InBIO RESEARCH NETWORK
Biodiversity and Evolutionary Biology

EnvMetaGen

Deliverable 4.4 (D4.4)

Protocol for next-gen analysis 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 the 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 (WP) 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). This report provides an overview of best practices for analysing eDNA samples, focusing on vertebrate faeces, water samples and bulks of invertebrates. It covers all steps from DNA extraction to amplification and sequencing in high throughput platforms, while addressing the challenges identified in each step. Moreover, it describes protocols already optimised and currently under development at InBIO, with particular reference to the detection of single species, diet assessments and biodiversity assessments. These three different outputs are fully aligned with the three key application areas that have been targeted in EnvMetaGen for a wider application of environmental metagenomics, and on which the development of strategic triple helix initiatives will focus on. These will therefore contribute to mainstream this technology as a new cost-effective solution to tackle current challenges faced by business partners and other stakeholders, while fostering InBIO-Industry-Government relationships (WP5 - Strengthening the triple helix). The protocols described herein were developed in close connection to, and build on accomplishments of, the other WPs in the project, including: i) the recruitment of the ERA Chair team (WP2), ii) secondments and researcher exchanges through collaborations with international networks (WP3), iii) the enhancement of the computational infrastructure at InBIO (WP4), and iv) participation in and organization of workshops and conferences (WP6). Future directions in eDNA analyses at InBIO, and other potential laboratory pipelines for future optimisation are identified. Together, Deliverables D4.2-D4.5 (this document; Egeter et al. 2018; Ferreira et al. 2018; Galhardo 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 main 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 is expected to strengthen the research potential of human resources, lab facilities and next-generation sequencing equipment funded by a previous FP7 CAPACITIES project. By expanding the research capacity and innovation of InBIO and strengthening knowledge transfer activities, EnvMetaGen intends to increase the ability of InBIO to tackle pressing societal challenges related to the loss of biodiversity, degradation of ecosystem services, and sustainable development.

One of the specific goals of EnvMetaGen project is the deployment of a reference environmental DNA (eDNA) lab of international prestige (objective iii), for improving innovation capacity and performance and act as a vehicle to attract new research collaborations and establish partnerships with top research institutes, international organizations, industry corporations, and the public administration. This objective is built into one of the seven work packages (WP) of EnvMetaGen Project – WP4 – *Deployment of an eDNA lab*. In this WP, for the deployment of the eDNA lab two tasks were defined, that include i) *Enhancing the computational and data storage infrastructure* (Task 4.1., Deliverable D4.1 reports on equipment acquisition to fulfil this task) and ii) *Building capacity for eDNA analysis* (Task 4.2.). These two tasks are expected to unlock the full research potential of InBIO in the field of environmental metagenomics, and strongly contribute to increase the research competitiveness in the ERA and the capacity to promote eDNA approaches to address relevant problems in ecology and environmental sciences.

The current document, Deliverable D4.4, reports on one of the four aspects of capacity building considered pivotal to boost the performance of InBIO in environmental metagenomics - the workflow for the laboratory analyses of eDNA samples. Together with the protocols for building and organising reference collections of DNA sequences (Deliverable D4.2; Ferreira et al. 2018), for collecting and preserving eDNA samples (Deliverable D4.3; Egeter et al. 2018) and for analysing the outputs of next-gen DNA sequencing platforms (Deliverable D4.5; Galhardo et al. 2018), it establishes a standardized set of best-practices that will be widely adopted at InBIO's laboratories, thus accomplishing Task 4.2 and a major goal of EnvMetaGen project. The four reports also describe the work developed for reaching two of the project's

milestones: MS6 - Collections from sampling campaigns, and MS7 Meta-genomics protocols and tools developed.

The development of the laboratory protocols described herein, was accomplished through a combination of activities developed 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 networking with international research teams (WP3, see completed Deliverables D3.3 and D3.5 and upcoming Deliverables D3.4 and D3.6, due at M48), an enhancement of the 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 deployment of the eDNA Lab and the setting up of the protocols were designed considering the interest of stakeholders from academia, in particular InBIO, but also from industry and governmental organisations to foster mainstreaming of environmental metagenomics in different domains, and in this way contribute to a major objective of WP5, *Strengthening the triple helix: InBIO – Government – Industry relations*. This is expected 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 and D5.5, due at M48). Three key application areas have been targeted for the development of triple helix initiatives in EnvMetaGen: 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 have been considered when designing the eDNA projects and protocols, thereby contributing to achieve objectives v) and vi) of the EnvMetaGen Work Plan. Therefore, the developed protocols focus mainly on the analyses of DNA from samples taken from freshwater, bulks of invertebrates, and vertebrate faeces. The primary laboratory approach relies on metabarcoding methods for the identification of species present in a sample using next-generation sequencing. All of the detailed protocols were developed within the scope of current EnvMetaGen-affiliated projects. These are summarized in the Appendix, including a description of the goals and framework, as well as their applicability to the triple-helix initiatives and EnvMetaGen objectives.

1.2. Overview of eDNA and applications

eDNA consists of the DNA extracted from an environmental sample without accessing the target organism or organisms (Lodge et al. 2012), and it has become a valuable tool for documenting biodiversity and for conservation. In the recent years, there has been a major development of the methods of analyses of eDNA, including quantitative Polymerase Chain Reaction (qPCR) techniques and high throughput sequencing (HTS), expanding greatly on its applications. The major applications include: single species detection, including endangered and invasive species (e.g. Biggs et al. 2015); biodiversity assessments (e.g. Yu et al. 2012); diet analyses (e.g. Pompanon et al. 2012); populations genetics (e.g. De Barba et al. 2016); parasitology (Srivathsan et al. 2016); disease detection (Huver et al. 2015); and microbiome characterisations (e.g. Gellie et al. 2017).

Several types of samples have already been used to recover eDNA, including water, air, soil, faeces, saliva, hair, pollen (in the study of invertebrates), and natural products (e.g. honey, herbal medicines). The advances in HTS allow the analyses of multiple samples and of complex mixtures, where DNA of several individuals and species may be found. The term 'DNA metabarcoding' was introduced by Taberlet et al. (2012) to label high-throughput multi-taxa identification using environmental DNA. These authors also expanded on this definition, including species identification from bulk samples of whole organisms, as for instance samples taken from aquatic or terrestrial environments and including multiple species of invertebrates.

The development and implementation of methods for the analyses of eDNA at InBIO started some years ago, mainly with the analyses of faecal samples for species detection and populations genetics of species of conservation concern (e.g. Barbosa et al. 2013, Godinho et al. 2014). However, within the frame of EnvMetaGen, namely of Task 4.2, there have been major advances in eDNA analyses at InBIO, with the broadening of analyses to other types of environmental samples (e.g. water samples, bulk samples, preservative ethanol) and targeting a wider range of applications, going from single species detection to diet and biodiversity assessments. Currently, several projects are ongoing (see Appendix), in the frame of which several protocols have already been developed and others are still being optimised. There has been an effort to test different methodologies and to perform proper optimisations to assure the best practices in the analyses of eDNA.

1.3. Structure of the report

This report details the protocols being developed at InBIO laboratories for the analyses of eDNA, following best practices. For accomplishing Task 4.2, *Building capacity for eDNA*

analysis, a workflow has been set at InBIO in the context of Deliverables D4.2 to D4.5 (this document; Egeter et al. 2018; Ferreira et al. 2018; Galhardo et al. 2018; Figure 1). This report focuses on the analyses of environmental samples from DNA extraction, through amplification of targeted sequences and library preparation to sequencing. Initial steps of the workflow, concerning field collection and preservation of eDNA samples are detailed in Deliverable D4.3 (Egeter et al. 2018). The bioinformatic protocols for processing the sequence data generated are detailed in Deliverable D4.5 (Galhardo et al. 2018). The identification of haplotypes through the preparation of a reference collection of barcodes is dealt with in Deliverable D4.2 (Ferreira et al. 2018).

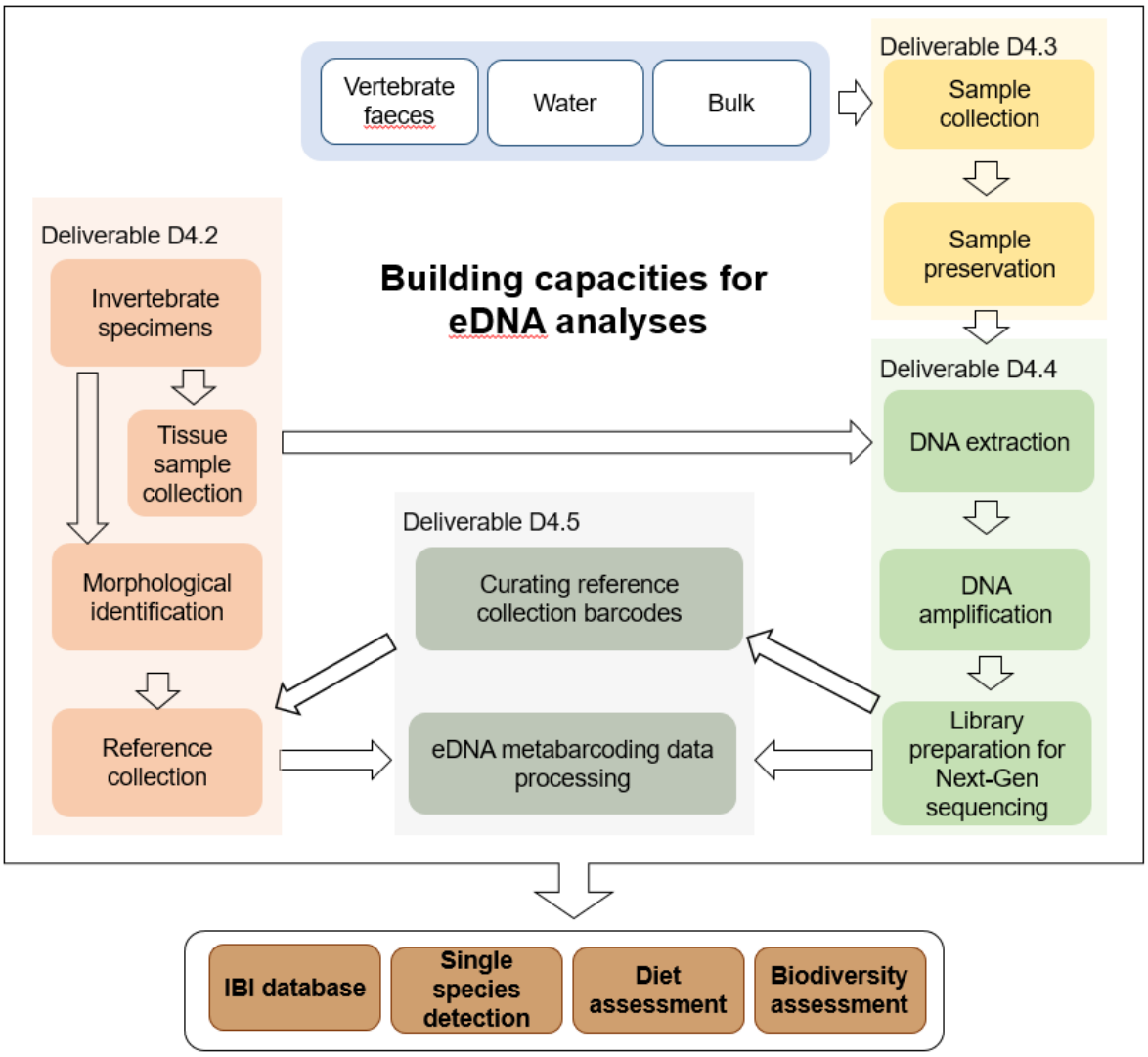

Figure 1. EnvMetaGen eDNA Lab workflow: steps are coloured according to the deliverable in which they are addressed (Deliverables D4.2 - D4.5; this document; Egeter et al. 2018; Ferreira et al. 2018; Galhardo et al. 2018). The type of eDNA samples (blue) and project applications (brown) require a range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

The report starts by discussing the general aspects concerning eDNA laboratory analyses, focusing on the challenges of the different steps (Section 2). Then the procedures developed for the analyses of eDNA are detailed (Section 3), and anticipated future work is discussed (Section 4).

2. LABORATORY WORKFLOW FOR THE ANALYSES OF eDNA

In this section we describe the laboratory workflow for the analyses of eDNA that is being optimised and applied within the scope of EnvMetaGen (see Figure 1), while discussing the challenges in each step. The laboratory procedures start with DNA extraction from the environmental samples collected in the field (see Deliverable 4.3 for details on sample collection and preservation). The eDNA obtained is then subjected to Polymerase Chain Reaction (PCR) targeting specific DNA regions, which are then indexed during library preparation, and the final library is sequenced. The following sections describe each step of the workflow.

2.1. DNA extraction

DNA extraction from the environmental samples is the first step within the laboratory workflow. There are multiple protocols in the literature that have been developed and may be applied to the different types of samples (e.g. Beja-Pereira et al. 2009; Deiner et al. 2015), but the protocol selection needs to be carefully considered depending on the goals of the study and the type of sample being analysed.

eDNA extraction protocols that are being optimised within the frame of EnvMetaGen-affiliated projects target three main applications: single species detection, diet assessments and biodiversity assessments (see Section 1 and Figure 1). From the outset of each project, a pilot study is usually performed to determine the extraction protocol that provides optimum results, considering the specific goals and types of samples being analysed. eDNA sample types being utilised by EnvMetaGen-affiliated projects may be split in three categories: vertebrate faeces, water samples and bulk invertebrate samples.

There are several relevant aspects that should be considered when selecting the best strategy and protocol for DNA extraction. One of these aspects is the proportion of sample that is used for extraction. For samples of small size (e.g. small rodents or bat faeces), the extraction of the whole sample is preferable as it is feasible and provides the maximum amount of information. For larger samples, such as large mammal faeces, large water filters or bulk samples, subsampling may be needed. The sampling or subsampling strategy for DNA extraction needs to be carefully considered, as it affects downstream analyses and results (see for example Yoccoz 2012 and Evans et al. 2017 and also Deliverable D4.3; Egeter et al. 2018). In the

particular case of diet studies involving small vertebrate faecal samples, one issue is the pooling of samples for analyses. This has been the focus of a study developed within EnvMetaGen, where the analyses of individual faeces were compared with the analyses of pooled samples (Mata et al. 2018). The results of this study indicated that biological replication is critical for dietary studies, as the pooled samples performed poorly, compared to individual samples.

Another aspect is the potential presence of substances in the samples that may cause amplification inhibition (PCR inhibitors), which is frequently an issue in environmental samples. This issue may be dealt with at the extraction phase, using DNA extraction protocols that include specific steps for the removal of inhibitors (see e.g. Beja-Pereira et al. 2009). However, if the extraction protocol used is not efficient in the removal of PCR inhibitors, it is also possible to deal with this issue at the amplification phase, by for example using bovine serum albumin in the PCR (Jiang et al. 2005) or using PCR methods more tolerant to inhibitors (digital PCR, Tréguier et al. 2014).

Below we describe the DNA extraction workflow considering the three main types of samples collected within the frame of the EnvMetaGen-affiliated projects.

2.1.1. Vertebrate Faeces

Vertebrates faeces are considered to be a good source of DNA for obtaining data concerning the detection of species, as well as for studying diets and for population genetic studies (Beja Pereira et al. 2009). In fact, these types of samples have been used widely at InBIO for answering questions concerning species distribution ranges (Mestre et al. 2015), population densities (Sabino-Marques et al. 2018), connectivity and dispersal (Ferreira et al. 2018), hybridization (Godinho et al. 2015), etc. Within the frame of EnvMetaGen-affiliated projects, we have also started to focus on the trophic niches of several taxa, through the analyses of faeces coupled with metabarcoding approaches (e.g. Mata et al. 2016, Appendix 1).

The extraction protocols being applied for vertebrate faeces depend not only on the goal of the study (see Section 2.1 and Figure 1), but also on the target species. Both the size of the scat and its main composition may impact on the extraction method and DNA yield for high throughput sequencing. Laboratory protocols for extracting DNA from scats of different species have been previously optimised at InBIO, mainly for single species detection and population genetic studies. Considering that for single species detection and population genetics, the target is the

DNA from the 'consumer', the DNA extraction is usually performed targeting the outer layer of the scat (that holds higher amount of cells from the 'consumer') and using either only a part of the scat (carnivore scats, e.g. Nakamura et al. 2017), or a small group of scats (rodent faeces, e.g. Ferreira et al. 2018). Pilot studies are currently being performed both for carnivore and rodent faeces, to assess whether DNA extraction from the outer layer of the scat, yields similar DNA extracts (considering amount, quality and prey diversity) to whole homogenised scats, allowing for the characterisation of the diets of the species. One other possibility for DNA extraction from large scats would be to homogenise the scat, and then use a subsample for DNA extraction, as has been done in other mammal studies (Egeter et al. 2015; Kartzinel et al. 2015). When subsampling is performed for extraction, the need of performing extraction replicates should be evaluated.

For small insectivore species (including both mammals and birds), as the scats usually break apart upon immersion on lysis buffer, the extraction protocols use the whole homogenised scat and a protocol that may include a step for adsorption of inhibitors (e.g. InhibitEX tablets, Qiagen). More details on the extraction protocols used within EnvMetaGen for the different types of vertebrate faeces are provided in Section 3.1.1.

2.1.2. Water samples

eDNA collected from water is currently being widely used for species detection and biodiversity assessments. Collection of eDNA samples in the field is usually performed either by water filtering or by precipitation (see Deliverable D4.3; Egeter et al. 2018; for details). Concerning DNA extraction from these types of samples there is a large heterogeneity in the approaches as many protocols have been optimised and applied (see e.g. Deiner et al. 2015 for details). The choice of protocols may affect the detectability of species and the sample biodiversity, and different approaches may be required depending on the target of the study (Deiner et al. 2015). Therefore, within the scope of EnvMetaGen-affiliated projects, several DNA extraction protocols of precipitated water and filters are being tested for species detection and biodiversity assessments of different taxa, in both lentic and lotic environments. The extraction protocols being tested include protocols with different lyses (mechanic and enzymatic) and DNA cleaning procedures (spin columns and magnetic beads and inclusion of a step of adsorption of inhibitors). Moreover, in the case of filter samples, the replicability

between different extractions of the same filter (extraction replicates) is also being tested. Details on the extraction protocols being tested are provided in Section 3.1.2.

2.1.3. Bulk invertebrate samples

Bulk invertebrate samples comprise a mixture of individuals that are representative of the invertebrate community (either in freshwater or above-ground habitats; see Deliverable D4.3; Egeter et al. 2018; for details). Within the frame of EnvMetaGen, two main types of bulk invertebrate samples are being analysed: i) bulk samples from macroinvertebrate sampling for water quality assessments and ii) bulk samples of flying insects, for the assessment of insect diversity and resource availability for bat diet.

These types of samples do not share the same issues regarding DNA extraction as in the previous cases of vertebrate faeces and water samples, since the amount and quality of DNA obtained from bulk samples is high. The challenge in the case of bulk samples is retrieving this large amount of DNA, while representing the diversity therein. Since samples are usually large, subsampling is commonly performed and several extraction replicates per sample may be done.

Regarding bulk samples of macroinvertebrates, there are already some studies focusing on the analyses of these samples (e.g. Elbrecht et al. 2017b). However, these mostly focus on the analyses of clean and sorted samples (Elbrecht et al. 2017a). Furthermore, DNA extraction of bulk samples, often requires their destruction. This may be an issue, due to regulatory processes that impede the destruction of the specimens. Therefore, within the context of EnvMetaGen-affiliated projects a protocol is being optimised that focus on the DNA extraction from the preservative ethanol of these samples. It has been previously shown that preservative ethanol of bulk samples of insects may be a good source of DNA for biodiversity assessments (Hajibabaei et al. 2012). However, this previous study was based also in the analyses of sorted and clean bulk samples. Therefore, we have been focusing on the development of a rapid and simple extraction protocol for metabarcoding analyses of macroinvertebrate samples for water quality assessments based on preservative ethanol. These protocols, being used for DNA extraction from preservative ethanol of macroinvertebrate bulk samples and bulk samples of flying insects, are described in Section 3.1.3.

2.2. DNA Amplification

After DNA extraction, the workflow used will depend on the main goal of the study. Most of the workflows being applied within the scope of EnvMetaGen-affiliated projects are based on targeted amplification (PCR) either using qPCR or sequencing through HTS. However, other PCR-free protocols may also be used, namely reduced representation library sequencing (e.g. double digest restriction enzyme-associated DNA sequencing, ddRAD-seq) for population genomic studies, or targeted capture techniques (see Section 4.2 for more information on how these methods may be applied in the future within the frame of EnvMetaGen-affiliated projects).

qPCR approaches are usually performed for species detection and imply the use of species-specific primer sets. Besides the power of detecting species in eDNA samples, this technique also allows the quantification of the target DNA, which has been shown, in some cases, to correlate with species abundance in the environment (Lodge et al. 2012, Thomsen et al. 2012, Lacoursière-Roussel et al. 2016). qPCR approaches are hence good for detecting rare species or a small range of species, and even for estimating their abundance. However, to detect a wide range of species using qPCR would imply the use of a high number of primer sets and increasing costs. In these situations, the use of HTS methods is valuable. Within EnvMetaGen-affiliated projects, workflows involving qPCR are starting to be optimised for targeted species detection in water samples (see Section 4.2 for details).

Approaches involving the targeted amplification and multispecies identification using HTS (DNA metabarcoding) are the more commonly applied within the scope of EnvMetaGen-affiliated projects. These are mostly used for biodiversity assessments and the analyses of diets. For DNA metabarcoding, both species-specific or general primers may be used, according to the goal of the study and available budget. Species-specific primers may be needed to assure detection of specific species (endangered, invasive or rare taxa). General primers allow the detection of a wide range of species but may fail to detect less abundant taxa (Thomsen et al. 2012). The power of detection of general primers will depend on i) the affinity to the targeted taxa sequences and ii) the availability of DNA reference collection databases needed for species identification. Within EnvMetaGen, DNA reference collections are being constructed for some taxonomic groups to cover the knowledge gaps (see Deliverable D4.2, Ferreira et al. 2018, for details) and pilot experiments are performed for the optimisation of the primers and the assessment of the power of detection of the target species.

The advantages of DNA metabarcoding include the ability of simultaneously sequencing several amplicons from multiple samples, allowing the collection of data pertaining to a large number of target species, while possibly obtaining additional data on non-target taxa. However, this technique is limited by technical biases (e.g. primer bias) that hamper reliable assessments of abundance from eDNA samples (Elbrecht and Leese 2015, Deiner et al. 2017). Although it is possible to apply some relative correction factors to DNA metabarcoding data, this is dependent upon the availability of known species mixtures to determine these correction factors (Thomas et al. 2016). There are other methods, such as capture enrichment, that are being tested, and show promising results in assessing species abundance (e.g. Dowle et al. 2016), as they avoid amplification biases. However, the design of capture probes implies extensive knowledge of the taxonomic diversity in the study area and is still very costly (Deiner et al. 2017). Therefore, within the frame of EnvMetaGen-affiliated projects, these methods are not currently being applied.

Regardless of the workflow selected for each project (qPCR or DNA metabarcoding) there are several aspects that need to be considered. The use of positive and negative controls is transversal to the used workflow, as positive controls allow to determine sequencing efficiency and negative controls allow the detection of contamination issues (Deiner et al. 2017). The number of positive and negative controls performed will depend on the study frame (e.g. number of samples, experiments, primer sets). Replication is another important aspect within the workflow, as technical replicates should reduce the rates of false negatives and help in the detection of false positives (Ficetola et al. 2015). The amount of technical replicates needed will also vary with the study goals and the probability of detection of the targeted taxa (Ficetola et al. 2015).

Within EnvMetaGen project the laboratory workflows being optimised are dependent on the main application, here summarized as single species detection, diet assessments and biodiversity assessments (see Figure 1). In the following sub-sections, we describe the general workflows according to the main applications.

2.2.1. Single species detection

Single species detection includes both the detection of endangered or invasive species in water or bulk samples as well as species identification in single samples, namely the ‘consumer’ in vertebrate faeces.

For species identification in vertebrate faeces, the most common approach is amplification with general primers (group specific), targeting mitochondrial or nuclear regions, followed by Sanger sequencing (e.g. Barbosa et al. 2013, Silva et al. 2015). These approaches target small gene fragments that show high amplification success in samples with low amounts and degraded DNA. There are a number of pipelines that were previously optimised at InBIO for single species identification of non-invasive samples, that rely on the amplification and sequencing of mitochondrial (e.g. D-loop, Cytochrome *b*) and nuclear markers (e.g. Interphotoreceptor Retinoid Binding Protein - IRBP, Kappa Casein - KCAS), for some groups of organisms, namely carnivores (Oliveira et al. 2010), rodents (Barbosa et al. 2013) and ungulates (Silva et al. 2015). Currently, these protocols are still applied in the frame of EnvMetaGen-affiliated projects, mainly for the identification of the 'consumer' in diet studies. Other approaches in these types of samples include the use of species-specific primers and/or qPCR (e.g. Alasaad et al. 2012, Thomsen et al. 2012, Davy et al. 2015). There are also protocols already used at InBIO using species-specific primers (e.g. *Galemys pyrenaicus*, Igea et al. 2013). Further protocols are being optimised using group specific primers for the identification of the 'consumer' in bat and bird diet studies. However, these new protocols use preferentially a DNA metabarcoding approach instead of Sanger sequencing, which allows species ID concurrently with the diet analyses in an effective way (see Section 2.3 for details on the DNA metabarcoding workflow).

For water or bulk samples, both qPCR protocols (see section 4) and DNA metabarcoding approaches are currently being optimised for the detection of amphibians and invasive fishes. Details on the amplification and sequencing protocols regarding single species detection are provided in Section 3.2.1.

2.2.2. Diet assessment

Several studies involving diet analyses of different taxa, including invertebrates, birds, bats, small mammals and carnivores, are being developed within the frame of EnvMetaGen. The general workflow applied in these studies is a DNA metabarcoding approach (see Section 2.3). Differences in the workflow of the different projects reside with the primer sets used, that are dependent upon the target taxa. In general, the genes targeted are Cytochrome *c* Oxidase subunit I (COI) and 16S ribosomal RNA for insects, amphibians and mammals, 12S ribosomal RNA for vertebrates in general and the chloroplast trnL intron for plants (see Section 3.2.2 for

details). For a general overview of diet composition, we also use the universal primers for eukaryotes targeting the nuclear 18S ribosomal RNA (Jarman et al. 2013). Though this gene allows the detection of a wide range of taxa, it often does not allow identification to the species level. Therefore, depending on the specific goal of the project it may be worth amplifying more than one gene fragment, allowing for a higher resolution in species identification of different taxonomic groups. Considering that most of the diet studies are based on analyses of faeces, the size of the targeted gene fragments is usually small, to overcome the effects of DNA degradation. Furthermore, as this type of samples contain predominantly DNA from the 'consumer', depending on the species diet and on the selected marker for analyses, there may be the need to use blocking primers to prevent or at least reduce the amplification and sequencing of the 'consumer' DNA. Annealing inhibiting primers have been shown to be the most efficient in blocking the amplification of 'consumer' DNA, while allowing for the detection of prey DNA previously not detected (Vestheim and Jarman 2008).

Nonetheless, during the initial phase of each project, an optimisation step is always performed to choose the best primer set(s) to use. Initially, reference sequence data is gathered from public databases (NCBI, BOLD) concerning the taxonomic group of interest and study area, for *in silico* testing of the possible primer pairs. Unfortunately, there are still several groups for which the public databases have very limited information, so an effort has been made to construct DNA barcodes reference libraries for those groups (see Deliverable D4.2; Ferreira et al. 2018; for further details on the reference collection being developed in the scope of EnvMetaGen). After an initial selection of two or three sets of primers, mock communities are created for *in vitro* testing of the primer sets. The PCR conditions are optimised to allow for the best amplification results with the least primer dimer. After optimisation, the products are sequenced to determine the coverage and the resolution of the possible primer sets considering the targeted taxa. The amplification protocols regarding diet assessment are detailed in Section 3.2.2.

2.2.3. Biodiversity assessment

Biodiversity assessments performed in the scope of EnvMetaGen-affiliated projects include determining freshwater diversity, based on water samples and bulk samples of macroinvertebrates, and assessing the diversity of flying insects, based on bulk samples, in the scope of projects studying the ecosystem services provided by bats in agroforestry systems (Appendix 1, Figure 1).

The general workflow applied in these projects is the same, and consists of a DNA metabarcoding approach. Differences in the workflow of the different projects reside with the primer sets used, which are dependent upon the target taxa (see Section 2.2.2 for details). The size of the gene fragment targeted also depends on the type of sample, with shorter fragments usually used for samples with low DNA quantity and quality, such as water samples, and larger fragments being used for samples with high DNA quantity and quality, such as bulk invertebrate samples. Further details on the amplification protocols regarding biodiversity assessments are provided in Section 3.2.3.

2.3. Library preparation for high throughput sequencing

Library preparation for sequencing is performed for all protocols that follow a metabarcoding approach. The workflow is generally the same for all projects, and consists of a DNA metabarcoding approach, based on a double indexing, two-step PCR pipeline (Figure 2). In order to allow pooling of several samples in the same sequencing run, we are using double indexing, where small barcodes ('indexes') are incorporated in both the forward and reverse Illumina adapter sequences. The first PCR is performed with the primers that target the region of interest (see Section 2.2 for details), but these primers already include the overhang adapter for the indexes and sequencing adapters. In the second PCR the sequencing adapters and indexes are incorporated into the amplicons. Two different indexes have been used in EnvMetaGen projects, Illumina indexes and custom made indexes. When Illumina indexes are used (96 combinations available at InBIO), a small multiplex identifier (MID, following Shokralla et al. 2015) is added to the overhang to allow additional pooling of samples for sequencing. Custom indexes were adapted from Kircher et al. (2012) and Gansauge and Meyer (2013) and allow pooling a maximum of 1920 samples, without the need of adding the MID to the overhang sequence, and these are currently preferentially used at InBIO. After indexing, briefly, the samples are cleaned with AMPure beads (Beckman Coulter), quantified with NanoDrop 1000 (Thermo Scientific), normalized and pooled for sequencing. Details on the library preparations protocols are provided in Section 3.3.

Sequencing is the final step of the laboratory workflow. DNA metabarcoding libraries prepared within the scope of EnvMetaGen-affiliated projects are usually sequenced on either a MiSeq or HiSeq platform, depending on the number of samples and desired coverage.

When planning a sequencing run, one important aspect to consider is the sequencing depth. Desired sequencing depth, that is number of reads per targeted sample, depends on the type of

sample and goal of the study. Highly diverse samples (e.g. for biodiversity assessments) should have higher coverage per sample, than that needed for single species identification or even diet analysis. Different libraries may be combined in the same sequencing run, but attention should be given to fragment sizes and diversity of the libraries.

In every sequencing run PhiX Control v3 library (Illumina) is included, as a library control. Moreover, using high proportions of PhiX, increases the library diversity, and sequencing performance, so usually between 15 to 20% is added to the final pool before sequencing.

The details on bioinformatics analyses of the sequencing runs are included in Deliverable D4.5 (Galhardo et al. 2018).

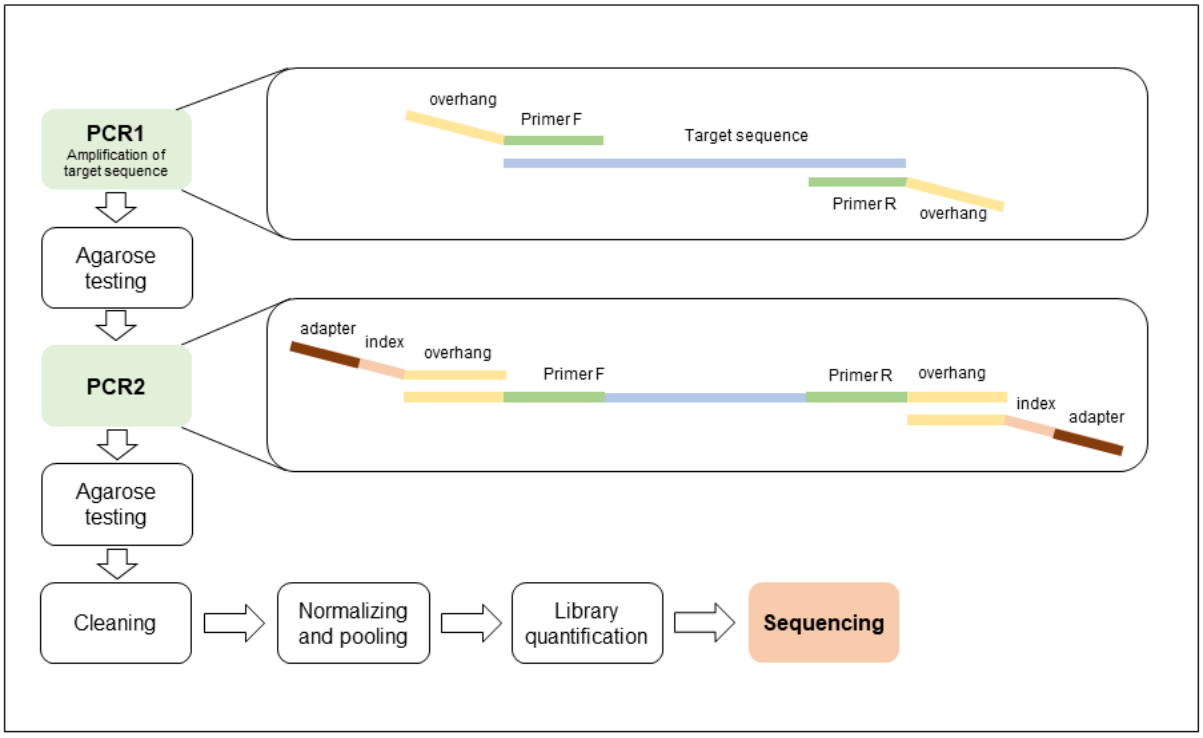

Figure 2. Amplification and library preparation workflow used in the EnvMetaGen project.

3. LABORATORY PROTOCOLS

3.1. DNA extraction protocols

Within the frame of the EnvMetaGen-affiliated projects, various extraction protocols were tested and optimised for different types of samples. The subsections below describe the different protocols tested, considering three major groupings of sample types, namely, vertebrate faeces, water samples and bulk invertebrate samples (Figure 1).

3.1.1. Vertebrate Faeces

This section includes a general description of the methods optimised for different types of vertebrate faeces: carnivore scats, rodent scats and small insectivore scats (including small mammals, bats and birds). All extractions of vertebrate faeces are performed in a dedicated room at InBIO, and following procedures to reduce the risk of contamination.

3.1.1.1. Carnivore scats

Laboratory protocols for extracting DNA from carnivore scats were optimised in the scope of previous InBIO projects for single species detection and population genetic studies. Considering that for these applications the target is the DNA from the 'consumer', the DNA extraction is usually performed targeting the outer layer of the scat. For the optimisation of the extraction protocols for diet analyses, pilot studies are being performed to assess if the extraction from the outer layer of the scat yields equivalent results to extracting homogenised scats. For the extraction from the outer layer of the scat, initially the sample, if preserved in ethanol, is dried in the oven and a portion of the outer layer of the scat is taken for DNA extraction. The protocol follows Frantz et al. (2003), after the guanidinium thiocyanate (GuSCN)/silica method (Boom et al. 1990). Briefly, the faecal material is suspended in 2ml of extraction buffer, containing GuSCN (Frantz et al. 2003), and incubated overnight at room temperature with rotation. The extracts are then centrifuged and 250ul of silica is added to the supernatant. The mix is then vortexed and centrifuged and the supernatant is discarded. The pellet is then washed twice with a guanidine washing buffer and once with an ethanol washing buffer (see Frantz et al. 2003 for details on the buffer composition). After washing, the pellet is dried and then hydrated with pure water and incubated overnight. Finally, the DNA is purified with spin columns (Frantz et al. 2003). For samples that are preserved in silica gel, instead of

ethanol, DNA is usually extracted using the same method, but with an additional prewash in Phosphate Buffered Saline (PBS) solution to release the cells into the solution (see e.g. Nakamura et al. 2017).

Tests involving the extraction of homogenised scats apply a similar protocol, but the scat is ground either manually or via bead beating, prior to DNA extraction.

Applications: Single species detection, diet assessment

EnvMetaGen Projects using this approach: IRANVERT, WOLFDIET

3.1.1.2. Rodent scats

DNA extraction protocols for rodent scats were previously optimised at InBIO, for single species detection and population genetic studies in the scope of previous projects. These protocols are based on the DNA extraction of the cells on the outer layer of a small group of faecal pellets (usually three to eight pellets from the same individual). Briefly, DNA extraction is performed using the E.Z.N.A. ® Tissue DNA Kit (Omega) using the manufacturer's protocol with modifications for non-invasive samples, namely an initial step of 15 minutes incubation at 56ºC using 300 µL of lysis washing buffer as described in Maudet et al. (2004), and without homogenizing the samples. This extraction procedure has proven to be efficient in obtaining DNA for the amplification of mitochondrial DNA for species identification and of microsatellite markers for individual identification (Ferreira et al. 2018, Sabino-Marques et al. 2018).

Applications: Single species detection, diet assessment

EnvMetaGen Projects using this approach: AGRIVOLE, MATEFRAG

3.1.1.3. Small insectivore vertebrate scats

Laboratory protocols for the extraction of DNA from small insectivore scats (including small mammals, bats and small birds) were optimised at InBIO, within the scope of EnvMetaGen-affiliated projects (Appendix 1). Three different protocols, using homogenised individual whole scats, were applied in these studies. The first protocol, applied in Mata et al. (2016), used the QIAamp DNA Stool Kit (Qiagen) following the standard protocol with adjustments suggested by Zeale et al. (2011). In the second protocol, described in Mata et al. (2018), DNA was extracted using the Stool DNA IsolationKit (Norgen Biotek Corporation), following the

manufacturer's protocol. The third protocol consists of DNA extraction with the E.Z.N.A. ® Tissue DNA Kit (Omega) using the manufactures' protocol with modifications for non-invasive samples, including two initial steps: i) 30 minutes incubation at 56ºC using 800 µL of lysis washing buffer (Maudet et al. 2004) and ii) addition of one quarter of an InhibitEX tablet (Qiagen) for adsorption of inhibitors. These protocols provide similar results in metabarcoding analyses. Currently the third protocol is being applied, as it is more cost-effective.

Applications: Single species detection, diet assessment

EnvMetaGen Projects using this approach: CHASCOS, ECOLIVES, GALEMYS, SABOR, TUA

3.1.2. Water samples

Herein we detail the extraction protocols tested for the two types of water samples: precipitation and water filtration. All extractions of water samples are performed in a dedicated room at InBIO, following procedures to reduce the risk of DNA contamination.

3.1.2.1. Precipitates

Water samples subjected to DNA precipitation are stored with ethanol and sodium acetate (see Deliverable D4.3; Egeter et al. 2018). DNA extraction from these samples follows Tréguier (2014) and Ficetola et al. (2008), with some adjustments: the samples are centrifuged at ~3000g for 1 hour; the remaining pellet is dried at 37ºC and the E.Z.N.A. ® Tissue DNA Kit (Omega) is used, with reagent volumes and incubation times adjusted for the volume of the sample.

Applications: Single species detection, biodiversity assessment

EnvMetaGen Projects using this approach: AZORES, CRAYFISH, FILTURB, XENOPUS

3.1.2.2. Filters

The vast majority of eDNA studies using water filtration use a standard 47 mm disc filter, with the pore size and material varying among studies (see Deliverable D4.3; Egeter et al. 2018). Initially, different extraction protocols were tested to evaluate which would provide the best results for assessing vertebrate species diversity using such filters, including the DNAeasy Tissue Kit (Qiagen) with QiaSchredder spin colums (Qiagen), the E.Z.N.A. ® Tissue DNA Kit (Omega) with and without QiaSchredder spin colums (Qiagen) and the E.Z.N.A. ® Water DNA

Kit (Omega). The DNA obtained through these different extraction procedures was amplified for markers used for fish detection. Results of this pilot study, indicated that E.Z.N.A. ® Tissue DNA Kit (Omega) alone would provide the best results concerning fish diversity. Therefore, current DNA extraction from filters is being performed using a quarter of each filter and the E.Z.N.A. ® Tissue DNA Kit (Omega) with the reagents volumes and incubation times adjusted for this type of sample. An additional step of 60 minutes' incubation at 54ºC using 2000 µL of lysis washing buffer as described in Maudet et al. (2004) may also be included.

Applications: Single species detection, biodiversity assessment

EnvMetaGen Projects using this approach: AZORES, FILTURB, FRESHING, GUELTA, ICVERTS, IRANVERTS

3.1.2.3. High-capacity filters

Within EnvMetaGen-affiliated projects, two high-capacity filters were trialled: one using a Cintropur NW18 filtering capsule (Airwatertech) and one FHT-45 enclosed capsule filter (Waterra). The trials revealed that the FHT-45 filters were far easier to use in the field and resulted in higher numbers species detections. Therefore, the use of the NW18 capsule filters was discontinued. The extraction protocol for high-capacity capsule filters is adapted from Valentini et al. (2016). Capsules are filled with 100 ml of resuspension buffer (50 mM Tris, 10 mM EDTA). Both ends are covered and capsules are shaken manually for 5 min. The buffer is subsequently filtered through a 0.45 µm 47 mm disc filter. Optionally, this process is repeated with a further 100 ml resuspension buffer. From this step, the extraction protocol is identical to that for standard 47 mm disc filters.

Applications: Single species detection, biodiversity assessment

EnvMetaGen Projects using this approach: FILTURB

3.1.3. Bulk invertebrate samples

Below is the description of the different methods used for DNA extraction from i) preservative ethanol of bulk macroinvertebrates samples and ii) bulk samples of flying insects.

3.1.3.1. Preservative ethanol

Initial optimisation steps included evaluating different extraction protocols for retrieving DNA from preservative ethanol of low diversity and high diversity samples. DNA extraction of samples of 2ml ethanol from the low and high diversity sites were tested with the following protocols: Nucleospin Soil Kit (Macherey Nagel), a modified E.Z.N.A. ® Tissue DNA Kit (Omega) with InhibitEX Buffer (Qiagen), and a new protocol using Agencourt® AMPure® XP beads (Beckman Coulter Company, Massachusetts, United States) and Qiagen® buffers, developed within the frame of the project. The DNA obtained through these different extraction procedures was amplified with a COI primer set (313bp, see Table 3 in Section 3.2.) designed for a group of macroinvertebrate insects (Ephemeroptera, Plecoptera, Trichoptera and Odonata). This pilot study showed that the protocol using beads was the one that provided the best results concerning the overall coverage and diversity detected.

Applications: Single species detection; biodiversity assessment

EnvMetaGen Projects using this approach: FRESHING

3.1.3.2. Bulk samples of flying insects

These bulk samples contain mostly flying insects which are first dried in the oven at 37ºC and then homogenised with a Bullet Blender 50 DX homogeniser (Next Advance) with glass beads. DNA extraction is then performed on the homogenised samples with the E.Z.N.A. ® Tissue DNA Kit (Omega) with the reagents volumes and incubation times adjusted for the volume of the sample, and optionally including two initial steps: i) 30 minutes' incubation at 56ºC using 1000 µL of lysis washing buffer (Maudet et al. 2004), and ii) addition of half an InhibitEX tablet (Qiagen) for adsorption of potential inhibitors.

Applications: Single species detection; biodiversity assessment

EnvMetaGen Projects using this approach: ECOLIVES, TUA

3.2. DNA Amplification protocols

Laboratory protocols following DNA extraction are dependent upon the main application of the study. This section describes the various laboratory protocols applied within the different projects.

3.2.1. Single species detection

Several protocols have been optimised at InBIO for single species detection. Herein we describe those currently being applied in the scope of EnvMetaGen-affiliated projects.

3.2.1.1. PCR with species-specific or group-specific primers for Sanger sequencing

After DNA extraction, PCR amplification is performed using primer sets previously optimised to target gene regions that allow species identification of the target taxa. In Table 2 we summarize the primers currently used for species ID. PCRs are performed in a 10ul volume, using 5uL of Qiagen© PCR Multiplex Kit Master Mix (Qiagen, Hilden, Germany) or Mytaq™ Mix (Bioline), 0.3 to 0.4uLof each primer and 1 to 2uLof DNA. The PCR thermo profile is also detailed in Table 2. If samples fail to amplify, PCR replicates are performed. Two negative controls are included in the PCR, the negative control from the extraction and a PCR negative control. The obtained PCR products are then purified with ExoSAP-IT® PCR clean-up Kit (GE Healthcare, Piscataway, NJ, USA) and sequences are generated with the amplification primers. Sequencing reactions are carried out using BigDye® Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems, Carlsbad, CA, USA) and samples are subsequently sequenced for the reverse strand on a 3130xl Genetic Analyser Sequencer (Applied Biosystems/HITASHI).

Sample type: Vertebrate faeces

EnvMetaGen Projects using this approach: AGRIVOLE, GALEMYS, MATEFRAG

Table 2. Primer sets and PCR conditions used for single species identification for different targeted taxa.

Taxa	Gene	Forward primer	Reverse primer	Ta	Reference	Seq[1]
Rodents	Cyt*b*	CYTB.F2S	CYTB.R2S	TD 56º-51ºC	Barbosa et al. 2013	Sanger
Microtus cabrerae	D-loop	Pro+	MicoMico	60ºC	Alasaad et al. 2011	Sanger
Galemys pyrenaicus	Cyt*b*	GpCytb-F2	GpCytb-R2	54ºC	Igea et al. 2013 Quaglietta et al. 2018	Sanger
Vertebrates	12S	12S-V5F	12S-V5R	54ºC	Riaz et al. 2011	HTS
Mammals	16S	16SmamF	16SmamR	54ºC	Ficetola et al. 2010	HTS
Bats	COI	SFF 145 F	SFF 351 R	52ºC	Walker et al. 2016	HTS

[1] – Sequencing workflow; HTS – High throughput sequencing

3.2.1.2. Amplification with group specific primers for HTS sequencing

In this situation the amplification protocol is the same as described above for Sanger sequencing, but the primers used for the initial amplification already have the overhang adapters for indexes and sequencing adapters (see workflow for DNA metabarcoding in Section 2.3 and Figure 2). This protocol has been applied for identifying bat and large mammal species, with primers and conditions shown in Table 2. The PCR products, including negative controls, follow then the library preparation protocols described in Section 3.3.

Sample type: Vertebrate faeces

EnvMetaGen Projects using this approach: ECOLIVES, IRANVERT, SABOR, TUA

3.2.2. Diet assessment

Laboratory protocols for application in diet studies follow the general DNA metabarcoding workflow as describe in Section 2.3. Several primers sets were already optimised for amplification of the different components of the diet of the target species, including insects, vertebrates and plants (see Table 3 for details on primers, fragment size and annealing temperatures). The PCRs are performed with primers modified to contain the overhang adapters for indexes and sequencing adapters (Figure 2). PCRs are performed in a 10uLvolume, using 5uL of Qiagen© PCR Multiplex Kit Master Mix (Qiagen, Hilden, Germany), 0.3 to 0.4uLof each primer and 1 to 2uLof DNA. Negative controls are always used in the PCR reaction, and usually 1 to 3 PCR replicates are performed. The PCR products, including negative controls, are then subject to the library preparation protocols described in section 3.3.

Sample type: Vertebrates faeces

EnvMetaGen Projects using this approach: AGRIVOLE, CHASCOS, ECOLIVES, GALEMYS, SABOR, WOLFDIET, TUA, NZFROG

3.2.3. Biodiversity assessment

Laboratory protocols for application in biodiversity assessments follow the general DNA metabarcoding workflow as describe in Section 2.3.

Amplification protocols have been optimised according to the goals of the different projects using this approach (see below) and target in general mitochondrial genes. The PCRs are

performed with primers modified to contain the overhang adapters for indexes and sequencing adapters. The primer sets already optimised for biodiversity assessments, as well as the annealing temperatures used, are shown in Table 3. PCRs are performed in a 10uLvolume, using 5uLof Qiagen© PCR Multiplex Kit Master Mix (Qiagen, Hilden, Germany), 0.3 to 0.4uL of each primer and 1 to 2uL of DNA. In some situations, it is necessary to perform DNA dilutions before amplification. Negative controls are always used in the PCR reaction, and usually 1 to 3 PCR replicates are performed.

The PCR products, including negative controls, are then subject to the library preparation protocols described in Section 3.3.

Sample type: Water samples (filters and precipitates), bulk samples

EnvMetaGen Projects using this approach: AZORES, ECOLIVES, FILTURB, FRESHING, GUELTA, ICVERTS, IRANVERTS, TUA, XENOPUS

Table 3. Primer sets and PCR conditions used for diet and biodiversity assessments, for targeted taxa. All primers were modified to contain the overhang adapters for indexes and sequencing adapters.

Taxa	Gene	Forward primer	Reverse primer	Length[1] (bp)	Ta	Reference
Arthropods	COI	BF2[2]	BR2[2]	421	45ºC	Elbrecht and Leese, 2017
Arthropods	COI	ZBJ-ArtF1c	ZBJ-ArtR2c	157	TD 61ºC-53ºC	Zeale et al. 2011 Mata et al. 2016
EPTO[3]	COI	CR_Tricho[2]	BR2[2]	313	54ºC	Elbrecht and Leese, 2017 This study
Insects	COI	fwhF2	fwhR2	205	50ºC	Vamos et al. 2017
Eukaryotes	18S	18Sf-msq	18Sf-msq	170	67ºC	Jarman et al. 2013
Vertebrates	12S	12S-V5F	12S-V5R	110	54ºC	Riaz et al. 2011
Mammals	16S	16SmamF	16SmamR	140	54ºC	Ficetola et al. 2010
Frogs	16S	Egeter-Frog-16SF	Egeter-Frog-16SR	170	63ºC	This study
Fishes	Mifish	MiFish-U-F	MiFish-U-R	185	65ºC	Miya et al. 2015
Plants	Trnl	trnL-g-F	trnL-h-R	143	50ºC	Taberlet et al. 2007

[1] – maximum length of the fragment; [2] – only for biodiversity assessments; [3] – Ephemeroptera, Plecoptera, Trichoptera and Odonata.

3.3. Library preparation for high throughput sequencing

After the first targeted PCR has been performed, all studies follow the same general workflow (see Section 2.3. and Figure 2).

Initially, all PCR products are analysed by agarose gel electrophoresis. Samples selected for sequencing are then organised for assignment of sequencing 'indexes' (refers to the sequences including sequencing adapters and indexes). Negative and positive controls are usually included in the sequencing run. If Illumina indexes are used, then samples are pooled into one plate (96 combinations of indexes available) before the second PCR reaction. When using custom indexes, there are a maximum of 1920 unique index combinations available at InBIO laboratories, hence, in one sequencing run, there is the possibility of mixing a maximum of 1920 samples per primer set. In this situation, there is no need for pooling samples before the second PCR reaction.

A second PCR reaction, index PCR, is then performed in a volume of 10ul, including 5uL of Phusion® High-Fidelity PCR Kit (New England Biolabs) or KAPA HiFi PCR Kit (KAPA Biosystems, USA), 0.5uLof each 'index' and 2uL of diluted PCR product (usually 1:4). This is a shorter PCR reaction, with only 10 cycles and an annealing temperature of 55ºC. A few randomly selected samples, are then tested by agarose gel electrophoresis, to ensure incorporation of the index in the PCR product (a difference of c. 100bp in size is expected).

Then, the amplicons are purified using AMPure XP beads (New England Biolabs). The proportion of beads used will vary according to the length of the target amplicon - higher concentrations of beads allow the recovery of smaller fragments of DNA. Following DNA purification, the amplicons are quantified for normalization and pooling of samples. Quantification is performed using NanoDrop 1000 (Thermo Scientific) following the manufacturer instructions. Then the samples are normalized (diluted to approximately similar concentrations) and pooled per amplicon.

Final quantification of the amplicon libraries is performed with qPCR using the KAPA Library Quantification Kit Illumina® Platforms (KAPA Biosystems, USA). Finally, the libraries per amplicon are pooled together in proportions that take into account the coverage needed for each amplicon in the sequencing run. Usually 15 to 20% PhiX Control v3 library (Illumina) is included in the final pool, as a control library and to increase the diversity of the sample, hence balancing the representation of the bases in the initial cycles and improving cross-talk and phasing calculations in the sequencing run.

4. OTHER DEVELOPING AREAS FOR THE PROJECT

In addition to the already optimised laboratory workflows within the frame of EnvMetaGen-affiliated projects, there are other protocols in early development and optimisation stages or which may be implemented in a near future. Below is a brief description of some of these areas and applications.

4.1. Obtaining DNA from other types of samples

4.1.1. Analyses of stomach contents

Molecular stomach content analyses have been widely used to assess vertebrate diets (Pompanon et al. 2012). Some studies have found that DNA extracted from stomach contents provides far greater prey detection success than DNA extracted from faecal samples (e.g. Egeter et al. 2015). However, a consideration is that prey DNA detection times following prey ingestion are likely to be shorter using stomach content samples, and stomach collection is highly invasive, involving the euthanasia of individuals.

Within EnvMetaGen-affiliated projects, there are two projects that target diet analyses using stomach contents. NZFROG, is already ongoing, and though the EnvMetaGen team hasn't been involved in the sample collection nor DNA extraction (see Appendix), the amplification and library preparation for sequencing was performed at InBIO, as described in Sections 3.2.2. and 3.3. CRAYFISH, is a project in its early stage of development, that targets the diet of the invasive crayfish (see Appendix). Within the frame of this project the optimisation of protocols for DNA extraction from crayfish stomachs will be carried out.

4.1.2. Analyses of invertebrate faeces

Assessing diets of invertebrates is challenging, as 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). Hence, the field of invertebrate ecology has largely pioneered research in the area of diet assessments using molecular tools (King et al. 2008; Sheppard and Harwood 2005; Symondson 2002). One EnvMetaGen-affiliated project, MANTIDS, aims to apply metabarcoding methods to characterise the diet of Mantids (see Appendix). 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. Therefore, in the frame of this project, which is in an early stage, the laboratory methods for the DNA extraction and further analyses of invertebrate faeces will be optimised. If successful, this approach may be extended to other projects.

4.1.3. eDNA from soil

Soil is a relevant source of DNA of multiple taxa. Several protocols have already been developed to recover extracellular DNA from soil samples (e.g. Taberlet et al. 2012). Moreover, DNA metabarcoding from soil samples has already been successfully used for assessing plant diversity (Yoccoz et al. 2012).

There are two EnvMetaGen-affiliated projects that include the analyses of soil for assessing plant diversity (AGRIVOLE) and for studying the phosphorus cycle (SOILPHOS). SOILPHOS is an ongoing project, that is using 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 (see Appendix). These samples were made available by collaborators, who also carried out the initial PCRs, while the library preparation, sequencing and further analyses was performed by the EnvMetaGen team, using the protocols described on Section 3.3. Nevertheless, it is possible that EnvMetaGen will continue both field and lab research in this area, for which the optimisation of protocols for DNA extraction will be relevant. AGRIVOLE is a project in its early steps, that 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 (see Appendix). Therefore, it is expected that, in the frame of these two projects, the laboratory procedures for DNA extraction and amplification from soil samples will be fully optimised.

4.2. Other laboratory protocols

All the laboratory protocols currently optimised are based in the amplification of targeted DNA fragments. Although these methods are quite effective in determining the biodiversity, diets and for species identification, they have some limitations, mainly regarding the assessment of

abundance of several taxa in mixed samples. Determining abundance of populations and species is an important aspect for accurate estimates of diversity and for conservation and management.

Within the methods involving amplification, qPCR approaches have the advantage of allowing the quantification of the species DNA, and have been shown to correlate with species abundance in the environment (Lodge et al. 2012, Thomsen et al. 2012, Lacoursière-Roussel et al. 2016). Within EnvMetaGen-affiliated projects, there is one project FILTURB, that aims to test the efficiency of different DNA capture methods in turbid waters for eDNA recovery and species detection (see Appendix) and apply qPCR methods. Another project, XENOPUS, aims to use qPCR to detect the presence of an invasive amphibian species across numerous sites in Portugal. However, as these procedures are in an early phase, they are not included in this deliverable, but are expected to be available for application in other projects soon.

There are also other methods, currently being tested, that do not rely on PCR amplification, such as capture enrichment, that show promising results in assessing species abundance (e.g. Dowle et al. 2016), as they avoid amplification biases. However, the design of capture probes implies extensive knowledge on the taxa diversity on the study area and is still highly costly (Deiner et al. 2017).

Other laboratory HTS protocols, may be considered valuable for other applications, namely population genomics of non-invasive samples, such as vertebrate faeces. Populations genetics analyses of non-invasive samples (e.g. faeces) is becoming increasingly used for population monitoring, since it allows to assess reproductive behaviour (e.g Schmidt et al. 2016), determine connectivity (e.g Stansbury et al. 2016) and estimate population size (Woodruff et al. 2016, Sabino-Marques et al. 2018). At InBIO, there are already several pipelines optimised for genetic non-invasive sampling, that rely on genotyping a small panel of microsatellite with high variability and low genotyping errors (e.g. Ferreira et al. 2018). Recently, it has been shown that HTS methods applied for microsatellite genotyping of non-invasive samples, provide higher genotyping success rates and lower genotyping error rates (de Barba et al. 2016). Moreover, the use of HTS may help reduce costs and time, as it provides the opportunity to overcome the technical difficulties in allele scoring. Currently, microsatellite libraries for HTS sequencing are being optimised at InBIO in the scope of the EnvMetaGen-affiliated project GALEMYS (see Appendix). The laboratory workflow for HTS sequencing of microsatellites, follows the same pipeline as the DNA metabarcoding for biodiversity assessment and diet studies (see above), using a double indexing, two-step PCR approach. However, as this

workflow is still in the optimisation phase it is not included in this deliverable. We expect that this methodology, once fully optimised will be highly relevant for other studies with species of conservation concern.

In addition to this, we are also aiming to optimize PCR free, HTS techniques for population genomics. Initial trials using reduced representation library sequencing (Genotyping-by-Sequencing, GBS) with pools of faecal samples show promising results in the recovery of DNA from target taxa and in the identification of SNPs for population genetic analyses. However, the feasibility of application of these methods to individually tagged samples remains untested, due to the potentially high proportion of missing data caused by DNA degradation of the individual samples. If feasible, this would be an extremely useful tool for population genomic studies on elusive endangered species. It is expected that further testing will be performed to assess the reliability of these methods in eDNA.

5. CONCLUDING REMARKS

This report provides a description of the current protocols for the next-gen analyses of eDNA samples, focusing on eDNA collected from three main sample types (vertebrate faeces, water samples and bulk samples) and targeting three main applications (single species detection, diet assessment and biodiversity assessment), which are fully aligned with the key application areas for triple helix initiatives of the EnvMetaGen project.

An overview was included in this report, regarding the issues and challenges that need to be considered in the selection of each protocol, as these choices affect greatly the downstream results. The implementation and optimisation of best-practices for the analyses of eDNA is one of the main goals of EnvMetaGen, and the assembly of information to support decisions and choices of protocols for each application has been one of the focal concerns in this project. Hence, the EnvMetaGen project, along with its numerous affiliated projects and collaborators, has been developing efforts for assessing the best methods of DNA extraction and subsequent workflow. A number of projects have compared different DNA extraction methods for the different types of samples, including AZORES, FRESHING, IRANVERT, MATEFRAG, WOLFDIET, SABOR. FRESHING has been trialling methods for DNA extraction from preservative ethanol from unclean and unsorted samples, which will be highly valuable considering the regulatory processes related with water quality assessment. SABOR focused on the effects of extracting DNA from either pooled or single faecal samples for downstream results, concerning the assessment of the diet of bats.

All EnvMetaGen-affiliated 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 contribute to strengthen triple-helix initiatives in all three key areas of application that have been targeted: freshwater species detection, natural pest control services and biomonitoring. This report compiles a list of working best practice protocols for the next-gen analyses of eDNA samples. Together, Deliverables D4.2-D4.5 (this document; Egeter et al. 2018; Ferreira et al. 2018; Galhardo 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.

6. HOW TO CITE

Paupério J, Fonseca N, Egeter B, Galhardo M, Ferreira S, Oxelfelt F, Aresta S, Martins F, Mata V, Veríssimo J, Puppo P, Pinto JC, Chaves C, Garcia-Raventós A, Peixoto S, da Silva LP, Vasconcelos S, Gil P, Khalatbari L, Jarman S and Beja P (2018). Deliverable 4.4 (D4.4): Protocol for next-gen analysis of eDNA samples, EnMetaGen project (Grant Agreement No 668981). European Union Horizon 2020 Research & Innovation Programme – H2020-WIDESPREAD-2014-2 doi: 10.5281/zenodo.2586885

7. REFERENCES

Admassu B, Juen A, 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.

Alasaad S, Soriguer RC, Jowers MJ, Marchal JA, Romero I, Sanchez A (2011) Applicability of mitochondrial DNA for the identification of Arvicolid species from faecal samples: a case study from the threatened Cabrera's vole. Molecular Ecology Resources, 11: 409-414.

Alasaad S, Sánchez A, García-Mudarra JL, Jowers MJ, Pérez JM, Marchal JA, Romero I, Garrido-García JA, Soriguer RC (2012) Single-tube HotSHOT technique for the collection, preservation and PCR-ready DNA preparation of faecal samples: The threatened Cabrera's vole as a model. European Journal of Wildlife Research, 58: 345-350.

Barbosa S, Paupério J, Searle JB, Alves PC (2013) Genetic identification of Iberian rodent species using both mitochondrial and nuclear loci: application to noninvasive sampling. Molecular Ecology Resources, 13: 43–56.

Beja-Pereira A, Oliveira R, Alves PC, Schwartz MK, Luikart G (2009) Advancing ecological understandings through technological transformations in noninvasive genetics. Molecular Ecology Resources, 9:1279–1301.

Biggs J, Ewald N, Valentini A, Gaboriaud C, Dejean T, Griffiths RA, Foster J, Wilkinson JW, Arnell A, Brotherton P, Williams P, Dunn F (2015) Using eDNA to develop a national citizen science-based monitoring programme for the great crested newt (Triturus cristatus). Biological Conservation, 183: 19–28.

Boom R, Sol CJA, Salimans MMM, Jansen CL, Wertheim-van Dillen, PME, van der Noordaa, J (1990) Rapid and simple method for purification of nucleic acids. Journal of Clinical Microbiology, 28: 495-503.

Davy CM, Kidd AG, Wilson CC (2015) Development and validation of environmental DNA (eDNA) markers for detection of freshwater turtles. PLoS ONE, 10: 1-13.

De Barba M, Miquel C, Boyer F, Mercier C, Rioux D, Coissac E, Taberlet P (2014) DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: Application to omnivorous diet. Molecular Ecology Resources, 14: 306–323.

Deiner K, Walser J-C, Mächler E, Altermatt F (2015) Choice of capture and extraction methods affect detection of freshwater biodiversity from environmental DNA. Biological Conservation, 183: 53-63.

Deiner K, Bik HM, Mächler E, Seymour M, Lacoursière-Roussel A, Altermatt F, Creer S, Bista I, Lodge DM, de Vere N, Pfrender M, Bernatchez L (2017) Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. Molecular Ecology, 26: 5872-5895.

Dowle E, Pochon X, Banks J, Shearer K, Wood SA (2016) Targeted gene enrichment and high-throughput sequencing for environmental biomonitoring: a case study using freshwater macroinvertebrates. Molecular Ecology Resources, 16: 1240-1254.

Egeter B, Bishop PJ, 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, 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) Deliverable 4.3 (D4.3): Protocol for field collection and preservation of eDNA samples, EnvMetaGen project (Grant Agreement No 668981). European Union Horizon 2020 Research & Innovation Programme - H2020-WIDESPREAD-2014-2 doi: 10.5281/zenodo.2579806.

Elbrecht V, Leese F (2015) Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass-sequence relationships with an innovative metabarcoding protocol. PLoS ONE, 7.

Elbrecht V, Leese F (2017) Validation and Development of COI Metabarcoding Primers for Freshwater Macroinvertebrate Bioassessment. Frontiers in Environmental Science, 5: 11.

Elbrecht V, Peinert B, Leese F (2017a) Sorting things out: Assessing effects of unequal specimen biomass on DNA metabarcoding. Ecology and Evolution, 7: 6918-6926.

Elbrecht V, Vamos EE, Meissner K, Aroviita J, Leese F (2017b) Assessing strengths and weaknesses of DNA metabarcoding-based macroinvertebrate identification for routine stream monitoring. Methods in Ecology and Evolution, 8: 1265-1275.

Evans NT, Li Y, Renshaw MA, Olds BP, Deiner K, Turner CR, Jerde CL, Lodge DM, Lamberti GA, Pfrender ME (2017). Fish community assessment with eDNA metabarcoding: effects of sampling design and bioinformatic filtering. Canadian Journal of Fisheries and Aquatic Sciences, 1–13.

Ferreira CM, Sabino-Marques H, Barbosa S, Costa P, Encarnação C, Alpizar-Jara R, Pita R, Beja P, Mira A, Searle JB, Paupério J, Alves PC (2018) Genetic non-invasive sampling (gNIS) as a cost effective tool for monitoring elusive small mammals. European Journal of Wildlife Research, 64: 46.

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) Deliverable 4.2 (D4.2): Protocol for building and organising reference collections of DNA sequences, EnvMetaGen project (Grant Agreement No 668981). European Union Horizon 2020 Research & Innovation Programme - H2020-WIDESPREAD-2014-2 doi: 10.5281/zenodo.2586893.

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

Ficetola GF, Coissac E, Zundel S, Riaz T, Shehzad W, Bessière J, Taberlet P and Pompanon F (2010) An in silico approach for the evaluation of DNA barcodes. BMC Genomics, 11: 434.

Ficetola GF, Pansu J, Bonin A, Coissac E, Giguet-Covex C, De Barba M, Gielly L, Lopes CM, Boyer F, Pompanon F, Rayé G, Taberlet P (2015) Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. Molecular Ecology Resources, 15: 543-556.

Frantz AC, Pope LC, Carpenter PJ, Roper TJ, Wilson GJ, Delahay RJ, Burke T (2003) Reliable microsatellite genotyping of the Eurasian badger (*Meles meles*) using faecal DNA. Molecular Ecology, 6: 1649-1661.

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) Deliverable 4.5 (D4.5): Protocol for the processing of DNA sequence data generated by next-gen platforms, EnMetaGen project (Grant Agreement No 668981). European Union Horizon 2020 Research & Innovation Programme – H2020-WIDESPREAD-2014-2 doi: 10.5281/zenodo.2586889.

Gansauge M.T., Meyer M (2013) Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. Nature Protocols, 8: 737-748.

Gellie NJC, Mills JG, Breed MF, Lowe AJ (2017) Revegetation re-wilds the soil bacterial microbiome of an old field. Molecular Ecology 26, 2895–2904.

Godinho R, López-Bao JV, Castro D, Llaneza L, Lopes S, Silva P, Ferrand N (2015) Real-time assessment of hybridization between wolves and dogs: Combining noninvasive samples with ancestry informative markers. Molecular Ecology Resources, 15: 317–328.

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 JL, Shokralla S, Konynenburg S (2012) Assessing biodiversity of a freshwater benthic macroinvertebrate community through non-destructive environmental barcoding of DNA from preservative ethanol. BMC Ecology 2012, 12: 28.

Huver JR, Koprivnikar J, Johnson PTJ, Whyard S (2015) Development and application of an eDNA method to detect and quantify a pathogenic parasite in aquatic ecosystems. Ecological Applications, 25:991-1002.

Igea J, Aymerich P, Fernández-González A, González-Esteban J, Gómez A, Alonso R, Gosálbez J, Castresana J (2013) Phylogeography and postglacial expansion of the endangered semi-aquatic mammal Galemys pyrenaicus. BMC evolutionary biology, 13: 115.

Jarman SN, McInnes JC, Faux C, Polanowski AM, Marthick J, Deagle BE, Southwell C, Emmerson L (2013) Adélie Penguin Population Diet Monitoring by Analysis of Food DNA in Scats. PLoS ONE, 8: e82227.

Jiang J, Alderisio KA, Singh A, Xiao L (2005) Development of Procedures for Direct Extraction of *Cryptosporidium* DNA from Water Concentrates and for Relief of PCR Inhibitors. Applied and Environmental Microbiology, 71: 1135-1141.

Kartzinel TR, Pringle RM (2015) Molecular detection of invertebrate prey in vertebrate diets: Trophic ecology of Caribbean island lizards. Molecular Ecology, 15: 903-914.

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

Kircher M, Sawyer S, Meyer M (2012) Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Research, 40: 1-8.

Lacoursière-Roussel G, Côté G, Leclerc V, Bernatchez L (2016) Quantifying relative fish abundance with eDNA: a promising tool for fisheries management. Journal of Applied Ecology, 53: 1148-1157.

Lodge DM, Turner CR, Jerde CL, Barnes MA, Chadderton L, Egan SP, Feder JL, Mahon AR, Pfender ME (2012) Conservation in a cup of water: Estimating biodiversity and population abundance from environmental DNA. Molecular Ecology, 21: 2555-2558.

Mata VA, Amorim F, Corley MFV, McCracken GF, Rebelo H, Beja P (2016) Female dietary bias towards large migratory moths in the European free-tailed bat (Tadarida teniotis). Biology Letters, 12: 20150988.

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

Maudet C, Luikart L, Dubray D, von Hardenberg A, Taberlet P (2004) Low genotyping error rates in wild ungulate faeces sampled in winter. Molecular Ecology Notes, 4: 772-775.

Mestre F, Pita R, Paupério J, Martins FMS, Alves, PC, Mira A, Beja P (2015) Combining distribution modelling and non-invasive genetics to improve range shift forecasting. Ecological Modelling, 297: 171–179.

Miya M, Sato Y, Fukunaga T, Sado T, Poulsen JY, Sato K, Minamoto T, Yamamoto S, Yamanaka H, Araki H, Kondoh M, Iwasaki W (2015). MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: Detection of more than 230 subtropical marine species. Royal Society Open Science, 2.

Nakamura M, Godinho R, Rio-Maior H, Roque S, Kaliontzopoulou A, Bernardo J, Castro D, Lopes S, Petrucci-Fonseca F, Álvares F (2017) Evaluating the predictive power of field variables for species and individual molecular identification on wolf noninvasive samples. European Journal of Wildlife Resources, 63: 53-63.

Oliveira R, Castro D, Godinho R, Luikart G, Alves PC (2010) Species identification using a small nuclear gene fragment: Application to sympatric wild carnivores from South-western Europe. Conservation Genetics, 11: 1023-1032.

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

Quaglietta L, Paupério J, Martins FMS, Alves PC, Beja P (2018) Recent range contractions in the globally threatened Pyrenean desman highlight the importance of stream headwater refugia. Animal Conservation. *In press*

Riaz T, Shehzad W, Viari A, Pompanon F, Taberlet P, Coissac E (2011) ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis. Nucleic Acids Research, 39: e145. doi: 10.1093/nar/gkr732.

Sabino-Marques H, Ferreira CM, Paupério J, Costa P, Barbosa S, Encarnação C, Alpizar-Jara R, Alves PC, Searle JB, Mira A, Beja P, Pita R (2018) Combining non-invasive genetics with spatially explicit capture-recapture models for density estimation of a patchily distributed small mammal. European Journal of Wildlife Research, 64: 44.

Schmidt K, Davoli F, Kowalczyk R, Randi E (2016) Does kinship affect spatial organization in a small and isolated population of a solitary felid: The Eurasian lynx? Integrative Zoology, 11: 334–349.

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

Shokralla S, Porter TM, Gibson JF, Dobosz R, Janzen DH, Hallwachs W, Golding GB, Hajibabaei M (2015) Massively parallel multiplex DNA sequencing for specimen identification using an Illumina MiSeq platform. Scientific Reports, 5: 9687.

Silva T, Godinho R, Castro D, Abáigar T, Brito JC, Alves PC (2015) Genetic identification of endangered North African ungulates using noninvasive sampling. Molecular Ecology Resources, 15: 652-661.

Srivathsan A, Ang A, Vogler AP Meier R (2016) Fecal metagenomics for the simultaneous assessment of diet, parasites, and population genetics of an understudied primate. Frontiers in Zoology, 13: 17.

Stansbury CR, Ausband DE, Zager P, Mack CM, Waits LP (2016) Identifying gray wolf packs and dispersers using noninvasive genetic samples. Journal of Wildlife Management, 80: 1408–1419.

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, Pompanon F, Gielly L, Miquel C, Valentini A, Vermat T, Corthier G, Brochmann C, Willerslev E (2007). Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. Nucleic Acids Research, 35: e14.

Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. Molecular Ecology, 21: 2045-2050.

Thomas AC, Deagle BE, Eveson JP, Harsch CH, Trites AW (2016) Quantitative DNA metabarcoding: Improved estimates of species proportional biomass using correction factors derived from control material. Molecular Ecology Resources, 16: 714-726.

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

Tréguier A, Paillisson JM, Dejean T, Valentini A, Schlaepfer MA, Roussel JM (2014) Environmental DNA surveillance for invertebrate species: advantages and technical limitations to detect invasive crayfish *Procambarus clarkii* in freshwater ponds. Journal of Applied Ecology, 51: 871–879.

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

Walker FM, Williamson CHD, Sanchez DE, Sobek CJ, Chambers CL (2016) Species From Feces: Order-Wide Identification of Chiroptera From Guano and Other Non-Invasive Genetic Samples. PLoS ONE 11: e0162342.

Woodruff SP, Lukacs PM, Christianson D, Waits LP (2016) Estimating Sonoran pronghorn abundance and survival with fecal DNA and capture–recapture methods. Conservation Biology, 30: 1102–1111.

Vamos EE, Elbretch V, Leese F (2017) Short COI markers for freshwater macroinvertebrate metabarcoding. PeerJ Preprint.

Vestheim H, Jarman SN (2008) Blocking primers to enhance PCR amplification of rare sequences in mixed samples – a case study on prey DNA in Antarctic krill stomachs. Frontiers in Zoology, 5: 12.

Yoccoz N (2012) The future of environmental DNA in ecology. Molecular Ecology, 21: 2031-2038.

Yoccoz NG, Brathen KS, Gielly L, Haile J, Edwards ME, Goslar T, Von Stedingk H, Brysting AK, Coissac E, Pompanon F, Sønstebø JH, Miquel C, Valentini A, De Bello F, Chave J, Thuiller W, Wincker P, Cruaud C, Gavory F, Rasmussen M, Gilbert MTP, Orlando L, Brochmann C, Willerslev E, Taberlet P (2012) DNA from soil mirrors plant taxonomic and growth form diversity. Molecular Ecology, 21: 3647–3655.

Yu DW, Ji Y, Emerson BC, Wang X, Ye C, Yang C, Ding Z (2012) Biodiversity soup: Metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. Methods in Ecology and Evolution, 3: 613–623.

Zeale MRK, Butlin RK, Barker GLA, Lees D, Jones G. (2011) Taxon-specific PCR for DNA barcoding arthropod prey in bat faeces. Molecular Ecology Resources, 11: 236–244.

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 (Egeter et al. 2018; Ferreira et al. 2018; Galhardo 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), 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). 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).