Naturhistoriska riksmuseet

Slutrapport

Datum
2015-04-27

Dnr
4.1–511-2014 (NRM)
2014.5.1-249 (SLU.aqua)



*Centrum för genetisk identifiering*

## Slutrapport om uppdrag

# Test av eDNA-metodik för miljöövervakning av fisk, kräftor och stormusslor i insjöar

*Rodrigo Esparza-Salas*

*Centrum for Genetisk Identifiering*

*Naturhistoriska Riksmuseet*

Naturhistoriska riksmuseet

Postadress:
Box 50007
104 05 Stockholm

Besöksadress:
Frescativägen 40
114 18 Stockholm

Telefon: 08-519 540 00
Telefax. 08-519 540 85
registrator@nrm.se

# Final report

# Test of eDNA methodology for environmental monitoring of fish, crayfish and large bivalves in lakes

**Rodrigo Esparza-Salas**

**Centre for Genetic Identification**
**Swedish Museum of Natural History**

## Summary

A series of methodologies for the use of DNA-based environmental monitoring of fish, crayfish and mussels in water bodies were developed. New DNA barcoding markers were developed for the simultaneous detection of fish and selected mussel species, as well as species-specific markers for the detection of two ecologically important crayfish species. In addition, DNA barcoding markers that were previously developed for the identification of fish in seal diet were tested in extracts from filtrated water samples. The different barcoding markers were tested in known DNA extracts of different target species, as well as in DNA extracted from water samples collected in different freshwater environments in central Sweden. The identification of such organisms in DNA extracts from filtered water was possible, although with very limited success. A series of possible difficulties to overcome the limitations in the use of the barcoding markers in eDNA studies are discussed, as well as possible steps to the eventual use of the methods described here in environmental monitoring activities in the near future.

# Test of eDNA methodology for environmental monitoring of fish, crayfish and large bivalves in lakes

## Background

The institution for aquatic resources at the Swedish University of Agricultural Sciences (Sveriges Lantbruksuniversitet; hereby SLU-aqua) requested the service of the Centre for Genetic Identification at the Swedish Museum of Natural History (Naturhistoriska Riksmuseet; hereby NRM) in the development and testing of environmental DNA techniques (hereby eDNA) in the study and monitoring of fish, crayfish and large mussel populations in freshwater environments in Sweden.

Although the use of eDNA methodology (defined as DNA obtained from environmental samples such as ice, water, soil or air) has been gaining popularity in environmental monitoring, there are no universal standards for its application. When successfully applied, eDNA methodology can be a source of information on diversity of extant organisms in an environment, both qualitatively and quantitatively (Pedersen et al. 2015). The use of DNA barcoding (a method that uses a short fragment of the DNA of an organism to assign it to a particular taxon) allows the simultaneous identification of related taxa in an eDNA sample. Additionally, single species can be detected in an eDNA sample by using species specific DNA markers.

## Methods

The survey consisted of different phases:

- Design and selection of DNA markers for each of the target groups
- Collection and preservation of samples in the field
- Sample filtration
- DNA extraction from sample filtrates
- Test of DNA markers for each of the target groups and a priori testing
- PCR amplification of DNA barcoding markers
- DNA sequencing
- Matching sequence data against a reference DNA database of known taxa.

Each of these steps conveys technical challenges, which can be intensified due to the particular limitations of environmental DNA. Some of the potential limitations in the application of eDNA in ecological studies include:

- Difficulties in sample preservation in the field
- Inappropriate filtration of the sample
- Low concentration of DNA in the environmental sample
- The presence of PCR inhibitors in the sample
- Absence of target taxon/barcoding marker data in DNA reference databases

With technical limitations in consideration, field and laboratory protocols were agreed on, in order to maximise the quality of the results.

### Barcoding primer design

We used DNA barcoding techniques to aid in the detection and identification of species for this project. Barcoding genes are relatively short fragments of DNA that are conserved within a species, while differing between species. Sequences of DNA barcoding genes have been constantly obtained from a wide range of plants and animals, which has helped building of reference DNA libraries that can be used in the DNA-based identification of species. In this project we used barcoding genes of mitochondrial origin. The use of mitochondrial genes is preferred in eDNA methodology because the volume of mitochondrial DNA recovered is usually much larger than that of nuclear DNA. In the case of fish and mussels, we used group specific barcoding markers that allow the simultaneous amplification of DNA from different species of the respective group. For the two species of crayfish (European crayfish, *Astacus astacus*; and Signal crayfish, *Pacifastacus leniusculus*), we designed species-specific markers that are better suited for the detection of presence or absence of the respective species, avoinding false identification of the other.

The 16s gene is widely used in eDNA studies that require amplification of different related organisms in a sample. It has advantages over other barcoding markers. Namely it is often possible to design markers that are conserved in a taxonomic group (e.g. fish or mussels in this project) while targeting short DNA fragments that differ in each species. Therefore one set of markers can be used to a wide range of species. In addition, assuming that all organisms present in a sample contain similar number of mitochondria, and that all mitochondrial genomes have the same probability to be amplified during the PCR reaction, 16s can be used as a semi-quantitative measure of diversity by combining PCR with next-generation sequencing (see below).

The COI gene is perhaps the most widely used barcoding marker for animals and other eukaryotes. There are universal markers available for a wide range of animal groups. The length of the barcoding COI DNA fragment has limited use in eDNA and other methods that rely on limited and fragmented DNA. However, since the COI sequence is unique for each animal species, it is well suited for the

design of shorter species-specific markers for presence/absence detection, as well as for quantitative assessment of species DNA in a sample.

We used previously developed 16s genetic markers for the identification of fish species, that have been used successfully in the study of diet in Grey seals (*Halichoerus grypus*) was used (Lundström et al., unpublished data). The primers amplify an approximately 250 base pair region of the mitochondrial 16s rDNA gene. The resolution in the identification of some fish groups was limited, namely cyprinids (Cyprinidae) and whitefish species (*Coregonus spp.*) cannot be identified to the genus or species level. For this reason, a complementary set of fish primers was designed in a different section of the 16s gene which overlaps slightly with the first set of fish primers, and which would improve the resolution in species identification for closely related species, such as cyprinid fishes (Cyprinidae). In the case of Crayfish, barcoding markers were designed in sections that were conserved in one species, but different in the other species, as to avoid undesired cross-amplification.

The use of eDNA methodology poses important challenges. Namely, the amount of DNA that can be isolated from an environmental sample (e.g. soil, soft bottoms, water, air or diet sample) is limited, particularly for target species that do not comprise a significant part of the biomass of the sampled environment; and the recovered DNA tends to be degraded as a result of environmental exposure, enzymatic degradation and other factors. These challenges were taken into account in the design of field and laboratory protocols for the objectives of the project.
A prior evaluation of the designed markers was performed by testing PCR reactions using reference DNA extracted from known individuals of all the target mussel and crayfish species.

**Field sampling.**

Five liter water samples were collected at six different locations: Norasjön (n=10), Svartälven (n=10), Svennedamsån (n=10), Stensjön (n=10), Öresjön (n=10) and Mälaren (n=5). All samples were filtered on location using a vacuum system which consisted of disposable 0,45 µm filter units (MoBio laboratories Inc., Carlsbad CA, USA) connected to a filtering flask and a water aspirator that provided the vacuum. The filter membrane was kept in 5ml tubes and kept frozen for subsequent extraction in the laboratory. The samples were taken at different depths in the respective body of water, as to maximise the detection of species.

**Laboratory analyses**

*DNA extraction.*

DNA was isolated using the PowerWater DNA isolation kit (MoBio laboratories), following the manufacturer protocols with a previous breakage treatment of the filters with zirconia-silica beads for maximising recovery of DNA. Each sample was kept in 100 uL of the elution buffer provided in the DNA isolation kit.

*PCR amplification*

All target DNA fragments were amplified using polymerase chain reaction (hereby PCR) using ready-to go Hot start Illustra PCR beads (GE Healthcare) in a final volume of 25 µl containing 0.4 nmol of each primer (Table 1), and 2 µl of DNA extract. Thermal cycling consisted of an initial denaturing of five minutes; followed by 40 cycles of 94 degrees C for 15 seconds, annealing at 51 degrees for 15 seconds and extension at 72 degrees for 30 seconds. A final extension of 10 minutes at 72 degrees was added at the end of the thermal cycling.

**Table 1**. PCR primers used in this survey. The individual barcodes are represented by NNNNNN.

| Primer name | Target | | Sequence 5’ – 3’ | Reference |
|---|---|---|---|---|
| 16sPrey | Vertebrate 16s rDNA gene | F: | NNNNNNcgtgcraaggtagcg | Lundström et al. (Unpublished) |
| | | R: | NNNNNNcctygggcgccccaac | |
| 16sKarp | Fish 16s rDNA gene | F: | NNNNNN | This study |
| | | R: | NNNNNN | |
| Mussel16s | Mussel (Unionida) 16s rDNA gene | F: | NNNNNNtgcatcatggagaagcaac | This study |
| | | R: | NNNNNNccataagctagcactttgatgc | |
| AstacusCOI | Noble crayfish COI gene | F: | ttttgattgctcccctttc | This study |
| | | R: | tcgaagatacacctgccaag | |
| PacifastacusCOI | Signal crayfish COI gene | F: | aagattttgattacttccattttcttt | This study |
| | | R: | gaagaaacacccgctaaatga | |

Every sample extract was amplified using a unique combination of forward and reverse primers that contained a six base-pair long sequence of nucleotides that served as “individual barcodes” for subsequent de-multiplexing of the DNA sequences from the pyrosequencing output (see below). The

success of the DNA amplification was evaluated by gel electrophoresis. Three microliters of the PCR product were run on 1% agarose gels stained with GelRed (Biotium Inc.) and visualised under UV light.

*Sequencing*

Prior to sequencing, all PCR products were purified using a SequalPrep kit (Invitrogen/ Life Technologies). The kit retains up to 25 ng of each PCR product, thus avoiding some of the stronger PCR samples from being overrepresented in the final dataset.

The use of next-generation sequencing technology (hereby NGS), apart from producing tens of thousands of sequences in parallel, can provide a quantitative dimension to multiple-species DNA studies. In this project we evaluated the usefulness of NGS methods in the qualitative and quantitative estimation of diversity in eDNA samples.

The PCR products of fish and mussel 16s genes were pooled together and sequenced in parallel using next generation sequencing/ Pyrosequencing, on a GS-Junior instrument (454 Life Sciences/Roche, Branford, CT, USA). Sequencing libraries were prepared according to the Lib-l protocols for adaptor ligation. The different PCR reactions (fish and mussel 16s genes, respectively) were labelled with a different MID primer to facilitate data sorting. Emulsion PCR was performed according to the protocols for the GS-Junior instrument, with the addition of one molecule per bead. Sequencing was performed according to the current GS-Junior protocols.

**Data analyses**

De-multiplexing of the individual PCR products was performed according to the previously assigned combination of forward and reverse individual barcodes, using the program find_tag_pairs (Johan Nylander, personal communication). In order to perform species identification, All DNA sequences in the output of the pyrosequencing run were matched against the GeneBank database using the MegaBLAST algorithm implemented in the BLAST+ package (Morgulis et al. 2008 : http://blast.ncbi.nlm.nih.gov/).

.

## Results and discussion

### *DNA marker performance in reference samples*

The newly designed barcoding primers for mussels tested positive for all the reference samples of large mussels, with the exception of the zebra mussel (figure 1 A) which belongs to a separate subclass of bivalves. Out of the two sets of mussel markers tested, we chose the shorter one that performed best for subsequent analyses. The specific crayfish primers tested positive for their respective target species and failed to amplify in the non-target species (figure 1 B).

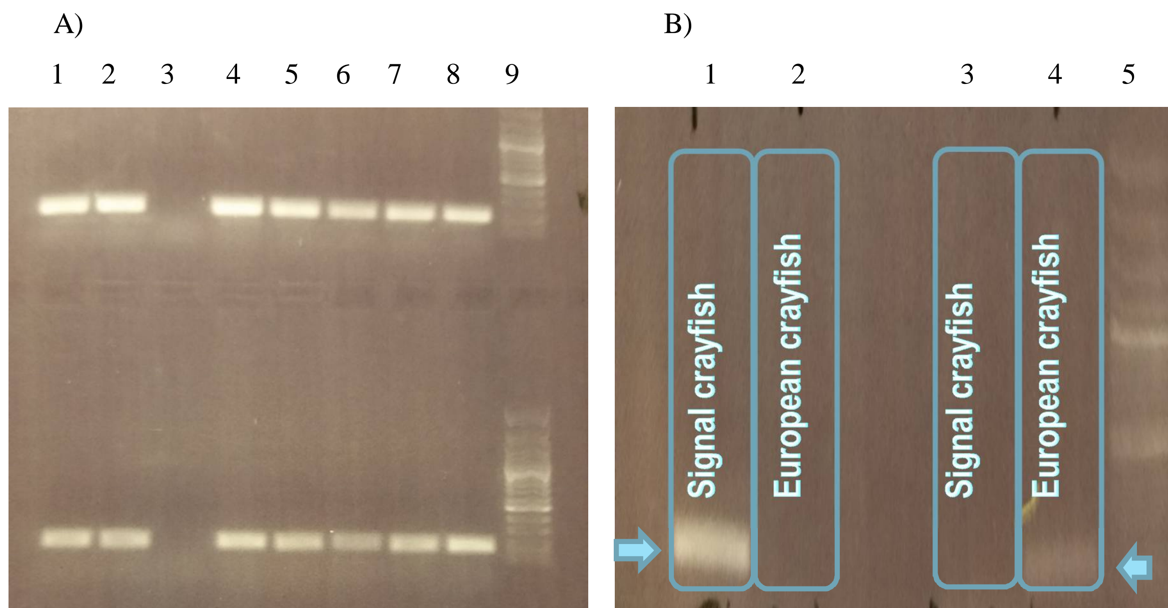


**Figure 1**. Agarose gel images of amplified PCR products of barcoding genes visualised under UV light. Positive PCR reactions are shown as bright bands on the gel image. Negative PCRs fail to produce a band.

A) Two different sections of the 16s gene in Large mussels:1) Duck mussel (*Anodonta anatina*); 2) Swan mussel (*Anodonta cygnea*); 3) Zebra mussel (*Dreissena polymorpha*); 4) Freshwater pearl mussel (*Margaritifera margaritifera*); 5)Depressed river mussel (*Pseudanodonta complanata*); 6) Thick-shelled river mussel (*Unio crassus*); 7) Painter's mussel (*Unio pictorum*); 8) Swollen river mussel (*Unio tumidus*). ); 9) DNA size reference.

B) A shorter section of the COI gene in Crayfish: Lanes 1 and 2 use specific signal crayfish marker; lanes 3 and 4 use specific European crayfish marker. Lane 5 is a DNA size reference. Positive PCR products are indicated with arrows.

### *DNA marker performance in eDNA samples*

The success of the barcoding markers varied when tested on the eDNA samples (Figure 2): The fish marker showed 12 positive products out of 55, while the mussel marker showed 13 out of 55 positive products. The second set of fish primer designed during this study, and the crayfish markers failed to produce any positive results.

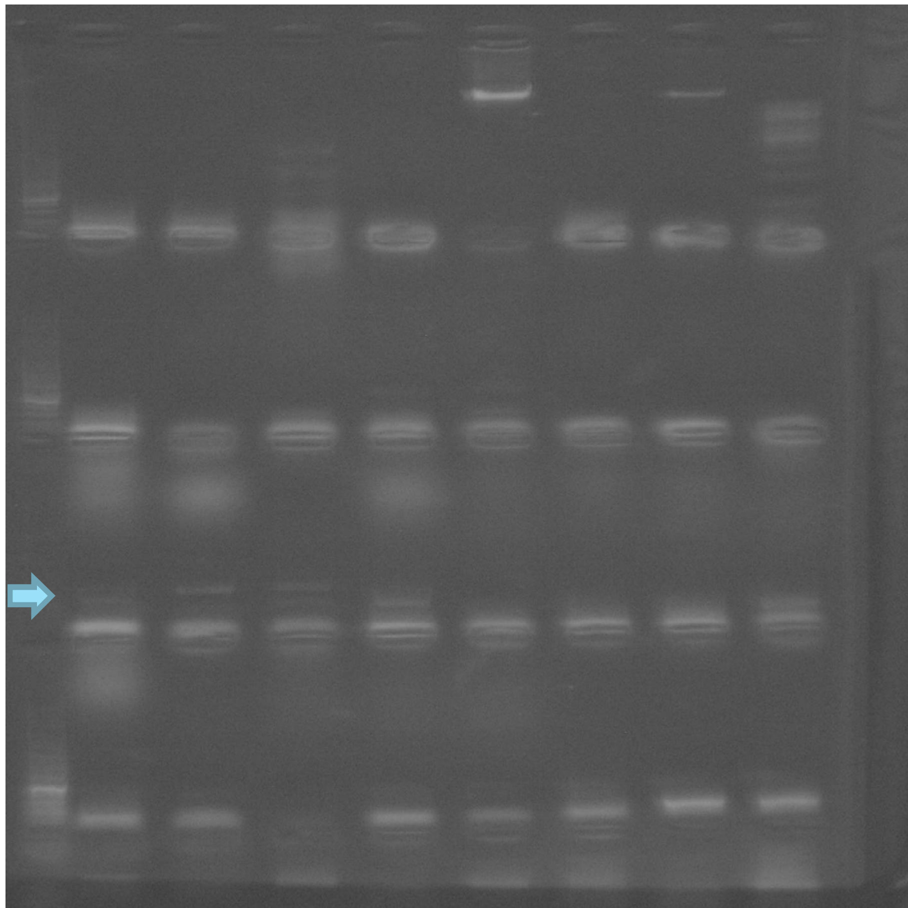

**Figure 2**. Example of PCR amplification of environmental DNA samples with mussel barcoding markers. The faint bands indicated in line with the arrow correspond to the expected positive PCR products. Other diffuse bands of different sizes correspond to non-specific amplification of other organisms present in the samples.

*Parallel sequencing*

The parallel pyrosequencing produced a large amount of non-specific DNA sequences, corresponding to DNA of other organisms present in the samples and a large proportion of non-identified DNA. Some of the organisms identified among the sequences included various bacteria, vascular plants, algae, protozoans, fungi, cyanobacteria and various animal groups including planktonic crustaceans, and annelids.

A small proportion of the sequences (close to 1% of the total) corresponded to six target fish species (European bullhead *Cottus gobio*; Pike *Esox Lucius*; Ruffe *Gymnocephalus cernuus*; Perch *Perca fluviatilis*; Roach *Rutilus rutilus*; and Tench *Tinca tinca*) and one species of mussel (Swollen river mussel *Unio tumidus*). All such fish and mussel species were detected in Svennevadsån, and one fish species (Roach) was detected in Mälaren. The rest of the samples failed to produce any target species sequences (Table 1).

**Table 2**. Presence of fish and mussel species in DNA extractions from different freshwater bodies, indicated with an "x"

| | Unio tumidus | Cottus gobio | Esox lucius | Gymnocephalus cernuus | Perca fluviatilis | Rutilus rutilus | Tinca tinca |
|---|---|---|---|---|---|---|---|
| Mälaren | | | | | | x | |
| Norasjön | | | | | | | |
| Öresjön | | | | | | | |
| Stensjön | | | | | | | |
| Svartälven | | | | | | | |
| Svennevadsån | x | x | x | x | x | x | x |

The failure to obtain a greater success in the amplification can be attributed to different factors:

*Sample filtration:*

It is possible that the 0.45 µm pore size of the filter membranes used was not able to retain small fragments of DNA in the samples. In a test performed subsequently in our laboratories for a different project, we noticed that water samples filtered with a 0,22 µm membrane performed better than a 0,45 µm membrane (data not shown), in spite of blocking the filter after a volume of only one liter. This suggests that a large proportion of the useful DNA might have been lost during filtration. However, the filtration of five liters of water through a 0,22 µl membrane would be impractical in the field because the filters tends to block rapidly. A better solution would be to perform filtration in two steps, where the larger water volume would be passed through a larger filter, to retain large particles, seeds, plant material and other debris, and the filtrate of this first step can be passed to a 0,22 µm membrane for collection of DNA.

*DNA degradation:*

DNA in water samples can be rapidly degraded due to exposure to UV light, enzymatic activity in the water and other factors. The fragments of DNA amplified in this study were 250 to 300 base pairs long, which might be too large for some fragmented DNA. Using other markers to amplify smaller

sections of the 16s genes might be more useful for environmental samples, but at the same time the resolution of species identification can be compromised.

*Absence of target species in the sample:*

Although many species of fish, mussel and crayfish have been detected previously in the sampled bodies of water, it is practically impossible to predict whether traces of any of such species will be present in the relatively small volumes of water sampled. Some organisms might be easier to identify because either their relative biomass is larger; they live in close contact with the particular sampled location in the lake or river (e.g. bottom vs. surface); or because they shed traces of DNA more easily than other organisms, in the form of mucosa, eggs and scats amongst others. By sampling larger volumes or a larger number of samples, the probability of detection is increased, although such a scheme must be evaluated in terms of budget and logistic restrictions.

*Limitations of the sequencing techniques:*

The DNA sequencing technology has advanced considerably in recent years. The Pyrosequencing methodology used in this study, although it has been used successfully, might not be the most efficient in sequencing a large amount of organisms in mixed samples such as eDNA samples. More recently developed techniques can produce outputs up to 200 times larger, which can be more informative in cases where the target organism DNA is limited.

## Conclusions

Even though the success of the techniques used in this study was limited, it is clear that they can be a step towards designing standardised field protocols for the use of eDNA methodologies in environmental monitoring. And even considering all the limitations, we were able to detect target species to some degree, which suggest that after correcting for other factors the usefulness of these techniques is improved. The methods described here can easily be adapted to new sequencing technologies, which will allow the creation of efficient DNA-based environmental monitoring protocols in the near future.

## References

Morgulis A et al. 2008. Database indexing for production MegaBLAST searches. Bioinformatics 24: 1757-1764

Pedersen MW et al. 2015. Ancient and modern environmental DNA. Philosophical Transactions of the Royal Society B: Biological Sciences. 370: 20130383

## Electronic Appendixes

Alignment of 16s gene sequences of Cyprinid fishes for primer design, in FastA format:
Mito_16s_Cyprinid.FAS

Alignment of 16s gene sequences of large Mussels found in Sweden for primer design, in FastA format:
16s_stormusslor.FAS

Alignment of COI Crayfish sequences for specific primer design, in FastA format:
KräftorCOI.FAS

Summary of Next-generation sequencing results:
APPENDIX04.xlsx

Raw data from Next-generation sequencing run in FastA format:
eDNAReads.fas