LifeWatch ERIC

ENVRI FAIR

LifeWatch ERIC e-Science for NIS research workshop | 20-21 May 2021

An event organised within the framework of the LifeWatch ERIC Internal Joint Initiative
and with the support of the ENVRI-FAIR project

eDNA metabarcoding as a tool for early detection of nonindigenous freshwater species

Ana Filipa Filipe, Filipa MS Martins, Mafalda Galhardo,
Nuno Fonseca, Rui Figueira & Pedro Beja

CIBIO InBIO RESEARCH NETWORK
Biodiversity and Evolutionary Biology

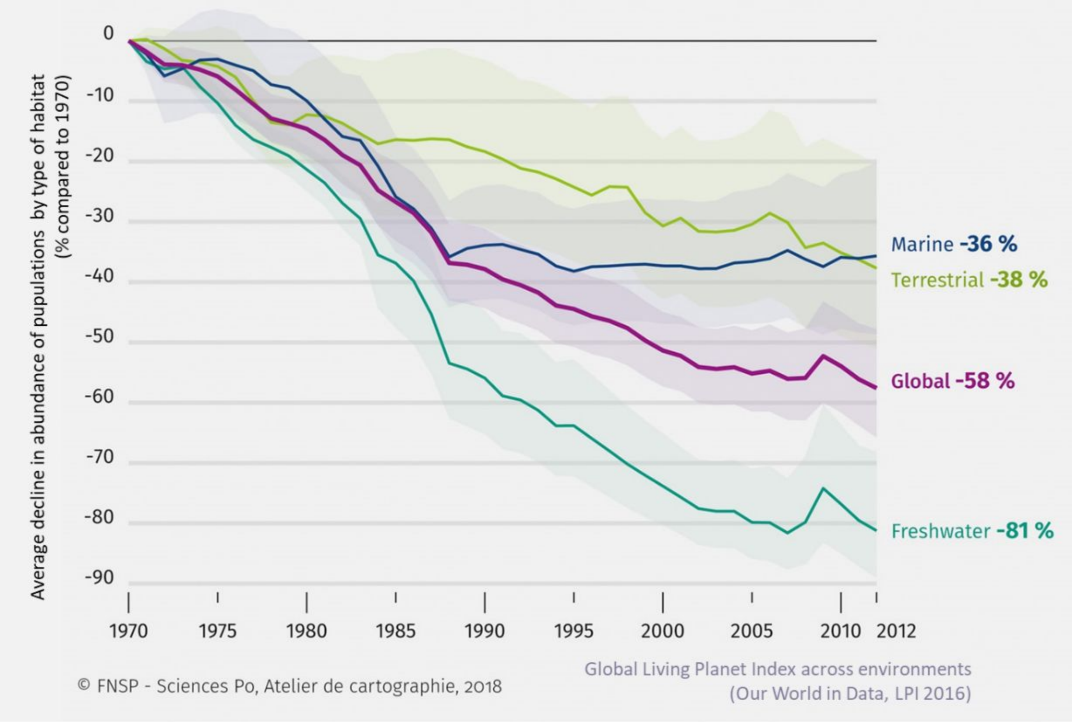
José F. Aldana Montes

khaos RESEARCH

Summary

1. BACKGROUND
   - Freshwater biodiversity
   - Nonindigenous invasive species (NIS)
2. APPROACH
   - eDNA metabarcoding
   - Reference collection
   - Steps
3. Validation Case #5: Metabarcoding for detecting NIS
   - Aim
   - Workflow
   - Service level isolation
4. CONCLUSIONS and CHALLENGES

Freshwater biodiversity



© FNSP - Sciences Po, Atelier de cartographie, 2018

Global Living Planet Index across environments
(Our World in Data, LPI 2016)

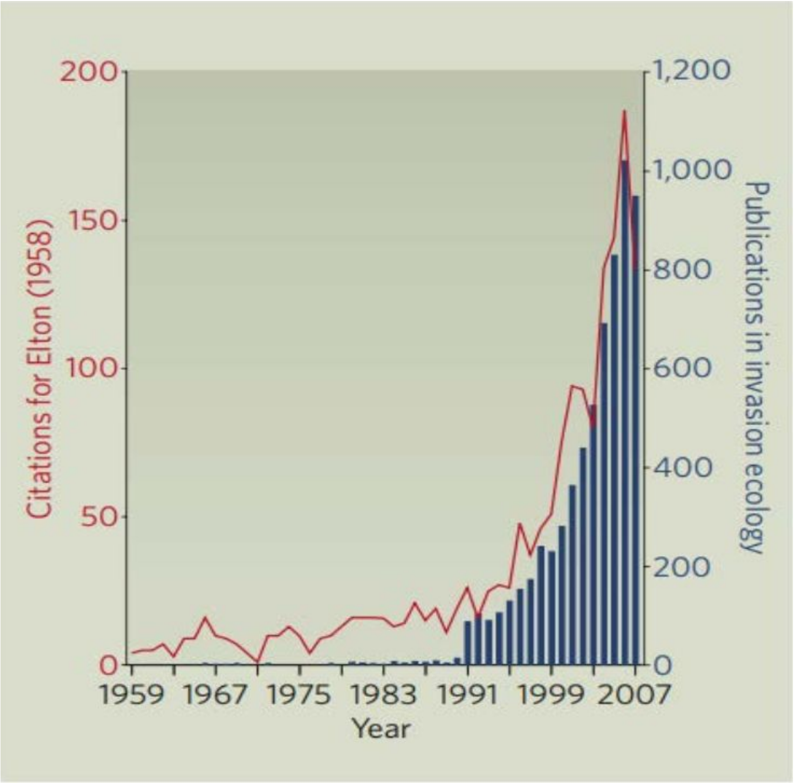
According to the Living Planet Index, **freshwater species** show the **greatest decline**, falling 81% between 1970 and 2012

The main threats to freshwater species are habitat loss and degradation, **non native invasive species** and overexploitation

Nonindigenous invasive species (NIS)

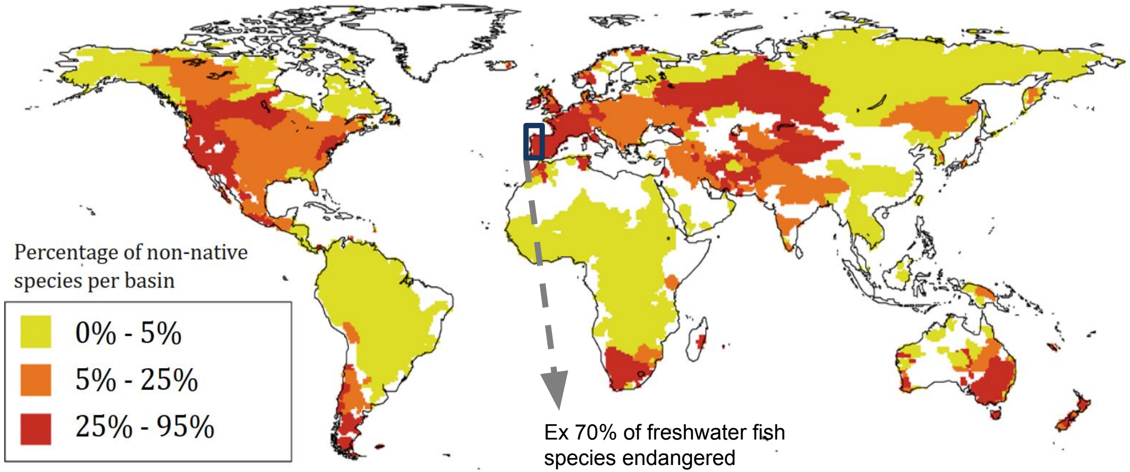
With the publication of '*The Ecology of Invasions*' by Animals and Plants', by Charles S. Elton,1958, demonstrated the profound influence of human activities in reshaping species distributions, with an early warning about the impacts of the transport of organisms

Since then, **biological invasions** have rapidly **increased** over the past century, with NIS constituting a major threat to biodiversity and ecosystem services

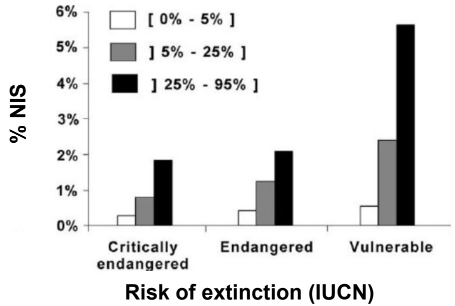


Ricciardi & MacIsaac (2008) The book that began invasion ecology. *Nature*

Non native invasive species (NIS)



Ex 70% of freshwater fish species endangered (Portugal)



The river basins that currently harbor more nonnatives have the higher levels of extinction risk for native fish species

Worldwide Distribution of Non-Native Freshwater Fish (Leprieur et al. 2008)

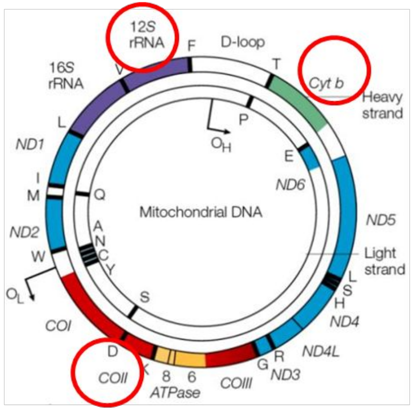
Environmental DNA (eDNA)

DNA traces from water, faeces, bulk, sediment samples to detect species (or OTUS) present in the surveyed ecosystem

Fish DNA reference database

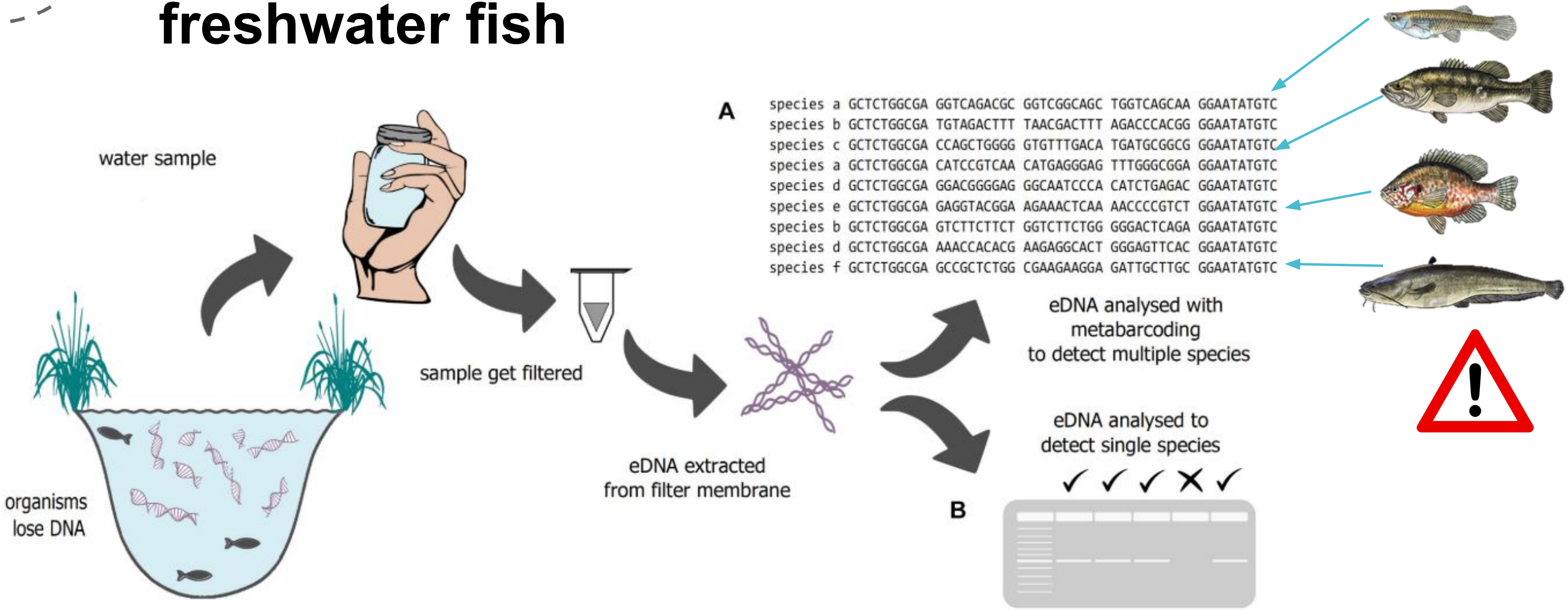
1. Tissue samples (fishes fin clips) from geographically distant sites
2. DNA extraction, amplification and sequencing (Sanger)
3. Complement with sequences of European fish species (BOLD and GenBank)

Markers used for species identification:
COI *cytochrome c oxidase subunit I*
CYTB *cytochrome B*



Marker used for metabarcoding:
12S – MiFish
~170bp
(Miya et al. 2015)

eDNA Metabarcoding as a tool to detect invasive freshwater fish



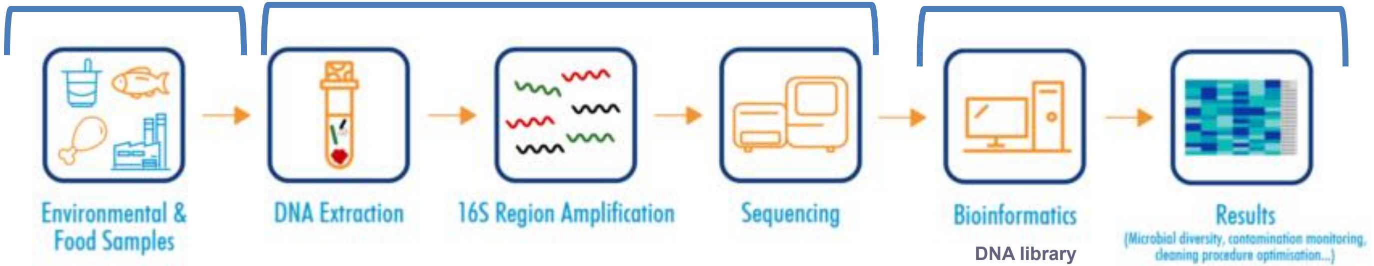
```
species a GCTCTGGCGA GGTCAGACGC GGTCGGCAGC TGGTCAGCAA GGAATATGTC
species b GCTCTGGCGA TGTAGACTTT TAACGACTTT AGACCCACGG GGAATATGTC
species c GCTCTGGCGA CCAGCTGGGG GTGTTTGACA TGATGCGGCG GGAATATGTC
species a GCTCTGGCGA CATCCGTCAA CATGAGGGAG TTTGGGCGGA GGAATATGTC
species d GCTCTGGCGA GGACGGGGAG GGCAATCCCA CATCTGAGAC GGAATATGTC
species e GCTCTGGCGA GAGGTACGGA AGAAACTCAA AACCCCGTCT GGAATATGTC
species b GCTCTGGCGA GTCTTCTTCT GGTCTTCTGG GGGACTCAGA GGAATATGTC
species d GCTCTGGCGA AAACCACACG AAGAGGCACT GGGAGTTCAC GGAATATGTC
species f GCTCTGGCGA GCCGCTCTGG CGAAGAAGGA GATTGCTTGC GGAATATGTC
```

FIGURE 1 | Schematic illustration of eDNA sampling. DNA fragments are filtered out of water and extracted. The fragments can be analyzed either to investigate **(A)** community structure or **(B)** the presence or absence of certain species.

Schadewell and Adams 2021

Metabarcoding steps

Field Laboratory Office



DNA library (BOLD+GeneBank)

Taxonomic assignment

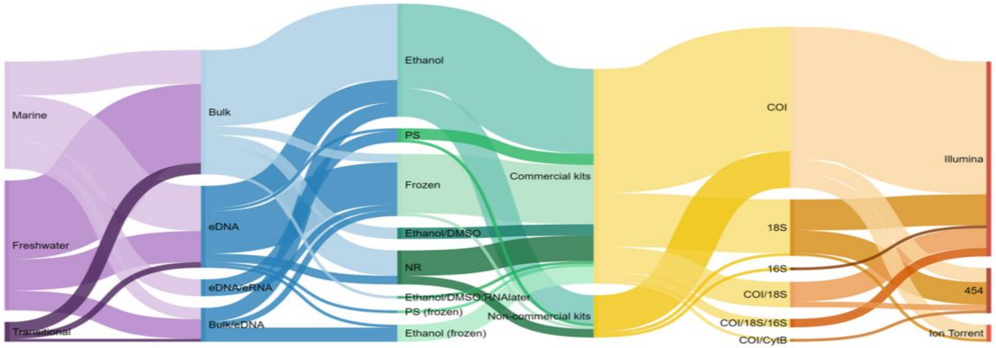


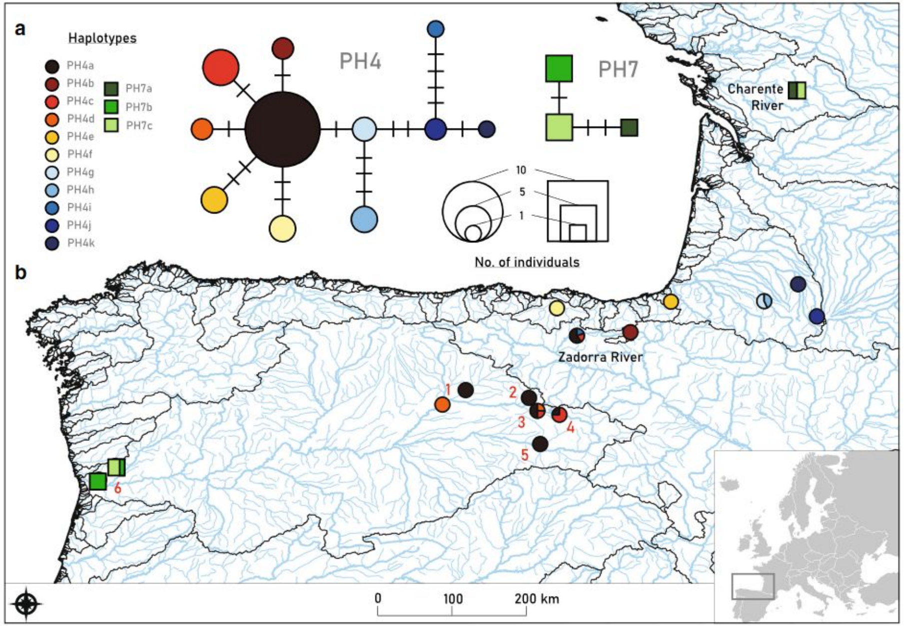
Figure 8. Diagram resuming the methods employed at each stage of the analytical chain of the metabarcoding workflow from the 90 publications analyzed in the current study. The thickness of the lines corresponds to the number of publications using a method. NR represents studies where the corresponding stage/method was not reported and PS represents preservation solution.

Duarte et al 2021. *Water*

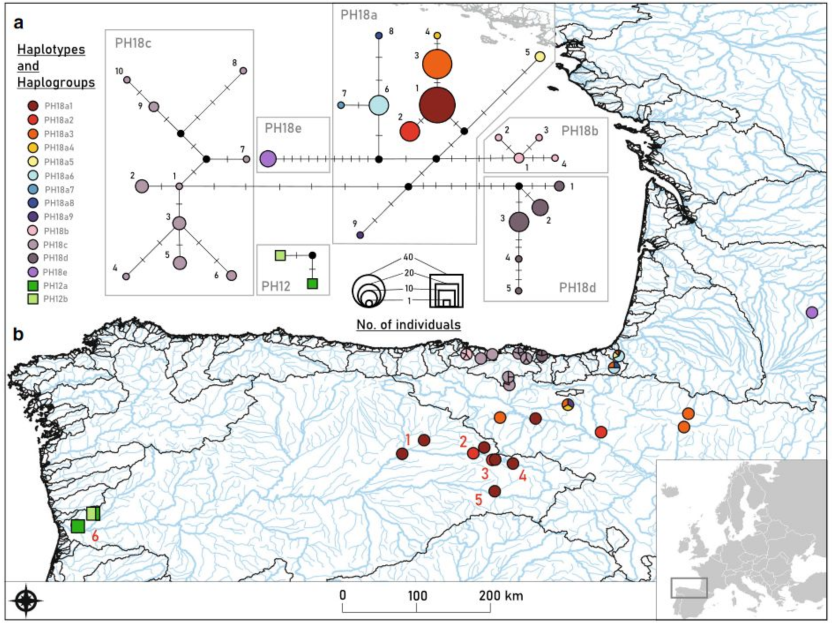
AS-VRE Lifewatch
Alien and Invasive **S**pecies
Virtual **R**esearch **E**nvironment

Data processing and management

Example: new *Phoxinus* in Portugal



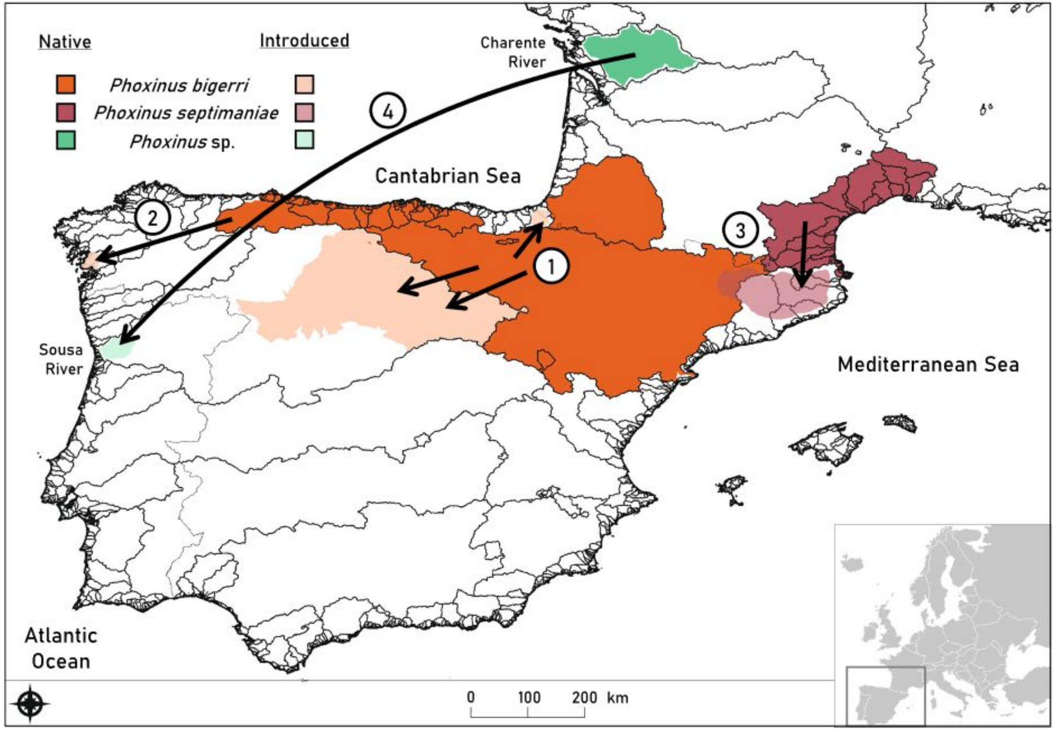
COI marker



Cyt b marker

Garcia-Raventós et al. 2020

Example: new *Phoxinus* in Portugal



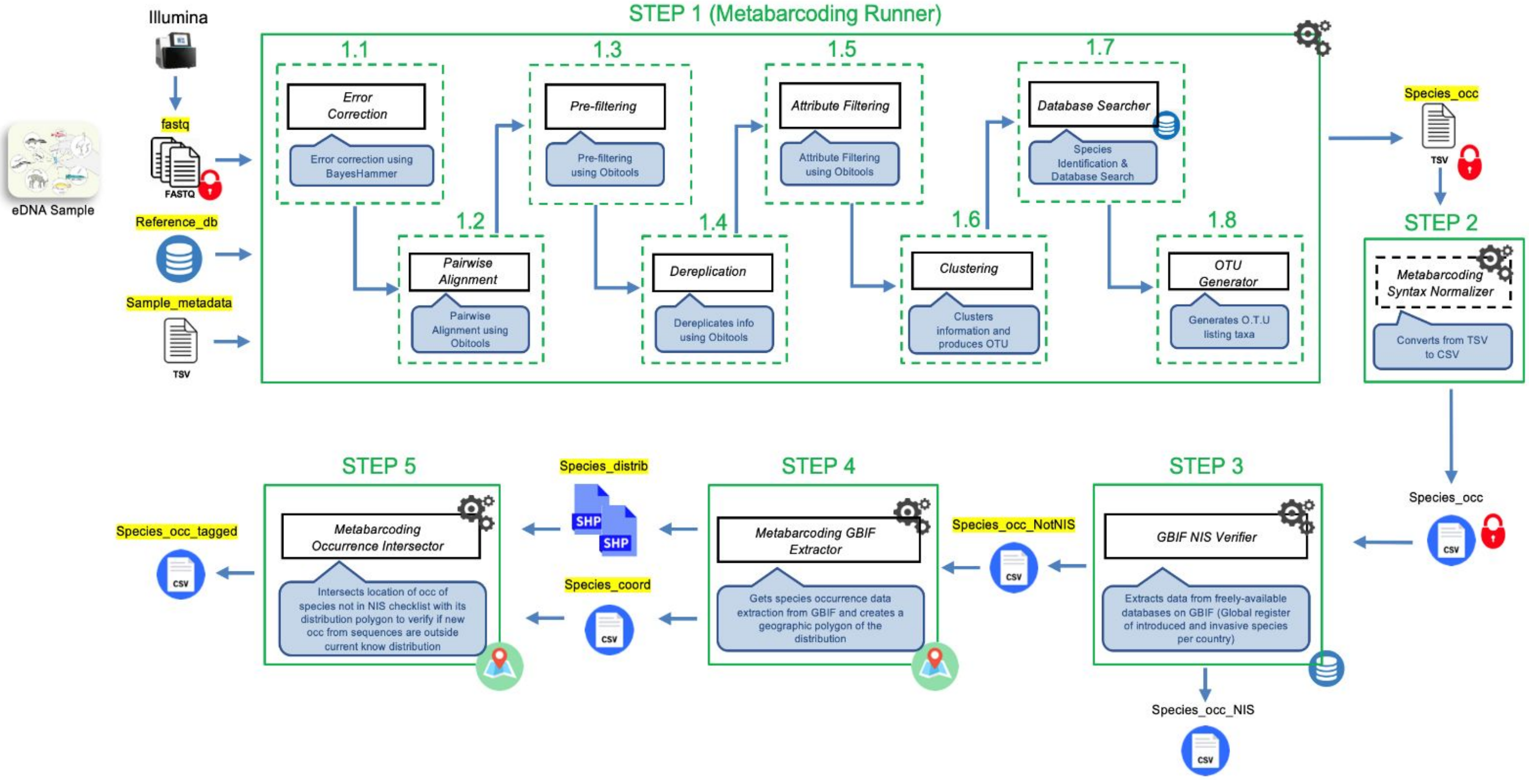
Garcia-Raventós et al. 2020

Aim of Validation Case #5

TO DETECT NEW NIS WITH eDNA METABARCODING

- ✔ Run pipeline for taxonomic assignment of a metabarcoding run, and obtain the list of species in the sample
- ✔ Produce the lists of NIS and of remnant species in the sample (native or new NIS for the region) with the GBIF NIS lists for the country
- ✔ Build the mapping (shapefiles) for all species detected, and a .csv with geographic coordinates
- ✔ Check if remnant species are outside the current known distributions

Validation Case #5: Metabarcoding workflow desk



Illumina

eDNA Sample

fastq

FASTQ

Reference_db

Sample_metadata

TSV

STEP 1 (Metabarcoding Runner)

1.1 *Error Correction*: Error correction using BayesHammer

1.2 *Pairwise Alignment*: Pairwise Alignment using Obitools

1.3 *Pre-filtering*: Pre-filtering using Obitools

1.4 *Dereplication*: Dereplicates info using Obitools

1.5 *Attribute Filtering*: Attribute Filtering using Obitools

1.6 *Clustering*: Clusters information and produces OTU

1.7 *Database Searcher*: Species Identification & Database Search

1.8 *OTU Generator*: Generates O.T.U listing taxa

Species_occ

TSV

STEP 2

Metabarcoding Syntax Normalizer: Converts from TSV to CSV

Species_occ

CSV

STEP 3

GBIF NIS Verifier: Extracts data from freely-available databases on GBIF (Global register of introduced and invasive species per country)

Species_occ_NIS

CSV

Species_occ_NotNIS

CSV

STEP 4

Metabarcoding GBIF Extractor: Gets species occurrence data extraction from GBIF and creates a geographic polygon of the distribution

Species_distrib

SHP

SHP

Species_coord

CSV

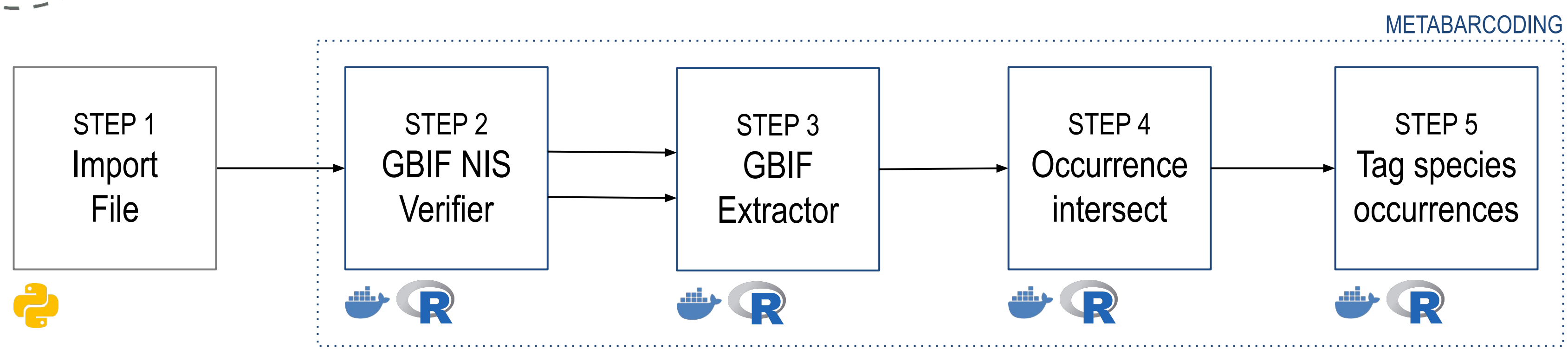
STEP 5

Metabarcoding Occurrence Intersector: Intersects location of occ of species not in NIS checklist with its distribution polygon to verify if new occ from sequences are outside current know distribution

Species_occ_tagged

CSV

Metabarcoding workflow



METABARCODING

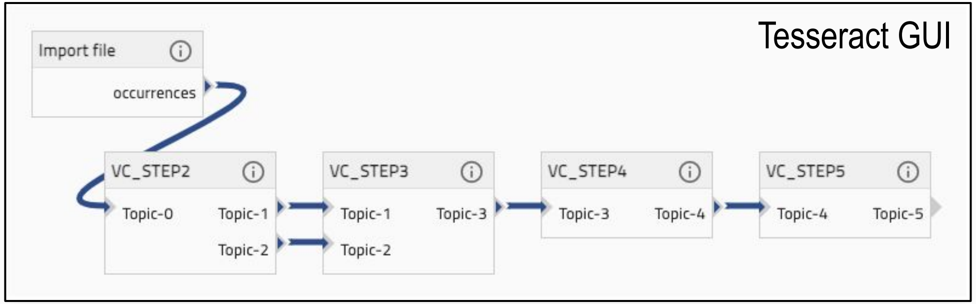
STEP 1
Import
File

STEP 2
GBIF NIS
Verifier

STEP 3
GBIF
Extractor

STEP 4
Occurrence
intersect

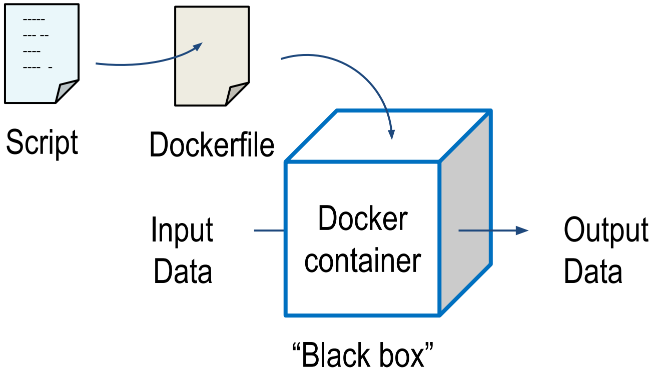
STEP 5
Tag species
occurrences



Tesseract GUI

Service level isolation

- Docker provides **service isolation** on the same machine.
- New services can be easily integrated into the platform with **minor efforts**.



Script

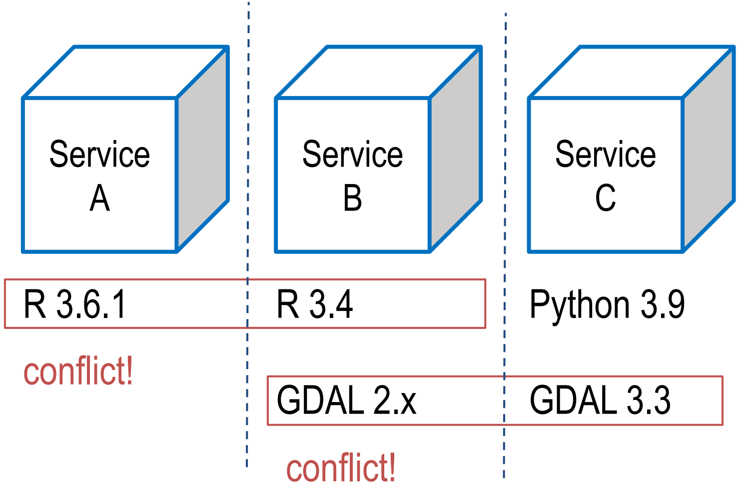
Dockerfile

Input Data

Docker container

Output Data

“Black box”



Service A

Service B

Service C

R 3.6.1

R 3.4

Python 3.9

conflict!

GDAL 2.x

GDAL 3.3

conflict!

Conclusions

- ✔ eDNA Metabarcoding allows to sample whole fish communities, and to identify NIS species
- ✔ VRE will help the difficult task of bioinformatic treatments of samples

With this VRE it will be possible to scale up the use of eDNA metabarcoding for monitoring large regions, and respond to EU directives and other needs across the world

Combining multiple methods (eDNA metabarcoding, electrofishing, fishing nets) will give a more complete view of freshwater fish communities

Challenges

- ✔ Automatic link to online DNA reference libraries (BOLD, GeneBank)
- ✔ Flexible enough to incorporate a user build-up DNA reference library (e.g. bash script “BOLD_NCBI_Merger” of Macher et al 2017)
- ✔ Including several pipeline options for filtering sequencing data (Number of minimum reads, …)
- ✔ Include a certainty index of NIS detection in the sample (e.g. specimen images, field observations,…)
- ✔ Combining multiple methods (eDNA metabercoding, electrofishing, fishing nets) will give a more complete view of freshwater fish communities

Acknowledgements and funding

FRESHING: Next-generation biomonitoring: freshwater bioassessment and species conservation improved with metagenomics
(PTDC/AAGMAA/2261/2014 – POCI-01-0145-FEDER-356016824)

EnvMetaGen: Capacity building at InBio for research and innovation using environmental metagenomics
(EUH2020-WIDESPREAD-2014-2)

FRESHCO: Multiple implications of invasive species on FRESHwater Mussel COextinction processes
(PTDC/AGR-FOR/1627/2014 – 04/SAICT/2015)

FCT
Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

COMPETE 2020

PORTUGAL 2020

UNIÃO EUROPEIA
Fundo Europeu de Desenvolvimento Regional

Thank you!

Ana Filipa Filipe affilipe@gmail.com CIBIO-inBIO and ISA, Lisboa, Portugal

José F. Aldana Montes jfam@lcc.uma.es Khaos Research, UMA, Málaga, Spain