, according to one possible contributing factor of faecal DNA metabarcoding. Therefore, this work aims to characterise the gut microbiota of the Southern River terrapin, *B. affinis* ssp. and gain a comprehensive understanding of the community composition of the microbiome. Faeces samples data were taken to have an extensive indication of gut microbiota taxonomic configuration.

Materials and Methods

Sampling collection and sequence production

The faecal microbial community structure of adult *B. affinis* ssp. populations from the East and West coasts of Peninsular Malaysia was studied. A total of seven samples of individuals were analysed (Table 1). The faeces samples of adult Southern River terrapins ($N = 5$) were collected from a captive population at the Bota Kanan (BK), Perak (4.3489° N, 100.8802° E) in 2020. Meanwhile, an adult Southern River terrapin ($N = 2$) was collected from a wild population located at Bukit Paloh, Kuala Berang (KB), Terengganu (5.0939° N, 102.7821° E) in 2021. The samples consisted of five *B. affinis affinis* (BK27, BK28, BK29, BK30, and BK31) samples and two *B. affinis edwarmolli* (KBW2 and KBW3) samples. Furthermore, KBW2 and KBW3 were faeces from two different wild river terrapin mothers. The microbial community in the faeces sample was sorted and identified using standard taxonomic keys [13]. Briefly, samples were collected and transferred using a sterile spatula into a sterile 50-ml Falcon tube and stored on ice during transportation to the laboratory. All samples were immediately stored at −20◦C before the extraction of DNA. The field permit approval number is B-00335-16-20, issued by the Department of Wildlife and Parks, Peninsular Malaysia.

Table 1. Sample details and sequencing statistics were obtained from *B. affinis* ssp. faeces samples using the DADA2 pipeline.

Sample Code	Species Name	Sample Source	Sex	Sampling Site	Sampling Date	Raw Reads	Reads Passing Quality Filtering	Reads Passing Quality Filtering (%)
BK27	*B. a. affinis*	Captivity	Undetermined	Bota Kanan, Perak	2/4/2020	60000	40124	66.87
BK28	*B. a. affinis*	Captivity	Undetermined	Bota Kanan, Perak	2/4/2020	60000	40586	67.64
BK29	*B. a. affinis*	Captivity	Undetermined	Bota Kanan, Perak	2/4/2020	60000	42151	70.25
BK30	*B. a. affinis*	Captivity	Undetermined	Bota Kanan, Perak	2/4/2020	60000	40846	68.08
BK31	*B. a. affinis*	Captivity	Undetermined	Bota Kanan, Perak	2/4/2020	60000	41411	69.02
KBW2	*B. a. edwardmolli*	Wild	Female	Kuala Berang, Terengganu	15/3/2021	60000	35603	59.34
KBW3	*B. a. edwardmolli*	Wild	Female	Kuala Berang, Terengganu	15/3/2021	60000	38602	64.34

DNA extraction and metagenome sequencing

The NucleoSpin® Soil Kit (Macherey-Nagel, Germany) is commonly used to extract DNA from the soil. However, in this study, it was used to extract DNA from the faeces sample. Briefly, from the 300 mg input volume of the faeces sample, a final extraction volume of 50 µl of DNA sample was achieved and further stored at -20 °C. A 1% (w/v) agarose gel electrophoresis was used to test the reliability of purified DNA. The DNA concentration was measured using a spectrophotometer (Implen NanoPhotometer® N60/N50, Germany) and fluorometric quantification using an iQuant™ Broad Range dsDNA Quantification Kit (GeneCopoeia, Inc., USA).

Two purified gDNA samples that passed DNA sample quality control and were amplified with locus-specific primers for bacterial 16S V3-V4 (5' to 3') were used for the 16S rRNA amplicon PCR quality control. To amplify (the V3-V4 region) of the 16S rRNA gene, the primer pair 16S V3-V4 Forward: CCTACGGGNGGCWGCAG and 16S V3-V4 Reverse: GACTACHVGGGTATCTAATCC was used [14].

The area of 16S rRNA was sequenced using the pair-end (PE) Illumina MiSeq platform, which provides raw reads of 300 bp. BBDuk (version 39.92) [15] removed sequence adapters and low-quality reads from the raw reads. QIIME2 (version 2019.10) [16] aligns and integrates the raw readings. The forward and reverse reads were then combined using QIIME2 [17, 18]. The Divisive Amplicon Denoising Algorithm 2 (DADA2) pipeline (version 1.14) [19,20] was used to denoise in an attempt to remove and/or correct incorrect reads, low-quality areas, and chimeric errors to provide Amplicon Sequence Variants (ASV) data [21]. In this study, the DADA2 pipeline was used as the standard ASV method [19].

Value of the Data

This is the first data release on the bacterial diversity of faeces from *Batagur affinis* ssp. Establishing a dataset of the bacterial diversity found in *B. affinis* ssp. faeces can be viewed as the first step toward sustainability. The data supplied here could be used to conduct comparative studies on the microbiota of vertebrates. Additionally, the data presented here can be used to understand host-microbiome interactions better and to determine whether the bacterial pathogen *B. affinis* ssp. is a threat to the critically endangered species.

Registration and availability of data

The data for this metagenome can be found on the BioSample Submission Portal under the BioProject PRJNA767629, and the Sequence Read Archive (SRA) accession numbers SAMN21919713–SAMN21919722. Available online at GenBank, https://www.ncbi.nlm.nih.gov/sra/PRJNA767629.

Data interpretation

This data report defines the bacterial diversity of faeces from *Batagur affinis* ssp., is a freshwater turtle listed as critically endangered on the IUCN Red List since 2000 [22] and among 24 species of turtles in Malaysia [23]. [24] stated that 60% of threats to *B. affinis* ssp. were caused by the exploitation of local eggs and meat consumption. However, diseases caused by microbiota may be a silent killer of *B. affinis* ssp., the leading cause this study tries to work out.

The 16S metabarcoding analysis is a powerful tool for analysing bacteria in a fraction of the time required by a traditional culturomic experiment, which needs bacteria to be isolated prior to characterisation [25]. The datasets presented here were obtained using Illumina MiSeq sequencing of the 16S rRNA V3-V4 region. The obtained data were further analysed using the SILVA and DADA2 pipelines.

As illustrated in Supplementary Figure 1, *B. affinis* ssp. faeces taxonomic composition contains 21 phyla, 28 classes, 39 orders, 70 families, and 140 genera. This study of the faecal microbiota of *B. affinis* ssp. revealed a taxonomic configuration that includes well-known members of the vertebrate gut microbiota. In particular, the two most abundant phyla found (Firmicutes and Bacteroidetes) are also plentiful in the human gut [26,27] as well as in other land vertebrates and reptiles [27-29].

The dataset had a total of 21 bacterial phyla, as illustrated in Figure 1a. Firmicutes (39%), Bacteroidetes (24%), and Proteobacteria (15%) were the most common phyla in this faecal sample. At the genus level of taxonomic determination, Cetobacterium (18%), Clostridium (13%), and Akkermansia (11%) were dominant. 22% of ASV lacked significant hits against the taxonomic database and were thus classified as uncultured. Figure 1b depicts the top ten genera, including uncultured ones; the complete list of genera is also identified.

This study aimed to generate genomic data resources for faecal bacteria associated with *B. affinis affinis* and *B. affinis edwardmolli* in Peninsular Malaysia using metabarcoding techniques. The Firmicutes phylum's dominating members were chosen as representatives of bacteria. Our genomic dataset analysis of the *Cetobacterium* and *Clostridium* genera revealed here will stimulate future studies by providing a fresh lens to investigate their lifestyles and aid in deciphering novel biological insights into bacteria-host relationships. Furthermore, as the first study on the faecal DNA metabarcoding of two *B. affinis* subspecies, this is the starting point for further research into gut microbial community patterns and their potential roles in *B. affinis* ssp. health and disease development may lead to possible population extinction in some cases. Thus, this will help us in the future conservation of *B. affinis* ssp. towards sustainability.

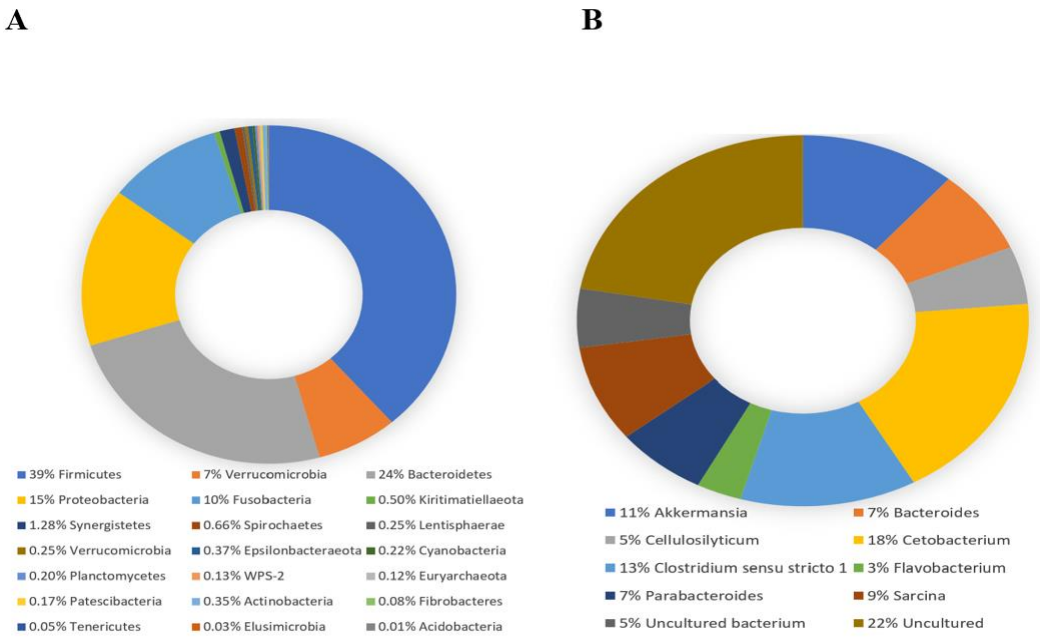

Figure 1. Shows the taxonomic composition of the *B. affinis* ssp. faecal microbiota at various taxonomic levels. (A) Taxonomic classification of ASVs at phylum level for the *Batagur affinis* ssp. faeces sample; (B) Taxonomic classification of ASVs at the genus level for the *Batagur affinis* ssp. faeces sample. Only the top ten genera are summarised here.

Conflicts of Interest

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper.

Acknowledgement

We would also like to express our gratitude to the Breeding Genetics Laboratory, Department of Aquaculture, Faculty of Agriculture, Universiti Putra Malaysia, for the facilities and support of this research project. Furthermore, we thank the Turtle Conservation Society of Malaysia and the Department of Wildlife and National Parks, Peninsular Malaysia, for their collaboration in this study. Finally, we thank our anonymous reviewers for their insights and suggestions to improve this paper.

References

[1] Baird, D. J., Hajibabeil, M. (2012). Biomonitoring 2.0: A new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. *Molecular Ecology, 21*, 2039-2044. https://doi.org/10.1111/j.1365-294X.2012.05519.x.

[2] Winand, R., Bogaerts, B., Hoffman, S., Lefevre, L., Delvoye, M., Van Braekel, J., Fu, Q., Roosens, N. H., De Keersmaecker, S. C. & Vanneste, K. (2020). Targeting the 16S rRNA gene for bacterial identification in complex mixed samples: comparative evaluation of second (Illumina) and third (Oxford Nanopore Technologies) generation sequencing technologies. *International Journal of Molecular Sciences, 21*, 298. https://doi.org/10.3390/ijms21010298.

[3] Ando, H., Mukai, H., Komura, T., Dewi, T., Ando, M., Isagi, Y. (2020). Methodological trends and perspectives of animal dietary studies by non-invasive fecal DNA metabarcoding. *Environmental DNA, 2*(4), 391-406. https://doi.org/10.1002/edn3.117.

[4] Ahasan, M. S., Kinobe, R., Elliott, L., Owens, L., Scott, J. (2019). Bacteriophage versus antibiotic therapy on gut bacterial communities of juvenile green turtle, *Chelonia mydas*. *Environ Microbiol*. https://doi.org/10.1111/1462-2920.14644.

[5] Arizza, V., Vecchioni, L., Caracappa, S., Sciurba, G., Berlinghieri, F. (2019). New insights into the gut microbiome in loggerhead sea turtles *Caretta caretta* stranded on the Mediterranean coast. *PLoS One.* 14:e0220329. https://doi.org/10.1371/journal.pone.0220329.

[6] Biagi, E., D'Amico, F., Soverini, M., Angelini, V., Barone, M. (2019). Fecal bacterial communities from Mediterranean loggerhead sea turtles (*Caretta caretta*). *Environ Microbiol Rep, 11*, 361-371. https://doi.org/10.1111/1758-2229.12683.

[7] Mohd Salleh, M. H., Esa, Y., Ngalimat, M. S., Chen, P. N. (2022). Faecal DNA metabarcoding reveals novel bacterial community patterns of critically endangered Southern River Terrapin, *Batagur affinis*. *PeerJ, 10*, e12970. https://doi.org/10.7717/peerj.12970.

[8] Valentini, A., Pompanon, F., Taberlet, P. (2009). DNA barcoding for ecologists. *Trends in Ecology and Evolution, 24*, 110-117. https://doi.org/10.1016/j.tree.2008.09.011.

[9] Pompanon, F., Deagle, B. E., Symondson, W. O. C., Brown, D. S., Jarman, S. N., Taberlet, P. (2012). Who is eating what: Diet assessment using next generation sequencing? *Molecular Ecology Resources*, *21*(8), 1931-1950. https://doi.org/10.1111/j.1365-294X.2011.05403.x.

[10] Ducotterd, C., Crovadore, J., Lefort, F., Rubin, J. F., Ursenbacher, S. (2021). A powerful long metabarcoding method for the determination of complex diets from fecal analysis of the European pond turtle (*Emys orbicularis*, L. 1758). *Molecular Ecology Resources, 21*(2), 433-447. https://doi.org/10.1111/1755-0998.13277.

[11] Goldberg, A. R., Conway, C. J., Tank, D. C., Andrews, K. R., Gour, D. S, Waits, L. P. (2020). Diet of a rare herbivore based on DNA metabarcoding of feces: Selection, seasonality, and survival. *Ecology and Evolution*, *10*(14), 7627-7643. https://doi.org/10.1002/ece3.6488.

[12] Ingala, M. R., Simmons, N. B., Wultsch, C., Krampis, K., Provost, K. L., Perkins, S. L. (2021). Molecular diet analysis of neotropical bats based on fecal DNA metabarcoding. *Ecology and Evolution, 11*, 7474-7491. DOI:10.1002/ece3.7579.

[13] Zemb, O., Achard, C. S., Hamelin, J., De Almeida, M. L., Gabinaud, B., Cauquil, L., Verschuren, L. M. & Godon, J. J. 2020. Absolute quantitation of microbes using 16S rRNA gene metabarcoding: A rapid normalisation of relative abundances by quantitative PCR targeting a 16S rRNA gene spike-in standard. *Microbiologyopen, 9*(3), p.e977. https://doi.org/10.1002/mbo3.977.

[14] Zhang, J., Ding, X., Guan, R., Zhu, C., Xu, C., Zhu, B., Zhang, H., Xiong, Z., Xue, Y., Tu, J. & Lu, Z. (2018). Evaluation of different 16S rRNA gene V regions for exploring bacterial diversity in a eutrophic freshwater lake. *Science of the Total Environment, 618*, 1254-1267. https://doi.org/10.1016/j.scitotenv.2017.09.228.

[15] Bushnell, B. (2018). *BBduk version 39.92*. [software]. Available from: https://sourceforge.net/projects/bbmap/. Retrieved on 14/05/2021.

[16] Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology, 37*(8), 852-857. https://doi.org/10.1038/s41587-019-0209-9.

[17] Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E.K., Fierer, N., Peña, A. G., Goodrich, J. K., Gordon, J. I. & Huttley, G. A. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods, 7*(5), 335-336. https://doi.org/10.1038/nmeth.f.303.

[18] Lawley, B., Tannock, G. W. (2017). *Analysis of 16S rRNA gene amplicon sequences using the QIIME software package. In Oral Biology* (pp. 153-163). Humana Press, New York, NY.

[19] Callahan, B. J, McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J., Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods, 13*(7), 581-583. https://doi.org/10.1038/nmeth.3869.

[20] Callahan, B., McMurdie, P., Holmes, S. (2019). DADA2 pipeline tutorial (1.8). Retrieved May 26, 2020, from https://benjjneb.github.io/dada2/tutorial_1_8.html.

[21] Nearing, J. T., Douglas, G. M., Comeau, A. M., Langille, M. G. I. (2018). Denoising the denoisers: an independent evaluation of microbiome sequence error-correction approaches. *PeerJ, 6*, e5364. https://doi.org/10.7717/peerj.5364.

[22] International Union for Conservation of Nature, International Union for Conservation of Nature, Natural Resources. Species Survival Commission, & IUCN Species Survival Commission. (2001). *IUCN Red List categories and criteria*. IUCN.

[23] Mohd Salleh, M. H., Esa, Y., Salleh, S. M., & Mohd Sah, S. A. (2022). Turtles in Malaysia: A Review of Conservation Status and a Call for Research. *Animals*, *12*(17), 2184. http://dx.doi.org/10.3390/ani12172184.

[24] Stanford, C. B., Iverson, J. B., Rhodin, A. G., van Dijk, P. P., Mittermeier, R. A., Kuchling, G., Berry, K. H., Bertolero, A., Bjorndal, K. A., Blanck, T. E. & Buhlmann, K. A. (2020). Turtles and tortoises are in trouble. *Current Biology, 30*(12), R721-R735. https://doi.org/10.1016/j.cub.2020.04.088.

[25] Hugon, P., Lagier, J. C., Colson, P., Bittar, F., Raoult, D. (2017). Repertoire of human gut microbes. *Microbial Pathogenesis, 106*, 103-112. https://doi.org/10.1016/j.micpath.2016.06.020.

[26] Ley, R. E., Lozupone, C. A., Hamady, M., Knight, R., Gordon, JI. (2008). Worlds within worlds: evolution of the vertebrate gut microbiota. *Nature Reviews Microbiology, 6*, 776-788. DOI: 10.1038/nrmicro1978.

[27] Abdelrhman, K. F. A., Bacci, G., Mancusi, C., Mengoni, A., Serena, F., Ugolini, A. (2016). A First Insight into the Gut Microbiota of the Sea Turtle *Caretta caretta*. *Frontiers Microbiology, 7*, 1060. DOI: 10.3389/fmicb.2016.01060.

[28] Costello, E. K., Gordon, J. I., Secor, S. M., Knight, R. (2010). Postprandial remodeling of the gut microbiota in Burmese pythons. *ISME Journal, 4*, 1375-1385. DOI: 10.1038/ismej.2010.71.

[29] Keenan, S. W., Engel, A. S., Elsey, R. M. (2013). The alligator gut microbiome and implications for archosaur symbioses. *Scientific Report, 3*, 2877. DOI: 10.1038/srep02877.