Universidad de Oviedo
*Universidá d'Uviéu*
*University of Oviedo*

Departamento de Biología Funcional

Programa de Doctorado en Ingeniería Química, Ambiental y Bioalimentaria

# Assessing DNA metabarcoding and environmental DNA for biomonitoring fluvial ecosystems using macroinvertebrates in peninsular Spain

## Evaluación del uso de metabarcoding y ADN ambiental para la biomonitorización de ecosistemas fluviales utilizando macroinvertebrados en la España peninsular

**TESIS DOCTORAL**

Álvaro Fueyo Rodríguez

**Directores:**

Dr. Yaisel J. Borrell Pichs y Dr. Javier Granero Castro

Oviedo, 2024

Universidad de Oviedo

# RESUMEN DEL CONTENIDO DE TESIS DOCTORAL

| 1.- Título de la Tesis | |
|---|---|
| Español/Otro Idioma: **Evaluación del uso de Metabarcoding y ADN ambiental para la biomonitorización de ecosistemas fluviales utilizando macroinvertebrados en la España peninsular.** | Inglés: **Assessing DNA metabarcoding and eDNA for biomonitoring fluvial ecosystems using macroinvertebrates in peninsular Spain.** |

| 2.- Autor | |
|---|---|
| Nombre: **Álvaro Fueyo Rodríguez** | |
| Programa de Doctorado: en Ingeniería Química, Ambiental y Bioalimentaria | |
| **Órgano responsable:** Centro Internacional de Prostgrado, Universidad de Oviedo | |



## RESUMEN (en español)

Los ecosistemas fluviales desempeñan un papel vital en el mantenimiento de la biodiversidad, proporcionando agua dulce y sosteniendo a las sociedades humanas. A pesar de su importancia, los ríos del mundo están cada vez más degradados, lo que plantea desafíos tanto para la integridad ecológica como para el bienestar humano. Con el fin de proteger estos ecosistemas de los impactos antropogénicos, los Estados Miembros de la Unión Europea han adoptado la Directiva Marco del Agua (DMA - 2000/60/EC), que requiere que los Estados Miembros implementen un sistema de monitoreo diseñado para proporcionar datos precisos y comparables sobre el estado de los cuerpos de agua en toda Europa. Este sistema incluye el biomonitoreo de indicadores biológicos como los macroinvertebrados para inferir el estado ecológico de los ríos.

Los métodos tradicionales de biomonitoreo, que se basan en la identificación morfológica de macroinvertebrados, son laboriosos, consumen mucho tiempo y requieren una considerable experiencia taxonómica, mostrando limitaciones en su uso rutinario. Los métodos basados en ADN ofrecen una alternativa prometedora. Esta tesis investiga la efectividad, exactitud y aplicabilidad de implementar el metabarcoding de muestras bulk y de ADN ambiental (eDNA) para el biomonitoreo de comunidades de macroinvertebrados en los ríos de la península ibérica.

Se estudió la exhaustividad de las bases de datos genéticas de referencia y se secuenciaron cientos de especímenes. Se analizó el efecto de este esfuerzo de secuenciación local en los resultados de metarcoding y la inferencia del estado ecológico del río. Los resultados mostraron que el 21% de las morfoespecies secuenciadas carecían de secuencias de referencia en las bases de datos BOLD o GenBank. El enriquecimiento de bases de datos con nuevas secuencias condujo a una mayor detección de taxones, causando cambios en la inferencia molecular del estado ecológico.

Se realizaron análisis comparativos para evaluar la congruencia entre los métodos basados en ADN y la identificación morfológica, destacando discrepancias y posibles causas. Los estudios de campo involucraron la recolección de muestras para identificación morfológica y muestras bulk y de eDNA para análisis de metabarcoding en varios ríos del noroeste de España. Los resultados mostraron que la metacodificación de ADN de muestras masivas y de eDNA proporcionan perspectivas complementarias sobre la diversidad de macroinvertebrados, con cada método mostrando fortalezas y limitaciones únicas. El metabarcoding de muestras bulk generalmente mostró una mayor resolución taxonómica y mayor congruencia con los datos morfológicos en comparación con el eDNA. En general, se pudieron detectar especies protegidas y exóticas, proporcionando los primeros registros de la presencia de algunas especies en la región.

Finalmente, se investigaron las causas de las diferencias existentes entre los métodos moleculares y morfológicos, identificando aquellas que representan falsos positivos o negativos en la detección, y se elaboró una hoja de ruta para corregirlas y desarrollar una técnica fiable para su aplicación en la España peninsular. Los hitos necesarios son la finalización de una lista de especies de macroinvertebrados de agua dulce en la península, la generación de secuencias de referencia para todas ellas, el desarrollo y prueba de primers que amplifiquen

correctamente todas las especies de la península, y el diseño de un nuevo sistema de muestreo que maximice la recuperación de la diversidad específica de macroinvertebrados.
En general, esta tesis muestra el potencial de integrar técnicas basadas en ADN en los programas de biomonitoreo para proporcionar un enfoque más escalable y detallado de evaluación ambiental. Por otro lado, muestra que el metabarcoding de muestras bulk y eDNA todavía requiere una serie de mejoras para aumentar la fiabilidad y precisión antes de que pueda presentarse como una alternativa viable para el biomonitoreo de ríos.

**RESUMEN (en Inglés)**

Fluvial ecosystems play a vital role in supporting biodiversity, providing freshwater, and sustaining human societies. Despite their importance, the world's rivers are increasingly degraded, posing challenges to both ecological integrity and human well-being. In order to protect environments from anthropogenic impacts, European Union Member States have adopted the Water Framework Directive (WFD - 2000/60/EC) which requires Member States to implement a comprehensive monitoring and assessment system designed to provide accurate and comparable data on the status of water bodies across Europe. This system includes the biomonitoring of biological indicators such as macroinvertebrates for infer the ecological status of rivers.
Traditional biomonitoring methods, which rely on morphological identification of macroinvertebrates, are labour-intensive, time-consuming, and require significant taxonomic expertise, showing limitations in routine use. DNA-based methods offer a promising alternative, potentially providing a scalable and more detailed approach. This thesis investigates the readiness, accuracy, and applicability of implementing DNA metabarcoding of bulk and environmental DNA (eDNA) samples for biomonitoring macroinvertebrate communities in the rivers of peninsular Spain.
The completeness of the reference genetic databases was studied, and hundreds of specimens were sequenced. The effect of this local sequencing effort on metabarcoding results and inference of the ecological status of the river was analysed. The results showed that 21% of sequenced morphospecies lacked reference sequences in the BOLD or GenBank databases. Enriching databases with new sequences led to more taxa being detected, causing changes in the molecular ecological status inference.
Comparative analyses were carried out to assess the congruence between DNA-based methods and morphological identification, highlighting discrepancies and possible causes. Field studies involved the collection of samples for morphological identification and bulk and eDNA samples for metabarcoding analysis from various rivers in northwest Spain. The results showed that DNA metabarcoding of bulk samples and eDNA samples provided complementary insights into macroinvertebrate diversity, with each method exhibiting unique strengths and limitations. Metabarcoding of bulk samples generally showed higher taxonomic resolution and more congruence with morphological data compared to eDNA. In general, they were able to detect protected and exotic species, providing the first records of the presence of some species in the region.
Finally, the causes of the existing differences between molecular and morphological methods were investigated, identifying those that represent false positives or negatives in detection, and a roadmap was drawn up to correct them and develop a reliable technique for its application in peninsular Spain. The necessary milestones are the completion of a checklist of freshwater macroinvertebrate species in the peninsula, the generation of genetic barcodes for all of them, the development and testing of primers that correctly amplify all species from the peninsula, and the design of a new sampling system that maximises the recovery of specific macroinvertebrate diversity.
Overall, this thesis shows the potential of integrating DNA-based techniques into biomonitoring frameworks to provide a more scalable and detailed approach to environmental assessment. On the other hand, it shows that bulk sample and eDNA metabarcoding still require a number of improvements to increase reliability and accuracy before they can be presented as a viable alternative for river biomonitoring.

**SR. PRESIDENTE DE LA COMISIÓN ACADÉMICA DEL PROGRAMA DE DOCTORADO EN INGENIERÍA QUÍMICA, AMBIENTAL, Y BIOALIMENTARIA**

Universidad de Oviedo
*Universidá d'Uviéu*
*University of Oviedo*

Departamento de Biología Funcional

Programa de Doctorado en Ingeniería Química, Ambiental y Bioalimentaria

# Assessing DNA metabarcoding and environmental DNA for biomonitoring fluvial ecosystems using macroinvertebrates in peninsular Spain

## Evaluación del uso de metabarcoding y ADN ambiental para la biomonitorización de ecosistemas fluviales utilizando macroinvertebrados en la España peninsular

**TESIS DOCTORAL**

Álvaro Fueyo Rodríguez

**Directores:**

Dr. Yaisel J. Borrell Pichs y Dr. Javier Granero Castro

Oviedo, 2024

## Abstract

Fluvial ecosystems play a vital role in supporting biodiversity, providing freshwater, and sustaining human societies. Despite their importance, the world's rivers are increasingly degraded, posing challenges to both ecological integrity and human well-being. In order to protect environments from anthropogenic impacts, European Union Member States have adopted the Water Framework Directive (WFD - 2000/60/EC) which requires Member States to implement a comprehensive monitoring and assessment system designed to provide accurate and comparable data on the status of water bodies across Europe. This system includes the biomonitoring of biological indicators such as macroinvertebrates for infer the ecological status of rivers.

Traditional biomonitoring methods, which rely on morphological identification of macroinvertebrates, are labour-intensive, time-consuming, and require significant taxonomic expertise, showing limitations in routine use. DNA-based methods offer a promising alternative, potentially providing a scalable and more detailed approach. This thesis investigates the readiness, accuracy, and applicability of implementing DNA metabarcoding of bulk and environmental DNA (eDNA) samples for biomonitoring macroinvertebrate communities in the rivers of peninsular Spain.

The completeness of the reference genetic databases was studied, and hundreds of specimens were sequenced. The effect of this local sequencing effort on metabarcoding results and inference of the ecological status of the river was analysed. The results showed that 21% of sequenced morphospecies lacked reference sequences in the BOLD or GenBank databases. Enriching databases with new sequences led to more taxa being detected, causing changes in the molecular ecological status inference.

Comparative analyses were carried out to assess the congruence between DNA-based methods and morphological identification, highlighting discrepancies and possible causes. Field studies involved the collection of samples for morphological identification and bulk and eDNA samples for metabarcoding analysis from various rivers in northwest Spain. The results showed that DNA metabarcoding of bulk samples and eDNA samples provided complementary insights into macroinvertebrate diversity, with each method exhibiting unique strengths and limitations. Metabarcoding of bulk samples generally showed higher taxonomic resolution and more congruence with morphological data compared to eDNA. In general, they were able to detect protected and exotic species, providing the first records of the presence of some species in the region.

Finally, the causes of the existing differences between molecular and morphological methods were investigated, identifying those that represent false positives or negatives in detection, and a roadmap was drawn up to correct them and develop a reliable technique for its application in peninsular Spain. The necessary milestones are the completion of a

checklist of freshwater macroinvertebrate species in the peninsula, the generation of genetic barcodes for all of them, the development and testing of primers that correctly amplify all species from the peninsula, and the design of a new sampling system that maximises the recovery of specific macroinvertebrate diversity.

Overall, this thesis shows the potential of integrating DNA-based techniques into biomonitoring frameworks to provide a more scalable and detailed approach to environmental assessment. On the other hand, it shows that bulk sample and eDNA metabarcoding still require a number of improvements to increase reliability and accuracy before they can be presented as a viable alternative for river biomonitoring.

**Keywords**: Iberian Peninsula, Bulk Sample, IBMWP, Ecological Status, River, Stream.

## Resumen

Los ecosistemas fluviales desempeñan un papel vital en el mantenimiento de la biodiversidad, proporcionando agua dulce y sosteniendo a las sociedades humanas. A pesar de su importancia, los ríos del mundo están cada vez más degradados, lo que plantea desafíos tanto para la integridad ecológica como para el bienestar humano. Con el fin de proteger estos ecosistemas de los impactos antropogénicos, los Estados Miembros de la Unión Europea han adoptado la Directiva Marco del Agua (DMA - 2000/60/EC), que requiere que los Estados Miembros implementen un sistema integral de monitoreo y evaluación diseñado para proporcionar datos precisos y comparables sobre el estado de los cuerpos de agua en toda Europa. Este sistema incluye el biomonitoreo de indicadores biológicos como los macroinvertebrados para inferir el estado ecológico de los ríos.

Los métodos tradicionales de biomonitoreo, que se basan en la identificación morfológica de macroinvertebrados, son laboriosos, consumen mucho tiempo y requieren una considerable experiencia taxonómica, mostrando limitaciones en su uso rutinario. Los métodos basados en ADN ofrecen una alternativa prometedora, proporcionando potencialmente datos de biodiversidad más detallados y precisos. Esta tesis investiga la efectividad, exactitud y aplicabilidad de implementar el metabarcoding de muestras bulk y de ADN ambiental (eDNA) para el biomonitoreo de comunidades de macroinvertebrados en los ríos de la península ibérica.

Se estudió la exhaustividad de las bases de datos genéticas de referencia y se secuenciaron cientos de especímenes. Se analizó el efecto de este esfuerzo de secuenciación local en los resultados de metabarcoding y la inferencia del estado ecológico del río. Los resultados mostraron que el 21% de las morfoespecies secuenciadas carecían de secuencias de referencia en las bases de datos BOLD o GenBank. El enriquecimiento de bases de datos con nuevas secuencias condujo a una mayor detección de taxones, causando cambios en la inferencia molecular del estado ecológico.

Se realizaron análisis comparativos para evaluar la congruencia entre los métodos basados en ADN y la identificación morfológica, destacando discrepancias y posibles causas. Los estudios de campo involucraron la recolección de muestras para identificación morfológica y muestras bulk y de eDNA para análisis de metabarcoding en varios ríos del noroeste de España. Los resultados mostraron que el metabarcoding de ADN de muestras bulk y de eDNA proporcionan perspectivas complementarias sobre la diversidad de macroinvertebrados, con cada método mostrando fortalezas y limitaciones únicas. El metabarcoding de muestras bulk generalmente mostró una mayor resolución taxonómica y mayor congruencia con los datos morfológicos en comparación con el eDNA. En general, se pudieron detectar especies protegidas y exóticas, proporcionando los primeros registros de la presencia de algunas especies en la región.

Finalmente, se investigaron las causas de las diferencias existentes entre los métodos moleculares y morfológicos, identificando aquellas que representan falsos positivos o negativos en la detección, y se elaboró una hoja de ruta para corregirlas y desarrollar una técnica fiable para su aplicación en la España peninsular. Los hitos necesarios son la finalización de una lista de especies de macroinvertebrados de agua dulce en la península, la generación de secuencias de referencia para todas ellas, el desarrollo y prueba de primers que amplifiquen correctamente todas las especies de la península, y el diseño de un nuevo sistema de muestreo que maximice la recuperación de la diversidad específica de macroinvertebrados.

En general, esta tesis muestra el potencial de integrar técnicas basadas en ADN en los programas de biomonitoreo para proporcionar un enfoque más escalable y detallado de evaluación ambiental. Por otro lado, muestra que el metabarcoding de muestras bulk y eDNA todavía requiere una serie de mejoras para aumentar la fiabilidad y precisión antes de que pueda presentarse como una alternativa viable para el biomonitoreo de ríos.

**Palabras clave:** Península Ibérica, Muestra Bulk, IBMWP, Estado ecológico, Río, Arroyo.

# Contents

# I. Context of this work

This thesis has been developed in the context of an industrial PhD at the company TAXUS Medio Ambiente S.L. (hereafter 'Taxus') in collaboration with the ARENA (Natural Resources Research Classroom) research group of the University of Oviedo. The ARENA research group is an accredited research group of the University of Oviedo, with more than 15 years of work in national and international projects in different applications of DNA-based tools. Taxus is a company based in Asturias, founded in 2003, which, among many other services, is dedicated to assessing the status of rivers. Taxus is accredited by the National Accreditation Body (ENAC), according to UNE-EN ISO/IEC 17025 for the sampling and processing of benthic macroinvertebrates and the calculation of the IBMWP, METI and IASPT indices (Accreditation no. 1280/LE2425). Taxus has been linked to R&D since its foundation, carrying out numerous scientific projects and producing publications. Prior to the start of this thesis (2020), Taxus did not have or offer any DNA-based molecular services.

DNA-based molecular techniques are of significant interest to environmental companies, governments, and environmental policymakers due to their potential ability to provide precise, rapid, and cost-effective solutions for monitoring and managing biodiversity and ecosystem health. Consequently, environmental companies and public administrations are keen to incorporate DNA-based techniques into their standard practices. However, offering these new molecular techniques as a commercial service or incorporating them into legislation requires a high level of development and validation of each technique for each application to ensure reliable, accurate, and reproducible results.

At the beginning of this thesis, environmental DNA (eDNA) and metabarcoding techniques for river biomonitoring, using macroinvertebrates as bioindicators in the Spanish part of the Iberian Peninsula, were at a Technology Readiness Level (TRL) of between 4 and 5 (European Commission, 2014). These levels of development coincide with what is known as "valley of death", an analogy used to describe a discontinuity in innovation processes, as it lies between the levels of development assumed by academia and the levels of profitable development for industry (Markham, 2002). This work falls into this gap and has therefore been co-funded by TAXUS MEDIO AMBIENTE S.L., SEKUENS (formerly IDEPA) with the R&D Projects

2021 programme (IDE/2021/000527), and by the Spanish Ministry of Science and Innovation with an Industrial PhD Grant for Álvaro Fueyo (DIN2019- 010834).

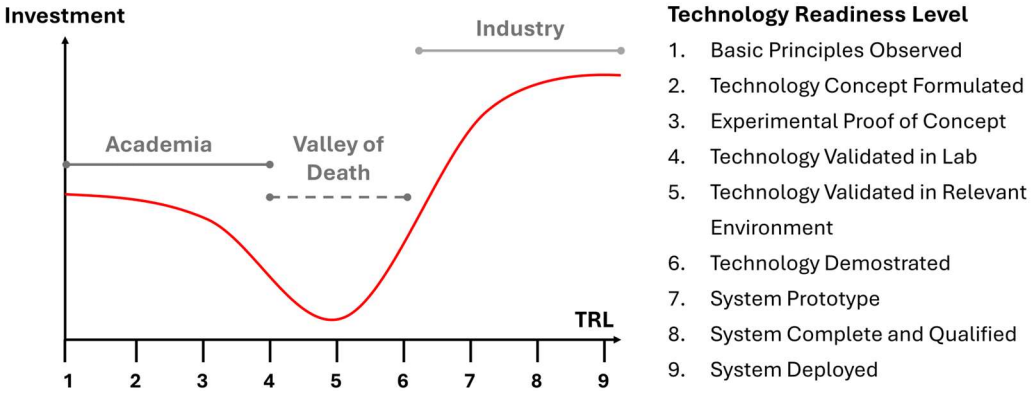



**Figure I-1** Investment vs. technology readiness level (TRL), showing the technological 'valley of death'. Adapted from: (Tübitak, 2020)

# II. Introduction

Fluvial ecosystems play a vital role in supporting biodiversity, providing fresh water, and sustaining human societies (Dudgeon et al., 2006; Hanna et al., 2018). More than 2 billion people depend directly on rivers for their water needs (WWF, 2024), and at least 12 million tons of freshwater fish are caught annually, providing food and livelihoods for hundreds of millions of people (FAO, 2021; Fluet-Chouinard et al., 2018). Fluvial ecosystems are home to a wealth of life, including 55% of all fish species and at least 10% of all known species (UN Environment Programme, 2022). Despite their importance, the world's rivers are increasingly degraded, posing challenges to both ecological integrity and human well-being (Millennium Ecosystem Assessment, 2005). This degradation is leading to an overall loss of freshwater biodiversity (Figure II-1) and has a profound impact on the social and ecosystem services provided (Basak et al., 2021), resulting in the protection and restoration of water-related ecosystems becoming a United Nations Sustainable Development Goal (SDG 6.6).

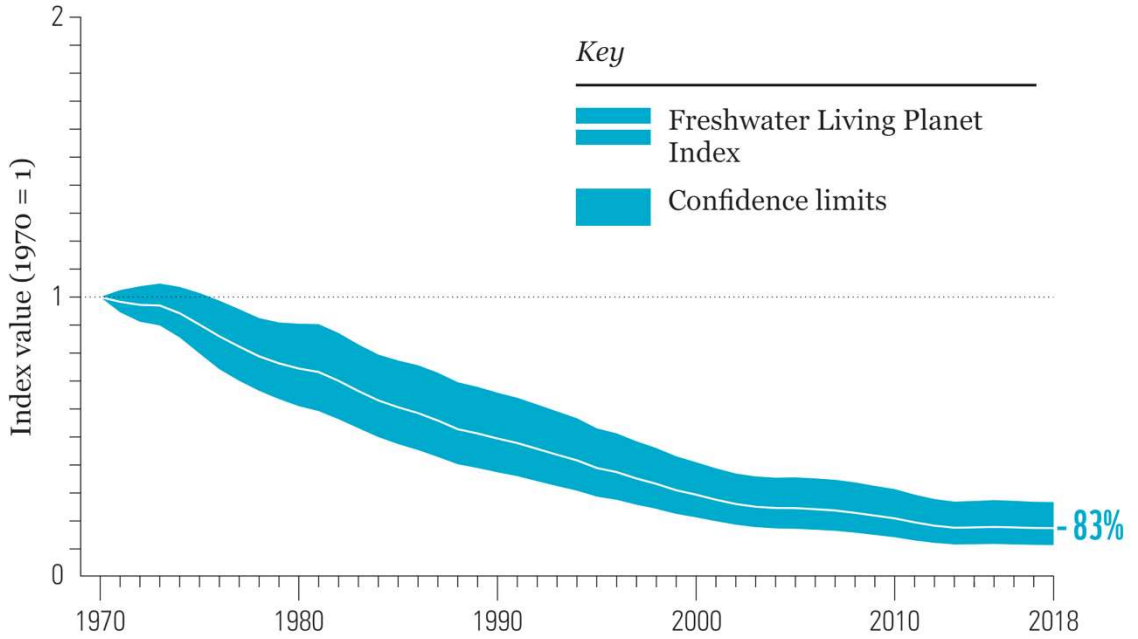



**Figure II-1** The freshwater living planet Index (1970-2018). The average abundance of 6,617 freshwater populations across the globe, representing 1,398 vertebrate species (mammals, birds, amphibians, reptiles and fish), declined by 83%. The white line shows the index's values, and the shaded areas represent the statistical certainty surrounding the trend (95% statistical certainty, range 74% to 89%). Source: (WWF-ZSL, 2022)

In Europe, according to the European Environment Agency (2021), the main threats to river ecosystems include pollution from urban and industrial wastewater, diffuse pollution from agriculture (22% of Europe's surface water bodies are significantly affected by both nutrients and pesticides), and pollution from mining and non-sewered housing. In addition, approximately 34% of surface water bodies are significantly affected by structural changes, such as river channel stabilisation, dam constructions, flood protection, or irrigation. Other minor pressures include aquaculture and invasive alien species. Water scarcity and

droughts, both permanent and seasonal, are an increasing problem in many areas of Europe. Roughly 6% of Europe's surface water bodies are significantly affected by water abstraction, mainly for agriculture, public water supply and industry.

## Water monitoring under the European legal framework.

In order to protect environments from anthropogenic impacts, European Union Member States have adopted a legal framework that includes the Council Directive 92/43/EEC on the conservation of natural habitats and of wild fauna and flora, the Directive 2024/1203 on the protection of the environment through criminal law, and the Directive 2004/35/EC on environmental liability with regards to the prevention and remedying of environmental damage. The latter is known as Environmental Liability Directive (ELD) and establishes a comprehensive EU-wide liability regime for environmental damage based on the 'polluter pays' principle. It entered into force in 2007 and makes those who have caused environmental damage liable for remediation. In its article 2(14), the Law defines 'Environmental damage' as "a **measurable** adverse change in a natural resource or the measurable impairment of a natural resource service that may occur directly or indirectly". The law also states that where environmental damage has occurred, the operator (e.g. industry) "shall take and bear the cost of the necessary remedial measures to restore the damaged natural resources and/or impaired services to, or towards, the baseline condition". 'Baseline condition' is defined in the law as the condition of the natural resources and services that would have existed if the environmental damage had not occurred, **estimated on the basis of the best information available**. The determination of the baseline condition of the natural environment, together with the assessment of the damage caused and the remedial measures required to repair the damage, can make a huge difference to the total amount of money spent on remedial measures and whether there is a real and complete recovery of the damaged environment. There is, therefore, a need for an objective and accurate system to measure the state of ecosystems, allowing all dimensions of their condition and the magnitude of possible impacts to be measured. Moreover, ecosystem monitoring is not only useful when damage occurs. Objective measurements of biotic and abiotic conditions are essential for the implementation of effective management strategies and policies aimed at restoring and maintaining ecosystems. At the same time, this information facilitates the evaluation of the effectiveness of conservation efforts and policy implementation, providing a scientific basis for adjusting management practices and ensuring compliance with environmental legislation. Without accurate and consistent monitoring, it is difficult to track progress or identify areas for improvement.

Therefore, in the year 2000, the Member States of the European Union provided themselves with a legal framework for **monitoring, managing, and restoring all their water bodies** (inland, transitional, and coastal surface waters, as well as groundwaters). This is the

Directive 2000/60/EC of the European Parliament and of the Council establishing a framework for Community action in the field of water policy, also known as Water Framework Directive (hereafter WFD).

The WFD requires Member States to implement the necessary measures to monitor and prevent deterioration of the status of all water bodies and to protect, enhance and restore them. To achieve these goals, WFD obliges Member States to:

- identify the individual river basins lying within their national territory and assign them to individual River Basin Districts (RBDs) by 2003.
- differentiate the relevant surface water bodies by type within the river basin district. (In Spain, rivers have been classified into different river types according to biogeographical and abiotic characteristics (CEDEX, 2004, 2007)).
- establish type-specific reference conditions for surface water body types and collect and maintain information on the type and magnitude of the relevant anthropogenic pressures.
- establish water status monitoring programmes by 2006.
- produce and publish six-year River Basin Management Plans (RBMPs) for each RBD by 2009 (Figure II-2).
- implement the programmes of measures and achieve the environmental objectives (good status of all water bodies) by 2015.

If Member States do not reach good status for all water bodies of a river basin district by 2015, the Water Framework Directive offers the possibility to Member States to engage in two further six-year cycles of planning and implementation appropriate measures. Therefore, the directive allows for the achievement of good statuses of all water bodies in three water planning cycles, 2009-2015 (1$^{st}$ RBMP), 2016-2021 (2$^{nd}$ RBMP), 2022-2027 (3$^{rd}$ RBMP). The achievement of good statuses for surface water bodies was not achieved by 2015 (Figure II-3)(European Environmental Agency, 2018). Thus, a second hydrological planning cycle was implemented, and the third planning cycle is currently being developed (at the time of the publication of this work).

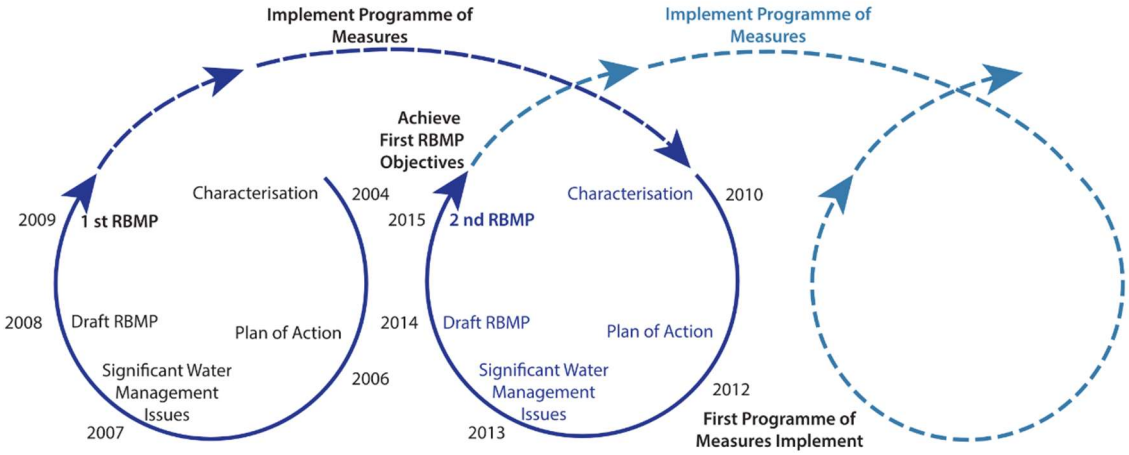
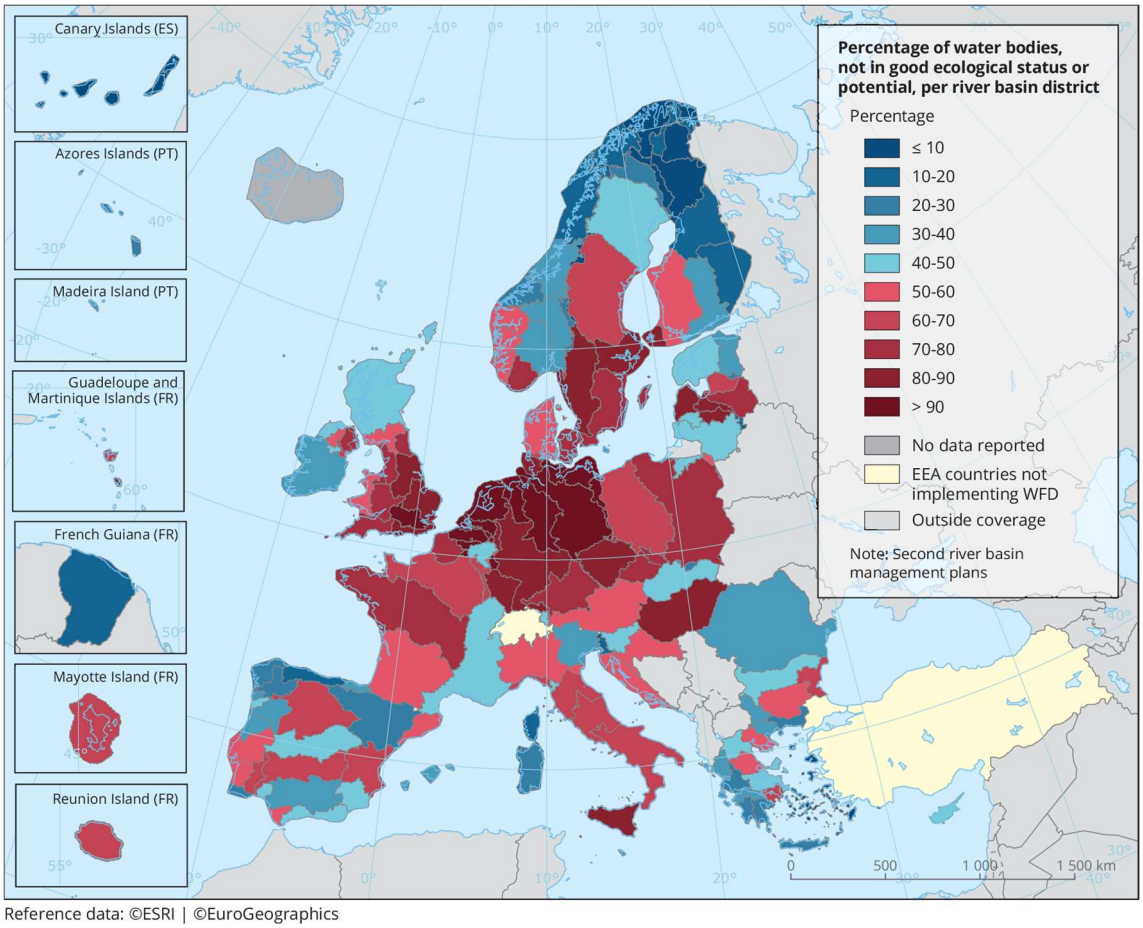



**Figure II-2** WFD river basin management planning process. Source: Kaspersen (2015).

In Spain, this European regulation has been incorporated into national legislation, mainly through the Real Decreto (hereafter RD) 1/2001, of 20 July, approving the revised text of the Water Law, RD 907/2007, of 6 July, approving the Regulation on Hydrological Planning and RD 817/2015, of 11 September, which updates and extends the previous legislation and establishes the criteria for monitoring and assessing the status of surface waters and the environmental quality standards.




Reference data: ©ESRI | ©EuroGeographics

**Figure II-3** Percentage of surface water bodies not in good ecological status or potential, per river basic district. The map is based on second RBMP. These plans were finalised in 2015 and reported between 2016 and 2018. Source: European environment Agency.

## River Status and Potential

The WFD also implemented a comprehensive monitoring and assessment system designed to provide accurate and comparable data on the status of water bodies across Europe. This system includes both biological and chemical indicators, ensuring a thorough evaluation of water quality and ecosystem health. The WFD defines the status of a water body as the degree of alteration to its natural conditions, and it is determined by **the worst** of its chemical and ecological status.

- The chemical status is an expression of surface water quality that reflects the degree of compliance with environmental quality standards for priority substances and other pollutants. Annex IV of RD 817/2015 establishes the annual average and maximum allowable concentration of pollutants for different matrices (water and biota). If any of these concentrations are exceeded, the water body is not considered to be in good status.
- The ecological status is an expression of the quality of the structure and functioning of aquatic ecosystems associated with surface waters in relation to reference conditions. In order to classify the ecological status of surface water bodies, the indices and indicators of the quality elements established in Annex II of RD 817/2015 are applied. Ecological status can be classified as Bad, Poor, Moderate, Good or Very Good. In the case of artificial water bodies or heavily modified water bodies, ecological status is renamed as 'ecological potential', and the category 'Very Good' is renamed 'Maximum'.

The quality elements established by RD 817/2015 to infer the ecological status of rivers are biological (aquatic flora, benthic macroinvertebrates, and fish), physicochemical (pH, dissolved oxygen, nutrients, etc.) and hydromorphological (hydrological regime, continuity, etc.). Once the quality elements have been measured in the water body and the values of the official indices have been calculated, an iterative process is followed to determine the final ecological status classification (Figure II-4).

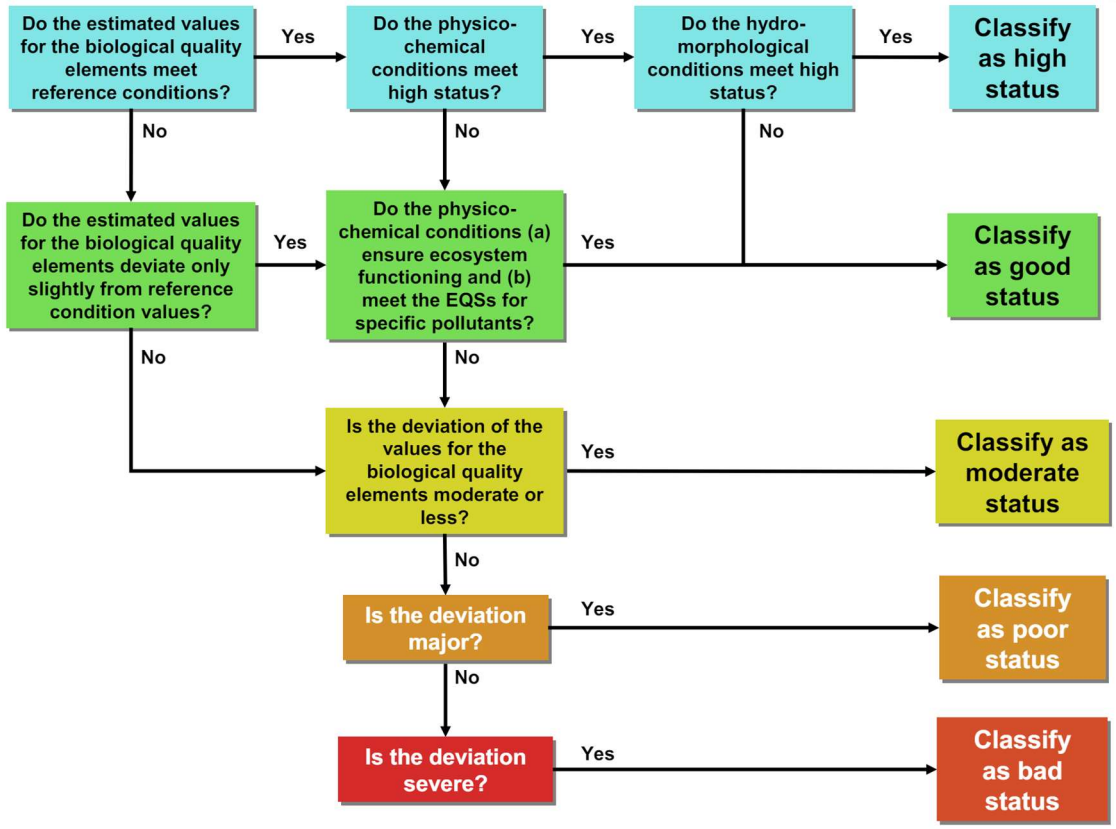



**Figure II-4** Decision-tree illustrating the criteria determining the different ecological status classes. Source: WFD- Annex V.

However, although several Biological Quality Elements (BQE) are mentioned in the legislation, the characterisation of Spanish rivers during the second planning cycle revealed that of the 3,531 points where BQEs were analysed, 90% measured only two BQEs or fewer. These were mainly benthic macroinvertebrates and phytobenthos. This can be partly attributed to the fact that internationally intercalibrated methods and indices are available almost exclusively for these two BQEs (MITECO, 2019). Despite over a decade of intensified research efforts following the adoption of the WFD to develop methods and indices for characterising the ecological status of European water bodies (e.g. Birk et al., 2012), not all of these have yet resulted in officially standardised and intercalibrated measurement methods and indices (Figure II-5).

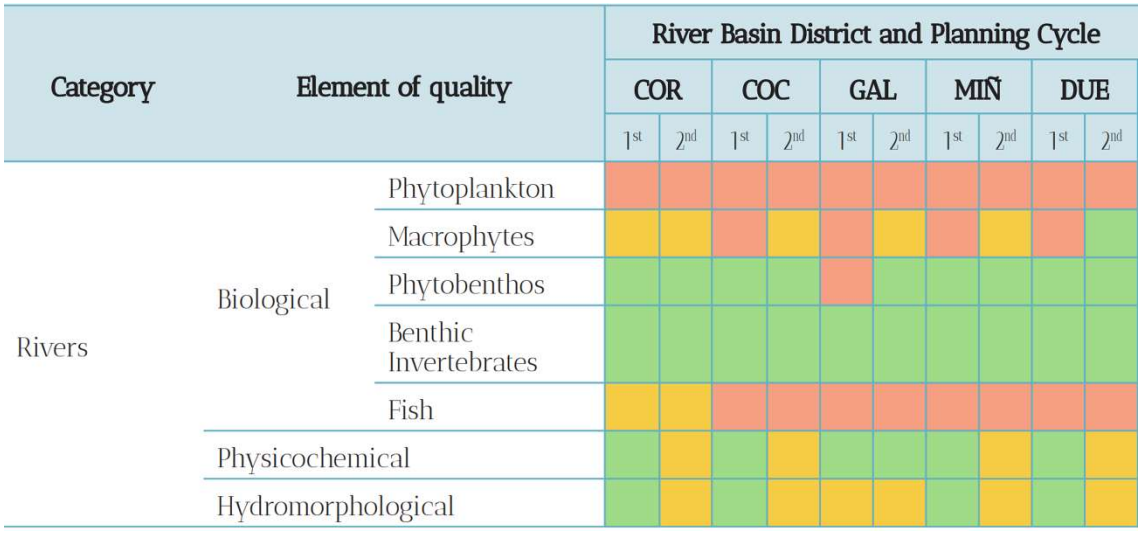



**Figure II-5** Availability of assessment methods on the ecological status of rivers of northwestern Spain. Green: Assessment methods developed. Yellow: Assessment methods partially developed or under developing. Red: Non-developed assessment methods. COR: Cantábrico Oriental, COC: Cantábrico Occidental, GAL: Cuencas interiores de Galicia, MIÑ: Miño-Sil, DUE: Duero. Source: (MITECO, 2019) (Partial).

## Macroinvertebrates as Bioindicators

River benthic macroinvertebrates (hereafter 'macroinvertebrates') are a complex group formed by species of 4 different phyla: arthropods, molluscs, annelids, and platyhelminthes (as considered by Spanish official agencies and legislation, MAGRAMA, 2013a, 2013b) which live permanently or at certain stages of their life cycle in aquatic environments and are at least 0.5 cm long. They are a faunal group of great diversity and ecological importance (Roldán Pérez, 2012; Stubbington et al., 2017; Wallace & Webster, 1996).

Macroinvertebrates have certain characteristics that make them very useful as bioindicators (Roldán Pérez, 2012; Tachet et al., 2010), including:

- they are practically ubiquitous in all continental aquatic systems, allowing comparative studies.
- they are mostly sedentary and therefore representative of local conditions.
- they respond quickly to environmental stressors.
- they provide information to integrate cumulative effects.
- they offer a wide range of responses to different impacts, as their communities are heterogeneous, with high taxonomic richness and different functional groups.

Macroinvertebrates have long been recognised as valuable bioindicators for assessing the ecological status of rivers. Their use in environmental monitoring can be traced back to the

early 20th century, with the pioneering work of Kolkwitz and Marsson (1909). Since then, several indices and sampling methodologies have been developed to assess the ecological status of rivers using benthic macroinvertebrate communities. One of the most widely used indices is the Biological Monitoring Working Party (BMWP) score system. The BMWP was developed in the UK during the 1970s and was created to provide a straightforward, yet effective, method for evaluating water quality in rivers and streams by assigning scores to different macroinvertebrate families based on their sensitivity to pollution (Hawkes, 1997). The higher the score of a given family, the more sensitive it is to pollutants. The cumulative score of a sampled site thus offers a measure of its ecological health, with higher scores indicating better water quality. This system allowed for more consistent and reliable monitoring across various regions. Recognising the effectiveness of the BMWP index, a great number of countries adopted and adapted it for monitoring their rivers (Cota et al., 2002; Mustow, 2002; Naranjo López et al., 2005; Roldan Pérez, 2003). In Spain, researchers and environmental agencies adapted it to the specific conditions of Iberian rivers, resulting in the Iberian Biological Monitoring Working Party (IBMWP) index (Alba-Tercedor et al., 2004). This adaptation involved recalibrating the sensitivity scores to reflect the unique ecological and hydrological characteristics of Spanish water bodies. Additionally, the IBMWP incorporated regional taxa that were not present in the original BMWP framework, ensuring a more accurate assessment of water quality in Spain.

**Table II-1** Count/percentage of rivers grouped by RBD, and its ecological status inferred using macroinvertebrates during the second RBMP. COR: Cantábrico Oriental, COC: Cantábrico Occidental, GAL: Cuencas interiores de Galicia, MIÑ: Miño-Sil, DUE: Duero. Source: Wise Freshwater

| | High | Good | Moderate | Poor | Bad |
|---|---|---|---|---|---|
| **Europe** | 12750 (23%) | 20477 (36%) | 15807 (28%) | 5829 (10%) | 1755 (3%) |
| **Spain** | 1101 (35%) | 1275 (41%) | 468 (15%) | 228 (7%) | 75 (2%) |
| **COR** | 31 (29%) | 44 (41%) | 20 (19%) | 9 (9%) | 4 (4%) |
| **COC** | 79 (33%) | 134 (56%) | 24 (10%) | 3 (1%) | |
| **GAL** | 86 (22%) | 238 (60%) | 45 (11%) | 24 (6%) | 3 (1%) |
| **MIÑ** | 113 (47%) | 76 (31%) | 27 (11%) | 20 (8%) | 6 (2%) |
| **DUE** | 283 (46%) | 221 (36%) | 73 (12%) | 28 (5%) | 5 (1%) |

## Official methodology in Spain

In 2013, an official methodology for macroinvertebrate sampling and processing was established in Spain (ML-Rv-I-2013, MAGRAMA, 2013b). Even so, other regions of the peninsula, as the Basque country, have specific sampling protocols (Agencia Vasca del Agua, 2021).

**Figure II-6** Kicksampling of macroinvertebrates in the River Nalón (Asturias, upstream of the Tanes reservoir and downstream of the Deboyu cave).

The official methodology for macroinvertebrate sampling and processing (ML-Rv-I-2013, MAGRAMA, 2013b) consists of a multihabitat, stratified and semiquantitative sampling of 20 kicks (sampling units) distributed over a 100 m section of the river. The sampling units are known as kicks because they consist mainly of removing the established area in front of the net with the foot, in order to free the macroinvertebrates from the bottom and allow them to enter the net dragged by the current. This sampling method is known as kicksampling and the resulting sample is known as kicksample or bulk sample. Each kick covers an estimated stream bottom area of 0.125 $m^2$ (semiquantitative). Habitats that represent ≥ 5 % of the total surface are sampled (multihabitat). The quantity of samples collected from each habitat is determined by its proportion relative to the entire study area (stratified). A macroinvertebrate sampling net of 500 µm mesh size, with a frame having a base of 0.25 m and a height of 0.25 m or more is used for sampling.

**Figure II-7** Taxus operators taking and processing benthic macroinvertebrate samples. Top-left: Kicksampling of macroinvertebrates in the River Del Valle (Somiedo, Asturias). Top-right: Detail of underwater sampling "kick". Bottom-left: Physico-chemical data collection of river water. Bottom-right: Sample processing under magnification for macroinvertebrate family identification.

All biota and material collected in the 20 kicks are stored together with a preservative liquid (e.g. ethanol). Once the sample arrives at the laboratory, it is subjected to a sieving process which separates the sample into 3 sieves (5 mm, 1 mm and 0.5 mm pore size) and removes the remaining stones and organic material. A portion of the individuals present in each sieve is then randomly sub-sampled and a minimum of 100 individuals from each sieve are identified under the magnifier. Only those taxonomic groups listed in the protocol are considered in the identification. In terms of resolution, the identification of macroinvertebrate taxa is multilevel as it includes identifications at different taxonomic levels, with the family level being the most common (1 genus, 121 families, 1 superorder, 1 subclass, 1 class; MAGRAMA, 2013a,Table XIV-1). Hereafter we will refer to all these taxa as 'IBMWP taxa'. These IBMWP taxa are clustered by the index in groups mainly at order level, hereafter considered as 'IBMWP groups' (1 suborder, 8 orders, 2 subclasses, 2 classes, 1 subphylum, 1 phylum; MAGRAMA, 2013a, Table XIV-1).

Once the macroinvertebrate composition and abundance data for each sample is known, the values of the official indices are calculated. In the RD 817/2015, two indices are considered for this biological quality element, one qualitative (IBMWP) and one semi-quantitative-multimetric (METI).

**The IBMWP index** (Iberian Biomonitoring Working Party) is the most widely used index for calculating river ecological status in Spain. The IBMWP index is qualitative, i.e. the abundance of different taxa is not taken into account in the calculation of the index, and cumulative, i.e. the index value increases with each detection of a new taxon included in the

index (IBMWP taxa). However, not all taxa contribute the same value when detected in the sample, since the IBMWP assigns different values to taxa in relation to their sensitivity/tolerance to anthropogenic impacts, where 1 is the value that indicates lower sensitivity and 10 is the value that indicates higher sensitivity. Therefore, the increase in diversity detected at the taxonomic resolution set by the index always translates into an increase in the index values, but in a weighted manner.

The WDF requires macroinvertebrate indices to be quantitative, that is, they must include data on the abundance of the different families, and multimetric, that is, their final value must be obtained by combining different metrics. However, RD 817/2015 envisages the use of the IBMWP index to determine the ecological status of water bodies, although it does not meet these two requirements as it is a qualitative index and only takes into account the presence or absence of the different families of macroinvertebrates. In order to comply with the requirements of the WFD, the NORTI index (North Spanish Indicators system) was designed to evaluate the ecological status using macroinvertebrates of all the rivers present in the Cantabrian and Miño-Sil hydrographic confederations (Pardo et al., 2010). Sometime later it was decided to extend it to all the rivers in the north of Spain, thus creating the METI index.

**The METI index** (Multimétrico Específico Del Tipo De Invertebrados Bentónicos) is a mutimetric index used mainly in rivers in the northwest of the Iberian Peninsula as it can only be applied to 9 of the 32 river typologies recognised in RD 817/2015. Furthermore, the metrics used vary depending on the typology to which it is applied. The index integrates metrics using sensitive families specific to each river typology, measures of abundance of the different families as well as diversity indices such as Margalef or Bray-Curtis.

For each of these indices, the annex II of the RD 817/2015 includes the reference values for each type of river. These reference values are the result of applying the indices to sites with little or no anthropological impact, stipulated in accordance with the WFD guidelines, and represent an ideal condition against which the values obtained at the sites under study can be compared (European Commission, 2003; Pujante, Ana María et al., 2016). Therefore, once the value of the indices has been obtained, the Ecological Quality Ratio (EQR) is calculated, which consists of dividing the value of the index at the sampled point by the reference value for the corresponding river typology. EQR values usually range from 0 to 1 but sometimes values above the reference value can be obtained and values slightly above 1 can be found. In turn, each of the indices for each river typology has cut-off values for this EQR that divide the characterisation of the ecological status of the river into the 5 categories established by the WFD as bad, poor, moderate, good and very good. These values have been internationally inter-calibrated to make status classifications intercomparable between EU countries (Bennett et al., 2011; Birk et al., 2018; M. Furse, Hering, Brabec, et al., 2006; Munné & Prat, 2009).

### Limitations

As previously mentioned, routine biomonitoring of macroinvertebrates has typically relied on morphological identification techniques to identify the different specimens collected in each sampling station at high taxonomic levels. This method based on morphological identification, even when well established, has been shown to have certain limitations when used routinely. These include long sample processing times, low taxonomic resolution, the need for expert taxonomists and difficulties in identifying some cryptic species at certain levels, among others (Blancher et al., 2022; Bush et al., 2019; Haase et al., 2006, 2010; Hering et al., 2010; Ntislidou et al., 2020). These limitations could potentially be overcome by the development and implementation of DNA-based techniques such as metabarcoding (Blackman et al., 2019a). This, together with the substantial reduction in sequencing costs over the last decades (Wetterstrand, 2023) has catalysed research into genomic tools in aquatic ecosystems biomonitoring (Cordier et al., 2021; Deiner et al., 2017; Pawlowski et al., 2018).

## Biomonitoring 2.0: DNA based river biomonitoring.

### Theoretical framework of genetic species identification

Five main characteristics of the DNA molecule make it an extremely useful tool for molecular species identification:

- It is universal as almost all organisms possess DNA, allowing for broad application in the identification of any type of organism, from bacteria to plants and animals.
- It is present in all biological tissues or fluids containing nucleated cells (or non-nucleated cells with plastids and/or mitochondria), which allows its analysis from almost all types of biological substrates (saliva, faeces, plant seeds, milk, etc).
- It is a stable biological molecule that can be recovered from biological material, even after extreme stress conditions (processed food products, faeces, bone remains, blood stains, etc).
- It is a molecule that can be shed into the environment by organisms in what is known as environmental DNA (hereafter ‘eDNA’) and can therefore be collected for indirect detection/identification.
- DNA sequences contain both highly conserved and highly variable regions.

Taking advantage of these characteristics, several techniques for the identification and differentiation of species using DNA were developed in the last two decades of the 20th century (Busch & Nitschko, 1999; Kumar et al., 2015; Rasmussen & Morrissey, 2008). These

include techniques such as e.g. amplified fragment length polymorphism (AFLP), random amplified polymorphic DNA (RAPD), restriction fragment length polymorphisms (RFLP), Specific PCR (PCR) or PCR-linked single stranded conformation polymorphism (PCR-SSCP), etc. However, these techniques had some limitations. They required a specific design for each new species to be analysed, they could be time consuming and some, such as RAPD, had low reproducibility (Pérez et al., 1998).

A molecular species identification technique that emerged as significantly more efficient, reliable, and reproducible is DNA barcoding (hereafter 'barcoding'). The technique was popularised by Hebert P. et al in 2003, who described it as the use of specific DNA sequences that act as genetic 'barcodes' embedded in each cell. The advantages of this technique include its high reproducibility, cost-effectiveness, potential applicability to the identification of all species using uniform procedures and being relatively simple to perform. This technique involves the amplification by PCR and sequencing of a relatively short region of the genome (200-800 bp) using universal primers. The region must have a sufficiently variable part to allow differentiation between species, surrounded by two evolutionarily highly conserved sequences for which complementary primers to amplify for many different species can be designed. The most used region for metazoans is a fragment of the mitochondrial gene encoding cytochrome oxidase I (COI). Specifically, a 658 bp region known as the Folmer region, defined by primers LCO1490 and HCO2189 (Folmer et al., 1994).

Hebert et al. (2003) explained the power of barcoding using combinatorial principles: Since DNA consists of only four possible nucleotides at each position, examining just 15 nucleotide positions can generate 4^15 (approximately 1 thousand million) unique codes, which is 100 times the number needed to uniquely identify each species if each had a unique barcode. However, a protein-coding gene as COI keeps most variations at the third nucleotide position of codons as they are weakly constrained by selection due to their degeneracy. Moreover, the nucleotide composition at third-position sites often shows a strong bias (A–T in arthropods, C–G in chordates), which reduces information content. Nevertheless, even if the A–T or C–G proportion were 1, examining just 90 base pairs (hereafter 'bp') would still provide 1 thousand million possible alternatives (2^(90/3) = 4^15). In comparisons of closely related species, given a modest mutation rate (e.g. 2% per million years), 13 diagnostic nucleotide differences can be expected in a 658 bp comparison of species with a million-year history of reproductive isolation.

Since Hebert et al. (2003), more than three thousand of scientific papers have been published referring to barcoding (DeSalle & Goldstein, 2019). Additionally, the necessary

infrastructure has been developed to expand and establish DNA barcoding as a universal tool. This includes major sequencing diversity programs and projects, such as the International Barcode of Life Project (iBOL) and the Consortium for the Barcode of Life (CBOL). These have also been instrumental in generating and curating large-scale DNA barcode data. Comprehensive reference databases such as GenBank (Benson et al., 2013a) and the Barcode of Life Data System (BOLD) (Ratnasingham & Hebert, 2007), which store and provide access to extensive DNA sequence data. Sequence identification tools such as BLAST (Camacho et al., 2009) and the BOLD Identification System facilitate the matching of unknown sequences to known references. Together, these resources provide a solid foundation for advancing the field of molecular taxonomy and have provided the necessary basis for barcoding to become a routine tool in many laboratories around the world.

However, barcoding has a major limitation for use in routine river biomonitoring. Namely barcoding uses capillary electrophoresis sequencing (Sanger method), which is not suitable for mixed samples containing amplified DNA from multiple organisms. This methodology requires specimens to be individualised before DNA extraction or use a specific primer pair for each species, making it inefficient. Advances in sequencing techniques with the development of high-throughput sequencing (HTS) have made it possible to overcome this limitation, giving rise to a new molecular tool known as 'metabarcoding'.

## Metabarcoding

DNA metabarcoding (hereafter 'metabarcoding') is a powerful and efficient method for biodiversity assessment that combines barcoding with HTS technologies such as Illumina, Ion Torrent or PacBio (Liu et al., 2019). This approach allows for the simultaneous identification of multiple species in a mix sample without the need for prior separation of individual organisms. It was first used to study bacterial communities (Sogin et al., 2006), and over time, it has been adapted to investigate macroorganism communities (Aylagas et al., 2014; Deiner et al., 2017).

In all cases, metabarcoding is based on the amplification and sequencing of one or more barcoding markers from samples containing DNA from several different organisms. The metabarcoding workflow varies depending on the sequencing platform used. Even for the same platform, there are different strategies that significantly modify pre- and post-sequencing protocols (Bohmann et al., 2022). The most common procedure in the application of metabarcoding for biomonitoring is the two-step PCR labelling strategy and sequencing on Illumina sequencers (Figure II-8). In this approach, a first PCR is performed to amplify the barcode region of the genome and incorporate a sequence for the primer recognition in the second PCR. During the second PCR the two indices that differentiate

each sample and the Illumina adapters (i5 and i7) are incorporated to the amplicon for sequencing. The products of the second PCR are purified and pooled together in an equimolar concentration for sequencing (multiplexing). Depending on the size of the fragment and the sequencing depth required (number of sequencing reads per sample) different Illumina devices can be used (Miseq, Hiseq, Novaseq, etc). After sequencing, reads are demultiplex (separated per sample for the downstream analyses) and bioinformatically processed (quality control, dereplication, chimera removal, denoising, etc) until a list of ASVs (amplicon sequence variant) or OTUs (Operational Taxonomic Unit) are obtained. These are then taxonomically assigned to species using the sequences available in genetic reference databases.

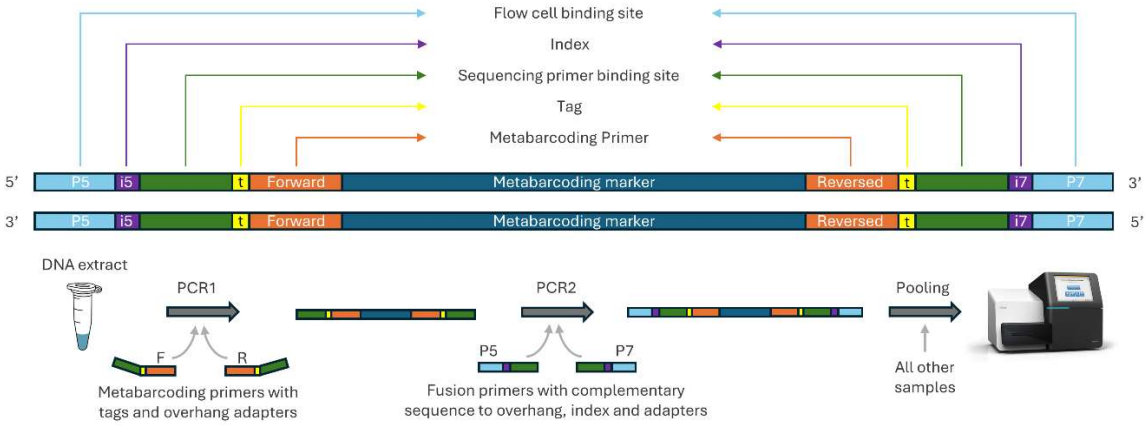



**Figure II-8** General scheme of the metabarcoding process (two-step tagged strategy) and the composition of a dual-tagged and dual-indexed Illumina metabarcoding library sequence.

Although there are many different methodological approaches to metabarcoding, and they are still being improved, some of the general advantages and disadvantages of using metabarcoding, compared to traditional morphological identification techniques, can already be highlighted.

*Taxonomic resolution, scalability, automatization, and optimization.*

Processing macroinvertebrate samples by morphological identification has a trade-off between taxonomic resolution (taxonomic level of identification) and the required expertise, cost, and time of the identification process (F. C. Jones, 2008). This limitation has forced the taxonomic resolution of the indices to be adapted to financial and logistical realities (Hawkes, 1997). The new metabarcoding processing of these samples overcomes this limitation, allowing species-level identifications to be obtained without further effort.

Another major advantage of DNA metabarcoding is its scalability, allowing the simultaneous analysis of multiple samples, which is particularly useful for large-scale monitoring programmes, providing a comprehensive insight into the biodiversity of river ecosystems (Ficetola & Taberlet, 2023). Sample processing protocols for morphological identification require a hands-on processing of samples one at a time through washing,

sieving and identification. Whereas the extraction, amplification and sequencing processes are characteristic of molecular techniques which can be optimised to process several samples simultaneously. Additionally, depending on the protocols used, some of the metabarcoding sample processing steps can be automated using liquid handling robots, significantly reducing the hands-on sample processing time (Buchner, Macher, et al., 2021). The identification and status calculation process can be fully automated if the analysis, values, and criteria for bioinformatic processing and ecological status inferring are standardised (Buchner et al., 2022; Buchner & Leese, 2020).

Finally, it should be noted that morphological identification techniques have been optimised, evaluated and intercalibrated over the years in Europe (Friberg et al., 2006; M. Furse, Hering, Brabec, et al., 2006). In contrast, DNA metabarcoding protocols are gradually being optimised, both in terms of efficacy (Buchner, Macher, et al., 2021; Leese et al., 2020) and in terms of processing time and cost (Buchner, Beermann, et al., 2021; Buchner, Haase, et al., 2021; Buchner, Macher, et al., 2021), but they are still far from reaching their optimisation ceiling. DNA metabarcoding requires further research and development to reach a sufficiently high technology readiness level (TRL) for routine application in biomonitoring (Blackman et al., 2019).

*Abundance data and sample types.*

One of the outputs of HTS sequencing is the number of reads per sequence (ASV), which can then be clustered in OTUs, or in species (to which different ASV and OTUs may be assigned). This number of reads has been widely used as an indirect measure of the relative abundance of each species in a sample (see Lamb et al., 2019). Nevertheless, the relative abundances estimated from the number of reads and those truly observed often differ. This is because several biases are introduced in the process (biomass variation, copy number variation, differences in extraction efficiency, etc), but the one that has been highlighted in literature as the main problem is the bias introduced by PCR (Elbrecht & Leese, 2015; Shelton et al., 2023). PCR amplification efficiencies can differ between species, resulting in changes to the final relative composition of the sample. Additionally, the efficiency may be influenced by the sample composition, preventing the use of general correction factors (Sickel et al., 2023). This can be considered the main limitation of DNA metabarcoding for biomonitoring, as official indices in many countries use abundance data to infer the ecological status of rivers as WFD demands. Some countries, such as Denmark and Germany, where qualitative indices are not available, have tested the use of presence/absence data (1/0) to calculate quantitative indices. They have produced similar results, highlighting the possibility of only using presence/absence data without losing considerable information (Beentjes et al., 2018; Buchner et al., 2019). In the case of Spain, there is an official qualitative index which is valid for calculating the ecological status of almost all rivers in the Iberian Peninsula, the IBMWP (Alba-Tercedor et al., 2004; MAGRAMA,

2013a). Therefore, the presence/absence data obtained by metabarcoding can be used directly for river condition assessment with existing official indices.

Different types of samples can provide the DNA used for metabarcoding. Bulk samples would be the equivalent for metabarcoding of kicksamples taken for morphological identification, except that all the material coming into contact with the sample must be sterilised beforehand to avoid any contamination. Bulk samples can be processed in different ways. The most direct way is to homogenise the content and take a subsample for DNA extraction, but this destroys the individual's morphology and makes it impossible to carry out a morphological analysis or recover any species of interest at a later time (Buchner, Haase, et al., 2021). To solve this problem, two other methods of obtaining DNA from bulk samples without homogenising the samples have been proposed. These are extracting DNA from the ethanol or liquid used to preserve bulk samples (Martins, Galhardo, et al., 2019) or performing a DNA extraction from tissue that does not affect the morphology of the individuals (Batovska et al., 2021). The latter two techniques present additional complications, as DNA extraction without homogenisation implies different efficacy depending on the sclerotization or the use of cases or protective wrappings by the different species (Martins, Porto, et al., 2021; Martoni et al., 2022). Bulk sampling is invasive and time-consuming, therefore the use of environmental samples is also being studied as a faster and indirect way of sampling macroinvertebrate DNA for biomonitoring (Vourka et al., 2023a). These environmental samples may be water (e.g.: Fernández et al., 2018, 2019), sediment (Blackman et al., 2017) or biofilm (Rivera et al., 2021) and the DNA obtained from them is known as environmental DNA.

## Environmental DNA

In brief, environmental DNA (hereafter 'eDNA'), refers to genetic material from species that can be captured and isolated through diverse methods from an environmental sample, i.e. water, air, soil, biofilm, etc (Power et al., 2023; Taberlet et al., 2018). eDNA can be classified into two main types: organismal and extra-organismal DNA (Rodriguez-Ezpeleta et al., 2021). Organismal DNA is extracted directly from whole organisms (bacteria, plankton, etc), whereas extra-organismal DNA includes cellular debris, secretions, or other biological material shed by organisms (Barnes & Turner, 2016; Taberlet et al., 2018).

Studies have demonstrated the usefulness of eDNA in monitoring biodiversity and detecting invasive species without the need for physical captures or visual observations, thus reducing the impact on natural habitats (Beng & Corlett, 2020; Bernos et al., 2023; Fonseca et al., 2023; Kestel et al., 2022; Rees et al., 2014). However, the accurate application of eDNA requires a thorough understanding of the environmental processes affecting DNA persistence and transport, as well as stringent methodological controls to minimise contamination and ensure reliable results (Mauvisseau et al., 2022; Rodriguez-Ezpeleta et al., 2021; Sepulveda, Hutchins, et al., 2020). Currently, there is no consensus

regarding eDNA sampling, capture, preservation, and extraction methods, or the choice of PCR primers, sequencing platforms, reading depths and bioinformatic pipelines (Vourka et al., 2023a). In the case of macroinvertebrate monitoring using eDNA, the most commonly used workflow consists of filtering the river water through a 0.22 to 0.45 µm filter and applying metabarcoding to the DNA extracted from the filter. As an example of the application of eDNA for macroinvertebrate biomonitoring in the Iberian Peninsula, Fernández et al. (2018 and 2019) applied this technique in the River Nalón, located in a protected area, using different markers and sequencing platforms and comparing the results with those of the morphological approach.

## Developed globally, applied locally.

As previously shown, much progress has been made in the application of macroinvertebrate metabarcoding to river biomonitoring. Much of this work has been carried out and optimised for application in other countries, with a view to global application (Blackman et al., 2019b; Elbrecht, 2017; Elbrecht & Steinke, 2019; Martins, Beja, et al., 2019). However, certain characteristics of the application of macroinvertebrate metabarcoding for river biomonitoring vary depending on the country in which it is applied:

- species, genera, and even families of macroinvertebrates found in rivers and the knowledge about them.
- sampling protocols and indices, even within countries of the European Union. Indices can vary not only in the metrics used to calculate ecological status, but also in the families considered, the sensitivity/tolerance value of each, the taxonomic resolution, etc.
- river typology and reference values for each.

These differences between countries require further work to identify and overcome the barriers that prevent the full local integration of this new technology. Overcoming these obstacles will ensure that Spain is not left out of the general development of the technology. To this end, it is a worthwhile to explore the incorporation of metabarcoding into local routine river biomonitoring projects, and to analyse the resulting data with a view to developing a roadmap for full local implementation.

# III. Objectives

The general objective of this thesis is to assess the readiness, accuracy, and applicability of using metabarcoding for macroinvertebrate-based river biomonitoring in Northwest Spain, using bulk and eDNA samples.

## Specific objectives

Within the legal framework of river biomonitoring in Spain, this thesis aims to:

- Assess the role of the completeness of macroinvertebrate genetic databases and its impact on the use of metabarcoding for the biomonitoring of the Iberian rivers.

- Explore the application of metabarcoding and eDNA for studying macroinvertebrate species diversity and detecting invasive and protected species.

- Study the accuracy of metabarcoding using bulk and eDNA samples in river biomonitoring compared to the gold standard approach of morphological identification and identify the sources of variation in detection between these two methodological approaches.

- Propose a roadmap to address the identified limitations, improve the performance and accuracy of metabarcoding, and facilitate its effective implementation in river biomonitoring in Spain.

# IV. Objetivos

El objetivo general de esta tesis es evaluar la preparación, exactitud y aplicabilidad del uso de *metabarcoding* para la biomonitorización de macroinvertebrados fluviales en el Noroeste de España, utilizando muestras *bulk* y *eDNA*.

## Objetivos específicos

Dentro del marco legal de la biomonitorización fluvial en España, esta tesis pretende:

- Evaluar el papel de la exhaustividad de las bases de datos genéticas de macroinvertebrados y su impacto en el uso del metabarcoding para la biomonitorización de los ríos ibéricos.

- Explorar la aplicación del *metabarcoding* y el *eDNA* para el estudio de la diversidad de especies de macroinvertebrados y la detección de especies invasoras y protegidas.

- Estudiar la precisión del *metabarcoding* utilizando muestras *bulk* y *eDNA* en la biomonitorización de ríos en comparación con el enfoque estándar de la identificación morfológica e identificar las fuentes de variación en la detección entre estos dos enfoques metodológicos.

- Proponer una hoja de ruta para abordar las limitaciones identificadas, mejorar el rendimiento y la exactitud del *metabarcoding*, y facilitar su implementación efectiva en la biomonitorización fluvial en España.

# V. Study on the availability of reference sequences of freshwater macroinvertebrate species from the Iberian Peninsula.

*The content of this chapter is published as: Fueyo, Á., Granero-Castro, J., Villazán Peñalosa, B., & Borrell Pichs, Y. (2021). Stones in the road prevents effective implementation of eDNA-based freshwater quality monitoring in the Iberian Peninsula. ARPHA Conference Abstracts, 4. https://doi.org/10.3897/aca.4.e64974*

In recent years, the use and development of the metabarcoding technique for biodiversity studies and the biomonitoring of surface waters have greatly increased (Ficetola & Taberlet, 2023; Keck, Blackman, et al., 2022). Despite its potential, the metabarcoding sequencing technique is still undergoing refinement and optimization. One crucial area for improvement is taxonomic assignment, as the process relies on matching the sequencing data obtained from samples against genetic databases such as BOLD or GenBank to identify species (Benson et al., 2013b; Ratnasingham & Hebert, 2007). The lack of genetic sequences for some macroinvertebrate species in these databases can result in a failure to correctly identify the newly generated operational taxonomic units (OTUs), this can lead to misassignments and false negatives, undermining the accuracy of the technique (Keck, Couton, et al., 2022). The incompleteness of reference databases is a recurrent issue mentioned in various studies (Martins, Feio, et al., 2021; Martins, Galhardo, et al., 2019; Múrria et al., 2020; Weigand et al., 2019). Indeed, available databases are geographically biased, with certain areas being extensively represented due to numerous research projects and funding, resulting in abundant genetic sequences for those species. However, native species from other less-surveyed regions face an underrepresentation in these databases. Furthermore, a sequence obtained from a specimen in one region may not be suitable for identifying another individual of the same species from a different region, as intraspecific divergence tends to increase with geographical scale, resulting in more uncertain genetic species identification (Bergsten et al., 2012). Thus, to use metabarcoding as a truly effective tool for assessing water quality in freshwater ecosystems, it is essential to conduct a preliminary check between the checklist of macroinvertebrate species used as bioindicators in a region and their representation in the available genetic databases.

In the case of the Iberian Peninsula, there is a major obstacle when testing the completeness of the genetic databases for freshwater macroinvertebrate species, as there is no complete checklist for this group. In other words, we do not have a comprehensive inventory of all the bioindicator species present in the rivers of the peninsula, unlike other neighbouring regions such as the United Kingdom, Madeira and the Azores island (Gunn et al., 2018; Hughes et al., 1998; Raposeiro et al., 2012). This, combined with the fact that many of the macroinvertebrate bioindicator families are composed of not only freshwater

species but also terrestrial or marine species, makes it difficult to directly approximate the actual percentage of freshwater macroinvertebrate species in the region represented in the available reference databases. Two studies have previously attempted to estimate the coverage of Iberian species in reference databases using partial species inventories, yielding notably divergent outcomes (Múrria et al., 2020; Weigand et al., 2019), 37.2% and 75.2%, respectively.

The aim of this section is to directly assess the level of completeness of one of the main genetic databases, specifically the BOLD database, for sequences of species of the different IBMWP taxa. For this purpose, freshwater river species of IBMWP taxa were compiled from two European (Schmidt-Kloiber & Hering, 2015; Weigand et al., 2019) and one Iberian (Múrria et al., 2020) database/list, creating an inventory as complete as possible with the available information (Figure V-1).

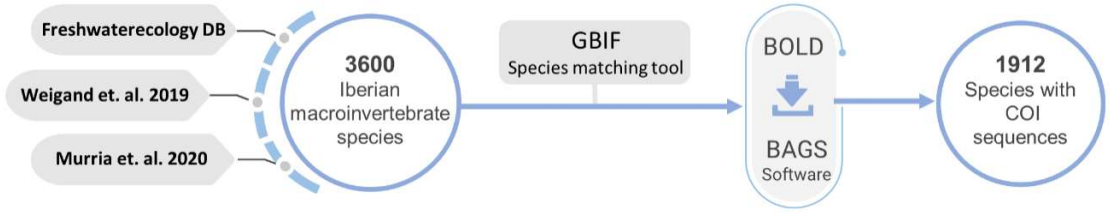



**Figure V-1** Data analysis workflow.

The resulting list was checked against the GBIF database using the 'species matching tool' to search for duplicates by synonyms or fuzzy names (typographical errors) (GBIF, 2024). The comparison analyses between the databases resulted in a new list of 3,600 species representing the 125 taxonomic quality indicator macroinvertebrate groups covered by the IBMWP. Eleven of these groups (9%) had no species records and the remaining 34 groups (27%) had fewer than five species records. All available COI sequences were then retrieved from the BOLD database using the BAGS software in January 2021 (Fontes et al., 2020). We found that of the 3,600 species in the resulting list, only 1,912 (53%) had at least one COI barcode sequence (Figure V-2). This value is intermediate between the two previous studies but is only an estimate, as it is not derived from the analysis of a complete list of species.

| Taxonomic group | Number of species in databases | % of sequencing coverage |
|---|---|---|
| Acariformes | 0 | |
| Chrysomelidae | 20 | 75% |
| Curculionidae | 27 | 37% |
| Dryopidae | 16 | 38% |
| Dytiscidae | 178 | 70% |
| Elmidae | 32 | 69% |
| Gyrinidae | 10 | 40% |
| Haliplidae | 19 | 79% |
| Helophoridae | 36 | 47% |
| Hydraenidae | 144 | 78% |
| Hydrochidae | 9 | 67% |
| Hydrophilidae | 91 | 67% |
| Hygrobiidae | 0 | |
| Noteridae | 4 | 50% |
| Psephenidae | 0 | |
| Scirtidae | 26 | 42% |
| Asellidae | 31 | 81% |
| Astacidae | 3 | 100% |
| Atyidae | 2 | 100% |
| Corophiidae | 8 | 63% |
| Gammaridae | 39 | 51% |
| Ostracoda | 0 | |
| Palaemonidae | 1 | 0% |
| Anthomyiidae | 9 | 67% |
| Athericidae | 3 | 33% |
| Blephariceridae | 10 | 10% |
| Ceratopogonidae | 167 | 37% |
| Chironomidae | 500 | 62% |
| Culicidae | 55 | 85% |
| Dixidae | 7 | 57% |
| Dolichopodidae | 181 | 15% |
| Empididae | 67 | 22% |
| Ephydridae | 116 | 38% |
| Limoniidae | 62 | 48% |
| Psychodidae | 62 | 32% |
| Ptychopteridae | 2 | 100% |
| Rhagionidae | 17 | 29% |
| Scatophagidae | 0 | |
| Sciomyzidae | 0 | |
| Simuliidae | 36 | 69% |
| Stratiomyidae | 56 | 21% |
| Syrphidae | 64 | 55% |

| Taxonomic group | Number of species in databases | % of sequencing coverage |
|---|---|---|
| Tabanidae | 96 | 14% |
| Thaumaleidae | 9 | 22% |
| Tipulidae | 160 | 39% |
| Baetidae | 47 | 57% |
| Caenidae | 10 | 80% |
| Ephemerellidae | 17 | 24% |
| Ephemeridae | 4 | 75% |
| Heptageniidae | 44 | 45% |
| Leptophlebiidae | 20 | 45% |
| Oligoneuriidae | 4 | 25% |
| Polymitarcyidae | 1 | 100% |
| Potamanthidae | 1 | 100% |
| Prosopistomatidae | 1 | 0% |
| Siphlonuridae | 8 | 38% |
| Aphelocheiridae | 3 | 33% |
| Corixidae | 65 | 66% |
| Gerridae | 19 | 79% |
| Hydrometridae | 2 | 100% |
| Mesoveliidae | 3 | 67% |
| Naucoridae | 4 | 25% |
| Nepidae | 11 | 55% |
| Notonectidae | 20 | 55% |
| Pleidae | 2 | 50% |
| Veliidae | 23 | 48% |
| Erpobdellidae | 4 | 100% |
| Glossiphoniidae | 7 | 86% |
| Hirudidae | 0 | |
| Piscicolidae | 1 | 100% |
| Crambidae | 16 | 63% |
| Ancylidae | 0 | |
| Bithyniidae | 2 | 50% |
| Ferrissia | 1 | 100% |
| Hydrobiidae | 10 | 40% |
| Lymnaeidae | 10 | 100% |
| Neritidae | 4 | 75% |
| Physidae | 4 | 100% |
| Planorbidae* | 18 | 78% |
| Sphaeriidae | 14 | 86% |
| Thiaridae | 0 | |
| Unionidae | 8 | 100% |
| Valvatidae | 4 | 100% |
| Viviparidae | 3 | 100% |

| Taxonomic group | Number of species in databases | % of sequencing coverage |
|---|---|---|
| Sialidae | 3 | 100% |
| Aeshnidae | 11 | 100% |
| Calopterygidae | 4 | 100% |
| Coenagrionidae | 14 | 100% |
| Cordulegastridae | 4 | 100% |
| Corduliidae | 5 | 80% |
| Gomphidae | 12 | 67% |
| Lestidae | 7 | 100% |
| Libellulidae | 35 | 94% |
| Platycnemididae | 3 | 67% |
| Oligochaeta | 14 | 86% |
| Capniidae | 19 | 63% |
| Chloroperlidae | 15 | 33% |
| Leuctridae | 68 | 47% |
| Nemouridae | 48 | 54% |
| Perlidae | 11 | 64% |
| Perlodidae | 23 | 70% |
| Taeniopterygidae | 21 | 62% |
| Beraeidae | 10 | 60% |
| Brachycentridae | 13 | 85% |
| Calamoceratidae | 1 | 100% |
| Ecnomidae | 2 | 100% |
| Glossosomatidae | 28 | 64% |
| Goeridae | 12 | 83% |
| Hydropsychidae | 40 | 50% |
| Hydroptilidae | 67 | 33% |
| Lepidostomatidae | 5 | 60% |
| Leptoceridae | 47 | 55% |
| Limnephilidae | 109 | 58% |
| Molannidae | 0 | |
| Odontoceridae | 2 | 100% |
| Philopotamidae | 28 | 46% |
| Phryganeidae | 6 | 67% |
| Polycentropodidae | 31 | 39% |
| Psychomyiidae | 29 | 48% |
| Rhyacophilidae | 44 | 48% |
| Sericostomatidae | 9 | 78% |
| Uenoidae | 3 | 67% |
| Dendrocoelidae | 0 | |
| Dugesiidae | 4 | 50% |
| Planariidae | 3 | 0% |
| Total | 3600 | 53% |

**Figure V-2** Number of species names available in the three reference databases/list used and the percentage of species with a COI sequence in BOLD database. Yellow flag: Less than five species names. Red flag: no species names. Access to the BOLD database was granted in January 2021.

As can be seen from the results, a complete inventory of Iberian river macroinvertebrate species and reference sequences for all of them is far from being available. Achieving this goal is certainly necessary for a full and effective implementation of metabarcoding for river biomonitoring, but it will require work beyond the scope and possibilities of this thesis.

Despite the lack of sequences for thousands of species, Murria et al. (2020) mention in their work that the species sequenced are probably the most common. This, together with the fact that current river quality indices are collapsed to include only family level detections, raises the question of whether sequencing the remaining species could really have a significant impact on river biomonitoring. Furthermore, it is worth asking whether the gradual increase in reference databases thanks to various sequencing projects, such as the

InBio barcoding project (S. A. Ferreira et al., 2020) or IBOL (International barcode of Life), can gradually complete the databases, or whether a specific and focused sequencing effort is needed.

# VI. The influence of reference genetic databases enrichment using local freshwater macroinvertebrate for metabarcoding based biodiversity studies in river monitoring.

*The content of this chapter is published as: Fueyo, Á., Sánchez, O., Coya, R., Carleos, C., Escudero, A., Cordón, J., Fernández, S., Granero-Castro, J., & Borrell, Y. J. (2024). The influence of databases enrichment using local macroinvertebrate genetic references for metabarcoding based biodiversity studies in river monitoring. Ecological Indicators, 158, 111454. https://doi.org/10.1016/j.ecolind.2023.111454*

To answer the questions raised in the previous section, we undertook a local sequencing effort and assessed whether it had a significant impact on biodiversity and biomonitoring results. We reviewed 92 macroinvertebrate samples collected from rivers in the north-west Iberian Peninsula in 2020-2022 (Figure VI-1) to identify all possible macroinvertebrate morphospecies present in routine biomonitoring samples, and then used them to generate new DNA sequences for the cytochrome oxidase I (COI) 5P region.

The main objectives of this study were to estimate the percentage of macroinvertebrate species present in routine biomonitoring samples that cannot be identified by barcoding due to a lack of sequences in reference databases. Additionally, we assessed the impact of this region-specific sequencing effort on the metabarcoding taxonomic assignments for both biodiversity detection and ecological status inference in freshwater biomonitoring in the northwestern Iberian Peninsula. Moreover, we aimed to investigate the temporal evolution of metabarcoding taxonomic assignment using the BOLD public records from previous years while exploring how this trend is affected by the implementation of local sequencing efforts.

## Materials and methods

### Sampling collection

Since 2007, a major sampling initiative has been undertaken by TAXUS, mainly in the north-west region of the Iberian Peninsula, to collect and analyse a significant number of macroinvertebrate samples from rivers each year. These samples are processed according to the sampling protocol of the Spanish Ministry of Environment (MAGRAMA, 2013b) and the macroinvertebrate taxa are identified to the level required by the IBMWP index (MAGRAMA, 2013a). 34 of these samples, collected in 2020-2021, were post-processed for a review in detail (checking all the individuals) to identify specimens at lower taxonomic levels, searching for all possible morphospecies present in the samples. An additional 58 samples from the period 2021-2022 were then processed to search for new morphospecies,

focusing on the IBMWP taxa that have not appeared in the previous samples. Finally, specimens were collected from a total of 45 samples (Figure VI-1, Figure XV-2 in Annex II).

For 11 of the first 34 samples taken to obtaining morphospecies, duplicate samples were taken as bulk samples for metabarcoding analysis. (Figure VI-1). The sampling protocol of the Spanish Ministry of the Environment (MAGRAMA, 2013b) was followed, and all materials contacting bulk samples were sterilised with 10% bleach and washed twice with distilled water before each sampling.

## Macroinvertebrate sequencing and database improvement

All the morphospecies collected were identified by two independent taxonomists at the most specific taxonomic level possible using the available taxonomic keys (e.g.: Smit, 2020; Tachet et al., 2010; Vieira Lanero, 2000). For those taxa where morphological variability was detected among different individuals, but identification could not be achieved at a lower taxonomic level, individuals were separated into different morphospecies. Finally, up to 3 specimens from each species/morphospecies were collected for subsequent genotyping, for a total of 414 specimens (Fig. 1B).

The COI-5P region was amplified using the primers LCO1490/HCO2198 (Folmer et al., 1994). In case no amplification was achieved with these primers, amplification was attempted with two other primer pairs, jgLCO1490/jgHCO2198 (Geller et al., 2013) as a first alternative and BF3/BR2 (Elbrecht et al., 2019; Elbrecht & Leese, 2017a) as a last alternative. PCRs were carried out using the GoTaq® G2 Flexi DNA Polymerase in a total volume of 40 µL. A total of 2.5 µL of sample was used in a final mix with a final concentration of: 0.25 µM for each primer, 1X for Buffer, 2.5 mM MgCl2, 0.25 mM DNTPs and 0.5 U of Taq. The PCR profile was 95 °C for 5 min initial denaturation, 35 cycles of 95 °C for 45 s denaturation, annealing for 45 s at 48 °C/ 48 °C / 50°C respectively for the different primer pairs, and 72 °C for 30 s elongation, ending with 72 °C for 7 min as a final elongation step. PCR amplifications were checked on a 2% agarose gel, enzymatically purified, and sequenced forward and reverse using the sequencing service from Macrogen Spain. Chromatographs were manually checked (trimmed, primers removed, base-calling errors corrected) in Geneious software v2022.2. The consensus sequences resulting from the alignment of the forward and reverse reads were used for taxonomic assignment against the GenBank database using BLAST and against the BOLD database using BOLD ID ENGINE. The pairwise identity threshold for BLAST and BOLD-ID was set at 97%, as this is the most commonly used threshold in macroinvertebrate metabarcoding studies for taxonomic assignment (Bruce et al., 2021). After all the processing, all the sequences obtained were uploaded to BOLD and GenBank.

## Metabarcoding analysis

For this study, eleven different rivers were sampled in duplicate. One of these two samples was used to assess the ecological status of the river according to the official taxonomical method (MAGRAMA, 2013b), and the other was used as a bulk sample for molecular identification. Bulk samples were homogenised according to the method proposed by Buchner et al. (2021), and 10 g of each sample was taken for extraction using the DNeasy PowerMax Soil kit (Qiagen), with two replicate extractions per sample. For extraction, the manufacturer's instructions were followed, with the first four steps of the protocol replaced by an overnight digestion process with 405 µL of 20 mg/mL Proteinase K in a rotary incubator at 56 °C. The extracted DNA was amplified using a double PCR strategy with BF3/BR2 primers (Bohmann et al., 2022) using the Qiagen Multiplex PCR Plus Kit in a total volume of 25 µL. A total of 2.5 µL of sample was used in a final mix with a final primer concentration of 0.2 µM each (BF3/BR2 in four length-varying versions and a universal tail attached as in Elbrecht et al. (2019) and Elbrecht & Leese (2017)) and 1 x of Qiagen Mastermix. The PCR profile used in PCR1 was 95 °C for 5 min initial denaturation, 30 cycles of 95 °C for 30 s denaturation, 50 °C for 30 s annealing and 72 °C for 30 s elongation, ending with 68 °C for 10 min as a final elongation step. PCR2 was conducted using 1 µL of PCR1 in a final volume of 25 µL with a primer concentration of 0.2 µM each (primers matching the universal tail with an i5/i7 index and P5/P7 Illumina adapters attached were used). The PCR profile used in PCR2 was 95 °C for 5 min initial denaturation, 20 cycles of 95 °C for 30 s denaturation, 61 °C for 30 s annealing and 72 °C for 42 s elongation, ending with 68 °C for 10 min as a final elongation step. All products from the second PCR were purified using Agencourt AMPure XP beads at a 0.7x ratio (20 µL PCR2 product + 14 µL beads). After clean-up, the DNA concentration of each individual product was measured using the Qubit 4 dsDNA BR assay kit and then pooled equimolar with the negative control added at the maximum volume for each individual library (up to 15 µL). The library was sequenced by an external sequencing service (Macrogen, Korea) on an Illumina MiSeq using a 600 cycle V3 kit with 300 bp paired-end sequencing and loaded at 12 pM with 5% Phi-X. Raw reads for the library were received from Macrogen as demultiplexed fastq files (NCBI BioProject Accession: PRJNA956464).

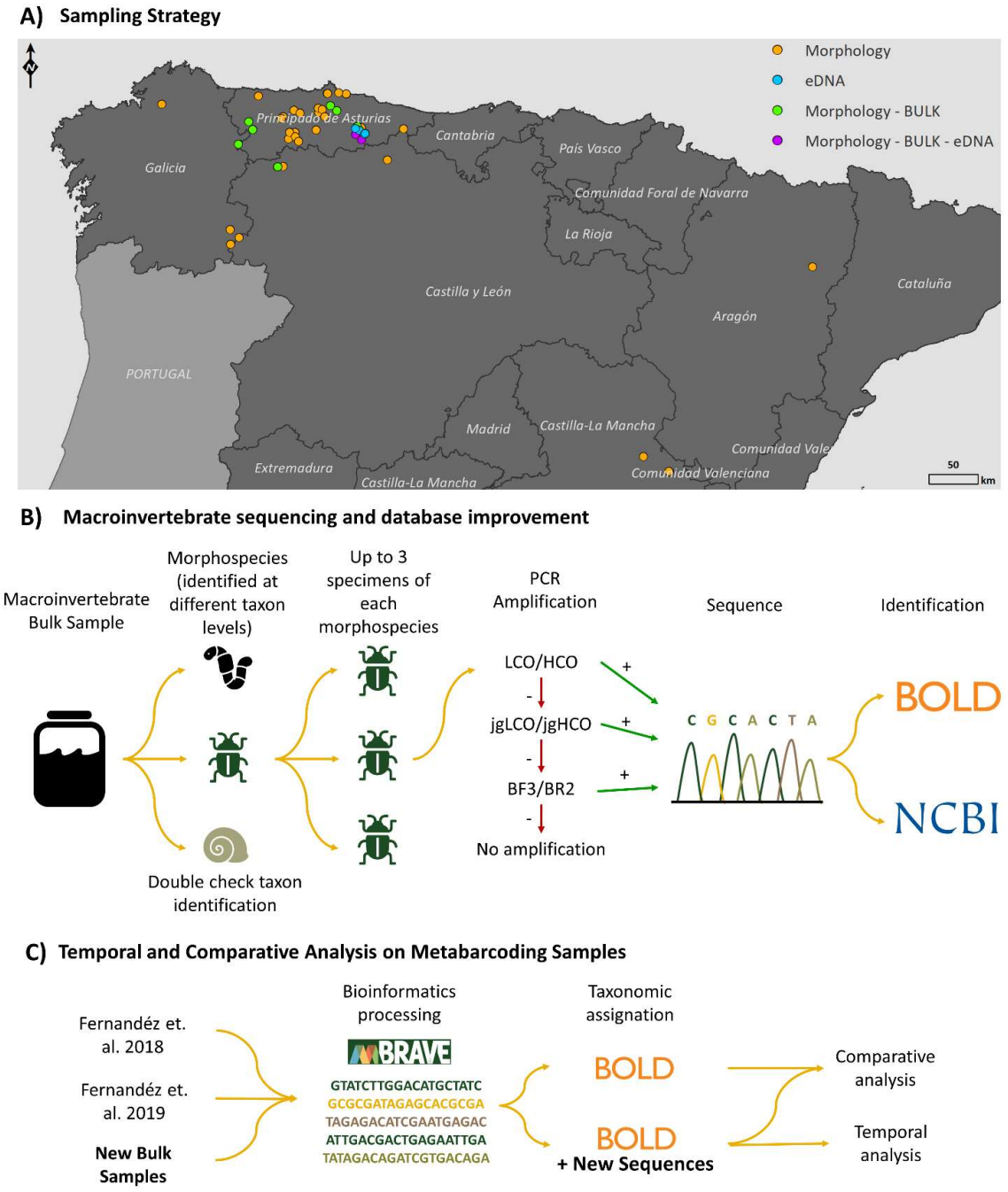



**Figure VI-1** Scheme of the experimental design. A) Map of the northern Iberian Peninsula showing the sampling locations and sample types. Of the morphological samples, only those from which specimens have been taken are shown. B) Methodological workflow for collection, identification, amplification, and sequencing of all morphospecies detected within routine monitoring samples of macroinvertebrates. Each distinct morphospecies encountered in the samples was collected and identified by two expert taxonomists. Subsequently, three individual specimens were chosen from each morphospecies for further analysis. An attempt was made to amplify the DNA of each selected individual using a variety of primers. Sequencing of the amplification products followed, and the resulting sequences were uploaded to the BOLD and GenBank databases. C) Metabarcoding comparative and temporal analysis. Metabarcoding data from three sources are processed using the BOLD genetic database plus a custom reference database using new local species genetic sequences.

## Bioinformatics

Three different sets of metabarcoding data for macroinvertebrate samples were used in this study. A set of eleven bulk-type samples was collected and metabarcoded specifically for this study (hereafter project BULK-ILLUMINA), with corresponding duplicates for morphological identification (see details above). In addition, two sets of six and seven metabarcoded environmental DNA (hereafter eDNA) water samples published by Fernández et al. (2018; 2019) were also used (SRA: SRP124881 and SRP128681). The metabarcoding approach in these two previous works was conducted using the mlCOIintF/jgHCO2198 primer pair (Geller et al., 2013; Leray et al., 2013) and sequenced using Illumina (hereafter project eDNA-Illumina) and ION torrent methods (hereafter project eDNA-ION), respectively (Fernández et al., 2018; Fernández et al., 2019). These eDNA samples also had a corresponding kicknet macroinvertebrate sample taken on the same day and time for morphological identification and inference of ecological status, analysed in their respective studies.

Sequencing results of all projects were uploaded and bioinformatically processed in mBRAVE separately (Multiplex Barcode Research and Visualization Environment, http://mbrave.net/). For Illumina projects (BULK and eDNA), reads were merged with a minimum overlap of 50/25 bp and a maximum substitution value of 20/10 bp respectively. For all projects, primers and adapter sequences were removed by trimming 26 bp from each end, and sequences were then filtered to a minimum quality score of QV20. A size selection of the reads was made for each project, considering the expected amplicon size with an allowed deviation of ± 10 base pairs. Sequences with more than 4% of bases of low-quality value (<20 QV) and those with more than 1% of ultra-low-quality value (<10 QV) were discarded. To remove singletons, the minimum OTU size was set to at least 2 reads; OTU threshold 2% (maximum distance inside a generated OTU); and exclusion from the OTU threshold when sequencing error introduces spurious haplotypes 3%.

Different datasets were used for the taxonomic assignation. Three BOLD reference datasets were used together for taxonomic assignment (SYS-CRLNONARTHINVERT, SYS-CRLNONINSECTARTH and SYS_CRLINSECTA) since this allowed us to include all invertebrate reference sequences in BOLD, constituting what we called the "BOLD default dataset". The assignment analyses were repeated by adding to this default dataset a new "local dataset" (DS-IBMWP) made with all the new reference species sequences obtained in this study, resulting in a new and more comprehensive dataset called the "BOLD improved dataset". Furthermore, to disaggregate the results per year, we download public records from BOLD (all species included in the IBMWP taxa) to construct a historical series of public BOLD records that meet strict quality criteria as set out by the Consortium for DNA Barcoding (CBOL) (BOLD SYSTEMS, 2019). Several historical datasets were constructed by dividing by year the BOLD records from 2003 until 2021 for the taxa represented in the

IBMWP index. The taxonomic assignment made by mBRAVE was produced in the form of BINs (barcode index number). BINs are one of the most widely used molecular operational taxonomic unit (MOTU) delimitation approaches, and for most animal taxa, COI gene MOTUs match with species (Ratnasingham & Hebert, 2013). BIN assignments were accepted at a minimum of 97% identity. BINs from each extraction replicate that only appeared in one PCR replicate were removed.

The IBMWP index was calculated for each of the samples using the official taxonomical method and metabarcoding-based identifications, according to the reference values established for each river typology (Real Decreto 817/2015), which establishes the criteria for monitoring and evaluation of the state of surface waters and environmental quality standards.

## Statistical analysis

All statistical analyses were conducted using R software 4.2.1. As the number of sequences (reads) of macroinvertebrate species obtained from metabarcoding procedures is not proportional to the number of individuals of those species (Elbrecht & Leese, 2015), metabarcoding data were scored as presence/absence for each BIN. A permutation paired sample t test (100000 permutations) was used to examine the significance of differences for ecological status inference between the ecological and morphological approaches.

In the analyses carried out to study the impact of the new sequence records in metabarcoding taxonomical assignment, the mean values of BINs and IBMWP taxa detections (before and after database enrichment procedures, and in comparison, with the classical taxonomic method) were compared using a pairwise t-test after checking their normality using the Shapiro–Wilk test. To study the temporal overview of taxonomic assignment using public BOLD records, we plotted species accumulation curves to compare the changes in taxonomic assignment due to the expansion of the databases over the years. The model used to parameterise the process was a loess model with a 95% confidence interval. The analyses for the comparison between temporal evolution of the ecological status inferences between the morphological and the molecular methods were conducted using the Friedman test (Friedman, 1937). The Friedman test was chosen because it is usually marketed as nonparametric repeated-measures ANOVA and is a natural candidate for comparing diagnostic methods. A Monte Carlo resampling was implemented to effectively compute correct p-values by means of 10,000 shuffles of the within-individual ranks.

## Results

A total of 414 specimens representing 77 different IBMWP taxa (out of the 125 taxonomic groups covered by the IBMWP index) were finally collected from 45 macroinvertebrate samples. The specimens were grouped into 175 morphotypes and identified at the lowest possible taxonomic level (below IBMWP taxonomic resolution) into 121 different taxa. DNA COI sequences were obtained from 156 of morphospecies (89%) (300 sequences). These sequences were blasted against the BOLD and GenBank databases, and 123 out of the 156 morphotypes successfully amplified (79%) were molecularly identified (Table VI-1), resulting in a total of 128 species identified with 97% or greater pairwise identity. The other 33 morphotypes sequenced (21%) obtained had less than 97% pairwise identity, so they are new records in genetic databases for NW macroinvertebrate species. Among the EPT orders (Ephemeroptera, Plecoptera, and Trichoptera), the identification rate in the databases of the sequences exceeded 90% in all three cases. However, in other IBMWP groups where a large number of families are considered, lower identification percentages were obtained, as in the case of Coleoptera (84%) or Diptera (68%).

**Table VI-1** Number of different macroinvertebrate morphospecies amplified and finally identified using the genetic reference databases BOLD, GenBank, and both databases combined. The percentage of morphospecies identified was calculated using the number of morphospecies identified with both genetic databases. *We consider Oligochaeta as Oligochaeta sensu stricto, which refers to clitellates that do not include branchiobdellids and leeches (Martin et al., 2008).

| ***IBMWP Group*** | **Amplified** | **Identified BOLD** | **Identified GenBank** | **Identified Total** | **Identified /Amplified** |
|---|---|---|---|---|---|
| Arachnida | 11 | 3 | 2 | 3 | 27% |
| Coleoptera | 25 | 20 | 19 | 21 | 84% |
| Crustacea | 3 | 1 | 1 | 1 | 33% |
| Diptera | 37 | 24 | 19 | 25 | 68% |
| Ephemeroptera | 10 | 9 | 4 | 9 | 90% |
| Heteroptera | 4 | 3 | 3 | 3 | 75% |
| Hirudinea | 2 | 1 | 1 | 1 | 50% |
| Megaloptera | 2 | 2 | 2 | 2 | 100% |
| Mollusca | 14 | 13 | 13 | 13 | 93% |
| Odonata | 3 | 2 | 3 | 3 | 100% |
| Oligochaeta* | 4 | 4 | 4 | 4 | 100% |
| Plecoptera | 14 | 12 | 12 | 13 | 93% |
| Trichoptera | 26 | 23 | 23 | 25 | 96% |
| Turbellaria | 1 | 0 | 0 | 0 | 0% |
| ***Total*** | **156** | **117** | **106** | **123** | **79%** |

## Metabarcoding results for new bulk samples.

A total of 72 different IBMWP taxa and 565 different BINs were detected using metabarcoding. An average of 24 taxa were detected per sample, with a maximum of 38 taxa per sample and a minimum of 10. The average number of BINs detected per sample was 108, with a maximum of 183 and a minimum of 44 BINs (Figure VI-2). Three IBMWP taxa (Chironomidae, Oligochaeta, and Simuliidae) occur in all samples and are the most abundant families, together with Acariformes and Baetidae.

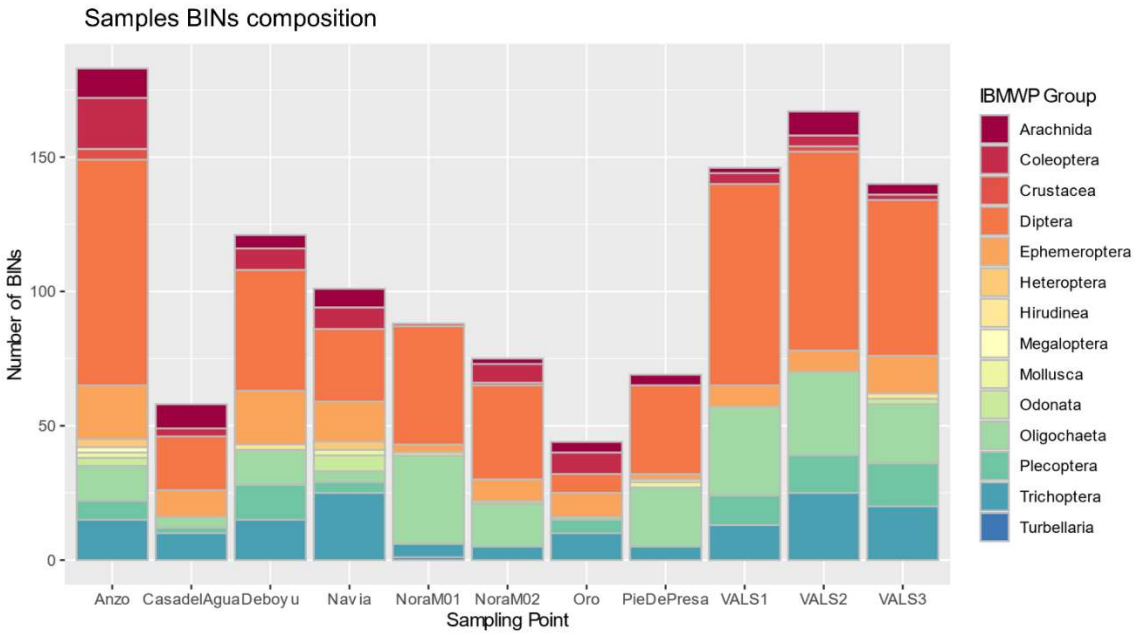



**Figure VI-2** Analysis of the 11 new bulk metabarcoding samples. Composition and number of BINs per IBMWP group of each of the samples.

## Impact of new sequence records on metabarcoding taxonomic assignment

To elucidate the potential impact of expanded reference databases with new local specimen sequences on routine biomonitoring, we analyzed how the availability of these sequences could improve the taxonomic assignment of the 24 metabarcoding samples (11 datasets from this study and 6 + 7 datasets from Fernandez et al. (2018; 2019)) when using both private (mBRAVE) and public BOLD records together. To do so, we compared the number of IBMWP taxa, BINs, and resulting ecological status for the IBMWP index from those 24 metabarcoded samples using the standard mBRAVE datasets (as the BOLD default dataset) and the new dataset (BOLD improved dataset) enriched with the new sequences generated in this study (Figure VI-3). An average of 17.25 more BINs (Figure VI-3A) and 1.54 more IBMWP taxa (Figure VI-3B) were detected when using the BOLD improved dataset. These new detections resulted in a change (improvement) in the ecological status assignments for four of the samples in the study (16.6%) (Figure VI-3C).

The results also indicated that there were no significant differences in the number of detected families when comparing molecular (neither using the BOLD default dataset nor the BOLD improved dataset) and morphological approaches (pairwise t-test: p = 0.219 and p = 0.139).

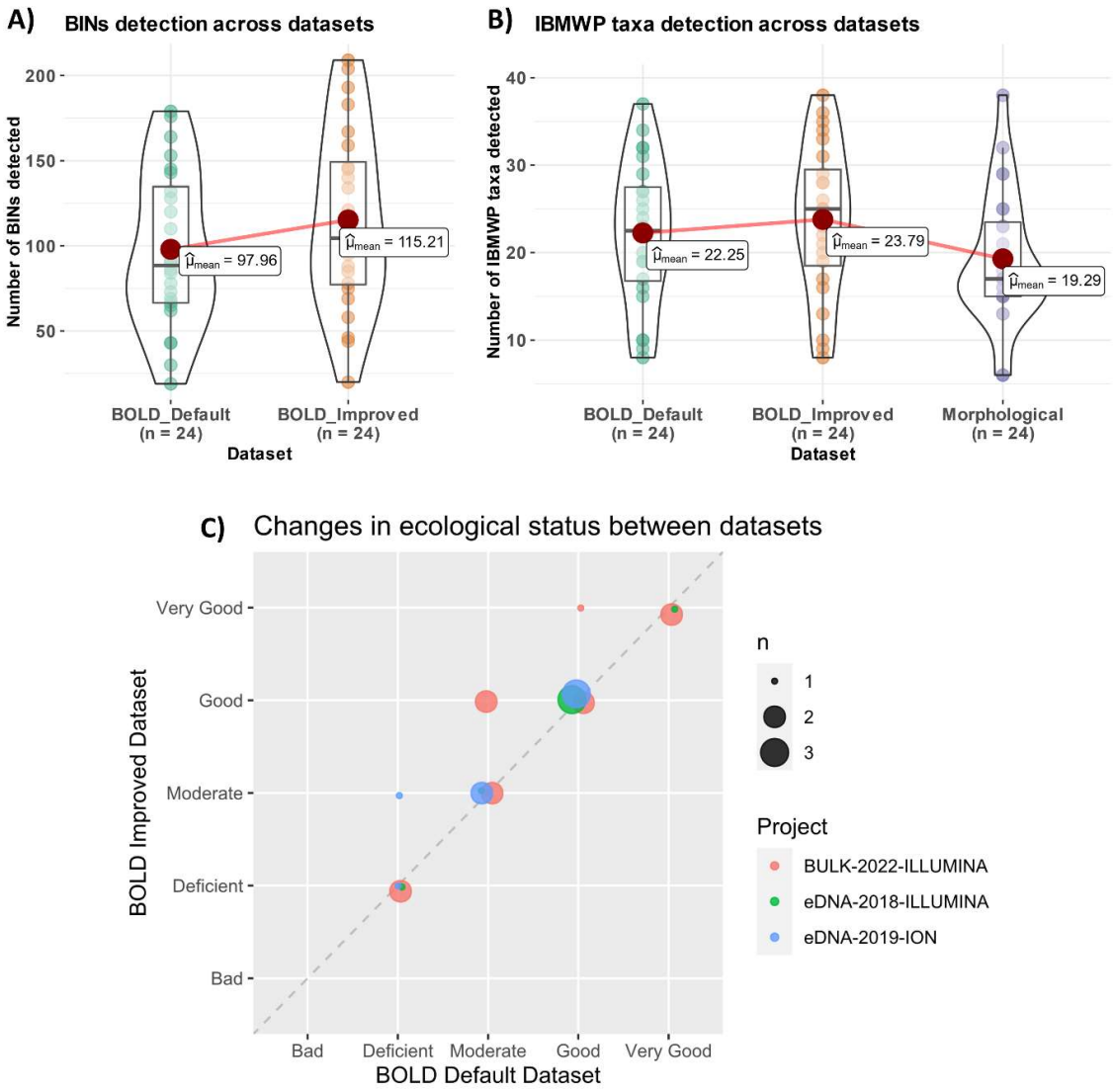



**Figure VI-3** Comparison of the number of different BINs (A) and IBMWP taxa (B) molecularly identified by including the new macroinvertebrate sequences generated in this study for the northern Iberian Peninsula with the standard mBRAVE dataset (BOLD default and improved datasets) and using the morphological classical approach. C) Comparison of the ecological status of 24 samples according to the IBMWP index obtained by using the BOLD default and improved datasets.

## Temporal overview of taxonomic assignment using public BOLD barcoding compliance records on metabarcoding samples.

Genetic databases such as BOLD and GenBank are dynamic in terms of incorporating new records over the years. To assess the effect of this improvement on macroinvertebrate diversity detection and routine biomonitoring, we disaggregated the public BOLD records by year (not available for the private mBRAVE system) and analysed the improvement in

taxonomic assignment in the metabarcoding datasets used in this work over time. As expected, the results show an increase in the detection of macroinvertebrate biodiversity at both the BIN and IBMWP taxa over the years in these samples (Figure VI-4 A & B). However, while the average number of new BINs discovered per sample increased at a rate of seven BINs per year between 2006 and 2011, this rate decreased to only one BIN per year in 2021-2022 (see Figure XV-2 in Annex II). In terms of IBMWP taxa detected, we found a similar pattern, but there was a much more accelerated decline in detections after 2014, when an average value of less than 0.5 new taxa detected per sample per year was reached and maintained until today. When sequences obtained from local samples were added to the taxonomic assignment (BOLD improved dataset), the average assignment per sample increased by 21.5 BINs (32.7 % increase) and 9.17 families (51% increase) (Figure VI-4 A & B). This improved detectability directly affects the inference of the ecological status of the rivers where the samples were taken (Figure VI-4C). If we compare the state inferred from the morphological data (fixed values) and that obtained from the molecular data for the different samples over time, we observe significant temporal differences ($p < 0.05$) among the ecological statuses obtained by the two different methods until that difference reached its minimum value (zero = same ecological status) when the new local sequences (BOLD improved datasets) were included ($p = 1.00$) (Figure VI-4C).

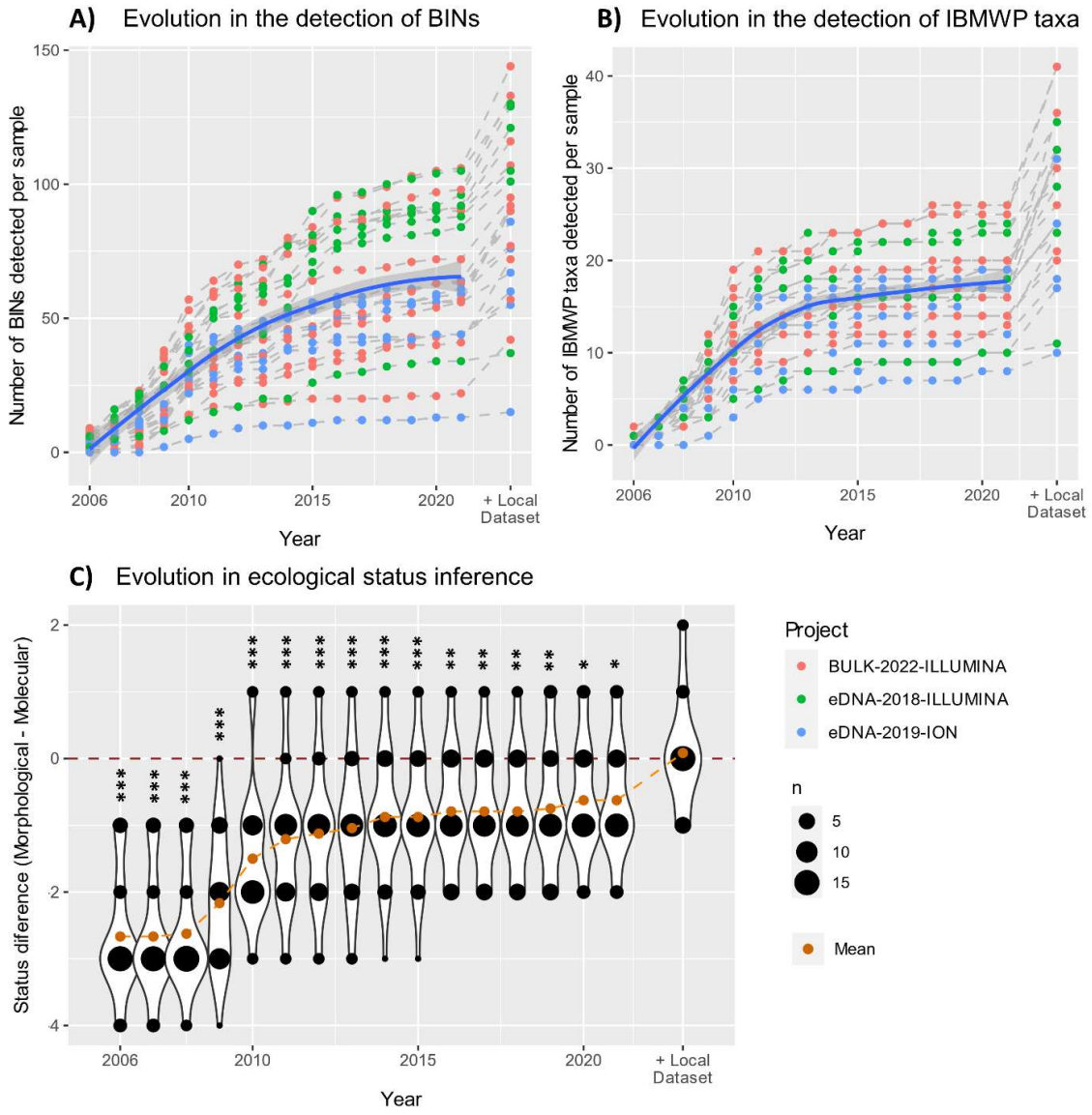



**Figure VI-4** Evolution of the number of BINs (A) and IBMWP taxa (B) molecularly identified using BOLD public records available over the years and including the new macroinvertebrate sequences generated in this study for the northern Iberian Peninsula. Colours indicate the source of each sample from different projects, and the dashed lines link the same sample along the time scale. The blue line represents a smoothed trend line (local polynomial regression, LOEES) that fits the data shown in the figure, and the grey shade is its 95% confidence interval. C) Comparison of the ecological status of 24 samples according to the IBMWP index obtained by the morphological approach and the molecular approach. A zero value means the same ecological status, whereas the status difference is the number of status degrees between approaches. Negative values indicate poorer ecological status in the molecular approach than in the morphological approach, while positive values indicate the opposite. Molecular ecological status inferences were made using the temporally disaggregated public BOLD datasets and were finally improved with the new sequences generated in this study (local dataset). A t-test comparison was carried out to check whether the mean of the distribution was equal to 0 for each of the years, and the p values are shown as asterisks between the following intervals: '*' $p < 0.05$, '**' $p < 0.01$, $p < 0.001$ '***'.

## Discussion

Database incompleteness is a well-known limitation of metabarcoding studies, both for routine biomonitoring and for diversity analyses (Martins, Feio, et al., 2021; Martins,

Galhardo, et al., 2019; Múrria et al., 2020; Weigand et al., 2019). However, only a few studies focusing on freshwater macroinvertebrates have attempted to sequence local species prior to taxonomic assignment (e.g., Elbrecht et al., 2016). The number of sequences available in public repositories is increasing, and more species are being genetically characterised for the genes commonly used in molecular identification (Porter & Hajibabaei, 2018). Nonetheless, the percentage of species already sequenced, as well as the number and size of sequencing projects for new species, varies depending on the geographical area (Behrens-Chapuis et al., 2021; Gaytán et al., 2020; Ge et al., 2021; Weigand et al., 2019). It is therefore worth considering how complete the databases are and how useful barcoding sequencing can be in areas where macroinvertebrate metabarcoding analyses are to be carried out (Csabai et al., 2023).

Two studies have partially verified the completeness of the genetic databases for macroinvertebrate species in the Iberian Peninsula. Weigand et al. (2019) provided an estimate of the sequencing coverage for macroinvertebrate species, among other groups, from several European countries. However, this work only included a species-level checklist from Portugal, comprising a total of 291 species. According to Weigand et al. (2019), 75.2% of their Iberian species list had sequences in reference databases. The high coverage in this case could be explained by the relatively low number of species studied. Múrria et al. (2020) compiled a partial checklist of 3,348 macroinvertebrate species from the Iberian Peninsula and found that only 37.2% of them had sequences for COI-5P in the BOLD and GenBank databases (Múrria et al., 2020). Despite this, they noted that the species sequenced were indeed the most common. Finally, in section V of this thesis, we repeated these analyses by combining the two previously mentioned partial checklists with a third new partial checklist (Schmidt-Kloiber & Hering, 2015) and found that only 53% of the total number of species included in the checklist had been sequenced. In all three studies, a complete checklist of all macroinvertebrate species from the Iberian Peninsula was lacking, as it has not yet been published. This gap prevents a complete analysis of the availability of reference sequences for many macroinvertebrate species of the peninsula, including some relevant IBMWP taxa such as Acariformes or Ostracods. Since these works, the InBIO initiative published sequences of Trichoptera (Pauperio et al., 2023), Plecoptera (S. A. Ferreira et al., 2020) and Diptera (S. A. Ferreira et al., 2020, 2021; Oosterbroek et al., 2020), among other macroinvertebrate groups (S. A. Ferreira et al., 2019; P. Sousa et al., 2021), from across the Iberian Peninsula, though with a prominent representation from northern Portugal.

The results obtained in this study show that, of all the taxa sequenced, 79% had at least one sequence in one of the reference databases. The Ephemeroptera, Plecoptera and Trichoptera orders had high identification percentages of over 90%, in contrast to the IBMWP group with the highest number of morphospecies collected here, Diptera, which had an identification percentage of only 65%. Among those groups with more than 10

morphospecies identified in this study, the group Arachnida, consisting entirely of the superorder Acariformes, had the lowest sequencing rate, with only 3 out of 11 amplified morphospecies being molecularly identified. These results indicate that for the specific case of the northwestern Iberian Peninsula, the major sequencing effort should be focused on Diptera or Acariformes. The globally high percentage of coverage achieved in this work compared to previous studies may be due to two combined factors. First, there may be a bias in the species analysed, as the sampling design likely resulted in the most common species being sampled and sequenced, as discussed in the article by Murria et al. (2020). On the other hand, this study was conducted after a large sequencing effort (InBIO project) for specific groups, which we found to have higher sequencing representativeness.

Regarding the metabarcoding analysis, it is important to acknowledge the limitations of the current study, which lacks a temporal and seasonal framework that could potentially influence the observed results. Nevertheless, our findings unequivocally show that even a slight increase in the average detectability of new taxa exerts a considerable impact on the ecological status assessment when employing molecular data. Notably, all changes observed in ecological status were in favour of improvements. Specifically, the inclusion of only 33 new local species sequences increased the detection of taxa considered by the IBMWP index by an average of 1.54 more per sample, and this small increase, in turn, led to a change in the inference of the ecological status of the river studied in 16.6% of the samples. These results clearly indicate that the sequencing of species not yet included in genetic databases may affect the performance of metabarcoding techniques for routine macroinvertebrate biomonitoring.

In the study of the temporal evolution of taxonomic assignment in metabarcoding samples using public records from BOLD, we observed a fast and significant trend toward higher detectability after the first years of database creation, however this trend gradually slows down to almost asymptotic values in recent years. This asymptotic trend clearly changes with the inclusion of the new sequences of local specimens ("local dataset"), leading to an increase in the detectability of new BINs and IBMWP taxa (35% and 51%, respectively), which had a direct impact on the inference of the ecological status of the rivers. In the temporal study applied to our samples, where we started with very few reference sequences, higher detectability reduced the difference between the ecological status values obtained by the molecular and morphological methods. However, it is expected that further completion of the reference databases will increase the detectability of families. Consequently, the inferred ecological status will, on average, exceed those inferred by morphological methods. This increased detectability will necessitate recalibrating the current indices to the new techniques or directly the need to create new indices to exploit the enhanced taxonomic resolution. Since, as previously discussed, when the detection is limited to the taxonomic level of the IBMWP index (families), the average increase of 17.25 detected BINs per sample only results in an additional 1.54 IBMWP taxa detected per

sample. In the same way, in the new bulk samples, a total of 565 different BINs were detected, but this number was reduced to only 72 taxa if we considered the IBMWP taxa level. This is due to the index's redundancy due to its low taxonomic resolution, which seems to lose much of the new information gained by using the new molecular techniques. Metabarcoding allows for species identification that goes beyond than the classical and redundant family criteria used, such as BMWP-like indices. Future indices may shift towards species rather than families. This shift opens up new possibilities for index design, provided that we improve our understanding of macroinvertebrate species and their specific tolerance to different types of pollution, using tolerance/sensitivity values at the species level rather than just at the family level (Beermann et al., 2018; Juvigny-Khenafou et al., 2021). This could improve the effective use of metabarcoding as a tool for biomonitoring in the coming years.

## Conclusions

The primary conclusion of this study is that sequencing macroinvertebrate species present on the Iberian Peninsula but not yet present in genetic databases can have a significant impact on the performance of metabarcoding techniques for diversity studies and routine biomonitoring and may potentially alter the inferred ecological status of rivers. In the context of a global process of development, validation and implementation of molecular techniques in routine biomonitoring (Blackman et al., 2019a; Blancher et al., 2022; Cordier et al., 2021), these results highlight the relevance of achieving completeness of the genetic databases before applying metabarcoding for routine biomonitoring. Furthermore, the results suggest that conducting a local sequencing effort may be a more effective and faster strategy to increase the sequencing coverage of local species than waiting for the gradual improvement and expansion of reference genetic databases.

Focusing on the Northwest Iberian Peninsula, our findings show that there were still a significant number of macroinvertebrate species used as river water quality indicators for which there were still no reference sequences in the two main genetic databases, such as BOLD and GenBank (approximately 21%). However, this percentage appears to be lower than those previously considered in the literature (25% to 63%). On the other hand, a comprehensive checklist of macroinvertebrates for the peninsula is still lacking, making it difficult to ascertain the exact number of unsequenced species and to conduct a targeted sequencing effort only on them. Therefore, we recommend prioritising the creation of such a checklist and initiating a coordinated effort to sequence any outstanding species.

# VII. Back to life after 50 years: Impact of environmental flow restoration on macroinvertebrate biodiversity and river ecological status. A case study of the Valseco stream, Spain.

*The contents of this chapter are currently under review in the journal Restoration Ecology as: Fueyo, Á; Sánchez, O; Villazán, B; Nicieza, A; Escudero, A; Cordón, J; Cabiedas, S; Fernández, L; Granero-Castro, J; Borrell, Y. (2024) Back to life after 50 years: Impact of Environmental Flow Restoration on Macroinvertebrate Biodiversity and River Ecological Status. A Case Study of the Valseco Stream, Spain. Restoration Ecology.*

The possibility of using molecular techniques to monitor the restoration of the Valseco stream (downstream of the Matalavilla reservoir dam) arose simultaneously with the study of the reference databases described in the previous section. This opportunity allowed us to test, as a pilot project, the use of metabarcoding with bulk samples and eDNA in a real operational environment. However, after analysing the data, the eDNA samples were not included in the restoration study. When designing the eDNA sampling in the Valseco stream, it was decided to take an additional sample of the effluent from the dam, directly from the bottom drains, before it came into contact with the stream. This sample was taken to determine the eDNA contribution of the reservoir water to the stream. Once the results were obtained, it was observed that the effluent from the dam contained eDNA of several macroinvertebrate species of the taxa included in the IBMWP index. This fact prevented the detection of species with eDNA at different points in the stream from being taken as evidence of their presence in the stream, as this eDNA could have originated upstream of the dam. Therefore, the molecular data analysed in this section are all from bulk samples.

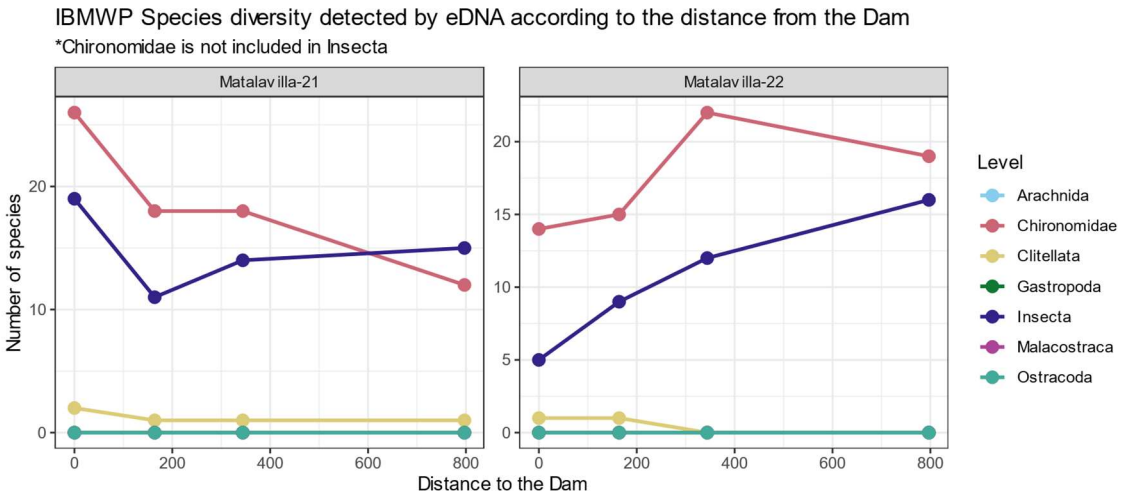



**Figure VII-1** Number of macroinvertebrate species detected by eDNA in the effluent from the Matalavilla dam (distance 0) and from the different sampling points in the Valseco stream.

## Introduction

Freshwater flowing waters cover a tiny fraction of the continental areas but have a disproportionate importance because of its multiple links with other ecosystems and the resources and services they provide to human societies. Water and energy demands have increased dramatically over the last century (Huđek et al., 2020; Kummu et al., 2016) leading to the construction of countless dams, reservoirs, and hydropower plants over the years to secure and regulate water supply and energy production. In consequence, widespread hydroelectric developments have significantly altered the natural river flows (AMBER Consortium, 2020; Belletti et al., 2020; Döll et al., 2009), compromising or directly disrupting many of the ecosystem services provided by lotic environments. Flow regimes support many fundamental ecological processes and functions necessary to maintain healthy rivers, and their artificial alteration threatens the ecological integrity of freshwater ecosystems (Fuller et al. 2015; Jones et al., 2020; van Puijenbroek et al. 2019). The building of hydropower dams has profound effects on stream temperatures, both increasing and decreasing depending on geographic location, size of dam, and time of the year (Ahmad et al. 2021), which can provoke the change of entire faunas (Heggenes et al., 2021; Lessard & Hayes, 2003; Poff & Zimmerman, 2010). Besides these changes in the temperature regimes, flow regulation often prevents the natural disturbance regime in streams and rivers, leading to human-induced progress in the process of ecological succession, and resulting in new "end" states characterized by different ecological communities (Palmer & Ruhi, 2019).

During the last decades of the 20th century, numerous scientific studies reported the impacts of the alteration of the rivers natural flow and provided evidence on the need and benefits of implementing environmental flows (hereafter e-flows) (Tharme, 2003). Subsequently, several countries around the world have introduced strategies in their water legislation to ensure adequate water supplies for natural ecosystems through e-flows

(Acreman & Ferguson, 2010; Mezger et al., 2019). In Spain, implementation of e-flows started in 2013 and 76.9% of rivers had minimum e-flows established by 2021 (ARM 2656/2008; RD 638/2016; RD 1159/2021; MITECO, 2007, 2022). However, an e-flow is not simply the implementation of a minimum flow rate. The definition of e-flows in European legislation (Water Framework Directive WFD) includes that they must contribute, among other aims, to the achievement of good ecological status of rivers (European commission, 2016). Unfortunately, the relationship between the implementation of new e-flows and the ecological response they generate in rivers is not being widely studied in some countries (Mezger et al. 2019, 2021).

The new legal requirements have prompted many new cases for the implementation of minimum e-flows in recent years. One of the most extreme examples of e-flows restoration is the case of the stream Valseco (León, Spain) in 2021. The flow of the last 900 metres of the stream was completely cut off by the construction of the Matalavilla dam in 1967 and was not restored until January 2021 when the environmental flow was implemented. It is known that implementation of e-flows can help to restore lost biodiversity in regulated rivers (Brooks et al., 2011; Growns, 2016; Mackie et al., 2013). However, the Valseco stream is a very suitable scenario to evaluate the recovery potential of freshwater communities after long-lasting, severe perturbations, as the water flow was completely absent and only restored after five decades. Characterization of the recolonization process, along with concurrent changes in species diversity and stream ecological status, can provide many clues for the implementation of minimum flow strategies.

One of the best approaches to assess the biological integrity of rivers and their ecological status is based on the analysis of the benthic macroinvertebrates (hereafter 'macroinvertebrates'), a group composed of several bioindicator species of annelids, flatworms, molluscs, and arthropods (Hynes 1959; Metcalfe 1989; Zamora-Múñoz & Alba-Tercedor 1996; Khamis et al. 2014; Crabot et al. 2021). Routine macroinvertebrate biomonitoring is based on morphological identification of the different indicator macroinvertebrate taxa at different high-ranking taxonomic levels. Nevertheless, routine monitoring methods based on morphological identification have several limitations. One of the main ones is its low taxonomic resolution. DNA-based methods such as metabarcoding can overcome this limitation and allow macroinvertebrates to be identified at more specific levels (Ji et al., 2022; Vourka et al., 2023a). However, these molecular techniques are not yet standardised, and much research is still ongoing to improve their performance and accuracy (Vourka et al., 2023a). For example, metabarcoding is sensitive to the lack of reference sequences in genetic databases, although it is estimated that for the northwest of the Iberian Peninsula approximately 79% of the species are sequenced (Fueyo, Sánchez, et al., 2024). But certainly, the main limitation of metabarcoding is not providing accurate quantification data (Shelton et al., 2023).

Previous research has reported that aquatic invertebrate communities typically show a relatively high level of resilience following drought events, especially those communities adapted to flow intermittency (Bogan et al., 2013, 2015; Datry et al., 2014; Schriever et al., 2015). These studies have considered as 'severe' drying events from 1-3 months (Vorste et al., 2016) up to 15 months (Bogan et al., 2013) and, consequently, their results might not be directly extrapolated to the case of large disturbance sustained over decades. Therefore, restoration of e-flows in streams affected by dam construction and severe flow restrictions, like the Valseco stream, offers the opportunity to 1) study the potential of recovery of small lotic systems after long periods of intense perturbation and 2) evaluate the impact of flow resumption on the ecological status of previously disturbed streams and their biodiversity.

Here, we provide baseline information regarding the effects of e-flow restoration on the community of aquatic invertebrates in a river of the Cantabrian Mountains and more broadly in small temperate streams. We will combine morphological identification techniques and a DNA-based approach to explore changes in macroinvertebrate diversity and the ecological status of the river after resumption of permanent surface flow, the natural state of the study stream. Specifically, we set out to answer two specific questions. First, do macroinvertebrate communities affected by long-lasting disturbances retain the resilience to quickly reach the ecological state of nearby, unaffected communities? To this end, we test the null hypotheses that there are no differences between affected and unaffected sections at the middle and at the end of the study period (12 and 24 months from e-flow recovery) in terms of richness, diversity, and measurements of freshwater ecological integrity. Second, do the results derived from morphological and DNA-based identification lead to consistent inferences? Morphological and DNA-based assessments differ in terms of quantitative and taxonomic resolution. In contrast to morphological assessment, the DNA-based approach used here lacked the potential to provide quantitative measures of abundance. However, it gives a higher taxonomic resolution. Our initial prediction is that OTUs and species composition may show differences in diversity between the different points that are overshadowed at the family level.

# Materials and methods

## Stream characteristics and flow regimes

The stream Valseco is a 9.29 km long stream located in León (Northern Spain; Figure VII-2). The catchment (49.23 $km^2$) has an orientation W-E, with maximum elevation around 2100 m and an average channel slope of 3.36%. This stream flowed uninterruptedly from its source to the confluence with Salentinos until 1967, when the construction of the Matalavilla reservoir was completed, cutting off its flow and restricting it almost completely in the last 900 metres of its course. The final stretch of the stream remained practically dry for more than 50 years, until the beginning of 2021, when the e-flow was restored after installation of two bottom valves in the Matalavilla dam. The established minimum e-flow

rates were 0.24 $m^3$/s in October-December, 0.36 $m^3$/s in January-March, 0.29 $m^3$/s in April-June and 0.18 $m^3$/s in July-September (MITECO, 2023). To have a reference of what would be expected in the river Valseco under current conditions but in the absence of the dam, sampling was conducted also in river Salentinos. This is a small stream (7.04 km) which drains an adjacent, similar catchment (26.76 km2; orientation: W-E; maximum elevations: 2100 m; average channel slope: 4.67 %). Downstream the confluence, at 850 m a.s.l., they form Las Vegas River, also named Salentinos II, which is a tributary of the River Sil. Though flows were historically higher in the Valseco than in the Salentinos, the situation reversed, even after re-implantation of e-flows. In addition, these sub-basins share climatic conditions (AEMET 2024) and have similar geological bedrocks (IGME 2024). Therefore, Salentinos can provide a conservative reference for the ecological status of the Valseco stream.

## Sampling and processing

We collected aquatic invertebrates periodically for 2 years (2021 and 2022) after the implementation of e-flows in Arroyo de Valseco in January of 2021. First, we established 3 sampling points distributed along a 900-meter stretch of the Valseco stream hat runs from the Matalavilla reservoir to its confluence with the Salentinos River (VALS1, VALS2 and VALS3; Figure VII-2). A change in the river channel during the first year of the study left VALS2 and VALS3 out of the stream and forced these points to be repositioned: VALS2BIS was located 300 m upstream the original VALS2, and VALS3BIS was repositioned 150 m downstream VALS3, very close to the confluence with the river Salentinos (Figure VII-2). In Arroyo Valseco, we collected macroinvertebrate samples monthly during the first year, and every 6 months during the second year. At the end of each year, an additional sample was collected at each point for DNA metabarcoding analysis. In addition, we set 3 sampling points as control stretches: two points (SA1, SA2) were located in river Salentinos 10 m and 2.3 km upstream the confluence of Arroyo Valseco, and the third (SA3) 2.5 km downstream the confluence (N42°49'44'', W6°27'36'', UTM30; Figure VII-2). We sampled these 3 points one month before and after the opening of the dam bottom valves (we obtained three more samples in SA2 during the April, May and June of 2021), and therefore they provided both a spatial reference for Arroyo de Valseco and a test for the effect of the implementation of e-flow downstream of the Valseco outlet (BACI design).

Macroinvertebrate samples for morphological identification were collected and processed according to official Spanish protocols (MAGRAMA, 2013a, 2013b, 2016). Samples for DNA metabarcoding analysis were collected and processed according to Fueyo et al. (2024b). Samples were homogenised and DNA extracted and amplified using BR2/BF3 primers (Elbrecht et al., 2019; Elbrecht & Leese, 2017b). Libraries were sequenced on Illumina Miseq aiming 100.000 reads per replicate. Only taxa present in the three PCR replicates were retained and reads from operational taxonomic units (OTUs) present in negative

controls were subtracted from the samples. The resulting taxa list was filtered to retain only those taxa included in the official ecological status indices.

## Biodiversity and Ecological status indices

The presence/absence and abundance data of macroinvertebrate taxa obtained by morphological identification was used to calculate both Shannon's alpha diversity index and the ecological status indices IBMWP and METI (MAGRAMA, 2013a, 2015).The IBMWP index is a qualitative and cumulative index, i.e. it only takes into account the presence/absence of different taxa and the index value increases with each new taxon detected. However, not all taxa contribute equally when detected in the sample, as the IBMWP index assigns different values to taxa in relation to their sensitivity/tolerance to anthropogenic impacts. On the other hand, the METI index is a quantitative and multi-metric index, i.e. it takes into account the abundance of the detected taxa and the final ecological status value is based on the calculation and weighting of different metrics. Both indices have reference values for each river typology set out in the legislation (MAGRAMA, 2015a).

## Statistical analysis

All data analyses were performed in R 4.3 (R Core Team, 2024). We used the 'vegan 2.6-4' package to conduct permutational multivariate analysis of variance (PERMANOVA) to determine whether the macroinvertebrate composition in the Valseco stream was related to sampling location and time after the implantation of e-flow. The abundances of the macroinvertebrate families were logarithmically transformed before analysis. A non-metric multidimensional scaling (NMDS) analysis was performed, grouping the samples from the Valseco stream by sampling site. In addition, to know the state of the recovery of the river Valseco after the restoration of the flow we compared the values obtained in December of 2021 and 2022 at Valseco stream with those obtained from November 2020 to January 2021 at the Salentinos river (considered here as the reference river). We used taxonomic richness (number of families), abundance (total number of individuals) and taxonomic diversity (Shannon index) to compare the structure of macroinvertebrate communities, and the EQR values of the IBMWP and the METI indices to evaluate the ecological status of the river. Shapiro-Wilk tests were performed to check the normality of the variables, and as they did not match normal distributions, Mann-Whitney U test analyses were performed for each variable.


**Figure VII-2** Detail of the three main sampling points in the Valseco stream and the hyporheic river section separating and disconnecting VALS 1 and VALS2BIS from VALS3BIS and the Salentinos river.

# Results

## Macroinvertebrate Diversity

The diversity of macroinvertebrate families, calculated by the Shannon index, (H) ranged between 1.5 and 2.5 along the Salentinos river (Figure VII-5A). During winter month, Shannon index ranged from 1.45 to 2.15 for SA1 and SA3. In the middle section (SA2), diversity was slightly higher (H = 2.25 – 2.40) and these values were maintained throughout spring and summer (Figure VII-5A). In the affected section in river Valseco, the diversity values of VALS1 and VALS2 were 1.1 and 0.7 at the initial sampling, then they decreased to a minimum (H =0.5) in March and April 2021 along with those observed in VALS3. Subsequently, the values of VALS1 gradually increase approaching the values of the reference river in October 2021. VALS3BIS showed a low initial diversity in March and then increased to steady values around 1.7 - 2.15 until the end of the study. In parallel, VALS2BIS diversity showed a minimum at the first sampling in May 2021 (H = 1.0) and then increased progressively to reach the values of the reference river (Figure VII-5A). The NMDS analysis (Figure VII-5B), revealed a high degree of similarity in the macroinvertebrate composition of the samples collected in the Valseco stream. Specifically, VALS1 and VALS2BIS cluster together while VALS3BIS approached the composition of Salentinos river (SA2). The analysis revealed a clear trend in temporal variation along the NMDS1 axis. The community composition of the Valseco stream samples gradually approached that of the Salentino stream samples (SA2). PERMANOVA analysis showed a significant effect of sampling date (Table VII-1) evidencing a clear pattern of temporal variation (Figure VII-5B), but also a spatial effect resulting from differences among sampling points within the Valseco stream. On average, taxonomic richness was about 40% lower in the Valseco in 2021 (approx. 15 families) than in the Salentinos river (≈25 families). However, richness in Valseco raised to 23 families in 2022, only about 8% lower than the reference samples in the Salentinos river. In terms of total abundance, in Salentinos river, there were on average 4525 ind/$m^2$, while in Valseco Stream, this value was 2807 indv/m2 in December 2021 and 675 ind/m2 in December 2022 (Figure XVI-1 and Figure XVI-2 in Annex III).

**Table VII-1** PERMANOVA analysis in the relation on family macroinvertebrate composition with sampling point and date. P-value: ≤0.05 * ≤ 0.01 ** ≤ 0.001 ***.

| | Df | SumOfSqs | $R^2$ | F | p-value |
|---|---|---|---|---|---|
| **Sampling point** | 2 | 0.4777 | 0.13103 | 4.409 | *** |
| **Sampling date** | 13 | 1.9536 | 0.53585 | 2.774 | *** |
| **Residual** | 18 | 0.9751 | 0.26746 | | |
| **Total** | 33 | 3.66458 | 1 | | |

## OTU and species diversity

Molecular analyses of macroinvertebrate OTU and species diversity (Fig. 3) showed values between 120 and 150 OTUs and 60 and 125 species per point in the stream Valseco in 2021.

It remains similar for VALS1 and VALS2BIS during the two years but increases in 2022 to more than 225 OTUs and 125 species in the case of VALS3BIS, mainly due to an increase in the diversity of the family Chironomidae (order Diptera). Overall, all the three sites in Valseco showed a decrease in oligochaete diversity between 2021 and 2022. At the same time, they showed a slight increase in the number of OTUs and species of Plecloptera, Thricoptera and Ephemoroptera (EPT group), whose families are classified as sensitive in the ecological status indices.

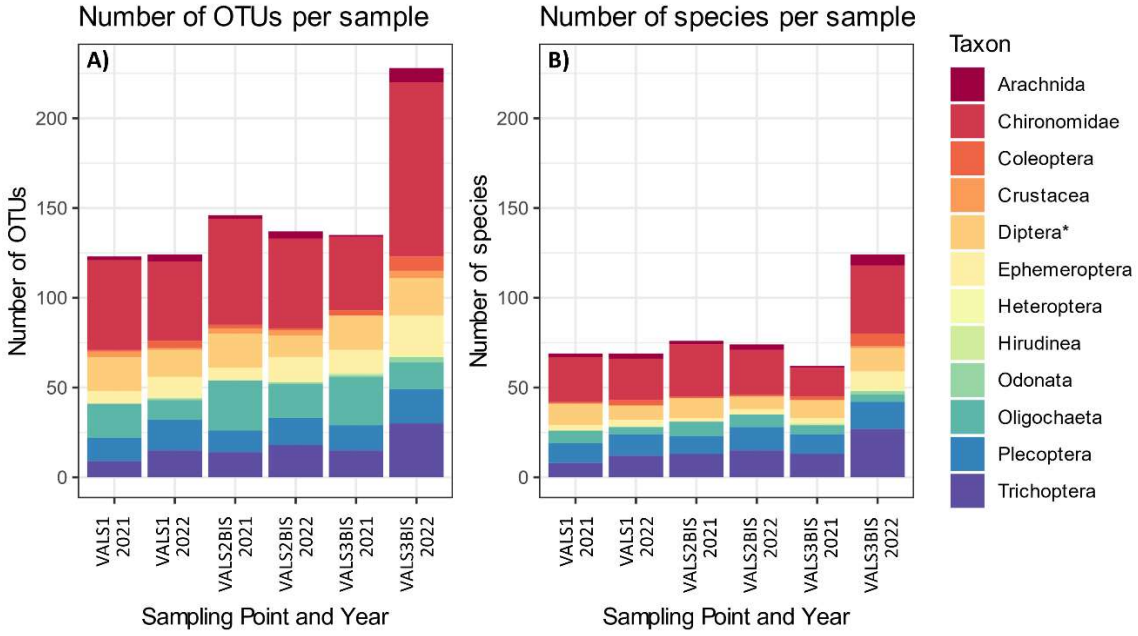



**Figure VII-3** Composition of OTUs (A) and species (B) in the three sampling points of the Valseco stream sampled in December 2021 (1 year after the restoration of the e-flow) and December 2022. *Excluding Chironomidae.

## Morphological vs Molecular identification

Considering detections by both methods together, an average of 34.7 IBMWP taxa were detected per sampling point (Figure IX-2). Of these, 53.4% were detected by only one of the two identification methods. Specifically, 44.2% of detections were made by molecular detection alone, while only 8.2% of detections were made by morphological identification alone.

The similarity of the composition of the IBMWP taxa obtained by molecular and morphological identification is 0.474 (Jaccard similarity score, Figure VII-4). If only the taxa identified by morphological identification are considered, the similarity score is 0.784, but if only those with more than 50 individuals are considered, the similarity score rises to 1, meaning the same composition at all points.

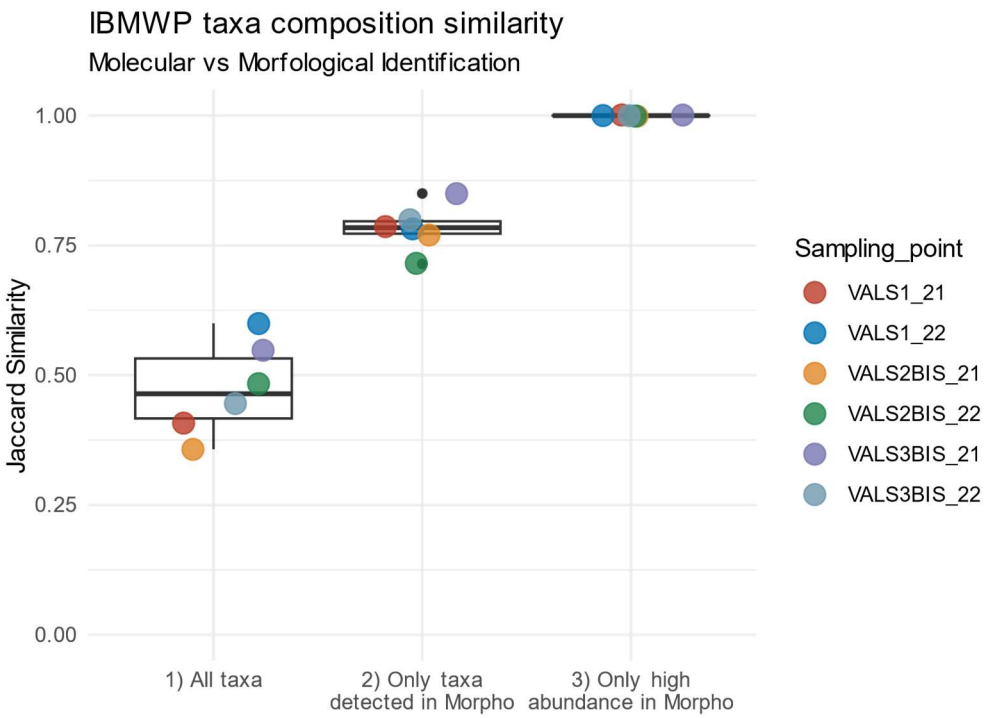



**Figure VII-4** Jaccard similarity of composition of IBMWP taxa between molecular and morphological identification. 1) Include all taxa. 2) Include only taxa detected by morphological identification. 3) Include only taxa detected by morphological identification with an abundance greater than 50 individuals per sample.

## Ecological Status

The ecological status of the Salentinos river ranged from 'good' to 'very good' over time according to the IBMWP and METI indices (Figure VII-5C and Figure VII-5A) Shannon Diversity Index for each of the samples taken over the entire time series of the study. Each of the points corresponds to the value of one sample, the samples of the same sampling point are connected by lines of the same colour. The horizontal dashed lines indicate the limits of interpretation of the diversity values. B) NDMS of macroinvertebrate taxa composition. Only sampling points with more than 4 samples are plotted. Stress = 0.159. C) D) Deviation from the reference values (EQR; Ecological Quality Ratio) for the C) IBMWP and D) METI indices of each of the samples taken throughout the time series: A) Valseco stream and Salentinos river; each of the points corresponds to the value of one sample, the samples of the same sampling point are connected by lines of the same colour. The horizontal dashed lines indicate the limits of change in ecological status for a given EQR value Figure VII-5D). For the entire study period, the ecological status in the Valseco stream improved progressively. VALS1 and VALS2BIS showed similar trends from an initial 'bad'/'poor' ecological status to a 'good'/'very good' status by December 2022. In turn, VALS3BIS shifted from a 'moderate' ecological status in March 2021 to a 'good' status in July 2021in both indices. VALS3BIS also approached and achieved a 'very good' status in December 2022 for the IBMWP and METI index respectively. The EQR values of the points from Valseco stream (VALS1, VALS2BIS and VALS3BIS) for IBMWP index, were significantly lower than those of the reference river at the end of the first year but they did not differ at

the end of the second year (Table VII-2). On the other hand, the EQR values for the METI index did not differ significantly neither at the end of the first or the second year.

**Table VII-2** Mann-Whitney U test one-way ANOVA analysis of the number of taza, abundance, EQR and Shannon diversity values of the Salentinos samples compared to the Valseco river results one and two years after e-flow restoration.

| | **Reference vs Valseco 2021** | | **Reference vs Valseco 2022** | |
|---|---|---|---|---|
| | W | P-value | W | P-value |
| **Number of taxa** | **18** | **0.03** | 13.5 | 0.30 |
| **Abundance** | 10 | 0.52 | **18** | **0.03** |
| **Shannon diversity** | 16 | 0.09 | 5 | 0.37 |
| **EQR IBMWP** | **18** | **0.03** | 16 | 0.09 |
| **EQR METI** | 15 | 0.16 | 9 | 1 |

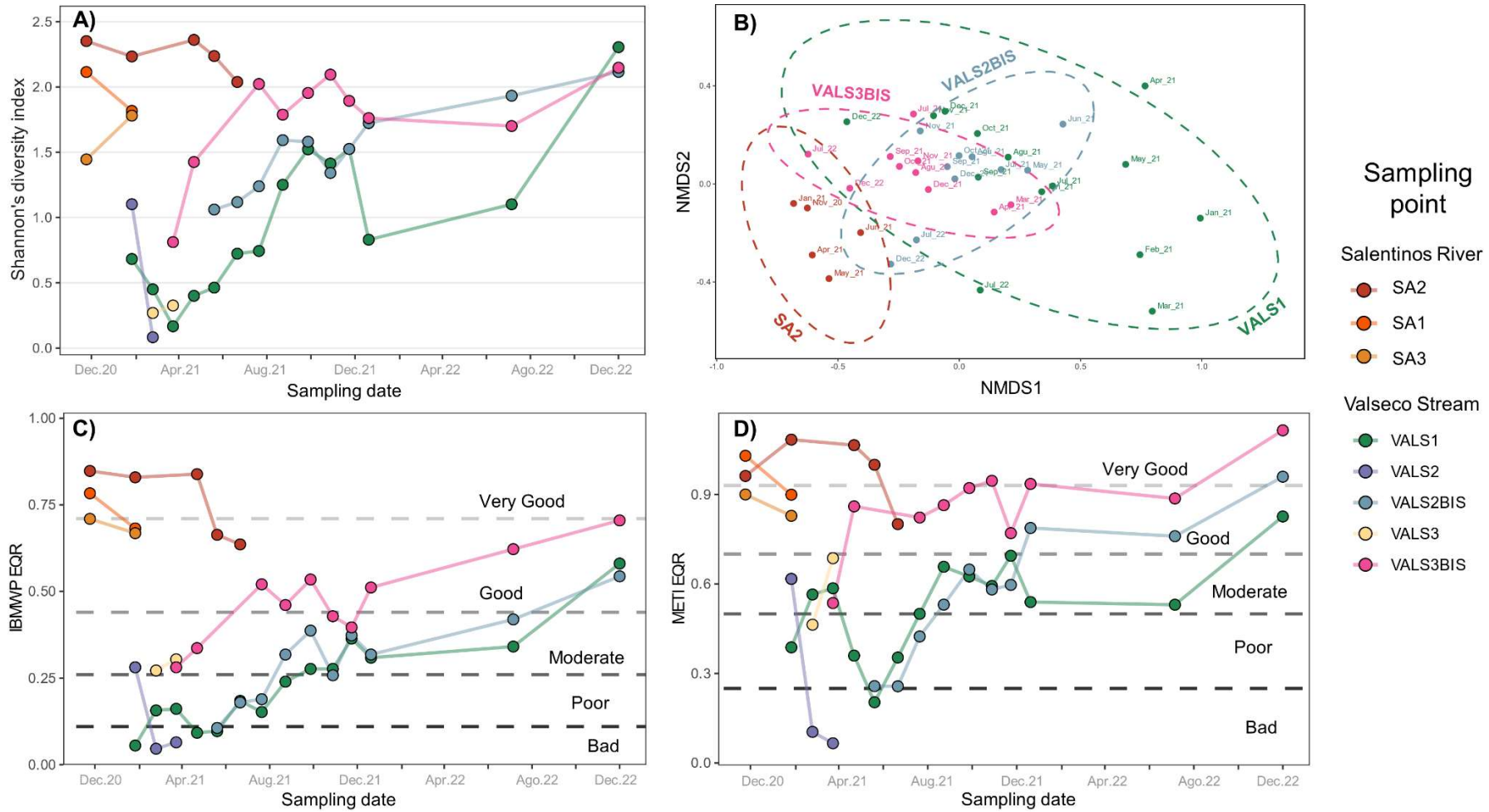



**Figure VII-5** A) Shannon Diversity Index for each of the samples taken over the entire time series of the study. Each of the points corresponds to the value of one sample, the samples of the same sampling point are connected by lines of the same colour. The horizontal dashed lines indicate the limits of interpretation of the diversity values. B) NDMS of macroinvertebrate taxa composition. Only sampling points with more than 4 samples are plotted. Stress = 0.159. C) D) Deviation from the reference values (EQR; Ecological Quality Ratio) for the C) IBMWP and D) METI indices of each of the samples taken throughout the time series: A) Valseco stream and Salentinos river; each of the points corresponds to the value of one sample, the samples of the same sampling point are connected by lines of the same colour. The horizontal dashed lines indicate the limits of change in ecological status for a given EQR value.

## Discussion

The implementation of e-flows in the lower river Valseco had an important positive impact in the restoration of the macroinvertebrate biodiversity. In comparison with a reference river exposed to similar constraints and environmental and conditions, the data revealed a rapid recovery of the diversity of the macroinvertebrate community in river Valseco, along with the attainment of a favourable ecological status. Broadly speaking, the results and conclusions derived from morphological and molecular approaches were consistent.

The main effects of flow alterations on macroinvertebrate communities have been studied for years, with these alterations causing a generalized decrease in the biodiversity and abundance of macroinvertebrate families (Dewson et al. 2007; Vaikasas et al. 2013; Mbaka & Wanjiru Mwaniki 2015). In some cases, it has also been studied how the implementation of e-flows can restore some of the biodiversity lost through the construction of dams and flow regulation (Brooks et al. 2011; Mackie et al. 2013; Growns 2016). However, our study system represents an extreme case where the water flow was completely cut off and restored after five decades. The return to life of the Valseco stream may have some similarity with the recovery of flow in intermittent rivers and ephemeral streams (IRES) where the surface water flow is restored after months of drought when the rainy season arrives (Stubbington et al. 2017). However, more rapid recolonisation is expected in the IRES, as macroinvertebrates can take refuge in pools or other temporal refuges after the river has dried out (Paltridge et al., 1997; Chester & Robson 2011). However, these 'dormant' communities are more unlikely to persist during prolonged droughts or permanent drying of the surface, since both the taxonomic and functional richness of dry riverbed communities decrease as the duration of the dry phase increases (Pařil et al. 2019). Even so, in the literature we can find cases of studies of recolonisation of IRES after droughts of several months in which there were no disconnected pools left. For example, Di Sabatino et al. (2023) studied the initial stages of recolonisation of a river that was dry during the summer and autumn months of 2021. Although they found the presence of eight different macroinvertebrate taxa the day after flow recovery and 11 taxa after 12 days, these values remained almost constant over the 5-month follow-up study, which is about 50% of total taxa diversity before desiccation and three times less for EPT taxa. In a similar study, Doretto et al. (2020) reported full recovery of the macroinvertebrate diversity after three months of flow restoration in an intermittent section compared to a perennial section of the same river. Studies carried out in other areas corroborate the high potential of IRES for rapid recovery of macroinvertebrate diversity and abundance after the dry season (Bogan et al. 2015; Dolédec et al. 2017; Hill et al. 2019). However, it should be noted that the composition of post-drought communities often remained different from that of pre-drought communities more than a year after the previous drought episode (Datry et al., 2014; Di Sabatino et al., 2023; Hill et al., 2019). This may suggest that these lotic communities may have multiple

stable points (Aguadé-Gorgorió et al., 2024; Sutherland, 1974) or simply that their composition can be influenced by interannual variation.

Between December 2020 (one month before the establishment of environmental flows) and July 2021, all but one of the estimates of diversity for the Salentinos river were within the range 1.7 – 2.50. In SA2 (the closest point upstream of the confluence of rivers) the diversity fluctuated between H = 2.25 and H = 2.40 until May, and then approached H = 2.00 in June. Therefore, H values between 2.00 and 2.50 can be used as a reference for ecological recovery and as a target to restore river functionality in the affected section.

The first sampling in the Valseco stream was carried out one month after the reestablishment of the environmental flow. At that time, we detected only 4 families in VALS1, a richness well below those often reported for IRES, while in VALS2 we found a total of 13 families, more in accordance with the dynamics observed in intermittent and ephemeral rivers. In March 2021, two months after restoration of environmental flow, family diversity remained very low at all sampling sites in Valseco ($H < 0.5$), except in VALS3BIS (Fig. 2A). From April onwards, diversity increased progressively at all the study sections and by the end of the first year (months 9-11) VALS1 and VALS2BIS reached diversity levels of $H \approx 1.50$ whereas VALS3BIS oscillated around $H \approx 2.00$. One year later the three sites showed values in the target range. As evidenced by NMDS and PERMANOVA analyses, the composition of Valseco fauna changed progressively and showed a clear convergence with the Salentinos community (especially in the case of VAL3BIS). Taken together, our results indicated that the affected section was less responsive to the recovery of surface water flow than a typical intermittent river, but the macroinvertebrate community fully recovered in terms of alpha diversity within two years from the flow restoration.

Colonization of newly created rivers as natural fishways and rewilding programs may provide a very suitable framework for assessing responses to flow recovery in long-term dry reaches. For example, Gustafsson et al. (2013) found that after two years, only 63 % of the benthic fauna families identified in the reference streams had colonized the new created fishway. Families exclusive to the reference streams and absent in the biocanal were mostly slow colonizers or associated with riparian vegetation, which was sparse in the biocanal. Macroinvertebrate recolonization of Valseco stream is not as fast as in the IRES, where resistance and resilience factors may come into play more strongly to facilitate recolonization (Van Looy et al. 2019). However, the recovery of the macroinvertebrate community in Valseco seemingly was better and faster than that of an artificial channel. This could be due to the fact that the Valseco stream flows in its former natural channel, with abundant vegetation on the banks and different types of habitats and substrates along its course, allowing it to host more diversity (Gustafsson et al. 2013).

Changes in the ecological status of the affected sections broadly paralleled changes in diversity (Fig. 2C-D): an initial progressive enhancement during the first year, when these

sections achieved levels of moderate or good status (very good in METI for VALS3BIS), which continued over the second period to achieve (METI) or approach (IBMWP) the reference status observed in river Salentinos (good/very good). We found that VALS3BIS, which is closer to the Salentinos river and therefore has a greater connectivity, recovered good ecological status faster than VALS1 and VALS2BIS. At the end of the second year these upper sections reached good ecological status and similar values of taxa (family) diversity to those of VALS3BIS, but OTU and species diversity of VALS3BIS was about twice that of VALS1 and VALS2BIS. These differences in recovery rates and diversity at OTU and species levels could arise from the proximity of VALS3BIS and the Salentinos River, but we cannot rule out that they could be due to differences in microhabitat complexity in these sections.

The comparison of molecular and morphological identification methods shows that the molecular identification technique can identify more IBMWP taxa at all points, as reported in other studies (Emmons et al., 2023; Kuntke et al., 2020). On the other hand, there are still taxa that were only detected by morphological examination. This may suggest some limitations of molecular identification, such as lack of sequences in the species genetic databases or primer (see section VI). However, it could also be due to sampling bias, as two independent samples were taken at each site and those taxa that were low in abundance at that site may have been captured in one sample and not the other. This last hypothesis is supported by the fact that the molecular approach can identify all those taxa that were detected in high abundance by the morphological approach. In any case, molecular techniques allowed us to obtain a higher taxonomic resolution, with some implications in the evaluation of the effectiveness of environmental flow restoration. While family-level diversity reached the reference values by the end of the second year, species diversity and OTU diversity did not. This suggests that the molecular approach may be more conservative, but the choice of one or another technique may depend on the importance we attribute to ecological and functional redundancies, which should be higher at the species level.

The differences in the number of OTUs and species for the same sampling point and year are mainly due to two reasons. Previous studies on marine macroinvertebrates have shown that OTUs diversity does not correspond to the total number of species at the site, as it is inflated by about 21% due to unintentional sequencing of NUMTs (Nuclear copies of Mitochondrial genes) (Schultz & Hebert, 2022). On the other hand, we know that the number of detected species is underestimated due to several factors, mainly the lack of reference sequences in the databases (Csabai et al., 2023). Therefore, the real number of species is likely an intermediate value.

In summary, our results support the idea that macroinvertebrate communities affected by long-lasting disturbances retain the resilience to quickly reach the ecological state of nearby, unaffected communities. In addition, this study shows that the implementation of e-flows is an effective management tool for river restoration that can help achieve the

objectives of the WFD. The results also support the role of the Salentinos river as a reservoir of macroinvertebrate biodiversity, facilitator of recovery and element of ecological resilience. This must be considered before extrapolation of our conclusions to other systems lacking this particular configuration. Finally, according to our initial prediction, the higher taxonomic resolution of the molecular approach may help uncover differences in diversity and community composition that are overshadowed at the family level, thus providing a more conservative tool for decision making.

# VIII. Unlocking rivers hidden diversity and ecological status using DNA metabarcoding in northwest Spain.

*Part of the content of this chapter is published as: Fueyo, Á., Sánchez, O., Carleos, C., Escudero, A., Cordón, J., Granero-Castro, J., & Borrell, Y. J. (2024). Unlocking rivers' hidden diversity and ecological status using DNA metabarcoding in Northwest Spain. Ecology and Evolution, 14(8). https://doi.org/10.1002/ece3.70110*

In the context of river ecosystems, the application of metabarcoding provide a scalable and highly sensitive approach for assessing, not only macrozoobenthic, but also global freshwater river diversity. Moreover, molecular species identification overcomes the challenges associated with the time and cost limitations of morphological identification. This species-level identification potentially allows the data obtained to also be used for detecting species of interest, whether they are exotic and invasive species or endangered and protected species, without further effort or cost. Therefore, this multitask approach would increases the suitability of metabarcoding for routine river biomonitoring.

This section aims firstly to apply metabarcoding of bulk and eDNA samples for a comprehensive assessment of macrozoobenthic-specific biodiversity and ecological status of rivers in northwestern Spain. In addition, this multitask approach would also help to provide valuable insights into the presence of other species relevant to river management (i.e. invasive and/or protected species), thus broadening the spectrum of metabarcoding possibilities for future biomonitoring strategies in Spain in a more holistic and effective framework.

## Materials and methods

### Sampling

A total of 27 sites of the NW of the Iberian Peninsula were sampled for this study, taking in parallel kicksamples for morphological analysis and bulk samples for metabarcoding. At 16 of these sites, water samples were also collected for eDNA metabarcoding analysis (Figure VIII-1). Kicksamples and bulk samples were collected following the sampling protocol of the Spanish Ministry of the Environment (MAGRAMA, 2013b) which consist of a semiquantitative, stratified and multihabitat sampling scheme with a 500 µm pore Surber net and then preserved in 96% ethanol with 0,01% BAC (benzalkonium chloride). All materials contacting bulk samples were sterilized before each sampling with ten times diluted commercial bleach (final concentration of 0.4% Cl) and washed twice with distilled water. For eDNA sampling, a transect was set up perpendicular to the river, with 5 one-liter water samples collected across the width of the river and taken close to the river bottom.

Samples from each point were pooled in a 5-litre bottle to which 1ml of 50% BAC was added and stored, no more than 24 hours, at 4 degrees until extraction (Jo et al., 2021; Takahara et al., 2020; Yamanaka et al., 2017).

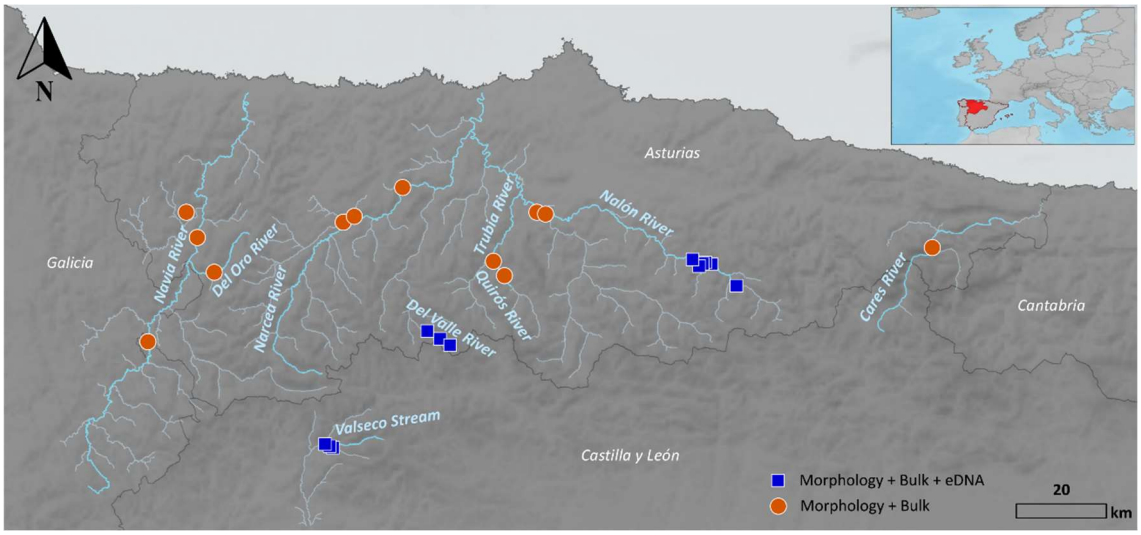



**Figure VIII-1**. Map showing sampling locations in rivers from northwest Spain. The colours and shapes vary depending on the sampling technique applied at each locality: Orange circles: Morphological identification and bulk metabarcoding; Blue Squares: Morphological identification, bulk metabarcoding and eDNA metabarcoding.

## DNA Extractions

Kicksamples for morphological identification were processed according to the Spanish official protocol (MAGRAMA, 2013a). Briefly, it consists of a sieving process that separates the sample into 3 sieves (5 mm, 1 mm and 0.5 mm pore size), then a portion of the individuals present in each sieve is randomly sub-sampled and a minimum of 100 individuals from each sieve are identified under the magnifying glass. Only the taxonomic groups listed in the protocol are considered for identification. In terms of resolution, the identification of macroinvertebrate taxa is multilevel because it includes identifications at different taxonomic levels, with the family level being the most common (1 genus, 121 families, 1 superorder, 1 subclass, 1 class).

Bulk samples were elutriated with distilled water to remove stones and sand and then homogenised according to the protocol of Buchner et al. (2021) consisting of grinding the elutriated sample in a blender, ensuring an effective decontamination between the different samples. Homogenised bulk samples were extracted twice (10 g per replicate), in addition to two negative controls, using the MaxPowerSoil Kit (Qiagen) following the manufacturer's instructions.

Water eDNA samples were filtered through a 0.45 µm CN filter in triplicate (1.25 L each) using a peristaltic bomb. In addition, three negative filtration controls of 1.25 L each of distilled water were filtered. Each filter was stored in 600 µL of ATL buffer (Qiagen) and subsequently extracted with the Blood and Tissue kit (Qiagen). Adjusting the volumes of AL

buffer and ethanol to 600 µL finally eluted with 100 µL of elution buffer. The elution was processed through the Zymo inhibitor removal kit following the manufacturer's instructions.

## DNA metabarcoding libraries preparations and sequencing

Two different library sets were constructed for different sets of samples using two different primer pairs and conditions in order to specifically detect the macroinvertebrates present in the samples. Libraries were all prepared in triplicate. For the bulk samples library preparation, a fragment of the mitochondrial COI gene of around 460 bp (including primers) was amplified using the following primers: Forward - BF3 (5' CCHGAYATRGCHTTYCCHCG 3') (Elbrecht et al., 2019) and Reverse - BR2 (5' TCDGGRTGNCCRAARAAYCA 3') (Elbrecht & Leese, 2017b). These primers also included the universal Illumina ligation sequences attached to their 5' ends. Forward universal tail (5' ACACTCTTTCCCTACACGACGCTCTTCCGATCT 3') and reverse universal tail (5' GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT 3'). A variable number of nucleotides (3 to 6) were inserted between the primer and the universal Illumina ligation sequences to increase sequence diversity, which in turn leads to better results on Illumina machines and allows for a reduced spike-in of ~5% PhiX (Elbrecht & Steinke, 2019; Wu et al., 2015). In the first amplification step, PCRs were carried out in a final volume of 25 µL, containing 2.5 µL of template DNA, 0.2 µM of the primers, 12.5 µL of 2x Multiplex PCR Master Mix (QIAGEN Multiplex PCR Plus kit), and ultrapure water up to 25 µL. The reaction mixture was incubated as follows: an initial denaturation step at 95 ºC for 5 min, followed by 30 cycles of 95 ºC for 30 s, 50 ºC for 90 s, 72 ºC for 30 s, and a final extension step at 68 ºC for 10 min. PCR products were purified using the Mag-Bind RXNPure Plus magnetic beads (Omega Biotek), following the instructions provided by the manufacturer.

For the eDNA library preparation, a fragment of the mitochondrial COI gene of around 191 bp (including primers) was amplified using the following primers: Forward - fwhF2 (5' GGDACWGGWTGAACWGTWTAYCCHCC 3') (Vamos et al., 2017) Reverse - EPTDr2n (5' CAAACAAATARDGGTATTCGDTY 3') (Leese et al., 2020). These primers also included the Illumina ligation primer sequences attached to their 5' ends and a variable number of nucleotides (4 to 6) in order to increase sequence diversity. In the first amplification step, PCRs were carried out in a final volume of 25 µL, containing 2.5 µL of template DNA, 0.3 µM of the primers, 12.5 µL of 2x Multiplex PCR Master Mix (QIAGEN Multiplex PCR Plus kit), and ultrapure water up to 25 µL. The reaction mixture was incubated as follows: an initial denaturation step at 95 ºC for 5 min, followed by 35 cycles of 95 ºC for 30 s, 50 ºC for 90 s, 72 ºC for 30 s, and a final extension step at 68 ºC for 10 min. PCR products were purified using the Mag-Bind RXNPure Plus magnetic beads (Omega Biotek), following the instructions provided by the manufacturer.

The oligonucleotide indices which are required for multiplexing different libraries in the same sequencing pool and the Illumina i-5 and i-7 sequences were attached in a second PCR round equal for both libraries. PCRs were carried out in a final volume of 25 µL, containing 1 µL of PCR product, 0.2 µM of the dual-indexed primers, 12.5 µL of 2x Multiplex PCR Master Mix (QIAGEN Multiplex PCR Plus kit), 1x CoralDYE, and ultrapure water up to 25 µL. The reaction mixture was incubated as follows: an initial denaturation at 95 ºC for 5 min, followed by 20 cycles of 95 ºC for 30 s, 61 ºC for 30 s, 72 ºC for 42 s, and a final extension step at 68 ºC for 10 min. A negative control that contained no DNA was included in every PCR round to check for contamination during libraries preparation.

The libraries were verified on 2% agarose gels and purified using the Mag-Bind RXNPure Plus magnetic beads (Omega Bio-tek), following the manufacturer's instructions. The purification was carried out twice for the libraries amplified with the fwhF2/EPTDr2n primer pair, and just once for the BF3/BR2 primer pair. Finished libraries were pooled in equimolar amounts according based on DNA quantification values determined using the Qubit 4 dsDNA HS Assay (Thermo Fisher Scientific) quantification. The pool was sequenced on a NovaSeq PE250 flow cell (Illumina) aiming for a total output of 100.000 reads per replicate. Raw sequencing data is available at NCBI SRA (Accession numbers: eDNA PRJNA1073854, bulk PRJNA1073752).

## Bioinformatic analyses

Demultiplexed reads were processed through APSCALE v1.6.3 pipeline with the default values (e.g. max sequence Expected Error: 1, min and max amplicon length: ±10 bp length expected, min size to pool sequences: 4, OUT clustering: 97%, denoising alpha: 2, denoising min size: 8) (Buchner et al., 2022). Generated OTUs were taxonomically assigned against BOLD system v4 database (Ratnasingham & Hebert, 2007) using BOLDigger v1.5.6 pipeline (Buchner & Leese, 2020). Resulting OTUs were processed with Taxon Table Tools (T.-H. Macher et al., 2020). For each sample extraction replicate reads were merged, but consistency of PCR replicates was required, so only OTUs present in all three PCR replicates were retained for bulk samples, and only taxa present in at least two of the three PCR replicates were retained for eDNA samples. OTUs reads present in negative controls were subtracted from every sample. OTUs tables were condensed in a presence/absence taxa list.

The taxa list was processed in R v4.3.1. We first looked for invasive or endangered taxa; then we filtered and removed OTUs that were not macrozoobenthos species and finally filtered and removed taxa not included in the IBMWP index.

## Statistical analyses

All statistical analyses were performed in R v4.3.1. A PERMANOVA test was run using the adonis2 function from the vegan v2.6 package to explore the effects of methodology and

sampling location on the variance of species composition. A total of 99,999 permutations were performed, and the marginal effects of each variable were tested. Only the samples with both molecular methods applied were tested in order to have a balanced design. Spearman's correlation analyses were performed to analyse the correlation between molecular and morphological EQR values. A linear model of the correlation of the EQR values of each molecular method with respect to the morphological EQR was generated and the differences were analysed with respect to the intercept and the slope of the bisector. Ecological status differences between molecular and morphological approach were tested with a Binomial and also with a Friedman test.

# Results

## Species Diversity

DNA metabarcoding of homogenised bulk samples resulted in a total of 460 detected species (643 OTUs) across all 27 samples with a mean value of 100 species per sample. A total of 360 species (405 OTUs) were detected by metabarcoding of water eDNA in 16 samples with a mean value of 84 species per sample. However, of all these species, only 160 OTUs were shared between both techniques (27.3% of the total), with 284 unique species detected in bulk samples (44.1%) and 184 in eDNA (28.6%) (Figure XVII-1 in Annex IV). The Jaccard dissimilarity in species composition between the two methodologies was 0.75 on average (Figure XVII-2 in Annex IV). The Class Insecta showed the highest species richness (bulk: 330 species, eDNA: 307) across all sites followed by Clitellata (bulk: 54, eDNA: 31) (Figure VIII-2). PERMANOVA results that species composition was significantly related to the two variables analysed were significantly related to the species composition: the sampling location (R2: 0.574, F= 2.3803, p-value<0.001) and the type of sample used (R2: 0.185, F= 11.4764, p-value<0.001).

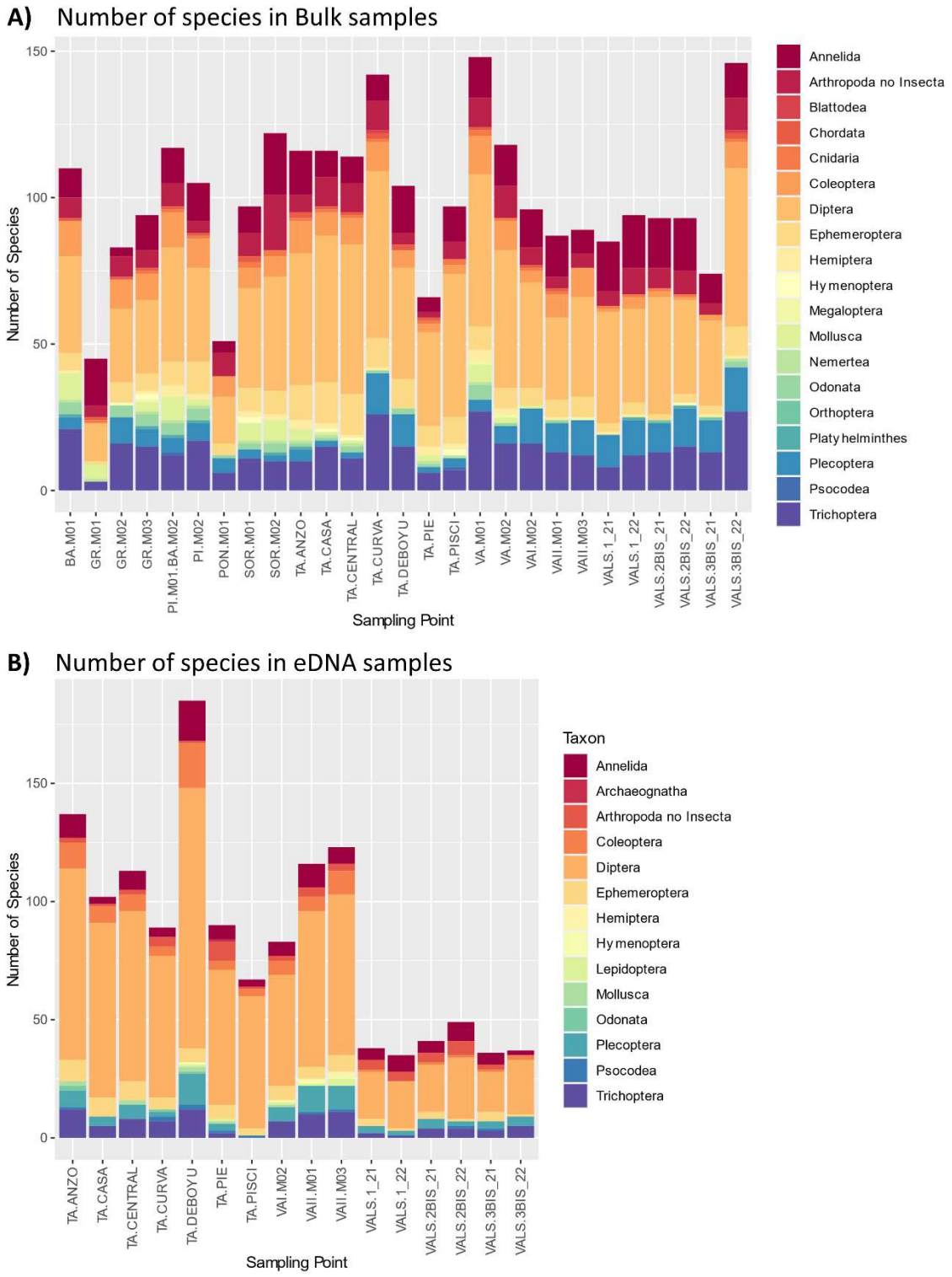



**Figure VIII-2** Species composition and number of species per sampling location. A) Bulk samples. B) eDNA samples.

Analysing all bulk samples together, 324 species were identified across the 81 IBMWP taxa, yielding an average of 4 species per IBMWP taxon (Figure VIII-3). This value is not evenly distributed among the different taxa, with a significant concentration in the family Chironomidae, which accounted for 92 species in the samples, followed by Acariformes with 20 species, and Oligochaeta with 15 species. In the case of eDNA samples, analysing

all samples together, 324 species were identified across 59 IBMWP taxa, yielding an average of 5,5 species per IBMWP taxon (Figure VIII-3). The family Chironomidae accounted for 116 species in the eDNA samples, followed by Muscidae with 16 species, and Empididae with 14 species.

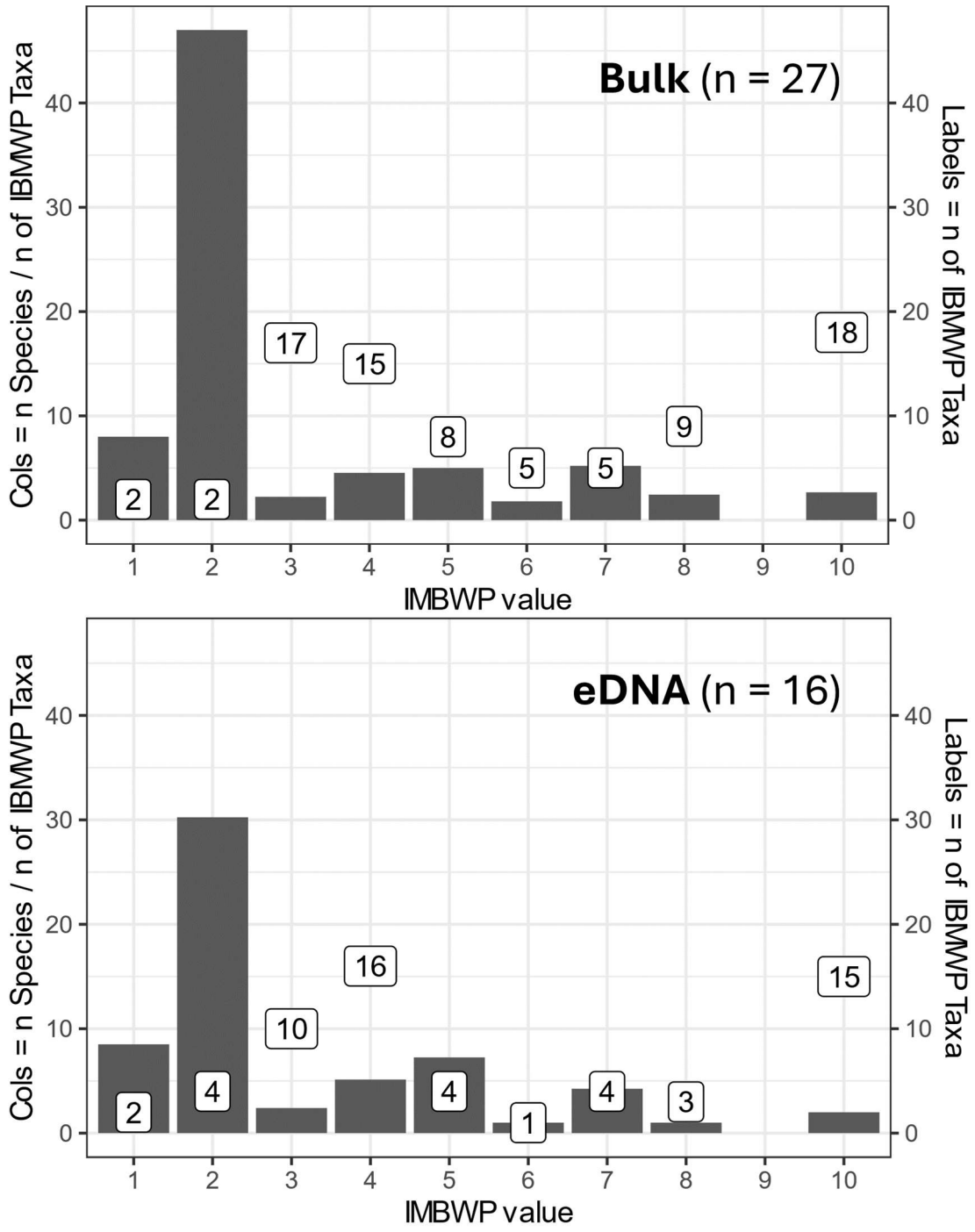



**Figure VIII-3** Number of IBMWP taxa and Species in bulk samples (above) and eDNA samples (below). Labels: Number of different IBMWP taxa present in all samples grouped by IBMWP value. Columns: Mean number of macroinvertebrate species per IBMWP taxa and IBMWP value. "n" represents the number of total samples of each type.

## Taxonomic Resolution

A total of 2,558 OTUs could not be identified to species level in the bulk samples, leaving 238 OTUs identified to genus level, 262 OTUs identified to Family level, 378 OTUs to Order level, 1600 to Class level and 60 to Phylum level. In the case of the eDNA samples 415 OTUs were identified to lower taxonomic level than species. Specifically, 182 OTUs were identified at genus level, 133 at Family level, 43 at Order level, 56 at Class level and 1 at Phylum level (Figure IX-3 in section IX).

## Exotic/invasive and Endangered/Protected Species detected.

A total of 11 exotic species were detected by molecular tools. These species were grouped into the phylum Arthropoda (6 species), Mollusca (2 species), Platyhelminthes, Chordata and Cnidaria (1 species each) (Table 1). Despite being introduced species, only three of them (*Vespa velutina*, *Potamopyrgus antipodarum* and *Pacifastacus leniusculus*) appear in the Spanish Catalogue of Invasive Alien Species (MAGRAMA, 2013c). As for the technique used, bulk proved to be the technique with the highest detection of exotic species, with a total of 9 species, while eDNA was only able to detect 4 species. However, there are two species that were only detected with eDNA and not by bulk (*Acanthocyclops americanus* and *Drosophila suzukii*). The species that have been detected in the greatest number of samples were *Craspedacusta sowerbii* (62.96% of the samples), *P. antipodarum* (44.44% of the samples) and *P. acuta* (14.81% of the samples).

**Table VIII-1** Number of samples in which Non-Indigenous Species (NIS) were detected. n: total number of samples processed by each method (in bold species declared as invasive in Spain).

| Non Indigenous Species (NIS) | Phylum (Class) | Native distribution | Invasive? | Bulk (n=27) | eDNA (n=16) |
|---|---|---|---|---|---|
| ***Pacifastacus leniusculus* (Dana, 1852)** | Arthropoda (Malacostraca) | North Western America | Yes (MAGRAMA, 2013c; Oliva-Paterna et al., 2021) | 3 | 2 |
| *Chydorus brevilabris* (Frey, 1980) | Arthropoda (Branchipoda) | North America | No | 1 | - |
| ***Acanthocyclops americanus* (Marsh, 1893)** | Arthropoda (Copepoda) | North America | Yes (Alekseev, 2021) | - | 3 |
| ***Vespa velutina* (Lepeletier, 1836)** | Arthropoda (Insecta) | South East Asia | Yes (MAGRAMA, 2013c) | 1 | - |

| | | | | | |
|---|---|---|---|---|---|
| ***Drosophila suzukii* (Matsumura, 1931)** | Arthropoda (Insecta) | South East Asia | Yes (Fiel et al., 2014) | - | 4 |
| *Ceratophysella communis* (Folsom, 1897) | Arthropoda (Collembola) | South East Asia | No | 1 | - |
| *Girardia sinensis* (Chen & Wang, 2015) | Platyhelminthes | North America | No | 1 | - |
| ***Onchorhynchus mykiss* (Walbaum, 1972)** | Chordata (Actinopterygii) | North Western America | Yes (Oliva-Paterna et al., 2021) | 1 | - |
| ***Craspedacusta sowerbii* (Lankester, 1880)** | Cnidaria (Actinopterygii) | South East Asia | Yes (Oliva-Paterna et al., 2021) | 17 | - |
| ***Potamopyrgus antipodatum* (Gray, 1843)** | Mollusca (Gastropoda) | New Zealand | Yes (Oliva-Paterna et al., 2021) | 12 | - |
| ***Physella acuta* (Draparnaud, 1805)** | Mollusca (Gastropoda) | North America | Yes (Oliva-Paterna et al., 2021) | 4 | 1 |

The species *Rana iberica* (Boulenger, 1879), included in the Spanish List of Wildlife Species under Special Protection Regime (MAGRAMA, 2011), and classified as Vulnerable (VU) under the criteria and assessments of The IUCN Red List of Threatened Species in 2020, was detected in sample GR-M02 by bulk sample metabarcoding.

## Freshwater Quality Status for River Biomonitoring

The EQR values obtained by the two molecular techniques and those obtained by morphological identification were compared for the IBMWP index (Figure VIII-4A). The results of Spearman's correlation analyses were significant for both the eDNA (rho = 0.773, p-value < 0.001) and bulk samples (rho = 0.631, p-value < 0.001). The linear model inferred for the eDNA samples shows no significant differences with either the slope (p-value = 0.450) or the intercept of the bisector (p-value = 0.543). In the case of the bulk samples there is no significant difference between the slope of the linear model obtained and the bisector (p-value = 0.377), however there is a significant difference in the intercept (p-value <0.01) which is inferred 0.37 points above the bisector origin. A total of 477 different IBMWP taxa were detected in the bulk samples (17.6 per sample) with an IASPT value of 6.04 (mean IBMWP value), while the morphological approach detected 359 IBMWP taxa (13.3 per sample) with an IASPT value of 5.64 per taxon. The Mann-Whitney U test performed to

compare the IBMWP values between the two techniques gave results that met the significance level (W= 92410.5, p-value <0.05).

The ecological states inferred from the molecular data were compared with those obtained from the morphological data (Figure VIII-4B). In the eDNA data, although there are variations in the categorization of the rivers between the two methods (8 equal states, 3 improved states, 5 worsened states), there is no significant tendency to infer a better or worse state than with the morphological data (p-value binom test: 0.726, p-value Friedman test Friedman test: 0.479). However, in the bulk sample data, a higher percentage of differences (10 equal states, 17 improved states) are observed, and these tend significantly to a better value of ecological state than inferred from the morphological data (p-value binom test < 0.001, p-value Friedman test < 0.001).

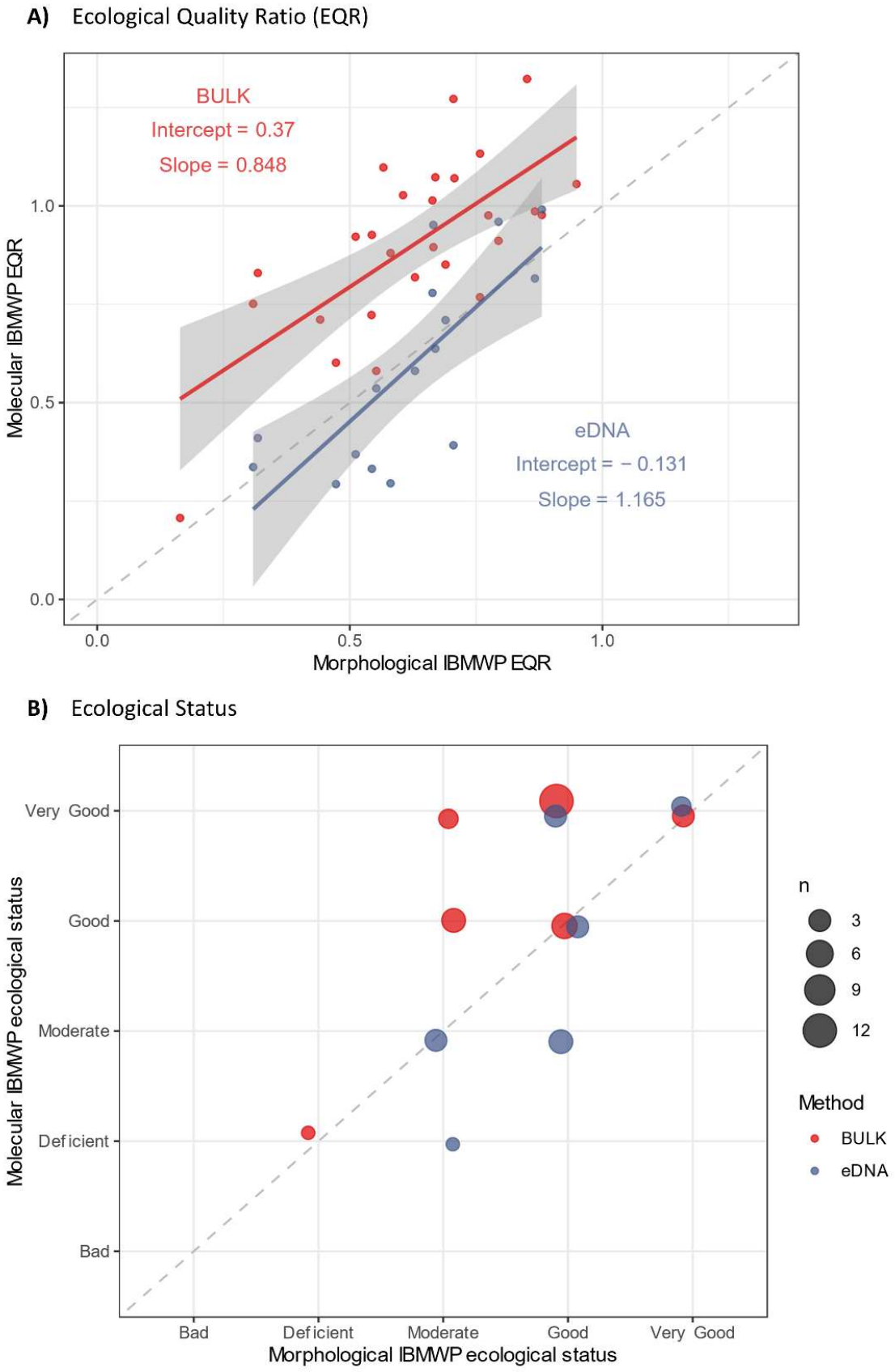



**Figure VIII-4** Comparison between A) EQR and B) ecological states obtained for the same sampling locations using different methods: molecular (eDNA and bulk samples) and morphological.

By simulating how species loss in bulk samples would affect the IBMWP EQR and the inference of ecological status for bulk samples (**¡Error! No se encuentra el origen de la referencia.**), we find that, on average, the loss of 7.5 species results in a shift from very good to good ecological status, while the loss of 15.4 species leads to a shift from good to moderate ecological status, when species removal is weighted by their IBMWP tolerance value (Table XVI-1). This represents the strictest scenario, as the highest scoring species would be removed first. In contrast, if species are removed randomly these thresholds increase to 8.8 and 18.5 species for changes from very good to good and good to moderate status, respectively (Table XVI-2).

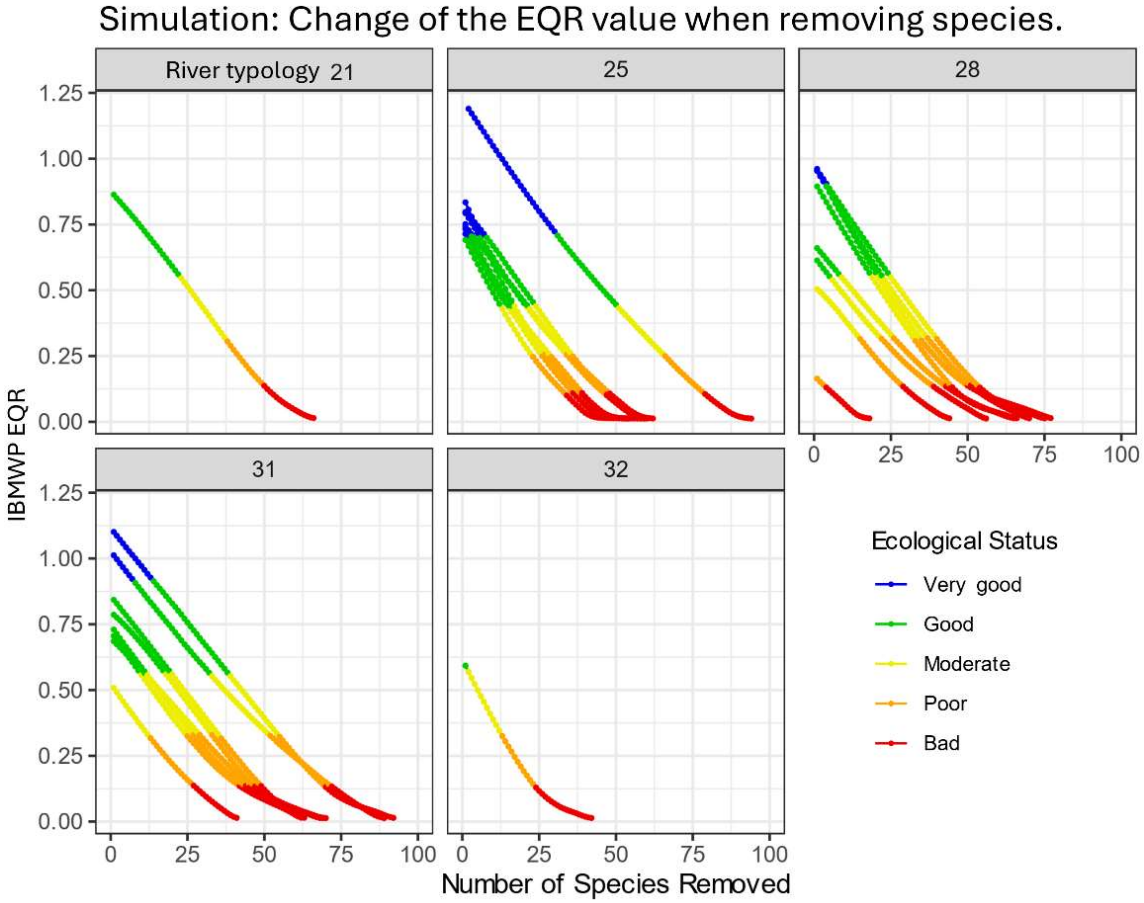



**Figure VIII-5** Each line corresponds to one bulk sample. The EQR value is the mean obtained after 1000 simulation replicates. Removal of species was weighted according to their sensitivity (i.e. a species with an IBMWP value of 10 was 10 times more likely to be removed than a species with an IBMWP value of 1).

# Discussion

## Metabarcoding unravelling rivers diversity: native, exotic, and protected species.

Achieving species-level resolution by morphological identification of biodiversity in river kicksamples is time-consuming and requires expert taxonomists. Therefore, in this study we have used molecular techniques such as DNA metabarcoding as a tool to identify species-level diversity in a more efficient way. With this approach we have been able to detect and identify a total of 628 different species from eDNA and bulk samples. However, only 160 species were detected with both methods, and the PERMANOVA analysis showed significant differences in species composition depending on the methodology used. The possibility that the genetic information in each type of sample differs could account for the variation in species composition found with each molecular technique. Bulk samples have been taken according to the official methodology of the ministry (MAGRAMA, 2013b), methodology adapted from (Barbour et al., 1998; Jáimez-Cuéllar et al., 2002) which consists of multihabitat, stratified and semiquantitative sampling of a 100-metre stretch of river with a Surber net. Similar sampling methodologies have been tested and have only managed to capture 50% of the species present in the stretch (M. T. Furse et al., 1981). On the other hand, the DNA captured in eDNA samples is not limited to the species present within the 100 meters sampled for the bulk sample but can provide DNA from several kilometers upstream of the sampling location (Deiner & Altermatt, 2014). In addition, DNA from terrestrial species is also detected in eDNA samples, which can be washed into the river by runoff or even from sewage treatment plants, which can 'pollute' the river with DNA from commercial species (Deiner et al., 2016; Inoue et al., 2023). Therefore, it is reasonable that the species composition of the genetic material contained in the two samples may vary greatly. Although there could be other possible reasons for these differences, such as the fragment amplified by the eDNA primer being too small and not powerful enough to correctly assign at the species level (Yeo et al., 2020), or that using different primers for each sample type leads to different amplification biases (Leese et al., 2021).

DNA based molecular techniques are a reliable tool for diversity studies and they are also widely used for species detection and identification of exotic and invasive species. Techniques such as qPCR or LAMP are used for the detection of single invasive species in aquatic environments (Carvalho et al., 2021; Cary et al., 2014; Keller et al., 2017; Peñarrubia et al., 2016; Sepulveda, Nelson, et al., 2020). But in addition to studying diversity as we have shown, the metabarcoding technique enables us to simultaneously identify several exotic and invasive alien species without additional work or cost. In this work we have reported a total of 11 species whose native distribution area does not correspond to the Iberian Peninsula, eight of them considered as invasive species (Table 1). All of them correspond to species whose presence in our waters had already been proven by previous morphological

and molecular studies, except for two species, the brachiopod *Chydorus brevilabris* and the springtail *Ceratophysella communis*. The species *Chydorus brevilabris* is a branchiopod crustacean with North American origin that has been reported in other European countries such as Belgium, France, Luxembourg, and the Netherlands (Soesbergen, 2020) and recently in the Asian country of Japan (Makino et al., 2023); however, its presence in the waters of the Iberian Peninsula was unknown. On the other hand, *C. communis* is a terrestrial springtail of Asian origin that has never been reported outside its native distribution. There seems to be a lack of documented impacts of these exotic species in the new environments where they have been introduced, thus both species should be the focus of future genetic and morphological studies to confirm their presence and impacts in the region.

As noted by Fiel et al. (2014) and Sánchez & Arias (2021), invasive species such as the insects Vespa velutina and Drosophila suzukii are highly prevalent in the area and it is therefore not unusual to find traces of genetic material in the samples. Similarly, the salmonid Onchorhynchus mykiss was introduced in Asturias for the purpose of salmonid aquaculture (Márquez, García-Vázquez and Borrell, 2014). On the other hand, the invasive aquatic species of gastropod molluscs, *Potamopyrgus antipodarum* and *Physella acuta*, as well as the invasive decapod crustacean *Pacifastacus leniusculus*, are widely distributed and very abundant in our waters (Vedia, Iván & Miranda, Rafael, 2013). They were the most detected exotic species by molecular techniques during this study, appearing; in the case of *P. antipodarum*, in more than 40% of the samples analyzed. The presence of the invasive cnidarian species *Craspedacusta sowerbii*, an active predator of zooplankton that can change animal communities in the water bodies it colonizes (Smith & Alexander, 2008), in more than 60% of the samples is also thought to be the most relevant data found (Medina-Gavilán & González-Duarte, 2018) While its presence was known in a large area of the Iberian Peninsula and in the surrounding autonomous communities of Galicia and Castilla y León, its presence in the Principality of Asturias was unknown.

Finally, the exotic flatworm *Girardia sinensis* and the copepod *Acanthocyclops americanus* were only known from the central and western fringe of the Iberian Peninsula where no impacts on the ecosystem have been reported (Alekseev, 2021; Benítez-Álvarez et al., 2023), corresponding, these records, to an increase in the knowledge of their distribution in this area. Since *A. americanus* can have less than 150 µm width, its presence in the sample can be lost when using a 500 µm Surber, the processing method of the bulk samples likely favors the species' presence in eDNA rather than bulk.

As these tools are useful for the detection of exotic and invasive species, they are also suitable for the monitoring and detection of endangered and protected species. In this work we report the detection of the Iberian frog (*Rana iberica*), a vulnerable species (IUCN, 2024) included in the Spanish List of Wildlife Species under Special Protection Regime

(MAGRAMA, 2011). This species, endemic to the Iberian Peninsula, is mainly distributed in the northwestern region with stable populations in the Basque Country and the Central System (Pleguezuelos et al., 2002). It was detected within its known distribution area, in the surroundings of the Grandas de Salime reservoir (Asturias). This indicates that metabarcoding (if using adequate primer pairs) can be used as a tool for follow-up campaigns for endemic and protected species in rivers.

Even with all the species detected, it was not possible to identify all the OTUs to species level. In the bulk samples only 17.43% of the OTUs generated could be identified to species level, in the case of the eDNA samples this percentage rose to 49.4%. This lack of taxonomic resolution can have several causes, such as the lack of sequences in the reference databases that prevent the taxonomic assignment of those species that are not currently found in the databases (Csabai et al., 2023; Fueyo, Sánchez, et al., 2024; Weigand et al., 2019). But it may also be due to the presence of NUMTs (nuclear copies of mitochondrial genes) or heteroplasmy, as they can provide OTU diversity that does not translate into real species diversity (Ožana et al., 2022). The lack of sequences in the reference databases of these variations or of reliable systems integrated in the bioinformatic pipelines of metabarcoding prevents the correct filtering of these variations. This results in the generation of spurious OTUs, many of which are identified at lower taxonomic levels (Anderson & Leite, 2012; Porter & Hajibabaei, 2021). These limitations in taxonomic resolution have little impact on the management of invasive, protected species or species of regional interest, as most of these species have sequences in the databases due to their importance. The number of OTUs that are not identified to species level and are not the result of artefacts, heteroplasmy or NUMTs can be gradually reduced by increasing the reference databases. To this end, there are many sequencing projects such as the International Barcode of Life (IBOL).

## Metabarcoding for biomonitoring the ecological status of river ecosystems.

In addition to diversity studies and the detection of exotic and endangered species, molecular techniques for biomonitoring the ecological status of river ecosystems are being proposed (Fernández, Rodríguez-Martínez, et al., 2019; Keck, Blackman, et al., 2022; Mortágua et al., 2019). Our results find significant correlations between the EQR values of each molecular technique with respect to those obtained by morphological identification for the IBMWP index showing the potential of these techniques for application in biomonitoring. However, the intercept of the linear model of the EQR values of the bulk samples has a significantly higher intercept than expected if the values were similar to those obtained by morphological identification. This indicates that the metabarcoding of bulk samples would be detecting more taxa or taxa with higher index values. In principle, metabarcoding of bulk samples is expected to have a higher detection rate (Kuntke et al.,

2020), as it is able to detect species from elements that do not easily allow their morphological identification, such as eggs, pupae, feces, parts of an individual (legs, abdomen, guts, etc), and it is even possible to detect the presence of a species by the presence of its DNA in the stomach contents of its predators (L. L. de Sousa et al., 2019). However, there is a lack of studies that go into detail to check all the false negative and false positive detections to differentiate ecologically relevant results from artefacts of the technique, as these will translate into changes in EQR values and hence inferred ecological status.

Only 50% of the statuses inferred from the eDNA results coincide with those inferred from the morphological data when the metabarcoding EQR values are converted into ecological status, but there are no significant trends towards poorer or better status. However, the bulk samples not only had 63% of samples with a different inferred status, but these tended to be significantly higher than those inferred from the morphologically detected families. These results clearly show that metabarcoding results cannot be directly applied on top of the current indices to infer ecological status since, among other things, the current indices are designed and calibrated with samples processed using morphological identification approach.

Both the difference in detectability and the differences in EQR and inferred ecological status are not a problem in the final implementation of these techniques for biomonitoring, as they still have to go through a long process of optimisation, fine tuning and intercalibration before they are ready for final testing for biomonitoring as when the currently used morphological techniques were standardized (Birk et al., 2018; Friberg et al., 2006; M. Furse et al., 2006). During this process, known problems in the application of these techniques should be corrected, such as the lack of sequences in the reference databases (Fueyo, Sánchez, et al., 2024), and intercomparison experiments should be carried out to determine the replicability of the technique (Blackman et al., 2019b). In addition, a more detailed study of the differences in detection between the metabarcoding and morphological identification and their possible causes is needed, as this can directly influence the results of the ecological status calculation.

## Macroinvertebrate biomonitoring under the new perspective of increased taxonomic resolution of DNA metabarcoding.

Species-level identification, while providing more detailed taxonomic information on macroinvertebrate composition, it is not necessarily the most appropriate approach for all biomonitoring studies. This perspective is summarised in the pragmatic concept of taxonomic sufficiency, where depth of identification is balanced against the need for information (Ferraro & Cole, 1992). Historically, this concept has been seen as an effort to reconcile the scientific ideal [sic] with the political, financial and logistic realities that affected the studies (F. C. Jones, 2008). Successive reviews on the subject indicate a

persistent lack of consensus around taxonomic sufficiency (Carter & Resh, 2001; F. C. Jones, 2008; Mueller et al., 2013; Resh & McElravy, 1993). One of the main arguments in the debate was the trade-off between taxonomic resolution and the cost/time needed (McGauley et al., 2018). Nevertheless, with the development of DNA metabarcoding, it is now possible to achieve species-level taxonomic resolution without increasing time or cost, bringing this debate back into focus.

The benefits of higher taxonomic resolution for the detection and identification of invasive or endangered species, as well as for mapping the specific macroinvertebrate biodiversity of sampling sites, were previously outlined. However, the new information provided for ecological status inference indices and its relevance to river biomonitoring has not yet been discussed here. It is important to consider the limitations which may affect the final list of species detected by metabarcoding in our samples (see section IX). These limitations include missing sequences in reference databases, the possible inclusion of terrestrial species, the small number of samples processed, and their limited geographical range. Despite these constraints, the data could offer a valuable preliminary approximation.

The species disappearance simulation shows that the changes in IBMWP status may reflect losses in specific diversity of between 8 and 18 species (Figure VIII-5). This does not take into account the possible appearance of impact-tolerant species replacing sensitive species, which, if counted, could increase the number of species present in the original community that would have to disappear in order for the ecological status of the river to deteriorate. In bulk samples, a comparison of the number of species detected (324) with the number of families (81) indicates that the low resolution of the IBMWP index results in a loss of approximately three quarters of the total available taxonomic information in the samples, this value increases in eDNA samples. Furthermore, compared to the total number of IBMWP taxa (81 in bulk and 56 in eDNA), more chironomid species were detected (92 in bulk and 116 in eDNA). It is well known that chironomids are responsible for much of the richness and abundance of aquatic macroinvertebrate communities, particularly in impacted environments (Serra et al., 2017) and are generally considered a pollution-resistant group (IBMWP value = 2). Conversely, many studies have found that some Chironomidae subfamilies, genera and species can be indicators of good water quality (Lenat, 1993; Lencioni et al., 2012; Paggi, 1999). In fact, river quality indices based only on the identification of chironomid larvae have been proposed (Molineri et al., 2020). The family Chironomidae is the most notable example, but this diverse response of different species within the same family to environmental impacts may be occurring in other groups (Figure XVI-3, and see page 54 of F. C. Jones, 2008). New indices based on more complete and detailed information can be developed using this improved taxonomic resolution (Whittier & Van Sickle, 2010). Additionally, considering that species may respond differently to different stressors, stressor-specific indices can be developed (Berger et al., 2018).

## Consideration for management

The scalability provided by molecular techniques promises to facilitate biomonitoring programs for diversity in the Iberian Peninsula. This, together with the progress in the study of other species groups (Apothéloz-Perret-Gentil et al., 2021) or the advancement of the eDNA technique with other types of samples (Garrett et al., 2023; T. H. Macher et al., 2023) shows the potential of this technique to achieve integrated biomonitoring of the global diversity of ecosystems in the future. As previously seen, the multiapproach properties of metabarcoding not only position it as a pathway for studying diversity in rivers but also as a technique for monitoring endangered and invasive species and inferring the ecological status of rivers. All of this within a single analysis. Nevertheless, the wide variation in results between the different methods casts doubt on their interpretation for use in management. Previous studies have shown that results from bulk samples are more similar than eDNA results to those from the morphological method, which is the gold standard (See Section IX). In addition, certain characteristics of bulk samples make them more reliable than eDNA for species identification. These are:

- The ability to use longer genetic markers, as the DNA in bulk samples is expected to be less degraded than in an eDNA sample. This allows more genetic information to be provided when identifying sequences, giving more confidence in species identification, although previous studies show that fragments larger than 200 bp are sufficient (Yeo et al., 2020).
- The possibility of using more degenerate primers as there is less risk of unspecific amplification in bulk samples. This makes it easier to avoid false negative detections due to primer bias.
- To have a higher concentration of DNA per individual in the sample as it is extracted directly from tissue. Potentially ensures the presence of DNA from all species in all extraction and PCR replicates.
- Much better delimitation of the spatial scale of the study, as eDNA can be washed to the sampling site from miles upstream of the river or from the land with run-off water (Deiner et al., 2016). This is a major problem when this DNA comes from sewage water as it may contain DNA from species that are not found in the river but rather species for human consumption or from aquariums (Inoue et al., 2023).

Yet eDNA also has its advantages compared to bulk samples in terms of species identification as it can detect species smaller than 500 µm which is the pore size of Surber nets, and the primer bias is different from the primers used in bulk samples. We consider it necessary to take all these characteristics into account when interpreting the results of these two techniques for management decisions. For the design of new projects, certain other differences between the two methods should be considered. The use of eDNA is a non-invasive and faster sampling method and can be carried out relatively easily in non-

wadeable rivers. Further efforts are needed to optimise and refine these two molecular techniques for use in diversity and/or biomonitoring studies. This includes species sequencing, development of informatics tools integrated into metabarcoding pipelines to detect and remove sequencing artefacts and NUMTs, development of better primers, implementation intercalibration projects, etc. Still, these DNA-based techniques continue to be useful alongside morphological identification. They allow higher taxonomic resolution, detect cryptic species, identify species that do not have the morphological characteristics that define them in good condition or for which there are no identification keys, identify species at different stages of development and have a second molecular confirmation of identifications, reducing human error.

The study presented in this section serves as a further step in implementing these molecular techniques for riverine biomonitoring in the Iberian Peninsula, one of many that need to be taken before a robust and reliable technique is available for routine use. The findings of this study serve as a baseline for future investigations into species diversity within this region.

# IX. General Discussion: Data-based considerations for implementing macroinvertebrate DNA metabarcoding for effective river biomonitoring in peninsular Spain.

*The contents of this section are currently under review in the journal Ecological Solutions and Evidence as: Fueyo, Á; Sánchez, O; Carleos, C; Granero-Castro, J; Borrell, Y. (2024) considerations for implementing macroinvertebrate DNA metabarcoding for effective river biomonitoring in peninsular Spain - advantages, limitations, and roadmap. Ecological Solutions.*

As previously seen, there are certain general advantages and disadvantages in the application of DNA metabarcoding for the biomonitoring of rivers with macroinvertebrates that are applicable universally. However, the application of these methods may vary depending on the country where they are applied, since the species and families to be studied may change, as well as the taxonomic and genetic information available about them, and other more formal elements such as official indices, sampling protocols, specific legislation, etc (Birk et al., 2012b; Friberg et al., 2006). Therefore, in this section we analyse the technical peculiarities of the implementation of this technique specifically in peninsular Spain. In this section, we (a) compare the results of detection of river macroinvertebrate taxa by morphological identification and molecular techniques in both bulk and water eDNA samples, (b) identify false positive and false negative detections of DNA metabarcoding and other sources of variation that explain the differences with the morphological identification method, (c) discuss the impact of these differences on the reliability of the use of these techniques for biomonitoring and offer solutions to resolve them when they could be a problem, (d) we analyse the new insights provided by the improved taxonomic resolution of DNA metabarcoding in biomonitoring using the IBMWP index, and (e) propose a roadmap for an effective implementation of the use of DNA metabarcoding for river biomonitoring in Peninsular Spain through the study of macroinvertebrates as bioindicators.

In addition, we illustrate this analysis with the detailed analysis of real data on the application of DNA metabarcoding in rivers in the northwest of the Iberian Peninsula, data from our previous section (VIII). Methodology for the new analyses applied to these data is given Annex V. In short, a total of 27 sites in the northwest of the Iberian Peninsula were sampled with parallel collection of kicksamples for morphological analysis and bulk samples for metabarcoding, following the sampling protocol of the Spanish Ministry (MAGRAMA, 2013b). At 16 of these sites, water samples were also collected for eDNA metabarcoding analysis. Two different library sets were constructed for bulk and eDNA

samples using two different primer pairs and conditions in order to specifically detect the macroinvertebrates present in each sample type. The taxa list was collapsed at the level of the different taxa considered in the IBMWP index (hereafter IBMWP taxa) which is mainly at family level (1 genus, 121 families, 1 superorder, 1 subclass, 1 class; MAGRAMA, 2013A, Table XIV-1 in Annex I), removing all taxa not included in the index.

## General overview of detection dissimilarities when using molecular methods or morphological identifications.

Previous research has shown significant differences in the detection of river macroinvertebrates by morphological identification and DNA metabarcoding, even when using different protocols, primers and taxonomic resolutions (A. Beermann et al., 2021; Deiner et al., 2016; Emmons et al., 2023; Gleason et al., 2020; Hajibabaei et al., 2019; Keck, Blackman, et al., 2022; Meyer et al., 2021). These findings are consistent with our results, which confirm a significant relationship between macroinvertebrate composition at the IBMWP taxa level and the methodology used (Table XVIII-1 in Annex V). However, the magnitude of the differences in macroinvertebrate composition between the molecular methods varies depending on the level at which the differences are analysed. For example, the differences between molecular methods appear to increase when they are analysed at higher levels, such as species level (Fueyo, Sánchez-Fernández, et al., 2024), OTU level (Gleason et al., 2020) or exact sequence variants (ESV) (Hajibabaei et al., 2019). Differences between molecular techniques and morphological identification also vary with the taxonomic resolution at which the data are compared, with differences increasing with increasing taxonomic resolution (A. Beermann et al., 2021; Emmons et al., 2023). Differences in detection between molecular and morphological approaches also vary depending on the type of sample used. For example, the Jaccard dissimilarity scores obtained by Gleason et al. (2020) and van der Lee et al. (2024) for macroinvertebrate composition show less dissimilarity between bulk and morphological samples than between eDNA samples and either of the other two methods. This is consistent with our results of Jaccard dissimilarity at IBMWP level (mainly family level) between different sample types (Figure IX-1). Compared to the studies by Gleason et al. (2020) and van der Lee et al. (2024), we obtained slightly lower dissimilarity scores when comparing the results of bulk and morphological data with the eDNA samples, that is, our eDNA results are more similar to the other two methods. The reason for this difference may be that we used a different pair of primers specifically designed for application with water eDNA samples, which improves detection of macroinvertebrate families compared to no target amplification (Leese et al., 2020).

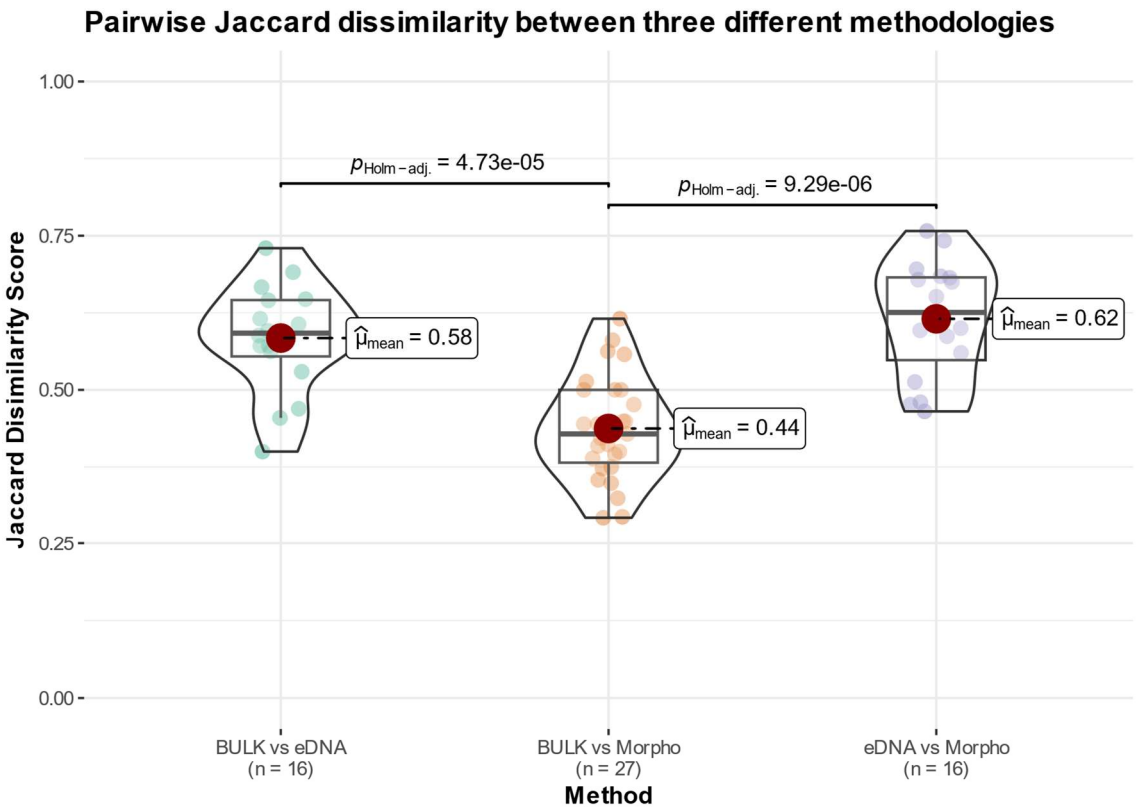



**Figure IX-1** Jaccard Dissimilarity Score between methodologies. Significant pairwise comparisons are displayed.

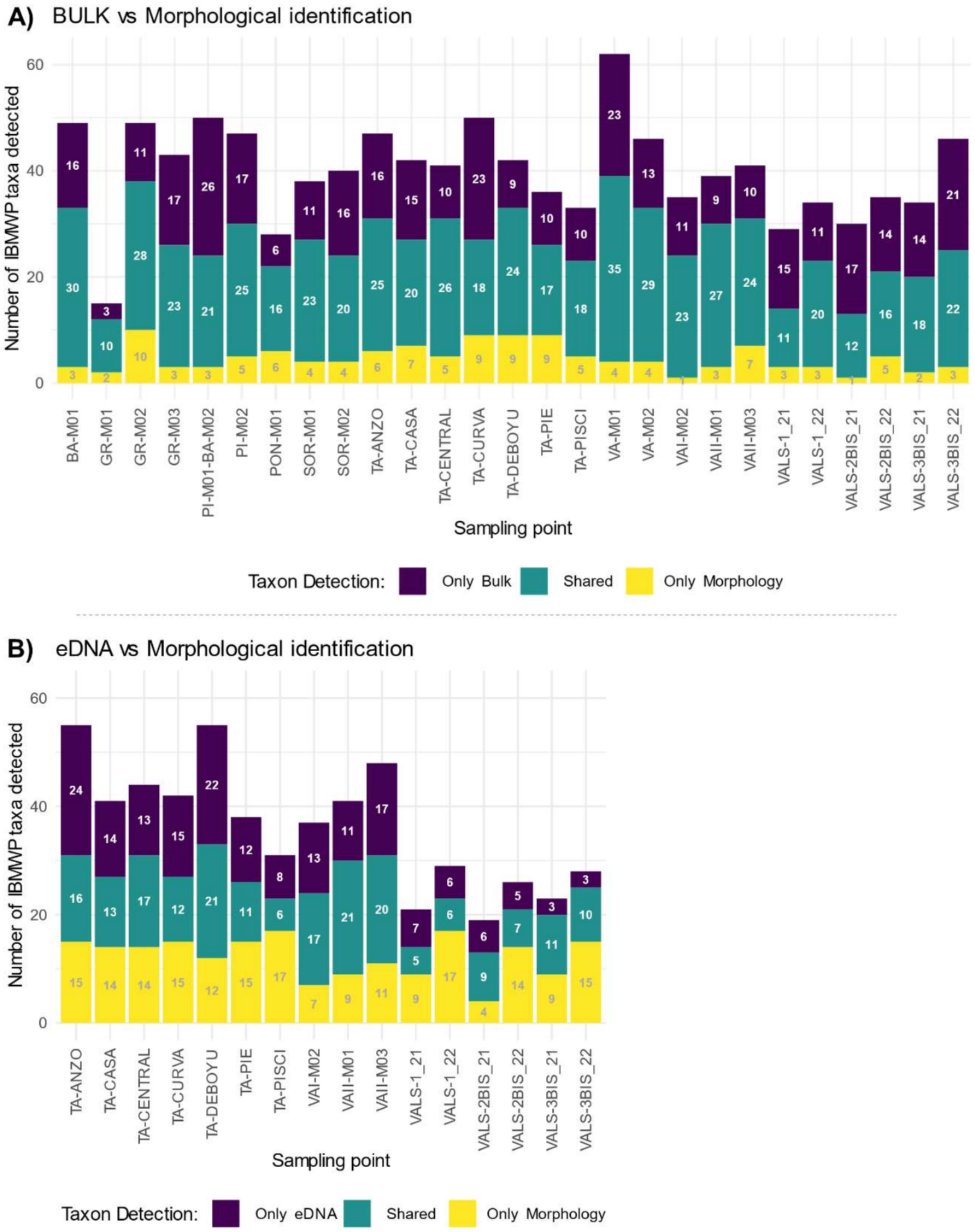



**Figure IX-2** IBMWP taxa shared between morphological identification and both metabarcoding approaches: A) Bulk, B) eDNA.

## Differences in detection between bulk sample metabarcoding and morphological identification in an 9-point overview.

Metabarcoding of bulk samples can potentially identify species from organic elements that are practically impossible to identify by morphological identification techniques. These are

parts of an individual (legs, abdomen, viscera, etc), pupae, eggs, feces, recent exuviates and it is even possible to detect the presence of a species through the presence of its DNA in the stomach contents of its predators (e.g.: de Sousa et al., 2019; Siegenthaler et al., 2019). That is, metabarcoding enables the detection and identification of species from almost all recent biological material in the sampled area. Thus, potentially increasing the detection capacity for elusive, cryptic, or low-abundance species. This increase in detection is considered one of the strengths of metabarcoding and can be illustrated in our data, where of the total of 80 IBMWP taxa detected by both methods, 16 were exclusively identified through metabarcoding, while only 5 were detected through morphological identification alone (Figure XVIII-1 in Annex V). However, the improvement in detection cannot fully explain all the differences found in the comparative results, as 26.25% of the taxa were only detected by one of the two methods (Figure IX-2A). There are other sources of variation that require further exploration:

## False positives.

1- Presence of terrestrial species.

Some of the families considered by IBMWP Index include both freshwater and terrestrial species. Some of these terrestrial species live relatively close to rivers and when sampling the riverbanks may fall into the sampling nets. These species should not be considered when calculating the presence/absence of the family in the index. And it is occasionally straightforward to identify some of the terrestrial species in morphological identification as they often have characteristic features. However, the metabarcoding results return a final list of species, where it would be easy to select only those that are aquatic if a complete checklist of the aquatic macroinvertebrate species of the Iberian Peninsula were available. But only a checklist at family level is available now (MAGRAMA, 2013a).

### Example.

This is the case for the subclass Oligochaeta s. str. (referring specifically to clitellates, excluding branchiobdellids and leeches). Approximately only one-third of the near 5.000 valid Oligochaeta species described to date inhabit freshwater environments. Luckily, a published checklist at the family level is available to help identify freshwater families (Martin et al., 2008). This has allowed us to remove 27 Oligochaeta OTUs (18 taxa) belonging to terrestrial families from our results. However, a family-level checklist is not enough, especially considering that Enchytraeidae family has two-thirds of its known species inhabiting environments other than freshwater (Martin et al., 2008). Therefore, it is unclear whether any of the 54 OTUs (25 taxa) identified in our data and classified as Enchytraeidae are terrestrial, which could lead to a false positive detection.

### Solution.

The most direct solution to this problem is construct a freshwater macroinvertebrate species checklist for the Iberian Peninsula as already exist for Azores Islands, Madeira and

Britain (Gunn et al., 2018; Hughes et al., 1998; Raposeiro, Pedro Miguel et al., 2012). Thus, terrestrial species could be filtered out, leaving only those with an aquatic phase in the river.

2- Erroneous taxonomic assigment.

As a final stage in the bioinformatic processing of metabarcoding data, the sequences obtained are compared with reference databases such as BOLD and GenBank to assign each one to a species. These reference databases have been built with contributions from the global scientific community, and although they have review and curation processes, they may contain errors, and these errors could be transferred to the metabarcoding assignment results (Baena-Bejarano et al., 2023; Meiklejohn et al., 2019; Pentinsaari et al., 2020; Sonet et al., 2013).

Example.

In our results, of the 723 OTUs assigned to IBMWP taxa in the bulk samples, 113 had two or more entries with different taxonomic information above the selected threshold (98% sequence similarity). In the case of eDNA samples it was 108 out of 577 OTUs. These results may be due to the presence of incorrectly identified sequences in the BOLD database, as discussed previously. But may also be due to the presence of synonymous names, species complexes (specimens belonging to morphologically distinct species that cannot be distinguished by DNA barcodes, including unconfirmed synonyms) or fuzzy species names (typographical errors in some of the entries in the database).

Solution.

A common solution is to generate a curated local/custom database from the reference databases and and/or from own sequencing data (Jeunen et al., 2023; Meglécz, 2023; Mugnai et al., 2023). However, this strategy has the limitation that not all BOLD sequences will be available, as many of them, although accessible by the BOLD ID engine, cannot be downloaded because they are marked as private (not yet published) and these gaps can have a major impact on taxonomic assignment (see section VI). On the other hand, reference databases have a gradual curation process in which errors are gradually corrected and conflicts between sequences and taxonomy are resolved. The ultimate long-term goal of this is to have reliable and error-free databases. So, one option is to manually check and report all errors in the results of each project, in order to gradually improve the reference databases.

## False negatives.

3- Taxonomic changes (synonyms and fuzzy names).

One of the steps in the bioinformatics processing of metabarcoding data is to assign the taxonomy of the species to each of the OTUs obtained. The problem arises because genetic databases contain certain errors in the taxonomic description. These errors can be synonyms due to the lack of updating to the new taxonomic changes or fuzzy names

(binomial names with missing or extra characters) due to data entry errors. These errors can cause mismatches during post-processing of the species checklist, resulting in a false negative detection.

Example.

In section VIII, taxonomic names were assigned to the OTUs using the BOLD database (Ratnasingham & Hebert, 2007). Upon comparing the nomenclature with the GBIF database (GBIF, 2024), we found certain discrepancies in some taxa (Table XVIII-3 in Annex V). Three of these discrepancies were observed at the taxonomic family level, potentially altering the value that the IBMWP index assigns to the detected species or even excluding it from the analysis. These cases involve the families Ancylidae/Planorbiidae, Thremmatidae/Uenoidae and Corixidae/Micronectidae. Firstly, traditionally the Ancylidae family was considered as a separate family within the superfamily Lymnaeoidea, however recent phylogenetic studies place it as a Tribe (Ancylini) within the family Planorbidae (Tinerella, 2006). This fact creates a confusion since the IBMWP index contemplates two different values for these two families, being 6 for the former family Ancylidae and 3 for the family Planorbidae. On the other hand, the taxonomic history of the family Thremmatidae has been quite variable over the years. First the genus *Thremma* (the only genus present in the Iberian Peninsula) was described as belonging to a new family, the family Thremmatidae. Later, some authors related this genus to the family Uenoidae, placing them as a new subfamily (Thremmatinae) within it. However, recent studies have again given the family Thremmatidae an independent category (Waringer et al., 2020). These recent changes have not been updated in many databases and indices, creating difficulties in manage the data. Finally, there is the case of individuals belonging to the current Micronectidae family, which some authors considered an independent category while others considered it a subfamily (Micronectinae) within the Corixidae family (Tinerella, 2008). However, later morphological and molecular studies affirmed that it was an independent family (Tinerella, 2006), but in many databases it is still assigned to the family Corixidae.

Solution.

Taxonomic correction tools such as GBIF species matching have been developed to address these issues, but are not yet ready for full automation as they require manual review. Another solution could be to include the most recent synonyms in the macroinvertebrate species checklist and consider them during the bioinformatic processing and create a curate and update genetic database for Iberian macroinvertebrates species which should be updated taxonomically on a regular basis.

4- Lack of reference sequences in genetic databases.

During the taxonomic assignment process, each of the OTUs obtained is identified by checking its sequences against genetic databases such as BOLD or GenBank. However,

these databases are far from complete (Csabai et al., 2023; Fueyo, Sánchez, et al., 2024; Múrria et al., 2020; Weigand et al., 2019). Therefore, if sequences of a species are missing, the taxonomic assignment cannot be performed correctly, resulting in a false negative detection. However, when species-level sequences are not found, the bioinformatic processing could address this problem by increasing the taxonomic range of assignment, reducing the required percentage of similarity between sequences (Buchner & Leese, 2020). This avoids false negative detection but imply a low taxonomic resolution for some OTUs.

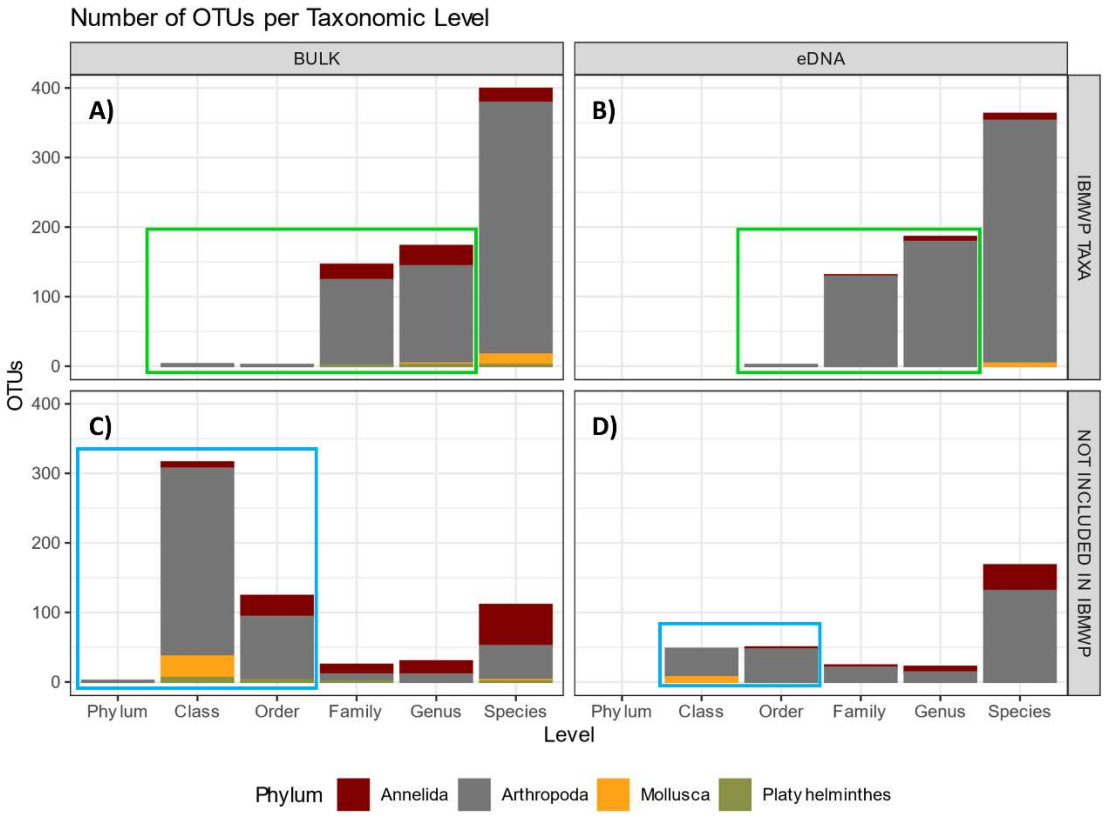



**Figure IX-3** OTUs Taxonomic resolution in metabarcoding data from the eDNA and bulk samples. OTUs are clustered depending on the type of sample and whether they are considered indicator taxa (IBMWP taxa) or not. Green square: OTUs included in the IBMWP taxa identified at higher taxonomic resolution than species level. Blue square: Potential IBMWP taxa discarded cause do not have enough taxonomic resolution.

Example.

In section VIII data, of all 1331 OTUs generated from the bulk samples only 38.3% were identified to species level and 66.4% were identified to, at least, family level (Figure IX-3). Of the 608 OTUs excluded from the index calculation, 72.6% were excluded because they were identified at low taxonomic resolution. That means that those OTUs detected in the bulk samples could have belonged to one of the indicator taxa considered in the index.

Solution.

As proposed in section VI one possible solution to this problem is to sequence all the Iberian macroinvertebrate species lacking a sequence in reference genetic databases for the COI gene. Otherwise, using molecular techniques to infer ecological status without reference sequences of all macroinvertebrate species would not only involve the risk of generating false negatives in detection, but as databases are completed over time, the results of the same sample could change depending on the year in which it is analysed, giving a false gradual increase in diversity. However, to carry out this sequencing effort directly, it is recommended to complete the checklist of macroinvertebrate species of the Iberian Peninsula first.

5- Low abundance and low biomass species.

The difference in number and/or biomass of different species in bulk samples can differ by several orders of magnitude. This difference translates into the number of copies of target DNA per individual in the final pool for sequencing, meaning that species with a low relative number or biomass in the sample may be overshadowed and therefore undetected, especially at lower sequencing depths (Braukmann et al., 2019; Shirazi et al., 2021). Sequencing depth is known to be a relevant factor for species detection in DNA metabarcoding studies, comparable to the sampling effort in field studies (Shirazi et al., 2021). Increasing sequencing read depth enhances the likelihood of recovering low abundance/biomass taxa from within the sequences pool although increasing it too much may simply result in more PCR and sequencing artifacts being recovered (Alberdi et al., 2018). Values of 100,000 sequencing reads per replicate have been proposed as an optimal value for bulk samples of macroinvertebrates (Elbrecht & Steinke, 2019), although each study has its own target (e.g. 120.000 reads per library Espinosa Prieto et al., 2024).

Example.

In our data from section VIII, we observed a significant relationship between the abundance of taxa detected in morphological samples and the probability of detection in bulk samples (p-value < 0.01), with taxa detected at lower abundances in morphological samples being the least likely to be detected in bulk samples (Figure XVIII-3). However, the detectability of low abundance taxa (n<5) was not significantly dependent on their biovolume, i.e. in which sieve these taxa were found in morphological identification (abundance 1: p-value= 0.1034, n= 124; abundance 4: p-value = 0.309, n = 56). The relationship between abundance and detectability could be explained by a sampling effect (see point 9- of this section). Despite this, the lack of a relationship between detectability and abundance could be due to either sufficient sequencing depth to avoid bias due to the biomass of the different taxa, or a lack of statistical power.

Solution.

A solution to the biomass problem could be to separate each sample of macroinvertebrates into groups by size prior to homogenisation and then reassemble the homogenisation products of each group proportionally (Elbrecht et al., 2017, 2021). However, this option is very time-consuming and not very scalable, making it one of the bottlenecks that make sample processing expensive and lengthy (Buchner et al., 2023). Another solution is to increase the number of sequencing reads per sample, which is a cheaper solution today and will become even cheaper as sequencing costs continue to fall (Elbrecht et al., 2021; Guareschi et al., 2017; Wetterstrand, 2023).

6- Primer bias.

The taxonomic groups of macroinvertebrates considered in the IBMWP index fall into 4 phyla (Molluscs, Annelids, Flatworms and Arthropods). The great genetic diversity of this group makes it difficult to develop primers that can amplify all of them without also amplifying other species as bacteria or diatoms. Although degenerate primers have been developed that work quite well, they still generate false negatives for some species or some groups due to the lack of affinity that results in what is known as primer bias.

Example.

The primer pair BF3/BR2 exhibits a known bias towards molluscs (Elbrecht et al., 2019; Szekeres et al., 2022), a bias that is reflected in our data where 4 out of the 5 families more frequently detected by morphological identification than by metabarcoding belong to the phylum Mollusca.

Solution.

There are some possible solutions to this problem. These include multiprimer approach (use different primers pairs with different bias for a more comprehensive amplification), multimarker approach (targeting different markers at the same time as 16S or 12S) or the development of new primer pair. However, any solution to this problem should be considered temporary until the sequences of all macroinvertebrate species on the peninsula become available.

## Other sources of variation.

Not all sources of variation that can explain the difference between the morphological approach and metabarcoding necessarily result in false positives or false negatives in the metabarcoding technique. Instead, the technique offers new information that was not considered in the morphological approach or directly resolves certain issues of the morphological approach:

7- Exotic species consideration.

The presence of exotic species in the rivers of the peninsula is increasing and widespread (Oficialdegui et al., 2023). It is therefore common for them to appear in routine biomonitoring samples. When exotic species are detected and identified in samples, most ecological status indices simply ignore the presence of the species for the index calculation. For summative indices, such as IBMWP (and numerous other BMWP-derived indices), it has been proposed to apply negative values to the detection of these species with values according to their impact on the ecosystem (Guareschi & Wood, 2019). Nevertheless, in order to both give them negative values and not count them as a detection of a native species, it is necessary to correctly identify the invasive species, and this is not always easy to do using a morphological approach.

Example.

Five exotic macroinvertebrate species appear in our data: The North American planarian *Girardia sinensis* (Girard, 1850) belongs to Class Turbellaria, a medium-tolerant family with a score of 5 on the IBMWP index. The Chinese peach blossom jellyfish *Craspedacusta sowerbii* Lankester, 1880, a macroinvertebrate belonging to the phylum Cnidaria which has no category or value in the IBMWP index. The North American signal crayfish *Pacifastacus leniusculus* (Dana, 1852) belongs to the family Astacidae, a non-tolerant family with relative high score on the IBMWP index (8 over 10). The North American tadpole snail *Physella acuta* (Draparnaud, 1805) belongs to the family Physidae, a tolerant family with a score of 3 in the IBMWP index. Finally, the New Zeland mud snail *Potamopyrgus antipodarum* (J.E. Gray, 1843) is not only an invasive species but also an example of taxonomic changes previously mentioned. It was previously classified under the family Hydrobiidae with a value of 3 out of 10 in the index but now belongs to the family Tateidae (Wilke et al., 2013). Although it is no longer considered by the index, it remains a problem because of its morphological similarity to other species in the family. It is usually identified as a member of the Hydrobiidae family by morphological identification, as was the case here.

Solution.

The use of molecular techniques such as metabarcoding is one solution to easily, quickly and correctly identify invasive species present in samples. For morphological identification techniques, in some groups specific identification keys exist and are being developed to help differentiate them from other species in the family, but a barcoding approach could also be included for those specimens that are difficult to resolve morphologically. It is also necessary to review the consideration of the presence of these species in the ecological status indices (Guareschi & Wood, 2019).

8- Identification errors.

Morphological identification techniques can also have false positives and false negatives. One of the sources of such errors is the process of identification by taxonomists as there are certain cryptic species that can be misleading.

Example.

Identifying the species of the Tricladida order can be challenging when ethanol is used as a preservative for the samples. Ethanol inclusion constricts and changes the colour of Tricladidae, making it difficult to observe their identifying characters. Misidentification of specimens from the Dugesiidae and Planariidae families is common, as occurred in our data. Morphological identification identified all Tricladidae in the samples as Planariidae, but molecular identification assigned them to different species of Dugesiidae.

9- Sampling variation.

The official methodology for macroinvertebrate sampling in rivers of the Peninsular Spain (MAGRAMA, 2013b) consists of a multihabitat, stratified and semiquantitative sampling of 20 kicks (sampling units) distributed over a 100 m section of the river. Each kick covers an estimated stream bottom area of 0.125 $m^2$ (semiquantitative). Habitats that represent ≥ 5% of the total surface are sampled (multihabitat). The quantity of samples collected from each habitat is determined by its proportion relative to the entire study area (stratified). This design has been shown to be incapable of capturing all macroinvertebrate biodiversity in a single sample. It was estimated that this sampling methodology can only recover on average between 55% and 83% of the total (family-level) macroinvertebrate diversity of the reach (Jáimez-Cuéllar et al., 2006; Ramos-Merchante & Prenda, 2017). Therefore, it is expected that when two samples are taken at the same site at the same time, different IBMWP taxa may be captured, especially for those taxa with low abundance, with implications not only for the study of diversity but also for the inference of the ecological status of the river (Guareschi et al., 2017). We have not found any studies testing the replicability of this sampling methodology in particular, but we have data from internal calibration exercises (Table XVIII-4 Figure XVIII-4 in Annex V). These exercises show an average similarity in IBMWP taxa composition between 2 sample replicates of only 64% and an average Jaccard dissimilarity of 0.36. These results are consistent with those obtained in the STAR-AQEM intercalibration projects between different sampling techniques in 14 European countries (Bush et al., 2019; M. Furse et al., 2006; Schmidt-Kloiber et al., 2014).

Example.

As mentioned in section 5-, in our data we observed a significant relationship between the abundance of taxa detected in morphological samples and the probability of detection in bulk samples (p-value < 0.01), with taxa detected at lower abundances in morphological samples being the least likely to be detected in bulk samples. This is consistent with the variability introduced due to sampling, as taxa at low abundance in the sampling area are the most likely to suffer from sampling stochasticity, being more likely not to be collected in two samples taken in parallel at the same sampling point.

Solution.

The macroinvertebrate sampling methods were developed to meet WFD requirements, considering the relative abundance of different taxa in the study area to infer the ecological status of the rivers (Jáimez-Cuéllar et al., 2002; Water Framework Directive (2000/60/EC), 2000). Additionally, to facilitate routine biomonitoring, these methodologies also prioritized optimization in terms of time and cost. However, the semi-quantitative and stratified nature of the sampling could present a challenge in developing a protocol that truly reflects the macroinvertebrate biodiversity of the study area. Recognising the limitations of the metabarcoding technique in accurately determining the relative abundance of species, there is an opportunity in exploring alternative sampling methods that prioritise maximising biodiversity representation in each sample, rather than taxon abundance (Elbrecht & Leese, 2015; Friberg et al., 2006; Jáimez-Cuéllar et al., 2006; Ramos-Merchante & Prenda, 2017; Sickel et al., 2023). Additionally, by eliminating stratified sampling, it is possible to avoid redundant sampling in already surveyed habitats, potentially enhancing metabarcoding performance by preventing unnecessary biomass increase of taxa already well-represented in the sample (Beentjes, et al., 2019; Elbrecht et al., 2017).

## The case of water eDNA metabarcoding vs morphological identification.

Macroinvertebrate diversity obtained from eDNA samples is not as similar to that obtained by morphological identification as that obtained from homogenised bulk samples (Figure IX-1) (Gleason et al., 2020; Keck, Blackman, et al., 2022; van der Lee et al., 2024). In our data, considering all shared samples, only 50.6% of taxa was detected by both methodologies, eDNA and morphological identification (Figure IX-2B and Figure XVIII-1). Some of the differences between eDNA and morphological results have the same sources as for bulk samples since they share some of the same processes. However, eDNA has more differences, which could be due to methodological issues such as eDNA sampling, non-specific amplification, or different primer bias (Blackman et al., 2019b; Bush et al., 2019; Leese et al., 2020; Vourka et al., 2023a), but another reason could simply be that the information contained in an eDNA sample is different from that contained in a kicksample. River eDNA studies show that rivers act as conveyor belts carrying not only genetic information from species present upstream of the study point, but also carry DNA from terrestrial species in the landscape surrounding the river (e.g.: Deiner et al., 2016; Reji Chacko et al., 2023). The detection of terrestrial species is considered a false positive, as in the case of bulk samples, and could be filtered out if a checklist of aquatic species is available. However, the eDNA of detected aquatic species may originate from several kilometre upstream (e.g.: Deiner & Altermatt, 2014; Jane et al., 2015; Nukazawa et al., 2018), much further than the upper boundary of the studied river section, which can vary from 10 metres to few hundred metres depending on biomonitoring program (Buss et al.,

2015). According to current biomonitoring standards, these detections from other stretches of the river, or even tributaries, would be considered false positives of the technique. Given the impossibility of filtering out these detections, eDNA would not be a suitable technique for river biomonitoring under current schemes. However, the potential of eDNA to provide information on stretches of river up to several kilometres from the sampling of a single transect could be exploited to develop monitoring systems with a broader scope than current biomonitoring schemes. Spatial scales at which eDNA results seems to be comparable with morphological identification results (Brantschen et al., 2021).

Another possible explanation for the dissimilarity of water eDNA samples is related to the ecology of the eDNA (Barnes & Turner, 2016). These are related not only to transport but also to the production, state, and degradation of eDNA in the river and how these variables could affect the detection of macroinvertebrate composition as well as the repeatability of results (Mauvisseau et al., 2022). In addition, the low concentration of eDNA in the water can compromise the robustness of metabarcoding, making the technique more sensitive to contamination and requiring more extractions and PCR replicates to recover full diversity (Beentjes et al., 2019; Sepulveda et al., 2020; Taberlet et al., 2012). However, thanks to extensive research into the technique, more is being understood about its ecology and applicability, and protocols are being developed to improve its effectiveness (Beentjes et al., 2019; Mauvisseau et al., 2022; Vourka et al., 2023b). This is gradually leading to an increasingly useful and robust technique, which may make it a powerful tool for macroinvertebrate biomonitoring in the future.

## Road Map

Metabarcoding has been shown to be a molecular technique that promises to change the way biomonitoring is done by providing scalability and higher taxonomic resolution (Biomonitoring 2.0). However, the technique still suffers from limitations for its application in the Iberian Peninsula. These limitations are feasible to overcome but require a comprehensive and coordinated effort. To begin with, regardless of the type of sample to be used, it is necessary to have available:

- A comprehensive checklist of freshwater macroinvertebrate species from the Iberian Peninsula.
- Reference sequences for the region of the genome used for the taxonomic assignment of these species (e.g.: Folmer region of the mitochondrial COI gene).

With these two elements, false positives due to the presence of terrestrial species and false negatives due to lack of taxonomic assignment and resolution could be avoid. In addition, synonyms of the different species should be added to the checklist which together with a fuzzy name detection tool such as GBIF species matching, would eliminate all false

negatives and duplicates due to taxonomic conflicts. In parallel, and independently of the development and application of molecular techniques, it is important to review and update existing indices, and to consider the inclusion of non-native and invasive species detection in current indices (Guareschi & Wood, 2019). Once these elements have been achieved, it is possible to consider the possibility of:

- Develop new primers to avoid possible amplification biases.

Given that once all the sequences of all the species to be detected are available, it will be possible to calculate in silico the impact of using currently available primers and to design and test new primers better adapted to peninsular species and optimised for both sample types.

At this point, and now specifically for bulk samples, the following should be considered:

- Develop a new sampling method, optimised to recover maximum specific diversity. Since it seems that reliable abundance data cannot be inferred from macroinvertebrates metabarcoding for the moment, if this limitation persists, the semi-quantitative and stratified characteristics of the current sampling method could be abandoned to focus on practical sampling optimized to recover diversity.
- Conduct intercomparison experiments of metabarcoding laboratory protocols and bioinformatics processing (Blackman et al., 2019b), including establishing appropriate sequencing depth values.
- Development of new indices that take advantage of the new information provided by metabarcoding thanks to its higher taxonomic resolution.

For eDNA samples, the spatial and temporal information on macroinvertebrate composition obtained from a water sample should be studied in more detail, as well as its repeatability and how this can be influenced by environmental variables or river elements such as water flow, presence of dams or obstacles, tributaries rivers, etc. Moreover, it is clear that the information provided by water eDNA samples does not fit into current biomonitoring schemes for stretches of tens or hundreds of metres, but it should not rule out the possibility of using the information provided by water eDNA samples for river biomonitoring in schemes with a wider spatial variable.

# X. Conclusions

This thesis assessed the **readiness**, **accuracy**, and **applicability** of metabarcoding for macroinvertebrate-based river biomonitoring in peninsular Spain, using bulk and eDNA samples. The key conclusions are as follows:

- Metabarcoding is not yet **ready** for direct implementation in river biomonitoring due to several limitations that need to be addressed:
    - A comprehensive checklist and reference genetic sequences for macroinvertebrate species from the Iberian Peninsula are lacking. This gap significantly affects the assessment of the ecological status of rivers.
    - Metabarcoding of bulk samples has shown higher detectability than morphological identification. This increased detectability translates into higher EQR values, requiring the recalibration of index reference values.
    - The spatial scale of current biomonitoring schemes needs to be reconsidered to incorporate the use of water eDNA, as it is likely to provide information on the presence of macroinvertebrates from upstream areas beyond the river reaches defined by official sampling protocols.
- Although metabarcoding of bulk samples is able to capture almost the entire diversity of macroinvertebrate families detected by morphological identification, potentially resolvable methodological issues limit its **accuracy** in detecting and identifying macroinvertebrate diversity in the samples:
    - Presence of errors in taxa identification in reference genetic databases.
    - Primer amplification bias.
    - Lack of reference sequences for some species.
    - Lack of a comprehensive checklist of Iberian freshwater macroinvertebrate species.
    - Lack of standardised sequencing depth values to ensure recovery of specific diversity.
- Metabarcoding, due to its the species-level taxonomic resolution, has been shown to provide better information and valuable additional **applications** when incorporated into routine biomonitoring than morphological identification:
    - It provides information on macroinvertebrate species diversity and allow the detection of invasive and protected species without extra effort.
    - It provides up to four times more taxonomic information on the macroinvertebrate composition than family-level identification, which could lead to the development of more comprehensive river quality indices.

# XI. Conclusiones

Esta tesis evaluó la preparación, exactitud y aplicabilidad del *metabarcoding* para la biomonitorización de macroinvertebrados en ríos de la España peninsular, utilizando muestras de ADN Ambiental (*eDNA*) y muestras *bulk*. Las principales conclusiones son las siguientes:

- El *metabarcoding* aún no está listo para su aplicación directa en la biomonitorización fluvial debido a varias limitaciones que deben ser abordadas:
    - Se carece de un inventario completo y de secuencias genéticas de referencia para las especies de macroinvertebrados de la Península Ibérica. Esta carencia afecta significativamente a la evaluación del estado ecológico de los ríos.
    - El *metabarcoding* de muestras *bulk* ha demostrado una mayor detectabilidad que la identificación morfológica. Esta mayor detectabilidad se traduce en valores de EQR más elevados, lo que obliga a recalibrar los valores de referencia de los índices.
    - Es necesario reconsiderar la escala espacial de los actuales sistemas de evaluación de estado ecológico de los ríos para incorporar el uso del *water eDNA*, ya que es probable que este tipod e muestra esté proporcionando información de la presencia de macroinvertebrados de zonas situadas aguas arriba, más allá del tramo definido por los protocolos oficiales de muestreo.
- Aunque el *metabarcoding* de muestras *bulk* es capaz de capturar casi toda la diversidad de familias de macroinvertebrados detectadas mediante identificación morfológica, existen cuestiones metodológicas potencialmente solucionables que limitan la exactitud en la detección e identificación de la diversidad de macroinvertebrados en las muestras:
    - Presencia de errores en la identificación de taxones en las bases de datos genéticas de referencia.
    - Sesgo de amplificación por PCR.
    - Falta de secuencias de referencia para algunas especies.
    - Falta de un inventario exhaustivo de especies ibéricas de macroinvertebrados de agua dulce.
    - Falta de valores estandarizados de profundidad de secuenciación para asegurar la recuperación de toda la diversidad específica.
- El *metabarcoding*, debido a su resolución taxonómica a nivel de especie, ha demostrado proporcionar mayor cantidad de información y valiosas aplicaciones adicionales cuando se incorpora a la biomonitorización rutinaria que la identificación morfológica:

- Proporciona información sobre la diversidad de especies de macroinvertebrados y permite la detección de especies invasoras y protegidas sin esfuerzo adicional.
- Proporciona hasta cuatro veces más información taxonómica sobre la composición de macroinvertebrados que la identificación a nivel de familia, lo que podría a la elaboración índices de calidad de los ríos más completos.

# XII. Acronyms

1. **ASV**: Amplicon Sequence Variant - Unique DNA sequences identified metabarcoding analysis, differing by as little as one nucleotide.
2. **BIN**: Barcode Index Number System. It clusters sequences using well established algorithms to produce operational taxonomic units that closely correspond to species.
3. **BOLD**: Barcode of Life Data System - A global database for DNA barcoding to facilitate species identification.
4. **BLAST**: Basic Local Alignment Search Tool - An algorithm for comparing biological sequence information to identify similarities.
5. **BQE**: Biological Quality Element - Biological components used to assess the ecological status of water bodies.
6. **BP**: Base Pair. It is a fundamental unit of double-stranded nucleic acids consisting of two nucleobases bound to each other by hydrogen bonds.
7. **CEDEX**: Centro de Estudios y Experimentación de Obras Públicas - A Spanish public research organization specializing in civil engineering and environmental studies.
8. **COI**: Cytochrome C Oxidase I - A mitochondrial gene widely used in DNA barcoding for species identification.
9. **DNA**: Deoxyribonucleic Acid.
10. **eDNA**: Environmental DNA - DNA collected from environmental samples, such as soil, air or water.
11. **ENAC**: National Accreditation Body - The Spanish organization responsible for accrediting testing and calibration laboratories.
12. **GBIF**: Global Biodiversity Information Facility - An international network providing open access to data about global biodiversity.
13. **IAS**: Invasive Alien Species - Non-native species that spread widely and cause harm to the environment, economy, or human health.
14. **IASPT**: Iberian Average Score per Taxon - An index measuring the mean pollution tolerance of macroinvertebrates in Iberian rivers.
15. **IBMWP**: Iberian Biological Monitoring Working Party – A qualitative index for the assessment of water quality on the basis of the presence of macroinvertebrates for Iberian rivers.

16. **ISO**: International Organization for Standardization - An international standard-setting body composed of representatives from various national standards organizations.
17. **METI**: Índice Multimétrico Específico de Tipo – An semiquantitative index used to evaluate the ecological status of Iberian rivers.
18. **NIS**: Non Indigenous Species are those species introduced outside their natural past or present range which might survive and reproduce.
19. **OTU**: Operational Taxonomic Unit.
20. **PCR**: Polymerase Chain Reaction - A method used to amplify a specific segment of DNA, creating thousands to millions of copies.
21. **qPCR**: Quantitative Polymerase Chain Reaction - A molecular technique that amplifies and quantifies a targeted DNA molecule.
22. **R&D**: Research and Development.
23. **RBDs**: River Basin Districts - Areas comprising one or more neighbouring river basins along with their associated groundwater and coastal waters.
24. **RBMP**: River Basin Management Plans - Strategic plans for the protection, improvement, and sustainable use of water resources within a river basin district.
25. **RNA**: Ribonucleic Acid.
26. **TRL**: Technology Readiness Level - A measurement system used to assess the maturity level of a particular technology from conception to deployment.
27. **URA**: Agencia Vasca del Agua - The Basque Water Agency responsible for comprehensive water management in the Basque region.
28. **WFD**: Water Framework Directive. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy

# XIII. References


A. Hawkes, H. (1997). Origin and development of the biologicalmonitoring working party score system. Wat. Res, 32(3), 964–968.

Acreman, M. C., & Ferguson, A. J. D. (2010). Environmental flows and the European Water Framework Directive. Freshwater Biology, 55(1), 32–48. https://doi.org/10.1111/j.1365-2427.2009.02181.x

Agencia Vasca del Agua. (2021). Protocolo de muestreo, análisis y evaluación de fauna bentónica macroinvertebrada en ríos vadeables Código: RW_MACROINVERTEBRADOS_URA_V_3.1.

Aguadé-Gorgorió, G., Arnoldi, J., Barbier, M., & Kéfi, S. (2024). A taxonomy of multiple stable states in complex ecological communities. Ecology Letters, 27(4). https://doi.org/10.1111/ele.14413

Alba-Tercedor, J., Jáimez-Cuéllar, P., Álvarez, M., Avilés, J., Bonada, N., Casas, J., Mellado, A., Ortega, M., Pardo, I., Prat, N., Rieradevall, M., Robles, S., Sáinz-Cantero, C. E., Sánchez-Ortega, A., Suárez, M. L., Toro, M., Vidal-Abarca, R., Vivas, S., & Zamora-Muñoz, C. (2004). Caracterización del estado ecológico de ríos mediterráneos ibéricos mediante el índice IBMWP (antes BMWP'). Limnetica, 21, 175–185.

Alberdi, A., Aizpurua, O., Gilbert, M. T. P., & Bohmann, K. (2018). Scrutinizing key steps for reliable metabarcoding of environmental samples. Methods in Ecology and Evolution, 9(1), 134–147. https://doi.org/10.1111/2041-210X.12849

Alekseev, V. R. (2021). Confusing invader: Acanthocyclops americanus (copepoda: Cyclopoida) and its biological, anthropogenic and climate-dependent mechanisms of rapid distribution in eurasia. Water (Switzerland), 13(10). https://doi.org/10.3390/w13101423

AMBER Consistorium. (2020, June 29). The AMBER Barrier Atlas. A Pan-European database of artificial instream barriers. Version 1.0.

Anderson, L., & Leite, R. (2012). Mitochondrial pseudogenes in insect DNA barcoding: differing points of view on the same issue. In Biota Neotrop (Vol. 12, Issue 3).

Apothéloz-Perret-Gentil, L., Bouchez, A., Cordier, T., Cordonier, A., Guéguen, J., Rimet, F., Vasselon, V., & Pawlowski, J. (2021). Monitoring the ecological status of rivers with diatom eDNA metabarcoding: A comparison of taxonomic markers and analytical approaches for the inference of a molecular diatom index. Molecular Ecology, 30(13), 2959–2968. https://doi.org/10.1111/mec.15646

Orden ARM/2656/2008, de 10 de Septiembre, Por La Que Se Aprueba La Instrucción de Planificación Hidrológica., Pub. L. No. ARM/2656/2008 (2008). https://www.boe.es/eli/es/o/2008/09/10/arm2656

Aylagas, E., Borja, Á., & Rodríguez-Ezpeleta, N. (2014). Environmental status assessment using DNA metabarcoding: Towards a genetics based marine biotic index (gAMBI). PLoS ONE, 9(3). https://doi.org/10.1371/journal.pone.0090529

Baena-Bejarano, N., Reina, C., Martínez-Revelo, D. E., Medina, C. A., Tovar, E., Uribe-Soto, S., Neita-Moreno, J. C., & Gonzalez, M. A. (2023). Taxonomic identification accuracy from BOLD and GenBank databases using over a thousand insect DNA barcodes from Colombia. PLoS ONE, 18(4 4). https://doi.org/10.1371/journal.pone.0277379

Barbour, M. T., Gerritsen, J., Snyder, B. D., & Stribling, J. B. (1998). Rapid Bioassessment Protocols For Use in Streams and Wadeable Rivers: Periphyton, Benthic Macroinvertebrates, and Fish. (Second Edition). U.S. EnvironmentalProtection Agency; Office of Water. http://www.epa.gov/OWOW/monitoring/techmon.html

Barnes, M. A., & Turner, C. R. (2016). The ecology of environmental DNA and implications for conservation genetics. In Conservation Genetics (Vol. 17, Issue 1, pp. 1–17). Springer Netherlands. https://doi.org/10.1007/s10592-015-0775-4

Basak, S. M., Hossain, M. S., Tusznio, J., & Grodzińska-Jurczak, M. (2021). Social benefits of river restoration from ecosystem services perspective: A systematic review. Environmental Science & Policy, 124, 90–100. https://doi.org/10.1016/j.envsci.2021.06.005

Batovska, J., Piper, A. M., Valenzuela, I., Cunningham, J. P., & Blacket, M. J. (2021). Developing a non - destructive metabarcoding protocol for detection of pest insects in bulk trap catches. Scientific Reports, 0123456789, 1–14. https://doi.org/10.1038/s41598-021-85855-6

Beentjes, K. K., Speksnijder, A. G. C. L., Schilthuizen, M., Hoogeveen, M., Pastoor, R., & van der Hoorn, B. B. (2019). Increased performance of DNA metabarcoding of macroinvertebrates by taxonomic sorting. PLoS ONE, 14(12), 1–19. https://doi.org/10.1371/journal.pone.0226527

Beentjes, K. K., Speksnijder, A. G. C. L., Schilthuizen, M., Hoogeveen, M., & van der Hoorn, B. B. (2019). The effects of spatial and temporal replicate sampling on eDNA metabarcoding. PeerJ, 7(7), e7335. https://doi.org/10.7717/peerj.7335

Beentjes, K. K., Speksnijder, A. G. C. L., Schilthuizen, M., Schaub, B. E. M., & Van der Hoorn, B. B. (2018). The influence of macroinvertebrate abundance on the assessment of freshwater quality in the Netherlands. Metabarcoding and Metagenomics, 2. https://doi.org/10.3897/mbmg.2.26744

Beermann, A., Buchner, D., Leese, F., Macher, T.-H., Ocadlik, M., & Weigand, A. (2021). New insights into Danube's macroinvertebrate communities from DNA metabarcoding as part of the Joint Danube Survey 4 (JDS4). ARPHA Conference Abstracts, 4. https://doi.org/10.3897/aca.4.e65206

Beermann, A. J., Elbrecht, V., Karnatz, S., Ma, L., Matthaei, C. D., Piggott, J. J., & Leese, F. (2018). Multiple-stressor effects on stream macroinvertebrate communities: A mesocosm experiment manipulating salinity, fine sediment and flow velocity. Science of the Total Environment, 610–611, 961–971. https://doi.org/10.1016/j.scitotenv.2017.08.084

Behrens-Chapuis, S., Herder, F., & Geiger, M. F. (2021). Adding DNA barcoding to stream monitoring protocols – What's the additional value and congruence between morphological and molecular identification approaches? PLoS ONE, 16(1 January). https://doi.org/10.1371/journal.pone.0244598

Belletti, B., Garcia de Leaniz, C., Jones, J., Bizzi, S., Börger, L., Segura, G., Castelletti, A., van de Bund, W., Aarestrup, K., Barry, J., Belka, K., Berkhuysen, A., Birnie-Gauvin, K., Bussettini, M., Carolli, M., Consuegra, S., Dopico, E., Feierfeil, T., Fernández, S., … Zalewski, M. (2020). More than one million barriers fragment Europe's rivers. Nature, 588(7838), 436–441. https://doi.org/10.1038/s41586-020-3005-2

Beng, K. C., & Corlett, R. T. (2020). Applications of environmental DNA (eDNA) in ecology and conservation: opportunities, challenges and prospects. In Biodiversity and Conservation (Vol. 29, Issue 7, pp. 2089–2121). Springer. https://doi.org/10.1007/s10531-020-01980-0

Benítez-Álvarez, L., Sluys, R., Leal-Zanchet, A. M., Leria, L., & Riutort, M. (2023). First molecular phylogeny of the freshwater planarian genus Girardia (Platyhelminthes: Tricladida) unveils hidden taxonomic diversity and initiates resolution of its historical biogeography. Zoological Journal of the Linnean Society, 197(1), 1–19. https://doi.org/10.1093/zoolinnean/zlac065

Bennett, C., Owen, R., Birk, S., Buffagni, A., Erba, S., Mengin, N., Murray-Bligh, J., Ofenböck, G., Pardo, I., van de Bund, W., Wagner, F., & Wasson, J. G. (2011). Bringing European river quality into line: An exercise to intercalibrate macro-invertebrate classification methods. Hydrobiologia, 667(1), 31–48. https://doi.org/10.1007/s10750-011-0635-2

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2013a). GenBank. Nucleic Acids Research, 41(D1). https://doi.org/10.1093/nar/gks1195

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2013b). GenBank. Nucleic Acids Research, 41(D1), 36–42. https://doi.org/10.1093/nar/gks1195

Berger, E., Haase, P., Schäfer, R. B., & Sundermann, A. (2018). Towards stressor-specific macroinvertebrate indices: Which traits and taxonomic groups are associated with vulnerable and tolerant taxa? Science of the Total Environment, 619–620, 144–154. https://doi.org/10.1016/j.scitotenv.2017.11.022

Bergsten, J., Bilton, D. T., Fujisawa, T., Elliott, M., Monaghan, M. T., Balke, M., Hendrich, L., Geijer, J., Herrmann, J., Foster, G. N., Ribera, I., Nilsson, A. N., Barraclough, T. G., & Vogler, A. P. (2012). The effect of geographical scale of sampling on DNA barcoding. Systematic Biology, 61(5), 851–869. https://doi.org/10.1093/sysbio/sys037

Bernos, T. A., Yates, M. C., Docker, M. F., Fitzgerald, A., Hanner, R., Heath, D., Imrit, A., Livernois, J., Myler, E., Patel, K., Sharma, S., Young, R., & Mandrak, N. E. (2023). Environmental DNA (eDNA) applications in freshwater fisheries management and conservation in Canada: overview of current challenges and opportunities. Canadian Journal of Fisheries and Aquatic Sciences, 80(7), 1170–1186. https://doi.org/10.1139/cjfas-2022-0162

Birk, S., Böhmer, J., & Schöll, F. (2018). Intercalibrating the national classifications of ecological status for very large rivers in Europe Biological Quality Element: Benthic invertebrates. https://doi.org/10.2760/443119

Birk, S., Bonne, W., Borja, A., Brucet, S., Courrat, A., Poikane, S., Solimini, A., van de Bund, W., Zampoukas, N., & Hering, D. (2012a). Three hundred ways to assess Europe's surface waters: An almost complete overview of biological methods to implement the Water Framework Directive. Ecological Indicators, 18, 31–41. https://doi.org/10.1016/j.ecolind.2011.10.009

Birk, S., Bonne, W., Borja, A., Brucet, S., Courrat, A., Poikane, S., Solimini, A., van de Bund, W., Zampoukas, N., & Hering, D. (2012b). Three hundred ways to assess Europe's surface waters: An almost complete overview of biological methods to implement the Water Framework Directive. Ecological Indicators, 18, 31–41. https://doi.org/10.1016/j.ecolind.2011.10.009

Blackman, R. C., Constable, D., Hahn, C., Sheard, A. M., Durkota, J., Hänfling, B., & Handley, L. L. (2017). Detection of a new non-native freshwater species by DNA metabarcoding of environmental samples – first record of gammarus fossarum in the UK. Aquatic Invasions, 12(2), 177–189. https://doi.org/10.3391/ai.2017.12.2.06

Blackman, R. C., Mächler, E., Altermatt, F., Arnold, A., Beja, P., Boets, P., Egeter, B., Elbrecht, V., Filipe, A. F., Iwan Jones, J., Macher, J., Majaneva, M., Martins, F. M. S., Múrria, C.,

Meissner, K., Pawlowski, J., Schmidt Yáñez, P. L., Zizka, V. M. A., Leese, F., … Deiner, K. (2019a). Advancing the use of molecular methods for routine freshwater macroinvertebrate biomonitoring - The need for calibration experiments. Metabarcoding and Metagenomics, 3, 49–57. https://doi.org/10.3897/mbmg.3.34735

Blackman, R. C., Mächler, E., Altermatt, F., Arnold, A., Beja, P., Boets, P., Egeter, B., Elbrecht, V., Filipe, A. F., Iwan Jones, J., Macher, J., Majaneva, M., Martins, F. M. S., Múrria, C., Meissner, K., Pawlowski, J., Schmidt Yáñez, P. L., Zizka, V. M. A., Leese, F., … Deiner, K. (2019b). Advancing the use of molecular methods for routine freshwater macroinvertebrate biomonitoring - The need for calibration experiments. Metabarcoding and Metagenomics, 3, 49–57. https://doi.org/10.3897/mbmg.3.34735

Blancher, P., Lefrançois, E., Rimet, F., Vasselon, V., Argillier, C., Arle, J., Beja, P., Boets, P., Boughaba, J., Chauvin, C., Deacon, M., Duncan, W., Ejdung, G., Erba, S., Ferrari, B., Fischer, H., Hänfling, B., Haldin, M., Hering, D., … Bouchez, A. (2022). A strategy for successful integration of DNA-based methods in aquatic monitoring. Metabarcoding and Metagenomics, 6, 215–226. https://doi.org/10.3897/mbmg.6.85652

Bogan, M. T., Boersma, K. S., & Lytle, D. A. (2013). Flow intermittency alters longitudinal patterns of invertebrate diversity and assemblage composition in an arid-land stream network. Freshwater Biology, 58(5), 1016–1028. https://doi.org/10.1111/fwb.12105

Bogan, M. T., Boersma, K. S., & Lytle, D. A. (2015). Resistance and resilience of invertebrate communities to seasonal and supraseasonal drought in arid-land headwater streams. Freshwater Biology, 60, 2547–2558. https://doi.org/10.1111/fwb.12522

Bohmann, K., Elbrecht, V., Carøe, C., Bista, I., Leese, F., Bunce, M., Yu, D. W., Seymour, M., Dumbrell, A. J., & Creer, S. (2022). Strategies for sample labelling and library preparation in DNA metabarcoding studies. Molecular Ecology Resources, 22(4), 1231–1246. https://doi.org/10.1111/1755-0998.13512

BOLD SYSTEMS. (2019). Barcode of Life Data Systems Handbook. (4th ed.). www.boldsystems.org

Brantschen, J., Blackman, R. C., Walser, J.-C., & Altermatt, F. (2021). Environmental DNA gives comparable results to morphology-based indices of macroinvertebrates in a large-scale ecological assessment. PLOS ONE, 16(9), e0257510. https://doi.org/10.1371/journal.pone.0257510

Braukmann, T. W. A., Ivanova, N. V., Prosser, S. W. J., Elbrecht, V., Steinke, D., Ratnasingham, S., de Waard, J. R., Sones, J. E., Zakharov, E. V., & Hebert, P. D. N. (2019). Metabarcoding a diverse arthropod mock community. Molecular Ecology Resources, 19(3), 711–727. https://doi.org/10.1111/1755-0998.13008

Brooks, A. J., Russell, M., Bevitt, R., & Dasey, M. (2011). Constraints on the recovery of invertebrate assemblages in a regulated snowmelt river during a tributary-sourced environmental flow regime. Marine and Freshwater Research, 62(12), 1407–1420. https://doi.org/10.1071/MF11128

Bruce, K., Blackman, R., Bourlat, S. J., Hellström, A. M., Bakker, J., Bista, I., Bohmann, K., Bouchez, A., Brys, R., Clark, K., Elbrecht, V., Fazi, S., Fonseca, V., Hänfling, B., Leese, F., Mächler, E., Mahon, A. R., Meissner, K., Panksep, K., … Deiner, K. (2021). A practical guide to DNA-based methods for biodiversity assessment. Pensoft Publishers. https://doi.org/10.3897/ab.e68634

Buchner, D., Beermann, A., Hörren, T., Enss, J., Frenzel, M., Li, Y., Müller, J., Pauls, S. U., Sorg, M., consortium, L., Haase, P., & Leese, F. (2023). German-wide Malaise trap metabarcoding estimates over 33,000 insect species. BioRxiv. https://doi.org/10.1101/2023.05.04.539402

Buchner, D., Beermann, A. J., Laini, A., Rolauffs, P., Vitecek, S., Hering, D., & Leese, F. (2019). Analysis of 13,312 benthic invertebrate samples from German streams reveals minor deviations in ecological status class between abundance and presence/absence data. PLOS ONE, 14(12), e0226547. https://doi.org/10.1371/journal.pone.0226547

Buchner, D., Beermann, A. J., Leese, F., & Weiss, M. (2021). Cooking small and large portions of “biodiversity-soup”: Miniaturized DNA metabarcoding PCRs perform as good as large-volume PCRs. Ecology and Evolution, 11(13), 9092–9099. https://doi.org/10.1002/ece3.7753

Buchner, D., Haase, P., & Leese, F. (2021). Wet grinding of invertebrate bulk samples - A scalable and cost-efficient protocol for metabarcoding and metagenomics. Metabarcoding and Metagenomics, 5, 73–81. https://doi.org/10.3897/MBMG.5.67533

Buchner, D., & Leese, F. (2020). BOLDigger - a Python package to identify and organise sequences with the Barcode of Life Data systems. Metabarcoding and Metagenomics, 4, 19–21. https://doi.org/10.3897/mbmg.4.53535

Buchner, D., Macher, T. H., Beermann, A. J., Werner, M. T., & Leese, F. (2021). Standardized high-throughput biomonitoring using DNA metabarcoding: Strategies for the adoption of automated liquid handlers. Environmental Science and Ecotechnology, 8. https://doi.org/10.1016/j.ese.2021.100122

Buchner, D., Macher, T. H., & Leese, F. (2022). APSCALE: advanced pipeline for simple yet comprehensive analyses of DNA metabarcoding data. Bioinformatics, 38(20), 4817–4819. https://doi.org/10.1093/bioinformatics/btac588

Busch, U., & Nitschko, H. (1999). Methods for the differentiation of microorganisms. Journal of Chromatography B: Biomedical Sciences and Applications, 722(1–2), 263–278. https://doi.org/10.1016/S0378-4347(98)00369-7

Bush, A., Compson, Z. G., Monk, W. A., Porter, T. M., Steeves, R., Emilson, E., Gagne, N., Hajibabaei, M., Roy, M., & Baird, D. J. (2019). Studying Ecosystems With DNA Metabarcoding: Lessons From Biomonitoring of Aquatic Macroinvertebrates. In Frontiers in Ecology and Evolution (Vol. 7). Frontiers Media S.A. https://doi.org/10.3389/fevo.2019.00434

Buss, D. F., Carlisle, D. M., Chon, T. S., Culp, J., Harding, J. S., Keizer-Vlek, H. E., Robinson, W. A., Strachan, S., Thirion, C., & Hughes, R. M. (2015). Stream biomonitoring using macroinvertebrates around the globe: a comparison of large-scale programs. Environmental Monitoring and Assessment, 187(1). https://doi.org/10.1007/s10661-014-4132-8

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. BMC Bioinformatics, 10. https://doi.org/10.1186/1471-2105-10-421

Carter, J. L., & Resh, V. H. (2001). After site selection and before data analysis: sampling, sorting, and laboratory procedures used in stream benthic macroinvertebrate monitoring programs by USA state agencies. Journal of the North American Benthological Society, 20(4), 658–682. https://doi.org/10.2307/1468095

Carvalho, J., Garrido-Maestu, A., Azinheiro, S., Fuciños, P., Barros-Velázquez, J., De Miguel, R. J., Gros, V., & Prado, M. (2021). Faster monitoring of the invasive alien species (IAS) Dreissena polymorpha in river basins through isothermal amplification. Scientific Reports, 11(1). https://doi.org/10.1038/s41598-021-89574-w

Cary, S. C., Coyne, K. J., Rueckert, A., Wood, S. A., Kelly, S., Gemmill, C. E. C., Vieglais, C., & Hicks, B. J. (2014). Development and validation of a quantitative PCR assay for the early detection and monitoring of the invasive diatom Didymosphenia geminata. Harmful Algae, 36, 63–70. https://doi.org/10.1016/j.hal.2014.04.003

CEDEX. (2004). Caracterización de los tipos de ríos y lagos.

CEDEX. (2007). Análisis de la correspondencia entre los tipos establecidos en la Caracterización de los ríos en España (v. 4.0, junio de 2005) y los definidos para el Ejercicio de Intercalibración Europea (GIGs de los ríos Centrales/Bálticos, Alpinos y Mediterráneos).

Chester, E. T., & Robson, B. J. (2011). Drought refuges, spatial scale and recolonisation by invertebrates in non-perennial streams. Freshwater Biology, 56(10), 2094–2104. https://doi.org/10.1111/j.1365-2427.2011.02644.x

Cordier, T., Alonso-Sáez, L., Apothéloz-Perret-Gentil, L., Aylagas, E., Bohan, D. A., Bouchez, A., Chariton, A., Creer, S., Frühe, L., Keck, F., Keeley, N., Laroche, O., Leese, F., Pochon, X., Stoeck, T., Pawlowski, J., & Lanzén, A. (2021). Ecosystems monitoring powered by environmental genomics: A review of current strategies with an implementation roadmap. Molecular Ecology, 30(13), 2937–2958. https://doi.org/10.1111/mec.15472

Cota, L., Goulart, M., Moreno, P., & Callisto, M. (2002). Rapid assessment of river water quality using an adapted BMWP index: a practical tool to evaluate ecosystem health. SIL Proceedings, 1922-2010, 28(4), 1713–1716. https://doi.org/10.1080/03680770.2001.11901915

Csabai, Z., Čiamporová-Zaťovičová, Z., Boda, P., & Čiampor, F. (2023). 50%, not great, not terrible: Pan-European gap-analysis shows the real status of the DNA barcode reference libraries in two aquatic invertebrate groups and points the way ahead. Science of the Total Environment, 863. https://doi.org/10.1016/j.scitotenv.2022.160922

Datry, T., Larned, S. T., Fritz, K. M., Bogan, M. T., Wood, P. J., Meyer, E. I., & Santos, A. N. (2014). Broad-scale patterns of invertebrate richness and community composition in temporary rivers: Effects of flow intermittence. Ecography, 37(1), 94–104. https://doi.org/10.1111/j.1600-0587.2013.00287.x

Deiner, K., & Altermatt, F. (2014). Transport distance of invertebrate environmental DNA in a natural river. PLoS ONE, 9(2). https://doi.org/10.1371/journal.pone.0088786

Deiner, K., Bik, H. M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., Creer, S., Bista, I., Lodge, D. M., de Vere, N., Pfrender, M. E., & Bernatchez, L. (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. Molecular Ecology, 26(21), 5872–5895. https://doi.org/10.1111/mec.14350

Deiner, K., Fronhofer, E. A., Mächler, E., Walser, J. C., & Altermatt, F. (2016). Environmental DNA reveals that rivers are conveyer belts of biodiversity information. Nature Communications, 7. https://doi.org/10.1038/ncomms12544

DeSalle, R., & Goldstein, P. (2019). Review and Interpretation of Trends in DNA Barcoding. Frontiers in Ecology and Evolution, 7. https://doi.org/10.3389/fevo.2019.00302

Dewson, Z. S., James, A. B. W., & Death, R. G. (2007). A review of the consequences of decreased flow for instream habitat and macroinvertebrates. In Journal of the North American Benthological Society (Vol. 26, Issue 3, pp. 401–415). https://doi.org/10.1899/06-110.1

Di Sabatino, A., Coscieme, L., & Cristiano, G. (2023). No post-drought recovery of the macroinvertebrate community after five months upon rewetting of an irregularly

intermittent Apennine River (Aterno River). Ecohydrology and Hydrobiology, 23(1), 141–151. https://doi.org/10.1016/j.ecohyd.2022.11.005

Dolédec, S., Tilbian, J., & Bonada, N. (2017). Temporal variability in taxonomic and trait compositions of invertebrate assemblages in two climatic regions with contrasting flow regimes. Science of the Total Environment, 599–600, 1912–1921. https://doi.org/10.1016/j.scitotenv.2017.05.057

Döll, P., Fiedler, K., & Zhang, J. (2009). Hydrology and Earth System Sciences Global-scale analysis of river flow alterations due to water withdrawals and reservoirs. In Hydrol. Earth Syst. Sci (Vol. 13). www.hydrol-earth-syst-sci.net/13/2413/2009/

Doretto, A., Bona, F., Falasco, E., Morandini, D., Piano, E., & Fenoglio, S. (2020). Stay with the flow: How macroinvertebrate communities recover during the rewetting phase in Alpine streams affected by an exceptional drought. River Research and Applications, 36(1), 91–101. https://doi.org/10.1002/rra.3563

Dudgeon, D., Arthington, A. H., Gessner, M. O., Kawabata, Z. I., Knowler, D. J., Lévêque, C., Naiman, R. J., Prieur-Richard, A. H., Soto, D., Stiassny, M. L. J., & Sullivan, C. A. (2006). Freshwater biodiversity: Importance, threats, status and conservation challenges. In Biological Reviews of the Cambridge Philosophical Society (Vol. 81, Issue 2, pp. 163–182). https://doi.org/10.1017/S1464793105006950

Elbrecht, V. (2017). Development of DNA metabarcoding methods for stream ecosystem assessment.

Elbrecht, V., Bourlat, S. J., Hörren, T., Lindner, A., Mordente, A., Noll, N. W., Schäffler, L., Sorg, M., & Zizka, V. M. A. (2021). Pooling size sorted Malaise trap fractions to maximize taxon recovery with metabarcoding. PeerJ, 9, e12177. https://doi.org/10.7717/peerj.12177

Elbrecht, V., Braukmann, T. W. A., Ivanova, N. V., Prosser, S. W. J., Hajibabaei, M., Wright, M., Zakharov, E. V., Hebert, P. D. N., & Steinke, D. (2019). Validation of COI metabarcoding primers for terrestrial arthropods. PeerJ, 2019(10). https://doi.org/10.7717/peerj.7745

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, 10(7), 1–16. https://doi.org/10.1371/journal.pone.0130324

Elbrecht, V., & Leese, F. (2017a). Validation and development of COI metabarcoding primers for freshwater macroinvertebrate bioassessment. Frontiers in Environmental Science, 5(APR). https://doi.org/10.3389/fenvs.2017.00011

Elbrecht, V., & Leese, F. (2017b). Validation and development of COI metabarcoding primers for freshwater macroinvertebrate bioassessment. Frontiers in Environmental Science, 5(APR), 1–11. https://doi.org/10.3389/fenvs.2017.00011

Elbrecht, V., Peinert, B., & Leese, F. (2017). Sorting things out: Assessing effects of unequal specimen biomass on DNA metabarcoding. Ecology and Evolution, 7(17), 6918–6926. https://doi.org/10.1002/ece3.3192

Elbrecht, V., & Steinke, D. (2019). Scaling up DNA metabarcoding for freshwater macrozoobenthos monitoring. Freshwater Biology, 64(2), 380–387. https://doi.org/10.1111/fwb.13220

Elbrecht, V., Taberlet, P., Dejean, T., Valentini, A., Usseglio-Polatera, P., Beisel, J. N., Coissac, E., Boyer, F., & Leese, F. (2016). Testing the potential of a ribosomal 16S marker for DNA metabarcoding of insects. PeerJ, 2016(4). https://doi.org/10.7717/peerj.1966

Emmons, S. C., Compson, Z. G., Malish, M. C., Busch, M. H., Saenz, V., Higgins, K. T., & Allen, D. C. (2023). DNA metabarcoding captures different macroinvertebrate biodiversity than morphological identification approaches across a continental scale. Environmental DNA. https://doi.org/10.1002/edn3.453

Espinosa Prieto, A., Hardion, L., Debortoli, N., & Beisel, J. (2024). Finding the perfect pairs: A matchmaking of plant markers and primers for multi-marker eDNA metabarcoding. Molecular Ecology Resources. https://doi.org/10.1111/1755-0998.13937

European Commission. (2003). River and lakes – Typology, reference conditions and classification systems. http://zotero.org/support/quick_start_guide

European Commission. (2014). Horizon 2020 – Work Programme 2014-2015–General Annexes, Technology Readiness Levels (TRL). https://ec.europa.eu/research/participants/data/ref/h2020/wp/2014_2015/annexes/h2020-wp1415-annex-g-trl_en.pdf

European Environment Agency. (2021). Drivers of and pressures arising from selected key water management challenges a European overview. https://doi.org/10.2800/059069

European Environmental Agency. (2018). European waters - assessment of status and pressures 2018. https://doi.org/10.2800/303664

FAO. (2021). A review of major river basins and large lakes relevant to inland fisheries. FAO. https://doi.org/10.4060/cb2827en

Fernández, S., Rodríguez, S., Martínez, J. L., Borrell, Y. J., Ardura, A., & García-Vázquez, E. (2018). Evaluating freshwater macroinvertebrates from eDNA metabarcoding: A river Nalón case study. PLoS ONE, 13(8). https://doi.org/10.1371/journal.pone.0201741

Fernández, S., Rodríguez-Martínez, S., Martínez, J. L., Garcia-Vazquez, E., & Ardura, A. (2019). How can eDNA contribute in riverine macroinvertebrate assessment? A metabarcoding approach in the Nalón River (Asturias, Northern Spain). Environmental DNA, 1(4), 385–401. https://doi.org/10.1002/edn3.40

Fernández, S., Rodríguez-Martínez, S., Martínez, J. L., Garcia-Vazquez, E., & Ardura, A. (2019). How can eDNA contribute in riverine macroinvertebrate assessment? A metabarcoding approach in the Nalón River (Asturias, Northern Spain). Environmental DNA, 1(4), 385–401. https://doi.org/10.1002/edn3.40

Ferraro, S. P., & Cole, F. A. (1992). Taxonomic Level Sufficient for Assessing a Moderate Impact on Macrobenthic Communities in Puget Sounds Washington, USA. Canadian Journal of Fisheries and Aquatic Sciences, 49(6), 1184–1188. https://doi.org/10.1139/f92-133

Ferreira, S. A., Andrade, R., Gonçalves, A. R., Sousa, P., Paupério, J., Fonseca, N. A., & Beja, P. (2020). The InBIO Barcoding Initiative Database: DNA barcodes of Portuguese Diptera 01. Biodiversity Data Journal, 8. https://doi.org/10.3897/BDJ.8.e49985

Ferreira, S. A., de Figueroa, J. M. T., Martins, F. M. S., Verissimo, J., Quaglietta, L., Grosso-Silva, J. M., Lopes, P. B., Sousa, P., Paupério, J., Fonseca, N. A., & Beja, P. (2020). The InBIO barcoding initiative database: Contribution to the knowledge on DNA barcodes of Iberian Plecoptera. Biodiversity Data Journal, 8. https://doi.org/10.3897/BDJ.8.e55137

Ferreira, S. A., Oosterbroek, P., Starý, J., Sousa, P., Mata, V. A., da Silva, L. P., Paupério, J., & Beja, P. (2021). The InBIO Barcoding Initiative Database: DNA barcodes of Portuguese Diptera 02 - Limoniidae, Pediciidae and Tipulidae. Biodiversity Data Journal, 9, 1–29. https://doi.org/10.3897/BDJ.9.e69841

Ferreira, S. A., Paupério, J., Grosso-Silva, J. M., & Beja, P. (2019). DNA barcoding of Sialis sp. (Megaloptera) in Portugal: the missing tool to species identification. Aquatic Insects, 40(2), 173–184. https://doi.org/10.1080/01650424.2019.1571612

Ficetola, G. F., & Taberlet, P. (2023). Towards exhaustive community ecology via DNA metabarcoding. Molecular Ecology. https://doi.org/10.1111/mec.16881

Fiel, R. A., Narganes, A. g., & Braña Arguelles, M. B. (2014). Incidencia de Drosophila suzukii en cultivos de arándano y frambuesa en Asturias. https://www.phytoma.com/la-revista/phytohemeroteca/258-abril-2014/incidencia-de-drosophila-suzukii-en-cultivos-de-arndano-y-frambuesa-en-asturias

Fluet-Chouinard, E., Funge-Smith, S., & McIntyre, P. B. (2018). Global hidden harvest of freshwater fish revealed by household surveys. Proceedings of the National Academy

of Sciences of the United States of America, 115(29), 7623–7628. https://doi.org/10.1073/pnas.1721097115

Folmer, O., Black, M., Hoeh, W., Lutz, R., & Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology, 3(5), 294–299. https://doi.org/10.1071/ZO9660275

Fonseca, V. G., Davison, P. I., Creach, V., Stone, D., Bass, D., & Tidbury, H. J. (2023). The Application of eDNA for Monitoring Aquatic Non-Indigenous Species: Practical and Policy Considerations. In Diversity (Vol. 15, Issue 5). MDPI. https://doi.org/10.3390/d15050631

Fontes, J. T., Vieira, P. E., Ekrem, T., Soares, P., & Costa, F. O. (2020). BAGS: An automated Barcode, Audit & Grade System for DNA barcode reference libraries. Molecular Ecology Resources, May, 1–11. https://doi.org/10.1111/1755-0998.13262

Friberg, N., Sandin, L., Furse, M. T., Larsen, S. E., Clarke, R. T., & Haase, P. (2006). Comparison of macroinvertebrate sampling methods in Europe. Hydrobiologia, 566(1), 365–378. https://doi.org/10.1007/s10750-006-0083-6

Fueyo, Á., Sánchez, O., Coya, R., Carleos, C., Escudero, A., Cordón, J., Fernández, S., Granero-Castro, J., & Borrell, Y. J. (2024). The influence of databases enrichment using local macroinvertebrate genetic references for metabarcoding based biodiversity studies in river monitoring. Ecological Indicators, 158, 111454. https://doi.org/10.1016/j.ecolind.2023.111454

Fueyo, Á., Sánchez-Fernández, O., Carleos, C., Escudero Marina, A., Cordón Ezquerro, J., Granero-Castro, J., & Borrell Pichs, Y. J. (2024). Unlocking Rivers Hidden Diversity and Ecological Status using DNA Metabarcoding in Northwest Spain. Authorea. https://doi.org/10.22541/au.171665353.30022618/v1

Fuller, M. R., Doyle, M. W., & Strayer, D. L. (2015). Causes and consequences of habitat fragmentation in river networks. Annals of the New York Academy of Sciences, 1355(1), 31–51. https://doi.org/10.1111/nyas.12853

Furse, M., Hering, D., Brabec, K., Buffagni, A., Sandin, L., & Verdonschot, P. F. M. (2006). The Ecological Status of European Rivers: Evaluation and Intercalibration of Assessment Methods (M. T. Furse, D. Hering, K. Brabec, A. Buffagni, L. Sandin, & P. F. M. Verdonschot, Eds.; 1st ed.). Springer Netherlands. https://doi.org/10.1007/978-1-4020-5493-8

Furse, M., Hering, D., Moog, O., Verdonschot, P., Johnson, R. K., Brabec, K., Gritzalis, K., Buffagni, A., Pinto, P., Friberg, N., Murray-Bligh, J., Kokes, J., Alber, R., Usseglio-Polatera, P., Haase, P., Sweeting, R., Bis, B., Szoszkiewicz, K., Soszka, H., … Krno, I.

(2006). The STAR project: context, objectives and approaches. In The Ecological Status of European Rivers: Evaluation and Intercalibration of Assessment Methods (pp. 3–29). Springer Netherlands. https://doi.org/10.1007/978-1-4020-5493-8_2

Furse, M. T., Wright, J. F., Armitage, P. D., & Moss, D. (1981). An appraisal of pond-net samples for biological monitoring of lotic macro-invertebrates. Water Research, 15(6), 679–689. https://doi.org/10.1016/0043-1354(81)90160-3

Garrett, N. R., Watkins, J., Simmons, N. B., Fenton, B., Maeda-Obregon, A., Sanchez, D. E., Froehlich, E. M., Walker, F. M., Littlefair, J. E., & Clare, E. L. (2023). Airborne eDNA documents a diverse and ecologically complex tropical bat and other mammal community. Environmental DNA, 5(2), 350–362. https://doi.org/10.1002/edn3.385

Gaytán, Á., Bergsten, J., Canelo, T., Pérez-Izquierdo, C., Santoro, M., & Bonal, R. (2020). DNA Barcoding and geographical scale effect: The problems of undersampling genetic diversity hotspots. Ecology and Evolution, 10(19), 10754–10772. https://doi.org/10.1002/ece3.6733

GBIF. (2024). Página de Inicio de GBIF. https://www.gbif.org

Ge, Y., Xia, C., Wang, J., Zhang, X., Ma, X., & Zhou, Q. (2021). The efficacy of DNA barcoding in the classification, genetic differentiation, and biodiversity assessment of benthic macroinvertebrates. Ecology and Evolution, 11(10), 5669–5681. https://doi.org/10.1002/ece3.7470

Geller, J., Meyer, C., Parker, M., & Hawk, H. (2013). Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys. Molecular Ecology Resources, 13(5), 851–861. https://doi.org/10.1111/1755-0998.12138

Gleason, J. E., Elbrecht, V., Braukmann, T. W. A., Hanner, R. H., & Cottenie, K. (2020). Assessment of stream macroinvertebrate communities with eDNA is not congruent with tissue-based metabarcoding. Molecular Ecology. https://doi.org/10.1111/mec.15597

Growns, I. (2016). The implementation of an environmental flow regime results in ecological recovery of regulated rivers. Restoration Ecology, 24(3), 406–414. https://doi.org/10.1111/rec.12330

Guareschi, S., Laini, A., & Sánchez-Montoya, M. M. (2017). How do low-abundance taxa affect river biomonitoring? Exploring the response of different macroinvertebrate-based indices. Journal of Limnology, 76(S1), 9–20. https://doi.org/10.4081/jlimnol.2016.1516

Guareschi, S., & Wood, P. J. (2019). Taxonomic changes and non-native species: An overview of constraints and new challenges for macroinvertebrate-based indices calculation in river ecosystems. Science of the Total Environment, 660, 40–46. https://doi.org/10.1016/j.scitotenv.2019.01.008

Gunn, I. D. M., Carvalho, L., Davies, C. E., Edwards, F. K., Furse, M. T., Maitland, P. S., Raper, C., Siriwardena, G. M., & Winfield, I. J. (2018). UK Checklist of freshwater species. NERC Environmental Information Data Centre. https://doi.org/10.5285/57653719-434b-4b11-9f0d-3bd76054d8bd

Gustafsson, S., Österling, M., Skurdal, J., Schneider, L. D., & Calles, O. (2013). Macroinvertebrate colonization of a nature-like fishway: The effects of adding habitat heterogeneity. Ecological Engineering, 61(PA), 345–353. https://doi.org/10.1016/j.ecoleng.2013.09.023

Haase, P., Murray-Bligh, J., Lohse, S., Pauls, S., Sundermann, A., Gunn, R., & Clarke, R. (2006). Assessing the impact of errors in sorting and identifying macroinvertebrate samples. Hydrobiologia, 566(1), 505–521. https://doi.org/10.1007/s10750-006-0075-6

Haase, P., Pauls, S. U., Schindehütte, K., & Sundermann, A. (2010). First audit of macroinvertebrate samples from an EU Water Framework Directive monitoring program: Human error greatly lowers precision of assessment results. Journal of the North American Benthological Society, 29(4), 1279–1291. https://doi.org/10.1899/09-183.1

Hajibabaei, M., Porter, T. M., Robinson, C. V., Baird, D. J., Shokralla, S., & Wright, M. T. G. (2019). Watered-down biodiversity? A comparison of metabarcoding results from DNA extracted from matched water and bulk tissue biomonitoring samples. PLoS ONE, 14(12), 1–16. https://doi.org/10.1371/journal.pone.0225409

Hanna, D. E. L., Tomscha, S. A., Ouellet Dallaire, C., & Bennett, E. M. (2018). A review of riverine ecosystem service quantification: Research gaps and recommendations. Journal of Applied Ecology, 55(3), 1299–1311. https://doi.org/10.1111/1365-2664.13045

Hawkes, H. A. (1997). Origin and development of the biological monitoring working party score system. Wat. Res, 32(3), 964–968.

Hebert, P. D. N., Cywinska, A., Ball, S. L., & DeWaard, J. R. (2003). Biological identifications through DNA barcodes. Proceedings of the Royal Society B: Biological Sciences, 270(1512), 313–321. https://doi.org/10.1098/rspb.2002.2218

Heggenes, J., Stickler, M., Alfredsen, K., Brittain, J. E., Adeva-Bustos, A., & Huusko, A. (2021). Hydropower-driven thermal changes, biological responses and mitigating measures in

northern river systems. In River Research and Applications (Vol. 37, Issue 5, pp. 743–765). John Wiley and Sons Ltd. https://doi.org/10.1002/rra.3788

Hering, D., Borja, A., Carstensen, J., Carvalho, L., Elliott, M., Feld, C. K., Heiskanen, A. S., Johnson, R. K., Moe, J., Pont, D., Solheim, A. L., & de Bund, W. van. (2010). The European Water Framework Directive at the age of 10: A critical review of the achievements with recommendations for the future. Science of the Total Environment, 408(19), 4007–4019. https://doi.org/10.1016/j.scitotenv.2010.05.031

Hill, M. J., Mathers, K. L., Little, S., Worrall, T., Gunn, J., & Wood, P. J. (2019). Ecological effects of a supra-seasonal drought on macroinvertebrate communities differ between near-perennial and ephemeral river reaches. Aquatic Sciences, 81(4). https://doi.org/10.1007/s00027-019-0659-7

Huđek, H., Žganec, K., & Pusch, M. T. (2020). A review of hydropower dams in Southeast Europe – distribution, trends and availability of monitoring data using the example of a multinational Danube catchment subarea. Renewable and Sustainable Energy Reviews, 117, 109434. https://doi.org/10.1016/j.rser.2019.109434

Hughes, S. J., Furse, M. T., Blackburn, J. H., & Langton, P. H. (1998). A checklist of Madeiran freshwater macroinvertebrates. Bol. Mus. Mun. Funchal, 50(284), 5–41. https://publications.cm-funchal.pt/jspui/handle/100/883

Inoue, Y., Miyata, K., Yamane, M., & Honda, H. (2023). Environmental Nucleic Acid Pollution: Characterization of Wastewater Generating False Positives in Molecular Ecological Surveys. ACS ES and T Water, 3(3), 756–764. https://doi.org/10.1021/acsestwater.2c00542

IUCN. (2024). The IUCN Red List of Threatened Species. Version 2023-1. https://www.iucnredlist.org

Jáimez-Cuéllar, P., Palomino Morales, J. A., Luzón Ortega, J., & Alba-Tercedor, J. (2006). Comparación de metodologías empleadas para la evaluación del estado ecológico de los cursos de agua. Tecnología Del Agua, 278, 42–57.

Jáimez-Cuéllar, P., Vivas, S., Bonada, N., Robles, S., Mellado, A., Álvarez, M., Avilés, J., Casas, J., Ortega, M., Pardo, I., Prat, N., Rieradevall, M., Sáinz-Cantero, C. E., Sánchez-Ortega, A., Suárez, M. L., Toro, M., Vidal-Abarca, R., Zamora-Muñoz, C., & Alba-Tercedor, J. (2002). Protocolo GUADALMED (PRECE). Limnetica, 21(4), 187–204.

Jane, S. F., Wilcox, T. M., Mckelvey, K. S., Young, M. K., Schwartz, M. K., Lowe, W. H., Letcher, B. H., & Whiteley, A. R. (2015). Distance, flow and PCR inhibition: EDNA dynamics in two headwater streams. Molecular Ecology Resources, 15(1), 216–227. https://doi.org/10.1111/1755-0998.12285

Jeunen, G. J., Dowle, E., Edgecombe, J., von Ammon, U., Gemmell, N. J., & Cross, H. (2023). crabs—A software program to generate curated reference databases for metabarcoding sequencing data. Molecular Ecology Resources, 23(3), 725–738. https://doi.org/10.1111/1755-0998.13741

Ji, F., Han, D., Yan, L., Yan, S., Zha, J., & Shen, J. (2022). Assessment of benthic invertebrate diversity and river ecological status along an urbanized gradient using environmental DNA metabarcoding and a traditional survey method. Science of The Total Environment, 806, 150587. https://doi.org/10.1016/j.scitotenv.2021.150587

Jo, T., Sakata, M. K., Murakami, H., Masuda, R., & Minamoto, T. (2021). Universal performance of benzalkonium chloride for the preservation of environmental DNA in seawater samples. Limnology and Oceanography: Methods, 19(11), 758–768. https://doi.org/10.1002/lom3.10459

Jones, F. C. (2008). Taxonomic sufficiency: The influence of taxonomic resolution on freshwater bioassessments using benthic macroinvertebrates. In Environmental Reviews (Vol. 16, pp. 45–69). https://doi.org/10.1139/A07-010

Jones, P. E., Consuegra, S., Börger, L., Jones, J., & Garcia de Leaniz, C. (2020). Impacts of artificial barriers on the connectivity and dispersal of vascular macrophytes in rivers: A critical review. Freshwater Biology, 65(6), 1165–1180. https://doi.org/10.1111/fwb.13493

Juvigny-Khenafou, N. P. D., Piggott, J. J., Atkinson, D., Zhang, Y., Macaulay, S. J., Wu, N., & Matthaei, C. D. (2021). Impacts of multiple anthropogenic stressors on stream macroinvertebrate community composition and functional diversity. Ecology and Evolution, 11(1), 133–152. https://doi.org/10.1002/ece3.6979

Kaspersen, B. S. (2015). The EU Water Framework Directive-Action Programmes and Climate Change Challenges. https://doi.org/10.13140/RG.2.1.2516.8249

Keck, F., Blackman, R. C., Bossart, R., Brantschen, J., Couton, M., Hürlemann, S., Kirschner, D., Locher, N., Zhang, H., & Altermatt, F. (2022). Meta-analysis shows both congruence and complementarity of DNA and eDNA metabarcoding to traditional methods for biological community assessment. Molecular Ecology, 31(6), 1820–1835. https://doi.org/10.1111/mec.16364

Keck, F., Couton, M., & Altermatt, F. (2022). Navigating the seven challenges of taxonomic reference databases in metabarcoding analyses. Molecular Ecology Resources. https://doi.org/10.1111/1755-0998.13746

Keller, S. R., Hilderbrand, R. H., Shank, M. K., & Potapova, M. (2017). Environmental DNA genetic monitoring of the nuisance freshwater diatom, Didymosphenia geminata, in

eastern North American streams. Diversity and Distributions, 23(4), 381–393. https://doi.org/10.1111/ddi.12536

Kestel, J. H., Field, D. L., Bateman, P. W., White, N. E., Allentoft, M. E., Hopkins, A. J. M., Gibberd, M., & Nevill, P. (2022). Applications of environmental DNA (eDNA) in agricultural systems: Current uses, limitations and future prospects. In Science of the Total Environment (Vol. 847). Elsevier B.V. https://doi.org/10.1016/j.scitotenv.2022.157556

Kolkwitz, R., & Marsson, M. (1909). Ökologie der tierischen Saprobien. Beiträge zur Lehre von der biologischen Gewässerbeurteilung. Internationale Revue Der Gesamten Hydrobiologie Und Hydrographie, 2(1–2), 126–152. https://doi.org/10.1002/iroh.19090020108

Kumar, A., Kumar, R. R., Sharma, B. D., Gokulakrishnan, P., Mendiratta, S. K., & Sharma, D. (2015). Identification of Species Origin of Meat and Meat Products on the DNA Basis: A Review. Critical Reviews in Food Science and Nutrition, 55(10), 1340–1351. https://doi.org/10.1080/10408398.2012.693978

Kummu, M., Guillaume, J. H. A., De Moel, H., Eisner, S., Flörke, M., Porkka, M., Siebert, S., Veldkamp, T. I. E., & Ward, P. J. (2016). The world's road to water scarcity: Shortage and stress in the 20th century and pathways towards sustainability. Scientific Reports, 6. https://doi.org/10.1038/srep38495

Kuntke, F., de Jonge, N., Hesselsøe, M., & Lund Nielsen, J. (2020). Stream water quality assessment by metabarcoding of invertebrates. Ecological Indicators, 111(November 2019), 105982. https://doi.org/10.1016/j.ecolind.2019.105982

Lamb, P. D., Hunter, E., Pinnegar, J. K., Creer, S., Davies, R. G., & Taylor, M. I. (2019). How quantitative is metabarcoding: A meta-analytical approach. Molecular Ecology, 28(2), 420–430. https://doi.org/10.1111/mec.14920

Leese, F., Sander, M., Buchner, D., Elbrecht, V., Haase, P., & Zizka, V. M. A. (2020). Improved freshwater macroinvertebrate detection from eDNA through minimized non-target amplification. BioRxiv, April, 2020.04.27.063545. https://doi.org/10.1101/2020.04.27.063545

Leese, F., Sander, M., Buchner, D., Elbrecht, V., Haase, P., & Zizka, V. M. A. (2021). Improved freshwater macroinvertebrate detection from environmental DNA through minimized nontarget amplification. Environmental DNA, 3(1), 261–276. https://doi.org/10.1002/edn3.177

Lenat, D. R. (1993). A Biotic Index for the Southeastern United States: Derivation and List of Tolerance Values, with Criteria for Assigning Water-Quality Ratings. Journal of the

North American Benthological Society, 12(3), 279–290. https://doi.org/10.2307/1467463

Lencioni, V., Marziali, L., & Rossaro, B. (2012). Chironomids as bioindicators of environmental quality in mountain springs. Freshwater Science, 31(2), 525–541. https://doi.org/10.1899/11-038.1

Leray, M., Yang, J. Y., Meyer, C. P., Mills, S. C., Agudelo, N., Ranwez, V., Boehm, J. T., & Machida, R. J. (2013). A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: Application for characterizing coral reef fish gut contents. Frontiers in Zoology, 10(1). https://doi.org/10.1186/1742-9994-10-34

Lessard, J. L., & Hayes, D. B. (2003). Effects of elevated water temperature on fish and macroinvertebrate communities below small dams. River Research and Applications, 19(7), 721–732. https://doi.org/10.1002/rra.713

Liu, M., Clarke, L. J., Baker, S. C., Jordan, G. J., & Burridge, C. P. (2019). A practical guide to DNA metabarcoding for entomological ecologists. In Ecological Entomology. https://doi.org/10.1111/een.12831

Macher, T. H., Schütz, R., Hörren, T., Beermann, A. J., & Leese, F. (2023). It's raining species: Rainwash eDNA metabarcoding as a minimally invasive method to assess tree canopy invertebrate diversity. Environmental DNA, 5(1), 3–11. https://doi.org/10.1002/edn3.372

Macher, T.-H., Beermann, A. J., & Leese, F. (2020). TaxonTableTools - A comprehensive, platform-independent graphical user interface software to explore and visualise DNA metabarcoding data. BioRxiv, 2020.08.24.264317. https://doi.org/10.1101/2020.08.24.264317

Mackie, J. K., Chester, E. T., Matthews, T. G., & Robson, B. J. (2013). Macroinvertebrate response to environmental flows in headwater streams in western Victoria, Australia. Ecological Engineering, 53, 100–105. https://doi.org/10.1016/j.ecoleng.2012.12.018

MAGRAMA. (2011). Real Decreto 139/2011, de 4 de febrero, para el desarrollo del Listado de Especies Silvestres en Régimen de Protección Especial y del Catálogo Español de Especies Amenazadas. https://www.boe.es/eli/es/rd/2011/02/04/139

MAGRAMA. (2013a). Protocolo de cálculo del índice IBMWP. Código: IBMWP-2013. Ministerio de Agricultura, Alimentación y Medio Ambiente. https://www.miteco.gob.es/content/dam/miteco/es/agua/temas/estado-y-calidad-de-las-aguas/IBMWP-2013_24_05_2013_tcm30-175292.pdf

MAGRAMA. (2013b). Protocolo De Muestreo Y Laboratorio De Fauna Bentónica De Invertebrados En Ríos Vadeables. Código: ML-Rv-I-2013. Ministerio de Agricultura, Alimentación y Medio Ambiente, 23.

MAGRAMA. (2013c). Real Decreto 630/2013, de 2 de agosto, por el que se regula el Catálogo español de especies exóticas invasoras. https://www.boe.es/eli/es/rd/2013/08/02/630

MAGRAMA. (2015a). PROTOCOLO DE CÁLCULO DEL ÍNDICE MULTIMÉTRICO ESPECÍFICO DEL TIPO DE INVERTEBRADOS BENTÓNICOS EN RÍOS METI-2015. Ministerio de Agricultura, Alimentación y Medio Ambiente. http://publicacionesoficiales.boe.es/

MAGRAMA. (2015b). Real Decreto 817/2015, por el que se establecen los criterios de seguimiento y evaluación del estado de las aguas superficiales y las normas de calidad ambiental. Boletín Oficial Del Estado, 50, 38–40.

MAGRAMA. (2016). Protocolo de cálculo del Índice Multimétrico específico del tipo de invertebrados bentónicos en ríos. 14. http://www.magrama.gob.es/es/agua/temas/estado-y-calidad-de-las-aguas/aguas-superficiales/programas-seguimiento/Protocolos-de-muestro-laboratorio-y-calculo-de-indices.aspx%5Cnhttp://files/2118/Protocolo de calculo del indice multimetrico espec?fico del ti

Makino, W., Suzuki, H., Otake, Y., Ban, S., & Urabe, J. (2023). The first report of the non-indigenous Chydorus brevilabris Frey, 1980 (Crustacea: Cladocera) in Asian freshwaters. Limnology, 24(3), 151–159. https://doi.org/10.1007/s10201-023-00719-4

Markham, S. K. (2002). Moving technologies from lab to market. Research Technology Management, 45(6), 31–42. https://doi.org/10.1080/08956308.2002.11671531

Márquez, I., García-Vázquez, E., & Borrell, Y. J. (2014). Possible effects of vaccination and environmental changes on the presence of disease in northern Spanish fish farms. Aquaculture, 431, 118–123. https://doi.org/10.1016/j.aquaculture.2013.12.030

Martin, P., Martinez-Ansemil, E., Pinder, A., Timm, T., & Wetzel, M. J. (2008). Global diversity of oligochaetous clitellates (‘Oligochaeta’; Clitellata) in freshwater. In Hydrobiologia (Vol. 595, Issue 1, pp. 117–127). https://doi.org/10.1007/s10750-007-9009-1

Martins, F. M. S., Beja, P., & Célio Alves, P. (2019). Towards Next-generation Biodiversity Monitoring: Improving Freshwater Quality Assessment using DNA Metabarcoding.

Martins, F. M. S., Feio, M. J., Porto, M., Filipe, A. F., Bonin, A., Serra, S. R. Q., Alves, P. C., Taberlet, P., & Beja, P. (2021). Assessing changes in stream macroinvertebrate communities across ecological gradients using morphological versus DNA

metabarcoding approaches. Science of the Total Environment, 797. https://doi.org/10.1016/j.scitotenv.2021.149030

Martins, F. M. S., Galhardo, M., Filipe, A. F., Teixeira, A., Pinheiro, P., Paupério, J., Alves, P. C., & Beja, P. (2019). Have the cake and eat it: Optimizing nondestructive DNA metabarcoding of macroinvertebrate samples for freshwater biomonitoring. Molecular Ecology Resources, 19(4), 863–876. https://doi.org/10.1111/1755-0998.13012

Martins, F. M. S., Porto, M., Feio, M. J., Egeter, B., Bonin, A., Serra, S. R. Q., Taberlet, P., & Beja, P. (2021). Modelling technical and biological biases in macroinvertebrate community assessment from bulk preservative using multiple metabarcoding markers. Molecular Ecology, 30(13), 3221–3238. https://doi.org/10.1111/mec.15620

Martoni, F., Piper, A. M., Rodoni, B. C., & Blacket, M. J. (2022). Disentangling bias for non-destructive insect metabarcoding. PeerJ, 10. https://doi.org/10.7717/peerj.12981

Mauvisseau, Q., Harper, L. R., Sander, M., Hanner, R. H., Kleyer, H., & Deiner, K. (2022). The Multiple States of Environmental DNA and What Is Known about Their Persistence in Aquatic Environments. In Environmental Science and Technology (Vol. 56, Issue 9, pp. 5322–5333). American Chemical Society. https://doi.org/10.1021/acs.est.1c07638

Mbaka, J. G., & Wanjiru Mwaniki, M. (2015). A global review of the downstream effects of small impoundments on stream habitat conditions and macroinvertebrates. In Environmental Reviews (Vol. 23, Issue 3, pp. 257–262). Canadian Science Publishing. https://doi.org/10.1139/er-2014-0080

McGauley, E., Tregunno, B., & Jones, F. C. (2018). Coarse taxonomy (tolerance-value averaging) biases Hilsenhoff's family-level biotic index. Environmental Monitoring and Assessment, 190(8). https://doi.org/10.1007/s10661-018-6817-x

Medina-Gavilán, J. L., & González-Duarte, M. M. (2018). Una síntesis de las localidades ibéricas conocidas para Craspedacusta sowerbii Lankester, 1880 (Cnidaria: Hydrozoa): nuevo registro para España procedente de la vega del Guadalquivir. Graellsia, 74(2), 072. https://doi.org/10.3989/graellsia.2018.v74.193

Meglécz, E. (2023). COInr and mkCOInr: Building and customizing a nonredundant barcoding reference database from BOLD and NCBI using a semi-automated pipeline. Molecular Ecology Resources, 23(4), 933–945. https://doi.org/10.1111/1755-0998.13756

Meiklejohn, K. A., Damaso, N., & Robertson, J. M. (2019). Assessment of BOLD and GenBank – Their accuracy and reliability for the identification of biological materials. PLoS ONE, 14(6). https://doi.org/10.1371/journal.pone.0217084

Meyer, A., Boyer, F., Valentini, A., Bonin, A., Ficetola, G. F., Beisel, J., Bouquerel, J., Wagner, P., Gaboriaud, C., Leese, F., Dejean, T., Taberlet, P., & Usseglio-Polatera, P. (2021). Morphological vs. DNA metabarcoding approaches for the evaluation of stream ecological status with benthic invertebrates: Testing different combinations of markers and strategies of data filtering. Molecular Ecology, 30(13), 3203–3220. https://doi.org/10.1111/mec.15723

Mezger, G., De Stefano, L., & González del Tánago, M. (2019). Assessing the Establishment and Implementation of Environmental Flows in Spain. Environmental Management, 64(6), 721–735. https://doi.org/10.1007/s00267-019-01222-2

Mezger, G., González del Tánago, M., & De Stefano, L. (2021). Environmental flows and the mitigation of hydrological alteration downstream from dams: The Spanish case. Journal of Hydrology, 598. https://doi.org/10.1016/j.jhydrol.2020.125732

Millennium Ecosystem Assessment. (2005). Ecosystems and human well-being : synthesis. Island Press.

Orden ARM/2656/2008, de 10 de Septiembre, Por La Que Se Aprueba La Instrucción de Planificación Hidrológica., Pub. L. No. RD 907/2007 (2007). https://www.boe.es/eli/es/rd/2007/07/06/907

MITECO. (2019). Summary of Spanish river basin management plans Second cycle of the WFD (2015-2021). https://www.miteco.gob.es/content/dam/miteco/es/agua/temas/planificacion-hidrologica/summary_book_rbmp_2nd_cycle_tcm30-508614.pdf

MITECO. (2022). Informe de seguimiento de Planes Hidrológicos y Recursos Hídricos en España. Año 2021.

MITECO. (2023). ANEXO III Disposiciones normativas del Plan Hidrológico de la parte española de la Demarcación Hidrográfica del Miño-Sil. www.chminosil.es.

Molineri, C., Tejerina, E. G., Torrejón, S. E., Pero, E. J. I., & Hankel, G. E. (2020). Indicative value of different taxonomic levels of Chironomidae for assessing the water quality. Ecological Indicators, 108, 105703. https://doi.org/10.1016/j.ecolind.2019.105703

Mortágua, A., Vasselon, V., Oliveira, R., Elias, C., Chardon, C., Bouchez, A., Rimet, F., João Feio, M., & F.P. Almeida, S. (2019). Applicability of DNA metabarcoding approach in the bioassessment of Portuguese rivers using diatoms. Ecological Indicators, 106(April), 105470. https://doi.org/10.1016/j.ecolind.2019.105470

Mueller, M., Pander, J., & Geist, J. (2013). Taxonomic sufficiency in freshwater ecosystems: effects of taxonomic resolution, functional traits, and data transformation. Freshwater Science, 32(3), 762–778. https://doi.org/10.1899/12-212.1

Mugnai, F., Costantini, F., Chenuil, A., Leduc, M., Gutiérrez Ortega, J. M., & Meglécz, E. (2023). Be positive: customized reference databases and new, local barcodes balance false taxonomic assignments in metabarcoding studies. PeerJ, 11, e14616. https://doi.org/10.7717/peerj.14616

Munné, A., & Prat, N. (2009). Use of macroinvertebrate-based multimetric indices for water quality evaluation in Spanish Mediterranean rivers: An intercalibration approach with the IBMWP index. Hydrobiologia, 628(1), 203–225. https://doi.org/10.1007/s10750-009-9757-1

Múrria, C., Väisänen, L. O. S., Somma, S., Wangensteen, O. S., Arnedo, M. A., & Prat, N. (2020). Towards an iberian dna barcode reference library of freshwater macroinvertebrates and fishes. Limnetica, 39(1), 73–92. https://doi.org/10.23818/limn.39.06

Mustow, S. E. (2002). Biological monitoring of rivers in Thailand: use and adaptation of the BMWP score. In Hydrobiologia (Vol. 479).

Naranjo López, C., González Lazo, D. D., Garcés González, G., Brandimarte, A. L., Muñoz Riveaux, S., & Musle Cordero, Y. (2005). Una metodología rápida y de fácil aplicación para la evaluación de la calidad del agua utilizando el índice BMWP-cub para ríos cubanos (1st ed., Vol. 9). Universidad Distrital Francisco José de Caldas.

Ntislidou, C., Bozatzidou, M., Argyriou, A. K., Karaouzas, I., Skoulikidis, N., & Lazaridou, M. (2020). Minimizing human error in macroinvertebrate samples analyses for ensuring quality precision in freshwater monitoring programs. Science of the Total Environment, 703. https://doi.org/10.1016/j.scitotenv.2019.135496

Nukazawa, K., Hamasuna, Y., & Suzuki, Y. (2018). Simulating the Advection and Degradation of the Environmental DNA of Common Carp along a River. Environmental Science and Technology, 52(18), 10562–10570. https://doi.org/10.1021/acs.est.8b02293

Oficialdegui, F. J., Zamora-Marín, J. M., Guareschi, S., Anastácio, P. M., García-Murillo, P., Ribeiro, F., Miranda, R., Cobo, F., Gallardo, B., García-Berthou, E., Boix, D., Arias, A., Cuesta, J. A., Medina, L., Almeida, D., Banha, F., Barca, S., Biurrun, I., Cabezas, M. P., … Oliva-Paterna, F. J. (2023). A horizon scan exercise for aquatic invasive alien species in Iberian inland waters. Science of the Total Environment, 869. https://doi.org/10.1016/j.scitotenv.2023.161798

Oliva-Paterna, F. J., Ribeiro, F., Miranda, R., Anastácio, P. M., García-Murillo, P., Cobo, F., Gallardo, B., GarcíaBerthou, E., Boix, D., Medina, L., Morcillo, F., Oscoz, J., Guillén, A., Arias, A., Cuesta, J. A., Aguiar, F., Almeida, D., Ayres, C., Banha, F., … Zamora-Marín, J. M. (2021). LIST OF POTENTIAL AQUATIC ALIEN SPECIES OF THE IBERIAN PENINSULA (2020) Updated list of potential aquatic alien species with high risk of invasion in

Iberian inland waters. https://lifeinvasaqua.com/wp-content/uploads/2021/04/TR2_Invasaqua_ING_PDF_interact.pdf

Oosterbroek, P., Starý, J., Andrade, R., Hancock, E. G., & Ferreira, S. A. (2020). The Craneflies of continental Portugal (Diptera, Limoniidae, Pediciidae, Tipulidae) including 28 species new for Portugal Moscas grulla de Portugal continental (Diptera, Limoniidae, Pediciidae, Tipulidae) incluyendo 28 especies nuevas para Portugal. Boln. Asoc. Esp. Ent, 44(4), 10–12.

Ožana, S., Dolný, A., & Pánek, T. (2022). Nuclear copies of mitochondrial DNA as a potential problem for phylogenetic and population genetic studies of Odonata. Systematic Entomology, 47(4), 591–602. https://doi.org/10.1111/syen.12550

Paggi, A. C. (1999). Los Chironomidae como indicadores de calidad de ambientes dulceacuícolas. Rev. Soc. Entomol. Argent., 58, 202–207.

Palmer, M., & Ruhi, A. (2019). Linkages between flow regime, biota, and ecosystem processes: Implications for river restoration. In Science (Vol. 365, Issue 6459). American Association for the Advancement of Science. https://doi.org/10.1126/science.aaw2087

Paltridge, R. M., Dostine, P. L., Humphrey, C. L., & Boulton, A. J. (1997). Macroinvertebrate recolonization after re-wetting of a tropical seasonally-flowing stream (Magela Creek, Northern Territory, Australia). Marine and Freshwater Research, 48(7), 633–645. https://doi.org/10.1071/mf97059

Pardo, I., Gómez Rodriguez, C., & Abraín, R. (2010). Sistema de clasificación del estado ecológico de los ríos en el ámbito de las Confederaciones Hidrográficas del Cantábrico y Miño–Sil.

Pařil, P., Polášek, M., Loskotová, B., Straka, M., Crabot, J., & Datry, T. (2019). An unexpected source of invertebrate community recovery in intermittent streams from a humid continental climate. Freshwater Biology, 64(11), 1971–1983. https://doi.org/10.1111/fwb.13386

Pauperio, J., Gonzalez, L. M., Martinez, J., González, M., Martins, F. M., Veríssimo, J., Puppo, P., Pinto, J., Chaves, C., Pinho, C. J., Grosso-Silva, J. M., Quaglietta, L., Silva, T. L., Sousa, P., Alves, P., Fonseca, N., Beja, P., & Ferreira, S. A. (2023). The InBIO barcoding initiative database: DNA barcodes of Iberian Trichoptera, documenting biodiversity for freshwater biomonitoring in a Mediterranean hotspot. Biodiversity Data Journal, 11. https://doi.org/10.3897/BDJ.11.e97484

Pawlowski, J., Kelly-Quinn, M., Altermatt, F., Apothéloz-Perret-Gentil, L., Beja, P., Boggero, A., Borja, A., Bouchez, A., Cordier, T., Domaizon, I., Feio, M. J., Filipe, A. F., Fornaroli, R., Graf, W., Herder, J., van der Hoorn, B., Iwan Jones, J., Sagova-Mareckova, M., Moritz,

C., … Kahlert, M. (2018). The future of biotic indices in the ecogenomic era: Integrating (e)DNA metabarcoding in biological assessment of aquatic ecosystems. Science of the Total Environment, 637–638, 1295–1310. https://doi.org/10.1016/j.scitotenv.2018.05.002

Peñarrubia, L., Alcaraz, C., Vaate, A. B. De, Sanz, N., Pla, C., Vidal, O., & Vinãs, J. (2016). Validated methodology for quantifying infestation levels of dreissenid mussels in environmental DNA (eDNA) samples. Scientific Reports, 6. https://doi.org/10.1038/srep39067

Pentinsaari, M., Ratnasingham, S., Miller, S. E., & Hebert, P. D. N. (2020). BOLD and GenBank revisited - Do identification errors arise in the lab or in the sequence libraries? PLoS ONE, 15(4). https://doi.org/10.1371/journal.pone.0231814

Pérez, T., Albornoz, J., & Domínguez, A. (1998). An evaluation of RAPD fragment reproducibility and nature. Molecular Ecology, 7(10), 1347–1357. https://doi.org/10.1046/j.1365-294x.1998.00484.x

Pleguezuelos, J. M., Márquez, R., & Lizana, M. (2002). Capítulo IX Las especies introducidas de Anfibios y Reptiles.

Poff, N. L., & Zimmerman, J. K. H. (2010). Ecological responses to altered flow regimes: A literature review to inform the science and management of environmental flows. In Freshwater Biology (Vol. 55, Issue 1, pp. 194–205). https://doi.org/10.1111/j.1365-2427.2009.02272.x

Porter, T. M., & Hajibabaei, M. (2018). Over 2.5 million COI sequences in GenBank and growing. PLoS ONE, 13(9). https://doi.org/10.1371/journal.pone.0200177

Porter, T. M., & Hajibabaei, M. (2021). Profile hidden Markov model sequence analysis can help remove putative pseudogenes from DNA barcoding and metabarcoding datasets. BMC Bioinformatics, 22(1). https://doi.org/10.1186/s12859-021-04180-x

Pujante, Ana María, Puig, Alejandra, Barrios, Elena, & Ruza, Javier. (2016). Contribución al establecimiento de condiciones de referencia y límites entre clases de estado ecológico en los ríos españoles. Limnetica, 35, 201–218. https://doi.org/10.23818/limn.35.17

Ramos-Merchante, A., & Prenda, J. (2017). Macroinvertebrate taxa richness uncertainty and kick sampling in the establishment of Mediterranean rivers ecological status. Ecological Indicators, 72, 1–12. https://doi.org/10.1016/j.ecolind.2016.07.047

Raposeiro, Pedro Miguel, Cruz, Ana Mafalda, Hughes, Samantha Jane, & Costa , Ana Cristina. (2012). Azorean freshwater invertebrates: Status, threats and biogeographic notes. Limnetica, 31, 13–22. https://doi.org/10.23818/limn.31.02

Rasmussen, R. S., & Morrissey, M. T. (2008). DNA-Based Methods for the Identification of Commercial Fish and Seafood Species. Comprehensive Reviews in Food Science and Food Safety, 7(3), 280–295. https://doi.org/10.1111/j.1541-4337.2008.00046.x

Ratnasingham, S., & Hebert, P. D. N. (2007). BOLD: The Barcode of Life Data System (www.barcodinglife.org). Molecular Ecology Notes , 7, 355–364.

Ratnasingham, S., & Hebert, P. D. N. (2013). A DNA-Based Registry for All Animal Species: The Barcode Index Number (BIN) System. PLoS ONE, 8(7). https://doi.org/10.1371/journal.pone.0066213

Real Decreto 638/2016, de 9 de Diciembre, Por El Que Se Modifica El Reglamento Del Dominio Público Hidráulico Aprobado Por El Real Decreto 849/1986, de 11 de Abril, El Reglamento de Planificación Hidrológica, Aprobado Por El Real Decreto 907/2007, de 6 de Julio, y Otros Reglamentos En Materia de Gestión de Riesgos de Inundación, Caudales Ecológicos, Reservas Hidrológicas y Vertidos de Aguas Residuales., Pub. L. No. RD 638/2016 (2016). https://www.boe.es/eli/es/rd/2016/12/09/638

Real Decreto 1159/2021, de 28 de Diciembre, Por El Que Se Modifica El Real Decreto 907/2007, de 6 de Julio, Por El Que Se Aprueba El Reglamento de La Planificación Hidrológica., Pub. L. No. RD/1159/2021 (2021). https://www.boe.es/eli/es/rd/2021/12/28/1159

Rees, H. C., Bishop, K., Middleditch, D. J., Patmore, J. R. M., Maddison, B. C., & Gough, K. C. (2014). The application of eDNA for monitoring of the Great Crested Newt in the UK. Ecology and Evolution, 4(21), 4023–4032. https://doi.org/10.1002/ece3.1272

Reji Chacko, M., Altermatt, F., Fopp, F., Guisan, A., Keggin, T., Lyet, A., Rey, P. L., Richards, E., Valentini, A., Waldock, C., & Pellissier, L. (2023). Catchment-based sampling of river eDNA integrates terrestrial and aquatic biodiversity of alpine landscapes. Oecologia, 202(4), 699–713. https://doi.org/10.1007/s00442-023-05428-4

Resh, V. H., & McElravy. (1993). Contemporary quantitative approaches to biomonitoring using benthic macroinvertebrates. In Freshwater biomonitoring and benthic macroinvertebrates (pp. 159–194). Champan and Hall.

Rivera, S. F., Vasselon, V., Mary, N., Monnier, O., Rimet, F., & Bouchez, A. (2021). Exploring the capacity of aquatic biofilms to act as environmental DNA samplers: Test on macroinvertebrate communities in rivers. Science of the Total Environment, 763. https://doi.org/10.1016/j.scitotenv.2020.144208

Rodriguez-Ezpeleta, N., Morissette, O., Bean, C., Manu, S., Banerjee, P., Lacoursiere, A., Ben, K., Alter, E., Roger, F., Holman, L., Stewart, K., Monaghan, M., Mauvisseau, Q., Mirimin, L., Wangensteen, O. S., Antognazza, C., Helyar, S., Boer, H. de, Monchamp, M.-E., … Deiner, K. (2021). Trade-offs between reducing complex terminology and

producing accurate interpretations from environmental DNA: Comment on “Environmental DNA: What’s behind the term?” by Pawlowski et al. (2020). Authorea Preprints, 1–11. https://doi.org/DOI: 10.1111/mec.15942

Roldan Pérez, G. (2003). Bioindicación de la calidad del agua en colombia. Propuesta para el uso del método BMWP/Col. (1st ed.). Universidad de Antioquia.

Roldán Pérez, G. (2012). Los macroinvertebrados como bioindicadores de la calidad del agua.

Sánchez, O., & Arias, A. (2021). All that glitters is not gold: The other insects that fall into the asian yellow-legged hornet vespa velutina ‘specific’ traps. Biology, 10(5). https://doi.org/10.3390/biology10050448

Schmidt-Kloiber, A., & Hering, D. (2015). www.freshwaterecology.info – An online tool that unifies, standardises and codifies more than 20,000 European freshwater organisms and their ecological preferences. Ecological Indicators, 53, 271–282. https://doi.org/10.1016/j.ecolind.2015.02.007

Schmidt-Kloiber, A., Strackbein, J., Vogl, R., Furse, M. T., & Hering, D. (2014). Description of the AQEM/STAR invertebrate database. Freshwater Metadata Journal, 1–8. https://doi.org/10.15504/fmj.2014.2

Schriever, T. A., Bogan, M. T., Boersma, K. S., Cañedo-Argüelles, M., Jaeger, K. L., Olden, J. D., & Lytle, D. A. (2015). Hydrology shapes taxonomic and functional structure of desert stream invertebrate communities. Freshwater Science, 34(2), 399–409. https://doi.org/10.1086/680518

Schultz, J. A., & Hebert, P. D. N. (2022). Do pseudogenes pose a problem for metabarcoding marine animal communities? Molecular Ecology Resources, 22(8), 2897–2914. https://doi.org/10.1111/1755-0998.13667

Sepulveda, A. J., Hutchins, P. R., Forstchen, M., Mckeefry, M. N., & Swigris, A. M. (2020). The Elephant in the Lab ( and Field ): Contamination in Aquatic Environmental DNA Studies. Frontiers in Evology and Evolution, 8(December), 1–12. https://doi.org/10.3389/fevo.2020.609973

Sepulveda, A. J., Nelson, N. M., Jerde, C. L., & Luikart, G. (2020). Are Environmental DNA Methods Ready for Aquatic Invasive Species Management? Trends in Ecology and Evolution, 35(8), 668–678. https://doi.org/10.1016/j.tree.2020.03.011

Serra, S. R. Q., Graça, M. A. S., Dolédec, S., & Feio, M. J. (2017). Chironomidae traits and life history strategies as indicators of anthropogenic disturbance. Environmental Monitoring and Assessment, 189(7). https://doi.org/10.1007/s10661-017-6027-y

Shelton, A. O., Gold, Z. J., Jensen, A. J., D'Agnese, E., Andruszkiewicz Allan, E., Van Cise, A., Gallego, R., Ramón-Laca, A., Garber-Yonts, M., Parsons, K., & Kelly, R. P. (2023). Toward quantitative metabarcoding. Ecology, 104(2). https://doi.org/10.1002/ecy.3906

Shirazi, S., Meyer, R. S., & Shapiro, B. (2021). Revisiting the effect of PCR replication and sequencing depth on biodiversity metrics in environmental DNA metabarcoding. Ecology and Evolution, 11(22), 15766–15779. https://doi.org/10.1002/ece3.8239

Sickel, W., Zizka, V., Scherges, A., Bourlat, S. J., & Dieker, P. (2023). Abundance estimation with DNA metabarcoding – recent advancements for terrestrial arthropods. Metabarcoding and Metagenomics, 7. https://doi.org/10.3897/mbmg.7.112290

Siegenthaler, A., Wangensteen, O. S., Soto, A. Z., Benvenuto, C., Corrigan, L., & Mariani, S. (2019). Metabarcoding of shrimp stomach content: Harnessing a natural sampler for fish biodiversity monitoring. Molecular Ecology Resources, 19(1), 206–220. https://doi.org/10.1111/1755-0998.12956

Smit, H. (2020). Water Mites of the World With Keys to the Families, Subfamilies, Genera and Subgenera (Acari: Hydrachnidia) (H. Siepel, Ed.; 1st ed.). Nederlandse Entomologische Vereniging.

Smith, A. S., & Alexander, J. E. (2008). Potential effects of the freshwater jellyfish Craspedacusta sowerbii on zooplankton community abundance. Journal of Plankton Research, 30(12), 1323–1327. https://doi.org/10.1093/plankt/fbn093

Soesbergen, M. (2020). Cladocera in the valley of the river Sûre (Luxemburg and Belgium), with a review of Chydorus brevilabris in Western Europe. Fauna Biol Stud, 7, 91–97. https://www.researchgate.net/publication/343323039

Sogin, M. L., Morrison, H. G., Huber, J. A., Welch, D. M., Huse, S. M., Neal, P. R., Arrieta, J. M., Herndl, G. J., Bay, J., & Center, P. (2006). Microbial diversity in the deep sea and the underexplored "'rare biosphere"'. www.pnas.orgcgidoi10.1073pnas.0605127103

Sonet, G., Jordaens, K., Braet, Y., Bourguignon, L., Dupont, E., Backeljau, T., De Meyer, M., & Desmyter, S. (2013). Utility of GenBank and the Barcode of Life Data Systems (BOLD) for the identification of forensically important Diptera from Belgium and France. ZooKeys, 365(SPEC.ISSUE), 307–328. https://doi.org/10.3897/zookeys.365.6027

Sousa, L. L. de, Silva, S. M., & Xavier, R. (2019). DNA metabarcoding in diet studies: Unveiling ecological aspects in aquatic and terrestrial ecosystems. In Environmental DNA (Vol. 1, Issue 3, pp. 199–214). Blackwell Publishing Inc. https://doi.org/10.1002/edn3.27

Sousa, P., Grosso-Silva, J., Andrade, R., Chaves, C., Pinto, J., Paupério, J., Beja, P., & Ferreira, S. A. (2021). The InBIO Barcoding Initiative Database: DNA barcodes of Portuguese

Hemiptera 01. Biodiversity Data Journal, 9, 1–25. https://doi.org/10.3897/BDJ.9.e65314

Stubbington, R., Bogan, M. T., Bonada, N., Boulton, A. J., Datry, T., Leigh, C., & Vander Vorste, R. (2017). The Biota of Intermittent Rivers and Ephemeral Streams: Aquatic Invertebrates. In Intermittent Rivers and Ephemeral Streams: Ecology and Management (pp. 217–243). Elsevier Inc. https://doi.org/10.1016/B978-0-12-803835-2.00007-3

Sutherland, J. P. (1974). Multiple Stable Points in Natural Communities. Source: The American Naturalist, 108(964), 859–873.

Szekeres, J., Beermann, A., Neubauer, T. A., Očadlik, M., Paunović, M., Raković, M., Csányi, B., Varga, A., Weigand, A., Wilke, T., & Fehér, Z. (2022). Rapid spread of a new alien and potentially invasive species, Clathrocaspia knipowitschii (Makarov, 1938) (Gastropoda: Hydrobiidae), in the Danube River. Archives of Biological Sciences, 74(1), 81–89. https://doi.org/10.2298/ABS220211006S

Taberlet, P., Bonin, A., Zinger, L., & Coissac, E. (2018). Enviromental DNA for Biodiversity Research and Monitoring (First Edition). Oxford University Press.

Taberlet, P., Coissac, E., Hajibabaei, M., & Rieseberg, L. H. (2012). Environmental DNA. In Molecular Ecology (Vol. 21, Issue 8). https://doi.org/10.1111/j.1365-294X.2012.05542.x

Tachet, H., Richoux, P., Bournaux, M., & Usseglio-Polatera, P. (2010). Invertébrés d'Eau Douce. Systématique, Biologie, Écologie. CNRS Éditions.

Takahara, T., Taguchi, J., Yamagishi, S., Doi, H., Ogata, S., Yamanaka, H., & Minamoto, T. (2020). Suppression of environmental DNA degradation in water samples associated with different storage temperature and period using benzalkonium chloride. Limnology and Oceanography: Methods, 18(8), 437–445. https://doi.org/10.1002/lom3.10374

Tharme, R. E. (2003). A global perspective on environmental flow assessment: Emerging trends in the development and application of environmental flow methodologies for rivers. River Research and Applications, 19(5–6), 397–441. https://doi.org/10.1002/rra.736

Tinerella, P. P. (2006). Phylogenetic ingroup relationships of the water boatmen (Insecta: Heteroptera: Corixoidea) and systematics of Australasian pygmy water boatmen (Corixoidea: Micronectidae) . North Dakota State University.

Tinerella, P. P. (2008). Taxonomic revision and systematics of New Guinea and Oceania pygmy water boatmen (Hemiptera: Heteroptera: Corixoidea: Micronectidae). Zootaxa, 1797(1), 1. https://doi.org/10.11646/zootaxa.1797.1.1

Tübitak, E. T. (2020). An impact assessment model for technology development programs. https://doi.org/10.13140/RG.2.2.27037.03040

UN Environment Programme. (2022). Freshwater Strategic Priorities 2022–2025 to implement UNEP's Medium-Term Strategy.

Vaikasas, S., Palaima, K., & Pliuraite, V. (2013). Influence of hydropower dams on the state of macroinvertebrates assemblages in the Virvyte river, Lithuania. Journal of Environmental Engineering and Landscape Management, 21(4), 305–315. https://doi.org/10.3846/16486897.2013.796956

van der Lee, G. H., Polling, M., van der Laan, I., Kodde, L., & Verdonschot, R. C. M. (2024). From DNA to diagnostics: A case study using macroinvertebrate metabarcoding to assess the effectiveness of restoration measures in a Dutch stream. Science of The Total Environment, 923, 171413. https://doi.org/10.1016/j.scitotenv.2024.171413

Van Looy, K., Tonkin, J. D., Floury, M., Leigh, C., Soininen, J., Larsen, S., Heino, J., LeRoy Poff, N., Delong, M., Jähnig, S. C., Datry, T., Bonada, N., Rosebery, J., Jamoneau, A., Ormerod, S. J., Collier, K. J., & Wolter, C. (2019). The three Rs of river ecosystem resilience: Resources, recruitment, and refugia. River Research and Applications, 35(2), 107–120. https://doi.org/10.1002/rra.3396

van Puijenbroek, P. J. T. M., Buijse, A. D., Kraak, M. H. S., & Verdonschot, P. F. M. (2019). Species and river specific effects of river fragmentation on European anadromous fish species. River Research and Applications, 35(1), 68–77. https://doi.org/10.1002/rra.3386

Vedia, Iván, & Miranda, Rafael. (2013). Review of the state of knowledge of crayfish species in the Iberian Peninsula. Limnetica, 32, 269–286. https://doi.org/10.23818/limn.32.22

Vieira Lanero, R. (2000). Las larvas de los tricópteros de Galicia (Insecta Trichoptera). Universidade de Santiago de Compostela.

Vinson, M. R. (2001). Long-term dynamics of an invertebrate assemblage downstream from a large dam. In Ecological Applications (Vol. 11, Issue 3). https://doi.org/https://doi.org/10.1890/1051-0761(2001)011[0711:LTDOAI]2.0.CO;2

Vorste, V. R., Corti, R., Sagouis, A., & Datry, T. (2016). Invertebrate communities in gravel-bed, braided rivers are highly resilient to flow intermittence. Freshwater Science, 35(1), 164–177. https://doi.org/10.1086/683274

Vourka, A., Karaouzas, I., & Parmakelis, A. (2023a). River benthic macroinvertebrates and environmental DNA metabarcoding: a scoping review of eDNA sampling, extraction, amplification and sequencing methods. In Biodiversity and Conservation (Vol. 32,

Issue 13, pp. 4221–4238). Springer Science and Business Media B.V. https://doi.org/10.1007/s10531-023-02710-y

Vourka, A., Karaouzas, I., & Parmakelis, A. (2023b). River benthic macroinvertebrates and environmental DNA metabarcoding: a scoping review of eDNA sampling, extraction, amplification and sequencing methods. Biodiversity and Conservation, 32(13), 4221–4238. https://doi.org/10.1007/s10531-023-02710-y

Wallace, J. B., & Webster, J. R. (1996). The Role of Macroinvertebrates in Stream Ecosystem Function. In Anna Rev. Emomd 19% (Vol. 41). www.annualreviews.org

Waringer, J., González, M. A., & Malicky, H. (2020). Discriminatory matrix for the larvae of the European Thremma species (Trichoptera: Thremmatidae). Zootaxa, 4718(4). https://doi.org/10.11646/zootaxa.4718.4.1

Water Framework Directive (2000/60/EC) (2000). https://data.europa.eu/doi/10.2779/75229

Weigand, H., Beermann, A. J., Čiampor, F., Costa, F. O., Csabai, Z., Duarte, S., Geiger, M. F., Grabowski, M., Rimet, F., Rulik, B., Strand, M., Szucsich, N., Weigand, A. M., Willassen, E., Wyler, S. A., Bouchez, A., Borja, A., Čiamporová-Zaťovičová, Z., Ferreira, S. A., … Ekrem, T. (2019). DNA barcode reference libraries for the monitoring of aquatic biota in Europe: Gap-analysis and recommendations for future work. In Science of the Total Environment (Vol. 678, pp. 499–524). https://doi.org/10.1016/j.scitotenv.2019.04.247

Wetterstrand, K. A. (2023). DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). www.genome.gov/sequencingcostsdata

Whittier, T. R., & Van Sickle, J. (2010). Macroinvertebrate tolerance values and an assemblage tolerance index (ATI) for western USA streams and rivers. Journal of the North American Benthological Society, 29(3), 852–866. https://doi.org/10.1899/09-160.1

Wilke, T., Haase, M., Hershler, R., Liu, H.-P., Misof, B., & Ponder, W. (2013). Pushing short DNA fragments to the limit: Phylogenetic relationships of 'hydrobioid' gastropods (Caenogastropoda: Rissooidea). Molecular Phylogenetics and Evolution, 66(3), 715–736. https://doi.org/10.1016/j.ympev.2012.10.025

Wu, L., Wen, C., Qin, Y., Yin, H., Tu, Q., Van Nostrand, J. D., Yuan, T., Yuan, M., Deng, Y., & Zhou, J. (2015). Phasing amplicon sequencing on Illumina Miseq for robust environmental microbial community analysis. BMC Microbiology, 15(1). https://doi.org/10.1186/s12866-015-0450-4

WWF. (2024). Freshwater. Supporting life on our planet by ensuring healthy, resilient freshwater systems. https://www.worldwildlife.org/initiatives/freshwater

WWF-ZSL. (2022). Living planet report 2022 - Building a nature-positive society (R. E. A. Almond, M. Grooten, D. Juffe Bignoli, & T. Petersen, Eds.).

Yamanaka, H., Minamoto, T., Matsuura, J., Sakurai, S., Tsuji, S., Motozawa, H., Hongo, M., Sogo, Y., Kakimi, N., Teramura, I., Sugita, M., Baba, M., & Kondo, A. (2017). A simple method for preserving environmental DNA in water samples at ambient temperature by addition of cationic surfactant. Limnology, 18(2), 233–241. https://doi.org/10.1007/s10201-016-0508-5

Yeo, D., Srivathsan, A., & Meier, R. (2020). Longer is Not Always Better: Optimizing Barcode Length for Large-Scale Species Discovery and Identification. Systematic Biology, 69(5), 999–1015. https://doi.org/10.1093/sysbio/syaa014

# XIV. Annex I

**Table XIV-1** List of macroinvertebrate taxa considered in Spanish legislation and their IBMWP score. Adapted from (MAGRAMA, 2013a). [1]The suborder Hydracarina has become the superorder Acariformes. [2]Anthomyiidae and Scatophagidae were formerly grouped as Muscidae. [3]Family Ferrissidae has become genus Ferrissia. [4]All genera except Ferrissia.

| | | | | | |
|---|---|---|---|---|---|
| **ARACHNIDS** | | **EPHEMEROPTERA** | | **ODONATA** | |
| Acariformes[1] | 1 | Baetidae | 4 | Aeshnidae | 8 |
| | | Caenidae | 4 | Calopterygidae | 8 |
| **COLEOPTERA** | | Ephemerellidae | 7 | Coenagrionidae | 6 |
| Chrysomelidae | 4 | Ephemeridae | 10 | Cordulegasteridae | 8 |
| Curculionidae | 4 | Heptageniidae | 10 | Corduliidae | 8 |
| Dryopidae | 5 | Leptophlebiidae | 10 | Gomphidae | 8 |
| Dytiscidae | 3 | Oligoneuriidae | 5 | Lestidae | 8 |
| Elmidae | 5 | Polymitarcidae | 5 | Libellulidae | 8 |
| Gyrinidae | 3 | Potamanthidae | 10 | Platycnemididae | 6 |
| Haliplidae | 4 | Prosopistomatidae | 7 | | |
| Helophoridae | 5 | Siphlonuridae | 10 | **OLIGOCHAETA** | |
| Hydraenidae | 5 | | | All | 1 |
| Hydrochidae | 5 | **HETEROPTERA** | | | |
| Hydrophilidae | 3 | Aphelocheiridae | 10 | **PLECOPTERA** | |
| Hygrobiidae | 3 | Corixidae | 3 | Capniidae | 10 |
| Noteridae | 3 | Gerridae | 3 | Chloroperlidae | 10 |
| Psephenidae | 3 | Hydrometridae | 3 | Leuctridae | 10 |
| Scirtidae (=Helodidae) | 3 | Mesoveliidae | 3 | Nemouridae | 7 |
| | | Naucoridae | 3 | Perlidae | 10 |
| **CRUSTACEA** | | Nepidae | 3 | Perlodidae | 10 |
| Asellidae | 3 | Notonectidae | 3 | Taeniopterygidae | 10 |
| Astacidae | 8 | Pleidae | 3 | | |
| Atyidae | 6 | Veliidae | 3 | **TRICHOPTERA** | |
| Corophiidae | 6 | | | Beraeidae | 10 |
| Gammaridae | 6 | **HIRUDINEA** | | Brachycentridae | 10 |
| Ostracoda | 3 | Erpobdellidae | 3 | Calamoceratidae | 10 |
| Palaemonidae | 6 | Glossiphoniidae | 3 | Ecnomidae | 7 |
| | | Hirudidae (=Hirudinidae) | 3 | Glossosomatidae | 8 |
| **DIPTERA** | | Piscicolidae | 4 | Goeridae | 10 |
| Anthomyiidae[2] | 2 | | | Hydropsychidae | 5 |
| Athericidae | 10 | **LEPIDOPTERA** | | Hydroptilidae | 6 |
| Blephariceridae | 10 | Crambidae (=Pyralidae) | 4 | Lepidostomatidae | 10 |
| Ceratopogonidae | 4 | | | Leptoceridae | 10 |
| Chironomidae | 2 | **MOLLUSCA** | | Limnephilidae | 7 |
| Culicidae | 2 | Ancylidae | 6 | Molannidae | 10 |
| Dixidae | 4 | Bithyniidae | 3 | Odontoceridae | 10 |
| Dolichopodidae | 4 | Ferrissia[3] | 6 | Philopotamidae | 8 |
| Empididae | 4 | Hydrobiidae | 3 | Phryganeidae | 10 |
| Ephydridae | 2 | Lymnaeidae | 3 | Polycentropodidae | 7 |

| | | | | | |
|---|---|---|---|---|---|
| Limoniidae | 4 | Neritidae | 6 | Psychomyiidae | 8 |
| Psychodidae | 4 | Physidae | 3 | Rhyacophilidae | 7 |
| Ptychopteridae | 4 | Planorbidae[4] | 3 | Sericostomatidae | 10 |
| Rhagionidae | 4 | Sphaeriidae | 3 | Uenoidae (=Thremmatidae) | 10 |
| Scatophagidae[2] | 4 | Thiaridae | 6 | | |
| Sciomyzidae | 4 | Unionidae | 6 | **TURBELLARIA** | |
| Simuliidae | 5 | Valvatidae | 3 | Dendrocoelidae | 6 |
| Stratiomyidae | 4 | Viviparidae | 6 | Dugesiidae | 3 |
| Syrphidae | 1 | | | Planariidae | 6 |
| Tabanidae | 4 | **NEUROPTERA** | | | |
| Thaumaleidae | 2 | Sialidae | 4 | | |
| Tipulidae | 5 | | | | |

# XV. Annex II

Supplementary information of section VI “The influence of reference genetic databases enrichment using local freshwater macroinvertebrate for metabarcoding based biodiversity studies in river monitoring.”

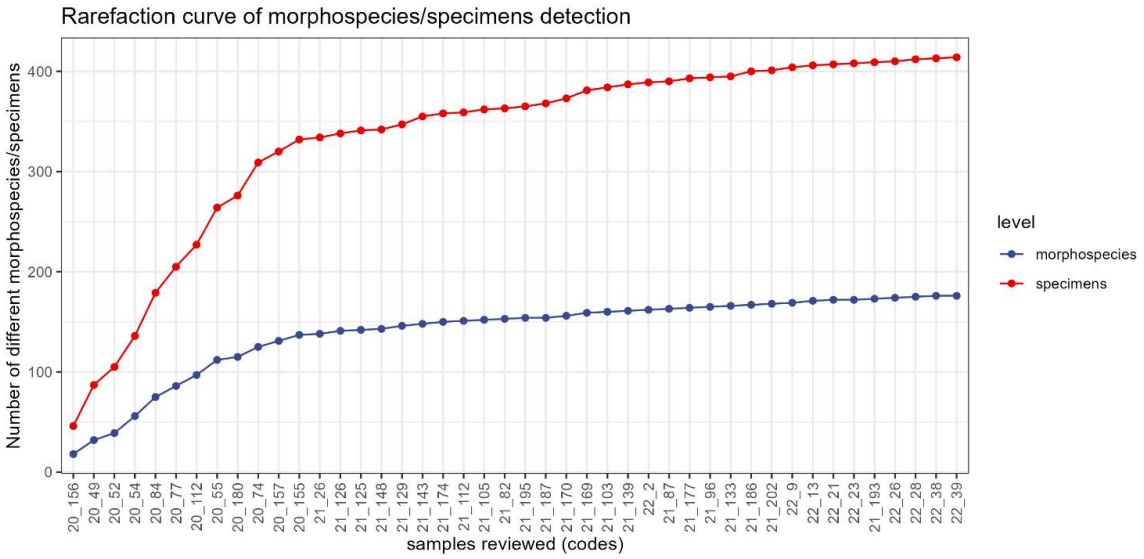



**Figure XV-1** Rarefaction curve of morphospecies and macroinvertebrate specimens collected for barcoding. Only samples from which specimens have been obtained are shown.

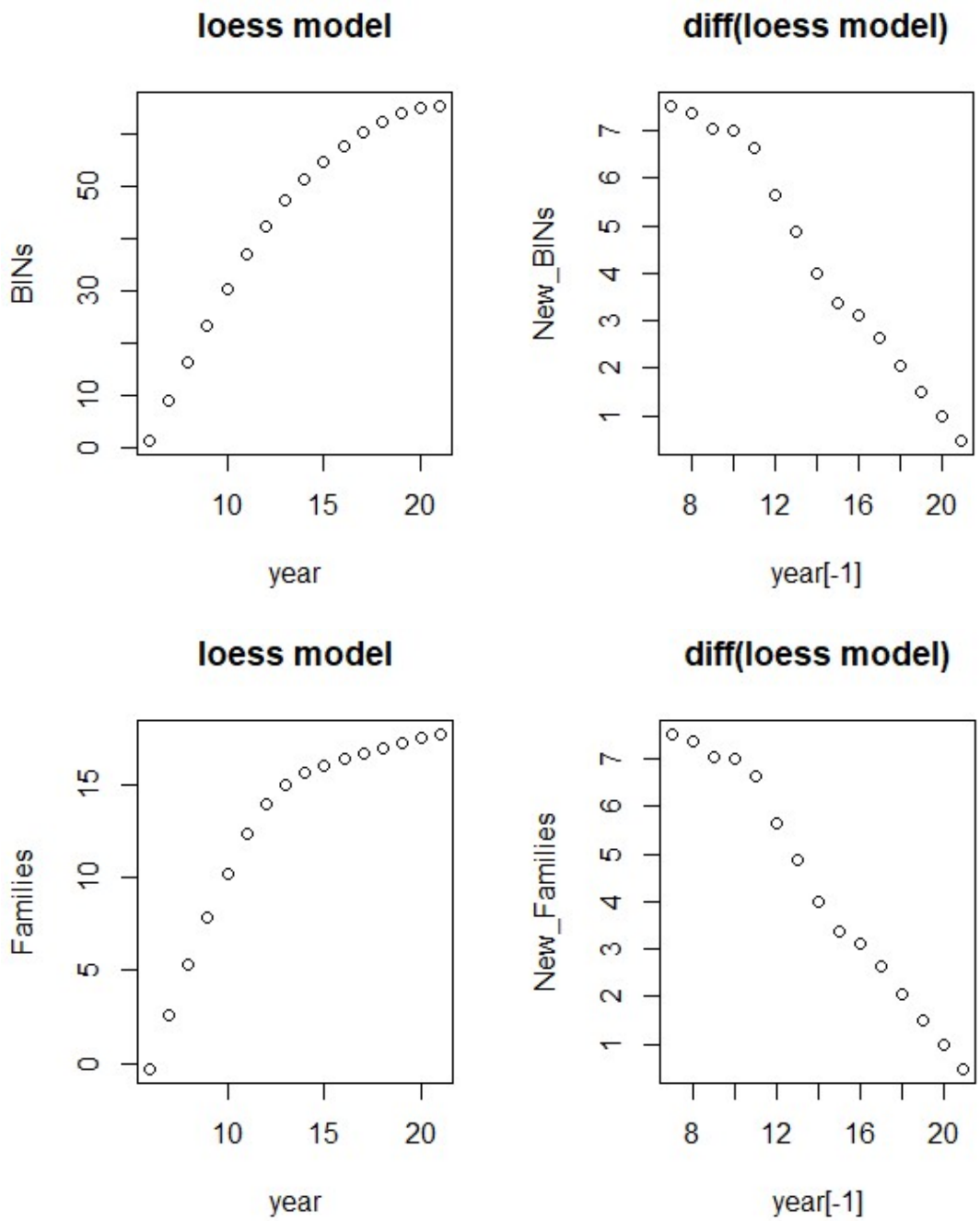



**Figure XV-2** Graphical representation of Loess models (left) of the increase in macroinvertebrate detection using public BOLD records for taxonomic assignment. The right plot is the result of subtracting one year's value (number of BINs or families) from the immediately preceding year to obtain the annual increase.

# XVI. Annex III

Supplementary information of section VIII “Unlocking rivers hidden diversity and ecological status using DNA metabarcoding in northwest Spain.”

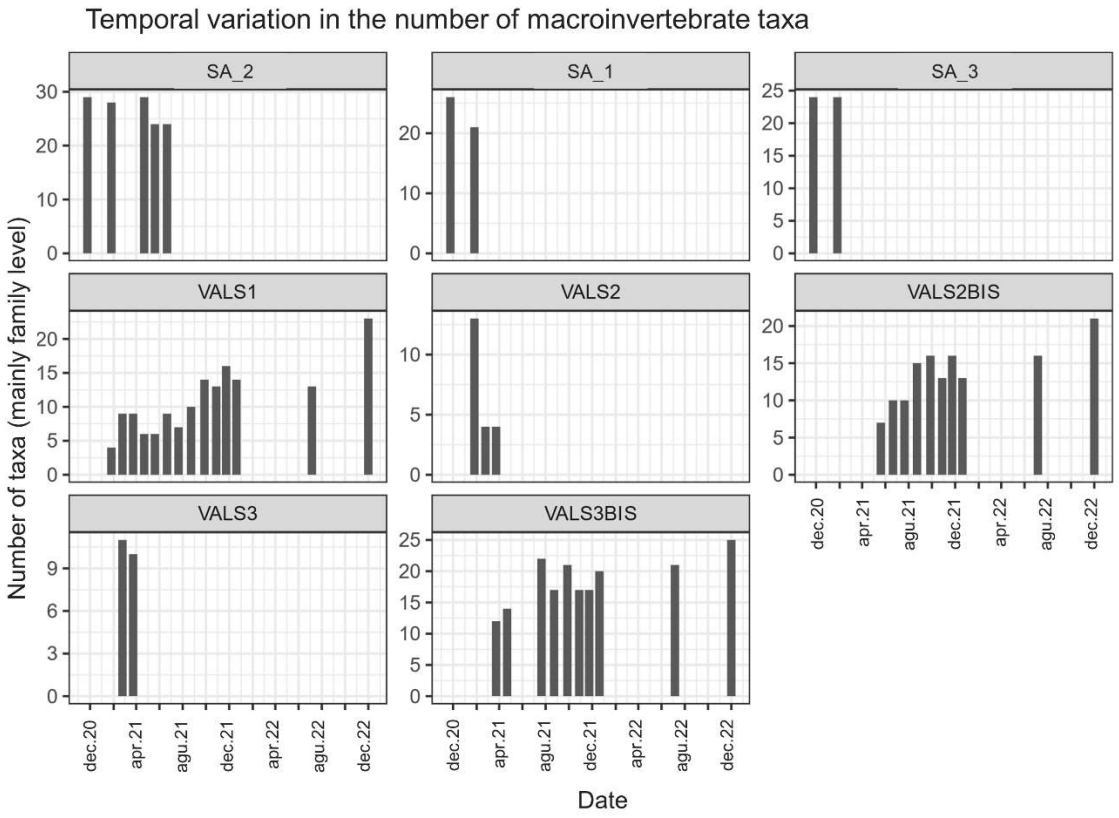



**Figure XVI-1** Variation in the number of different families over time in the sampling points of the Salentinos river (SA) and the Valseco stream (VALS).

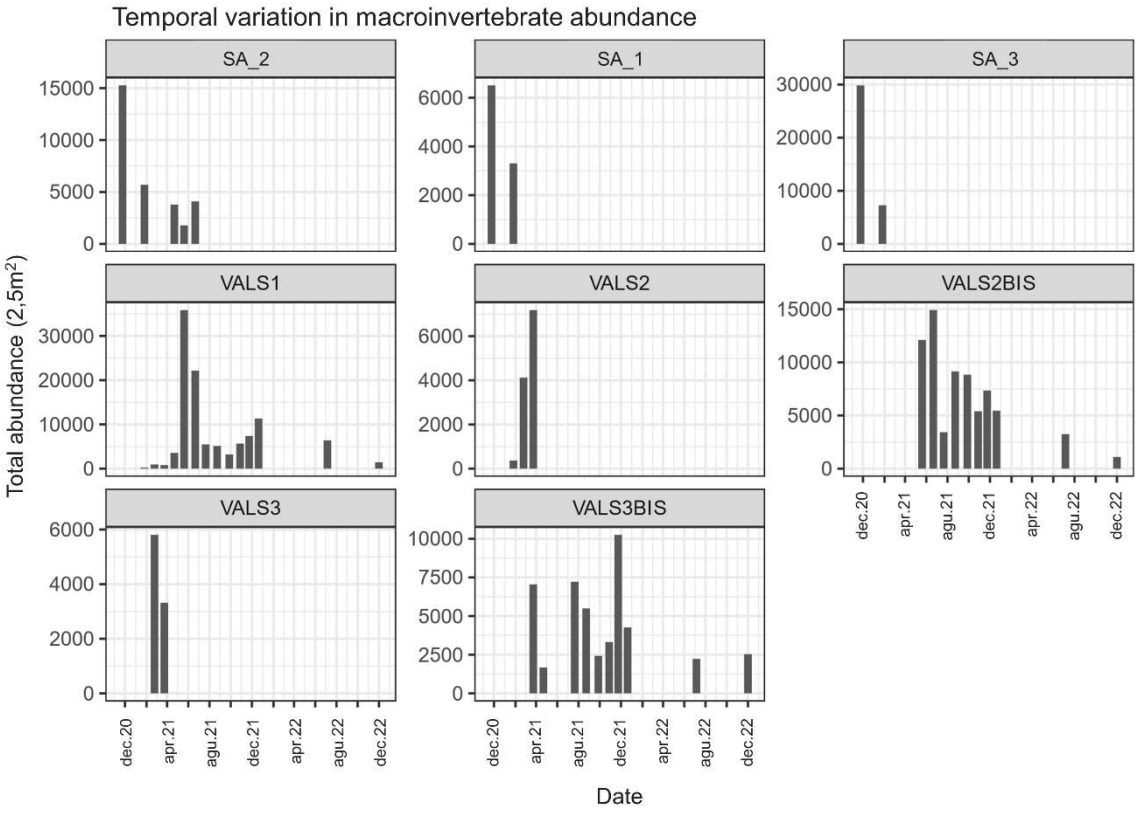



**Figure XVI-2** Total abundance of macroinvertebrates (individuals/2.5m2) of each of the samples taken throughout the time series. Valseco stream (VALS) and Salentinos river (SA).

## Species Analysis

The following graphs show the proportion of presence of the different species (detected in bulk samples by DNA metabarcoding) in all sampling points (27). Samples are categorised as being of good ecological status according to the official methodology (i.e. morphological identification) (those with good or very good status, 19 samples) and not good ecological status (those with moderate, poor or bad status, 8 samples). This approach is carried out with very few samples and from a very specific geographical area, with a low representation of samples that do not reach good status. In addition, the status categorisation was carried out with the same group as the one to be analysed (macroinvertebrates). This is a first approximation, which we recommend to repeat in the future, once most of the problems discussed in section IX have been solved, with a larger number of samples and complemented with data on anthropic pressure, pollutants, physico-chemical parameters and/or other bioindicator groups.

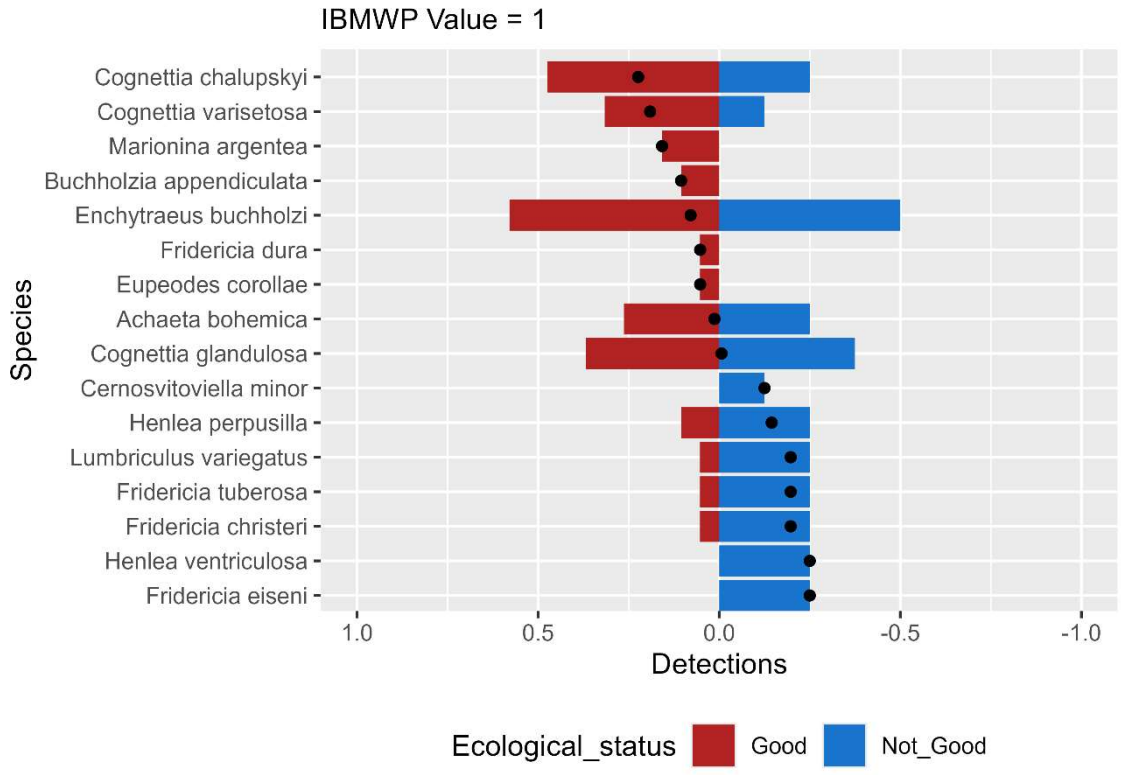
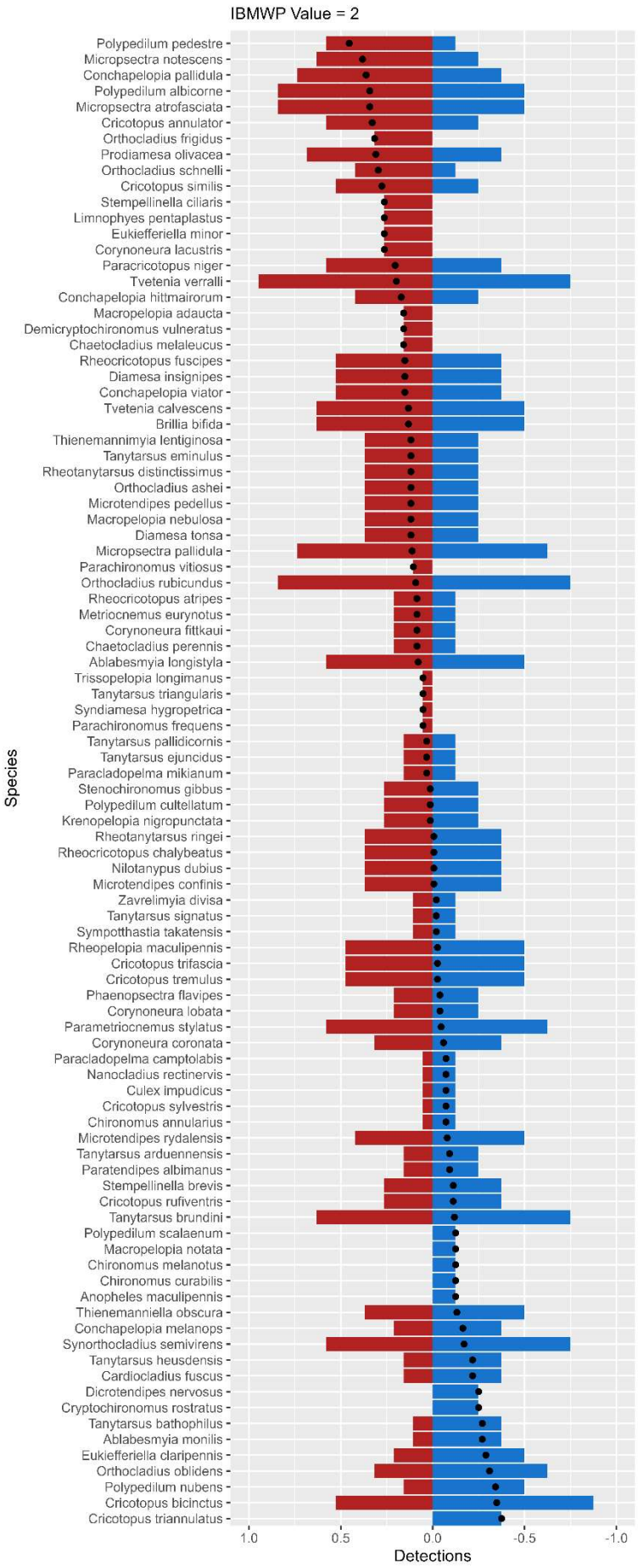
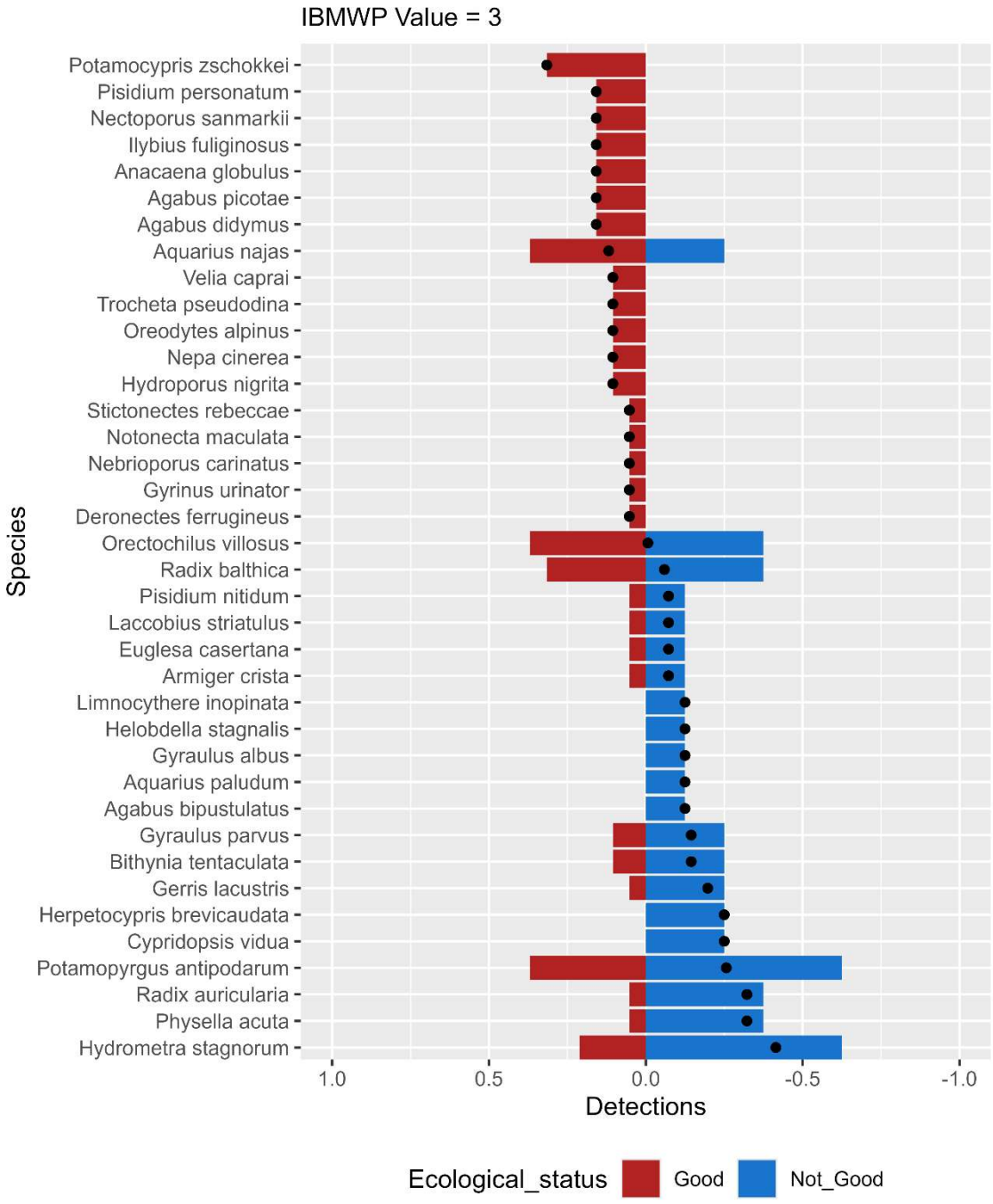
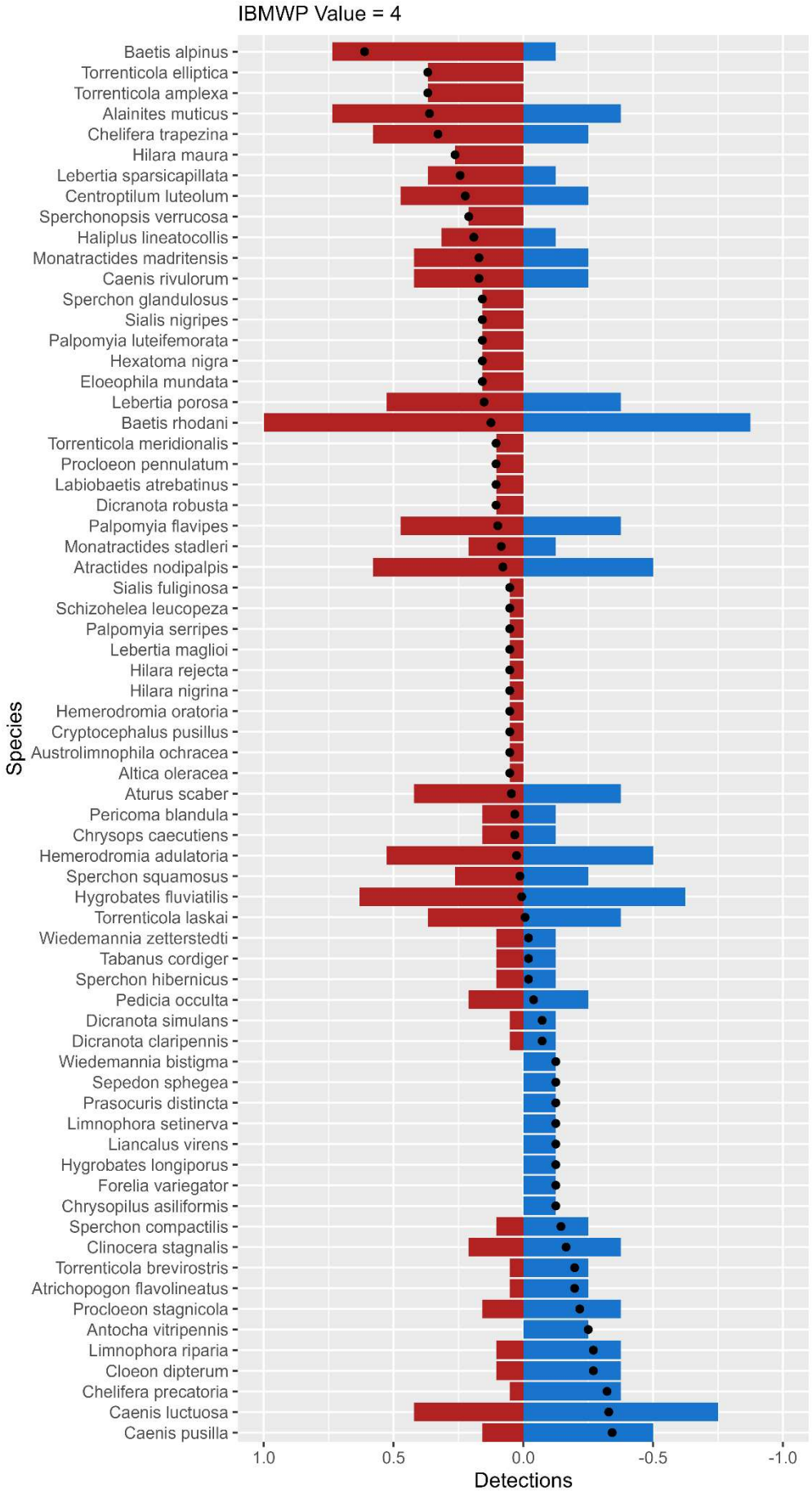
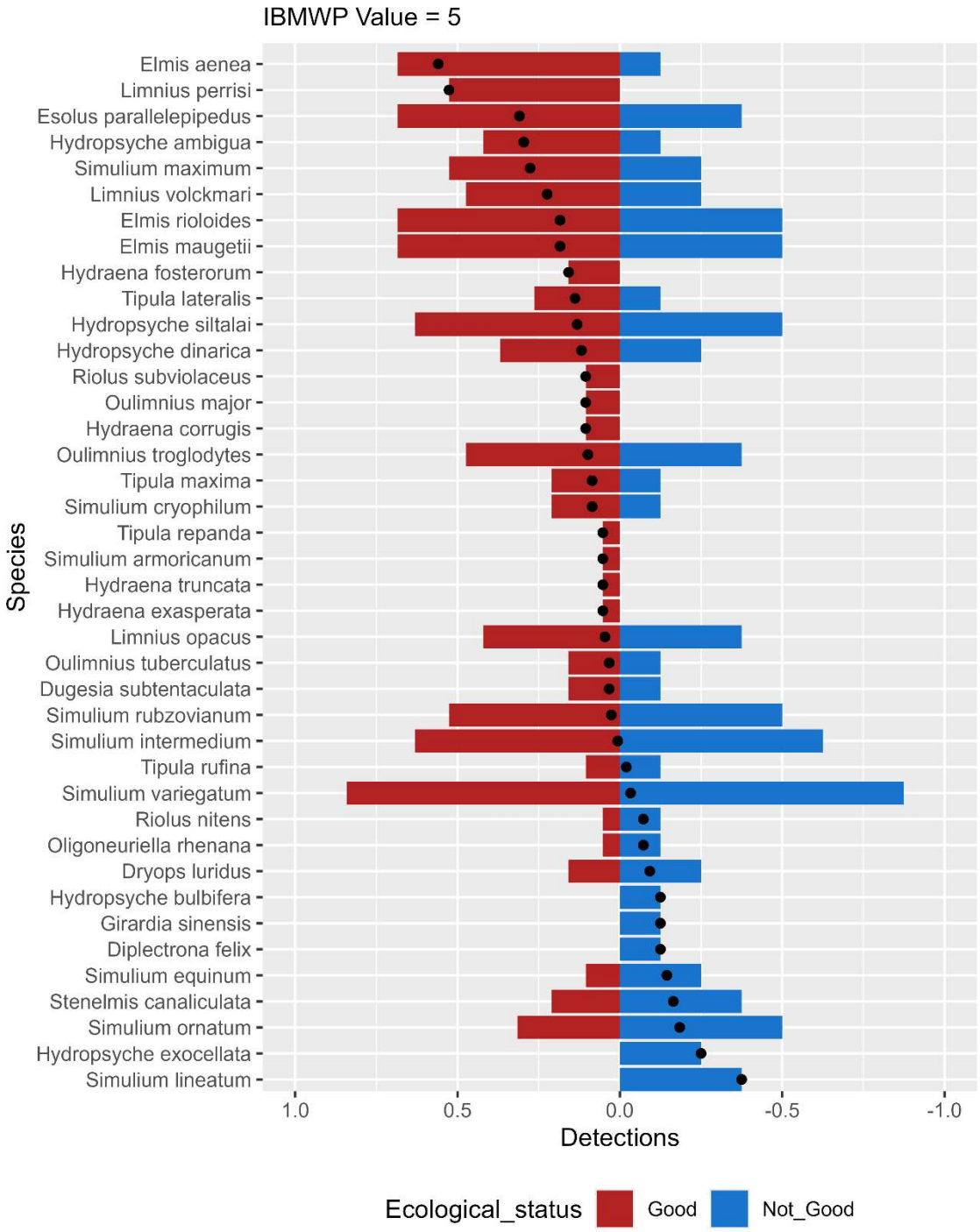


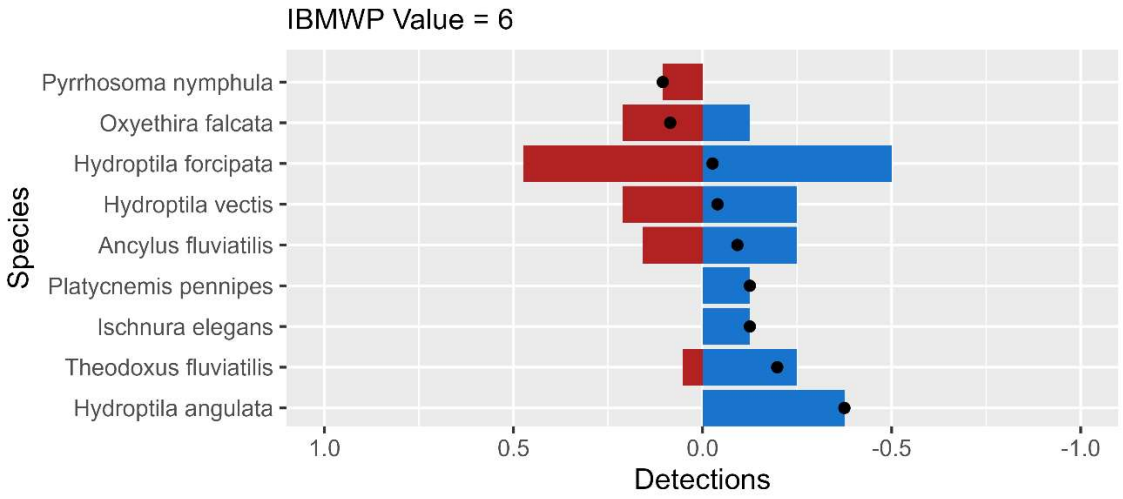

**Figure XVI-3** Proportion of occurrence of each species in bulk samples. Grouped into good and not good ecological status samples. Black dots are the difference between the proportions of occurrence in good and not good ecological status samples for each species. The figure continues on the following pages.

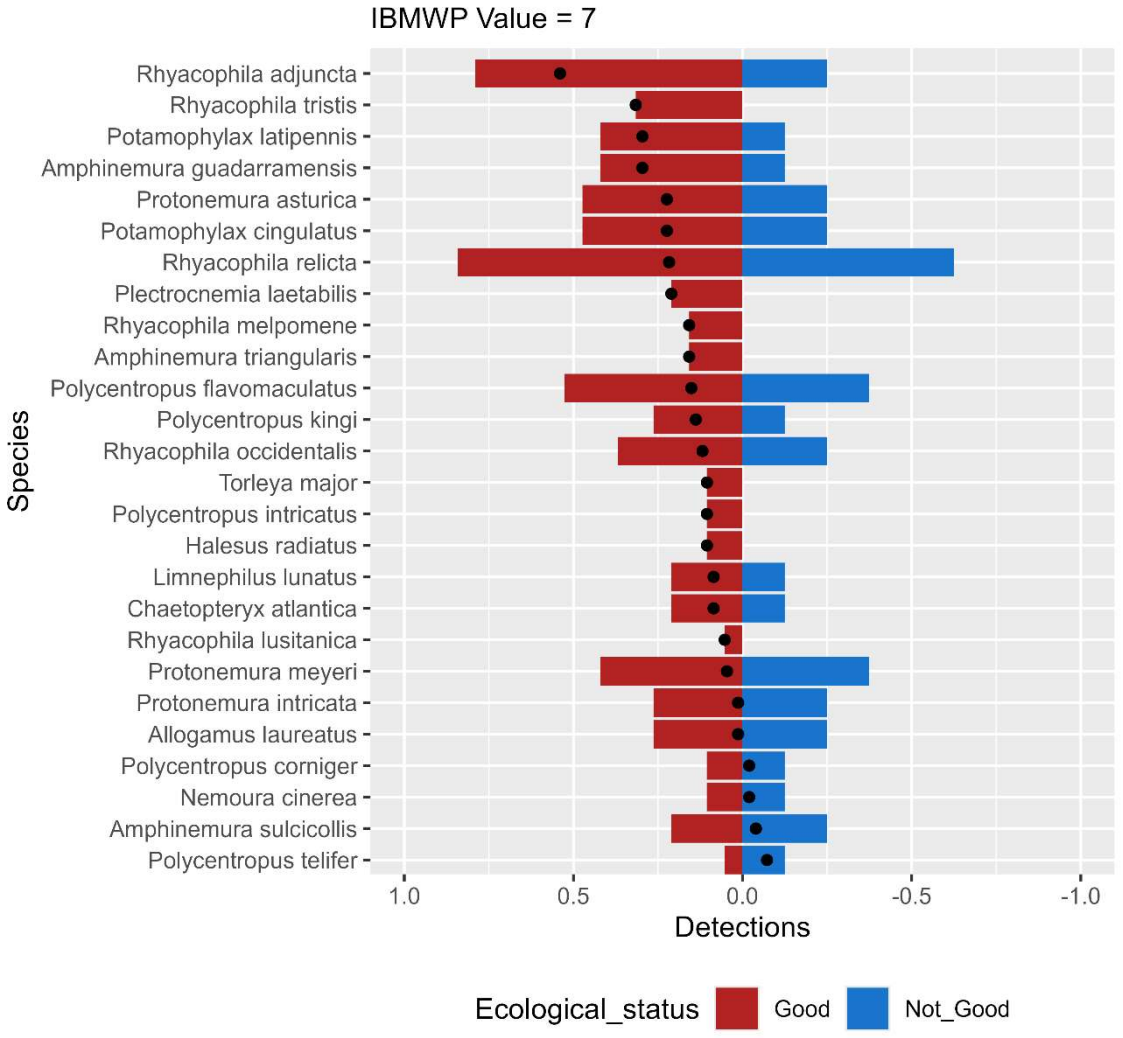

IBMWP Value = 2
Species
Polypedilum pedestre
Micropsectra notescens
Conchapelopia pallidula
Polypedilum albicorne
Micropsectra atrofasciata
Cricotopus annulator
Orthocladius frigidus
Prodiamesa olivacea
Orthocladius schnelli
Cricotopus similis
Stempellinella ciliaris
Limnophyes pentaplastus
Eukiefferiella minor
Corynoneura lacustris
Paracricotopus niger
Tvetenia verralli
Conchapelopia hittmairorum
Macropelopia adaucta
Demicryptochironomus vulneratus
Chaetocladius melaleucus
Rheocricotopus fuscipes
Diamesa insignipes
Conchapelopia viator
Tvetenia calvescens
Brillia bifida
Thienemannimyia lentiginosa
Tanytarsus eminulus
Rheotanytarsus distinctissimus
Orthocladius ashei
Microtendipes pedellus
Macropelopia nebulosa
Diamesa tonsa
Micropsectra pallidula
Parachironomus vitiosus
Orthocladius rubicundus
Rheocricotopus atripes
Metriocnemus eurynotus
Corynoneura fittkaui
Chaetocladius perennis
Ablabesmyia longistyla
Trissopelopia longimanus
Tanytarsus triangularis
Syndiamesa hygropetrica
Parachironomus frequens
Tanytarsus pallidicornis
Tanytarsus ejuncidus
Paracladopelma mikianum
Stenochironomus gibbus
Polypedilum cultellatum
Krenopelopia nigropunctata
Rheotanytarsus ringei
Rheocricotopus chalybeatus
Nilotanypus dubius
Microtendipes confinis
Zavrelimyia divisa
Tanytarsus signatus
Sympotthastia takatensis
Rheopelopia maculipennis
Cricotopus trifascia
Cricotopus tremulus
Phaenopsectra flavipes
Corynoneura lobata
Parametriocnemus stylatus
Corynoneura coronata
Paracladopelma camptolabis
Nanocladius rectinervis
Culex impudicus
Cricotopus sylvestris
Chironomus annularius
Microtendipes rydalensis
Tanytarsus arduennensis
Paratendipes albimanus
Stempellinella brevis
Cricotopus rufiventris
Tanytarsus brundini
Polypedilum scalaenum
Macropelopia notata
Chironomus melanotus
Chironomus curabilis
Anopheles maculipennis
Thienemanniella obscura
Conchapelopia melanops
Synorthocladius semivirens
Tanytarsus heusdensis
Cardiocladius fuscus
Dicrotendipes nervosus
Cryptochironomus rostratus
Tanytarsus bathophilus
Ablabesmyia monilis
Eukiefferiella claripennis
Orthocladius oblidens
Polypedilum nubens
Cricotopus bicinctus
Cricotopus triannulatus
1.0
0.5
0.0
-0.5
-1.0
Detections

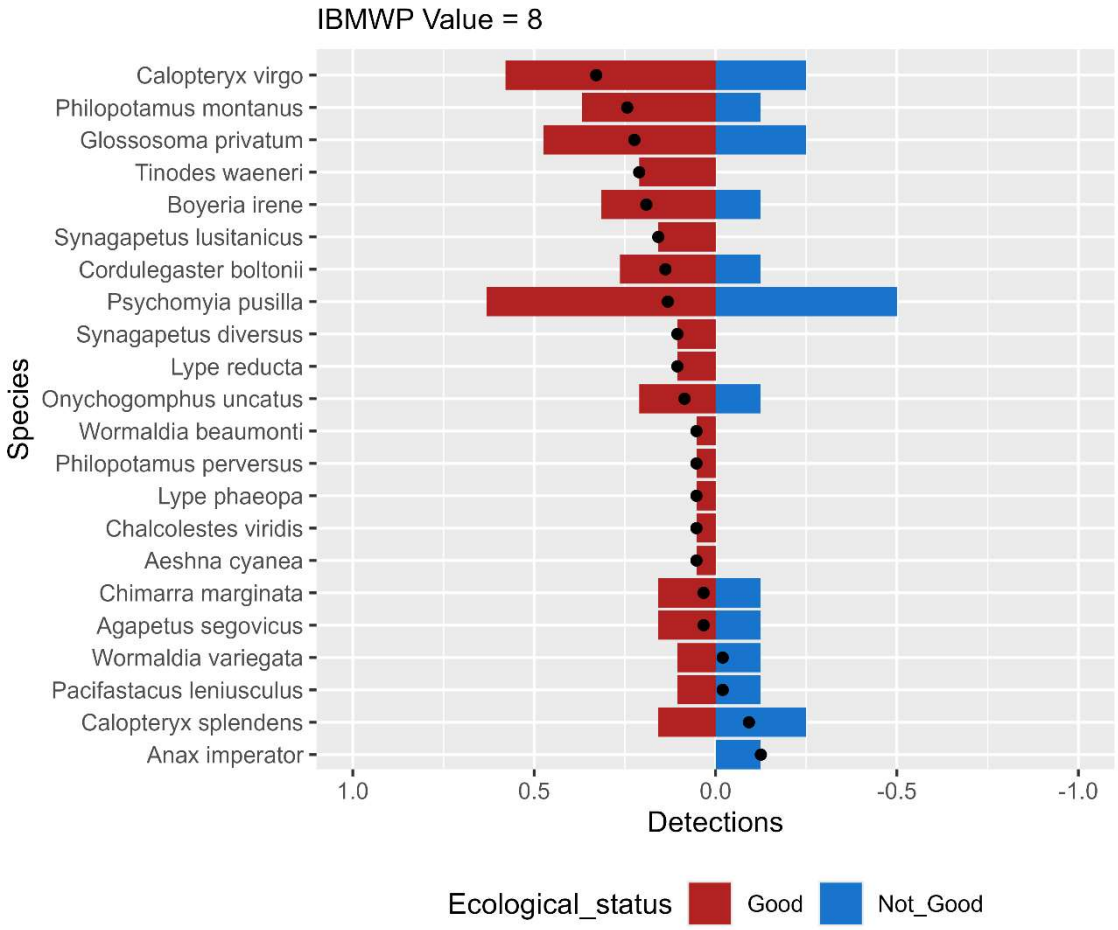

IBMWP Value = 3
Potamocypris zschokkei
Pisidium personatum
Nectoporus sanmarkii
Ilybius fuliginosus
Anacaena globulus
Agabus picotae
Agabus didymus
Aquarius najas
Velia caprai
Trocheta pseudodina
Oreodytes alpinus
Nepa cinerea
Hydroporus nigrita
Stictonectes rebeccae
Notonecta maculata
Nebrioporus carinatus
Gyrinus urinator
Deronectes ferrugineus
Orectochilus villosus
Radix balthica
Pisidium nitidum
Laccobius striatulus
Euglesa casertana
Armiger crista
Limnocythere inopinata
Helobdella stagnalis
Gyraulus albus
Aquarius paludum
Agabus bipustulatus
Gyraulus parvus
Bithynia tentaculata
Gerris lacustris
Herpetocypris brevicaudata
Cypridopsis vidua
Potamopyrgus antipodarum
Radix auricularia
Physella acuta
Hydrometra stagnorum
Species
1.0
0.5
0.0
-0.5
-1.0
Detections
Ecological_status
Good
Not_Good


IBMWP Value = 4
Species
Baetis alpinus
Torrenticola elliptica
Torrenticola amplexa
Alainites muticus
Chelifera trapezina
Hilara maura
Lebertia sparsicapillata
Centroptilum luteolum
Sperchonopsis verrucosa
Haliplus lineatocollis
Monatractides madritensis
Caenis rivulorum
Sperchon glandulosus
Sialis nigripes
Palpomyia luteifemorata
Hexatoma nigra
Eloeophila mundata
Lebertia porosa
Baetis rhodani
Torrenticola meridionalis
Procloeon pennulatum
Labiobaetis atrebatinus
Dicranota robusta
Palpomyia flavipes
Monatractides stadleri
Atractides nodipalpis
Sialis fuliginosa
Schizohelea leucopeza
Palpomyia serripes
Lebertia maglioi
Hilara rejecta
Hilara nigrina
Hemerodromia oratoria
Cryptocephalus pusillus
Austrolimnophila ochracea
Altica oleracea
Aturus scaber
Pericoma blandula
Chrysops caecutiens
Hemerodromia adulatoria
Sperchon squamosus
Hygrobates fluviatilis
Torrenticola laskai
Wiedemannia zetterstedti
Tabanus cordiger
Sperchon hibernicus
Pedicia occulta
Dicranota simulans
Dicranota claripennis
Wiedemannia bistigma
Sepedon sphegea
Prasocuris distincta
Limnophora setinerva
Liancalus virens
Hygrobates longiporus
Forelia variegator
Chrysopilus asiliformis
Sperchon compactilis
Clinocera stagnalis
Torrenticola brevirostris
Atrichopogon flavolineatus
Procloeon stagnicola
Antocha vitripennis
Limnophora riparia
Cloeon dipterum
Chelifera precatoria
Caenis luctuosa
Caenis pusilla
1.0
0.5
0.0
-0.5
-1.0
Detections


IBMWP Value = 5
Elmis aenea
Limnius perrisi
Esolus parallelepipedus
Hydropsyche ambigua
Simulium maximum
Limnius volckmari
Elmis rioloides
Elmis maugetii
Hydraena fosterorum
Tipula lateralis
Hydropsyche siltalai
Hydropsyche dinarica
Riolus subviolaceus
Oulimnius major
Hydraena corrugis
Oulimnius troglodytes
Tipula maxima
Simulium cryophilum
Tipula repanda
Simulium armoricanum
Hydraena truncata
Hydraena exasperata
Limnius opacus
Oulimnius tuberculatus
Dugesia subtentaculata
Simulium rubzovianum
Simulium intermedium
Tipula rufina
Simulium variegatum
Riolus nitens
Oligoneuriella rhenana
Dryops luridus
Hydropsyche bulbifera
Girardia sinensis
Diplectrona felix
Simulium equinum
Stenelmis canaliculata
Simulium ornatum
Hydropsyche exocellata
Simulium lineatum
Species
1.0
0.5
0.0
-0.5
-1.0
Detections
Ecological_status
Good
Not_Good


IBMWP Value = 6
Pyrrhosoma nymphula
Oxyethira falcata
Hydroptila forcipata
Hydroptila vectis
Ancylus fluviatilis
Platycnemis pennipes
Ischnura elegans
Theodoxus fluviatilis
Hydroptila angulata
Species
1.0
0.5
0.0
-0.5
-1.0
Detections


IBMWP Value = 7
Rhyacophila adjuncta
Rhyacophila tristis
Potamophylax latipennis
Amphinemura guadarramensis
Protonemura asturica
Potamophylax cingulatus
Rhyacophila relicta
Plectrocnemia laetabilis
Rhyacophila melpomene
Amphinemura triangularis
Polycentropus flavomaculatus
Polycentropus kingi
Rhyacophila occidentalis
Torleya major
Polycentropus intricatus
Halesus radiatus
Limnephilus lunatus
Chaetopteryx atlantica
Rhyacophila lusitanica
Protonemura meyeri
Protonemura intricata
Allogamus laureatus
Polycentropus corniger
Nemoura cinerea
Amphinemura sulcicollis
Polycentropus telifer
Species
1.0
0.5
0.0
-0.5
-1.0
Detections
Ecological_status
Good
Not_Good


IBMWP Value = 8
Calopteryx virgo
Philopotamus montanus
Glossosoma privatum
Tinodes waeneri
Boyeria irene
Synagapetus lusitanicus
Cordulegaster boltonii
Psychomyia pusilla
Synagapetus diversus
Lype reducta
Onychogomphus uncatus
Wormaldia beaumonti
Philopotamus perversus
Lype phaeopa
Chalcolestes viridis
Aeshna cyanea
Chimarra marginata
Agapetus segovicus
Wormaldia variegata
Pacifastacus leniusculus
Calopteryx splendens
Anax imperator
Species
1.0
0.5
0.0
-0.5
-1.0
Detections
Ecological_status
Good
Not_Good


IBMWP Value = 10
Micrasema longulum
Sericostoma flavicorne
Dinocras cephalotes
Sericostoma pyrenaicum
Epeorus assimilis
Leuctra aurita
Perla madritensis
Micrasema moestum
Leuctra fusca
Leuctra castillana
Odontocerum albicorne
Ephemera danica
Habroleptoides confusa
Habrophlebia eldae
Adicella reducta
Leuctra geniculata
Siphonoperla torrentium
Leuctra franzi
Isoperla grammatica
Thremma tellae
Habrophlebia lauta
Micrasema servatum
Triaenodes ochreellus
Chloroperla tripunctata
Leuctra hippopus
Electrogena lateralis
Thremma gallicum
Silo nigricornis
Perla grandis
Ecdyonurus dispar
Calamoceras marsupus
Oecetis testacea
Perlodes microcephalus
Mystacides azureus
Micrasema cenerentola
Calliarcys humilis
Atrichops crassipes
Athripsodes albifrons
Lepidostoma hirtum
Leuctra inermis
Capnioneura libera
Larcasia partita
Ecdyonurus venosus
Ceraclea dissimilis
Taeniopteryx schoenemundi
Setodes argentipunctellus
Atherix ibis
Choroterpes picteti
Species
1.0
0.5
0.0
-0.5
-1.0
Detections
Ecological_status
Good
Not_Good

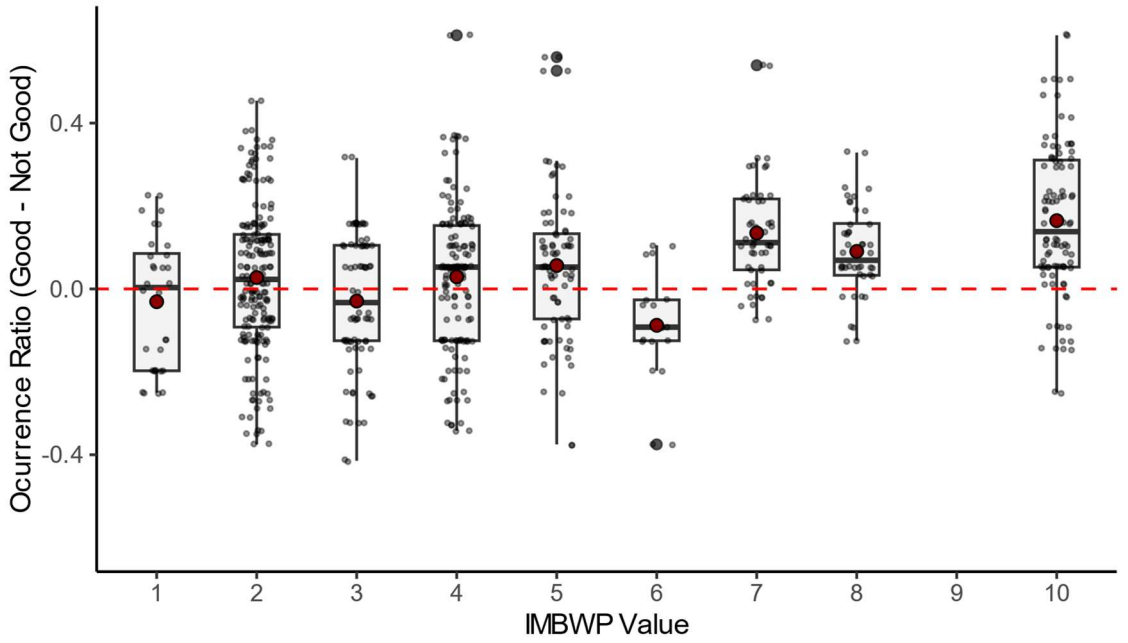



**Figure XVI-4** Summary graph of occurrence difference values at points in good and not good condition (black dots of graph Figure XVI-3). Each grey dot is a Species. Red dot represents the mean value. The occurrence ratio is the difference between the proportion of occurrences of each species at sampling sites with good (positive) or not good (negative) ecological status (inferred from morphological data). Dots above the red dashed line represent a species that occurs more often at sampling sites in good conditions than at sampling sites in not good conditions.

**Table XVI-1** Number of species losses required for a change in ecological status to occur in bulk samples. Mean values obtained after 1000 simulation replicates. Removal of species was weighted according to their sensitivity (i.e. a species with an IBMWP value of 10 was 10 times more likely to be removed than a species with an IBMWP value of 1).

| *Sampling_point* | *Very Good to Good* | *Good to Moderate* | *Moderate to Poor* | *Poor to Bad* |
|---|---|---|---|---|
| *BA-M01* | 5 | 20 | 15 | 14 |
| *PI-M01-BA-M02* | 4 | 17 | 14 | 15 |
| *PI-M02* | 4 | 19 | 14 | 14 |
| *TA-CURVA* | 8 | 25 | 19 | 20 |
| *VA-M01* | 14 | 25 | 16 | 15 |
| *VAI-M02* | 6 | 15 | 14 | 13 |
| *VAII-M01* | 8 | 16 | 12 | 11 |
| *VAII-M03* | 4 | 18 | 12 | 13 |
| *VALS-1_22* | 2 | 13 | 12 | 12 |
| *VALS-2BIS_21* | 3 | 13 | 10 | 10 |
| *VALS-2BIS_22* | 3 | 14 | 12 | 10 |
| *VALS-3BIS_21* | 6 | 11 | 10 | 12 |
| *VALS-3BIS_22* | 31 | 20 | 15 | 13 |
| *GR-M02* | - | 23 | 15 | 12 |
| *GR-M03* | - | 19 | 14 | 12 |
| *PON-M01* | - | 2 | 11 | 11 |
| *SOR-M01* | - | 6 | 16 | 17 |
| *SOR-M02* | - | 9 | 17 | 17 |

| | | | | |
|---|---|---|---|---|
| *TA-ANZO* | - | 10 | 15 | 17 |
| *TA-CASA* | - | 12 | 17 | 17 |
| *TA-CENTRAL* | - | 12 | 15 | 17 |
| *TA-DEBOYU* | - | 20 | 16 | 13 |
| *VA-M02* | - | 18 | 15 | 14 |
| *VALS-1_21* | - | 13 | 10 | 11 |
| *TA-PIE* | - | - | 13 | 14 |
| *TA-PISCI* | - | - | 15 | 14 |
| *GR-M01* | - | - | - | 4 |
| *Mean* | *7.5* | *15.4* | *14* | *13.4* |

**Table XVI-2** Number of species losses required for a change in ecological status to occur in bulk samples. Mean values obtained after 1000 simulation replicates. Species removal was completely random.

| *Sampling_point* | *Very Good to Good* | *Good to Moderate* | *Moderate to Poor* | *Poor to Bad* |
|---|---|---|---|---|
| *BA-M01* | 5 | 24 | 16 | 15 |
| *PI-M01-BA-M02* | 5 | 21 | 15 | 16 |
| *PI-M02* | 5 | 21 | 17 | 16 |
| *TA-CURVA* | 9 | 31 | 22 | 18 |
| *VA-M01* | 16 | 28 | 18 | 16 |
| *VAI-M02* | 7 | 20 | 14 | 15 |
| *VAII-M01* | 9 | 20 | 13 | 11 |
| *VAII-M03* | 4 | 23 | 14 | 12 |
| *VALS-1_22* | 2 | 17 | 14 | 13 |
| *VALS-2BIS_21* | 3 | 17 | 12 | 11 |
| *VALS-2BIS_22* | 4 | 16 | 14 | 13 |
| *VALS-3BIS_21* | 7 | 14 | 12 | 12 |
| *VALS-3BIS_22* | 38 | 21 | 15 | 13 |
| *GR-M02* | - | 27 | 17 | 14 |
| *GR-M03* | - | 22 | 15 | 15 |
| *PON-M01* | - | 3 | 13 | 13 |
| *SOR-M01* | - | 6 | 20 | 18 |
| *SOR-M02* | - | 11 | 21 | 17 |
| *TA-ANZO* | - | 13 | 18 | 17 |
| *TA-CASA* | - | 13 | 22 | 19 |
| *TA-CENTRAL* | - | 14 | 19 | 19 |
| *TA-DEBOYU* | - | 24 | 18 | 14 |
| *VA-M02* | - | 21 | 18 | 16 |
| *VALS-1_21* | - | 16 | 11 | 13 |
| *TA-PIE* | - | - | 16 | 15 |
| *TA-PISCI* | - | - | 19 | 15 |
| *GR-M01* | - | - | - | 4 |
| *Mean* | *8.8* | *18.5* | *16.3* | *14.4* |

# XVII. Annex IV

Supplementary information of section VII “Back to life after 50 years: Impact of environmental flow restoration on macroinvertebrate biodiversity and river ecological status. A case study of the Valseco stream, Spain.”

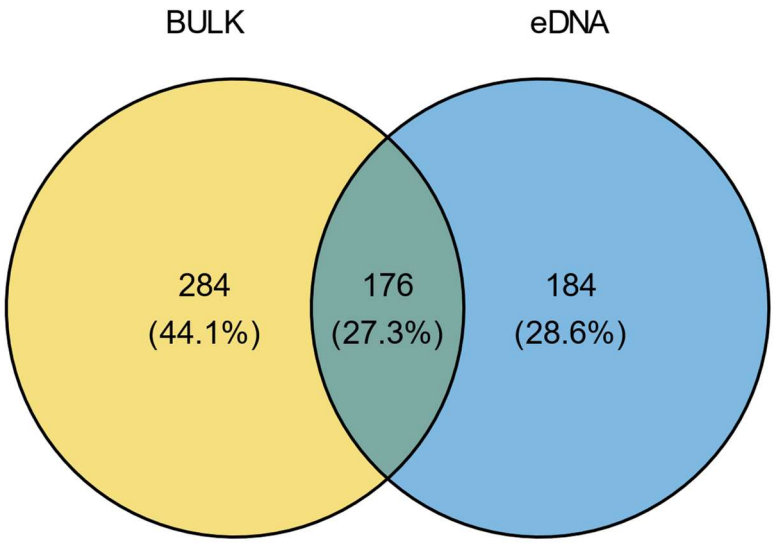



**Figure XVII-1** Venn diagram showing the species diversity shared between bulk and eDNA samples.

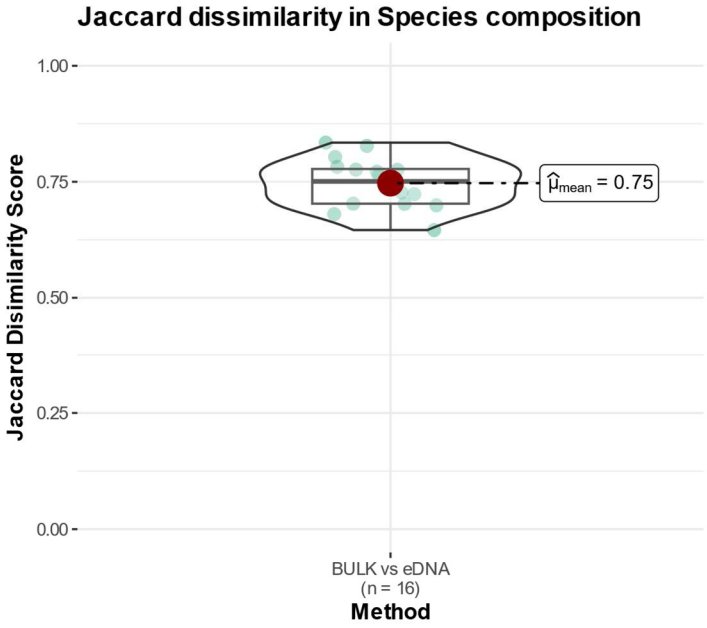



**Figure XVII-2** Jaccard dissimilarity scores between Species composition of eDNA and bulk samples collected at the same sampling point.

# XVIII. Annex V

Supplementary information of section IX "General Discussion: Data-based considerations for implementing macroinvertebrate DNA metabarcoding for effective river biomonitoring in peninsular Spain."

## Statistical analysis

All statistical analyses were performed in R v4.3.1. A PERMANOVA test was run using the adonis2 function from the vegan v2.6 package to explore the effects of methodology used on the variance of IBMWP taxa composition. A total of 99,999 permutations were performed, and the marginal effects of each variable were tested. Only the samples with both molecular methods applied were tested to have a balanced design. Jaccard dissimilarity was calculated using the vegan v2.6 package, normality of the results was tested using Shapiro-wilk test and values were plotted and analysed using ggstatsplot package. The relationships between the variable of macroinvertebrate family detection by metabarcoding and the abundance of these families in morphological samples were analysed statistically. Two different models were used: a linear model and a logistic model for binomial data. First, a logarithmic transformation of the abundance variable was performed. Then, linear and logistic models are fitted to the data using the lm and glm function, respectively. After fitting the models, a permutation test was carried out to assess the significance of the relationship between the explanatory variable and the response variable in both models and 10,000 replications were performed to generate a simulated p-value. To study the possible effect of the biovolume of individuals with low abundance in the sample, the relationship of the detection of the different families in the metabarcoding samples was analysed with respect to the sieve in which they were detected in the morphological identification (Large: 5 mm, Medium: 1 mm and Small: 0.5 mm). Chi-square analyses were performed for each abundance value less than 5 in which families were not detected by metabarcoding.

## Comparison of IBMWP taxa composition

Considering all the samples and techniques together of the 86 IBMWP taxa detected, 6 were detected only by eDNA, 5 by bulk and 4 by morphological identification (Figure XVIII-1). Morphotaxonomy detected a total of 64 different IBMWP taxa, fewer than the bulk approach which detected 75 taxa, but more than the eDNA samples which only detected 58 IBMWP taxa. However, it should be noted that the number of eDNA samples is lower than the other two sample types (16 eDNA, 27 bulk and morpho).

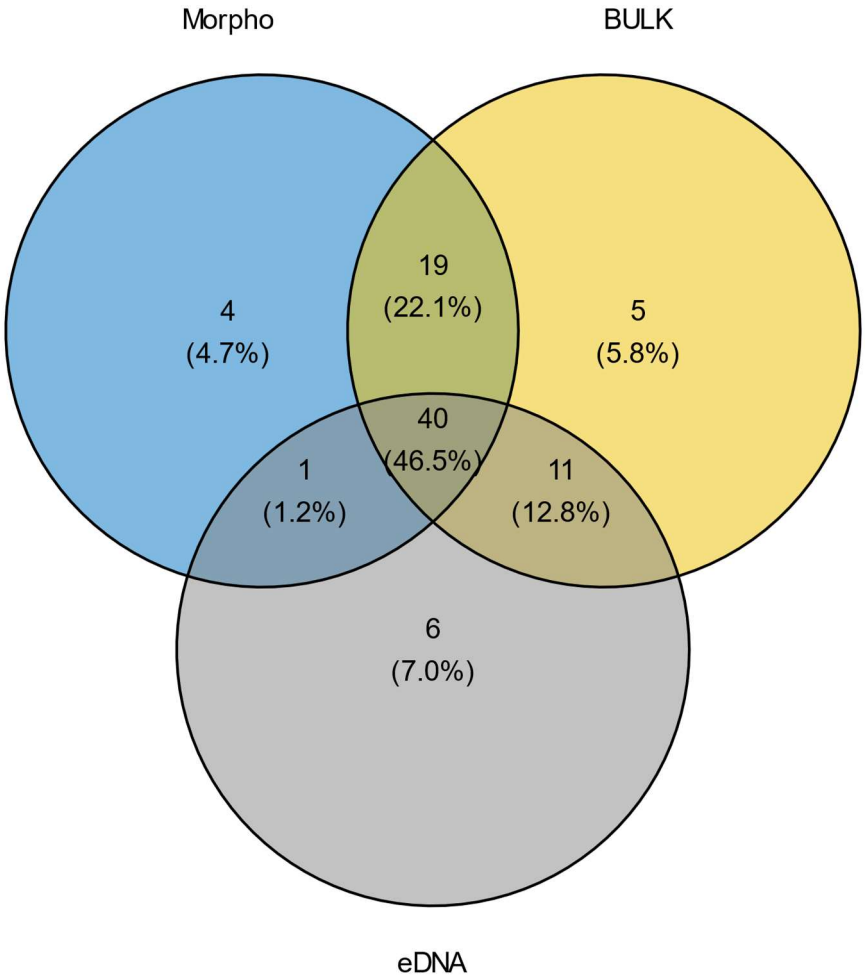



**Figure XVIII-1** Venn Diagram of IBMWP taxa composition of all samples together for each methodological approach. Number of samples considered: 27 bulk, 27 Morphotaxonomy, 16 eDNA.

**Table XVIII-1** PERMANOVA analysis of macroinvertebrate composition using different approaches.

| | R2 | P-VALUE |
|---|---|---|
| **MORPHO VS EDNA** | **0.19** | **0.0001** |
| **MORPHO VS BULK** | **0.07** | **0.002** |
| **BULK VS EDNA** | **0.26** | **0.0001** |

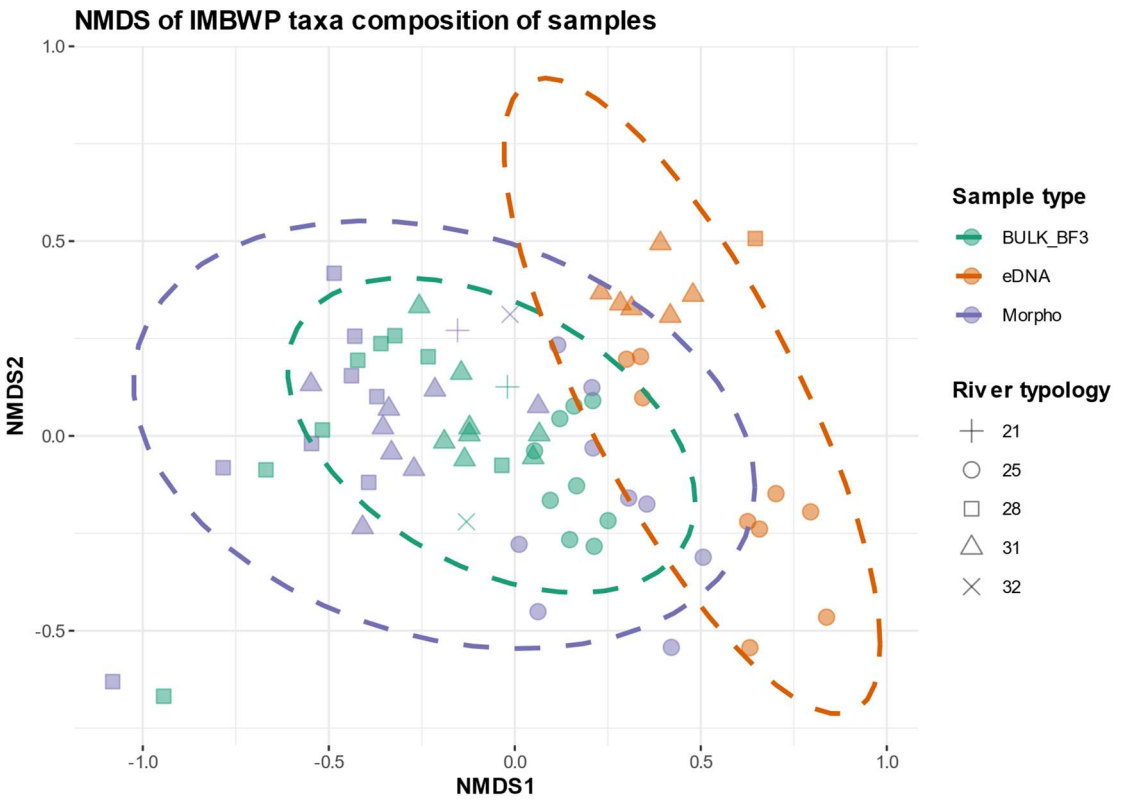



**Figure XVIII-2** NMDS analysis of IBMWP taxa composition. Stress: 0.173. Procrustes: rmse 6.283453e-05, max resid 0.000245

**Table XVIII-2** PERMANOVA results

| | DF | SUMOFSQS | $R^2$ | F | P-VALUE |
|---|---|---|---|---|---|
| **METHODOLOGY** | 2 | 1.4396 | 0.15910 | 14.2982 | 0.001 *** |
| **SAMPLING_POINT** | 26 | 5.0437 | 0.55744 | 3.8535 | 0.001 *** |
| **RESIDUAL** | 41 | 2.0640 | 0.22811 | | |

**Table XVIII-3** Inconsistencies in the taxonomy of the detected species at the level considered by the IBMWP index.

| | PHYLUM | CLASS | ORDER | FAMILY | GENUS | SPECIES |
|---|---|---|---|---|---|---|
| **BOLD:** | Arthropoda | Insecta | Trichoptera | Thremmatidae | Thremma | Thremma tellae |
| **GBIF:** | Arthropoda | Insecta | Trichoptera | Uenoidae | Thremma | Thremma tellae |
| **BOLD:** | Mollusca | Gastropoda | Pulmonata | Ancylidae | Ancylus | Ancylus fluviatilis |
| **GBIF:** | Mollusca | Gastropoda | | Planorbidae | Ancylus | Ancylus fluviatilis |
| **BOLD:** | Arthropoda | Insecta | Hemiptera | Corixidae | Micronecta | |
| **GBIF:** | Arthropoda | Insecta | Hemiptera | Micronectidae | Micronecta | |

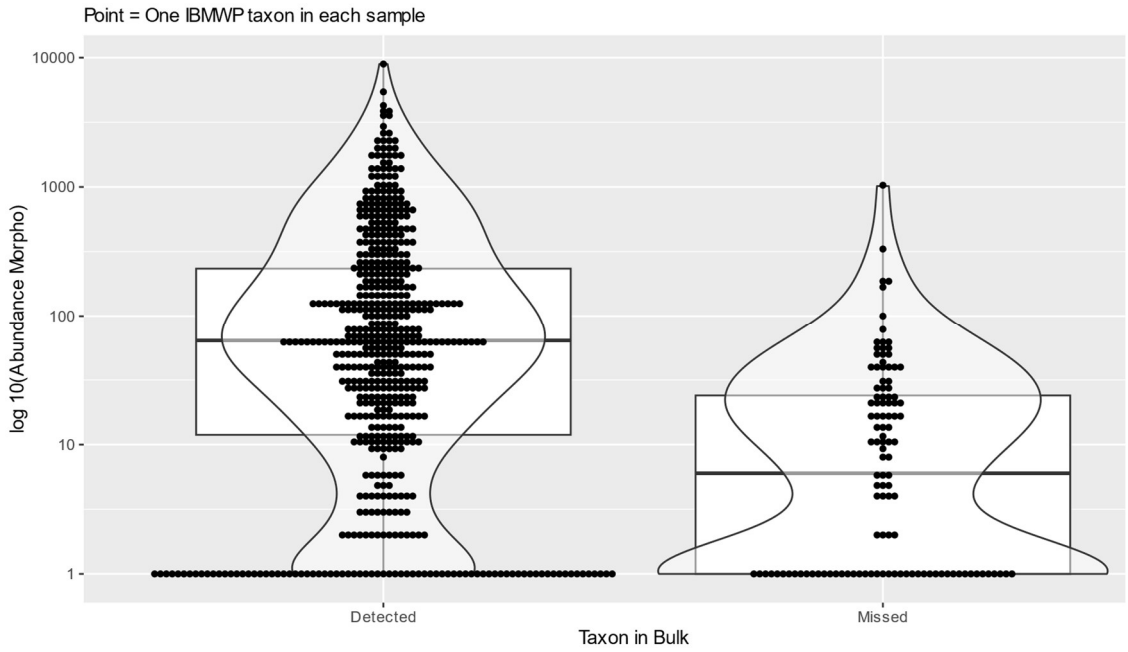



**Figure XVIII-3** Detectability by metabarcoding of different IBMWP taxa in relation to their abundance in morphological samples.

## Macroinvertebrate sampling comparative test

On 23 August 2017, a sampling was carried out at the same time and sampling point (coordinates UTM30; upper point X: 263895, Y: 4799614; lower point X: 263811, Y: 4799527) by 4 different samplers following the official sampling protocol of the Spanish Ministry of Environment (MAGRAMA, 2013b). While the values for number of taxa, EQR and abundance do not vary considerably (Table XVIII-4), the composition of families varies greatly (Figure XVIII-4). Only 21 families (44.7% of the total) are shared among the 4 samplers. While 14 families (almost 30%) have only been detected by one of the four samplers.

**Table XVIII-4** Sampling test. IBMWP values calculated according to the official Spanish Ministry of Environment protocol (MAGRAMA, 2013a). * Different samplers.

| | **JCE*** | **JGC*** | **AEM*** | **EMC*** | Mean $\bar{x}$ | Standard deviation **σ** | **σ/$\bar{x}$** |
|---|---|---|---|---|---|---|---|
| IBMWP | 156 | 167 | 154 | 162 | 159,75 | 5,909 | 0,037 |
| Number of Taxa | 30 | 34 | 32 | 32 | 32 | 1,633 | 0,051 |
| Abundance | 8.229 | 11.635 | 13.307 | 13.731 | 11725,5 | 2500,458 | 0,213 |
| EQR | 0,83 | 0,79 | 0,78 | 0,78 | 0,79 | 0,025 | 0,031 |
| Status | Good | Good | Good | Good | | | |

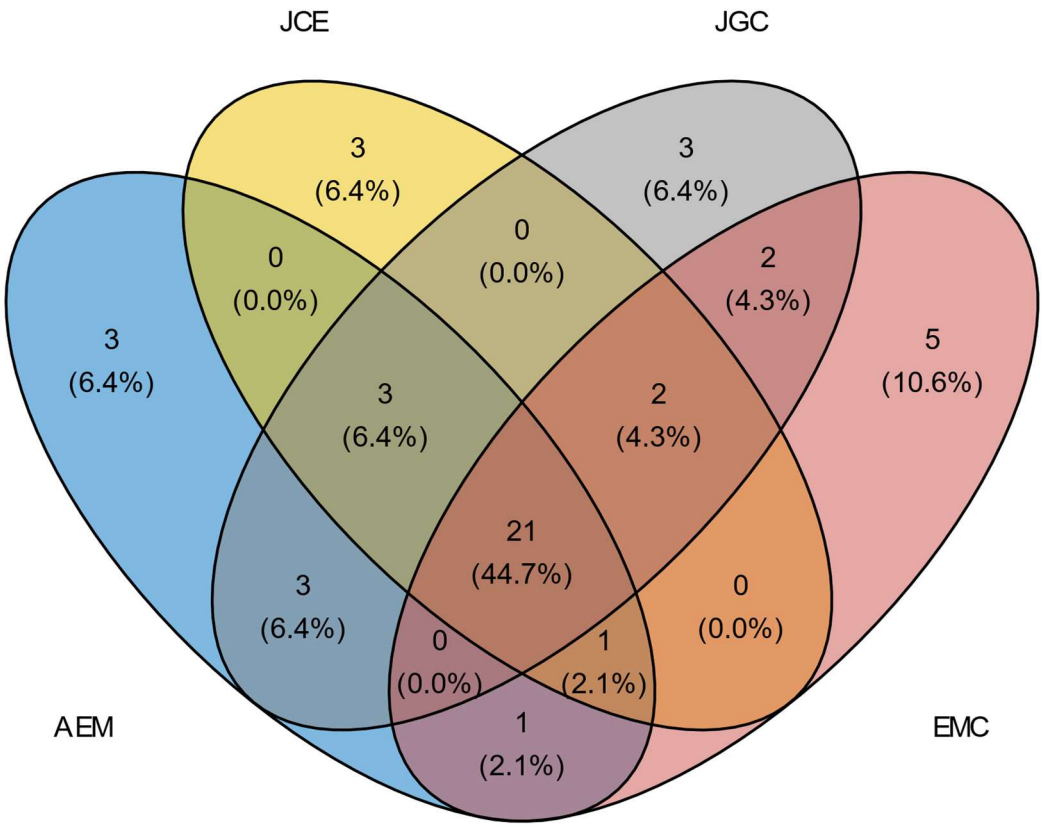



**Figure XVIII-4** Venn Diagram of IBMWP taxa composition of the sample collected by each sampler.