InBIO RESEARCH NETWORK
Biodiversity and Evolutionary Biology

EnvMetaGen

Deliverable 4.3 (D4.3)

Protocol for field collection and preservation of eDNA samples

Project acronym:	ENVMETAGEN
Project name:	Capacity Building at *InBIO* for Research and Innovation Using Environmental Metagenomics
Work Programme Topics Addressed:	H2020-WIDESPREAD-2014-2 (ERA CHAIRS)
Grant agreement:	668981
Project duration:	01/09/2015 – 31/08/2020 (60 months)
Co-ordinator:	ICETA - Instituto de Ciências e Tecnologias Agrárias e Agro-Alimentares
Delivery date from Annex I:	M36 (August 2018)
Actual delivery date:	M37 (September 2018)
Lead beneficiary:	ICETA
Project's coordinator:	Pedro Beja

Dissemination Level		
PU	Public	✓
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 668981

TABLE OF CONTENTS

SUMMARY

The overall goal of ERA Chair/EnvMetaGen project No 668981 is to expand the research and innovation potential of InBIO – Research network in Biodiversity and Evolutionary Biology, through the creation of an ERA Chair in Environmental Metagenomics. This field was selected as the focus of the ERA Chair, because Environmental DNA (eDNA) analysis is increasingly being used for biodiversity assessment, diet analysis, detection of rare or invasive species, population genetics and ecosystem functional analysis. In this context, the work plan of EnvMetaGen includes one work package dedicated to the *Deployment of an eDNA Lab* (WP4), which involves the training of InBIO researchers and technicians for implementing best practice protocols for the analysis of eDNA (Task 4.2). These protocols are essential for key application areas and to the development of research projects in association with business partners and other stakeholders, and thus to the strengthening of InBIO triple-helix initiatives (InBIO-Industry-Government; WP5). This report provides an overview of the current state of the art for collecting and preserving eDNA samples, with particular focus on vertebrate faecal samples, water samples and bulk invertebrate samples, which have been selected as key targets for the development of triple helix strategic initiatives (Task 5.3). The protocols already optimized and currently under development for the collection and preservation of eDNA samples are reported herein. Moreover, the future directions of sample collection and preservation at InBIO are discussed. This development was made possible through the recruitment of the ERA Chair team (WP2), secondments and Junior Researcher exchanges through the collaboration with international networks (WP3), an enhancement of computational infrastructure at InBIO (WP4) and participation of team members in workshops and conferences (WP6).

Together, Deliverables D4.2-D4.5 (Ferreira et al. 2018; Galhardo et al. 2018; Paupério et al. 2018) form a detailed account of the successful deployment of a fully functional eDNA lab under the EnvMetaGen project, and provide a valuable resource for eDNA practitioners in all spheres of the triple-helix model.

1. INTRODUCTION

1.1 The context

The overall goal of EnvMetaGen is to expand the research and innovation potential of InBIO – Research network in Biodiversity and Evolutionary Biology, through the creation of an ERA Chair in Environmental Metagenomics. The project strengthens the research potential of human resources, lab facilities and next-generation sequencing equipment funded by a previous FP7 CAPACITIES project (No 286431). Through research, innovation, and knowledge transfer, EnvMetaGen will increase the capacity of InBIO to tackle pressing societal challenges related to the loss of biodiversity, degradation of ecosystem services, and sustainable development.

The EnvMetaGen project is structured around seven interconnected Work Packages. Each Work Package has a number of Tasks designed to meet the respective Work Plan objectives. The primary objective of Work Package 4, *Deployment of an eDNA Lab*, is to deploy a fully functional environmental DNA (eDNA) lab, building upon the extant Illumina genomic platform funded by the previous FP7 project. To achieve this objective, the Work Package aims: to enhance the computational infrastructure to accommodate the massive amounts of data generated by the next-generation sequencing (Task 4.1) and to train InBIO researchers and technicians for implementing best practice protocols for the analysis of eDNA (Task 4.2). Together, these activities contribute to unlocking the full research potential of InBIO in the field of environmental metagenomics.

A Report on equipment acquisition (Deliverable D4.1), fulfilling the objectives of Task 4.1, has been submitted previously. This document, Deliverable D4.3, reports on one of the four aspects of capacity building considered pivotal to boost the future performance of InBIO in environmental genomics, which are the protocols for field collection and preservation of eDNA samples. Together with the protocols for building and organising reference collections of DNA sequences (Deliverable D4.2; Ferreira et al. 2018), for next-gen analysis of eDNA samples (Deliverable D4.4; Paupério et al. 2018) and for processing of DNA sequence data generated by next-gen platforms (Deliverable D4.5; Galhardo et al. 2018) it constitutes a standardised set of knowledge and skills that will be widely adopted in InBIO's genomic lab, achieving in this way Task 4.2 and a major objective of the EnvMetaGen project, and reaching in due time two

of the project's milestones: MS6 - Collections from sampling campaigns; and MS7 - Metagenomics protocols and tools developed.

The development of the protocols herein was made possible through a combination of activities planned within other Work Packages of the EnvMetaGen project, namely the Recruitment of the ERA Chair team (WP2; see completed Deliverables D2.1-D2.6), Secondments and Junior Researcher Exchanges through the collaboration with international networks (WP3; see completed Deliverables D3.3 & D3.5 and upcoming Deliverables D3.4 & D3.6, due at M48), an enhancement of computational infrastructure at InBIO (WP4; see above) and participation of team members in workshops and conferences (WP6; see completed Deliverable D6.6 and upcoming Deliverable D6.7, due at M48).

The protocols were designed considering the interests of stakeholders from academia, in particular InBIO, but also from industry and governmental organisations, to allow mainstreaming of environmental metagenomics to solve problems in the different domains, and in this way contribute to a major objective of Work Package 5, *Strengthening the triple helix: InBIO – Government – Industry relations*, which is to foster the contribution of InBIO for innovation and economic development, as one of the ways to ensure its long term sustainability (WP5; see completed Deliverable D5.3 and upcoming Deliverables D5.4 & D5.5, due at M48). EnvMetaGen has three key application areas: 1) Monitoring of freshwater eDNA for species detection; 2) Assessing natural pest control using faecal metagenomics and; 3) Next-generation biomonitoring using DNA metabarcoding. These key areas were proposed as strategic triple helix initiatives and have been considered when designing eDNA projects and protocols, and that is why they are directed to samples taken from freshwater, bulk invertebrate samples and vertebrate faecal samples. Metabarcoding, the identification of species present in a sample using next-generation sequencing, has been the primary approach. For details of current EnvMetaGen-affiliated projects, including their applicability to the triple-helix initiatives and EnvMetaGen objectives, see Appendix.

1.2 Overview of eDNA collection and preservation methods

eDNA usually describes DNA present in soil, water, sediments, air and other environmental samples, but can also be extended to any samples that contain DNA from a variety of sources, e.g. diet samples containing many food items or samples containing multiple invertebrate

individuals of different species. A primary characteristic of the samples is that target organisms are not isolated prior to DNA extraction (Taberlet et al. 2012). eDNA analysis is increasingly being used for biodiversity assessment, diet analysis, detection of rare or invasive species, population genetics and ecosystem functional analysis (Taberlet et al. 2012; Rees et al. 2014). As such, the range of sample types and research techniques using eDNA is extensive.

There are two major steps in obtaining samples for downstream DNA extraction and subsequent analyses: sample collection and sample preservation. Different environments require different sampling methodologies and this will depend not only on the environmental characteristics, but on the research priorities (Rees et al. 2014). There are a number of different protocols in the literature for these steps, to the extent that very few studies share exactly the same methods (Dickie et al. 2018).

In general, the goal of sample collection is to obtain samples containing good quality DNA in sufficiently high amounts and preserving this DNA until it can be processed in the laboratory, whilst also avoiding contamination (either cross-sample or from external sources).

1.3 Report structure

This report details the development of best practices at InBIO, regarding the collection and preservation of eDNA samples for downstream analyses. The report is structured according to the sample types being investigated: vertebrate faeces, freshwater, and bulk invertebrate samples. Each section is divided into two subheadings: collecting samples and preserving samples. Other sample types that are being considered for future EnvMetaGen projects are also discussed. Figure 1 provides an overview of the EnvMetaGen eDNA Lab workflow, with workflow steps arranged according to the reporting structure of Deliverables D4.2 - D4.5 (Ferreira et al. 2018; Galhardo et al. 2018; Paupério et al. 2018). Section 6 details all the field sampling protocols currently implemented by EnvMetaGen-affiliated projects. For details of current EnvMetaGen-affiliated projects, including their applicability to the triple-helix initiatives and EnvMetaGen objectives, see Appendix.

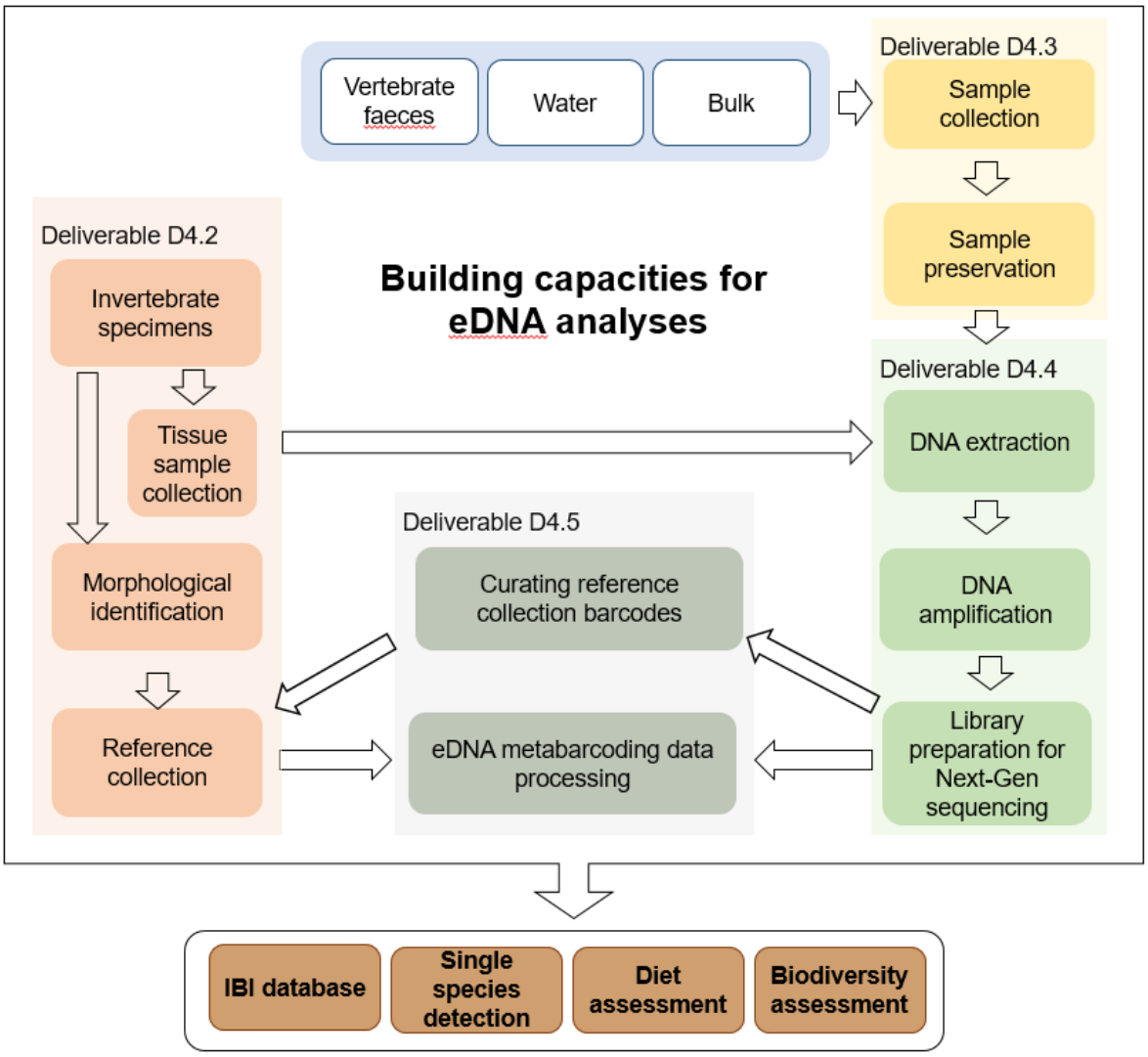

Figure 1. EnvMetaGen eDNA Lab workflow: steps are coloured according to the deliverable in which they are addressed (Deliverables D4.2 - D4.5). The type of eDNA samples (blue) and project applications (brown) require a range of tailored protocols within workflow steps, which are detailed in Deliverables D4.2 - D4.5 (Ferreira et al. 2018; Galhardo et al. 2018; Paupério et al. 2018).

2. SAMPLING VERTEBRATE FAECES FOR DNA

The investigation of faecal samples, coupled with advances in environmental metagenomics, was selected as one of the key areas to be developed in the EnvMetaGen project, as described in Work Package 2 (Task 2.2). Environmental metagenomics provides a unique opportunity to quantify trophic interactions on a large scale, through the massive sequencing and subsequent identification of DNA from faecal or other bodily extracts. The information thus generated is critical in a number of applications, including for instance the quantification of pest control services by insectivore bats and birds, the assessment of habitat loss and fragmentation effects on food web organization in human-modified landscapes, and the identification of critical food resources for endangered species. Developing methods and providing training for faecal analysis using eDNA approaches is integral to WP4, *Deployment of an eDNA Lab*, and to the strategic initiatives of the triple-helix model (WP5).

This section provides an overview of the protocols developed and used within EnvMetaGen-affiliated projects for field sampling of vertebrate faeces for downstream DNA-based analysis. The method of obtaining faecal samples will depend largely on the study species in question, in particular whether it is feasible to capture individuals or only search for faeces at field sites. Detailed protocols are provided in Section 6.

2.1 Collecting faecal samples

2.1.1 *Capturing individuals*

Live-capturing individuals has the advantage of providing individual level information such as sex, age and physical condition of the animal, all of which can be used for later analysis in the comparisons of diets. Capturing provides fresh faecal samples, as they are produced while the individual is being held. Fresh faecal samples are well documented to provide better quality DNA for downstream analysis than older samples (Deagle et al. 2005b; Santini et al. 2007; Panasci et al. 2011; McInnes et al. 2017). Live-capture methods are also used for monitoring populations through mark-recapture, which can allow for repeated measurements of diets of individuals to investigate temporal changes. One of the downsides of this approach is that fieldwork can be laborious, or unfeasible, for many taxa, such as large or dangerous vertebrates. Furthermore, live-handling of vertebrates requires permits and should only be carried out by

well-trained personnel to minimise the risk of injury or stress to the animal (see Sherwin et al. 2003).

2.1.1.1 Mist netting

Mist netting is a common technique for capturing bird and bat species (Keyes and Grue 1982; Dunn and Ralph 2004) and has been used to obtain faecal samples in many metabarcoding diet studies (e.g. Alberdi et al. 2012; Dodd et al. 2012; González-Varo et al. 2014; Galimberti et al. 2016). Mist nets are typically comprised of a mesh suspended between two poles. When properly installed the nets are virtually invisible. Mist nets have horizontally strung lines that create loose pockets. When a bird or bat hits the net, it falls into this pocket, where it becomes tangled.

- Pros: Provided good placement, many individuals can be captured in a short time frame.
- Cons: Can be difficult to set up and place correctly. Labour-intensive as nets need to be checked very regularly. There is some risk of injury to animals. Not suitable for all species
- EnvMetaGen-affiliated projects using this approach: SABOR

2.1.1.2 Tent spring trapping

Another method for capturing birds is tent spring trapping (Höglund and Lundberg 1987; Alatalo et al. 1996). These traps consist of a small net with a springed hinged frame. They are placed on the ground and baited. When a bird tugs at the bait, it triggers the frame to spring up and over it, forming a tent which encloses the individual. The individual can then be easily captured and handled.

- Pros: Can be used for species where mist trapping is not effective. Relatively easy to set up. Different bait types can be used to target different species (for example, using mealworms to attract insectivorous species or fruit to attract frugivorous species).
- Cons: Labour-intensive as traps need to be checked very regularly. There is some risk of injury to animals.
- EnvMetaGen-affiliated projects using this approach: CHASCOS

2.1.1.3 Box trapping

For small mammals, box trapping is a widely-used method for humane live capture (e.g. Cameron et al. 2005; Wilson et al. 2007a; Egeter et al. 2015). Commonly, the traps are

aluminium folding box traps, baited with peanut butter, rolled oats or other attractant. The traps have a pedal trigger that is released when the animal walks into the trap, closing the door behind it. Before releasing the captured animal, individual characteristics can be measured and faeces deposited in the tap can be collected.

- Pros: Depending on the species being studied, traps can be left overnight (12 hours) and checked each morning. Folding traps are easily transported and set up. There is very little risk of injury to animals.
- Cons: For some taxa (e.g. shrews) this can be labour-intensive, as traps need to be checked very regularly. Traps need to be soaked with bleach and thoroughly rinsed between each capture to ensure there is no DNA carry-over contamination of faeces. Can be difficult to entice trap-shy species/individuals to enter the traps.
- EnvMetaGen-affiliated projects using this approach: AGRIVOLE

2.1.1.4 Modified mesh trapping

For live-trapping Pyrenean desman, a semi-aquatic mammal, modified mesh traps based on the design of eel traps are used (Melero et al. 2009). These traps need to be placed at appropriate sites in the rivers, namely at narrow pathways within the river current. Traps are partially submerged in the water and checked regularly. The captured individuals are promptly manipulated to record individual characteristics and for tagging. Faecal samples are collected from the container where the individual is kept while being manipulated and the individual is release at the trapping site.

- Pros: Efficient at trapping desmans.
- Cons: Very labour-intensive, as traps need to be checked very regularly during the night.
- EnvMetaGen-affiliated projects using this approach: GALEMYS

2.1.2 *Field survey*

Where capturing individuals is not feasible, for example, due to a species´ large size, difficulty to handle or elusive nature, it is usual to obtain vertebrate faecal samples by field survey. Survey can include general walking surveys in areas inhabited by the study species, or more targeted surveys of nests, roosts, dens, latrine sites, known territory marking sites etc. Many species deposit faeces regularly at sites that are easy to survey, e.g. otters. The type of survey will

depend on the ecology of the species and knowledge of the habitat use by a population in the study area. Of course, samples obtained by this method can often be older and contain degraded DNA, negatively affecting downstream analyses (Deagle et al. 2005b; Santini et al. 2007; Panasci et al. 2011; McInnes et al. 2017). Even the substrate on which faecal samples are found can significantly affect detection of food item DNA (Oehm et al. 2011; McInnes et al. 2017). One way to be sure of obtaining fresh samples is to conduct regular (e.g. daily) site visits where all faeces are removed each day, either along transects, within study patches or at regularly used latrine sites.

- Pros: Sometimes the only feasible method for collecting faeces of a given study species. In some scenarios large numbers of relatively easy-to-find scats can be located. No risk of injury to animals.
- Cons: Individual level information such as sex, age and physical condition is not available. It may be difficult to find the number of scats needed for statistical analysis. Resolving which species deposited a particular scat can also require considerable expertise and even experienced researchers have been known to incorrectly assign scats to species based on appearance (Paxinos et al. 1997; Farrell et al. 2000; Davison et al. 2002). Scats are not guaranteed to be fresh. Faeces sitting in the environment prior to collection can become contaminated with DNA from other fauna, plants and fungi in the area, potentially causing false positives in downstream analyses. Unless sampling effort is standardised across sites, it can be difficult to compare sites in a statistically meaningful manner.
- EnvMetaGen-affiliated projects using this approach: AGRIVOLE, GALEMYS, IRANVERT, MATEFRAG, WOLFDIET

2.1.3 *Artificial refugia*

Artificial refugia are an ecological tool used to monitor wildlife populations or improve habitat availability through the provision of human-built constructions that attract target species (Zappalorti and Reinert 1994; Souter et al. 2004; Flaquer et al. 2006; Johnson et al. 2007; Lettink and Cree 2007; Wilson et al. 2007b). Individuals are free to come and go as they please. Examples include bat and bird boxes, reptile artificial cover objects, and artificial dens/retreat sites. Regarding EnvMetaGen-affiliated projects, bat boxes are the only artificial refugia being

used. These have been used to collect bat faeces in a number of previous studies (Shiel et al. 1998; Dondini and Vergari 2000; Siemers and Swift 2006).

- Pros: Can work well to obtain a larger sample size. Where only one bat is utilising bat boxes, individual level information such as sex, age and physical condition can be collected.
- Cons: Bat boxes are not always utilised by bats if sufficient habitat already exists locally. There can be a substantial initial cost in establishing bat boxes. If multiple individuals utilise a bat box, faecal pellets cannot be attributed to one individual. Unless bat boxes are thoroughly cleaned, or disposable sheets are used, there is a high risk for cross-contamination between faecal pellets. There is potential for certain cohorts of the population to be sampled, leading to bias in diet characterisation (i.e. bats using refuges may not represent the feeding habits of the population).
- EnvMetaGen-affiliated projects using this approach: TUA

2.2 Sample preservation

When the study species produces large faecal deposits, it can be necessary to collect sub samples of faeces. If subsampling is necessary, then the location within the faecal sample can affect downstream DNA analyses. Homogenising entire scats before subsampling can lead to a higher frequency of detection of food items than simply subsampling intact faeces (Deagle et al. 2005b; Stenglein et al. 2010). The types of faecal samples handled as part of EnvMetaGen-affiliated projects are all relatively small, fitting in 1.5 - 2 mL tubes, and as such this aspect does not currently pose an issue for these projects.

There have been many different approaches to stomach and faecal sample preservation (Frantzen et al. 1998; reviewed in King et al. 2008) including freezing, ethanol, ethanol plus freezing, silica gel, commercial kits and various buffers. A number of studies have reported success using 95% ethanol followed by storage at room temperature (Deagle et al. 2005a; Deagle et al. 2005b; Deagle et al. 2009; Deagle et al. 2010). Santini et al. (2007) showed that samples frozen in 95% ethanol outperformed storage in ethanol at room temperature as well as outperforming storing dried with silica at -20 ^{o}C and storing in a GUS buffer (guanidine thiocyanate) at room temperature. Silica gel generally appears to work well for samples that are already quite dry, but is often not suitable for very wet samples, which tend to develop mould

(Murphy et al. 2002). All the EnvMetaGen-affiliated projects use either 96% ethanol or silica for sample preservation. See Section 6 for preservation conditions associated with each sampling protocol.

Storage duration in previous studies has been reported anywhere from a few days to over two years, with successful DNA amplification even from long-term stored samples, but the effect of storage duration has not been studied extensively. Puechmaille et al. (2007) did not find any significant effect on amplification success of mtDNA for samples preserved using silica gel for either six or 18 months.

3. SAMPLING WATER FOR eDNA

The utilisation of water eDNA samples was also selected as one of the key areas to be developed in EnvMetaGen project, as described in Work Package 2 (Task 2.2). Aquatic habitats are often difficult to sample accurately, making the use of eDNA particularly promising for detecting invasive and rare species. Also, the use of eDNA is promising for the ecological assessment of freshwater ecosystems, which currently is based on "conventional" techniques that are costly, time consuming and require highly experienced taxonomists. As a consequence, improving eDNA techniques increases the capacity of InBIO to undertake biodiversity assessments of freshwater, particularly in poorly explored regions where InBIO researchers are increasingly working. Developing methods and providing training for eDNA analysis of water samples is integral to WP4, *Deployment of an eDNA Lab*, and to the strategic initiatives of the triple-helix model (WP5).

eDNA is being increasingly used to conduct biodiversity assessments (Jerde et al. 2011; Thomsen et al. 2012). Most studies using eDNA metabarcoding for detecting vertebrates from water-bodies concentrate eDNA prior to DNA extraction by filtering water (e.g. Robson et al. 2016; Lopes et al. 2017). Filtering generally results in higher amounts of eDNA being extracted than precipitating directly from water, as much higher volumes of water are processed when filtering (Adrian-Kalchhauser and Burkhardt-Holm 2016; Eichmiller et al. 2016; Hinlo et al. 2017). However, turbid water gives rise to limitations in filtering substantial volumes of water due to clogging of filter membranes (Turner et al. 2014; Robson et al. 2016; Spens et al. 2017), at least using low surface area filters. The primary alternative is using DNA precipitation methods on water samples directly, but filtering generally results in higher amounts of eDNA being extracted than precipitating directly from water, as much higher volumes of water are processed when filtering (Adrian-Kalchhauser and Burkhardt-Holm 2016; Eichmiller et al. 2016; Hinlo et al. 2017). As a result, the approaches used will vary according to environmental factors as well as the research objective of a project.

This section provides an overview of the protocols developed and used within EnvMetaGen and affiliated projects for field sampling of water for downstream DNA-based analysis. Detailed protocols are provided in Section 6.

3.1 Collecting water samples

3.1.1 *eDNA precipitation*

The most widely used precipitation method for collecting eDNA is to take a relatively small volume of water (usually 15 ml) in the field, and to add ethanol and sodium acetate (Ficetola et al. 2008). This is later transferred to the laboratory and stored at -20^{o}C. See Section 6 for EnvMetaGen field sampling protocols using eDNA precipitation.

- Pros: Can be used in turbid environments, where filtering can be unfeasible. Very easy to collect samples. Simple and cheap method with few equipment items needed.
- Cons: The low volume of water is generally considered to yield less eDNA, resulting in lower species detection rates.
- EnvMetaGen-affiliated projects using this approach: AZORES, CRAYFISH, FILTURB, XENOPUS

3.1.2 *Filtering*

The idea behind filtering water in the field is that high volumes of water can be passed through a membrane with a designated pore size, concentrating eDNA on the membrane. It has been used in the majority of water eDNA studies to date. Water is drawn through a filtering apparatus using a pump, which can be either manual or powered, and either vacuum or peristaltic. The filter is then removed and stored, either in preservation buffer, dry or in ethanol. There is currently no consensus on the exact pore size that should be used, but generally a range between 0.2 µm and 2 µm has been shown to capture eDNA at acceptable levels. A pore size of 0.45 µm appears to be the most commonly used. Wide pores make it easier to filter higher volumes of water, but are likely to involve a trade-off in amount / diversity of eDNA molecules captured. The membrane material has also been shown to affect eDNA biodiversity results (Majaneva et al. 2018), but again there are a large range of materials being used and there is currently no consensus on which material is most efficacious (e.g. mixed cellulose, cellulose nitrate, cellulose acetate, glass fibre, polyethersulfone, polyester, polypropylene, polyvinylidene difluoride, polytetrafluoroethylene).

- Pros: Can filter higher volumes of water (e.g. 250 ml – 2 L), compared to precipitation. Literature suggests filtering results in higher number of taxa detected.

- Cons: Special equipment required. More costly than precipitation method. Cannot be used in very turbid environments, or at least, turbidity can lead to different sampling volumes, making comparison among samples difficult.
- EnvMetaGen-affiliated projects using this approach: AZORES, FILTURB, FRESHING, GUELTA, ICVERTS, IRANVERTS

The vast majority of eDNA studies using filters use a 47 mm disc filter. A recent development has been the utilization of high-capacity filters: having narrow pores, but high surface areas (Civade et al. 2016; Eva et al. 2016; Valentini et al. 2016; Vences et al. 2016; Lopes et al. 2017; Spens et al. 2017). These are now commercially available, and are often packaged as an enclosed capsule, which reduces the potential for contamination in the field as there is no direct handling of the membrane itself.

From pilot studies conducted by EnvMetaGen, such capsules increase the water volume filtering capacity by ten-fold, even in extremely turbid environments, while maintaining the use of a narrow pore size of 0.45 µm.

- Pros: Can filter much higher volumes of water (1 L – 50 L).
- Cons: Very costly compared to either precipitation or standard filtering approaches.
- EnvMetaGen-affiliated projects using this approach: FILTURB

3.2 Sample preservation

Water samples collected by the precipitation method are stored in ethanol and sodium acetate. There are a wide variety of approaches used when storing filters and these have been compared in a number studies (Hinlo et al. 2017; Majaneva et al. 2018; Spens et al. 2017). These include freezing, lysis buffer, resuspension buffer, dry (with silica gel), dry (no additives), ethanol and others. The results to date have been varied, but most methods appear to preserve eDNA relatively well. As for all sample types for eDNA studies, the effect of storage duration has not been studied extensively. In general, storage duration has been reported anywhere from a few days to over a year, with successful DNA amplification even from long-term stored samples. See Section 6 for preservation conditions associated with each sampling protocol.

4. BULK SAMPLING INVERTEBRATES

The analysis of DNA from bulk samples collected in the field was not initially targeted as one of the key areas to be developed in EnvMetaGen project (WP2), but its importance soon become apparent for the development of the triple helix strategic initiatives, namely for the components of freshwater biomonitoring and biodiversity assessments (Task 5.3). While collecting eDNA directly from the environment has the advantages of being non-invasive and requiring relatively low effort in the field, it has a number of obstacles. eDNA extractions from environmental samples, such as water, generally result in highly complex samples containing eDNA from a diverse range of taxa. The presence of large amounts of eDNA from non-target groups (e.g. bacteria, when the study focus is invertebrate diversity) can pose difficulties for biodiversity analyses. Such difficulties include a reduced efficiency of Polymerase Chain Reaction (PCR) primers for species detection and the production of large amounts of unusable sequencing data. eDNA of the taxa of interest may also be highly diluted in water samples and thus difficult to detect using eDNA approaches. Bulk sampling (the capture of multiple individuals of a range of species in one sample) circumvents these issues by producing a more targeted starting sample that contains the study taxa in question in high quantities, while minimising the presence of non-target species. This method has been used in a number of DNA metabarcoding studies (e.g. Pfrender et al. 2010; Stein et al. 2014). For EnvMetaGen-affiliated projects this mainly concerns invertebrate bulk sampling in freshwater and above-ground environments. Freshwater invertebrates are caught by kick-sampling or other netting methods, while above-ground sampling is done using heath traps or vacuuming methods.

Bulk sampling is a cost-effective method to obtain samples that are representative of the invertebrate community (either in freshwater or above-ground habitats). DNA from bulk samples can be used for ecological assessment, detection of invasive or threatened species and biomonitoring. Developing methods and providing training for analysis of bulk samples is integral to WP4, *Deployment of an eDNA Lab*, and to strategic initiatives of the triple-helix model (WP5). Bulk samples can also be used to detect pest species, contributing to invasive species control (identified as an emerging eDNA research line, EnvMetaGen Objectives).

This section provides an overview of the protocols developed and used within EnvMetaGen for bulk sampling and preservation. Detailed protocols are provided in Section 6.

4.1 Collecting bulk invertebrate samples

4.1.1 *Kick-net sampling*

Kick-net sampling is often used for stream and small river habitats, and involves placing a net on the stream/river bed while disturbing the area immediately upstream of the net. This results in invertebrates on rocks and other sediments being caught in the net to give a representation of the invertebrate community of that site. It is common to identify different habitats within a site and to allocate sampling time proportionally to each habitat type.

- Pros: Simple and cheap method with few equipment items needed.
- Cons: In areas with deep sediment can result in excess sediment / detritus being sampled along with the invertebrates, leading to complex samples.
- EnvMetaGen-affiliated projects using this approach: FRESHING

4.1.2 *Light trapping*

Light traps have been used for monitoring invertebrate diversity of certain taxonomic groups (especially moths) for many years. In the case of heath traps the invertebrates are attracted to a UV light. Once they arrive at the trap they collide with the trap flight interception structure and fall into a collection receptacle, where they remain until the trap is disassembled. The traps are portable and autonomous and can be used for a standard time period in each sampling point. Their use in metabarcoding studies is relatively recent.

- Pros: Not very labour intensive.
- Cons: Only suitable for some taxonomic groups.
- EnvMetaGen-affiliated projects using this approach: TUA

4.1.3 *Vacuum sampling*

Vacuum sampling has been demonstrated as a suitable technique for sampling invertebrates, particularly diptera, homoptera and hymenoptera (Doxon et al. 2011). The technique involves the use of a specialized battery-powered backpack aspirator, which vacuums invertebrates present in vegetation. Invertebrates are captured in a removable container and can be stored or sorted for later identification or DNA extraction.

- Pros: Samples a broad taxonomic variety present on vegetation.

- Cons: Much more labour intensive than light trapping. Requires specialist equipment. Strong-flying invertebrates may escape capture causing biased sampling. Plant material vacuumed as by-product, which may necessitate downstream sorting.
- EnvMetaGen-affiliated projects using this approach: ECOLIVES

4.2 Sample preservation

There are very few research items investigating different preservation techniques for bulk samples. Although the use of ethanol is common worldwide (Hajibabaei et al. 2012; Shokralla et al. 2010), in the United Kingdom, Longmire´s Solution is being used in order to avoid the regulations pertaining to ethanol transportation (DNAqua-Net Working Group Meeting, Pécs, Hungary, 2018).

A consideration when choosing preservative is whether DNA will be extracted from the tissue collected, or from the preservative itself. Extracting DNA from bulk sample tissues involves mechanically homogenizing the tissues prior to DNA extraction. Extracting DNA from the preservative involves either filtering the preservative and extracting DNA from the filter or subsampling the preservative and extracting DNA directly. If DNA is to be extracted from preservative, the choice of preservative may affect downstream protocols. Recent research by EnvMetaGen-affiliated projects indicates that at least seven days should be allowed between collection of bulk samples and subsampling of the preservative (in this case ethanol) to allow DNA to enter the preservative from tissues. All the EnvMetaGen-affiliated projects use 96% ethanol for bulk sample preservation. See Section 6 for preservation conditions associated with each sampling protocol.

5. OTHER DEVELOPING AREAS FOR ENVMETAGEN

There are a number of other sample types that can be collected for eDNA studies. These are considered to be developing areas for the EnvMetaGen project, that are not detailed in this report as they are currently in the initial stages, and field collection protocols for these sample types are still being developed. However, overviews of all projects are provided in the Appendix of this report, as is the relevance of each project to the triple-helix initiatives and EnvMetaGen objectives. It should also be noted that such developing areas have the potential to identify and progress new strategic key areas within the triple-helix initiatives.

5.1 Sampling stomach contents

Although EnvMetaGen-affiliated projects to date have not directly involved the collection of stomach contents for diet analysis, one project, NZFROG, is utilising DNA that was extracted from stomach samples as part of another project. The processing of the eDNA samples for this project is included in Deliverables D4.4 (Paupério et al. 2018) and D4.5 (Galhardo et al. 2018). It is also anticipated that stomach contents will be collected as part of the CRAYFISH project, but as this project is in its early stages, detailed protocols are not provided at this point. Overviews of both projects are provided in the Appendix.

It is often easier to extract good quality DNA from stomachs than from faeces, as ingested items have been subjected to less digestion. There is also generally less predator DNA co-extracted than when using faeces (depending on the study species) and stomach contents samples are less prone to cross-contamination or contamination from external environmental sources. On the other hand, stomach collection is highly invasive, involving the euthanasia of individuals (although this is not a major ecological concern when dealing with invasive species), which also requires increased permitting from the relevant authorities. The collection of stomach samples may be extended to other species as part of future EnvMetaGen projects.

5.2 Sampling invertebrate faeces

Although major developments have been made for assessing vertebrate diets using metabarcoding, it is the field of invertebrate ecology that has largely pioneered research in this area of molecular ecology (King et al. 2008; Sheppard and Harwood 2005; Symondson 2002). One of the reasons for this is that many invertebrates either heavily masticate their prey or are fluid feeders, precluding morphological analysis (Admassu et al. 2006; Greenstone et al. 2007, Pompanon et al. 2012; Sunderland 1988; Symondson 2002). One EnvMetaGen-affiliated project, MANTID, aims to utilise metabarcoding methods to characterise the diet of Mantids. Mantids are efficient predators that capture and eat a wide variety of insects and other small prey. This project will assess mantid diets, through the collection of mantid faecal samples, focussing on their potential as agricultural pest controllers. If successful, this approach may be extended to other projects. As this project is in its early stages, detailed protocols are not provided at this point. A project overview is provided in the Appendix. The benefits of using faeces for invertebrate diet studies are generally the same considerations as for vertebrate

studies: it is non-invasive, does not require euthanasia of the animal and allows for repeat measurements on individuals.

5.3 Sampling soil

Analysis of soil eDNA has been used to assess biodiversity, ecosystem function and biomonitoring, which aligns well with existing EnvMetaGen objectives. One EnvMetaGen-affiliated project, SOILPHOS, has included the use of DNA extracted from an agricultural plant growth experiment focused on investigating the impacts of fertilizer on bacterial communities, in particular the bacterial component that is part of the phosphorus-cycling process. These samples were made available by collaborators, who also carried out the initial PCRs. The processing of the metabarcoding data for this project is included in Deliverable D4.5 (Galhardo et al. 2018). Another project, AGRIVOLE, which is currently in the early stages of development, aims to assess the responses of vole communities to agroecosystem structure and management practices, and includes the analyses of soil samples for analysing plant diversity. It is expected that, within the frame of these two projects, laboratory procedures for DNA extraction and amplification from soil samples will be optimised. Overviews of both projects are provided in the Appendix.

6. ENVMETAGEN FIELD SAMPLING PROTOCOLS

6.1 SAMPLING VERTEBRATE FAECES FOR DNA

6.1.1 *Mist netting – Bats and Birds*

Mist nets are assembled. The placement and length of nets varies according to habitat characteristics and the species being targeted. Nets are checked regularly, every 20 min for bats and every 30-60 min for birds. Captured individuals are carefully removed from the net and placed inside cotton bags (pre-cleaned by washing in bleach and rinsing thoroughly), one individual per bag. Variables of interested are recorded (age, sex, weight, wing length, etc.) and individuals are released. The maximum time an individual is kept in a bag is 30 min. Faeces deposited in the bag are removed and placed into a 2 mL screw cap tube, by using the tube itself to scoop up the faeces. For bats, tubes are pre-prepared in the laboratory by filling one fifth of the tube with silica beads. For birds, due to the faeces having a more liquid consistency, tubes

are either pre-prepared with 1 mL 96% ethanol in each tube or, once the sample has been collected, the remaining space in the tube is filled with 96% ethanol using a bottle with a dropper lid. Samples are brought to the laboratory and stored at -20ºC until DNA extraction.
EnvMetaGen-affiliated projects using this approach: SABOR, TUA

6.1.2 *Tent spring trapping – Birds*

Trap nets are set and baited in suitable areas. Chosen trap sites and bait types vary depending on the study species. For the CHASCOS project, mealworms are used as bait. Rocks are placed on the frame of the trap to ensure the trap stays in place once sprung and to provide some concealment of the trap. Traps are checked every 30 – 60 min. Captured individuals are carefully removed from the net and placed inside cotton bags (pre-cleaned by washing in bleach and rinsing thoroughly), one individual per bag. Variables of interested are recorded (age, sex, weight, wing length, etc.) and individuals are released. The maximum time an individual is kept in a bag is 30 min. Faeces deposited in the bag are removed and placed into a 2 mL screw cap tube, by using the tube itself to scoop up the faeces. Tubes are either prepared with 1 mL 96% ethanol in each tube or, once the sample has been collected, the remaining space in the tube is filled with 96% ethanol using a bottle with a dropper lid. In some cases, opportunistic sampling is undertaken whereby faeces deposited by the trapped individual on the rocks used during trap set up are similarly collected. Samples are brought to the laboratory and stored at -20ºC until DNA extraction.

EnvMetaGen-affiliated projects using this approach: CHASCOS

6.1.3 *Box trapping – Mammals*

Traps are placed in suitable areas according to the target species, in a grid or in transects. Distance between traps may vary between 2 to 10m and a mixture of sardine, oat and apple is used as bait. Traps are usually left open continuously during the night and day, and are checked at least twice per day. Pre-baiting for one or two days may be performed to increase trapping success. Captured individuals are removed from the trap to a plastic bag and the handled with care. Variables of interested are recorded (age, sex, weight, etc.) and individuals are released. Once the individual has been released, faecal pellets are collected from the trap or from the bag, using sterilized tweezers (cleaned with ethanol and flamed between samples). cats are handled

with care, in order to keep them as intact as possible and placed in a small 1.5 mL tube previously filled with 96% ethanol. Samples are brought to the laboratory and stored at -20ºC until DNA extraction.
EnvMetaGen-affiliated projects using this approach: AGRIVOLE

6.1.4 *Modified mesh trapping –Semi-aquatic small mammals*

Traps are placed at appropriate sites in the rivers, namely at narrows pathways within the river current. The traps are set at the end of the day partially submerged in the water and checked every 2 hours. No bait is used. Captured individuals are promptly manipulated to record individual characteristics and for tagging. Faecal samples are collected from the container where the individual is maintained while being manipulated, and the individual is release at the trapping site.

EnvMetaGen-affiliated projects using this approach: GALEMYS

6.1.5 *Field survey – Multiple classes of vertebrates*

6.1.5.1 General survey

Scats are searched for at known den sites, latrine sites, territorial marking sites or more generally within the study area, including scent-marking posts. Scats are visually assigned as belonging to the study species. Scats are carefully handled using disposable gloves. Entire scats are placed in tubes, to which 96% ethanol is added, or scats are placed in zip-lock bags, which are left open to dry the scat before closing. Samples are stored at room temperature until DNA extraction.
EnvMetaGen-affiliated projects using this approach: IRANVERT, WOLFDIET

6.1.5.2 Stratified survey

Study areas inhabited by the study species are defined into “patches”. Each patch is surveyed on four consecutive days, and all faecal pellets are collected. This allows for the collection of only freshly deposited faeces after the first day of survey. Time spent searching for faeces is directly proportional to size of the respective study patch. Faeces that are either isolated or in small latrines (<20 pellets) are preferentially targeted, to minimise the risk of different individuals contributing to the same sample. Each sample consists of between one and 20

pellets. Faecal pellets are collected using sterilized tweezers (cleaned with ethanol and flamed between samples). Scats are handled with care, in order to keep the outer layer as intact as possible. Scats are placed in a small 1.5 mL tube previously filled with 96% ethanol. Samples are brought to the laboratory and stored at -20°C until DNA extraction.
EnvMetaGen-affiliated projects using this approach: AGRIVOLE, MATEFRAG

6.1.5.3 Semi-aquatic mammal survey
Surveys are performed in sites at small streams, with shallow and flowing waters, abundant rocks and potential shelters. In each site, a stream sector with a maximum of 600m is surveyed. Faeces are searched along the stream bed and along margins, using a flashlight if needed in shady areas or in cavities. Fresh faecal pellets are collected using sterilized tweezers (cleaned with ethanol and flamed between samples). Scats are handled with care, in order to keep them as intact as possible. Scats are placed in a small 1.5 mL tube previously filled with 96% ethanol. Samples are brought to the laboratory and stored at -20°C until DNA extraction.
EnvMetaGen-affiliated projects using this approach: GALEMYS

6.1.6 *Artificial refugia – Bats*

Bat boxes are placed at sites representing dominant agricultural and natural habitats in the study area, and are visited once per month from April to October. At each box, faeces are collected until one 2 mL screw cap tube is full. Tubes are pre-prepared in the laboratory by filling one fifth of the tube with silica beads. Samples are stored in the laboratory at -20°C until DNA extraction.

EnvMetaGen-affiliated projects using this approach: TUA

6.2 SAMPLING WATER FOR eDNA

6.2.1 *eDNA precipitation*

At each site, 15 mL water is collected by submerging a sterile 50 mL centrifuge tube slightly below the water surface, allowing it to fill with 15 mL of water. A negative control (15 mL distilled water) is included. Immediately after collection, 1.5 mL of 3M sodium acetate and 33.5 mL ethanol (over 96%) is added to the tubes. Samples are stored at -20°C until DNA extraction.

EnvMetaGen-affiliated projects using this approach: AZORES, CRAYFISH, FILTURB, XENOPUS

6.2.2 *Filtering*

6.2.2.1 Filtering Protocol A: Standard Conditions

For each sample at a field location only equipment that has been sterilised by bleach is used, including all tubes and filter holders. Gloves are used throughout. A new sterile filter (of the desired pore size/material) is placed in the filter holder (Thermo Scientific Nalgene Polysulfone Filter Holder with Funnel) and secured firmly. One end of a silicon tube is attached to the output nozzle of the filter holder. The other end of the tubing is placed through the peristaltic pump head (Solinst 410). The pump is connected to a car battery. Up to 500 mL water is added to the filter holder and pumped through the apparatus. This is repeated until the desired volume is filtered. If 500 mL cannot be filtered due to turbidity, the volume is noted or the required volume is reduced. The filter is carefully removed, folded and placed in a tube (5+ mL) with 96+% ethanol, or placed in zip lock bag. At each site a negative control is used, by filtering up to 2 L dH_2O through a sterile filter. Samples are stored at -20 ^oC.

EnvMetaGen-affiliated projects using this approach: AZORES, FILTURB, FRESHING, ICVERTS

6.2.2.2 Filtering Protocol B: Extreme turbid environments in remote areas

For each sample at a field location only equipment that has been sterilised by bleach is used, including all tubes and filter holders. Gloves are used throughout. The entire desired volume is pre-filtered through a polypropylene filter (Airwatertech, Belgium) housed in a Cintropur NW18 filtering capsule (Airwatertech, Belgium). A dual action hand-powered vacuum pump (Tribord, France) is used to draw the water through the filtration system, capturing the filtrate. Filtrate is transferred to a filter holder (Thermo Scientific Nalgene Polysulfone Filter Holder with Funnel) fitted with a new sterile filter (of the desired pore size/material) and similarly drawn through the filter. The pre-filter and final filter are stored separately in a 50 mL and a 2 mL tube respectively, to which 96+% ethanol is added. At each site a negative control is included, by filtering up to 2 L dH_2O through a sterile filter. Upon return to the laboratory filters are stored at -20 ^oC.

EnvMetaGen-affiliated projects using this approach: GUELTA

6.2.2.3 Filtering Protocol C: Water with very low turbidity and low volume requirements

For each sample, 2 L water is collected in a sterile container. A 250 mL syringe is used to draw water from the bottle. A syringe filter of the desired pore size is connected and the water is forced through the filters. This is repeated until the desired volume is filtered or the filter becomes clogged. The syringe filter is removed, folded and placed in a tube (5+ mL) with 96+% ethanol, or placed in zip lock bag. At each site a negative control is included, by filtering dH_2O through a sterile filter. Upon return to the laboratory filters are stored at -20 ^oC.

EnvMetaGen-affiliated projects using this approach: IRANVERTS

6.2.2.4 Filtering Protocol D: High-capacity capsule filters

For each sample at a field location only equipment that has been sterilised by bleach is used, including all tubes and filter holders. Gloves are used throughout. One end of a length of a silicon tube is attached to the input nozzle of the capsule (FHT-45, Waterra). The other end is placed directly into the waterbody. One end of a second tube is attached to the output nozzle of the capsule. The other end is positioned in the peristaltic pump head (Solinst 410). The pump is connected to a car battery. The desired volume (or up to maximum capacity) is filtered. Tubes are carefully removed, any water remaining in the capsule is poured out, and the capsule is placed in a zip lock bag. At each site negative control is included, by filtering dH_2O through a new capsule. Upon return to the laboratory capsules are stored at -20 ^oC.

EnvMetaGen-affiliated projects using this approach: FILTURB

6.3 BULK SAMPLING INVERTEBRATES

6.3.1 *Kick-net sampling*

Each site is surveyed initially to assess the proportional coverage of each habitat type. Six one-metre transects are chosen, within a 50 m stream reach. At each transect a net (0.25 m x 0.25 m, 500 μm mesh) is placed on the stream / river bed. While standing upstream, the bed is disturbed with feet along the one-metre transect. Big rocks are manually handled to dislodge any invertebrates. Contents of all six transects are combined into one container, which is filled with 96% ethanol and stored at room temperature. On the 7^{th} to 14^{th} day post-collection,

between 2 and 10 mL of the ethanol from the container is subsampled and stored at -20 ºC until DNA extraction.

EnvMetaGen-affiliated projects using this approach: FRESHING

6.3.2 *Light-trapping*

A heath light trap is set up using UV LED lights with light sensors. This ensures that the UV lights are only active during night hours, optimizing battery life. The trap is visited within three hours of dawn and all trapped insects are collected by closing the mesh bag. In the laboratory or field station the mesh bag is placed in a -20 ºC freezer for 4+ hours. Invertebrates are transferred from the mesh bag to a container with 96% ethanol and stored at room temperature until DNA extraction.

EnvMetaGen-affiliated projects using this approach: TUA

6.3.3 *Vacuum sampling*

Selected trees are marked (five at each sampling site). The canopy of each tree is vacuumed for 1 minute. This is done using a battery-powered backpack aspirator with a collection cup. The material collected from the five trees is transferred to 100-ml containers, filled with 96% ethanol, and stored at room temperature.

EnvMetaGen-affiliated projects using this approach: ECOLIVES

7. CONCLUDING REMARKS

In this report, an overview of the current state of the art for collecting and preserving eDNA samples was provided, with particular focus on three sample types that are utilised to provide information on key application areas of the triple-helix initiatives. Decisions need to be taken early in project designs in order to collect eDNA samples in an efficient and suitable manner, as these will affect all subsequent downstream processes and consequently will have a substantial impact on the overall success of a project. To make these decisions, many factors must be considered pertaining to the study species in question, the environment being sampled, the resources available, and, ultimately, the questions that need answering.

While in some cases eDNA sample collection is relatively straightforward, in other cases, major technical challenges and questions persist, such as: *How fresh do faecal samples need to be? Which filtering methods are the most efficacious for filtering turbid water? How long should one wait before subsampling ethanol from invertebrate bulk samples?*

The EnvMetaGen project, along with its numerous affiliated projects and collaborators, has been developing methods to overcome such challenges. A number of projects have compared methods for eDNA collection from water samples, including AZORES, FILTURB, GUELTA, ICVERTS & XENOPUS. The FILTURB project successfully evaluated field collection methods for eDNA sampling in challenging turbid water environments, while GUELTA and ICVERTS have been trialing specific methods for poorly-explored remote regions (Egeter et al. 2018). The TUA project has been using novel methods for collection of invertebrate bulk samples, while the FRESHING project has successfully compared methods of bulk invertebrate preservation and subsequent subsampling. The SABOR project asked and answered the pertinent question of *How much is enough?* in terms of how many faecal pellets are necessary to characterize a diet and whether pooling of pellets prior to DNA extraction yields similar results (Mata et al. 2018).

All EnvMetaGen projects are generating ecological and environmental data to tackle pressing societal challenges related to the loss of biodiversity, degradation of ecosystem services, and sustainable development. These data feed into the triple-helix initiatives in the context of the strategic key areas of freshwater species detection, natural pest control services and biomonitoring. This report compiles a list of working best practice protocols for the collection and preservation of eDNA samples. Together, Deliverables D4.2-D4.5 (Ferreira et al. 2018; Galhardo et al. 2018; Paupério et al. 2018) form a detailed account of the successful deployment of a fully functional eDNA lab under the EnvMetaGen project, and provide a valuable resource for eDNA practitioners in all spheres of the triple-helix model.

8. HOW TO CITE

Egeter B, Fonseca NA, Paupério J, Galhardo M, Ferreira S, Oxelfelt F, Aresta S, Martins FMS, Mata VA, da Silva LP, Peixoto S, Garcia-Raventós A, Vasconcelos S, Gil P, Khalatbari L and Beja P (2018). Protocol for field collection and preservation of eDNA samples, D4.3 of the H2020-funded project EnvMetaGen (European Union Horizon 2020 Research & Innovation

Programme - H2020-WIDESPREAD-2014-2, Grant Agreement No 668981 - Capacity Building at InBIO for Research and Innovation Using Environmental Metagenomics). Porto, Portugal. doi: 10.5281/zenodo.2579807

9. REFERENCES

Admassu B, Juen A and Traugott M (2006). Earthworm primers for DNA-based gut content analysis and their cross-reactivity in a multi-species system. Soil Biology and Biochemistry 38: 1308-1315.

Adrian-Kalchhauser I and Burkhardt-Holm P (2016) An eDNA assay to monitor a globally invasive fish species from flowing freshwater. PloS One 11: e0147558.

Alatalo RV, Höglund J, Lundberg A, Rintamäki PT and Silverin B (1996) Testosterone and male mating success on the black grouse leks. Proceedings of the Royal Society B 263: 1697-1702.

Alberdi A, Garin I, Aizpurua O and Aihartza J (2012) The foraging ecology of the mountain long-eared bat Plecotus macrobullaris revealed with DNA mini-barcodes. PLoS One 7: e35692.

Cameron BG, van Heezik Y, Maloney RF, Seddon PJ and Harraway JA (2005) Improving predator capture rates: analysis of river margin trap site data in the Waitaki Basin, New Zealand. New Zealand Journal of Ecology 29: 117-128.

Civade R, Dejean T, Valentini A, Roset N, Raymond J-C, Bonin A, Taberlet P and Pont D (2016) Spatial representativeness of environmental DNA metabarcoding signal for fish biodiversity assessment in a natural freshwater system. PloS One 11: e0157366.

Davison A, Birks JD, Brookes RC, Braithwaite TC and Messenger JE (2002) On the origin of faeces: morphological versus molecular methods for surveying rare carnivores from their scats. Journal of Zoology 257: 141-143.

Deagle B, Chiaradia A, McInnes J and Jarman S (2010) Pyrosequencing faecal DNA to determine diet of little penguins: is what goes in what comes out? Conservation Genetics 11: 2039-2048.

Deagle BE, Jarman SN, Pemberton D and Gales NJ (2005a) Genetic screening for prey in the gut contents from a giant squid (*Architeuthis* sp.). Journal of Heredity 96: 417-423.

Deagle BE, Kirkwood R and Jarman SN (2009) Analysis of Australian fur seal diet by pyrosequencing prey DNA in faeces. Molecular Ecology 18: 2022-2038.

Deagle BE, Tollit DJ, Jarman SN, Hindell MA, Trites AW and Gales NJ (2005b) Molecular scatology as a tool to study diet: analysis of prey DNA in scats from captive Steller sea lions. Molecular Ecology 14: 1831-1842.

Dickie IA, Boyer S, Buckley H, Duncan RP, Gardner P, Hogg ID, Holdaway RJ, Lear G, Makiola A and Morales SE (2018) Towards robust and repeatable sampling methods in eDNA based studies. Molecular Ecology Resources.

Dodd LE, Chapman EG, Harwood JD, Lacki MJ and Rieske LK (2012) Identification of prey of *Myotis septentrionalis* using DNA-based techniques. Journal of Mammalogy 93: 1119-1128.

Dondini G and Vergari S (2000) Carnivory in the greater noctule bat (*Nyctalus lasiopterus*) in Italy. Journal of Zoology 251: 233-236.

Doxon ED, Davis CA and Fuhlendorf SD (2011), Comparison of two methods for sampling invertebrates: vacuum and sweep-net sampling. Journal of Field Ornithology 82: 60-67.

Dunn EH and Ralph CJ (2004) Use of mist nets as a tool for bird population monitoring. Stud Avian Biology 29: 1-6.

Egeter B, Bishop PJ and Robertson BC (2015) Detecting frogs as prey in the diets of introduced mammals: a comparison between morphological and DNA-based diet analyses. Molecular Ecology Resources, 15: 306-316.

Egeter B, Peixoto S, Brito JC, Jarman S, Puppo P and Velo-Antón G (2018) Challenges for assessing vertebrate diversity in turbid Saharan water-bodies using environmental DNA. Genome 61: 807-814.

Eichmiller JJ, Miller LM and Sorensen PW (2016) Optimizing techniques to capture and extract environmental DNA for detection and quantification of fish. Molecular Ecology Resources 16: 56-68.

Eva B, Harmony P, Thomas G, Francois G, Alice V, Claude M and Tony D (2016) Trails of river monsters: Detecting critically endangered Mekong giant catfish Pangasianodon gigas using environmental DNA. Global Ecology and Conservation 7: 148-156.

Farrell LE, Roman J and Sunquist ME (2000) Dietary separation of sympatric carnivores identified by molecular analysis of scats. Molecular Ecology 9: 1583-1590.

Ferreira S, Fonseca N, Egeter B, Paupério J, Galhardo M, Oxelfelt F, Aresta S, Archer J, Corley M, Penado A, Pina S, Jarman S and Beja P (2018). Protocol for building and organising reference collections of DNA sequences, D4.2 of the H2020-funded project EnvMetaGen (European Union Horizon 2020 Research & Innovation Programme - H2020-WIDESPREAD-2014-2, Grant Agreement No 668981 - Capacity Building at InBIO for Research and Innovation Using Environmental Metagenomics). Porto, Portugal. doi: 10.5281/zenodo.2586893.

Ficetola GF, Miaud C, Pompanon F and Taberlet P (2008) Species detection using environmental DNA from water samples. Biology Letters 4: 423-425.

Flaquer C, Torre I and Ruiz-Jarillo R (2006) The value of bat-boxes in the conservation of *Pipistrellus pygmaeus* in wetland rice paddies. Biological Conservation 128: 223-230.

Frantzen MAJ, Silk JB, Ferguson JWH, Wayne RK and Kohn MH (1998) Empirical evaluation of preservation methods for faecal DNA. Molecular Ecology 7: 1423-1428.

Galhardo M, Fonseca NA, Egeter B, Paupério J, Ferreira S, Oxelfelt F, Aresta S, Muñoz-Merida A, Martins FMS, Mata VA, da silva L, Peixoto S, Garcia-Raventós A, Vasconcelos S, Gil P, Khalatbari L, Jarman S and Beja P (2018). Protocol for the processing of DNA sequence data generated by next-gen platforms, D4.5 of the H2020-funded project EnvMetaGen (European Union Horizon 2020 Research & Innovation Programme - H2020-WIDESPREAD-2014-2, Grant Agreement No 668981 - Capacity Building at InBIO for Research and Innovation Using Environmental Metagenomics). Porto, Portugal. doi: 10.5281/zenodo.2586889.

Galimberti A, Spinelli S, Bruno A, Mezzasalma V, Mattia F, Cortis P and Labra M (2016) Evaluating the efficacy of restoration plantings through DNA barcoding of frugivorous bird diets. Conservation Biology 30: 763-773.

González-Varo JP, Arroyo JM and Jordano P (2014) Who dispersed the seeds? The use of DNA barcoding in frugivory and seed dispersal studies. Methods in Ecology and Evolution 5: 806-814.

Greenstone MH, Rowley DL, Weber DC, Payton ME and Hawthorne DJ (2007) Feeding mode and prey detectability half-lives in molecular gut-content analysis: an example with two predators of the Colorado potato beetle. Bulletin of Entomological Research 97: 201-209.

Hajibabaei M, Spall J, Shokralla S and van Konynenburg S (2012) Assessing biodiversity of a freshwater benthic macroinvertebrate community through non-destructive environmental barcoding of DNA from preservative ethanol. BMC ecology: 12: 28.

Hinlo R, Gleeson D, Lintermans M and Furlan E (2017) Methods to maximise recovery of environmental DNA from water samples. PloS One 12: e0179251.

Höglund J and Lundberg A (1987) Sexual selection in a monomorphic lek-breeding bird: correlates of male mating success in the great snipe *Gallinago media*. Behavioural Ecology Sociobiology 21: 211-216.

Jerde CL, Mahon AR, Chadderton WL and Lodge DM (2011) "Sight-unseen" detection of rare aquatic species using environmental DNA. Conservation Letters 4: 150-157.

Johnson JR, Knouft JH and Semlitsch RD (2007) Sex and seasonal differences in the spatial terrestrial distribution of gray treefrog (*Hyla versicolor*) populations. Biology Conservation 140: 250-258.

Keyes B and Grue C (1982) Capturing birds with mist nets: a review. North American Bird Bander 7: 2-14.

King RA, Read DS, Traugott M and Symondson WOC (2008) Molecular analysis of predation: a review of best practice for DNA-based approaches. Molecular Ecology 17: 947–963.

Lettink M and Cree A (2007) Relative use of three types of artificial retreats by terrestrial lizards in grazed coastal shrubland, New Zealand. Applied Herpetology 4: 227-243.

Lopes CM, Sasso T, Valentini A, Dejean T, Martins M, Zamudio KR and Haddad CF (2017) eDNA metabarcoding: a promising method for anuran surveys in highly diverse tropical forests. Molecular Ecology Resources 17: 904-914.

Majaneva M, Diserud OH, Eagle SH, Boström E, Hajibabaei M and Ekrem T (2018) Environmental DNA filtration techniques affect recovered biodiversity. Scientific Reports 8: 4682.

Mata VA, Rebelo H, Amorim F, McCracken GF, Jarman S and Beja P (2018) How much is enough? Effects of technical and biological replication on metabarcoding dietary analysis. Molecular Ecology.

Melero Y, Aymerich P, Luque-Larena JJ, Gosàlbez J (2012) New insights into social and space use behaviour of the endangered Pyrenean desman (*Galemys pyrenaicus*). European Journal of Wildlife Research 58: 185-193.

McInnes JC, Alderman R, Deagle BE, Lea MA, Raymond B and Jarman SN (2017) Optimised scat collection protocols for dietary DNA metabarcoding in vertebrates. Methods in Ecology and Evolution 8: 192-202.

Murphy MA, Waits LP, Kendall KC, Wasser SK, Higbee JA and Bogden R (2002) An evaluation of long-term preservation methods for brown bear (*Ursus arctos*) faecal DNA samples. Conservation Genetics 3: 435-440.

Oehm J, Juen A, Nagiller K, Neuhauser S and Traugott M (2011) Molecular scatology: how to improve prey DNA detection success in avian faeces? Molecular Ecology Resources 11: 620-628.

Panasci M, Ballard WB, Breck S, Rodriguez D, Densmore LD, Wester DB and Baker RJ (2011) Evaluation of fecal DNA preservation techniques and effects of sample age and diet on genotyping success. The Journal of Wildlife Management 75: 1616-1624.

Paupério J, Fonseca N, Egeter B, Galhardo M, Ferreira S, Oxelfelt F, Aresta S, Martins F, Mata V, Veríssimo J, Puppo P, Pinto J, Chaves C, Garcia-Raventós A, Peixoto S, da Silva L, Vasconcelos S, Gil P, Khalatbari L, Jarman S and Beja P (2018). Protocol for next-gen analysis of eDNA samples, D4.4 of the H2020-funded project EnvMetaGen (European Union Horizon 2020 Research & Innovation Programme - H2020-WIDESPREAD-2014-2, Grant Agreement No 668981 - Capacity Building at InBIO for Research and Innovation Using Environmental Metagenomics). Porto, Portugal. doi: 10.5281/zenodo.2586885.

Paxinos E, McIntosh C, Ralls K and Fleischer R (1997) A noninvasive method for distinguishing among canid species: amplification and enzyme restriction of DNA from dung. Molecular Ecology 6: 483-486.

Pfrender M, Hawkins C, Bagley M, Courtney G, Creutzburg B, Epler J, Fend S, Ferrington Jr L, Hartzell P and Jackson S (2010) Assessing macroinvertebrate biodiversity in freshwater ecosystems: advances and challenges in DNA-based approaches. The Quarterly Review of Biology 85: 319-340.

Pompanon F, Deagle BE, Symondson WOC, Brown DS, Jarman SN and Taberlet P (2012) Who is eating what: diet assessment using next generation sequencing. Molecular Ecology 21: 1931-1950.

Puechmaille SJ, Mathy G and Petit EJ (2007) Empirical evaluation of non-invasive capture–mark–recapture estimation of population size based on a single sampling session. Journal of applied Ecology 44: 843-852.

Rees HC. Maddison BC, Middleditch DJ, Patmore JR and Gough KC (2014). The detection of aquatic animal species using environmental DNA–a review of eDNA as a survey tool in ecology. Journal of applied Ecology 51: 1450-1459.

Robson HL, Noble TH, Saunders RJ, Robson SK, Burrows DW and Jerry DR (2016) Fine-tuning for the tropics: application of eDNA technology for invasive fish detection in tropical freshwater ecosystems. Molecular Ecology Resources 16: 922-932.

Santini A, Lucchini V, Fabbri E and Randi E (2007) Ageing and environmental factors affect PCR success in wolf (*Canis lupus*) excremental DNA samples. Molecular Ecology Resources 7: 955-961.

Sheppard SK and Harwood JD (2005) Advances in molecular ecology: tracking trophic links through predator–prey food-webs. Functional Ecology 19: 751-762.

Sherwin CM, Christiansen SB, Duncan IJ, Erhard HW, Lay DC, Mench JA, O'Connor CE and Petherick JC (2003) Guidelines for the ethical use of animals in applied ethology studies. Applied Animal Behaviour Sci 81: 291-305.

Shiel C, Duvergé P, Smiddy P and Fairley J (1998) Analysis of the diet of Leisler's bat (*Nyctalus leisleri*) in Ireland with some comparative analyses from England and Germany. Journal of Zoology 246: 417-425.

Shokralla S, Singer GA, Hajibabaei M (2010) Direct PCR amplification and sequencing of specimens' DNA from preservative ethanol. Biotechniques 48: 233-234.

Siemers BM and Swift SM (2006) Differences in sensory ecology contribute to resource partitioning in the bats Myotis bechsteinii and Myotis nattereri (Chiroptera: Vespertilionidae). Behavioural Ecology Sociobiology 59: 373-380.

Souter NJ, Michael Bull C and Hutchinson MN (2004) Adding burrows to enhance a population of the endangered pygmy blue tongue lizard, *Tiliqua adelaidensis*. Biological Conservation 116: 403-408.

Spens J, Evans AR, Halfmaerten D, Knudsen SW, Sengupta ME, Mak SS, Sigsgaard EE and Hellström M (2017) Comparison of capture and storage methods for aqueous macrobial eDNA using an optimized extraction protocol: advantage of enclosed filter. Methods in Ecology and Evolution 8: 635-645.

Stein ED, Martinez MC, Stiles S, Miller PE and Zakharov EV (2014) Is DNA barcoding actually cheaper and faster than traditional morphological methods: results from a survey of freshwater bioassessment efforts in the United States? PLoS One 9: e95525.

Stenglein J, De Barba M, Ausband D and Waits L (2010) Impacts of sampling location within a faeces on DNA quality in two carnivore species. Molecular Ecology Resources 10: 109-114

Sunderland KD (1988) Quantitative methods for detecting invertebrate predation occurring in the field. Annals of Applied Biology 112: 201-224.

Symondson WOC (2002) Molecular identification of prey in predator diets. Molecular Ecology 11: 627-641.

Taberlet P, Coissac E, Hajibabaei M, Rieseberg LH (2012) Environmental DNA. Molecular Ecology 21: 1789-93.

Thomsen PF, Kielgast JOS, Iversen LL, Wiuf C, Rasmussen M, Gilbert MTP, Orlando L and Willerslev E (2012) Monitoring endangered freshwater biodiversity using environmental DNA. Molecular Ecology 21: 2565-2573.

Turner CR, Barnes MA, Xu CC, Jones SE, Jerde CL and Lodge DM (2014) Particle size distribution and optimal capture of aqueous macrobial eDNA. Methods in Ecology and Evolution 5: 676-684.

Valentini A, Taberlet P, Miaud C, Civade R, Herder J, Thomsen PF, Bellemain E, Besnard A, Coissac E and Boyer F (2016) Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. Molecular Ecology 25: 929-942.

Vences M, Lyra ML, Perl RGB, Bletz MC, Stanković D, Lopes CM, Jarek M, Bhuju S, Geffers R, Haddad CFB and Steinfartz S (2016) Freshwater vertebrate metabarcoding on Illumina

platforms using double-indexed primers of the mitochondrial 16S rRNA gene. Conservation Genetics Resources 8: 323-327.

Wilson DJ, Efford MG, Brown SJ, Williamson JF and McElrea GJ (2007a) Estimating density of ship rats in New Zealand forests by capture-mark-recapture trapping. New Zealand Journal of Ecology 31: 47-59.

Wilson DJ, Wiedemer RL and Clark RD (2007b) Sampling skinks and geckos in artificial cover objects in a dry mixed grassland-shrubland with mammalian predator control. New Zealand Journal of Ecology 31: 169.

Zappalorti RT and Reinert HK (1994) Artificial refugia as a habitat-improvement strategy for snake conservation. Contributions to Herpetology 11: 369-375.

APPENDIX: DESCRIPTION OF ENVMETAGEN-AFFILIATED PROJECTS

This section provides a description of current EnvMetaGen-affiliated projects. At present, there are 20 ongoing EnvMetaGen-affiliated projects. Through the development of field, laboratory and data analysis pipelines, each of the projects contributes to the deployment of an eDNA Lab, which is the main goal of Work Package 4 and the focus of Deliverables 4.2 to 4.5 (Ferreira et al. 2018; Galhardo et al. 2018; Paupério et al. 2018).

All of the projects are highly collaborative involving a total of six other InBIO research groups, five research groups from other Portuguese institutions and fourteen overseas research groups. Twelve of the projects are being led by the EnvMetaGen team. These collaborations build relationships with key national and international organisations and networks in the environmental area, fostering the establishment of long-term partnerships with leading research institutions, helping to fulfil the objectives of Work Package 3 *Development of Capacities to Participate in the ERA*.

All projects are within the focus of one or more of the three key areas being developed under the triple-helix model of innovation (WP5):

1. Monitoring of freshwater eDNA for species detection
2. Assessing natural pest control using faecal metagenomics
3. Next-generation biomonitoring using DNA metabarcoding

The applicability of each project to EnvMetaGen Work Packages and Objectives is highlighted. Overall, the projects´ contributions to the deployment of an eDNA Lab, by developing analyses within the scope of the triple-helix key areas, as well as fostering networks among institutional, national and international collaborators, substantially increase InBIO´s capacity for research and innovation using environmental metagenomics.

AGRIVOLE

The role of voles in agroecosystems: linking pest management to biodiversity conservation under environmental change

Agroecosystem services are being threatened worldwide by biodiversity loss. Biological pest management is one of the main ecosystem services often supported by agroecosystems, as non-

crop habitats can provide resources for species that may act as natural controllers of agricultural pests, responsible for huge losses in crop yields. However, there is still limited understanding on how biodiversity levels relate with biological control, particularly considering current trends in agricultural land use change. AGRIVOLE project aims to assess the responses of vole communities to agroecosystem structure and management practices, by combining ecological tools and high throughput DNA sequencing techniques. The project will analyse the effects of different population regulatory processes and evaluate how community responses may affect the potential for pest outbreaks or impact the resilience of vole species of conservation concern. The focus will be on the vole community of northeastern Portugal agroecosystems, a species rich system where vole pests have significant economic impact on fruit tree orchards. The project will use data previously collected on voles' distribution in the region, complemented with detailed plant and vole surveys across agroecosystems with different structures and management treatments. We will also use high-throughput sequencing techniques, namely DNA metabarcoding, to determine voles' trophic niches based on their droppings. Overall, it is expected that the results obtained in this project contribute significantly to foster sustainable agricultural techniques linking pest management to biodiversity conservation. This project begun recently, but its progress will boost the development of the laboratory methods for analysing herbivore diets, using a metabarcoding approach, as well as the methods for collecting and analysing soil samples for determining plant diversity. Moreover, this project involves a collaboration with the University of Natural Resources and Life Sciences, Vienna, for building a reference collection for plants using high throughput sequencing, fundamental for the diet studies and vegetation surveys. Therefore, this project will contribute significantly for building capacity on the eDNA analyses in InBIO, while expanding its network of collaborations (WP3). AGRIVOLE is aligned with one of the key application areas of EnvMetaGen, *Assessing natural pest control using faecal metagenomics*, and it is expected that it provides relevant outcomes for practical applications in crop management. This may lead to the development of services, relevant to the farmers and Regional Agricultural Institutions, thereby fostering the triple helix (WP5).

AZORES

Assessing fish diversity in Azores freshwater lagoons using a metabarcoding approach

Eutrophication is a relevant issue for water quality in lagoons and is considered one of the main environmental problems in the Azorean archipelago, with high impacts on landscape, economy and the conservation of natural resources. Landscape changes and anthropogenic activities in general are considered as the main causes for eutrophication, and the lagoons in the island of São Miguel, are considered a good example of this situation, where land use changes have been associated with water quality degradation. Water quality of the Azorean lagoons has been monitored since 2003, and within this frame the development of efficient and cost-effective methods for monitoring biodiversity in the lagoons has become highly relevant. This project aims at developing a cost-effective monitoring program for fish diversity in the Azores freshwater lagoons. The main goal is the optimization of field and laboratory protocols for assessing the diversity of fish communities from environmental samples, using a metabarcoding approach. Samples have been collected by the University of Azores InBIO team, using both water filtering and precipitation techniques. The data is helping to refine best practices in collecting eDNA samples from water, while the optimisation of extraction and amplification protocols contribute to the development of capacities at InBIO. This project is aligned with the one of the key application areas of EnvMetaGen, *Next-generation biomonitoring using DNA metabarcoding*, and it is expected that it will help progress monitoring programs for fish diversity in freshwater ecosystems. The developed methodology is of relevance for the Regional Government of Azores, and applicable to other areas, with potential for application by other regional institutions and companies, thereby fostering the triple helix (WP5), and contributing to the expansion of InBIO´s collaboration network.

CHASCOS

Diet analysis of black wheatears (*Oenanthe leucura*)

The black wheatear (*Oenanthe leucura*) is the most threatened passerine in Portugal. Its distribution used to range from the Portuguese coast to the French Pyrenees. Nowadays it is extinct in France, while in Portugal it is restricted to the remote inner Douro and Tagus valleys, and in Spain its population decreased more than one third in recent years. To help understand the reasons for this severe decline, this project aims to study in detail the diet of this threatened bird. High throughput sequencing techniques have been shown to be able to characterise the diet of several animals in unprecedented detail. However, to study the diet of passerines and other large feeding spectrum animals is challenging for metabarcoding techniques due to

several constraints, such as molecular marker selection and secondary predation detection. High throughput sequencing is being used on droppings from captured birds in the Douro valley. As well as using traditional morphological analysis, several commonly used molecular markers are being used. All the information obtained from the molecular markers and the morphological identification are being compared. This has allowed the detailed description of the feeding requirements of the black wheatear, and given the observed large feeding spectrum and plasticity found, it has become apparent that it is unlikely that its decline is directly related to shortage of food. The project also identified the main problems and biases of some of the most commonly used molecular markers used in metabarcoding diet studies, and allowed for the development of techniques to minimize these problems. The project focuses on protecting biodiversity (identified as a societal challenge to be tackled by InBIO, EnvMetaGen Objectives) thereby contributing to the triple-helix initiatives (WP5). It focuses on identification of critical food resources for endangered species (identified as an emerging eDNA research line, EnvMetaGen Objectives). By comparing diet analysis protocols and molecular markers, it contributes substantially to the development of an eDNA lab by making technical advancements that have implications for eDNA best practices (WP4) and help to build capacity at InBIO.

CRAYFISH

Assessing the impact of invasive crayfish through diet analysis

The invasion of freshwater ecosystems by exotic species is a cause of concern worldwide due to their negative environmental and economic impacts. Invasive crayfish are one of the most detrimental alien species occurring in European freshwater ecosystems. Among the known, negative effects are bioturbation, competition with native species, predation on native biodiversity, effects on leaf and algae abundance, and trophic subsidizing for predators (which in turn can enhance predation on native species). To adequately assess the impact of these species, including their potential overlap with the trophic niche of native, threatened fauna, and provide information on their control and management, knowledge of their trophic ecology is essential. This project aims to characterize the diet of two invasive crayfish species in Northern Portugal (*Procambarus clarkii* and *Pacifastacus leniusculus*) using metabarcoding. As both species are thought to have a varied generalist diet, the project will involve conducting assays targeting a number of mitochondrial metabarcoding markers across multiple prey groups. The project will provide high resolution diet information for improved management of these

invasive species, which pose a widespread global threat to biodiversity. It should be noted that this project is in the early stages of development, and as such detailed protocols are not provided in these deliverables. The project will focus on biodiversity conservation and invasive species control (identified as an emerging eDNA research line, EnvMetaGen Objectives), producing data to inform governmental management for protecting biodiversity (identified as a societal challenge to be tackled by InBIO, EnvMetaGen Objectives) thereby contributing to the triple-helix initiatives (WP5). The project already has an associated InBIO MSc student, who will receive training in metagenomic techniques, helping to build InBIO's capacity (WP4).

ECOLIVES

Fostering sustainable management in Mediterranean olive farms: pest control services provided by wild species as incentives for biodiversity conservation

Efficient pest management is recognized as a major challenge for fostering economically profitable agroecosystems worldwide. Biocontrol services provide clear incentives for biodiversity conservation in agroecosystem as naturally occurring species can efficiently reduce populations of pests, thus reducing both crop losses to pests and the need for agrochemicals. Yet, the ecology of biocontrol services is poorly known, thus limiting our ability to understand its value and to plan their conservation and management. Using Mediterranean olive farms as case study, the overarching research goal of this project is to estimate the value of natural biological control of the Olive fruit fly (*Bactrocera oleae*) and the Olive fruit moth (*Prays oleae*) –the two major pests in olive farms worldwide–, in farms following distinct pest management strategies. The overall hypothesis is that the abundance and diversity of biocontrol providers will decline with increasing pest management intensity, which will be expressed in a non-negligible economic impact. Specifically, the project will focus on predatory insects (parasitoid wasps) as well as insectivorous vertebrates (birds and bats) as biocontrol providers. This is particularly relevant because, although birds and bats are thought to provide high levels of pest suppression, knowledge about their role as biocontrol providers is negligible compared to insect predators in Mediterranean olive farms in particular and in agroecosystems worldwide in general. The hypothesis will be tested by quantifying occurrence and abundance patterns both of biocontrol providers and insect pests in 2 olive farms following distinct types of pest management strategies: IPM (Integrated Pest Management), where producers apply agrochemicals when pest populations reach the economic threshold; and organic, where

producers rely completely on biocontrol services. The relative importance of each biocontrol provider on levels of pest infection will be investigated, and their economic value calculated. The data obtained at this local scale will be used to model potential scenarios of biocontrol services provision in olive farms at the whole Iberian Peninsula, with the aim to select priority conservation-management in the face of global environmental change. This project is based in Évora University and the EnvMetaGen team will participate on the development of molecular tools to identify prey items of key predators/parasitoids present in olive farms and to perform diet analysis. The project is likely to provide data to assist farmers finding better solutions to pest control than using high loads of pesticides. This project is of high relevance to existing and future InBIO-Industry-Government triple-helix initiatives (WP5), as it uses faecal eDNA samples to assess natural species as a form of pest control, addressing the provision of ecosystem services (identified as a promising eDNA research theme, WP2). The associated InBIO PhD student will receive training in metagenomic techniques, helping to boost InBIO's capacity (WP4).

FILTURB

Comparing methods to filter turbid water and modelling site occupancy based on eDNA detections

eDNA survey methods have been applied mainly in freshwater ecosystems, focusing on water without a high sediment load. This is largely due to difficulties with sampling suitable volumes of turbid water. One of the objectives of this project is to test the efficiency of different DNA capture methods in turbid waters, evaluating their performance on eDNA recovery and species detection. The project will compare the most common filtering and DNA precipitation methods with newer high-capacity filtering approaches. The latter have the potential to filter much higher volumes of water than the former, even in turbid environments. Using the information from this objective a second aspect of eDNA sampling will be investigated: modelling site occupancy based on eDNA detections. Once shed into the environment, the probability of detecting DNA of a target species will vary depending on environmental factors. By collecting eDNA samples multiple times at many sites, the probability of detection of amphibians will be estimated using site occupancy models. This will inform future studies on the number of samples that are required to detect a given species. The project is focussed on making technical advancements for cost-effective species detection and biodiversity assessment, contributing to existing and

future triple-helix initiatives in different areas (WP5). By comparing existing and emerging protocols, it will also help to implement best practice protocols for eDNA analysis (WP4). The project already has an associated InBIO MSc student, who will receive training in eDNA sampling and metagenomic techniques, helping to boost InBIO's capacity (WP4). This project is closely linked with GUELTA.

FRESHING

Next-generation biomonitoring: freshwater bioassessment and species conservation improved with metagenomics

Data collection of freshwater habitats is essential, allowing countries to fulfil legislation requirements, such as the European Union Habitat and Water Framework directives. However, collecting biotic data for freshwater monitoring implies extensive effort. This project aims to investigate the value of using latest metagenomic approaches and applied ecological tools to improve freshwater bioassessments and detection of species of conservation concern, and ultimately optimize monitoring programs. Objectives include: 1) developing metagenomic approaches to obtain reliable biodiversity data and species detections; 2) building metagenomic multimetric indexes for bioassessment of ecological quality; 3) validating rapid landscape predictions for monitoring bioassessment indices, and threatened and invasive species; and 4) designing a next-generation biomonitoring framework for freshwaters for an early warning system to alert authorities. The project will focus on fishes and macroinvertebrates, in the Douro Basin (North Portugal), because they are informative freshwater indicators and include many species of conservation concern. Ultimately, the project will use decision making and conservation tools to perform a cost-efficiency analysis, and design a framework for next-generation monitoring programs in freshwaters. The project is focussed on making technical advancements for cost-effective species detection, biodiversity assessment and biomonitoring. It will have implications for biodiversity conservation and invasive species control, contributing to the triple-helix initiatives (WP5) and the development of an emerging eDNA research line (EnvMetaGen Objectives), producing data to inform governmental management for protecting biodiversity (identified as a societal challenge to be tackled by InBIO, EnvMetaGen Objectives). The project tackles the pressing societal challenge of the loss of biodiversity (EnvMetaGen Objective). The project has an associated InBIO PhD student, who will receive

training in metagenomic techniques, and will include the comparison of existing and emerging protocols, helping to boost InBIO's capacity (WP4).

GALEMYS

Conservation genetics of a threatened semi-aquatic mammal: The Iberian desman (*Galemys pyrenaicus*) in northeast Portugal

The Iberian desman (*Galemys pyrenaicus*) is a threatened, elusive mammal endemic of the Iberian Peninsula and the Pyrenees. In Portugal, the species is restricted mostly to the North of the country and a recent survey revealed a marked reduction in the species distribution in Northeast Portugal. Besides the paucity of distributional data, baseline information relative to the ecology, genetic diversity and structure in Portugal is also scarce. However, this knowledge is crucial for understanding how river connectivity shapes the species ecology, particularly considering the threat posed by the recent construction of large hydroelectric infrastructures. Therefore, this project aims at determining the degree of genetic diversification and structuring of the desman population in Portugal and examining how species traits and trophic requirements together with river connectivity and other landscape features influence the species persistence in fragmented areas. This information is vital for an efficient conservation of this endangered, poorly known, semiaquatic mammal. For achieving this main goal, a set of microsatellites is being optimized using high throughput sequencing (HTS) for analysing the population genetic structure and diversity with tissues and non-invasive samples (faeces). Moreover, faeces collected in two river basins are being analysed using metabarcoding for assessing the species trophic niche in the study area. Therefore, this project is contributing for building capacities at InBIO, namely for the optimization of methods for genotyping microsatellites using HTS and for refining best practices in the diet analyses of insectivores using metabarcoding. GALEMYS project is related with one of the key application areas of EnvMetaGen, *Next-generation biomonitoring using DNA metabarcoding*, as it is expected that the results obtained with this project will help define conservation actions for this endangered species. Therefore, we expect this project to contribute with relevant information to the Portuguese administration strengthening the relation between InBIO and administration (WP5).

GUELTA

Assessing vertebrate diversity in turbid Saharan water-bodies using environmental DNA

The Sahara Desert is the largest warm desert in the world and a poorly-explored area. Small water-bodies occur across the desert, which are crucial habitats for vertebrate biodiversity, as well as providing resources for local human activities. The long-term conservation of these habitats requires a better assessment of local biodiversity and potential human-related conflicts. There is potential to use eDNA for monitoring vertebrate biodiversity in these areas. However, there are a number of difficulties with sampling eDNA from such turbid water-bodies and it is often not feasible to rely on electrical tools in remote desert environments. This project is trialling novel, manually-powered, water filtering methods in Mauritania to obtain eDNA samples. The project is focussed on making technical advancements for cost-effective biodiversity assessment, contributing to triple-helix initiatives in identified key areas (WP5), in poorly explored regions (identified as a promising eDNA research theme, WP2). As well as contributing to the deployment of an eDNA lab, it provides training for InBIO researchers as it involves the investigation and comparison of multiple field eDNA sampling methods (WP4). This project is also closely linked to FILTURB.

ICVERTS

Providing an eDNA tool for rapid assessment of ecological integrity through detection of rare indicator species in Western Africa

This project focuses on the detection of two iconic West African wetland species as bio-indicators: the Critically Endangered West African slender-snouted crocodile (*Mecistops cataphractus*) and the Endangered pygmy hippopotamus (*Choeropsis liberiensis*). The goal of the project is to assess whether an eDNA approach can provide a rapid assessment tool of ecological integrity by detecting the presence of these important indicator species. Such a tool would greatly reduce manpower and costs associated with traditional survey methods. High sensitivity qPCR species-specific assays have been developed to detect the DNA of these two high-value species. Water samples were collected throughout protected areas of Cote d'Ivoire, the last strongholds for these species in the Upper Guinea forests of West Africa. Although qPCR is often regarded as the most sensitive method of species detection, there is a current ideological shift towards the idea that metabarcoding methods may in fact detect rare species in eDNA samples with a similar efficacy. The project will compare both approaches of species detection. The project is focussed on developing biodiversity assessment tools, contributing to triple-helix initiatives in identified key areas (WP5), in a poorly-explored tropical region

(identified as a promising eDNA research theme, WP2), to be used by researchers and government for protecting biodiversity (identified as a societal challenge to be tackled by InBIO, EnvMetaGen Objectives).

IBI

InBIO Barcoding Initiative

DNA barcoding is an essential tool in a vast array of ecological and conservation studies. With the advent of Next Generation sequencing, it became possible to implement diet analysis and monitoring methods based on DNA metabarcoding. While such studies can include a range of environmental DNA sample types, such as faeces, saliva, blood meal, stomach contents, hair, water, air, pollen/natural by-products (e.g. honey), soil, bulk samples (or preservative), all demand the availability of a reference collection of DNA sequences in order to allow the correct identification of taxa found in each sample. Therefore, its applicability is hampered by the lack of comprehensive reference collections, particularly of invertebrates that are underrepresented in reference databases and this knowledge gap becomes greater in biodiversity hotspots. During the early stages of the EnvMetaGen project conception the need of developing a reference collection of DNA sequences for Portuguese invertebrates was identified and for this reason the Task 4.2. - Building capacity for eDNA analysis includes the construction and organisation of reference collections of DNA sequences as one of the pivotal capacity-building aspects. The InBIO Barcoding Initiative consists in the development of a DNA reference collection of voucher specimens identified by specialised taxonomists following the best practices, which is essential to develop and conduct consistent, reliable and repeatable research studies boosting the future performance of InBIO in environmental genomics. By combining field work and networking with taxonomists and ecologists, the project aims to produce DNA barcodes for thousands of species, covering over one hundred families of insects. The reference library will be a fundamental tool for long-term and large scale monitoring programs in Portugal and serve as base for ecological studies related with loss of biodiversity, degradation of ecosystem services, and sustainable development (EnvMetaGen Objectives) and to promising eDNA research themes (WP2). Along its construction the project contributes for the training in taxonomy and metagenomic techniques, helping to boost InBIO's capacity (WP4). Furthermore, it is likely to become a tool with significant relevance to the InBIO-Industry-Government triple-helix initiatives (WP5) by promoting the development of partnerships in all

key areas: Monitoring of freshwater eDNA for species detection; Assessing natural pest control using faecal metagenomics; and Next-generation biomonitoring using DNA metabarcoding.

IRANVERTS

Assessing diet of large felids in central deserts of Iran

Information on population structure, hormones, parasites and diets can all be produced using non-invasive faecal samples. Such information is highly valuable for conservation of elusive species such as Asiatic cheetah (*Acinonyx jubatus venaticus*). For this project scat samples are being collected from large carnivores across the distribution range of Asiatic cheetah. Using metabarcoding, scats will firstly be assigned to the predator species and secondly used to assess the diets of large felids. Two different extraction methods are being trialled to test for their efficacy in producing DNA suitable for predator species identification. Extracted DNA will be subject to PCR using a number of vertebrate-targeting PCR primers. Possible prey items include wild sheep (*Ovis orientalis*), wild goat (*Capra aegagrus*), gazelles (*Gazella bennettiii* and *Gazella subgutturosa*) and domestic livestock. This project is of relevance to the agricultural industry sector as well as for conservation of a threatened species, contributing to two key areas targeted for triple-helix initiatives (WP5). It tackles the pressing societal challenge of sustainable development (EnvMetaGen Objective) and includes assessment of habitat loss on trophic interactions in human-modified landscapes and management of wild and domestic herbivores (identified as promising eDNA research themes, WP2).

MANTIDS

Diet analysis of mantids

Modern molecular techniques have made it possible to assess species composition of complex samples, almost independently of individual density. In the last decades, DNA Metabarcoding together with High Throughput Sequencing (HTS) has allowed for diet assessment in several groups of animals, including insects. Although major developments have been made for assessing vertebrate diets using metabarcoding, it is the field of invertebrate ecology that has largely pioneered research in this area of molecular ecology. One of the reasons for this is that many invertebrates either heavily masticate their prey or are fluid feeders, precluding morphological analysis. This EnvMetaGen-affiliated project aims to utilise metabarcoding methods to characterise the diet of selected species of mantids in Portugal. Mantids (Order:

Mantodea) are highly-adapted predatory insects. Their diet is thought to be varied but no DNA-based assessment has been performed so far. This project will assess mantid diets in nature, through the collection of mantid faecal samples, focussing on their potential as agricultural pest controllers. This exploratory project might prove to be of high relevance to the InBIO-Industry-Government triple-helix activities (WP5), as it uses faecal eDNA samples to assess natural species as a form of pest control, addressing the provision of ecosystem services (identified as a promising eDNA research theme, WP2). The associated InBIO master student, will receive training in metagenomic techniques, helping to boost InBIO's capacity (WP4).

MATEFRAG

Impacts of habitat fragmentation on social and mating systems: testing ecological predictions for a monogamous vole through non-invasive genetics

Intensification of agriculture has caused severe loss and fragmentation of semi-natural habitats worldwide. Studies of the effects of habitat fragmentation on biodiversity have revealed large impacts on species distribution and abundance patterns. However, understanding demographic and behavioural processes that determine species vulnerability to fragmentation is important to properly understand population viability in human-dominated landscapes. Key, relevant, within-population processes affecting reproductive success and thus population persistence include social interactions, mating systems, and the formation of Kin-structures. In this project we aim to assess the effects of habitat fragmentation on mammalian social and mating systems, and how this affects population persistence. As it is expected that monogamous species are more susceptible to stochasticity and prone to extinction events, we have focused this project on a monogamous Iberian endemic mammal, the Cabrera vole (*Microtus cabrerae*). To achieve this main goal, this project is using genetic non-invasive sampling (faeces) for individual identification and for estimating kin-structure. The methods being used for species and individual identification from faeces were already optimized at InBIO (see Deliverable 4.4 for details; Paupério et al. 2018), hence this project has provided a relevant contribution in capacity building of eDNA (WP4).

NZFROG

Determining the impact of invasive mammals on frogs in New Zealand

Since the arrival of mammals, New Zealand's endemic frogs (*Leiopelma* spp.) have undergone a number of species extinctions and range contractions. Only two species now persist on the mainland. One of these, *Leiopelma archeyi*, is Critically Endangered and ranked as the world´s most evolutionarily distinct and globally endangered amphibian. Ship rats (*Rattus rattus*) have often been implicated in the decline of amphibians in New Zealand and worldwide, but prey from rodent stomach contents are notoriously difficult to identify. This project utilises metabarcoding to survey for predation by ship rats on the remaining mainland *Leiopelma* species. New PCR primers were developed that target all anuran species. This study has provided the first evidence of these frog species in mammalian stomach contents and this, along with evidence from other studies, has led to the the New Zealand government including certain important sites in their rodent control program. It should be noted that field samples for this project were collected as part of a separate project and as such the field collection protocols are not explicitly detailed, but the treatment of the eDNA samples and subsequent data are included in Deliverables 4.4 and 4.5 (Galhardo et al. 2018; Paupério et al. 2018). The project focuses on biodiversity conservation and invasive species control, contributing to the triple-helix initiatives (WP5) and an emerging eDNA research line (EnvMetaGen Objectives), producing data to inform governmental management for protecting biodiversity (identified as a societal challenge to be tackled by InBIO, EnvMetaGen Objectives). It also contributes to the deployment of an eDNA lab (WP4) by providing a new and validated primer set.

SABOR

Assessment of the role of bats as pest regulators in Mediterranean agriculture

Small vertebrate insectivores are judged to provide important ecosystem services by controlling insect pests. Bats, in particular, are major insect predators, suggesting that they play a vital role in protecting crops from pests. However, there’s a lack of basic information regarding bats’ diet and foraging behaviour. Traditional diet analyses use visual identification of arthropod fragments present in faecal or stomach contents, and are limited to order or family level identifications, not allowing the identification of possible pest species. When species level identifications are possible, these are usually restricted to hard-bodied insects, like Coleoptera. Recently, with the advancement of molecular methods, it became possible to identify at the species level both hard and soft-bodied insects, present in bat guano. In particular, the emergence of HTS techniques allows the barcoding of multiple insect species in complex

samples – metabarcoding. These novel methods are revolutionizing dietary studies and can give us precious insights into the role of bats as pest regulators. This project consists of a PhD thesis and aims to answer the following questions: i) What's the diet of a Mediterranean bat community? ii) How do bats group in terms of diet composition? iii) Is there a relationship between bat diet and bat/insect traits? IV) Which bats prey on pest insects and how often? This study will help enlightening the role of bats as pest regulators in Mediterranean agricultural fields. This will not only promote bat populations, but also help farmers finding better solutions to pest control than using high loads of pesticides. This project is of high relevance to develop InBIO-Industry-Government triple-helix initiatives (WP5), as it uses faecal eDNA samples to assess natural species as a form of pest control, addressing the provision of ecosystem services (identified as a promising eDNA research theme, WP2). The associated InBIO PhD student, has been receiving training in metagenomic techniques, helping to boost InBIO's capacity (WP4).

SOILPHOS

Assessing diversity of phosphorus-cycling bacteria in response to fertiliser treatments

Phosphorus is essential to crop and pasture growth and is added to soil in large volumes around the world. However, phosphorus is a scarce, finite resource with peak phosphorus expected as early as 2030 and high-quality rock phosphate estimated to be exhausted within 80 years. It has long been established that bacteria are involved in making phosphorus available to plants, but only recently have DNA-based technologies developed enough to study 1) bacterial soil community and 2) the prevalence of 'phosphorus-freeing' genes in the soil. The aim of this project is to investigate the prevalence and diversity of phosphorus-freeing genes in soil experimentally subjected to various phosphorus levels. The objective is to inform practitioners and researchers as to whether the global community should be trying to foster certain bacterial communities that will allow us to continue food production at its current rate whilst lowering the amount of phosphorus currently applied to agricultural land. This project is of high relevance to develop InBIO-Industry-Government triple-helix initiatives (WP5) as well as tackling the pressing societal challenge of sustainable development (EnvMetaGen Objective) and addressing the provision of ecosystem services (identified as a promising eDNA research theme, WP2). It should be noted that eDNA sampling and PCRs for this project were part of a

separate project and as such are not explicitly detailed, but the data processing is included in Deliverable 4.5 (Galhardo et al. 2018).

TUA

Promotion of ecosystem services in the Vale do Tua Regional Natural Park: Control of agricultural and forest pests by bats

The Vale do Tua Regional Natural Park (PNRVT) is an excellent example of the natural and patrimonial values that exist in the northern region of Portugal. Here the landscape is dominated by a mosaic of natural and semi-natural vegetation and agricultural areas with predominance of vineyards, olive groves and cork oak forests. Thus, as in other regions of the interior of Portugal, the region's economy is very dependent on agricultural productivity. In this context, one of the most relevant Ecosystem Services (ESs) potentially provided by biodiversity in the region may be the control of agricultural and forestry pests. Due to the high diversity of birds and bats in the region, it is expected that these groups may have great relevance in the provision of these ESs. Several studies have shown that large numbers of these flying vertebrates associated with high prey consumption (mostly insects) make birds and bats one of the most significant natural controllers of agricultural and forest pests populations, thus providing a high economic value, reduced use of pesticides and increased productivity. Therefore, this project aims to create conditions for the intensification of the provision of pest control services (identified as a promising eDNA research theme, WP2) by promoting the populations of the respective predators, focusing essentially on bats. In order to increase the number of bat colonies in the areas of interest, shelter boxes were placed in the most important agricultural and forestry systems in the PNRVT area, specifically vineyards, olive groves and cork oak forests. The evaluation of the effectiveness of this measure will be done by analysing the diet of bats in the shelters, checking which bat species are using the shelters and if they consume (and when) the existing agricultural and forest pests in the region. This project is a prime example of an InBIO-Industry-Government triple-helix initiative (WP5), as it involves stakeholders from administration (the Agency for Regional Development of the Tua Valley, in charge of the management of the park), academia (InBIO) and industry (landowners within the geographical limits of the park). Its results will allow the development of management plans optimizing the ESs provided by bats in the region, giving an example where the promotion and preservation of biodiversity will translate into economic gains for the stakeholders involved, thus waiting

for the PNRVT's management model to be disseminated at the regional and national levels, fostering sustainable development (EnvMetaGen Objective).

WOLFDIET

Describing the diet of African golden wolf (*Canis anthus*) and assessing human conflict

The African golden wolf (*Canis anthus*), previously considered as Golden jackal (*Canis aureus*), is now recognized as a new canid species occurring in North and East Africa. There is a lack of knowledge regarding most of the ecological traits of this medium-sized canid, particularly regarding feeding ecology. African wolves are reported as generalist feeders, consuming plants, insects and vertebrates, including livestock and poultry which raise important conflicts with humans. However, the few available studies are based on identification of macro-components found in scats rarely genetically validated, which may bias the results and underestimate some prey items. Based on 150 scats of African wolves collected in NW Senegal (comprising Djoudj National Park and a neighboring agricultural area) already available and genetically identified in a scope of another InBIO project, this study aims to adequately characterize the diet of African wolves using metabarcoding. The project will involve targeting metabarcoding markers across multiple prey groups and a methodological assay involving two different extractions performed for each scat. By using a high resolution approach, this project is expected to assess the diet of African wolves and their potential impact on threatened fauna (e.g. breeding and migrating birds) and domestic animals, providing essential information for an efficient management. This project is of relevance to the agricultural industry sector as well as for conservation of a threatened species, contributing to key areas identified for triple-helix initiatives (WP5). It tackles the pressing societal challenge of sustainable development (EnvMetaGen Objective) and includes assessment of habitat loss on trophic interactions in human-modified landscapes and management of wild and domestic herbivores (identified as promising eDNA research themes, WP2).

XENOPUS

Detecting the presence of invasive frogs (*Xenopus laevis*) in Portugal

The African clawed frog (*Xenopus laevis*) is a species that has been introduced to many parts of the world. Invasions are due to both accidental escape and voluntary release of laboratory animals in many cases. The predatory impacts of *X. laevis* on native populations of

amphibians and fish have been well documented. The species has been implicated in the global transmission of disease including chytridiomycosis, a disease cited as one of the principal causes for the global decline in amphibians. Under a protocol established between Portugal´s governmental conservation agency (ICNF), the Environmental Biology Centre of the Faculty of Sciences of the University of Lisbon and the Gulbenkian Institute of Science, a plan was developed that aims to control *X. laevis*. In order to assess whether the control protocol is effective, an eDNA experiment was planned which aims to detect *X. laevis* at sites where the species is present, sites where it has never been observed and sites where populations have been the subject of the control protocol. The aim is to simultaneously provide a reliable species detection tool and assess the efficacy of current control protocols. This project involves all three groups of the InBIO-Industry-Government triple-helix model (WP5). It focusses on invasive species detection and control (identified as an emerging eDNA research line, EnvMetaGen Objectives) as well as tackling the pressing societal challenge of the loss of biodiversity (EnvMetaGen Objective) and addressing the provision of ecosystem services (identified as a promising eDNA research theme, WP2).