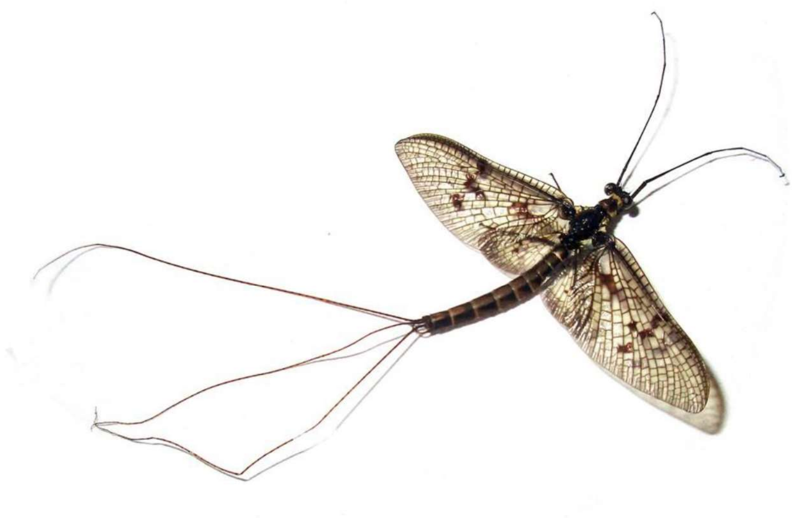
Master's degree in Ecology, Environmental Management and Restoration



WHERE ARE UNKNOWN AQUATIC INSECTS? LOOKING FOR UNKNOWN BIODIVERSITY HOTSPOTS ACROSS EUROPE

Carlota Sánchez Campaña

Thesis done under the direction of

Dr. Núria Bonada Caparrós

Department of Evolutionary Biology, Ecology and Environmental Sciences

Dr. Cesc Múrria i Farnós

Department of Evolutionary Biology, Ecology and Environmental Sciences

Barcelona – September 15th, 2021

Universitat de Barcelona

Departament de Biologia Evolutiva, Ecologia i Ciències Ambientals

FEHM lab RESEARCH GROUP

Freshwater Ecology, Hydrology & Management

WHERE ARE UNKNOWN AQUATIC INSECTS? LOOKING FOR UNKNOWN BIODIVERSITY HOTSPOTS ACROSS EUROPE

Author: Carlota Sánchez Campaña

Directors: Dr. Núria Bonada Caparrós & Dr. Cesc Múrria i Farnós

Department of Evolutionary Biology, Ecology and Environmental Sciences

University of Barcelona

Date: September 15th, 2021

Master's degree in Ecology, Environmental Management and Restoration

Carlota Sánchez Campaña

Dr. Núria Bonada Caparrós

Dr. Cesc Múrria i Farnós

Firmado digitalmente por CPISR-1 C NURIA BONADA CAPARROS Fecha: 2021.09.15 16:15:22 +02'00'

Cover picture: *Ephemera vulgata*. Source: B. Schoenmakers (https://waarneming.nl/photos/26369567/).

ABSTRACT

Aquatic insects are key for the correct functioning of freshwater ecosystems. Between 80 and 95% of insect species, however, remain to be formally described and most likely many of them will disappear before they are known. The objective of this thesis was to assess where are the unknown biodiversity hotspots of aquatic insects in Europe. To do so, a database with all the new aquatic insects' descriptions in the last twenty years (2000-2020) was compiled, and a set of sampling effort, environmental and socioeconomic variables was used to determine the factors behind the distribution of these recently described species. The results showed that the Mediterranean Basin was the region with the highest unknown biodiversity of aquatic insects, with Turkey being the country with more recently described species in the last 20 years. The three Orders with more described species were Trichoptera, Diptera and Coleoptera. This unknown biodiversity was mainly related to two environmental and socioeconomic factors; a high number of unknown species was found in mid-elevation areas and in regions with a high number of universities. Our findings suggest that future efforts to discover unknown biodiversity should be focused in regions with these characteristics located at south and eastern European regions. Surprisingly, the percentage of protected areas did not explain the unknown biodiversity, and more new species were found outside than inside protected areas. New protected areas should be established for the conservation of the still unknown species before they disappear by human impacts, especially in the Mediterranean Basin where freshwater biodiversity inventories are far from being complete and ecosystems are heavily impacted.

RESUM

Els insectes aquàtics són clau pel bon funcionament dels ecosistemes d'aigua dolça. Entre el 80 i 95% de les espècies d'insectes continuen sense estar formalment descrites i és molt probable que desapareguin abans que puguin ser-ho. L'objectiu d'aquest treball de fi de grau ha sigut determinar on es troben *hotspots* de biodiversitat d'insectes aquàtics a Europa. Per fer-ho, s'ha creat una base de dades amb totes les espècies d'insectes aquàtics descrites en els últims 20 anys (2000 – 2020) i s'han utilitzat un seguit de variables ambientals, socioeconòmiques i d'esforç de mostreig per establir els factors darrera la distribució d'aquestes espècies recentment descrites. Els resultats mostren que la conca Mediterrània era la amb més biodiversitat desconeguda d'insectes aquàtics, on Turquia és el país amb més espècies noves descrites en els últims 20 anys. Els Ordres amb un major nombre de descripcions han sigut, respectivament, Trichoptera, Diptera i Coleoptera. Aquesta biodiversitat desconeguda estava principalment relacionada amb dos factors ambientals i socioeconòmics; un gran nombre d'espècies desconegudes es va trobar en regions d'elevacions mitjanes i amb un alt nombre d'universitats. Els resultats suggereixen que futurs esforços per descobrir biodiversitat desconeguda s'haurien de focalitzar en zones amb aquestes característiques del sud i est d'Europa. Sorprenentment, el percentatge d'àrees protegides no explicava al biodiversitat desconeguda, i moltes espècies noves van ser trobades fora d'àrees protegides. S'han d'establir noves àrees protegides per la conservació de les especies encara desconegudes abans que aquesta desaparegui pels impactes humans, especialment a la conca Mediterrània on els inventaris de biodiversitat d'aigua dolça són lluny d'estar complerts i els ecosistemes estan greument afectats.

CONTENTS

INDEX OF FIGURES AND TABLES

1. INTRODUCTION

Even though our knowledge about biodiversity is growing, we are still far from establishing the exact number of species inhabiting the Earth. The vast majority of species are still not formally described (i.e. Linnean shortfall), and for many described species, there are several knowledge shortfalls related to their geographical distribution, biological characteristics, or ecological requirements (Bini et al., 2006; Hortal et al., 2015). Understanding biodiversity in a wide sense is key to maintain ecosystems and the services they provide to humans. However, anthropogenic impacts, such as habitat loss and degradation, are causing an unprecedented biodiversity loss (Payo-Payo & Lobo, 2016; Sánchez-Fernández et al., 2008), and many species will likely disappear before they are formally described.

Despite insects are key for the correct ecosystem functioning (Noriega et al., 2018; Schuldt & Assmann, 2010), 80-95% of the expected species are still unnamed according to Stork (2007. Several reasons including factors related to sampling effort, ecological or socioeconomic characteristics may explain the weak description of insect species. To find new species, taxonomists commonly sample regions that are already known for having a high biodiversity and, therefore, regions that are expected to be poor in diversity are mainly unexplored (sampling effort related factors) (Sastre & Lobo, 2009). Thus protected or low impacted, pristine areas are probably more explored than others areas (Sastre & Lobo, 2009). Society preferences also affect the priorities in research investments and, for this reason, funds are commonly centred to charismatic species such as birds or mammals, while insects (less charismatic) remain largely under-represented (Troudet et al., 2017). However, there are several exceptions; Lepidoptera, Orthoptera and Odonata are not under-represented when looking at protected species list, probably because of their size and vivid colouring (Leandro et al., 2017).

Even though the majority of insect taxonomists are based in North America and Europe (Gaston & May, 1992), 60% of the new descriptions of species are made by amateur taxonomists (Fontaine et al., 2012). The lack of funding, legal impediments, and the poor recognition of taxonomy by academy, society and policy-makers has brought us to the present situation, where taxonomy is in crisis and there is a very slow rate of publication of studies describing new species (Fontaine et al., 2012; Guerra-García et al., 2008). Over the past years, several initiatives have emerged to try to solve this problem. In Europe, *Fauna Europea* and the Pan-European Species directories Infrastructure (PESI) provide a checklist with all the multicellular terrestrial and freshwater animals recorded across Europe (Fontaine et al., 2012). Similarly, *Fauna Iberica* and the *Swedish Taxonomy Initiative* (STD) are two examples of initiatives at the country level that aim to build a national biodiversity inventory (Guerra-García et al., 2008). For the specific case of insects, the European Union has partnered with the Consortium of European Taxonomic Facilities (CETAF), the International Union for the Conservation of Nature (IUCN) and the Pensoft Publishers, *The Red List of Insect Taxonomists* (2021), which aims to create a database that will be the foundation of new policies to avoid the declining of insects.

The prime focus of this thesis was not the whole Insecta class, but the aquatics insects, which account for 60.4% of all freshwater animal species (Fenoglio et al., 2016). Freshwater ecosystems and their biodiversity are particularly affected by anthropogenic impacts, even more than terrestrial or marine ecosystem (e.g. see the Living Planet Index). The reason for this is that freshwater species have small geographical ranges, low dispersal abilities and high endemism levels, and the ecosystems where they are found are receivers and transmitters (i.e. effects are transported to the whole basin) and affected by multiple stressors, such as pollution, habitat destruction and biological invasions among others (Conti et al., 2014; Dudgeon et al., 2006). Future predictions also indicate that the rise in temperatures due to climate change will significantly impact the distribution and life cycles of freshwater species (Bonada et al., 2007; Conti et al., 2014) and local extinctions will be frequent (Múrria et al., 2020). In particular, aquatic insects account for a significant portion or freshwater biodiversity (i.e. 60.4% according to Balian, Segers et al. (2008)). They have a wide variety or biological and ecological traits, occupying many trophic niches and being present in almost all freshwater ecosystems (Fenoglio et al., 2014). Aquatic insects also contribute significantly to the process of organic matter and have effects on the nutrient cycling (Lundquist & Zhu, 2018). However, the knowledge we have on the ecology of aquatic insects at species level is still very poor because the challeging taxonomic identification for most groups (Tierno de Figueroa et al., 2013). Many species are still not known, there is a separated taxonomy between larvae (mostly aquatic) and adults (mostly terrestrial), and very few have recognised conservation concern (e.g. the IUCN Red List includes very few threaten aquatic insects).

1.1. OBJECTIVES AND HYPOTHESES

Our objectives were to (1) determine where are the hotspots of unknown biodiversity of aquatic insects in Europe (i.e. species that remain to be discovered), (2) know if protected areas protect unknown biodiversity and (3) understand the factors behind the distribution of this unknown biodiversity considering sampling effort, environmental and socioeconomic variables. This thesis also contributes in the completeness of a continental biodiversity inventory of aquatic insects. In this thesis is assumed that *unknown biodiversity* is the compendium of new species descriptions during the last 20 years (from 2000 to 2020). In other words, this approach supposes that this thesis started in 2000 and had the aim to know where new species would be described in the next 20 years in Europe.

Freshwater ecosystems around the Mediterranean Basin have been considered hotspots for biodiversity, also for aquatic insects. This is because the area has been subjected to several historical (e.g. glaciations) and ecological (e.g. seasonal and predictable drying periods) factors (Bonada & Resh, 2013; Tierno de Figueroa et al., 2013). In addition, this area has been traditionally less studied than central European countries, where taxonomy has a longer tradition. Other factors related to political or economic factors could also be important; for example, the Yugoslav Wars (1991 – 1999) (Radovic, 2004) or the mining policies in Turkey that did not prohibit mining activities in protected areas (Birben, 2019) could have slow down the rates of new species description. Therefore, **our first hypothesis** would be that hotspots for unknown biodiversity in Europe should be found in south and eastern areas.

Protected areas are one of the most useful tools to protect species and, as Guareschi et al. (2015) indicated, a big number of species' survival depends on their presence in protected areas. Actually, biodiversity hotspots has been used as a cost-effective strategy to establish conservation areas (Myers et al., 2000) although it should not be seen as the only way to protect and preserve biodiversity (Stork & Habel, 2014). In addition, protected areas are pristine areas with complex and heterogeneous landscapes that potentially host high levels of biodiversity. Therefore, **our second hypothesis** would be that protected areas should harbour more unknown biodiversity. Alternatively, unknown species would be found outside of protected areas, because these areas are usually much more explored.

Several factors related to sampling effort, ecological or socioeconomic characteristics can explain unknown biodiversity patterns and, therefore several hypotheses are related to our third objective. Firstly, we expect that areas with more sampling effort should have more species described in the last 20 years. Sampling effort is directly correlated with the socioeconomic level of each country; in those regions where the socioeconomic level is higher, such as the northern of Europe, the number of resources available to do sampling work is higher. Also, island regions such as England or Ireland, have their biodiversity inventories saturated due to their relatively small extent (which facilitates the required sampling effort). Secondly, environmental conditions also play a significant role in determining the patterns of distribution of the unknown biodiversity, since environmental conditions drive where the biodiversity is located and should indicate if there are still more taxa to describe. For instance, those habitats more holding higher aquatic insect biodiversity (Hershkovitz et al., 2015) should have high levels of unknown biodiversity unless they are completely explored. Finally, it is expected that those basins with a higher number of universities, research funding and development, education expenditures and researcher personnel should have a greater number of described species because more funds and researchers are present. For example, since 2008 to 2013, the economic crisis took a big toll on the budgets associated to science and environmental protection, which have made even bigger the differences in the efforts on diversity conservation between northern and southern countries (Landesmann, 2013). These consequences are still observable, the eastern and southern countries of the EU have less investment in research than its central and northern counterparts (Van Noorden & Butler, 2019).

2. MATERIALS AND METHODS

2.1. STUDY AREA

The study focuses in the European continent, including western Russia, Turkey and Cyprus (Fig. 1), and comprises an extension of 11.323.564 km^2, ranging across several bioclimatic regions from the Mediterranean to the Polar Artic. The Macaronesian islands, belonging politically to Europe, were not included given the drastic climatic differences with continental Europe. Countries in the middle east such as Georgia, Armenia and Azerbaijan were discarded because scarce data availability. For the whole area, HydroBASINS was used to select the spatial unit most appropriated for our study. HydroBASINS portrays the watershed boundaries and sub-basin delineations at a global scale (Lehner & Grill, 2013), and uses the Pfafstetter coding system (Verdin & Verdin, 1999), ranging from levels 1 to 12, to delineate the size of the spatial unit (12 being the higher resolution). The layers are based on a grid resolution of 15 arc-seconds (~ 500 m) and use the WGS84 datum. In this study, the level 4 was selected given the scale of the study. This corresponded in a total of 146 spatial units and an average polygon area of 80.234,22 km^2, larger or smaller spatial units than 4 would be unpractical to understand biodiversity patterns.

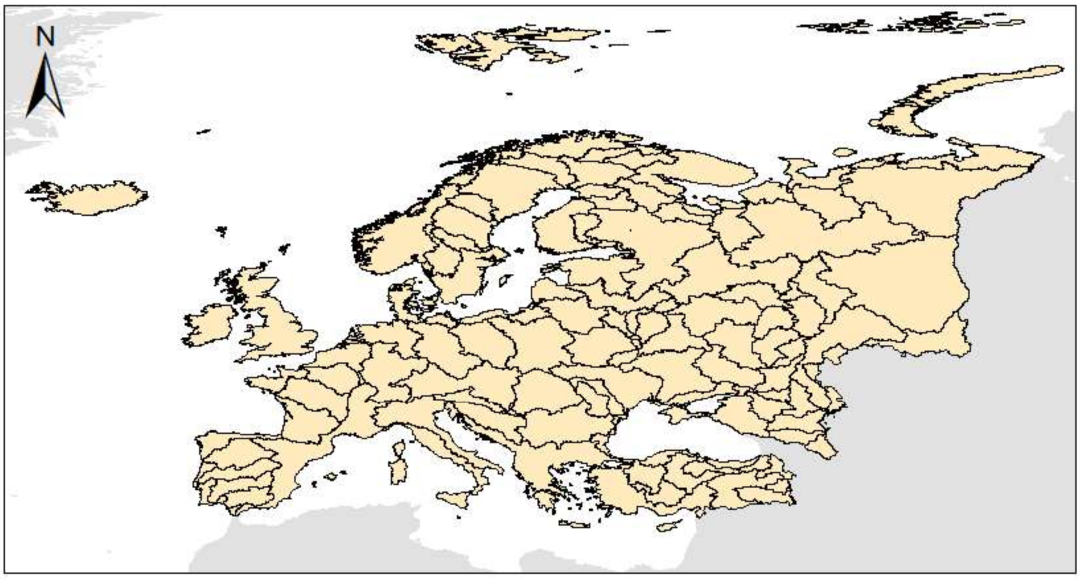


Fig. 1. Map showing the extent of this study and the spatial units considered by the level 4 of HydroBASIN (Lehner & Grill, 2013).

2.2. SPECIES DATA

A database of the number of new species of aquatic insects described in the study area in the last 20 years (2000 – 2020) was complied. Subspecies or species group were discarded. The initial selection of Orders and families was done based on the list of monophyletic freshwater lineages published by Múrria et al. (2018). Based on this list, a first research was conducted in general taxonomic and biodiversity web pages, including the *Taxa and Autecology Database for Freshwater Organisms* (https://www.freshwaterecology.info/), the *Index to Organisms Names* (http://www.organismnames.com/query.htm), PESI (http://www.eu-nomen.eu/portal/) and the *Barcode of Life Data System* (https://www.boldsystems.org/). A bibliographical research was done using search engines (e.g. *Google Scholar* and *Scopus*). In the case that I could not access to the original paper, the corresponding author of the article was contacted.

The next step was to search in specialized journals (e.g. *Graellsia* or *Braueria*) and Order-specific web portals, such as *Ephemeroptera of the world* (http://www.insecta.bio.spbu.ru/z/Eph-spp/index.htm), *Trichoptera World Checklist* (https://entweb.sites.clemson.edu/database/trichopt/index.php), *Systema Dipterorum* (http://diptera.org/Nomenclator), *Plecoptera species file* (http://Plecoptera.SpeciesFile.org) and the *Chironomid home page* (http://www.chironomidae.net). To ensure that all recently described species were included in the study, the database was sent to several taxonomic experts (listed in the acknowledgments) for comments and corrections. It was common that some of the holotype descriptions did not come with the geographic coordinates, so the database was sent again to the taxonomic experts to ask for the coordinates. In some cases, the holotype collector did not include the coordinates, and to solve it, an approximation of the coordinates was done using the holotype description and the web mapping platform *Google Maps*. This back and forth with the taxonomic experts continued for several months until we had the reassurance that the database was exhaustive and as much complete as possible.

2.3. SAMPLING EFFORT, ENVIRONMENTAL AND SOCIOECONOMIC DATA

To understand patterns of unknown biodiversity in the study area, we considered a series of variables related to sampling effort, environmental and socioeconomic factors.

A tentative preliminary list of these variables can be found in Table S1 in the supplementary materials. Using the original list, we removed some variables because they were represented, correlated or covered by other closely related variables (see details in the statistical analysis). The final selected variables can be found in Table 1. The sampling effort was provided by the Global Biodiversity Information Facility (GBIF) (https://www.gbif.org), the environmental information comes from HydroBASIN (https://hydrosheds.org/page/hydrobasins), while the socioeconomic data comes from HydroBASIN, the European Tertiary Education Register (ETER) project (https://www.eter-project.com) and Eurostat (https://ec.europa.eu/eurostat).

Table 1. Sampling effort (S), environmental (E) and socioeconomic (SE) variables used.

Variable	Meaning	Units	Rationale	Source
Species per spatial unit according to GBIF ^{S}	Number of recorded species' occurrences for each basin based on the data provided by GBIF	Number of species per country	Countries with more species could potentially have more species to describe	GBIF
Researchers ^{S}	Number of researchers per basin	Count	More naturalists going to the field to collect specimens, more chances to describe new species	Eurostat
Elevation (MDE) ^{E}	Height above sea level	m	More remote areas (such as those in higher elevation) have more chances to have undescribed species	HydroBASIN
Geographic coordinates of each basin ^{E}	Latitude and longitude of the centroid of each unit	Decimal degrees	Some latitudes or longitudes may harbour more species than others (e.g. south vs north)	HydroBASIN
Precipitation ^{E}	Annual precipitation average	mm	Precipitation level play a big role in species distributions. Higher precipitation could give more speciation opportunities	HydroBASIN
Temperature ^{E}	Annual air temperature average	ºC	Temperature ranges play a big role in species distributions. Some species could favour mild temperature ranges against more cold ones	HydroBASIN
Population density ^{E}	Population per unit area	People per km^2	Higher population densities, more impacts on freshwater ecosystems and less chances to discover new species (probably already extinct)	HydroBASIN
Percentage of protected areas ^{E}	Areas under protection figures	km^2	Protected areas could also protect undiscovered species	HydroBASIN
Number of universities per country $^{S, SE}$	University institutions in each country	Universities per country	More universities per country, more researchers working on taxonomy	ETER Project
Expenditure on education (EU) ^{SE}	Resources dedicated towards education amongst the EU members	Euros	More money dedicated on education could increase the number of experts dedicated to taxonomy	Eurostat
Research and development expenditure (EU) ^{SE}	Money intended for R&D projects	Euros	Countries that assign more money on R&D projects might invest also more on taxonomy	Eurostat

2.4. STATISTICAL ANALYSIS

The spatial analysis was done using the Environmental Systems Research Institute software (ESRI, 2017) and the statistical analysis was performed using the R programming language (R Core Team, 2020). All the graphics are presented using the *ggplot2* package (Wickham, 2016). A first analysis was carried out to establish if the discovered species in the last 20 years were more frequent inside or outside the protected areas, according to their taxonomic Order and using information from Protected Planet (UNEP-WCMC & IUCN, 2021).

To prevent multicollinearity problems (Alin, 2010), the correlations between the explanatory variables were checked using the *Hmisc* package (Harrell, 2020). Those pairs with an R-square value over 0.6 were removed to assure that the selected independent variables (Table 1) could explain the best our dependent variable, i.e. the number of total species discovered in the last 20 years per HydroBASIN unit. Considering that independent variables covered different range and magnitude, they were scaled.

Our dependent variable is a discrete variable that follows a Poisson distribution. Since we used parametric statistics, we transformed the dependent variable using a logarithmic transformation (log(x+1)). Because the data does not follow a Normal distribution, the relationship between the species richness and the sampling effort, environmental and socioeconomic variables were tested using Generalized Linear Models (GLM). Although our data follows a Poisson distribution, the values of the dependent variable are numeric (not integer) and a Gaussian distribution was assumed. A Likelihood Ratio Test (LRT) was performed to establish which of the independent variables could explain part of the data distribution (Johnson & Omland, 2004).

An additive complex model was built with all non-correlated variables, and also two-way models were built with all possible combinations. Since elevation and species richness tend to have a quadratic relationship, which means that species richness peaks at mid-elevations (Sanders & Rahbek, 2012), models where the elevation had a quadratic term were also tested. The Akaike Information Criterion (AIC) was used to determine which model was more accurate in predicting the distribution of the insect species described in the last 20 years.

In parallel, the three Orders with the biggest number of species described between 2000 – 2020 were analysed separately: Trichoptera (360 sp.), Diptera (223 sp.) and Coleoptera (105 sp.). The main reason of this analysis is because factors that explain the number of species described could be different across Orders. For instance, Diptera is a diverse Order able to tolerate a wide range of environmental conditions, whereas most Trichoptera species require waters that are clean, cool and well-oxygenated (Resh & Carde, 2009). Coleoptera individuals are found in the ecotone between land and inland waters, a habitat known for its rich biodiversity and sensitivity to environmental changes (Resh & Carde, 2009; Ribera, 2000). The models for these individual Orders were built following the same process as for the total species richness'.

3. RESULTS

3.1. SPECIES DATABASE

The final database included 789 species described in the last 20 years, belonging to the Orders Coleoptera (105 sp.), Diptera (223 sp.), Ephemeroptera (28 sp.), Lepidoptera (2 sp.), Neuroptera (3 sp.), Odonata (2 sp.), Plecoptera (66 sp.) and Trichoptera (360 sp.) (Fig. 2). No new species from the Order Hemiptera (Heteroptera) were found. Although the initial database included 799 species, several records were eliminated because the localities of the holotype was missed. The majority of the recently described species corresponded to Trichoptera, whereas Lepidoptera and Odonata were the Orders with the lowest number of described species (two in both cases) (Fig. 4).

Turkey was the country where more descriptions have been made (199 sp.), followed by Spain (90 sp.) and Italy (89 sp.). Meanwhile, in the northern part of Europe a small number of recently described species were recorded, mostly Diptera. Therefore, the highest number of recently described species was found in southern Europe, in particular around the Mediterranean Basin (Fig. 4 & 5).

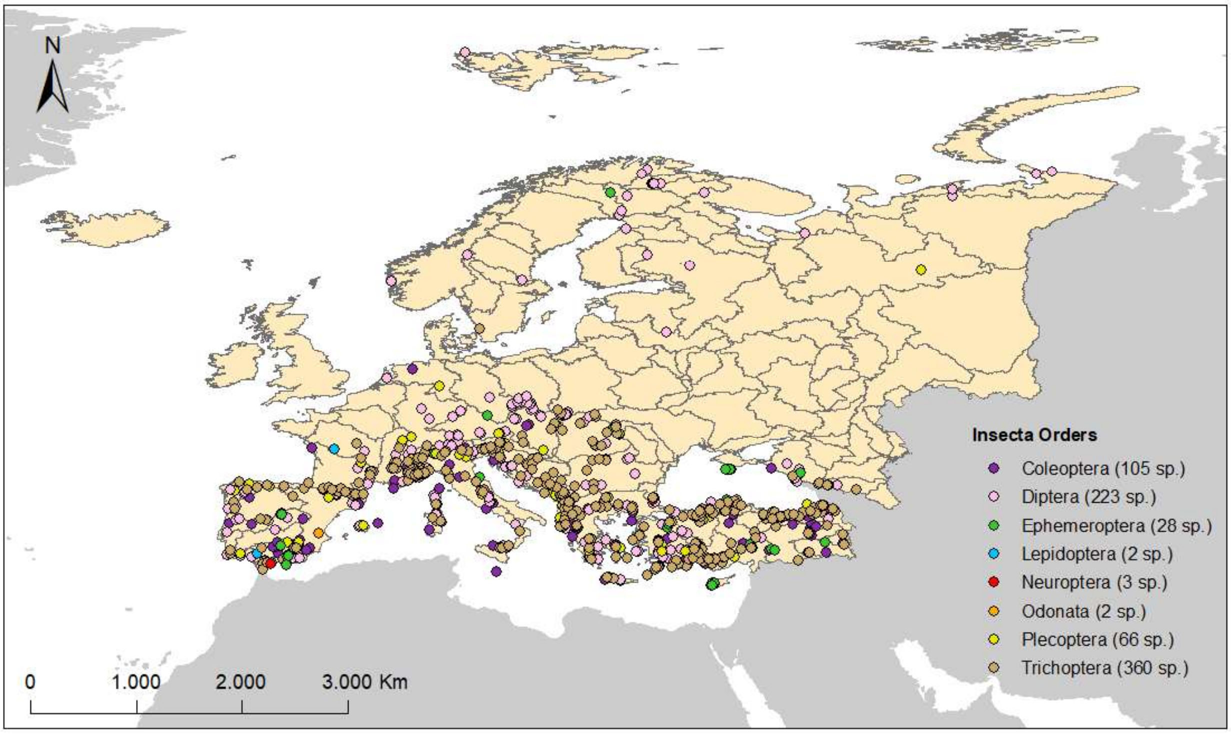


Fig. 2. Map of all the recently described species (2000-2020) of aquatic insects in Europe. The different colours correspond to different Orders.

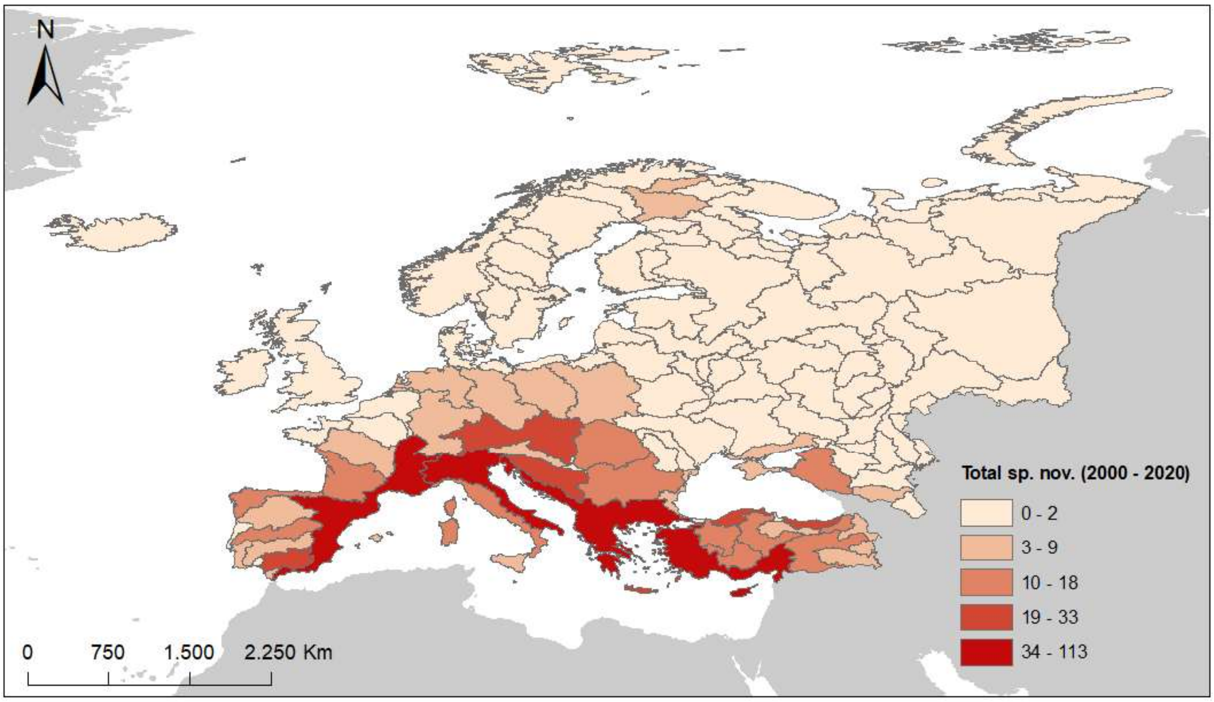


Fig. 3. Number of recently described species (2000-2020) of aquatic insects per sampling unit in Europe.

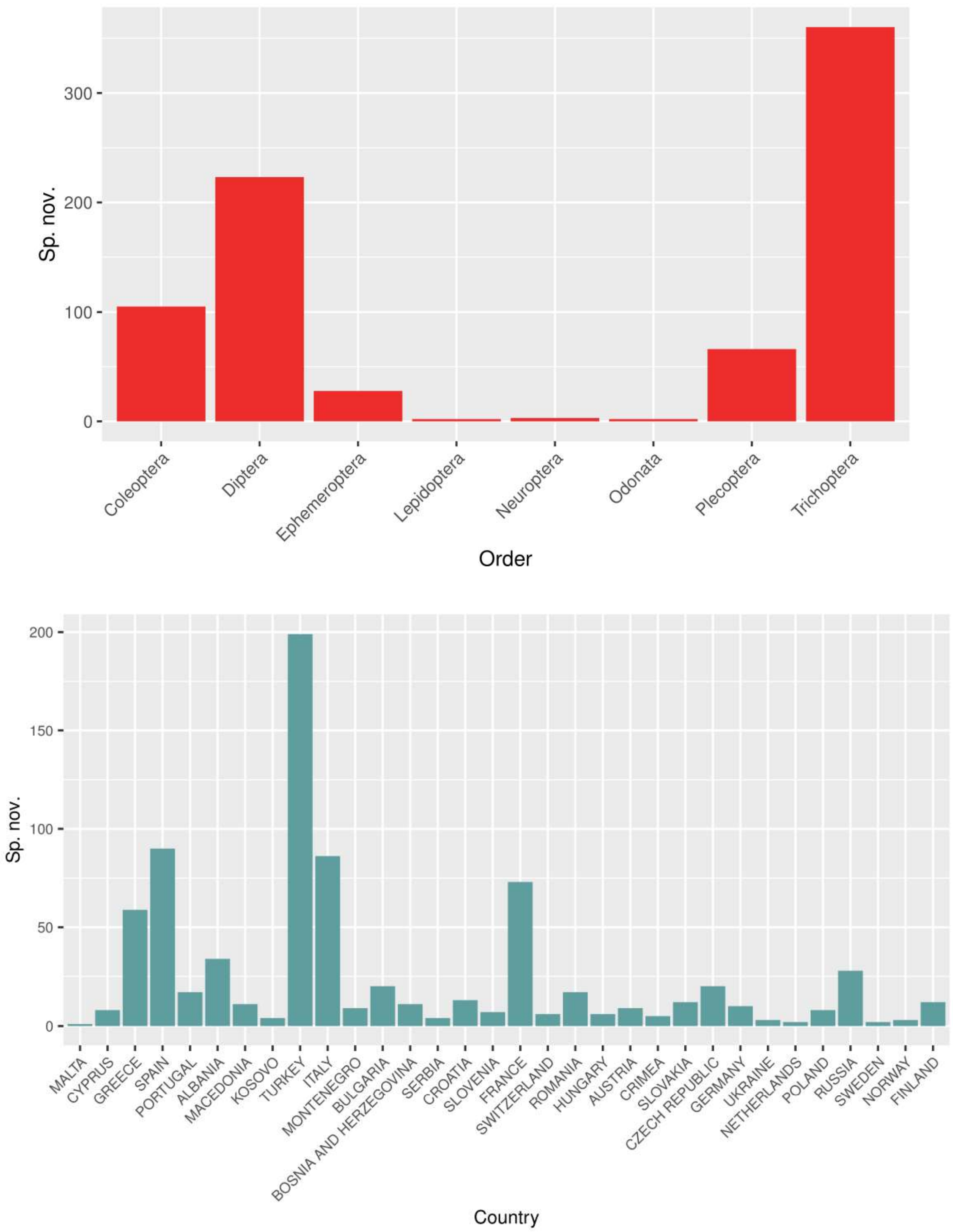


Fig. 4. Number of recently described species (Sp. nov.) in Europe per Order (2000-2020) (arranged alphabetically) (top) and per country (2000-2020) (arranged from lowest to highest latitude) (bottom).

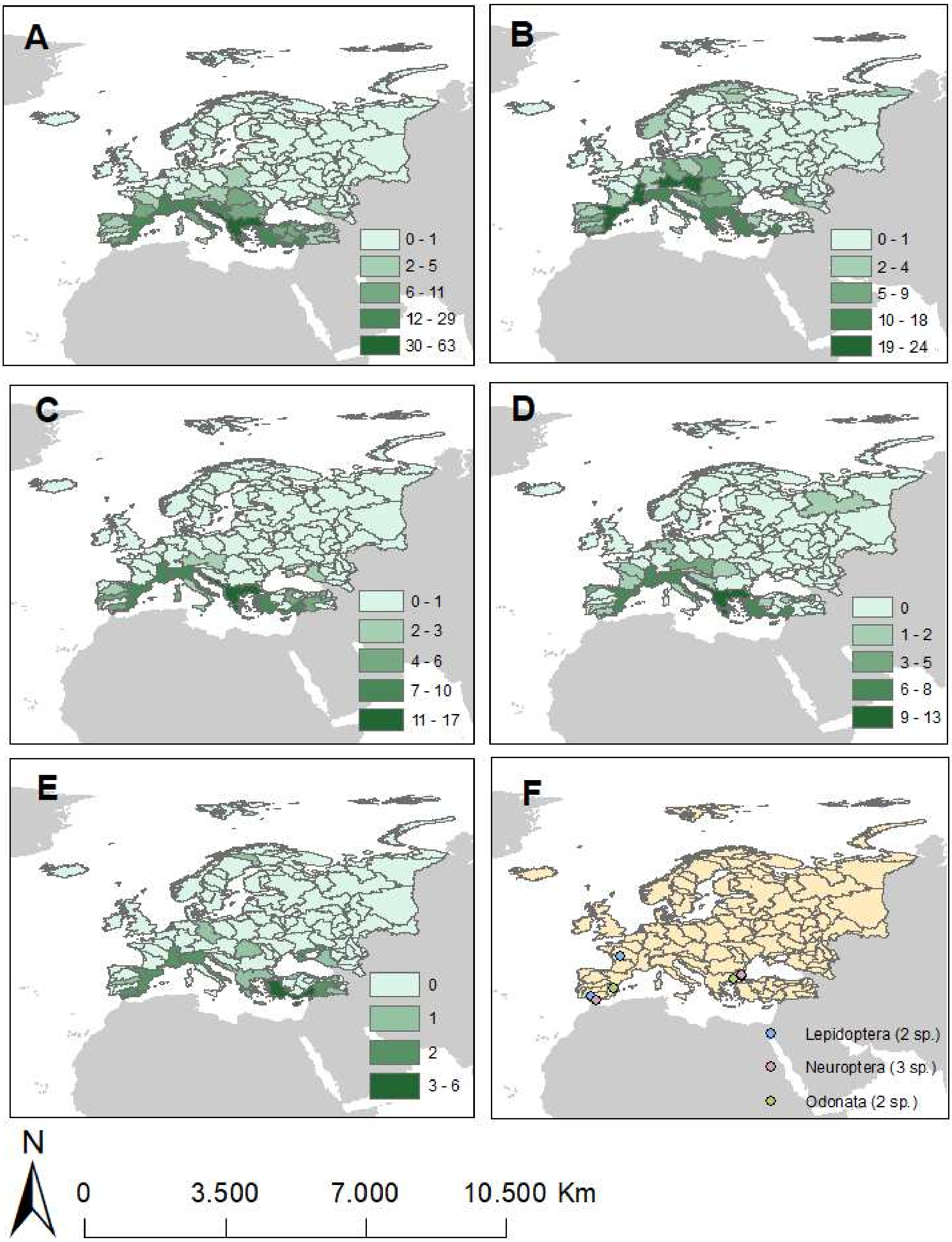


Fig. 5. Number (in range) of recently described species (2000-2020) per spatial unit in Europe according to Order as follows: (A) Trichoptera, (B) Diptera, (C) Coleoptera, (D) Plecoptera, (E) Ephemeroptera, (F) Lepidoptera, Neuroptera, and Odonata.

3.2. PROTECTED AREAS

In general, a higher number of recently described species were found outside than inside protected areas. Figure 6 showed that the majority of the recently described species were found outside the protected areas (580), especially for Trichoptera (mainly found in Turkey, where a great portion of its territory is not protected), with 278 species outside protected areas. This trend was similar in almost all orders and, therefore, the majority of recently described species was found in spatial units where the 25% or less of the surface was a protected area (Fig. 7).

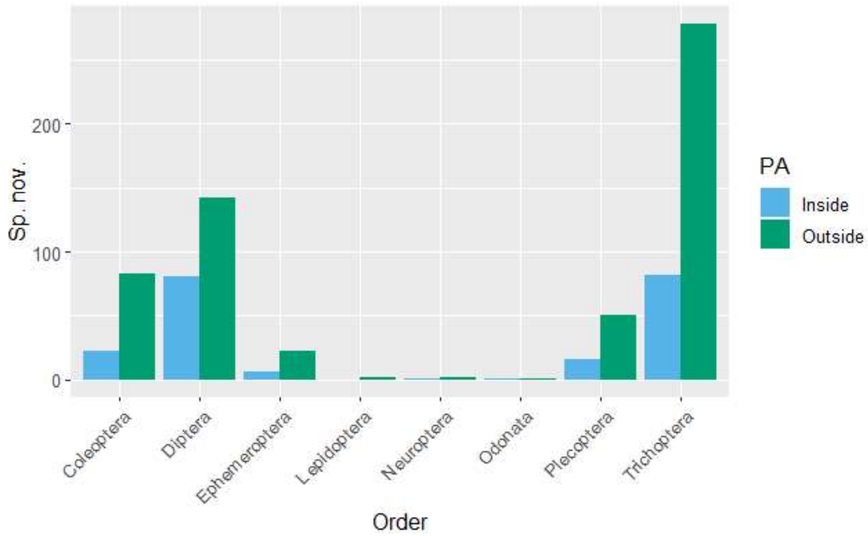


Fig. 6. Number of recently described species (Sp. nov.) inside and outside protected areas (PA). Protected areas include.

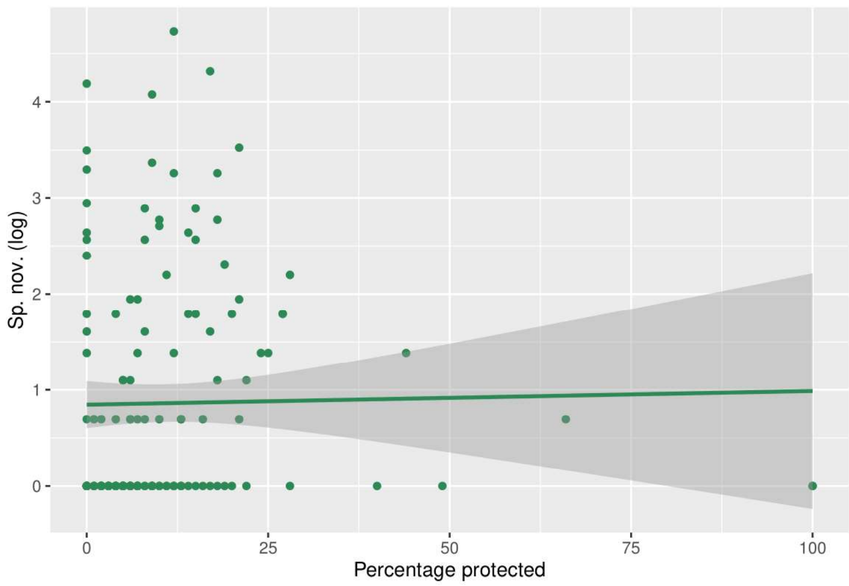


Fig. 7. Log-transformed number of recently described species (Sp. nov.) as a function of the percentage of protected area in each spatial unit.

3.3. DISTRIBUTION OF UNKNOWN SPECIES

ALL SPECIES RECENTLY DESCRIBED

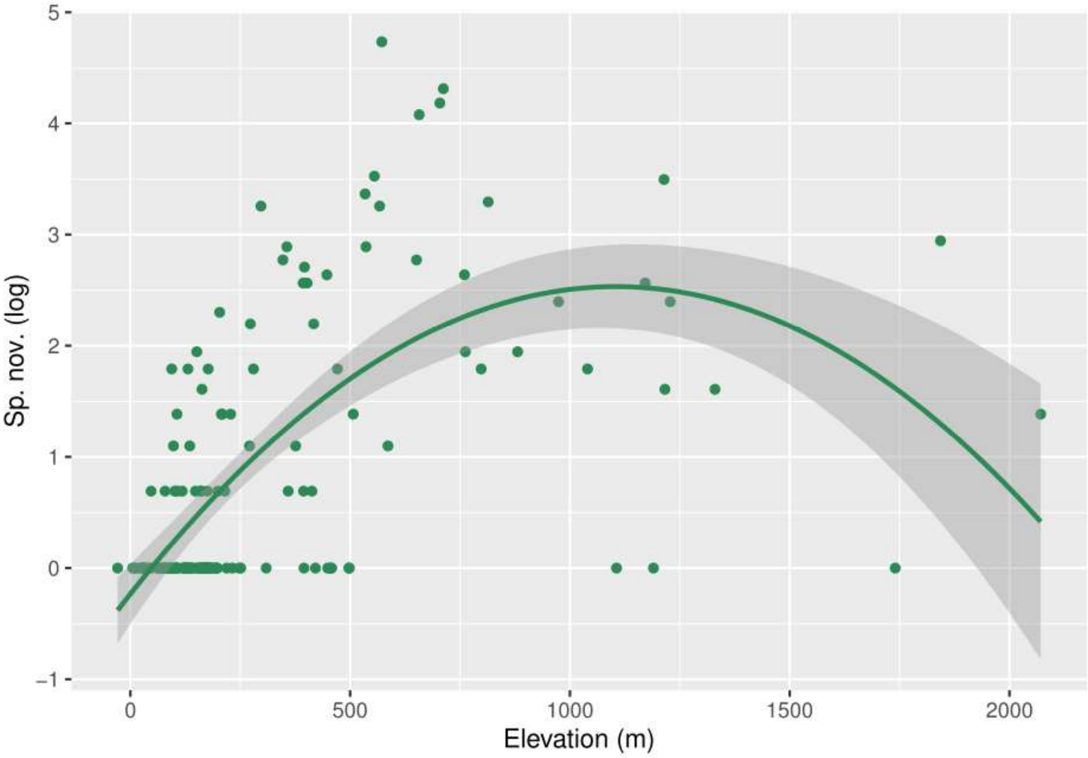
A Shapiro–Wilk normality test on the dependent variable informed that data was not compatible with a normal distribution (p-value = 1.543e-14), likely because most of the spatial units did not have species described in the last 20 years. As a result, the models were performed using a GLM Gaussian distribution. As the economic variables (R&D expenditure, education expenditure, and researches) had some NA values that can invalidate the GLM, these values we replaced by the mean of values for that variable.

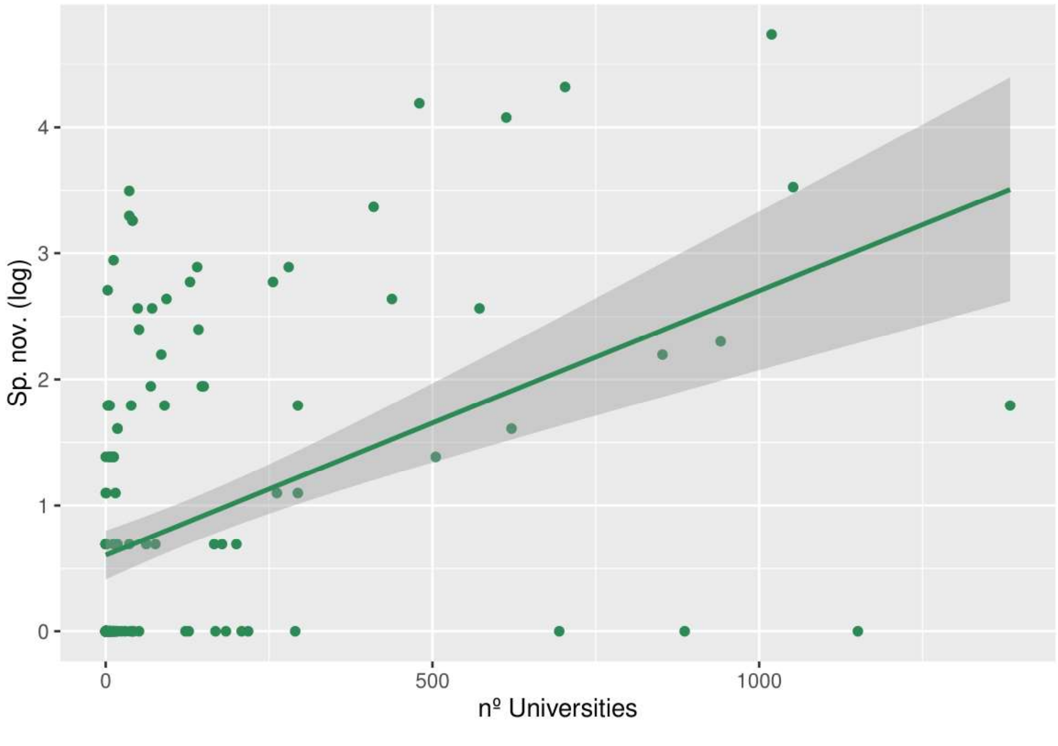
Elevation, number of universities, and GBIF data were the three predictor variables that had significant p-values values (Table 2), and therefore explained the distribution of recently described species. A big proportion of recently described species were found in mid-elevations, around 250 and 700 meters. The number of universities located in each basin and the species richness described in that basin followed a positive correlation, and therefore, spatial units with more universities had more species recently described. The relationship between the recently described species and the GBIF occurrence data followed a unimodal distribution. Hence, in basins where the number of GBIF occurrences is nor low nor extremely high – so when the sampling effort is mostly intermediate – the number of recently described species was the highest (Fig. 8).

Table 2. Results of the individual GLMs (Gaussian distribution) for each independent variable.

Predictor variables	AIC	Hypothesis test*	
		df	Pr (>Chi) ≤ 0.05
Longitude	472.27	1	0.4384
Latitude	471.75	1	0.2914
Elevation	431.53	1	**6.631e-12**
Precipitation	472.18	1	0.4069
Population density	470.02	1	0.09186
Percentage protected	472.83	1	0.8383
Researchers	472.73	1	0.708
Universities	440.53	1	**2.286e-09**
GBIF**	457.88	1	**7.912e-05**

**Analysis of Deviance Table, Likelihood-ratio test (LRT). **Transformed using log(1+x).*


Sp. nov. (log)
Elevation (m)
5
4
3
2
1
0
−1
0
500
1000
1500
2000


Sp. nov. (log)
nº Universities
4
3
2
1
0
0
500
1000

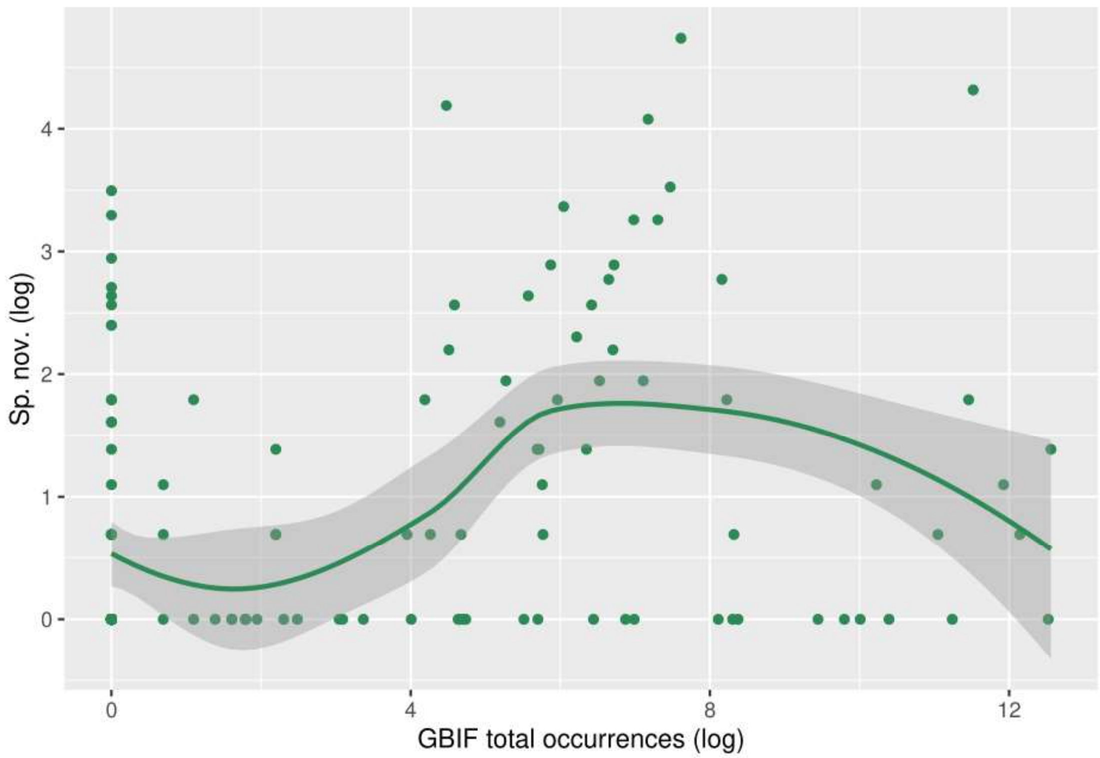


Fig. 8. Relationship between the log-transformed number of recently described species (Sp. nov.) per spatial unit and the three variables that best fitted the model.

Out of the 43 models tested, the one that had a lower AIC is the model with the additive effect of the elevation (quadratic factor) and the number of universities, (AIC = 369.29), indicating that the distribution of the number of species recently described in a basin is explained by the number of universities at mid-elevations. For this best model, adding the quadratic term to the elevation improved the fitting models and, therefore, reduced the AIC (Table S2). The number of occurrences of species according to GBIF's data (a sampling effort variable) was individually relevant (Table 2), but it did not contribute significatively in the global model.

TRICHOPTERA, DIPTERA AND COLEOPTERA RECENTLY DESCRIBED

Results for the richest Orders were similar to the ones obtained considering all species. For **Trichoptera**, four predictor variables explained the distribution of the recently described species: elevation, latitude, GBIF occurrence data and number of universities. Elevation was the most significant variable (p-value = 1.948e-14). For the GBIF data, the specific data occurrences for Trichoptera was used, whereas this data was unavailable for Diptera and Coleoptera, since it was not possible to discriminate the total amount of species by type of ecosystem in the GBIF database. Surprisingly, this variable was not significant (p-value = 0.09318), even though GBIF's total occurrences are significant (p-value = 0.01279). In general, the recently described species of Trichoptera were located in mid-elevation zones at latitudes between the 40 and 65 degrees. The distribution of recently described species according to the number of universities and the GBIF occurrences followed a similar pattern as the ones obtained for the entire dataset of recently described species.

For **Diptera** and **Coleoptera**, the significant variables were the same three as for the total of recently described species: elevation, GBIF occurrences, and the number of universities. For Diptera, the number of universities was the most significant variable (p-value = 6.084e-15), whereas for Coleoptera the elevation was the most significant variable (p-value = 2.891e-07). In these two cases, the distribution of the number of recently described species followed a similar pattern as previously described for the entire dataset: a higher description of species in areas that had a higher number of universities and streams located at mid-elevations. Overall, the most accurate model that explained the data distribution was the additive model with the elevation in quadratic factor and the number of universities.

4. DISCUSSION

Our results revealed that during the last 20 years, the areas with the highest number of described species corresponded to the southern areas in Europe, in particular to the Mediterranean Basin (Fig. 3). These results are in accordance to our first hypothesis and are not surprising because the Mediterranean Basin is a well-known biodiversity hotspot (Ivković & Plant, 2015; Myers, 1990) also for aquatic insects (Bonada et al., 2013; Tierno de Figueroa et al., 2013). One of the reason can be the cyclic Pleistocene glacial periods, that are associated to a high concentration of endemisms in the Iberian, Italic and Balkan peninsulas (Blondel et al., 2010; Ivković & Plant, 2015; Tierno de Figueroa et al., 2013) and the seasonal and predictable hydrological conditions (i.e. flow intermittence), which results in more niche spaces and higher temporal diversity (Tonkin et al., 2017).

The Orders with the highest number of recently described species are the ones that have more aquatic lineages in Europe (Balian, Lévêque, et al., 2008). In contrast, other groups such as Lepidoptera and Neuroptera, have a small number of aquatic lineages and, therefore, a low number of recently described. In the case of damselflies and dragonflies, the number of recently described species was low because these are well-studied groups, with high dispersal abilities and relatively low total species numbers, so it is rare to find new species (Fontaine et al., 2012).

One of the most surprising result of our study is the lack of significance of the protected areas in explaining the unknown aquatic insects biodiversity, in contrast to our second hypothesis. Around 73% of the described species during the past 20 years were found outside protected areas. Despite the percentage of protected areas and the number of new species should be related since these areas tend to have the highest sampling effort, our results showed that they do not protect unknown biodiversity. Unfortunately, freshwater ecosystems and aquatic insects are seldom considered when conceiving the conservation plans (Ivković & Plant, 2015), and current protected areas fail in protecting freshwater biodiversity (Guareschi et al., 2015; Hermoso et al., 2015). Protected areas are not designed considering aquatic insects and, therefore, it is not surprising that an important part of the recently discovered species are recorded in unprotected areas (Ivković & Plant, 2015; Payo-Payo & Lobo, 2016). Despite these recent efforts, we still need more initiatives to protect freshwater ecosystems. Sadly, the Iberian Peninsula is one example of the poor protection of the freshwater habitats and the diversity that they harbor (Sánchez-Fernández et al., 2008). The lack of specific legislation to protect invertebrates (including aquatic insects) is also critical for ensuring the conservation of freshwater biodiversity (Schuldt & Assmann, 2010). Turkey deserves a special mention, given the high rates of endemisms and because it hosts between 60.000 and 80.000 invertebrate species (Kucuk & Erturk, 2013). However, the lack of resources and poor conservation policies in this country resulted in a decreased of the number of protected areas from 10.1% to 8.9% between 2013 and 2018 (Birben, 2019). Finally, Russia is also particular case because the political and socioeconomic consequences caused by the collapse of the Soviet Union had a tremendous negative impact on the creation and safeguard of protected areas (Müller, 2013; Wendland et al., 2015). Therefore, our study suggests that future biodiversity conservation plans should focus on non-protected areas, which could still hold unknown and highly vulnerable species. The creation of entomologic (micro)reserves could be a promising approach to conserve unknown freshwater biodiversity. This concept appeared in 1992, proposed by Emílio Lagunas Lumbreras, and the objective is to *create small protected areas to try to preserve natural habitats and flora and fauna species of interest, due to their rarity or risk of extinction* (Quartau & Simões, 2014). For example, this figure was used in Portugal to create (micro)reserves to protect *Eurypha contentei* (Insecta, Hemiptera, Cicadoidea) and through the Spanish Entomological Association (AEE) five entomologic (micro)reserves have been recently created in Spain (Galante et al., 2015).

An alternative reason that could explain the failure of protected areas in protecting unknown biodiversity is that 199 species are from Turkey, a country with a low percentage of protected areas, skewing our data. Also, our results could also be distorted by the high number of zeros in our data; 81 out of the 146 spatial units did not have any recently described species. The resolution used could also blur the possible spatial effects, since we are homogenizing several basins within a unit. Future analysis exploring finer resolutions and using alternative zero-inflated statistical models would be required to confirm the obtained results in current.

In our third objective, we hypothesized that a higher sampling effort would result into a higher number of species descriptions, that new species would be found in areas with environmental conditions that favour high biodiversity, and in regions with more economic resources. Our results agreed with these hypothesis, with elevation (an environmental variable) and the number of universities per spatial unit (an sampling effort and socioeconomic variable) explaining unknown biodiversity patterns. The recently described species are found at mid-elevation zones, which is in agreement with our hypothesis, and may contain higher diversity values. There are several reasons that could explain this phenomenon, the first one being the protective effect of mountain areas. Through time, mountains have become biodiversity refugia due to their intricate topography, increasing isolation and speciation, moreover, changes in rainfall, soil type and vegetation could occur in short distances (Elsen et al., 2018; Perrigo et al., 2020). As Pereira et al. (2007) pointed out, elevations allow for more ecological niches to coexist which prompts for more speciation (i.e. high species diversity may drive high speciation rates). Importantly, at lower elevations there are fewer river sections in pristine conditions due to urban expansion and anthropic pressure that reduce the expected richness in these areas. On the contrary, higher elevations present scarce opportunities since the topographic complexity prevents the developing of large human activities, which preserves the quality of the ecosystems (Elsen et al., 2020). Another factor that could explain the presence of aquatic insects at mid-elevations,which very often correspond to mid-order sections,is related to the River Continuum Concept (RCC). The RCC postulates high alfa diversity in mid-order sections (but see Finn et al. (2011) for beta diversity) because the increase of the width, depth, flow characteristics, temperature, and the complexity of the water from headwater to mid-order sections (Vannote et al., 1980). At high elevations the environmental conditions are harsher and the stream smaller, which should reduce the abundance of taxa that inhabit high elevations streams. On another hand, a higher number of universities means that the probability of having more taxonomists is higher, i.e. there is a higher chance of finding new species. This emphasizes the importance of investing in taxonomy, a relatively inexpensive field. Our results suggested that more money dedicated towards research did not necessarily resulted into more species descriptions, but more number of universities did. The northern and central European territories have higher GBIF occurrences (i.e. an individually significant variable), meaning that these are regions well sampled and the probability of finding new species is low. Diptera is a really diverse Order that requires more taxonomic effort, this could have generated a delay in Diptera descriptions and could explain why in these 20 years new Dipterans have been described in central and northern Europe (Fig. 2 & 6).

The species database generated in this thesis is a useful resource of information to complete freshwater biodiversity inventories in Europe and to know where are located the unknown biodiversity of aquatic insects in Europe, aiding in the decision process of conservation and management efforts. Despite we did not apply complex statistical models to know where new species of aquatic insects will be found in the future, our findings suggest that taxonomic efforts to find new species should be directed towards south and eastern European areas, with a high number of universities and at mid-elevations. The creation of new protected areas should be especially focused in the Mediterranean Basin, where freshwater biodiversity inventories are still incomplete and ecosystems suffer from heavy human impacts.

5. CONCLUSIONS

In regards to our hypotheses we can say that (1) the Mediterranean basin is, indeed, a hotspot of unknown aquatic insect biodiversity, (2) protected areas do not contain more biodiversity, (3) those basins with a higher number of GBIF records, and consequently, that have more sampling efforts over extended time, are not the ones with a higher number of described species in the period considered, and (4) Mediterranean basins with a higher number of universities have more species discovered, even though it does not necessary mean that the given basin is richer since the taxonomist and the described species do not have to be from the same country.

The results highlight the importance of the protector effect of the mountainous areas, especially at mid-elevations, do for preserving biodiversity, whether it is known or not. Moreover, our findings also emphasize the importance of re-thinking our protected areas and the criteria behind them, as it is highlighted in this thesis by the fact that more than half of the described species between 2000 and 2020 are located outside protected areas. Our results indicate that more efforts are needed for protecting the mid-elevation streams, which still harbor a high number of unknown species.

6. ACKNOWLEDGEMENTS

First and foremost, I would like to thank my directors, Núria Bonada and Cesc Múrria, for all their patience, assistance and aid through this process. Also, I would like to acknowledge all the help and support I received from Dr. Virgilio Hermoso López, Dr. David Sánchez Fernández and Dr. J. Manuel Tierno de Figueroa. Last but not least, this thesis could not have been possible without all the data, comments and corrections I got from Adolfo Cordero-Rivera (Odonata), Andrés Millán (Coleoptera and Hemiptera), Craig Macadam (miscellaneous), Dávid Murányi (Plecoptera), Francesc Uribe Porta (miscellaneous), Füsun Sipahiler (Tricoptera), Gennaro Coppa (Trichoptera), Henri Tachet (Trichoptera), Horst Aspöck (Megaloptera and Neuroptera), Jani Heino (Diptera, Chironomidae), János Oláh (Trichoptera), Joel Breil-Moubayed (Diptera, Chironomidae), KD Dijkstra (Odonata), Marcos González (Trichoptera), Marija Ivković (Diptera), Martin Spies (Diptera, Chironomidae), Petr Pařil (Diptera), Roman Godunko (Ephemeroptera), Romain Sarremejane (miscellaneous), Stamatis Zogaris (miscellaneous), Tomáš Derka (Ephemeroptera), Vedula Polaskova (Diptera), Victor Montserrat (Megaloptera and Neuroptera), Wolfram Graf (Trichoptera) and Wolfram Mey (Lepidoptera).

7. REFERENCES

Alin, A. (2010). Multicollinearity. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*(3), 370–374. https://doi.org/10.1002/wics.84

Balian, E. V., Lévêque, C., Segers, H., & Martens, K. (Eds.). (2008). *Freshwater Animal Diversity Assessment*. Springer Netherlands. https://doi.org/10.1007/978-1-4020-8259-7

Balian, E. V., Segers, H., Lévèque, C., & Martens, K. (2008). The Freshwater Animal Diversity Assessment: an overview of the results. *Hydrobiologia*, *595*(1), 627–637. https://doi.org/10.1007/s10750-007-9246-3

Bini, L. M., Diniz-Filho, J. A. F., Rangel, T. F. L. V. B., Bastos, R. P., & Pinto, M. P. (2006). Challenging Wallacean and Linnean shortfalls: knowledge gradients and conservation planning in a biodiversity hotspot. *Diversity and Distributions*, *12*(5), 475–482. https://doi.org/10.1111/j.1366-9516.2006.00286.x

Birben, Ü. (2019). THE EFFECTIVENESS OF PROTECTED AREAS IN BIODIVERSITY CONSERVATION: THE CASE OF TURKEY. *CERNE*, *25*(4), 424–438. https://doi.org/10.1590/01047760201925042644

Blondel, J., Aronson, J., Bodiou, J.-Y., & Boeuf, G. (2010). *The Mediterranean Region: Biological Diversity in Space and Time* (2nd ed.). Oxford University Press.

Bonada, N., & Resh, V. H. (2013). Mediterranean-climate streams and rivers: geographically separated but ecologically comparable freshwater systems. *Hydrobiologia*, *719*(1), 1–29. https://doi.org/10.1007/s10750-013-1634-2

Conti, L., Schmidt-Kloiber, A., Grenouillet, G., & Graf, W. (2014). A trait-based approach to assess the vulnerability of European aquatic insects to climate change. *Hydrobiologia*, *721*(1), 297–315. https://doi.org/10.1007/s10750-013-1690-7

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. *Biological Reviews*, *81*(02), 163. https://doi.org/10.1017/S1464793105006950

Elsen, P. R., Monahan, W. B., & Merenlender, A. M. (2018). Global patterns of protection of elevational gradients in mountain ranges. *Proceedings of the National Academy of Sciences*, *115*(23), 6004–6009. https://doi.org/10.1073/pnas.1720141115

Elsen, P. R., Monahan, W. B., & Merenlender, A. M. (2020). Topography and human pressure in mountain ranges alter expected species responses to climate change. *Nature Communications*, *11*(1), 1974. https://doi.org/10.1038/s41467-020-15881-x

Fenoglio, S., Bonada, N., Guareschi, S., López-Rodríguez, M. J., Millán, A., & Tierno de Figueroa, J. M. (2016). Freshwater ecosystems and aquatic insects: a paradox in biological invasions. *Biology Letters*, *12*(4), 20151075. https://doi.org/10.1098/rsbl.2015.1075

Fenoglio, S., Merritt, R. W., & Cummins, K. W. (2014). Why do no specialized necrophagous species exist among aquatic insects? *Freshwater Science*, *33*(3), 711–715. https://doi.org/10.1086/677038

Finn, D. S., Bonada, N., Múrria, C., & Hughes, J. M. (2011). Small but mighty: headwaters are vital to stream network biodiversity at two levels of organization. *Journal of the North American Benthological Society*, *30*(4), 963–980. https://doi.org/10.1899/11-012.1

Fontaine, B., van Achterberg, K., Alonso-Zarazaga, M. A., Araujo, R., Asche, M., Aspöck, H., Aspöck, U., Audisio, P., Aukema, B., Bailly, N., Balsamo, M., Bank, R. A., Belfiore, C., Bogdanowicz, W., Boxshall, G., Burckhardt, D., Chylarecki, P., Deharveng, L., Dubois, A., … Bouchet, P. (2012). New Species in the Old World: Europe as a Frontier in Biodiversity Exploration, a Test Bed for 21st Century

Taxonomy. *PLoS ONE*, *7*(5), e36881. https://doi.org/10.1371/journal.pone.0036881

Galante, E., Numa Valdez, C., & Verdú, J. R. (2015). La conservación de los insectos en España, una cuestión no resuelta. *Revista IDE@ - SEA*, *7*, 1–13.

Gaston, K. J., & May, R. M. (1992). Taxonomy of taxonomists. *Nature*, *356*(6367), 281–282. https://doi.org/10.1038/356281a0

Guareschi, S., Bilton, D. T., Velasco, J., Millán, A., & Abellán, P. (2015). How well do protected area networks support taxonomic and functional diversity in non-target taxa? The case of Iberian freshwaters. *Biological Conservation*, *187*, 134–144. https://doi.org/10.1016/j.biocon.2015.04.018

Guerra-García, J., Espinosa, F., & García, J. (2008). Trends in taxonomy today. *Zoologica Baetica*, *19*, 15–49.

Hermoso, V., Filipe, A. F., Segurado, P., & Beja, P. (2015). Effectiveness of a large reserve network in protecting freshwater biodiversity: a test for the Iberian Peninsula. *Freshwater Biology*, *60*(4), 698–710. https://doi.org/10.1111/fwb.12519

Hershkovitz, Y., Dahm, V., Lorenz, A. W., & Hering, D. (2015). A multi-trait approach for the identification and protection of European freshwater species that are potentially vulnerable to the impacts of climate change. *Ecological Indicators*, *50*, 150–160. https://doi.org/https://doi.org/10.1016/j.ecolind.2014.10.023

Hortal, J., de Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., & Ladle, R. J. (2015). Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, *46*(1), 523–549. https://doi.org/10.1146/annurev-ecolsys-112414-054400

Ivković, M., & Plant, A. (2015). Aquatic insects in the Dinarides: identifying hotspots of endemism and species richness shaped by geological and hydrological history using Empididae (Diptera). *Insect Conservation and Diversity*, *8*(4), 302–312. https://doi.org/10.1111/icad.12113

Johnson, J. B., & Omland, K. S. (2004). Model selection in ecology and evolution. *Trends in Ecology & Evolution*, *19*(2), 101–108. https://doi.org/10.1016/j.tree.2003.10.013

Kucuk, M., & Erturk, E. (2013). Biodiversity and Protected Areas in Turkey. *Sains Malaysiana*, *42*, 1455–1460.

Landesmann, M. (2013). “The New North-South Divide in Europe: Can the European Convergence Model be Resuscitated?” *Vienna Institute Monthly Report*, *2013*, 3–

13.

Leandro, C., Jay-Robert, P., & Vergnes, A. (2017). Bias and perspectives in insect conservation: A European scale analysis. *Biological Conservation*, *215*, 213–224. https://doi.org/10.1016/j.biocon.2017.07.033

Lehner, B., & Grill, G. (2013). Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. *Hydrological Processes*, *27*(15), 2171–2186. https://doi.org/10.1002/hyp.9740

Lundquist, M. J., & Zhu, W. (2018). Aquatic insect functional diversity and nutrient content in urban streams in a medium- sized city. *Ecosphere*, *9*(5). https://doi.org/10.1002/ecs2.2284

Müller, M. (2013). From Sacred Cow to Cash Cow: The Shifting Political Ecologies of Protected Areas in Russia. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2365248

Múrria, C., Dolédec, S., Papadopoulou, A., Vogler, A. P., & Bonada, N. (2018). Ecological constraints from incumbent clades drive trait evolution across the tree-of-life of freshwater macroinvertebrates. *Ecography*, *41*(7), 1049–1063. https://doi.org/10.1111/ecog.02886

Múrria, C., Sáinz-Bariáin, M., Vogler, A. P., Viza, A., González, M., & Zamora-Muñoz, C. (2020). Vulnerability to climate change for two endemic high-elevation, low-dispersive Annitella species (Trichoptera) in Sierra Nevada, the southernmost high mountain in Europe. *Insect Conservation and Diversity*, *13*(3), 283–295. https://doi.org/10.1111/icad.12387

Myers, N. (1990). The biodiversity challenge: Expanded hot-spots analysis. *Environmentalist*, *10*(4), 243–256. https://doi.org/10.1007/BF02239720

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, *403*(6772), 853–858. https://doi.org/10.1038/35002501

Noriega, J. A., Hortal, J., Azcárate, F. M., Berg, M. P., Bonada, N., Briones, M. J. I., Del Toro, I., Goulson, D., Ibanez, S., Landis, D. A., Moretti, M., Potts, S. G., Slade, E. M., Stout, J. C., Ulyshen, M. D., Wackers, F. L., Woodcock, B. A., & Santos, A. M. C. (2018). Research trends in ecosystem services provided by insects. *Basic and Applied Ecology*, *26*, 8–23. https://doi.org/10.1016/j.baae.2017.09.006

Payo-Payo, A., & Lobo, J. M. (2016). The unpredictable characteristics of the localities where new Iberian species will be discovered. *Diversity and Distributions*, *22*(1), 73–82. https://doi.org/10.1111/ddi.12386

Pereira, H. M., Proença, V. M., & Vicente, L. (2007). Does species diversity really drive speciation? *Ecography*, *30*(3), 328–330. https://doi.org/10.1111/j.0906-7590.2007.04779.x

Perrigo, A., Hoorn, C., & Antonelli, A. (2020). Why mountains matter for biodiversity. *Journal of Biogeography*, *47*(2), 315–325. https://doi.org/10.1111/jbi.13731

Quartau, J., & Simões, P. (2014). *Microrreservas em entomologia: uma abordagem à conservação de euryphara contentei (Insecta, Hemiptera, Cicadoidea) em Portugal* (pp. 99–119).

Resh, V. H., & Carde, R. T. (Eds.). (2009). *Encyclopedia of Insects* (2nd Editio). Elsevier. https://doi.org/10.1016/B978-0-12-374144-8.X0001-X

Ribera, I. (2000). Biogeography and conservation of Iberian water beetles. *Biological Conservation*, *92*(2), 131–150. https://doi.org/10.1016/S0006-3207(99)00048-8

Sánchez-Fernández, D., Bilton, D. T., Abellán, P., Ribera, I., Velasco, J., & Millán, A. (2008). Are the endemic water beetles of the Iberian Peninsula and the Balearic Islands effectively protected? *Biological Conservation*, *141*(6), 1612–1627. https://doi.org/10.1016/j.biocon.2008.04.005

Sanders, N. J., & Rahbek, C. (2012). The patterns and causes of elevational diversity gradients. *Ecography*, *35*(1), 1–3. https://doi.org/10.1111/j.1600-0587.2011.07338.x

Sastre, P., & Lobo, J. M. (2009). Taxonomist survey biases and the unveiling of biodiversity patterns. *Biological Conservation*, *142*(2), 462–467. https://doi.org/10.1016/j.biocon.2008.11.002

Schuldt, A., & Assmann, T. (2010). Invertebrate diversity and national responsibility for species conservation across Europe – A multi-taxon approach. *Biological Conservation*, *143*(11), 2747–2756. https://doi.org/10.1016/j.biocon.2010.07.022

Stork, N. E. (2007). World of insects. *Nature*, *448*(7154), 657–658. https://doi.org/10.1038/448657a

Stork, N. E., & Habel, J. C. (2014). Can biodiversity hotspots protect more than tropical forest plants and vertebrates? *Journal of Biogeography*, *41*(3), 421–428.

https://doi.org/10.1111/jbi.12223

Tierno de Figueroa, J. M., López-Rodríguez, M. J., Fenoglio, S., Sánchez-Castillo, P., & Fochetti, R. (2013). Freshwater biodiversity in the rivers of the Mediterranean Basin. *Hydrobiologia*, *719*(1), 137–186. https://doi.org/10.1007/s10750-012-1281-z

Tonkin, J. D., Bogan, M. T., Bonada, N., Rios-Touma, B., & Lytle, D. A. (2017). Seasonality and predictability shape temporal species diversity. *Ecology*, *98*(5), 1201–1216. https://doi.org/10.1002/ecy.1761

Troudet, J., Grandcolas, P., Blin, A., Vignes-Lebbe, R., & Legendre, F. (2017). Taxonomic bias in biodiversity data and societal preferences. *Scientific Reports*, *7*(1), 9132. https://doi.org/10.1038/s41598-017-09084-6

Van Noorden, R., & Butler, D. (2019). Science in Europe: by the numbers. *Nature*, *569*(7757), 470–471. https://doi.org/10.1038/d41586-019-01565-0

Vannote, R. L., Minshall, G. W., Cummins, K. W., Sedell, J. R., & Cushing, C. E. (1980). The River Continuum Concept. *Canadian Journal of Fisheries and Aquatic Sciences*, *37*(1), 130–137. https://doi.org/10.1139/f80-017

Verdin, K. ., & Verdin, J. . (1999). A topological system for delineation and codification of the Earth's river basins. *Journal of Hydrology*, *218*(1–2), 1–12. https://doi.org/10.1016/S0022-1694(99)00011-6

Wendland, K. J., Baumann, M., Lewis, D. J., Sieber, A., & Radeloff, V. C. (2015). Protected Area Effectiveness in European Russia: A Postmatching Panel Data Analysis. *Land Economics*, *91*(1), 149–168. http://www.jstor.org/stable/24243797

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. URL: https://CRAN.R-project.org/package=ggplot2

SUPPLEMENTARY MATERIALS

A) SAMPLING EFFORT, ENVIRONMENTAL AND SOCIOECONOMIC VARIABLES (USED AND NOT USED)

Table S1. Preliminary list of variables.

Variable	Meaning	Units	Rationale
Soil uses	Type of soils uses according to the CORINE classification	Type of soil	Changes in soil usage and the type of soil (forests, shrub-lands...) could affect the presence of new species of aquatic insects
Global extent of ice sheets at the Last Glacial Maximum (LGM)	The LGM was the last time, during the Last Glacial Period, where the ice sheets were at their greatest extent	km	There may be differences regarding the presence of species depending on whether or not an area was affected by the LGM
Global surface temperature anomalies	Departure from a reference value or long-term average	ºC	Changes in the temperature could affect the discovery of new species of aquatic insects, as well as undermining the survival rates
Increase in the number of droughts	Increase in the number of days that water availability is lower than what is required	months	An increase in the number of droughts and the increase in their severity (especially due to climate change) can cause changes in the distribution of species
Elevation (MDE)	Height above sea level	m	More remote areas (such as those in higher elevation), more chances to have undescribed species.
IUCN Red List of threatened species	The IUCN Red List is a classification of the status of plants, animals and other organisms threatened with extinction	Number of species endangered in each country	More threatened species more chances to have threaten/extinct species before describe them (what is bad for one species could be bad for others)

Species per spatial unit according to GBIF	Number of recorded species for each spatial unit (HydroSHEDS, 4) base on the data provided by GBIF	Number of species per country	Countries with more species could potentially have more species to describe
Projected change in seasonal streamflow	Changes in streamflow	m^3/s	Drastic changes in the seasonal streamflow could cause loss of habitats
Terrestrial and aquatic ecoregions	Area that has a unique ecology, climate, geomorphology, soil, hydrology, flora and fauna	Km^2	Ecoregions could determine the number of hotspots, so certain ecoregions could have more potential hotspots than others
Natura 2000 network	Terrestrial and Marine Areas under the protection of the Natura 2000 network (only applicable to EU members)	Km^2	Sites under this protection figure could be shelter for new species (although freshwater ecosystems are underrepresented in the Natura 2000 network)
Protected Areas	World Database on Protected Areas (WDPA)	%	As well as with the Natura 2000 network, protected areas could be shelter for new species
Precipitation	Monthly precipitation	mm	Precipitations levels play a big role in species distributions. Certain level of annual precipitation could favour the presence of certain species
Minimum and maximum temperatures	Monthly temperatures	ºC	Similar to precipitation, temperature ranges also play a big role in species distributions. Species tolerate specific temperature ranges, for example, some species could favour mild temperature ranges against more cold ones
Annual Mean Temperature	Annual mean temperature	ºC	Bioclimatic variables could give a better explanation about species distribution. For that it is considered annual trends, seasonality and limiting factors
Mean Diurnal Range	Mean of the monthly temperature ranges	ºC	
Isothermality	How large the day-to-night temperatures oscillate relative to the annual oscillations	%	

Temperature Seasonality	Amount of temperature variation over a year based on the variation of monthly temperature averages	ºC
Max Temperature of Warmest Month	Maximum monthly temperature occurrence over a given year	ºC
Min Temperature of Coldest Month	Minimum monthly temperature occurrence over a given year	ºC
Temperature Annual Range	Temperature variation over a given period	ºC
Mean Temperature of Wettest Quarter	Quarterly index that approximates mean temperatures that prevail during the wettest season	ºC
Mean Temperature of Driest Quarter	Quarterly index that approximates mean temperatures that prevail during the driest quarter	ºC
Mean Temperature of Warmest Quarter	Quarterly index that approximates mean temperatures that prevail during the warmest quarter	ºC
Mean Temperature of Coldest Quarter	Quarterly index that approximates mean temperatures that prevail during the coldest quarter	ºC
Annual Precipitation	Sum of all total monthly precipitation values	mm
Precipitation of Wettest Month	Total precipitation that prevails during the wettest month	mm
Precipitation of Driest Month	Total precipitation that prevails during the driest month	mm

Precipitation Seasonality	Measure of the variation in monthly precipitation totals over the course of the year. This index is the ratio of the standard deviation of the monthly total precipitation to the mean monthly total precipitation	%	
Precipitation of Wettest Quarter	Quarterly index that approximates total precipitation that prevails during the wettest quarter	mm	
Precipitation of Driest Quarter	Quarterly index that approximates total precipitation that prevails during the driest quarter	mm	
Precipitation of Warmest Quarter	Quarterly index that approximates total precipitation that prevails during the warmest quarter	mm	
Precipitation of Coldest Quarter	Quarterly index that approximates total precipitation that prevails during the coldest quarter	mm	
Future tendencies of temperature and precipitation	Model predictions about changes in precipitation and temperature regimes	ºC and mm	Future tendencies in precipitation and temperatures could be used as a way to predict where new species of aquatic insects may be discovered
Class of Köppen-Geiger climate classification of the site where the species was first discovered	Climatic classification in 5 big groups, subdivided in 30 classes that indicate the levels of precipitation and temperatures that characterized each climate and the type of vegetation	ºC and mm	The behaviour of temperatures and rainfall can condition the presence of biodiversity hotspots, with combinations of temperatures and temperatures more favourable than others

Natural discharge	Volume of water flowing through a river channel	m^3/s	Fluctuations of natural discharge could determine the presence of a hotspot
Number of Universities per country (Tertiary education)	University institutions in each country	Universities per country	More universities per country, more researchers working on taxonomy
Expenditure on education in the EU	Resources dedicated towards education amongst the EU members	€	More money dedicated on education could increase the number of experts dedicated to taxonomy
Research and development expenditure (EU)	Money intended for R&D projects	€	Countries that assign more money on R&D projects might invest also more on taxonomy
Gross domestic product (GDP) growth	Monetary value of the production of goods and services of final demand of a country. Indicates how fast the economy of a country is growing	annual %	Countries with a higher GDP growth could host more hotspots of biodiversity because said countries have more resources to spend on taxonomic work
Soil sealing index EU	Soil surfaces sealed with impervious materials	% of total surface or Km^2	An increase in the impervious of the ground could reduce the presence of aquatic insects
Water Exploitation Index plus (WEI+)	Measure of total fresh water use as a percentage of the renewable fresh water resources (groundwater and surface water) at a given time and place	% of long term average available water	More exploitation of water resources could reduce suitable habitats for aquatic insects
Water use and stress (global)	Global demand of water resources being higher than the available resources	m^3	Countries with higher water stress and/or use could endanger/reduce aquatic insects' habitats
Surface of sites under Natura 2000	Terrestrial and Marine Areas under the protection of the Natura 2000 network (only applicable to EU members)	Km^2	Protected areas could also protect undiscovered species

Expenditure & investments in environmental protection	Money spent by the government in environmental protection (taxes, fees, etc.)	Millions € or million units of national currency	A higher spending on environmental policies could help protect potential hotspots
World population	Global population	People per country	An increase in the number of population could cause a decrease in insect populations due to urbanization and other anthropic impacts
Population density	Population per unit area	people/km^2	Higher population densities, more impacts on freshwater ecosystems and less chances to discover new species (probably already extinct)
Rivers - Urban areas distance	Distance between rivers and urban areas	Km	Urbanization near rivers could cause a drop in the number of aquatic insects
Obstacles in the river channel	Infrastructures that could potentially modify river's streamflow	Number of infrastructures	Places with more irrigation infrastructures and/or dams could reduce the number of hotspots, since changes in river regimes affect habitats availability
Distance between descriptor and species	-	Km	Sometimes people catch insects and then send it to specialists, meaning that the descriptor and the described specie could be in different countries
Naturalists per unit	Number of naturalists per HydroSHEDS unit	Count	More naturalists going to the field to collect specimens, more chances to describe new species.
Human footprint	Human influence in every biome on the land's surface	Index value	The more human footprint, the less probability of finding a hotspot (habitat fragmentation/destruction)

B) TESTED MODELS

Table S2. Models equation and AIC for all complex models tested.

Response variable	Model equation	AIC
Recently described species per study basin*	Universities + Elevation + Latitude + Population density + Percent protected + Researchers + GBIF*	398.17
	Universities + Elevation (quad.) + Latitude + Population density + Percent protected + Researchers + GBIF*	376.03
	Longitude + Latitude	473.37
	Longitude + Elevation	431.98
	Longitude + Elevation (quad.)	394.31
	Longitude + Population density	471.65
	Longitude + Researchers	474.10
	Longitude + Universities	442.53
	Longitude + Percent protected	474.24
	Longitude + GBIF*	459.80
	Latitude + Elevation	432.75
	Latitude + Elevation (quad.)	398.24
	Latitude + Precipitation	473.19
	Latitude + Population density	470.59
	Latitude + Researchers	473.47
	Latitude + Universities	434.66
	Latitude + GBIF*	452.47
	Latitude + Percent protected	473.54
	Elevation + Population density	429.30
	Elevation (quad.) + Population density	394.31
	Elevation + Researchers	432.09
	Elevation (quad.) + Researchers	396.21

	Elevation + Universities	394.81
	Elevation (quad.) + Universities	**369.29**
	Elevation + Precipitation	433.48
	Elevation (quad.) + Precipitation	398.80
	Elevation + Percent protected	433.09
	Elevation (quad.) + Percent protected	398.87
	Elevation + GBIF*	410.51
	Elevation (quad.) + GBIF*	387.15
	Population density + Universities	442.46
	Population density + Researchers	471.90
	Population density + Precipitation	471.51
	Population density + Percent protected	472.00
	Population density + GBIF*	458.23
	Researchers + GBIF*	459.82
	Researchers + Precipitation	473.97
	Researchers + Universities	441.32
	Researchers + Percent protected	474.71
	Universities + GBIF*	441.05
	Universities + Precipitation	442.24
	Universities + Percent protected	442.00
	Percent protected + Precipitation	474.16

**Transformed using log(1+x).*