

Species detection using environmental DNA from water samples

Gentile Francesco Ficetola^{1,2,*}, Claude Miaud², François Pompanon¹ and Pierre Taberlet¹

¹Laboratoire d'Ecologie Alpine, CNRS-UMR 5553, Université Joseph Fourier, BP 53, 38041 Grenoble Cedex 09, France

²Laboratoire d'Ecologie Alpine, CNRS-UMR 5553, Université de Savoie, 73376 Le Bourget du Lac Cedex, France

*Author and address for correspondence: Dipartimento di Scienze dell'Ambiente e del Territorio, Università Milano Bicocca, Piazza della Scienza 1, 20126 Milano, Italy (francesco.ficetola@unimi.it).

The assessment of species distribution is a first critical phase of biodiversity studies and is necessary to many disciplines such as biogeography, conservation biology and ecology. However, several species are difficult to detect, especially during particular time periods or developmental stages, potentially biasing study outcomes. Here we present a novel approach, based on the limited persistence of DNA in the environment, to detect the presence of a species in fresh water. We used specific primers that amplify short mitochondrial DNA sequences to track the presence of a frog (*Rana catesbeiana*) in controlled environments and natural wetlands. A multi-sampling approach allowed for species detection in all environments where it was present, even at low densities. The reliability of the results was demonstrated by the identification of amplified DNA fragments, using traditional sequencing and parallel pyrosequencing techniques. As the environment can retain the molecular imprint of inhabiting species, our approach allows the reliable detection of secretive organisms in wetlands without direct observation. Combined with massive sequencing and the development of DNA barcodes that enable species identification, this approach opens new perspectives for the assessment of current biodiversity from environmental samples.

Keywords: biodiversity inventories; biological invasion; conservation genetics; DNA barcoding; secretive species

1. INTRODUCTION

The assessment of species distribution is a first critical phase of biodiversity studies and is necessary for several disciplines such as biogeography, conservation biology and ecology (Margurran 2004). However, several species are difficult to detect, especially during particular time periods or developmental stages, potentially biasing study outcomes (Gotelli & Colwell 2001; MacKenzie *et al.* 2006). The extraction of DNA from environmental samples allows the characterization of their micro-organisms (Venter *et al.* 2004). It can also provide information on extinct communities of macro-organisms, since short DNA sequences can persist for

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rsbl.2008.0118> or via <http://journals.royalsociety.org>.

long time periods, as shown by the studies on old sediments, permafrost and ice cores (Hofreiter *et al.* 2003; Willerslev *et al.* 2003, 2007). While short DNA sequences may be present at high density in the environment, their potential for the study of present-day communities of macro-organisms remains substantially unexplored. Here we present a novel approach, based on the persistence of DNA in the environment, to detect the presence of a species in fresh water. We examined whether DNA fragments are preserved in the aquatic environment and whether they can be used for a reliable assessment of current species presence. We first used the method in controlled environments and then evaluated whether it could be applied under natural field conditions. The model species was the American bullfrog *Rana catesbeiana* (= *Lithobates catesbeianus*), an invasive amphibian for which high-quality census data exist (Ficetola *et al.* 2007a,b). This allowed reliable field validation. The American bullfrog is native to western North America, but has been introduced into ecosystems around the globe. It is considered one of the world's most harmful invasive species, since it is responsible for the decline of native amphibians by direct predation, competition, diffusion of diseases and complex biotic interactions (Blaustein & Kiesecker 2002; Kats & Ferrer 2003; Garner *et al.* 2006).

2. MATERIALS AND METHODS

We performed experiments in both controlled conditions and natural populations. Tadpoles were reared in aquariums filled with 3 l of water, collected in a natural alpine spring at 1000 m above sea level and at 80 km from the nearest bullfrog record. We used 0, 1, 5 and 10 tadpoles per aquarium; each density was replicated six times. After 24 hours, we collected a 15 ml water sample from each aquarium.

Bullfrog distribution in natural wetlands was assessed in France through traditional surveys of more than 2500 wetlands over 4 years (Ficetola *et al.* 2007b). We selected three ponds (surface 1000–10 000 m²) where bullfrogs were present at low density (one to two adults seen, no reproduction), three ponds where bullfrogs were present at high density (dozens of adults and thousands of tadpoles) and three ponds where bullfrogs have never been detected. These latter ponds were more than 30 km from the nearest bullfrog record. For each pond, we collected three 15 ml water samples from different points.

Immediately after collection, we added 1.5 ml of sodium acetate 3 M and 33 ml absolute ethanol to the water samples; these were then stored at –20°C until DNA extraction. To recover the precipitated DNA and/or the cellular remains, we centrifuged the mixture (5500g, 35 min, 6°C) and discarded the supernatant (Valiere & Taberlet 2000). The pellet was then subjected to a classical DNA extraction using QIAamp Tissue Extraction Kit (Qiagen).

Polymerase chain reaction (PCR) amplification was performed with primers 5'-TGCCAACGGAGCATCATTC-3' and 5'-ATAAAGGTAGGAGCCGTAGT-3' especially designed for this experiment. These primers amplify a 79 bp segment of mitochondrial *cyt-b*, which is monomorphic in all 397 individuals analysed by population genetic studies covering the whole native and European range of the species (Ficetola *et al.* 2008). The basic local alignment search tool showed that these primers do not match with high scores to any other sequences stored in GenBank. Moreover, we used these primers to try amplifying tissue-extracted DNA of all other frog species living in France (*Rana ridibunda*, *Rana kleptonesculenta*, *Rana lessonae*, *Rana dalmatina*, *Rana temporaria*), with at least two individuals from different localities per each species. None of these tests resulted in DNA amplification. The PCRs were conducted in a 25 µl total volume with 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 2.5 mM MgCl₂, 0.8 µM of each primer, BSA (5 µg), 0.5 U of AmpliTaq Gold DNA polymerase (Applied Biosystems) and 5 µl of DNA extract. The amplification of each water sample was repeated three to five times, using the multi-tube approach (Taberlet *et al.* 1996), which gave 9–15 repeats per pond. The number of replicates was similar in ponds with different bullfrog densities (mixed model, $F_{2,6}=0.506$, $p=0.51$). The PCR programme included an initial 10 min denaturation step at 95°C, 55 cycles of denaturation at 95°C for 30 s and annealing at

61°C for 30 s. PCR products were visualized using electrophoresis on 2% agarose gel. Sequencing reactions were performed in both directions using BigDye Terminator Cycle Sequencing Kit v. 1.1 and following standard procedures. Moreover, the PCR product of one pond was sequenced using the 454 pyrosequencing technology (GS20, Roche, Basel; Margulies *et al.* 2005). This massive parallel sequencing technique allows a large number of sequences to be obtained from a single PCR product, comparable with large-scale cloning protocols.

We used generalized mixed models assuming a binomial error to compare the amplification rates among ponds with different bullfrog densities. The identities of ponds and water samples were included as random factors, nested within density. Mixed models were fitted using lme4 in R (R Development Core Team 2006) and followed by *post hoc* orthogonal contrasts through the origin.

3. RESULTS

Using selective primers and PCR, we successfully amplified DNA from water samples of all 18 aquariums where we reared tadpoles at different densities (0.3, 1.7, 3.3 tadpoles per l), while we never amplified DNA from the aquariums where bullfrog tadpoles were absent. We also amplified DNA from water samples of all the natural ponds where the species was present, while we never amplified DNA from the ponds where bullfrogs were absent (table 1). Almost all PCR products were sequenced and corresponded perfectly to the published bullfrog *cyt-b* sequence (Austin *et al.* 2004). Moreover, 673 fragments from one PCR product were sequenced using the 454 pyrosequencing technology (Margulies *et al.* 2005). Apart from clearly recognizable sequencing and PCR errors (mostly in a poly A repeat), all obtained sequences perfectly matched the bullfrog *cyt-b*, indicating that the PCR products were not mixes of different sequences. The probability that all nine replicates from a pond with a low density of frogs would be negative (i.e. false negative) is approximately 1.5% (see below).

These results were obtained using a multi-tube approach and ancient DNA precautions, which are suitable for analysing DNA that is degraded and/or at low concentrations (Taberlet *et al.* 1996). In water samples collected in ponds where bullfrogs were present, 22–89% of replicates were positive (table 1). The average amplification success (\pm s.e.) was 0 in ponds where bullfrogs were absent, 0.37 ± 0.1 in ponds where bullfrogs were present at low densities and 0.79 ± 0.08 in ponds where bullfrogs were present at high densities. Owing to the protocol used (multi-tube approach and high number of PCR cycles), it is necessary to ensure that the positive amplifications were not due to artefacts. This requires that all controls be negative and that this lack of amplification not be due to chance. Differences in amplification rates among ponds with differing densities of target species were significant (generalized linear mixed model, $\chi^2_2 = 19.5$, $p < 0.0001$). The amplification success was significantly higher in ponds with bullfrogs than in control ponds (orthogonal contrast, $\chi^2_1 = 18.2$, $p < 0.0001$); it is thus unlikely that all the control samples were negative just by chance. Moreover, the amplification rate was significantly higher in ponds with high bullfrog densities than in ponds with low densities ($\chi^2_1 = 5.7$, $p = 0.017$).

Table 1. Rate of bullfrog detection in water samples.

pond	bullfrog presence and relative abundance	water samples positives at least once	positive PCRs
1	yes-low	2/3	2/9
2	yes-low	3/3	6/9
3	yes-low	2/3	2/9
4	yes-high	3/3	8/9
5	yes-high	3/3	6/9
6	yes-high	3/3	8/10
7	no	0/3	0/9
8	no	0/3	0/9
9	no	0/3	0/15

4. DISCUSSION

We showed that environmental DNA (either in solution or in cellular debris) can be used to ascertain species presence in a wetland, and that this technique is able to discriminate between absence and presence, even at low densities. As the environment can retain the molecular imprint of inhabiting species (Hofreiter *et al.* 2003), our approach allows the reliable detection of secretive organisms in wetlands without direct observation. This is possible because environmental DNA can be detected at very low concentrations (see electronic supplementary material). The same approach could be useful for studying secretive aquatic or semi-aquatic species, which release DNA into the environment through mucus, faeces, urine and remains.

This approach can be the answer to many situations where traditional census techniques give low-quality results and/or require a huge sampling effort. This is the case when trying to quantify secretive harmful, invasive (such as bullfrog) or threatened species. Detecting invasive species at the early stages of invasions, and when they are at low densities, is the key to timely interventions to control them (Hulme 2006). Moreover, this technique allows the assessment of distribution of rare threatened species that are the target of conservation plans, or of harmful species that are kept at low density by management, with a reduced monitoring effort.

The use of such a method, like all techniques dealing with the detection of DNA traces, requires several precautions (Taberlet *et al.* 1996; Cooper & Poinar 2000). First, several factors could affect the amount of DNA in environmental samples, such as volume of water, size and density of the organism and volume of secretions. Second, as it is difficult to evaluate how long DNA fragments persist in water, this method could lead to the detection of a species after it has left the wetland. Short DNA fragments can persist a long time under dry cold conditions and in the absence of light. In one extreme example, DNA from extinct vertebrates has been amplified from 10 000-year old dry cave sediments (Willerslev *et al.* 2003). Actually, DNA fragments of approximately 400 bp may persist up to one week at 18°C in lake water (Matsui *et al.* 2001). This suggests that detecting a species that is no longer present is unlikely, owing to both the DNA's fast decomposition rate and its degradation from UV radiation exposure.

However, the presence of cells or particle-bound DNA may lead to a longer persistence of detectable DNA in water.

In addition to its applications for the study and management of secretive freshwater species, this approach opens new avenues for the study of biodiversity. The ongoing effort to develop DNA barcodes for identifying species from degraded DNA (Hajibabaei *et al.* 2006; Taberlet *et al.* 2007) will make our approach applicable to more and more plant and animal species. Moreover, massive sequencing techniques could be used to analyse PCR products generated with universal primers working on degraded substrates (e.g. Hofreiter *et al.* 2003; Willerslev *et al.* 2003; Taberlet *et al.* 2007). These factors will soon make possible the assessment of the current biodiversity of macro-organisms from environmental samples.

G.F.F. was partially founded by a grant of the French Ministry for Research for young foreign researchers.

- Austin, J. D., Lougheed, S. C. & Boag, P. T. 2004 Discordant temporal and geographic patterns in maternal lineages of eastern north American frogs, *Rana catesbeiana* (Ranidae) and *Pseudacris crucifer* (Hylidae). *Mol. Phylogenet. Evol.* **32**, 799–816. (doi:10.1016/j.ympev.2004.03.006)
- Blaustein, A. R. & Kiesecker, J. M. 2002 Complexity in conservation: lessons from the global decline of amphibian populations. *Ecol. Lett.* **5**, 597–608. (doi:10.1046/j.1461-0248.2002.00352.x)
- Cooper, A. & Poinar, H. N. 2000 Ancient DNA: do it right or not at all. *Science* **289**, 1139–1139. (doi:10.1126/science.289.5482.1139b)
- Ficetola, G. F., Thuiller, W. & Miaud, C. 2007a Prediction and validation of the potential global distribution of a problematic alien invasive species—the American bullfrog. *Divers. Distrib.* **13**, 476–485. (doi:10.1111/j.1472-4642.2007.00377.x)
- Ficetola, G. F., Coïc, C., Detaint, M., Berroneau, M., Lorvelec, O. & Miaud, C. 2007b Pattern of distribution of the American bullfrog *Rana catesbeiana* in Europe. *Biol. Invasions* **9**, 767–772. (doi:10.1007/s10530-006-9080-y)
- Ficetola, G. F., Bonin, A. & Miaud, C. 2008 Population genetics reveals origin and number of founders in a biological invasion. *Mol. Ecol.* **17**, 773–782. (doi:10.1111/j.1365-294X.2007.03622.x)
- Garner, T. W. J., Perkins, M. W., Govindarajulu, P., Seglie, D., Walker, S., Cunningham, A. A. & Fisher, M. C. 2006 The emerging amphibian pathogen *Batrachochytrium dendrobatidis* globally infects introduced populations of the North American bullfrog, *Rana catesbeiana*. *Biol. Lett.* **2**, 455–459. (doi:10.1098/rsbl.2006.0494)
- Gotelli, N. J. & Colwell, R. K. 2001 Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecol. Lett.* **4**, 379–391. (doi:10.1046/j.1461-0248.2001.00230.x)
- Hajibabaei, M., Smith, M. A., Janzen, D. H., Rodriguez, J. J., Whitfield, J. B. & Hebert, P. D. N. 2006 A minimalist barcode can identify a specimen whose DNA is degraded. *Mol. Ecol. Notes* **6**, 959–964. (doi:10.1111/j.1471-8286.2006.01470.x)
- Hofreiter, M., Mead, J. I., Martin, P. & Poinar, H. N. 2003 Molecular caving. *Curr. Biol.* **13**, R693–R695. (doi:10.1016/j.cub.2003.08.039)
- Hulme, P. E. 2006 Beyond control: wider implications for the management of biological invasions. *J. Appl. Ecol.* **43**, 835–847. (doi:10.1111/j.1365-2664.2006.01227.x)
- Kats, L. B. & Ferrer, R. P. 2003 Alien predators and amphibian declines: review of two decades of science and the transition to conservation. *Divers. Distrib.* **9**, 99–110. (doi:10.1046/j.1472-4642.2003.00013.x)
- MacKenzie, D. I., Nichols, J. D., Royle, J. A., Pollock, K. H., Bailey, L. A. & Hines, J. E. 2006 *Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence*. Amsterdam, The Netherlands: Elsevier.
- Margulies, M. *et al.* 2005 Genome sequencing in micro-fabricated high-density picolitre reactors. *Nature* **437**, 376–380. (doi:10.1038/nature03959)
- Margurran, A. E. 2004 *Measuring biological diversity*. Malden, MA: Blackwell Science.
- Matsui, M., Honjo, M. & Kawabata, Z. 2001 Estimation of the fate of dissolved DNA in thermally stratified lake water from the stability of exogenous plasmid DNA. *Aquat. Microb. Ecol.* **26**, 95–102. (doi:10.3354/ame026095)
- R Development Core Team 2006 *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Taberlet, P., Griffin, S., Goossens, B., Questiau, S., Manceau, V., Escaravage, N., Waits, L. P. & Bouvet, J. 1996 Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Res.* **24**, 3189–3194. (doi:10.1093/nar/24.16.3189)
- Taberlet, P. *et al.* 2007 Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucleic Acids Res.* **35**, e14. (doi:10.1093/nar/gkl938)
- Valiere, N. & Taberlet, P. 2000 Urine collected in the field as a source of DNA for species and individual identification. *Mol. Ecol.* **9**, 2150–2152. (doi:10.1046/j.1365-294X.2000.11142.x)
- Venter, J. C. *et al.* 2004 Environmental genome shotgun sequencing of the Sargasso sea. *Science* **304**, 66–74. (doi:10.1126/science.1093857)
- Willerslev, E. *et al.* 2003 Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science* **300**, 791–795. (doi:10.1126/science.1084114)
- Willerslev, E. *et al.* 2007 Ancient biomolecules from deep ice cores reveal a forested Southern Greenland. *Science* **317**, 111–114. (doi:10.1126/science.1141758)