

# Towards an Iberian DNA barcode reference library of freshwater macroinvertebrates and fishes

Cesc Múrria<sup>1,2,\*</sup>, Leif O. S. Väisänen<sup>3</sup>, Simona Somma<sup>1</sup>, Owen S. Wangensteen<sup>4</sup>, Miquel A. Arnedo<sup>2</sup> and Narcís Prat<sup>1</sup>

<sup>1</sup> Grup de Recerca Freshwater Ecology and Management, Departament de Biologia Evolutiva, Ecologia i Ciències Ambientals, Universitat de Barcelona, Catalonia.

<sup>2</sup> Institut de Recerca de la Biodiversitat (IRBio), Departament de Biologia Evolutiva, Ecologia i Ciències Ambientals, Universitat de Barcelona, Catalonia.

<sup>3</sup> Stream Ecology Research Group, Department of Ecology and Genetics, University of Oulu, Finland.

<sup>4</sup> Norwegian College of Fishery Science, UiT The Arctic University of Norway, Tromsø, Norway.

\* Corresponding author: cmurria@ub.edu

Received: 16/05/18

Accepted: 20/05/19

## ABSTRACT

### Towards an Iberian DNA barcode reference library of freshwater macroinvertebrates and fishes

Freshwater macroinvertebrates and fishes are focal groups in major ecosystem biomonitoring programs in the Iberian Peninsula. Yet, their use as bioindicators is sometimes constrained by the time and cost needed for sorting macroinvertebrates specimens and their challenging taxonomic identification, and the huge sampling procedures for capturing rare or incipient fish species, respectively. Given the increasing cost-effectiveness of metazoan identification based on metabarcoding [i.e., high-throughput sequencing (HTS) of DNA barcodes] and reliability of species-level identification and the high number of samples that can be processed, its use in biomonitoring of freshwater communities can provide an alternative to morphology-based approaches. However, the accuracy of species assignment in metabarcoding approaches relies on the availability of a comprehensive DNA barcode reference library. Because of the high level of endemism in the Iberian Peninsula, current public repositories for DNA barcodes may not be informative enough to identify the Iberian fauna to species level. Here, we compiled the Iberian freshwater macroinvertebrates and fishes taxonomic list (including indigenous and non-indigenous species) and the available molecular data for the cytochrome oxidase I DNA barcode (*cox1*, COI-5P) in public repositories to assess the extent of DNA barcode coverage. The DNA barcode coverage was reported for DNA fragments within the Folmer region (658 bp). Given that HTS platforms provide DNA sequence in the range of 50-400 bp in length, we also reported the second half of the DNA barcode (313 bp, Leray region) and the first part of the Leray region (285 bp, Leray-285), which are short DNA barcodes useful to assign metabarcoding *cox1* data. For macroinvertebrates, the final taxonomic checklist comprises 3348 species including Mollusca (65 species), Crustacea (101 species) and Insecta (3182 species). We present an initial DNA barcode reference library, with an overall coverage of ~ 35 % of the Iberian taxa. Exploring this data, we find a strong taxonomic bias. Based on Leray-285, Odonata (43 of 79 species barcoded, 54.43 %) and Hemiptera (44 of 81 species barcoded, 54.32 %) were the best represented lineages. In contrast, Diptera (393 of 1693 species barcoded, 23.21 %), and Plecoptera (42 of 135 species barcoded, 31.11 %) were underrepresented. For fishes, the available DNA barcode data covered 98.11 % of the indigenous (76) and non-indigenous (30) species. By revealing and quantifying current gaps on the available data (~ 65 %), we aim to improve efficiency in designing the next steps towards the ambitious yet necessary goal of compiling a complete DNA barcode reference library for Iberian macroinvertebrates and fishes.

**Key words:** bioassessment, bioindicator, conservation, DNA metabarcoding, eDNA, species richness, streams, taxonomic resolution, water quality

## RESUMEN

**Hacia una biblioteca ibérica de referencia de códigos de barras genéticos de los macroinvertebrados y peces de agua dulce**  
*Los macroinvertebrados y peces de agua dulce son los organismos más usados en los principales programas de biomonitorio*

de ecosistemas en la Península Ibérica. Todavía su uso como bioindicadores se ha visto a veces obstaculizado por el tiempo y el coste necesarios para clasificar los especímenes de macroinvertebrados y su difícil identificación taxonómica, y las dificultades de muestreo para capturar especies de peces raras o incipientes, respectivamente. Dada la reducción en coste y tiempo que supone la identificación de metazoos mediante metabarcoding [es decir, secuenciación de alto rendimiento (HTS) de códigos de barras de ADN] la confiabilidad de la identificación a nivel de especie y el alto número de muestras que se pueden procesar, su uso en biomonitorización de comunidades de agua dulce, puede proporcionar una alternativa a las aproximaciones tradicionales basadas en la morfología. Sin embargo, la precisión de la asignación de especies en dichas técnicas de metabarcoding requiere disponer de una exhaustiva biblioteca de referencia de códigos de barras de ADN. Debido al alto nivel de endemismo en la Península Ibérica, los repositorios públicos actuales de códigos de barras de ADN pueden no ser lo suficientemente informativos para identificar la fauna ibérica a nivel de especie. En este estudio hemos compilado la lista taxonómica de macroinvertebrados y peces ibéricos de agua dulce (incluidas las especies autóctonas y no autóctonas) y los datos moleculares disponibles en repositorios públicos para el código de barras genético Citocromo Oxidasa C Subunidad I (cox1, COI-5P), para evaluar el alcance y extensión de su cobertura. La cobertura del código de barras del ADN se compiló para fragmentos de ADN ubicados dentro de la región de Folmer (658 pb). Dado que las plataformas HTS proporcionan una secuencia de ADN de un rango de 50-400 pb de longitud, también compilamos información de la segunda mitad del código de barras de ADN (313 pb, región de Leray) y la primera parte de la región de Leray (285 pb, Leray-285), que son los fragmentos de ADN más cortos y todavía útiles para asignar secuencias de cox1 obtenidas por metabarcoding. Para los macroinvertebrados, la lista taxonómica comprendió 3348 especies, incluidos Mollusca (65 especies), Crustacea (101 especies) e Insecta (3182 especies). Proporcionamos la primera biblioteca de referencia de código de barras de ADN, con una cobertura global del 35 % de los taxones ibéricos. Al explorar estos datos, encontramos un fuerte sesgo taxonómico. Sobre la base de Leray-285, Odonata (43 de 79 especies presentan código de barras, 54.43 %) fueron los linajes mejor representados. Por el contrario, Diptera (393 de 1.693 especies presentan código de barras, 23.21 %) y Plecoptera (42 de 135 especies presentan código de barras, 31.11 %) estaban subrepresentados. Para los peces, los datos de códigos de barras de ADN disponibles cubrieron el 98.11 % de las especies autóctonas (76) y no autóctonas (30). Mediante la identificación y cuantificación de la proporción de especies sin códigos de barras de DNA (~ 65 %), pretendemos proporcionar unas guías para el diseño eficiente de los próximos pasos hacia el objetivo ambicioso pero necesario de compilar una biblioteca completa de referencia de código de barras de ADN para los macroinvertebrados y peces ibéricos.

**Palabras clave:** monitorización, bioindicador, conservación, DNA metabarcoding, eDNA, riqueza de especies, ríos, resolución taxonómica, calidad del agua

## INTRODUCTION

Given the unprecedented rate of biodiversity loss due to human-driven global change, understanding species distribution and community composition is essential for ecological management and conservation biology (Dirzo & Raven, 2003; Cardinale *et al.*, 2012). However, the implementation of a planetary-scale assessment of biodiversity patterns and habitat health is a daunting task. To date, biodiversity assessment methods are mostly based on a morphology-based taxonomic identification of collected specimens, which is a direct approach. However, the individual sorting and counting of a large sample of specimens and its morphological identification is very time consuming, which has slowed down the pace of on-going global-scale biomonitoring programs. With more sites and regions being monitored and the need to achieve finer taxonomic resolution of target taxa,

bio-inventorying initiatives require ever-growing funding and human resources. In this context, novel genomic tools have been developed to accelerate and reduce the costs of large-scale biodiversity surveys and assessments (Hajibabaei *et al.*, 2007; 2011).

The mitochondrial “DNA-barcode” marker Cytochrome Oxidase subunit I (COI, *cox1*) is sufficiently variable to cluster sequences into molecular entities that roughly approximate taxonomic species in many eukaryotic groups, allowing discrimination between morphologically similar species and life stages (Hebert *et al.*, 2003). Therefore, the quantitative analysis of the nucleotide variation in the *cox1* gene of entire communities may be used to speed up species identification in rapid biodiversity assessments (Tänzler *et al.*, 2012; Joly *et al.*, 2014; Múrria *et al.*, 2015). Additionally, this approach allows revealing potential cases of cryptic species and

identify groups with challenging taxonomy as well as life stages, including immatures and body parts (Hebert *et al.*, 2004; Hajibabaei *et al.*, 2007; Múrria *et al.*, 2010; Múrria *et al.*, 2014). However, an accurate assignment of a given *cox1* sequence to any known taxonomic species (i.e. species identification) requires a comprehensive genetic dataset of most of known taxa from across the focal region and all known sequence variants that are to be considered part of that species. Building up such a global scale barcode reference library is the main objective of the International Barcode for Life (<http://ibol.org>) (Ratnasingham & Hebert, 2007). Nevertheless, the implementation of a global biodiversity assessment has been hindered by the exorbitant economical costs of sequencing entire communities using the standard Sanger sequencing technology, which requires sequencing single genes from a single specimen in each run (Hajibabaei *et al.*, 2007). The advent of high-throughput sequencing (HTS) technologies, has exponentially reduced the cost by sequencing tens of millions of DNA fragments in parallel. Novel metabarcoding methods, which use gene-specific PCR primers to amplify DNA barcodes from a collection of organisms, have been developed to take advantage of the cost-efficient massive sequencing provided by HTS technologies, facilitating the identification of large collections of organisms in a single sequencing run (Baird & Hajibabaei, 2012). For instance, metabarcoding of the DNA isolated from a river water sample (environmental DNA, eDNA) can provide a complete characterization of the local community (Hajibabaei *et al.*, 2011). Ongoing improvements in specific metabarcoding protocols have been devised to develop a standardised protocol for global-scale biodiversity assessments (Ji *et al.*, 2013; Creer *et al.*, 2016). However, the advances in sequencing technology do not alleviate the need for a complete DNA barcode library, which is a well curated, comprehensive reference database of identified DNA barcodes. To guarantee automatic identification at the global scale, such a library should eventually include all taxa on Earth.

Macroinvertebrates and fishes are keystone organisms in freshwater ecosystems. Macroinvertebrates contribute to important functional

processes (e.g. litter decomposition, filtration and sediment retention) (Usseglio-Polatera *et al.*, 2000), whereas fishes are top predators that accumulate and transport energy and matter across long distances and regulate predators, herbivores and plants in rivers (Power, 1990). The introduction of non-indigenous invasive fish species for recreational uses, have resulted in the extirpation of many indigenous fish species worldwide and a biotic homogenization of freshwater fish, as observed in the Iberian Peninsula, among other regions (Cambray, 2003; Clavero & Garcia-Berthou, 2006). Changes in macroinvertebrate and fish community composition along environmental gradients have been shown to be associated to hydrology, water quality, habitat degradation and stream geomorphology (Bonada *et al.*, 2006). As a result, freshwater macroinvertebrate and fish biomonitoring has been included in standardised protocols to assess the health of aquatic ecosystems, detect anthropogenic impact and determine policy intervention (Karr, 1981; Prat & Munné, 2004; Bonada *et al.*, 2006; Ferreira *et al.*, 2007; Schmutz *et al.*, 2007).

In contrast to the most fish biomonitoring programs, which are based on taxonomic resolution at species level, biomonitoring programs for freshwater macroinvertebrates usually rely on a coarser taxonomic resolution, generally at the family level, mostly due to the time-consuming identification process or lack of taxonomic expertise (Hawkes, 1998; Alba-Tercedor *et al.*, 2004; Prat & Munné, 2004). However, the use of a coarser taxonomic resolution may compromise the detection of subtle ecosystem responses in community composition that can be observable at species or genus level, which is especially critical for detecting complex stressors other than organic pollution or non-indigenous species (Schmidt-Kloiber & Nijboer, 2004; Baird & Hajibabaei, 2012; Woodward *et al.*, 2013). Moreover, standard biomonitoring programs assume that life-history traits within families are phylogenetically conserved, but these traits could be species-specific and hence closely related species could differ on their responses to ecological conditions, habitat alteration or stressors (Usseglio-Polatera *et al.*, 2000; Carew *et al.*, 2007; Graf *et al.*, 2008; Múrria *et al.*, 2012). Poor taxonomic

knowledge may overlook key and abundant components of biodiversity, which is especially critical for frequently ignored groups such as dipterans, characterised by high abundance and richness in all freshwater habitats and particularly their species-specific responses to environmental quality (Carew *et al.*, 2007; Puntí *et al.*, 2009; Carew & Hoffmann, 2015).

In addition to the completeness of the reference database, other factors such as the length of the amplicon and the universality of the primer sets used are essential for a successful species detection and assignment. The standard metazoan DNA barcode implementation relies on a 658 bp fragment of the 5' half of the COI (Hebert *et al.*, 2003), which is commonly amplified using the "universal" LCOI1490/HCO2198 primers (Folmer *et al.*, 1994). However, the Folmer primer set is unable to amplify a significant portion of taxonomic groups, in these cases specific primers have been designed for amplifying the same region (e.g., Lobo *et al.*, 2013). In addition, the Illumina based HTS platforms, usually preferred in metabarcoding approaches, yield DNA sequence reads in the range of 50 to 400 bp in length, which are too short for amplifying the entire Folmer region. Similarly, due to post-mortem DNA degradation, shorter and poor quality DNA fragments (~ 150-300 bp) are more easily traceable either in the environment (eDNA samples), museum specimens or in gut contents (Valentini *et al.*, 2009). Taking all the above considerations, the most effective DNA target fragment length for metabarcoding studies is commonly around 250-350 bp. Additionally, it has been shown that the "universal" Folmer primers (Lobo *et al.*, 2013) actually fail to amplify a significant portion of freshwater macroinvertebrates due to primer bias (Elbrecht & Leese, 2015). Recent studies demonstrated the suitability of different universal primers sets for targeting short hypervariable regions of the DNA barcode with fragments ranging 178 to 205 bp long (e.g., Vamos *et al.*, 2017). Currently, 313 bp fragment at the 3' end of the DNA barcode is likely one of the most common sequenced regions (Leray *et al.*, 2013; Macías-Hernández *et al.*, 2018; Kemp *et al.*, 2019).

Both freshwater macroinvertebrates and

fishes are highly endemic in the Iberian Peninsula (Hernando & Soriguer, 1992; Ribera, 2000). Moreover, macroinvertebrates frequently feature vast molecular diversity and overlooked (cryptic) species, as previously recorded across Iberian streams (Ribera & Vogler, 2004; Abellan *et al.*, 2009; Múrria *et al.*, 2012, 2014). Overall a significant amount of ecological information is neglected in biomonitoring due to the lack of fine-scale taxonomic resolution for macroinvertebrates. For fishes, difficulties in comparison of catch per effort (e.g., complete or partial electrofishing, benthic or pelagic multimesh gillnets) and removal procedures for sampling fishes assemblages in rivers and large water bodies reduce biomonitoring power and sensitivity. Application of novel metabarcoding tools may overcome many of the former limitations, by providing massive, fine-scale identification at the community level, and hence complement and improve traditional bio-assessment protocols (Pilgrim *et al.*, 2011, Baird & Hajibabaei, 2012, Leese *et al.*, 2016).

In this study, we assess public availability of *cox-1* DNA sequences for freshwater macroinvertebrates (Mollusca and Arthropoda) and fishes of the Iberian Peninsula, considering indigenous and non-indigenous species, and generate an exhaustive list of Iberian DNA barcodes for future implementation in biomonitoring programs based on *cox-1*-metabarcoding approaches. We report the current taxonomic coverage of DNA barcodes, provide direct access to sequence data (including also GenBank reference numbers, NCBI taxonomy ID: taxid, which are compiled in fasta files in Supplementary Material), make this dataset available on the Limnetica website, and suggest further guidelines to complete an Iberian reference DNA barcode library of freshwater macroinvertebrates.

## METHODS

A comprehensive list of all Iberian freshwater Mollusca and Arthropoda was compiled using information from the multi-authored Fauna Iberica (<http://www.fauna-iberica.mncn.csic.es/faunaib/index.php>) and the Pan-European Species directories Infrastructure (PESI) ([www.eu-nomen.eu/portal/](http://www.eu-nomen.eu/portal/)). The two lists were compared, and cross

referenced with the Fauna Europaea checklist (<https://fauna-eu.org>) for detecting and removing duplicates (websites accessed on March 2018). For fishes, a list of species captured on the Iberian Peninsula was downloaded from fishbase (<https://www.fishbase.se/search.php>) on February 2019. The final species list included indigenous and non-indigenous species and was used as primary reference in targeted searches on GenBank ([www.ncbi.nlm.nih.gov/genbank/](http://www.ncbi.nlm.nih.gov/genbank/)) and BOLD (<http://www.boldsystems.org>) for matching *cox1* sequences. One *cox1* sequence per species was downloaded and therefore intraspecific genetic variability was overlooked. Molecular searches were run between June 2017 and March 2018 depending on the taxonomic group; for fishes the final search was done on February 2019. Available DNA barcodes were ranked by sequence length to select fragments similar to the standard 658 bp of Animal DNA barcode. For each barcode, the GenBank accession number and NCBI taxonomy ID (taxid), which is the reference number for each species in the Taxonomy Database that is a curated classification and nomenclature for all of the organisms in the public sequence databases (<https://www.ncbi.nlm.nih.gov/taxonomy>), were recorded when available. For those Iberian genera with no representative species in GenBank, a close-relative species within the same genus from Europe, or, alternatively, from North America, was downloaded and included in the final dataset to increase the accuracy of species assignment in metabarcoding analyses. However, only Iberian species were used to evaluate the extent of the DNA barcode coverage. A final round of taxonomic misidentification assessment was conducted by building phylogenetic trees separately for Mollusca, Crustacea, each order of Insecta and fishes to assess phylogenetic congruence across taxa (i.e., whether all sequences of species within a genus were monophyletic). Phylogenetic trees were inferred using maximum likelihood as implemented in the program RAxML 7.0.4 (Stamatakis, 2006) under the GTR+ $\Gamma$  substitution model.

We present results based on DNA fragments located within the complete DNA barcode region (658 bp), hereafter referred as the Folmer region (Folmer *et al.*, 1994, Hebert *et al.*, 2003). Given

HTS platforms provide DNA sequence in the range of 50-400 bp in length, we also report the coverage for sequences including the complete Leray fragment (313 bp) designed specifically for metabarcoding approaches (Leray *et al.*, 2013; Wangensteen *et al.*, 2018a, 2018b), which corresponds to the second half of the Folmer region, and hereafter referred as the Leray fragment; as well as the first 285 bp of the Leray fragment, hereafter referred as Leray-285 fragment. The last fragment has also been shown to provide resolution for metabarcoding data in specimens missing the complete Leray fragment, with little loss of taxonomic resolution (author's unpublished results).

## RESULTS

The taxonomic checklist for Iberian freshwater Mollusca and Arthropoda comprised a total of 3346 species. The DNA barcode dataset considering any DNA fragment located within the Folmer region included 1245 species, representing 37.21 % of the total recorded fauna. Similarly, the overall coverage for the Leray fragment was 32.67 % and the Leray-285 fragment was 35.09 % (Table 1). Although the majority of species were sequenced for the entire Folmer region, the DNA barcode fragments for some groups such as Odonata, Mollusca and Crustacea covered mainly the initial 5' segment and hence the DNA barcode coverage was lower for both the Leray and the Leray-285 fragments.

In the Supplementary Material, we report the taxonomic checklist for Mollusca (Table S1, available at <http://www.limnetica.net/en/limnetica>), Crustacea (Table S2, available at <http://www.limnetica.net/en/limnetica>) and each order of freshwater insects: Odonata (Table S3, available at <http://www.limnetica.net/en/limnetica>), Ephemeroptera (Table S4, available at <http://www.limnetica.net/en/limnetica>), Plecoptera (Table S5, available at <http://www.limnetica.net/en/limnetica>), Diptera (Table S6, available at <http://www.limnetica.net/en/limnetica>), Hemiptera (Table S7, available at <http://www.limnetica.net/en/limnetica>), Lepidoptera and Megaloptera (Table S8, available at <http://www.limnetica.net/en/limnetica>), Coleoptera (Table S9, available at

**Table 1.** Number of species and sequences available considering the three different DNA barcode lengths (Folmer; Leray; and Leray-285) per each family ordered by subphylum (Crustacea), phylum (Mollusca) and order (Insecta). *Número de especies y secuencias disponibles considerando tres longitudes diferentes del código de barras genético (Folmer; Leray y Leray-285) por cada familia ordenada por subfilo (Crustacea), filo (Mollusca) y orden (Insecta).*

GROUP (Phylum or Order)	Family	Species number	Folmer	Leray	Leray-285
Mollusca	Unionidae	7	7	6	7
	Margaritiferidae	2	2	2	2
	Corbiculidae	1	1	1	1
	Sphaeriidae	10	8	1	1
	Dreissenidae	1	1	1	1
	Neritidae	2	1	1	1
	Viviparidae	1	1	1	1
	Hydrobiidae	10	3	0	2
	Valvatidae	2	2	0	1
	Physidae	3	3	0	2
	Planorbidae	11	7	6	6
	Ancylidae	1	1	1	1
	Succineidae	1	1	0	1
	Acroloxidae	1	1	1	1
	Ferrissidae	1	0	0	0
	Lymnaeidae	6	6	5	6
	Melanopsidae	5	1	1	1
	TOTAL	65	46 (70.77%)	27 (41.54%)	35 (53.85%)

Cont.

Table 1. (cont.)

GROUP (Phylum or Order)	Family	Species number	Folmer	Leray	Leray_285
Crustacea	Cambaridae	2	2	2	2
	Astacidae	3	3	3	3
	Atyiidae	3	3	3	3
	Palaemonidae	1	0	0	0
	Potamidae	1	1	1	1
	Asellidae	31	24	12	18
	Gammaridae	40	21	14	19
	Niphargidae	7	1	1	1
	Corophiidae	11	2	1	2
	Triopsidae	2	2	1	2
	TOTAL	101	59 (58.42%)	38 (37.62%)	51 (50.5%)
Odonata	Calopterygidae	3	1	0	0
	Lestidae	7	4	2	4
	Platycnemididae	3	2	2	2
	Coenagrionidae	14	12	8	9
	Aeshnidae	10	9	2	5
	Gomphidae	8	6	4	4
	Cordulegaster	2	2	0	0
	Macromiidae	1	0	0	0
	Corduliidae	3	3	1	3
	Libellulidae	28	22	6	16
	TOTAL	79	61 (77.22%)	25 (31.65%)	43 (54.43%)

Cont.

Table 1. (cont.)

GROUP (Phylum or Order)	Family	Species number	Folmer	Leray	Leray_285
Ephemeroptera	Baetidae	45	27	23	25
	Caenidae	10	7	5	7
	Ephemerellidae	10	2	2	2
	Ephemeridae	4	3	3	3
	Heptageniidae	41	20	17	18
	Isonychiidae	1	1	1	1
	Leptophlebiidae	20	9	9	9
	Oligoneuriidae	4	1	1	1
	Polymitarcyidae	1	1	1	1
	Potamanthidae	1	1	1	1
	Prosopistomatidae	1	0	0	0
	Siphonuridae	8	3	2	2
	TOTAL	146	75 (51.37%)	65 (44.52%)	70 (47.95%)
Plecoptera	Perlodidae	16	5	3	3
	Perlidae	9	3	3	3
	Chloroperlidae	7	3	3	3
	Taeniopterygidae	13	6	6	6
	Nemouridae	34	13	13	13
	Capniidae	10	4	4	4
	Leuctridae	46	8	9	10
	TOTAL	135	42 (31.11%)	41 (30.37%)	42 (31.11%)

Cont.



Table 1. (cont.)

GROUP (Phylum or Order)	Family	Species number	Folmer	Leray	Leray_285
Diptera	Dolichopodidae	183	7	5	7
	Empididae	66	4	4	4
	Anthomyiidae	9	3	3	3
	Athericidae	2	1	1	1
	Ceratopogonidae	167	40	38	40
	Chaoboridae	2	0	0	0
	Chironomidae	549	211	181	194
	Culicidae	57	45	43	44
	Dixidae	7	2	2	2
	Simuliidae	36	20	20	20
	Thaumaleidae	9	2	2	2
	Ephydriidae	109	16	12	14
	Syrphidae	54	13	13	13
	Psychodidae	64	17	16	17
	Ptychopteridae	2	1	1	1
	Stratiomyidae	53	5	4	5
	Rhagionidae	18	3	2	2
	Tabanidae	104	5	2	5
	Cylindrotomidae	1	0	0	0
	Limoniidae	66	13	10	11
Tipulidae	135	9	8	8	
	TOTAL	1693	417 (24.63%)	367 (21.68%)	393 (23.21%)

Cont.

Table 1. (cont.)

GROUP (Phylum or Order)	Family	Species number	Folmer	Leray	Leray_285
Hemiptera	Hebridae	1	1	1	1
	Gerridae	12	8	8	8
	Hydrometridae	1	1	1	1
	Mesoveliidae	1	1	1	1
	Veliidae	9	3	3	3
	Ochteridae	1	1	0	1
	Notonectidae	8	5	5	5
	Pleidae	1	1	1	1
	Aphelocheiridae	3	1	1	1
	Naucoridae	2	1	1	1
	Nepidae	2	2	2	2
	Corixidae	40	20	18	19
	TOTAL	81	45 (55.56%)	42 (51.85%)	44 (54.32%)
Megaloptera	Sialidae	4	3	3	3
	TOTAL	4	3 (75%)	3 (75%)	3 (75%)

Cont.

Table 1. (cont.)

GROUP (Phylum or Order)	Family	Species number	Folmer	Leray	Leray_285
Coleoptera	Hydroscaphidae	2	1	1	1
	Microsporidae	1	0	0	0
	Gyrinidae	10	4	4	4
	Halplidae	19	11	10	11
	Noteridae	3	1	1	1
	Hygrobiidae	1	0	0	0
	Dytiscidae	179	84	82	83
	Hydrophilidae	140	63	62	63
	Hydraenidae	146	74	72	73
	Scirtidae	34	11	10	11
	Elmidae	32	17	17	17
	Dryopidae	17	6	4	6
	Limnichidae	6	1	1	1
	Heteroceridae	17	7	6	7
	Psephenidae	1	1	1	1
	Chrysomelidae	18	9	9	9
	Curculionidae	24	6	6	6
	TOTAL	650	296 (45.54%)	286 (44%)	294 (45.23%)
Lepidoptera	Crambidae	4	2	2	2
	TOTAL	4	2 (50%)	2 (50%)	2 (50%)

Cont.

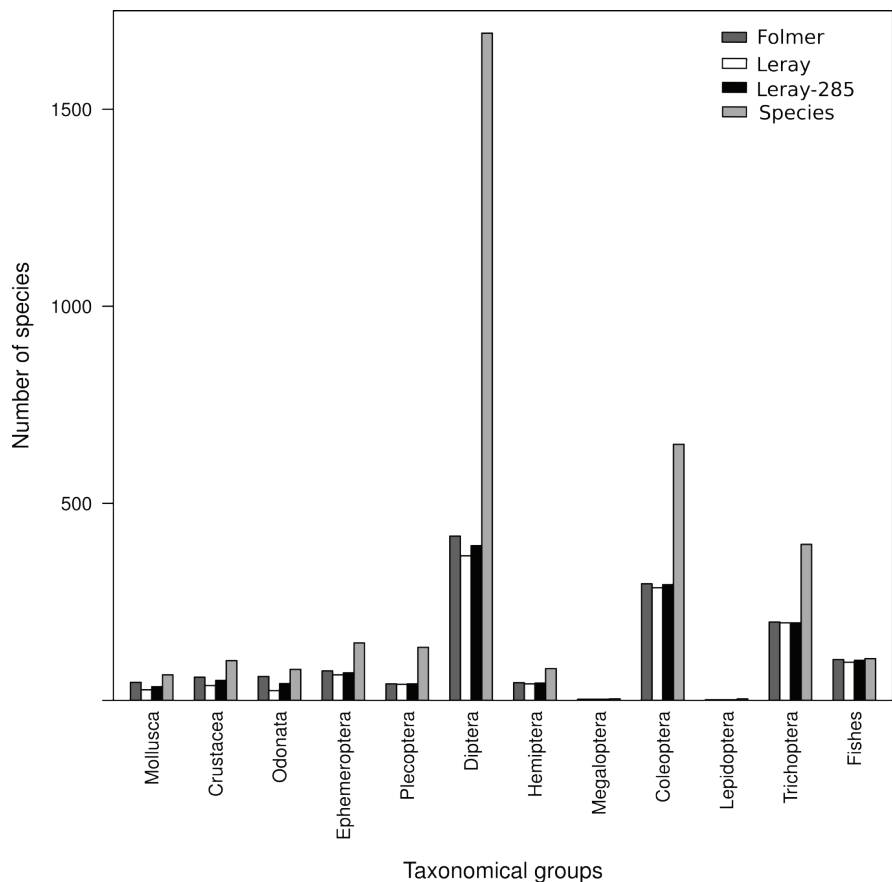
Table 1. (cont.)

GROUP (Phylum or Order)	Family	Species number	Folmer	Leray	Leray_285
Trichoptera	Rhyacophilidae	35	15	15	15
	Glossosomatidae	22	9	9	9
	Hydroptilidae	49	13	12	12
	Ptilocolepidae	3	2	2	2
	Philopotamidae	19	10	10	10
	Ecnomidae	2	2	2	2
	Hydropsychidae	32	15	15	15
	Polycentropodidae	23	9	9	9
	Psychomyiidae	21	9	9	9
	Brachycentridae	11	5	4	4
	Lepidostomatidae	3	3	3	3
	Phryganeidae	3	2	2	2
	Apataniidae	3	0	0	0
	Goeridae	10	9	9	9
	Limnephilidae	96	58	58	58
	Uenoidae	2	2	2	2
	Calamoceratidae	1	1	1	1
	Leptoceridae	40	22	22	22
	Odontoceridae	2	2	2	2
	Beraeidae	10	6	6	6
Helicopsychidae	1	0	0	0	
Sericostomatidae	8	6	6	6	
	TOTAL	396	199 (50.25%)	197 (49.74%)	197 (49.74%)
<b>TOTAL</b>		<b>3352</b>	<b>1245 (37.14%)</b>	<b>1093 (32.6%)</b>	<b>1174 (35.02%)</b>
Fishes	Iberian	76	74 (97.36%)	67 (88.15%)	72 (94.73%)
	Non-indigenous	30	30 (100%)	30 (100%)	30 (100%)

<http://www.limnetica.net/en/limnetica>) and Trichoptera (Table S10, available at <http://www.limnetica.net/en/limnetica>). Each table includes the taxonomic information, GenBank reference numbers, NCBI taxonomy ID (taxid) and geographical distribution. Moreover, all *cox1* sequences compiled were included in three different fasta files each one including the Folmer region, Leray fragment and Leray-285 fragment (Supplementary Material, files 1-3, available at <http://www.limnetica.net/en/limnetica>).

The recorded taxa and coverage of DNA barcodes varied across phylum, orders and families (Fig. 1). Table 1 summarizes the number of species and DNA barcodes available for the 130 Iberian freshwater Mollusca, Arthropoda and

fishes. Mollusca (46 of 65 species barcoded, 70.77 %) and Crustacea (59 of 101 species barcoded, 58.42 %) for the Folmer regions were highly represented. For insects, the most diverse order was Diptera (1693 species), which also had the lowest DNA barcode coverage (417 species barcoded, 24.63 %). Other highly diverse orders such as Coleoptera (296 of 650 species barcoded) or Trichoptera (200 of 390 species barcoded), had similar DNA barcode availability, 45.54 % and 51.28 % of the total number of taxa, respectively. Orders of moderate diversity such as Ephemeroptera (75 of 146 species barcoded) or Plecoptera (42 of 135 species barcoded), greatly differed in the DNA barcode coverage, 52.37 % and 31.11 %, respectively. Insect orders with few aquatic



**Figure 1.** Number of species (light grey) and sequences available considering the three different DNA barcode lengths: Folmer (dark grey), Leray (white) and Leray-285 (black) for Mollusca, Crustacea, orders of Insecta and fishes. *Número de especies (gris claro) y secuencias disponibles considerando las tres longitudes diferentes del código de barras genéticos: Folmer (gris oscuro), Leray (blanco) y Leray-285 (negro) para Mollusca, Crustacea, los órdenes de Insecta y peces.*

representative showed contrasting levels of DNA barcode coverage, 77.22 % of the 79 Odonata species were barcoded but only 55.56 % of the 81 Hemiptera species were represented in currently available databases. Only 9 of the 130 Iberian freshwater Mollusca and Arthropoda families were not represented by any DNA barcode: Palaemonidae (Crustacea); Ferrissiidae (Mollusca); Macromiidae (Odonata); Prosopistomatidae (Ephemeroptera); Chaoboridae and Cylindrotomidae (Diptera); Microsporidae and Hygrobiidae (Coleoptera) and Helicopsychidae (Trichoptera).

The taxonomic checklist for Iberian freshwater fishes covered a total of 98.11 % of the indigenous (76) and non-indigenous (30). Only the indigenous species *Chondrostoma olisiponensis* and *Squalius palaciosi* (Cypriniformes; Cyprinidae) lacked DNA barcodes. Since several species were involved in synonymies, the accepted name in Pan-European Species directories Infrastructure (PESI) was used in the fasta files, but all synonymies were added in the final table that included the NCBI taxonomy ID, GenBank reference numbers, geographical distribution and a classification as indigenous and non-indigenous in the Supplementary Material (Tables S11).

## DISCUSSION

For the highly diverse Iberian freshwater macroinvertebrate fauna, available genetic resources at species level are limited (~ 35 %), and hence we are still far from having a comprehensive DNA barcode reference library. However, although approximately 65 % of the Iberian freshwater macroinvertebrate species are lacking DNA barcodes, the reference library here provided covers most important family-level indicator groups that are usually included in ecological status assessments. The asymmetric distribution of DNA barcodes coverage across orders mostly reflects the existence of DNA barcoding initiatives focused on either specific taxa, e.g. the global initiative for the Trichoptera barcode of life contributed by Iberian entomologists (Zhou *et al.*, 2016), or particular regions, e.g. the German Barcode of Life (Raupach *et al.*, 2014, Morinière *et al.*, 2017). Moreover, the reference DNA barcode library here provided reveals differences

in species coverage depending on the DNA fragment considered, namely any portion within the Folmer region (37.22 %), the Leray (32.68 %) and the Leray-285 fragments (35.1 %). These differences are especially critical for detecting Odonata, Mollusca and Crustacea in metabarcoding approaches, due to the poor representation of the Leray fragment for species within these taxa. Some widely used algorithms for taxonomic assignment of metabarcoding data, e.g. ecotag (Boyer *et al.*, 2016), can only accurately assign DNA sequences that entirely overlap the target fragment, and therefore studies that use shorter or only partially overlapping fragments result in poor species assignment (Múrria *et al.*, in prep.). Future efforts to increase the coverage of the DNA barcode library should focus on amplifying the Folmer region or, even better, the whole mitochondrial genome (e.g. Iwasaki *et al.*, 2013; Salleh *et al.*, 2017), which not only would allow for species assignment using any mitochondrial primer set, but would facilitate the design of improved universal or group-specific metabarcoding primers.

For fishes, the recovered DNA barcode reference library is almost complete and only two indigenous species were missed. Even the most recently updated checklist of Iberian freshwater macroinvertebrates, as the one compiled here, is far from complete, due to the lack of modern taxonomic revisions for many taxa (e.g. Chironomidae, Prat *et al.*, 2016), the current rate of newly described species (e.g., Zamora-Muñoz *et al.*, 2017), or the estimated level of cryptic species in some groups such as *Baetis* (Ephemeroptera) (Williams *et al.*, 2006; Múrria *et al.*, 2014). In this regard, it is important to highlight that the more species represented by DNA barcodes in the reference database, the greater the chances of discovering cryptic species and resolving taxonomic conundrums. From a more applied standpoint, given that the current DNA barcode reference library covers all families, it should already provide taxonomic assignments for monitoring at least at the same taxonomic resolution as implemented in standard morphological biomonitoring programs. However, it should be highlighted that biomonitoring protocols based on metabarcoding approaches will greatly benefit from fine-scale resolution reference database, and hence

programs focused on obtaining DNA barcodes at the ecosystem scale are critical for completing a reference DNA barcode library for Iberian freshwater taxa (Zhou *et al.*, 2009, Webb *et al.*, 2012; Carew *et al.*, 2017). Specially critical is low DNA barcode coverage for Plecoptera (31.11 %), despite this order is one of the most intolerant to any presence of pollutants in the water bodies and among the most important taxa in biomonitoring programs, which suggests that more investments are needed in this group.

Improved taxonomic resolution provided by metabarcoding tools has advantages beyond biomonitoring programs (Bohmann *et al.*, 2014, Pedersen *et al.*, 2015, Deiner *et al.*, 2017). In conservation biology, for instance, detecting DNA traces in the environment (eDNA) can provide new records of rare or threatened species in inaccessible habitats, or for species with seasonal shifts of habitat uses or sensitive to commonly used sampling methods (Thomsen *et al.*, 2011). Non-indigenous species have been included in our dataset, and future eDNA studies conducted in Iberian rivers would also allow for detection of introduced species such as the American crayfish *Procambarus clarkii* (Girard, 1859) (Tréguier *et al.*, 2014) or the numerous introduced fishes (Keskin, 2014). Such data could be used to establish an early warning system for native biodiversity conservation (Rees *et al.*, 2014). Similarly, analysis of faecal samples or gut contents could inform on the relevance of freshwater macroinvertebrate communities in local trophic networks, including the diet of focal predators (Pompanon *et al.*, 2012). Other potential applications include ecotoxicological studies, which may quantify the effects of toxic chemicals on entire biological communities, rather than a limited small number of identified species, which is critical for understanding how sedimentary communities respond to a range of environmental stressors (Chariton *et al.*, 2014). Since ecological processes deposit DNA into the environment, a high-quality DNA barcode reference library should also contribute to advances in palaeoecology, i.e. the composition of ancient communities (Anderson-Carpenter *et al.*, 2011).

Ultimately, DNA barcoding methods are dependent on taxonomic expertise to create the

essential high-quality, complete reference libraries. Unfortunately, like in many other regions, many freshwater macroinvertebrate taxonomists in the Iberian Peninsula are either retired or are close to retirement, without having any guarantee that they will be replaced by a new generation of well-trained taxonomist. This will pose challenges for building and improving future DNA barcode reference libraries. Diptera provides a good example of this situation, as it represents over 50 % of all aquatic taxa, but it is often underrepresented in public repositories of species and genetic data.

The DNA barcode library here assembled should provide the guidelines to coordinate on-going efforts and spark new initiatives aimed at completing the Iberian freshwater DNA barcode reference library. Future DNA barcoding projects should guarantee, and rely on, solid taxonomic expertise for the correct specimen identification of reference barcodes and build on a well-preserved and curated voucher collection for species validation. They would also benefit from a prioritised taxonomic sampling strategy, for example by giving priority to underrepresented, yet relevant lineages, such as Diptera. We also propose that further projects should obtain DNA barcodes for multiple individuals and covering different biogeographical regions to assess intraspecific variability and metapopulation structure, which may reveal recent expansions and genetic diversity that may be relevant in conservation planning.

## ACKNOWLEDGEMENTS

This study was supported by the Fundació Aigües de Barcelona. Additional funding was provided by project 2014SGR1604 from the Catalan Government (MA). To Carmen Zamora Muñoz for valuable comments on the taxonomic checklist of Trichoptera.

## REFERENCES

ABELLAN, P., A. MILLAN A. & I. RIBERA. 2009. Parallel habitat-driven differences in the phylogeographical structure of two independent lineages of Mediterranean saline water beetles. *Molecular Ecology*, 18:

- 3885-3902. DOI: 10.1111/j.1365-294X.2009.04319.x
- ALBA-TERCEDOR, J., P. JÁIMEZ-CUÉLLAR, M. ÁLVAREZ *et al.* 2004. Caracterización del estado ecológico de ríos mediterráneos ibéricos mediante el índice IBMWP (antes BMWP'). *Limnetica*, 21 (3-4) (2002): 175-185.
- ANDERSON-CARPENTER, L. L., J. S. MCLA-CHLAN, S. T. JACKSON, M. KUCH, C. Y. LUMIBAO & H. N. POINAR. 2011. Ancient DNA from lake sediments: bridging the gap between paleoecology and genetics. *BMC evolutionary biology*, 11: 30. DOI: 10.1186/1471-2148-11-30
- BAIRD, D. J. & M. HAJIBABAEI. 2012. Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. *Molecular ecology*, 21: 2039-2044. DOI: 10.1111/j.1365-294X.2012.05519.x
- BOHMANN, K., A. EVANS, M. T. P. GILBERT *et al.* 2014. Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology & Evolution*, 29: 358-367. DOI: 10.1016/j.tree.2014.04.003
- BONADA, N., N. PRAT, V. H. RESH & B. STATZNER. 2006. Developments in aquatic insect biomonitoring: a comparative analysis of recent approaches. *Annual Review of Entomology*, 51: 495-523. DOI: 10.1146/annurev.ento.51.110104.151124
- BOYER, F., C. MERCIER, A. BONIN, Y. LE BRAS, P. TABERLET & E. COISSAC. 2016. Obitools: A unix-inspired software package for DNA metabarcoding. *Molecular Ecology Resources*, 16: 176-182. DOI: 10.1111/1755-0998.12428
- CAMBRAY, J. A. 2003. Impact on indigenous species biodiversity caused by the globalisation of alien recreational freshwater fisheries. *Hydrobiologia*, 500: 217-230. DOI: 10.1023/A:1024648719995
- CARDINALE, B. J., J. E. DUFFY, A. GONZALEZ *et al.* 2012. Biodiversity loss and its impact on humanity. *Nature*, 486: 59-67. DOI: 10.1038/nature11148
- CAREW, M. E., V. PETTIGROVE, R. L. COX & A. A. HOFFMANN. 2007. The response of Chironomidae to sediment pollution and other environmental characteristics in urban wetlands. *Freshwater Biology*, 52: 2444-2462. DOI: 10.1111/j.1365-2427.2007.01840.x
- CAREW, M. E. & A. A. HOFFMANN. 2015. Delineating closely related species with DNA barcodes for routine biological monitoring. *Freshwater Biology*, 60: 1545-1560. DOI: 10.1111/fwb.12587
- CAREW, M. E., S. J. NICHOLS, J. BATOVS-KA, R. ST CLAIR, N. P. MURPHY, M. J. BLACKET & M. E. SHACKLETON. 2017. A DNA barcode database of Australia's freshwater macroinvertebrate fauna. *Marine and Freshwater Research*, 68: 1788-1802. DOI: 10.1071/MF16304
- CHARITON, A. A., K. T. HO, D. PROESTOU *et al.* 2014. A molecular-based approach for examining responses of eukaryotes in microcosms to contaminant-spiked estuarine sediments. *Environmental Toxicology and Chemistry*, 33: 359-369. DOI: 10.1002/etc.2450
- CLAVERO, M. & E. GARCÍA-BERTHOU. 2006. Homogenization dynamics and introduction routes of invasive freshwater fish in the Iberian Peninsula. *Ecological Applications*, 16: 2313-2324. DOI: 10.1890/1051-0761(2006)016[2313:HDAIRO]2.0.CO;2
- CREER, S., K. DEINER, S. FREY *et al.* 2016. The ecologist's field guide to sequence-based identification of biodiversity. *Methods in Ecology and Evolution*, 7: 1008-1018. DOI: 10.1111/2041-210X.12574
- DEINER, K., H. M. BIK, E. MÄCHLE *et al.* 2017. Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology*, 26: 5872-5895. DOI: 10.1111/mec.14350
- DIRZO, R. & P. H. RAVEN. 2003. Global state of biodiversity and loss. *Annual Review of Environment and Resources*, 28: 137-167. DOI: 10.1146/annurev.energy.28.050302.105532
- ELBRECHT, V. & F. LEESE. 2015. Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass—sequence relationships with an innovative metabarcoding protocol. *PLoS ONE*, 10: e0130324. DOI: 10.1371/journal.pone.0130324
- FERREIRA, T., N. CAIOLA, F. CASALS, F., J.



- M. OLIVEIRA & A. DE SOSTOA. 2007. Assessing perturbation of river fish communities in the Iberian Ecoregion. *Fisheries Management and Ecology*, 14: 519-530. DOI: 10.1111/j.1365-2400.2007.00581.x
- FOLMER, O., M. BLACK, W. HOEH, R. LUTZ & R. VRIJENHOEK. 1994. DNA primers for amplification of mitochondrial cytochrome C oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, 3: 294-299.
- GRAF, W., J. MURPHY, J. DAHL, C. ZAMORA-MUNOZ & M. J. LÓPEZ-RODRÍGUEZ. 2008. *Distribution and ecological preferences of European freshwater organisms*. Volume 1. Trichoptera (Vol. 1). Pensoft Publishing.
- HAJIBABAEI, M., SINGER G. A., HEBERT P. D. N. & HICKEY D. A. 2007. DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Trends in Genetics*, 24: 167-172. DOI: 10.1016/j.tig.2007.02.001
- HAJIBABAEI, M., S. SHOKRALLA, X. ZHOU, G. A. C. SINGER & D. J. BAIRD. 2011. Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. *Plos ONE*, 6: e17497. DOI: 10.1371/journal.pone.0017497
- HAWKES, H.A. 1998. Origin and development of the biological monitoring working party score system. *Water Research*, 32: 964-968.
- HEBERT, P. D. N., A. CYWINSKA, S. L. BALL & J. R. DEWAARD. 2003. Biological identifications through DNA barcodes. *Proceedings of the Royal Society B*, 270: 313-321. DOI: 10.1098/rspb.2002.2218
- HEBERT, P. D. N., E. H. PENTON, J. M. BURNS, D. H. JANZEN & W. HALLWACHS. 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America*, 101: 14812-14817. DOI: 10.1073/pnas.0406166101
- HERNANDO, J. A. & M. C. SORIGUER. 1992. Biogeography of the freshwater fish of the Iberian Peninsula. *Limnetica*, 8: 243-253.
- IWASAKI, W., T. FUKUNAGA, R. ISAGOZAWA R. *et al.* 2013. MitoFish and MitoAnnotator: a mitochondrial genome database of fish with an accurate and automatic annotation pipeline. *Molecular Biology and Evolution*, 30, 2531-2540. DOI: 10.1093/molbev/mst141
- JI, Y., L. ASHTON, S. M. PEDLEY *et al.* 2013. Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology Letters*, 16: 1245-1257. DOI: 10.1111/ele.12162
- JOLY, S., T. J DAVIES, A. ARCHAMBAULT *et al.* 2014. Ecology in the age of DNA barcoding: the resource, the promise and the challenges ahead. *Molecular Ecology Resources*, 14: 221-232. DOI: 10.1111/1755-0998.12173
- KARR, J. R. 1981. Assessment of biotic integrity using fish communities. *Fisheries*, 6: 21-27.
- KEMP, J., A. LÓPEZ-BAUCELLS, R. ROCHA, O., S. WANGENSTEEN, Z. ANDRIATAFIKA, A. NAIR & M. CABEZA. 2019. Bats as potential suppressors of multiple agricultural pests: a case study from Madagascar. *Agriculture, Ecosystems & Environment*, 269: 88-96. DOI: 10.1016/j.agee.2018.09.027
- KESKIN, E. 2014. Detection of invasive freshwater fish species using environmental DNA survey. *Biochemical Systematics and Ecology*, 56: 68-74. DOI: 10.1016/j.bse.2014.05.003
- LERAY, M., J. Y. YANG, C. P. MEYER *et al.* 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: 34. DOI: 10.1186/1742-9994-10-34
- LEESE, F., F. ALTERMATT, A. BOUCHEZ *et al.* 2016. DNAqua-Net: Developing new genetic tools for bioassessment and monitoring of aquatic ecosystems in Europe. *Research Ideas Outcomes*, 2: e11321. DOI: 10.3897/rio.2.e11321
- LOBO, J., P. M. COSTA, M. A. TEIXEIRA, M. S. FERREIRA, M. H. COSTA, F. O. COSTA. 2013. Enhanced primers for amplification of DNA barcodes from a broad range of marine metazoans. *BMC Ecology*, 13: 34. DOI: 10.1186/1472-6785-13-34
- MACÍAS-HERNÁNDEZ, N., K. ATHEY, V. TONZO, O. S. WANGENSTEEN, M. A. ARNEDO & J. D. HARWOOD. 2018.

- Molecular gut content analysis of different spider body parts. *PLoS One*, 5: e0196589. DOI: 10.1371/journal.pone.0196589
- MORINIÈRE, J., L. HENDRICH, M. BALKE *et al.* 2017. A DNA barcode library for Germany's mayflies, stoneflies and caddisflies (Ephemeroptera, Plecoptera & Trichoptera). *Molecular Ecology Resources*, 17: 1293-1307. DOI: 10.1111/1755-0998.12683
- MÚRRIA, C., C. ZAMORA-MUÑOZ, N. BONADA, C. RIBERA & N. PRAT. 2010. Genetic and morphological approaches to the problematic presence of three *Hydropsyche* species of the *pellucidula* group (Trichoptera: Hydropsychidae) in the westernmost Mediterranean Basin. *Aquatic Insects*, 32: 85-98. DOI: 10.1080/01650424.2010.482939
- MÚRRIA, C., N. BONADA, M. A. ARNEÑO, C. ZAMORA-MUÑOZ, N. PRAT & A. P. VOGLER. 2012. Phylogenetic and ecological structure of Mediterranean caddisfly communities at various spatio-temporal scales. *Journal of Biogeography*, 39: 1621-1632. DOI: 10.1111/j.1365-2699.2012.02729.x
- MÚRRIA, C., M. MORANTE, M. RIERADEVALL, C. RIBERA & N. PRAT. 2014. Genetic diversity and species richness patterns in Baetidae (Ephemeroptera) in the Montseny Mountain range (North-East Iberian Peninsula). *Limnetica*, 33: 313-26.
- MÚRRIA, C., A. T. RUGENSKI, M. R. WHILES & A. P. VOGLER. 2015. Long-term isolation and endemism of Neotropical aquatic insects limit the community responses to recent amphibian decline. *Diversity and Distributions*, 21: 938-949. DOI: 10.1111/ddi.12343
- PEDERSEN, M. W., S. OVERBALLE-PETERSEN, L. ERMINI *et al.* 2015. Ancient and modern environmental DNA. *Philosophical Transactions of the Royal Society B*, 370: 20130383. DOI: 10.1098/rstb.2013.0383
- PILGRIM, E. M., S. A. JACKSON, S. SWENSON *et al.* 2011. Incorporation of DNA barcoding into a large-scale biomonitoring program: opportunities and pitfalls. *Journal of the North American Benthological Society*, 30: 217-231. DOI: 10.1899/10-012.1
- POMPANON, F., B. E. DEAGLE, W. O. SYMONDSON, D. S. BROWN, S. N. JARMAN & P. TABERLET. 2012. Who is eating what: diet assessment using next generation sequencing. *Molecular Ecology*, 21: 1931-1950. DOI: 10.1111/j.1365-294X.2011.05403.x
- POWER, M. E. 1990. Effects of fish in river food webs. *Science*, 250: 811-814.
- PRAT, N. & A. MUNNÉ. 2004. Biomonitoring de la calidad del agua en los ríos ibéricos: lecciones aprendidas. *Limnetica*, 33: 47-64.
- PRAT, N., T. PUNTÍ & M. RIERADEVALL. 2016. The use of larvae and pupal exuviae to study the biodiversity of Chironomidae in Mediterranean streams. *Journal of Entomological and Acarological Research*, 48: 29-36. DOI: 10.4081/jear.2016.5529
- PUNTÍ, T., M. RIERADEVALL & N. PRAT. 2009. Environmental factors, spatial variation, and specific requirements of Chironomidae in Mediterranean reference streams. *Journal of the North American Benthological Society*, 28: 247-265. DOI: 10.1899/07-172.1
- RATNASINGHAM, S. & P. D. N. HEBERT. 2007. BOLD: the barcode of life data system. *Molecular Ecology Notes*, 7: 355-364. DOI: 10.1111/j.1471-8286.2007.01678.x
- RAUPACH, M. J., L. HENDRICH, S. M. KÜCHLER, F. DEISTER, J. MORINIÈRE & M. M. GOSSNER. 2014. Building-up of a DNA barcode library for true bugs (Insecta: Hemiptera: Heteroptera) of Germany reveals taxonomic uncertainties and surprises. *PLoS One*, 9: e106940. DOI: 10.1371/journal.pone.0106940
- REES, H. C., B. C. MADDISON, D. J. MIDDLEDITCH, J. R. PATMORE & K. C. GOUGH. 2014. The detection of aquatic animal species using environmental DNA—a review of eDNA as a survey tool in ecology. *Journal of Applied Ecology*, 51: 1450-1459. DOI: 10.1111/1365-2664.12306
- RIBERA, I. 2000. Biogeography and conservation of Iberian water beetles. *Biological Conservation*, 92: 131-150. DOI: 10.1016/S0006-3207(99)00048-8
- RIBERA, I. & A. P. VOGLER. 2004. Speciation of Iberian diving beetles in Pleistocene refugia (Coleoptera, Dytiscidae). *Molecular Ecology*, 13: 179-193.

- SALLEH, F.M., J. RAMOS-MADRIGAL, F. PEÑALOZA et al. 2017. An expanded mammal mitogenome dataset from Southeast Asia. *GigaScience*, 6: 1-8. DOI: 10.1093/gigascience/gix053
- SCHMIDT-KLOIBER, A. & R. NIJBOER. 2004. The effect of taxonomic resolution on the assessment of ecological water quality classes. *Hydrobiologia*, 516: 269-283. DOI: 10.1023/B:HYDR.0000025270.10807.10
- SCHMUTZ, S., A. MELCHER, C. FRANGEZ et al. 2007. Spatially based methods to assess the ecological status of riverine fish assemblages in European ecoregions. *Fisheries Management and Ecology*, 14: 441-452. DOI: 10.1111/j.1365-2400.2007.00582.x
- STAMATAKIS, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, 22: 2688-2690. DOI: 10.1093/bioinformatics/btl446
- TÄNZLER, R., K. SAGATA, S. SURBAKTI, M. BALKE & A. RIEDEL. 2012. DNA barcoding for community ecology - how to tackle a hyperdiverse, mostly undescribed Melanesian fauna. *Plos ONE*, 7: e28832. DOI: 10.1371/journal.pone.0028832
- THOMSEN, P. F., J. KIELGAST, L. L. IVERSEN et al. 2011. Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular Ecology*, 21: 2565-2573. DOI: 10.1111/j.1365-294X.2011.05418.x
- TRÉGUIER, A., J. M. PAILLISSON, T. DEJEAN, A. VALENTINI, M. A. SCHLAEPFER & J. M. ROUSSEL. 2014. Environmental DNA surveillance for invertebrate species: advantages and technical limitations to detect invasive crayfish *Procambarus clarkii* in freshwater ponds. *Journal of Applied Ecology*, 51: 871-879. DOI: 10.1111/1365-2664.12262
- USSEGLIO-POLATERA, P., M. BOURNAUD, P. RICHOUX & H. TACHET. 2000. Biological and ecological traits of benthic freshwater macroinvertebrates: relationships and definition of groups with similar traits. *Freshwater Biology*, 43: 175-205. DOI: 10.1046/j.1365-2427.2000.00535.x
- VALENTINI, A., F. POMPANON & P. TABERLET. 2009. DNA barcoding for ecologists. *Trends in Ecology & Evolution*, 24: 110-117. DOI: 10.1016/j.tree.2008.09.011
- VAMOS, E., V. ELBRECHT & F. LEESE. 2017. Short COI markers for freshwater macroinvertebrate metabarcoding. *Metabarcoding and Metagenomics*, 1: e14625. DOI: 10.3897/mbmg.1.146254625
- WANGENSTEEN, O. S., E. CEBRIAN, C. PALACÍN & X. TURON. 2018a. Under the canopy: community-wide effects of invasive seaweeds in marine protected areas revealed by metabarcoding. *Marine Pollution Bulletin*, 127: 54-66. DOI: 10.1016/j.marpolbul.2017.11.033
- WANGENSTEEN, O. S., C. PALACÍN, M. GUARDIOLA & X. TURON. 2018b. DNA metabarcoding of littoral hard-bottom communities: high diversity and database gaps revealed by two molecular markers. *PeerJ*, 6: e4705. DOI: 10.7717/peerj.4705
- WEBB, J. M., L. M. JACOBUS, D. H. FUNK et al. 2012. A DNA barcode library for North American Ephemeroptera: progress and prospects. *PLoS One*, 7: e38063. DOI: 10.1371/journal.pone.0038063
- WILLIAMS, H. C., S. J. ORMEROD & M. W. BRUFORD. 2006. Molecular systematics and phylogeography of the cryptic species complex *Baetis rhodani* (Ephemeroptera, Baetidae). *Molecular Phylogenetics and Evolution*, 40: 370-382. DOI: 10.1016/j.ympev.2006.03.004
- WOODWARD, G., C. GRAY & D. J. BAIRD. 2013. Biomonitoring for the 21st century: New perspectives in an age of globalisation and emerging environmental threats. *Limnetica*, 32: 159-174. DOI: 10.23818/limn.32.14
- ZAMORA-MUÑOZ, C., C. MÚRRIA, N. BONADA & M. A. GONZÁLEZ. 2017. The *Hydropsyche instabilis* group (Trichoptera: Hydropsychidae) on the Iberian Peninsula: evolutionary relationships, new species, taxonomical controversies, and a key to larvae. *Arthropods Systematics & Phylogeny*, 75: 159-172.
- ZHOU, X., S. ADAMOWICZ, L. JACOBUS, R. DEWALT & P. D. N. HEBERT. 2009. Towards a comprehensive barcode library for

arctic life –Ephemeroptera, Plecoptera, and Trichoptera of Churchill, Manitoba, Canada. *Frontiers in Zoology*, 6: 30–38. DOI: 10.1186/1742-9994-6-30  
ZHOU, X., P. B. FRANSEN, R. W. HOLZEN-

THAL *et al.* 2016. The Trichoptera barcode initiative: a strategy for generating a species-level Tree of Life. *Philosophical Transactions of the Royal Society B*, 371: 20160025. DOI: 10.1098/rstb.2016.0025