SURVEY FOR AFRICAN FRESHWATER FISH USING DNA BARCODING

***Authors:** Suzanne Coey, Environmental Resources Management Ltd; Carmelo Andújar Fernández, Imperial College London/Natural History Museum; Alfried P Vogler, Imperial College London/Natural History Museum; Douglas Yu, University of East Anglia and Kunming Institute of Zoology; Phil Atkinson, Combined Ecology; Andy Coates; Environmental Resources Management Ltd.*

Introduction

Collection of baseline survey data to inform ecological impact assessment (EcIA) requires experienced specialists to undertake field surveys. This can be time consuming and costly, particularly in remote areas. Recent technological advances have shown the promise of environmental DNA (eDNA) in saving costs during EcIA as a rapid screening tool and by improving the efficiency of monitoring schemes (Herder *et al*, 2014; Rees *et al*, 2014). For example, in the UK this technique, approved by Natural England, is >99% efficient in detecting the presence of great crested newts, compared to ~76% using traditional field survey techniques (Biggs *et al*, 2014).

Freshwater species have undergone a 76% decline globally since 1970 (WWF, 2014) and in Africa the rate of freshwater biodiversity loss is of concern due to the high levels of endemism and rapid pace of habitat degradation. The assessment of species distribution is critical to EcIA; however in much of Africa there is a lack of knowledge of even basic distributional data for freshwater fish (Darwall *et al*, 2011). Further research is required, particularly given the continued expansion of extractive industries and hydropower development. Freshwater species across much of Africa are understudied even in conventional terms, and although efforts to collect their DNA barcode information have begun (*eg* Lowenstein *et al*, 2011); work to date has been limited. This paper outlines a study by ERM Ltd, in partnership with Combined Ecology, the Natural History Museum (London) and the University of East Anglia. The study has two objectives:

1) to expand the library of DNA barcodes for freshwater fish in Africa; and
2) to investigate the performance of eDNA as a survey technique when undertaken concurrently with conventional survey methods.

What is eDNA?

Environmental DNA methods involve the study of DNA left behind by species in the environment (in water, soil *etc*), as free DNA or cell debris (Valentini *et al*, 2009). The methods rely on the amplification by Polymerase Chain Reaction (PCR) of small DNA fragments (typically <150 base pairs (bp)). Small fragments are more stable, and thus easier to recover from the environment than longer fragments (Deagle *et al*, 2006).

The cytochrome b (*cytb*) gene of mitochondrial DNA [(1)] is commonly used for determining phylogenetic relationships between organisms, due to its variability and diagnostic characters.

[(1)] The majority of eDNA studies use a mitochondrial gene as a marker, as mitochondrial DNA is much more abundant than nuclear DNA, enhancing the likelihood of detection in environmental samples. The *cytb* gene has an evolutionary rate which makes it appropriate for species identification based on genetic variation.

Fragments of this gene are suitable for amplification through PCR and constitute one of a number of options for use as a DNA barcoding marker.

In general, eDNA methods can be classified in two main groups (detailed below).

(1) Methods relying on quantitative PCR (qPCR) with specific primers [(1)] to detect the presence of a particular species (**eDNA-barcode methods**). With this method, species-specific primers must be designed for the particular ecosystem of interest (*ie* each individual species).

(2) Methods using universal primers, which amplify the DNA of multiple species in a single sample, subsequently sequenced using 'Next Generation Sequencer' (NGS) methods (**eDNA-metabarcoding methods**) (Taberlet *et al*, 2012). With this method it is still frequently necessary to adapt the primers for the species set expected (*eg* invertebrates, fish).

In remote areas with limited background information, the points above imply the need to obtain tissue samples from the species present, extract DNA, amplify and obtain the required sequences. Primers can then be designed and optimised [(2)] so that the environmental samples (water, soil *etc*) can be analysed for the presence of those species.

eDNA in Aquatic Systems

In aquatic ecosystems, eDNA has proven especially useful due to the characteristics of the aquatic environment (Herder *et al*, 2014). eDNA spreads in water and persists from days to weeks (Dejean *et al*, 2011; Pilliod *et al*, 2014; Thomsen *et al*, 2012b); enough to detect the recent presence of species and avoiding the detection of historic presence (Dejean *et al,* 2011).

Recent papers have proven the potential of eDNA barcoding to detect amphibians (Biggs *et al*, 2014; Thomsen *et al*, 2012b), arthropods (Jones *et al*, 2008; Thomsen *et al*, 2012b), gastropods (Goldberg *et al*, 2013; Lance and Carr, 2012), mammals (Thomsen *et al*, 2012b) and fish (Jerde *et al*, 2013; Mahon *et al*, 2013) (see Herder *et al*, 2014 for a complete review).

Regarding eDNA metabarcoding approaches, only a few studies have focused on aquatic eDNA and fish. Minamoto *et al,*(2011) used *cytb* universal primers, tested on DNA from aquaria and a river system, detecting four different fish species. A study by Thomsen *et al* (2012b) detected the presence of four of seven fish species in a freshwater system using specifically designed primers for a short fragment of the *cytb* gene. Thomsen *et al* (2012a) also adapted *cytb* primers for a set of marine fish detecting up to 15 species and outperforming/equaling a set of 9 conventional survey methods.

[(1)] Primers are chemically synthesised molecules which bind to the section of DNA of interest (in this case, a fragment of the *cytb* gene) which is to be amplified during PCR. Primer design must account for the variation within a species and the variation among species, to avoid false positive or false negative results.

[(2)] Testing of the primer to ensure that the qPCR reaction always results in a positive detection in the presence of target DNA and no amplification of non-target DNA occurs.

Building a Barcode Database for African Freshwater Fish

The first objective of this study is to add to the existing barcode database for freshwater fish in Africa, which is at present limited. To this end, fish tissue samples were collected during planned commercial survey visits to Cameroon (Congo and Ogooue river basins), Mozambique (Messalo river basin) and South Africa (Incomati river basin).

At each sampling location, fish were caught using conventional means (electrofishing, seine netting and collection from fishermen). 61 fish species were caught from 21 sample sites. Fin clips were taken from a subsample of each species and preserved in ethanol for later analysis. Where possible, samples from 3-5 individuals per species were collected to account for intra-species variation.

In the laboratory, the fish tissue samples were individually extracted using standard Qiagen DNA extraction methods and sequenced using Sanger technology to obtain a fragment of the *cytb* gene (285bp, Primers: L14912-CYB 5' TTCCTAGCCATACAYTAYAC; H15149-CYB 5' GGTGGCKCCTCAGAAGGACATTTGKCCYCA) (Miya and Nishida, 2000). These longer 285bp fragments contain more taxonomic information than the <150bp fragments which typically persist as eDNA in water; however the current trial attempted first to look for longer pieces before zeroing in on the proven protocol which targets the shorter pieces.

Barcodes were successfully sequenced for 59 of the 61 species sampled, giving a recovery rate of 97%. A phylogenetic tree was reconstructed using BEAST software (Drummond *et al*, 2012) and the Generalized Mixed Yule Coalescent (GMYC) method (Pons *et al*, 2006) was used to delimit the species.

Fish Survey Using In-stream DNA

For the second objective of the study, DNA samples were extracted from river water at approximately half the sampling sites in Cameroon to compare the detection rate using this method with the conventional sampling techniques. Samples were collected when water levels were subsiding following the wet season. Collection was first attempted in the dry season (when eDNA would have been more concentrated in the water); however an electric pump was not available at that time and filtration by hand pump was not possible due to high turbidity.

DNA was collected through filtration of water from 8 of the 14 sites. Collection from all 14 was not possible due to access problems following the heavy rains. Samples of 300 ml (12 subsamples of 25 ml each) were collected from each watercourse and filtered *in situ* using a battery operated vacuum pump and disposable funnels with 0.45 µm acetate cellulose filters (Goldberg *et al*, 2011). Filtered samples were stored in ethanol to await laboratory analysis.

In the laboratory, an initial test using eDNA metabarcoding was performed following Minamoto *et al* (2011), to amplify the same region of the *cytb* gene (285bp) previously sequenced for the tissue samples. This first test could not amplify the 285 bp sections, likely because the DNA would have broken up into shorter fragments in the river. The next step is to use the sequences obtained from the tissue extraction from the fish fin clips to design:

(a) species-specific primers for eDNA barcoding for each species for which fin clips were obtained; and
(b) a set of universal primers for a shorter fragment (~80bp) of the *cytb* gene.

The species-specific primers developed for each target species can then be used for the eDNA barcoding method using qPCR. Multiple species-specific probes could be used to assay for several species simultaneously. The universal primers will be used to amplify all species in the water sample together, targeting a very small DNA fragment, followed by sequencing of the mixed amplicons [(1)] *ie* a metabarcoding approach.

Discussion

Extraction of *cytb* sequences from the fish tissue samples has proved a success, with barcodes collected for 59 different fish species, expanding the reference library for Africa.

Following analysis of the tissue samples and comparison against existing records it was possible to match some sequences with barcodes on the Genbank database [(2)]. However it was noted that some, when related to the database, were assigned to an incorrect species. For example, the database suggested that the barcode for the species from Cameroon known to be *Hemichromus elongatus* was *Hemichromus fasciatus,* which is phenotypically distinct from the former and not known from Cameroon. The samples of *H. elongatus* collected were 98% similar to the barcode for *H. fasciatus* in Genbank. Searching the database produces the closest match and because *H. elongatus* has not been barcoded, whereas *H. fasciatus* has, the search returned *H. fasciatus* as being the closest match. Another possible explanation for mis-identification would be that the specimen from which the Genbank barcode was obtained was incorrectly identified. Some databases are tied to museum-curated specimens (so the identity can be checked) but many, including Genbank, are not necessarily, and some interpretation is required. This highlights the importance of continuing to collect barcodes for a greater number of species in Africa, and for verification of barcodes from voucher specimens, to avoid species being wrongly assigned when direct conventional sampling is not also undertaken.

Trials to amplify *cytb* sequences from the water filter samples are ongoing and initial efforts to amplify relatively long 285bp sequences may have failed due to the degradation of the eDNA into smaller fragments within the rivers, which is not unexpected. There is also a possibility that the eDNA concentrations in the rivers were not high enough to yield a positive result from 300 ml samples, which is possible as samples were collected immediately following the wet season when the greater volume of water would dilute the DNA present. Once complete, however, it is hoped that with further analysis of the filter samples it will be possible to amplify shorter (<150bp) strands of the *cytb* gene which are more likely to have persisted in the environment.

Despite limitations, eDNA is currently considered suitable as a supplementary technique to traditional survey, *eg* as a screening tool for particular species of conservation concern. Such a

---

[(1)] Amplified DNA fragments.

[(2)] http://www.ncbi.nlm.nih.gov/genbank/

technique would be applicable for gathering data on the wider distribution of species during EcIA, including offset studies. It is also suitable as a rapid post-consent monitoring tool to investigate the continued persistence of species of interest throughout a project life cycle.

The use of eDNA as a commercial survey technique is in its early stages; however as researchers worldwide are continuing to expand our knowledge base, the reliability of this technique is likely to improve rapidly. Further efforts to collect barcodes mean that reference libraries will continue to expand, improving the value of existing databases. With this in mind, it is very likely that the use of eDNA will have a significant role to play worldwide in the future of survey for impact assessment, offset studies and monitoring.

References

Biggs, J., Ewald, N., Valentini, A., Gaboriaud, C., Griffiths, R., Foster, J., Wilkinson, J., Arnett, A., Williams, P., Dunn, F., 2014. Analytical and methodological development for improved surveillance of the Great Crested Newt Appendix 5 . Technical advice note for field and laboratory sampling of great crested newt (Triturus cristatus) environmental DNA.

Darwall, W.R.T., Smith, K.G., Allen, D.J., Holland, R.A, Harrison, I.J., and Brooks, E.G.E. (eds.). 2011. The Diversity of Life in African Freshwaters: Under Water, Under Threat. An analysis of the status and distribution of freshwater species throughout mainland Africa. Cambridge, United Kingdom and Gland, Switzerland: IUCN. xiii+347pp+4pp cover.

Deagle, B.E., Eveson, J.P., Jarman, S.N., 2006. Quantification of damage in DNA recovered from highly degraded samples--a case study on DNA in faeces. Front. Zool. 3, 11. doi:10.1186/1742-9994-3-11.

Dejean, T., Valentini, A., Duparc, A., Pellier-Cuit, S., Pompanon, F., Taberlet, P., Miaud, C., 2011. Persistence of environmental DNA in freshwater ecosystems. PLoS One 6, e23398. doi:10.1371/journal.pone.0023398.

Drummond, A.J., Suchard, M. a, Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29, 1969–73. doi:10.1093/molbev/mss075.

Goldberg, C.S., Pilliod, D.S., Arkle, R.S., Waits, L.P., 2011. Molecular detection of vertebrates in stream water: a demonstration using Rocky Mountain tailed frogs and Idaho giant salamanders. PLoS One 6, e22746. doi:10.1371/journal.pone.0022746.

Goldberg, C.S., Sepulveda, A., Ray, A., Baumgardt, J., Waits, L.P., 2013. Environmental DNA as a new method for early detection of New Zealand mudsnails (Potamopyrgus antipodarum). Freshw. Sci. 32, 792–800. doi:10.1899/13-046.1.

Herder, J.E., Valentini, A., Bellemain, E., Dejean, T., Delft, J.J.C.W. van, Thomsen, P.F., Taberlet, P., 2014. Environmental DNA - a review of the possible applications for the detection of (invasive) species. doi:10.1111/j.1365-294X.2012.05542.x.

Jerde, C.L., Chadderton, W.L., Mahon, A.R., Renshaw, M.A., Corush, J., Budny, M.L., Mysorekar, S., Lodge, D.M., Bay, S., 2013. Detection of Asian carp DNA as part of a Great Lakes basin-wide 526, 522–526.

Jones, W.J., Preston, C.M., Marin Iii, R., Scholin, C. a, Vrijenhoek, R.C., 2008. A robotic molecular method for in situ detection of marine invertebrate larvae. Mol. Ecol. Resour. 8, 540–50. doi:10.1111/j.1471-8286.2007.02021.x.

Lance, R.F., Carr, M.R., 2012. Detecting eDNA of Invasive Dreissenid Mussels : Report on Capital Investment Project.

Lowenstein, J.H, Osmundson, T.W., Becker, S., Hanner, R., Stiassny, M.L., 2011. Incorporating DNA barcodes into a multi-year inventory of the fishes of the hyperdiverse Lower Congo River, with a multi-gene performance assessment of the genus Labeo as a case study. Mitochondrial DNA Suppl 1:52-70. doi: 10.3109/19401736.2010.537748.

Mahon, A.R., Jerde, C.L., Galaska, M., Bergner, J.L., Chadderton, W.L., Lodge, D.M., Hunter, M.E., Nico, L.G., 2013. Validation of eDNA surveillance sensitivity for detection of Asian carps in controlled and field experiments. PLoS One 8, e58316. doi:10.1371/journal.pone.0058316.

Minamoto, T., Yamanaka, H., Takahara, T., Honjo, M.N., Kawabata, Z., 2011. Surveillance of fish species composition using environmental DNA. Limnology 13, 193–197. doi:10.1007/s10201-011-0362-4.

Miya, M., Nishida, M., 2000. Use of mitogenomic information in teleostean molecular phylogenetics: a tree-based exploration under the maximum-parsimony optimality criterion. Mol. Phylogenet. Evol. 17, 437–55. doi:10.1006/mpev.2000.0839.

Pilliod, D.S., Goldberg, C.S., Arkle, R.S., Waits, L.P., 2014. Factors influencing detection of eDNA from a stream-dwelling amphibian. Mol. Ecol. Resour. 14, 109–16. doi:10.1111/1755-0998.12159.

Pons, J., Barraclough, T., Gomez-Zurita, J., Cardoso, A., Duran, D., Hazell, S., Kamoun, S., Sumlin, W., Vogler, A., 2006. Sequence-Based Species Delimitation for the DNA Taxonomy of Undescribed Insects. Syst. Biol. 55, 595–609. doi:10.1080/10635150600852011.

Rees, H.C., Maddison, B.C., Middleditch, D.J., Patmore, J.R.M., Gough, K.C., 2014. The detection of aquatic animal species using environmental DNA - a review of eDNA as a survey tool in ecology. J. Appl. Ecol. n/a–n/a. doi:10.1111/1365-2664.12306.

Taberlet, P., Coissac, E., Hajibabaei, M., Rieseberg, L.H., 2012. Environmental DNA. Mol. Ecol. 21, 1789–93. doi:10.1111/j.1365-294X.2012.05542.x.

Thomsen, P.F., Kielgast, J., Iversen, L.L., Møller, P.R., Rasmussen, M., Willerslev, E., 2012a. Detection of a diverse marine fish fauna using environmental DNA from seawater samples. PLoS One 7, e41732. doi:10.1371/journal.pone.0041732.

Thomsen, P.F., Kielgast, J., Iversen, L.L., Wiuf, C., Rasmussen, M., Gilbert, M.T.P., Orlando, L., Willerslev, E., 2012b. Monitoring endangered freshwater biodiversity using environmental DNA. Mol. Ecol. 21, 2565–73. doi:10.1111/j.1365-294X.2011.05418.x.

Valentini, A., Pompanon, F., Taberlet, P., 2009. DNA barcoding for ecologists. Trends Ecol. Evol. 24, 110–7. doi:10.1016/j.tree.2008.09.011.

WWF, 2014. Living Planet Report 2014: species and spaces, people and places. [McLellan, R., Iyengar, L., Jeffries, B. and N. Oerlemans (Eds)]. WWF, Gland, Switzerland.