A Review on Environmental DNA (eDNA) Metabarcoding Markers for Wildlife Monitoring Research

N Othman1, H Haris1, Z Fatin1, M F Najmuddin1, N H Sariyati1, B M Md-Zain2 and M A B Abdul-Latif1,3*

1Faculty of Applied Sciences and Technology, Universiti Tun Hussein Onn Malaysia, 84600, Johor, Malaysia
2Department of Biological Sciences and Biotechnology, Faculty of Science and Technology Universiti Kebangsaan Malaysia, Selangor, Bangi 43600, Malaysia
3Oasis Integrated Group (OIG), Institute for Integrated Engineering (I²E), Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Johor, Malaysia

*Corresponding author: latiff@uthm.edu.my

Abstract. Environmental DNA or eDNA utilizes traceable genetic materials in the environment for monitoring the presence of organisms in a given area and it is now gaining popularity as an alternative for traditional monitoring methods. Thus, the selection of genetic markers is crucial for identification of species in wildlife monitoring. This paper aims to review several DNA markers which can be utilized by next-generation sequencing (NGS) for biodiversity monitoring. Cytochrome C oxidase Subunit I (COI) are noted as the most widely used marker in metabarcoding research for detection of targeted species.

1. Introduction
In the 21st century, earth’s biodiversity shows a declining trend which become a major crisis and challenge for environmentalist to protect the ecosystem especially in identification of organisms. Traditionally, the conservation methods to access the distribution of wildlife can only be relied on morphological identification of species by visual surveys and manual counting the individuals. As time goes by, Sanger sequencing was introduced by Sanger et al [1] in 1970s with the development of polymerase chain reaction (PCR) have helped to identify the DNA of species by amplifying the genetic region of interest [2]. The amplification of DNA using PCR provides aid in species identification either for one or multiple species simultaneously depending on the genetic marker used. The genetic markers or DNA markers are identifiable DNA sequences found in genome that can be used to track the inheritance of the organism [3].

However, to recover DNA of thousand specimens from various environmental sample, usage of DNA metabarcoding approach may give a valuable insight for quantifying biodiversity [4] using high-throughput sequencing (HTS) or next-generation sequencing (NGS). NGS has the potential to generate several hundred thousand to multiple millions of sequencing reads for identification of species. Around 2000s, NGS evolved rapidly with various platforms released, where it initially started with Titanium pyrosequencer (Roche) followed by Illumina technology [5]. These platforms have been used in order...
to perform DNA sequencing which evolved significantly with increasing speed and minimizing the cost as the platforms have various principle of sequencing chemistry [6]. To overcome the problem in monitoring large-scale biodiversity, eDNA are used to extract the DNA in order to identify the species in an ecosystem using NGS technologies [4]. DNA from the environment holds various genetic material of the organism living in that habitat and they can be found and collected easily [7] which later can be used for analysis of the environment. With thousands of samples, NGS can compare the obtained sequences to a growing standard reference library of known organisms and taxa present in an environmental sample can be identified precisely [8].

Regardless of platform used, the development marker for eDNA metabarcoding is also important for eDNA assays if the amplified region is relatively short, effective and specific for the target species [9] as eDNA can be degraded in a short time [10]. Different primers and regions will give distinct results in coverage, resolution, and it can lead to bias between taxa in describing the communities in environment [11]. Hence, this review focuses on the suitable markers used in wildlife monitoring according to the eDNA research taxon. The findings can be used for future references during selecting the potential markers in biodiversity management and monitoring.

2. Materials and methods

The data mining on the selection of markers in eDNA study were done using SCOPUS database. According to the indexed title, abstract and keywords by using “eDNA” as the main keyword, the search of articles was conducted by adding some keywords involving markers used for detection of eDNA in wildlife monitoring, namely “COI”, “16S”, “Cytb”, “D-loop”, “18S”, “NADH” and “12S”. These markers were selected because most of the previous study exploited the region of mitochondrial DNA and nuclear DNA for detection eDNA among wildlife. In addition, “wildlife” keyword was included in data mining in order to determine the acceptable paper that can be used in this review paper. The selection of paper for wildlife in eDNA includes mammals, invertebrates, bird, reptiles and amphibian groups as presented in Figure 1.

3. Results

From the systematic review, the search of database on eDNA was sorted according to the types of markers used and according to the different taxon of animals. A total of 150 from 254 research articles related to eDNA of wildlife were included (Table 1). Figure 1 shows the literature count for all the studies of eDNA in wildlife from SCOPUS database. The suitable peer-reviewed articles were chosen according to the information that was related with wildlife monitoring, including information on detection of the invasive species and diseases among wildlife.

The result shows that most publications on eDNA are related with fishes and invertebrates compared to other taxons. Cytochrome C oxidase Subunit I (COI) marker is the most widely used genetic markers which correspond to 45 out of 150 articles across all the taxons. Relatively, the number of related studies for wildlife monitoring are far lower compared with the COI marker.

| Database | Search topics                  | Total number of research articles | Suitable peer-reviewed article |
|----------|--------------------------------|-----------------------------------|--------------------------------|
| Scopus   | “eDNA” AND “COI”              | 67                                | 45                             |
|          | “eDNA” AND “16S”              | 57                                | 23                             |
|          | “eDNA” AND “D-loop”           | 5                                 | 4                              |
|          | “eDNA” AND “Cytb”             | 5                                 | 4                              |
|          | “eDNA” AND “12S”              | 27                                | 22                             |
|          | “eDNA” AND “18S”              | 44                                | 19                             |
|          | “eDNA” AND “NADH”             | 4                                 | 3                              |
|          | “eDNA” AND “Wildlife”         | 45                                | 30                             |
| Total    |                                 | 254                               | 150                            |
4. Discussion

COI serves as the standard DNA barcode marker for animals and it has an extensive reference library such as Barcode of Life Data Systems (BOLD) and GenBank (NCBI), as well as taxonomic discriminatory power and predictable sequence variation [12]. Previously, most researchers used COI marker because it is capable to discern between closely related animals [13] which can be used in eDNA metabarcoding. However, from the systematic literature review of eDNA, for some targeted specific taxonomic groups in environmental samples, COI marker still requires a lot of research as the utilization of COI has yet to be fully explored [14]. Besides, the limitation of COI mini-barcode primers is still difficult to prove as 130 bp of COI by Ficetola et al. [15] are not sufficiently conserved to cover a broad range of taxa [16,17]. This is also supported by Horton et al. [18] in which DNA metabarcoding are not able to efficiently utilize the protein coding genes such as COI because the interspecific genetic variation impedes the use of universal primers.

Ribosomal RNAs (12S, 16S and 18S) are also strongly suggested to be used for DNA metabarcoding as rRNAs is highly conserved regions and evolve more slowly than the mitochondrial genome as a whole [19] which is important for ribosomes to provide a species-specific signature sequence [20, 21]. 12S and 16S are able to amplify highly variable mitochondrial from various animal species sequences but 12S rRNA might yield different results compared to 16S rRNA due to the conserved regions for both of the genes at the 3’ ends of sequences [22]. The short hypervariable regions of 12S and 16S are most commonly used as genetic markers in identifying the wide range of animals in environmental samples [23, 24] especially for easily degraded samples [25]. As for 18S rRNA region, the references database is still considered as the limiting factor due to the insufficient references for taxonomic identification of many species [26, 27]. Cytb and D-loop genes may be the most mtDNA markers used for identification of vertebrates especially in conventional PCR [28] but the exploration on using these markers are still lower in eDNA field. Therefore, multi-species approach using Cytb and D-loop regions are needed in the future to test the accuracy for metabarcoding which is highly dependent on marker choices.

*NA: not available

Figure 1. Types of eDNA markers used in the monitoring of different wildlife groups.
Eventually, there is no specific DNA marker that has all the features to be used as a standard metabarcoding marker and chosen as best marker with extensive reference databases. So, choices of marker regions for metabarcoding can be said solely depending on the type of animal taxons involved. Furthermore, shorter region of eDNA markers will be a suitable alternative but it will still unable to provide broad species-level resolution. Hence, a standard universal primer for mini-barcode amplification is difficult to design [29] in order to amplify the degraded DNA from the environment [30] as preferable amplicons length suggested by using Illumina or Ion Torrent platforms is only around 100-200 bp [31].

5. Conclusion
This systematic survey data helps the researchers to decide on the selection of markers for eDNA metabarcoding. COI is mostly used for invertebrates and fish, while other rRNAs markers (12S, 16S and 18S) are also used across other taxons in eDNA study. However, the validation of markers should be performed as the potential and limitation of each markers varies according to the targeted species. Future studies should focus on strengthening DNA database and experimenting with various new short primer for eDNA metabarcoding in order to unlock the immense potential of eDNA research in wildlife monitoring.

6. Acknowledgement
The authors acknowledge Ministry of Higher Education Malaysia and Universiti Tun Hussein Onn Malaysia for providing the necessary funding, facilities, and assistance. This research was supported by Malaysian Technical University Network Grant (UTHM-MTUN-K121), Fundamental Research Grant Scheme FRGS/1/2018/WAB13/UTM/03/2 provided by Ministry of Higher Education Malaysia and GPPS-UTHM-2019-H528 postgraduate grant by Universiti Tun Hussein Onn Malaysia (UTHM), Malaysia.

References
[1] Sanger F, Nicklen S and Coulson A R 1977 DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences of the United States of America 74 5463-5467.
[2] Freeland J R, Kirk H and Pedersen S 2011 Molecular Ecology (2nd ed.) Wiley & Sons, 601 Chichester, UK.
[3] Raza S, Shoaib M W and Mubeen H 2016 Genetic Markers: Importance, uses and applications, International Journal of Scientific and Research Publications 6.
[4] Taberlet P, Coissac E, Hajibabaei M and Rieseberg L H 2012 Environmental DNA. Molecular Ecology 21 1789-1793.
[5] Stefani S, Stankovic D, Borme D, Olazabal A, Juretic T, Pallavicini A and Tirelli V 2018 Multi marker metabarcoding approach to study mesozooplankton at basin scale. Scientific Reports 8 12085.
[6] Ambardar S, Gupta R, Trakroo D, Lal R and Vakhlu J 2016 High throughput sequencing: an overview of sequencing chemistry. Indian Journal Microbiology 56 (4) 394-404.
[7] Baird D J and Hajibabaei M 2012 Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. Molecular Ecology 21 2039–2044.
[8] Shokralla S, Spall J L, Gibson J F and Hajibabaei M 2012 Next-generation sequencing technologies for environmental DNA research. Molecular Ecology 21 1794–1805.
[9] Freeland J R 2017 The importance of molecular markers and primer design when characterizing biodiversity from environmental DNA (eDNA). Genome 60 (4) 358-374.
[10] Ruppert K M, Kline R J and Rahman M S 2019 Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. Global Ecology and Conservation 17 e00547.
[11] Thomsen P F and Willerslev E 2015 Environmental DNA - an emerging tool in conservation for
monitoring past and present biodiversity. *Biological Conservation* **183** 4–18.

[12] Collins R A, Bakker J, Wangensteen O S, Soto A Z, Corrigan L, Sims D W, Genner M J and Mariani S 2019 Non-specific amplification compromises environmental DNA metabarcoding with COI, *Methods in Ecology and Evolution* **10** 1-17.

[13] Deagle B E, Jarman S N, Coissac E, Taberlet P and Deagle B E 2014 DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. *Biology Letters* **10** 20140562.

[14] Hebert P D, Cywinska A and Ball S L 2003 Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences* **270** 313–321.

[15] Meusnier I, Singer G A, Landry J F, Hickey D A, Hebert P D and Hajibabaei M 2008 A universal DNA mini barcode for biodiversity analysis. *BMC Genomics* **9** pp 214.

[16] Ficetola GF, Coissac E, Zundel S, Riaz T, Shehzad W, Bessiere J, Taberlet P and Pompanon F 2010 An: in silico approach for the evaluation of DNA barcodes. *BMC Genomics*. **11** 434.

[17] Leray M, Yang J Y, Meyer C P, Mills S C, Agudelo N, Ranwez V, Boehm J T and Machida R J 2013 A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Frontiers in Zoology* **10** 34.

[18] Horton D J, Kershner M W and Blackwood C B 2017 Suitability of PCR primers for characterizing invertebrate communities from soil and leaf litter targeting metazoan 18S ribosomal or cytochrome oxidase I (COI) genes. *European Journal of Soil Biology* **80** 43-48.

[19] Valentini A, Taberlet P, Miaud C, Civade R, Herder J, Thomsen P F and Bellemain E 2016 Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. *Molecular Ecology* **33** 929–942.

[20] Riaz T, Shehzad W, Viari A, Pompanon F, Taberlet P and Coissac E 2011 ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis. *Nucleic Acids Research* **39** (21) e145.

[21] Elbrecht V, Taberlet P, Dejean T, Valentini A, Usseglio-polatera P, Beisel J-N, Coissac E, Boyer F and Leese F 2016 Testing the potential of a ribosomal 16S marker for DNA metabarcoding of insects. *Peer J Preprints* **4** e1855v1.

[22] Dell'Anno A, Carugati L, Corinaldesi C, Riccioni G and Danovaro R 2015 Unveiling the biodiversity of deep-sea nematodes through metabarcoding: Are we ready to bypass the classical taxonomy? *PLoS One* **10** e0144928.

[23] Foote A D, Thomsen P F, Svøegaard S, Wahlberg M, Kielgast J, Kyhn L A, Salling A B, Galatius A, Orlando L and Gilbert M T P 2012 Investigating the Potential Use of Environmental DNA (eDNA) for Genetic Monitoring of Marine Mammals. *Plos One* **7** (8) 41781.

[24] Staats M, Arulandhu A J, Gravendeel B, Holst-Jensen A, Scholtens I, Peelen T, Prins T W and Kok E 2016 Advances in DNA metabarcoding for food and wildlife forensic species identification. *Analytical and Bioanalytical Chemistry* **408** 4615–4630.
M T P and Willerslev E 2012 Meta-barcoding of “dirt” DNA from soil reflects vertebrate biodiversity. *Molecular Ecology* 21 1966–1979.

[31] Pavan-Kumar A, Gireesh-Babu P and Lakra W S 2015 DNA Metabarcoding: a new approach for rapid biodiversity assessment. *Journal of Cell Science & Molecular Biology* 2 111.