Short Communication

Use of a portable DNA sequencer for invertebrate species identification in Cambodia, a pilot study

Heather RITCHIE-PARKER[1,*,‡], KEATH Sophorn[2,‡], THI Sothearen[2], HUN Seiha[2], HAK Kosal[2], PHAK Satha[2], CHIM Samhaiy[2], MEI Sophea[2], THOU Sophak[2], PHANN Sam Ath[2], KHIN Chandara[2], SONG Iva[2], MEAS Seanghun[2,§] & Alexander BALL[1,§]

[1] RZSS WildGenes, Conservation Department, Royal Zoological Society of Scotland, Edinburgh, UK.

[2] Centre for Biodiversity Conservation, Department of Biology, Faculty of Science, Royal University of Phnom Penh, Phnom Penh, Cambodia.

‡ Joint first authors. § Joint last authors.

* Corresponding author. Email hritchieparker@rzss.org.uk

Paper submitted 11 August 2023, revised manuscript accepted 16 May 2024.

DNA barcoding is a well-established molecular genetics tool that uses a short, standardised reference gene, or genes, to create species-specific DNA sequences known as barcodes. It can be used for a wide range of applications such as creating genetic references for voucher specimens (Ward *et al*., 2009), species identification and taxonomic delimitation (DeSalle *et al*., 2005) and population genetics (Hajibabaei *et al*., 2007b). These applications can be used to aid conservation efforts by identifying illegally traded species (Rehman *et al*., 2015), contributing to population assessments (Wilson *et al*., 2016), identification of invasive species (Armstrong & Ball, 2005) and biodiversity monitoring (Hajibabaei *et al*., 2007a).

The accessibility of barcoding has changed over the years with the advent of portable sequencers increasing accessibility while decreasing costs and turnaround times for data generation. They allow for rapid, in-country identification of samples which is particularly important for conservation efforts as projects can be time sensitive, occur in remote locations, or focus on CITES-protected species where international shipping is restricted (Krehenwinkel *et al*., 2019). As well as providing the ability to sequence samples in situ, portable sequencers also provide opportunities for sequencing in locations where access to traditional sequencing methodologies is limited (Pomerantz *et al*., 2018) or prohibitively expensive (von Rintelen *et al*., 2017). This is pertinent to many conservation priorities as biodiversity hotspots are often located in regions facing these challenges, such as Southeast Asia (Myers *et al*., 2000).

We tested the utility of a portable DNA sequencer, the Oxford Nanopore Technologies (UK) MinION, for identification of invertebrate species in Cambodia during a workshop ran at the Centre for Biodiversity Conservation in February 2023. Briefly, invertebrate species were collected from the Royal University of Phnom Penh campus (Fig. 1) and euthanised before being transferred to a sample tube with absolute ethanol for storage. A basic taxonomic identification was recorded for each sample using simple morphological characteristics but for any future development of voucher specimens we would recommend an expert taxonomic identification and capturing high-resolution images of samples where possible. DNA was extracted from the samples (Fig. 2) using QuickExtract (LGC Biosearch Technologies, Middlesex, UK) following the manufacturers protocol. Approximately 365 bp of the COI barcoding locus were amplified using the mlCOIintF (5′-GGWACWG-

CITATION: Ritchie-Parker, H., Keath S., Thi S., Hun S., Hak K., Phak S., Chim S., Mei S., Thou S., Phann S.A., Khin C., Song I., Meas S. & Ball, A. (2024) Use of a portable DNA sequencer for invertebrate species identification in Cambodia, a pilot study. *Cambodian Journal of Natural History*, **2024**, 86–90.

Fig. 1 Invertebrate sample collection on campus of Royal University of Phnom Penh.

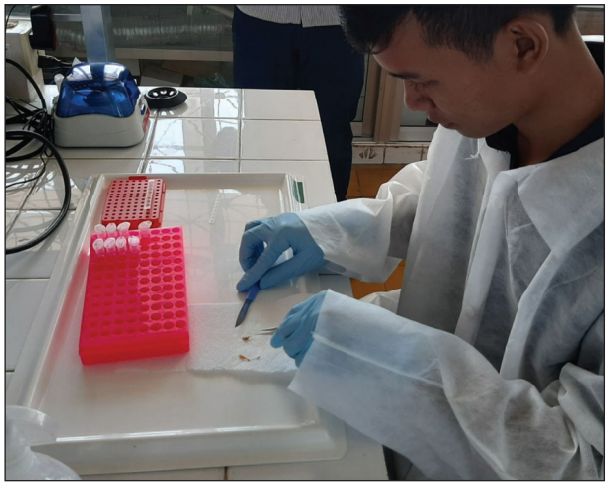

Fig. 2 Preparation of invertebrate samples for DNA extraction in laboratory.

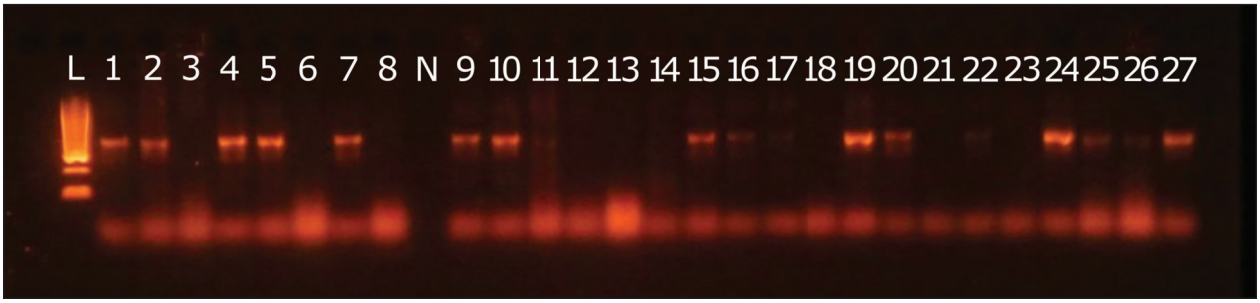

Fig. 3 Gel electrophoresis image visualising which samples were successfully amplified at the COI barcoding loci. The first well (L) contains a DNA ladder (HyperLadder 100 bp) used to determine the length of PCR product amplified. Samples were morphologically identified as: 1) ladybird, 2) fly, 3) treehopper, 4) yellow butterfly, 5) moth, 6) spider, 7) damselfly, 8) butterfly, 9) grasshopper, 10) cockroach, 11) hopper, 12) fly, 13) hopper, 14) beetle, 15) mosquito, 16) butterfly, 17) dragonfly, 18) worm, 19) spider, 20) yellow ant, 21) cockroach, 22) grasshopper, 23) leaf hopper, 24) spider, 25) fly, 26) ant, 27) red ant, and N indicates the negative control. Samples that were successfully amplified are shown by a band over 300 bp (e.g., sample 1) and those that failed have an absence of a band (e.g. sample three).

GWTGAACWGTWTAYCCYCC-3', Wangensteen *et al*., 2018) and jgHCO2198 (5'-TAIACYTCIGGRTGICCRAARAAYCA-3', Geller *et al*., 2013) primer set and the following 10 µl PCR master mix: 1.4X DreamTaq Hot Start PCR Master Mix (Thermo Scientific, Massachusetts, US), 1 µM forward primer (mlCOIintF), 1 µM reverse primer (jgHCO2198), 1.5 µl nuclease free water, 0.5 µl bovine serum albumin, and 1 µl DNA template. PCRs were conducted on a MiniPCR (MiniPCR, Massachusetts, US) and conditions were as follows: initial denaturation at 95 °C for five minutes; 35 cycles of denaturation at 95 °C for 30 seconds, annealing at 50 °C for 30 seconds and extension at 72 °C for 30 seconds, followed by final extension at 72 °C for two minutes. PCR products were visualised on a 2% agarose gel (Fig. 3). All 18 samples that successfully amplified at PCR were pooled together to create a sequencing library following the ligation sequencing amplicons protocol (SQK-LSK114, Oxford Nanopore Technologies).

The library ran for two hours and generated a total of 2.99 million reads. Reads were basecalled with Guppy v6.4.6 (Oxford Nanopore Technologies) using default parameters where a total of 2.19 million reads passed quality control. These sequences were compared to a local database containing COI reference sequences taken from NCBI GenBank (<www.ncbi.nlm.nih.gov/>) and BOLD (<www.boldsystems.org/>) repositories (Meglécz, 2023; database available from <https://zenodo.org/

Table 1 Details of 12 specimens identified using the sequences generated during the workshop. Each specimen was given a morphological ID and the consensus sequences generated were compared to NCBI GenBank and BOLD repositories where the top match was recorded alongside the query coverage and identity match. Supercript figures indictate the taxonomic level each specimen was identified to: [1] Species, [2] Genus, [3] Family, [4] Order.

Morphological ID	NCBI GenBank				BOLD			
	DNA Barcoding ID	Top Match Accession	Query Cover	Identity Match	DNA Barcoding ID	Top Match BIN	Query Cover	Identity Match
Spider	*Tetragnatha mandibulata* [1]	MK057477	100%	99.37%	*Tetragnatha mandibulata* [1]	AAK2567	100%	97.59%
Yellow butterfly	*Eurema hecabe* [1]	MN609532	100%	100%	*Eurema hecabe* [1]	AAA6082	100%	100%
Ladybird	*Illeis bistigmosa* [1]	MZ325765	100%	100%	*Illeis* sp. [2]	-	100%	97.78%
Fly sp.1	Dolichopodidae sp. [3]	KX052896	98%	90.42%	Dolichopodidae [3]	ACV0962	-	99.04%
Cockroach	*Pycnoscelus surinamensis* [1]	MW535117	100%	99.69%	*Pycnoscelus surinamensis* [1]	AAG9904	100%	99.68%
Dragonfly	*Brachythemis contaminata* [1]	MG885560	98%	100%	*Brachythemis contaminata* [1]	ADC3495	100%	98.41%
Grasshopper	*Pseudoxya diminuta* [1]	KC139999	100%	100%	*Pseudoxya diminuta* [1]	ACD4638	100%	100%
Fly sp.2	Diptera sp. [4]	GU675516	100%	97.17%	*Calliphora dispar* [1]	AAH7137	100%	97.17%
Yellow ant	*Anoplolepis gracilipes* [1]	MK482686	100%	99.69%	*Anoplolepis gracilipes* [1]	AAA9474	100%	100%
Red ant	*Oecophylla smaragdina* [1]	AB185478	100%	99.69%	*Oecophylla smaragdina* [1]	AAA5846	100%	99.68%
Moth	*Hypena* sp. [2]	KX860485	100%	99.69%	*Hypena simplicalis* [1]	AAG5843	100%	99.68%
Damselfly	*Agriocnemis pygmaea* [1]	MK506257	100%	100%	*Agriocnemis pygmaea* [1]	ABW0502	100%	100%

record/6555985#.ZHb953bMKUk>). Positive matches to the database were made for 1.78 million reads (81%). Matches were grouped by taxonomic ID and ranked by number of positive matches. All taxonomic ID groups that had less than 10,000 hits were discarded, and the remaining sequences were removed if they were less than 100 bp in length or had a BLAST identity match below 90%.

A total of 12 taxonomic ID groups passed quality control measures. Each group was mapped to a taxonomically appropriate reference sequence (i.e. the top GenBank or BOLD hit for each group) in Geneious Prime (v2021.1.1: Biomatters, Auckland, New Zealand) using the Geneious mapper algorithm and default parameters. Consensus sequences 318 bp in length were generated for each of the 12 taxonomic ID groups. Consensus nucleotide sequences were translated into protein sequences to check for the presence of NUMTs (nuclear mitochondrial DNA segments) and none were identified. Finally, consensus sequences were again compared to the NCBI GenBank and BOLD repositories for classification. We successfully classified 12 samples from the generated sequencing data (Table 1).

We were able to identify a total of 12 different specimens from the 18 samples that successfully amplified at PCR. The disparity between the number of samples amplified and the number of specimens identified may be due to several factors. Firstly, the level of PCR amplification was not consistent across all samples so weaker samples may not have sequenced at a sufficient depth to

pass quality control measures. Secondly, as samples were given a broad taxonomic classification based on basic morphology some samples may have been duplicates of the same species (e.g., two samples sequenced were identified as grasshoppers but only one species of grasshopper was identified). Lastly, our reference database may not have had sufficient taxonomic breadth to identify every sample, although it is unlikely that samples would not have been identified at a higher taxonomic level (i.e. Order).

Basecalled sequences used to generate the consensus sequence for each of the 12 taxonomic groups are available via Figshare (<https://figshare.com/>), as follows, Taxonomic ID Groups 1–12 (respectively): 10.6084/m9.figshare.23300720, –.23301173, –.23301197, –.23301203, –.23301212, –.23301224, –.23301230, –.23301239, –.23301242, –.23301251, –.23301263 and –.23301269. Consensus sequences are also available from Figshare (10.6084/m9.figshare.24525697).

We showed the utility of a portable DNA sequencer for invertebrate species identification in Cambodia. We were able to identify samples with a high degree of certainty but not all specimens could be identified to species level; this highlights the paucity of barcoding voucher sequences for species in Cambodia. Most samples identified to species level were taxa with widespread distributions that had been catalogued and barcoded from elsewhere in their ranges. Developing a barcoding database from morphologically-identified museum voucher specimens for species in Cambodia would be of vast benefit to conservation efforts in the region (Francis *et al*., 2010). As well as providing valuable information about the biology and ecology of native species, it would also provide the baseline data required for species identification and biodiversity monitoring (Krishnamurthy & Francis, 2012).

Portable DNA sequencers provide an important opportunity to increase genetics capacity for conservation in Cambodia where access to sequencing facilities has often been limited, prohibitively expensive, or where transportation of samples has been challenging. Specifically, the Oxford Nanopore Technologies portable sequencer, MinION, offers a competitively priced sequencing solution (Watsa *et al*., 2020) and the newly established distribution facility in Singapore will increase the accessibility of materials across the Asia-Pacific region (Oxford Nanopore Technologies, 2023). Given its portable nature, MinION also presents the opportunity for use in remote locations as well as acting as a teaching tool (Watsa et al., 2020) and while it is still often purported to produce data with a high error rate, recent developments in pore technology and chemistry have greatly improved sequencing accuracy (van Djik *et al*., 2023).

As with any new technology careful consideration does need to be made for the use of portable DNA sequencers in Cambodia. In particular, flow cells for MinION have a limited shelf life and the longevity of flow cells and other required reagents is affected by temperature storage. Preserving these items could be difficult as access to stable temperature storage is challenging in this region. It is also important that appropriate training is acquired on the use of flow cells, preparation of input materials and the bioinformatic pipelines required for analysis. Because our protocol was tested within a laboratory setting, additional development would be required to ensure all components operate effectively in field conditions before deployment in remote work locations. Overall, portable DNA sequencers provide an opportunity to increase capacity domestically and a means to develop genetic resources in country (Pomerantz *et al*., 2018; Krehenwinkel *et al*., 2019; Watsa *et al*., 2020).

Demonstrating the use of a portable sequencer in Cambodia is a first step to developing its potential for conservation in the region. Our pilot study used established protocols and acts as a proof of concept for the utility of portable sequencers in Cambodia. Its success has shown that with further development there is potential for in-country generation of barcoding data for invertebrate voucher specimens where protocols can be adapted to a range of taxa, as well as other applications such as identification of wildlife products that are traded illegally. While the protocols and techniques demonstrated here may not be novel to the wider scientific community, they represent the potential for a shift in the landscape of conservation genetics in Cambodia.

Acknowledgements

We would like to thank the Centre for Biodiversity Conservation at the Royal University of Phnom Penh, specifically Uk On Norong, for facilitating the workshop that led to our pilot study. We also thank the attendees that participated in the various stages of the workshop: Soum Savoeurn, Bun Phanna, Thoeun Leakhena, Hap Sophorn, Chhang Soknang, Thy Vandet, Khom Seiha, Uy Sophorn, Yan Phanna, Chroeng Chantha, Soth Phalla and Say Theavy. We are grateful to Kara Dicks for contributing to development of the workshop, Georgina Cole for advice on invertebrate euthanasia, Kriss Paul and Ade Worley for IT assistance and Mark Haddow for supporting the running of the workshop. We also thank the Royal Zoological Society of Scotland for funding.

References

Armstrong, K.F. & Ball, S.L. (2005) DNA barcodes for biosecurity: invasive species identification. *Philosophical Transactions of the Royal Society B*, **360**, 1813–1823.

DeSalle, R., Egan, M.G. & Siddall, M. (2005) The unholy trinity: taxonomy, species delimitation and DNA barcoding. *Philosophical Transactions of the Royal Society B*, **360**, 1905–1916.

Francis, C.M., Borisenko, A.V., Ivanova, N.V., Eger, J.L., Lim, B.L., Guillén-Servent, A., Kruskop, S.K., Mackie, I. & Hebert, P.D.N. (2010) The role of DNA barcodes in understanding and conservation of mammal diversity in Southeast Asia. *PLoS ONE*, **5**, e12575.

Geller, J., Meyer, C., Parker, M. & Hawk, H. (2013) Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys. *Molecular Ecology Resources*, **13**, 851–861.

Hajibabaei, M., Singer, G.A.C., Clare, E.L. & Hebert, P.D.N., (2007a) Design and applicability of DNA assays and DNA barcodes in biodiversity monitoring. *BMC Biology*, **5**, 24.

Hajibabaei, M., Singer, G.A.C., Hebert, P.D.N. & Hickey, D.A. (2007b) DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Trends in Genetics*, **23**, 167–172.

Krehenwinkel, H., Pomerantz, A. & Prost, S. (2019) Genetic biomonitoring and biodiversity assessment using portable sequencing technologies: current uses and future directions. *Genes*, **10**, 858.

Krishnamurthy, K.P. & Francis, R.A. (2012) A critical review on the utility of DNA barcoding in biodiversity conservation. *Biodiversity and Conservation*, **21**, 1901–1919.

Meglécz, E. (2023) COInr and mkCOInr: building and customizing a nonredundant barcoding reference database from BOLD and NCBI using a semi-automated pipeline. *Molecular Ecology Resources*, **23**, 933–945.

Myers, N., Mittermeier, R., Mittermeier, C. da Fonseca, G.A.B & Kent, J. (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858.

Oxford Nanopore Technologies (2023) *Oxford Nanopore Technologies Teams up with UPS Healthcare to Accelerate Delivery of New-Generation DNA Sequencing Technology to More Customers in Locations across Asia Pacific*. Https://nanoporetech.com/about-us/news/oxford-nanopore-technologies-teams-ups-healthcare-accelerate-delivery-new-generation [accessed 21 May 2023].

Pomerantz, A., Peñafiel, N., Arteaga, A., Bustamante, L., Pichardo, F., Coloma, L.A., Barrio-Amorós, C.L., Salazar-Valenzuela, D. & Prost, S. (2018) Real-time DNA barcoding in a rainforest using nanopore sequencing: opportunities for rapid biodiversity assessments and local capacity building. *GigaScience*, **7**, giy033.

Rehman, A., Jafar, S., Raja, N. A. & Mahar, J. (2015) Use of DNA barcoding to control the illegal wildlife trade: a CITES case report from Pakistan. *Journal of Bioresource Management*, **2**, 3.

van Dijk, E. L., Naquin, D., Gorrichon, K., Jaszczyszyn, Y., Ouazahrou, R., Thermes, C. & Hernandez, C. (2023) Genomics in the long-read sequencing era. *Trends in Genetics*, **39**, 649–671.

von Rintelen, K., Arida, E. & Häuser, C. (2017) A review of biodiversity-related issues and challenges in megadiverse Indonesia and other Southeast Asian countries. *Research Ideas and Outcomes*, **3**, e20860.

Wangensteen, O.S., Palacín, C., Guardiola, M. & Turon, X. (2018) DNA metabarcoding of littoral hard-bottom communities: high diversity and database gaps revealed by two molecular markers. *PeerJ*, **6**, e4705.

Ward, R.D., Hanner, R. & Herbert, P.D.N. (2009) The campaign to DNA barcode all fishes, FISH-BOL. *Journal of Fish Biology*, **74**, 329–356.

Watsa, M., Erkenswick, G.A., Pomerantz, A. & Prost, S. (2020) Portable sequencing as a teaching tool in conservation and biodiversity research. *PLoS Biol*, **18**, e3000667.

Wilson, J.J., Sing K.W., Lee P.S. & Wee A.K.S. (2016) Applications of DNA barcodes in wildlife conservation in tropical East Asia. *Conservation Biology*, **30**, 982–989.

