The distribution and richness of the Italian riverine fish provided by the BioFresh database

Authors: Sibilia, Andrea, Orioli, Valerio, Trasforini, Stefania, Puzzi, Cesare Mario, and Bani, Luciano

The distribution and richness of the Italian riverine fish provided by the BioFresh database

Andrea SIBILIA[1,2], Valerio ORIOLI[2], Stefania TRASFORINI[1]
Cesare Mario PUZZI[1] and Luciano BANI[2]*

[1] *Environmental Research in Fishing Waters, Via Repubblica 1, I-21020 Varano Borghi (Varese), Italy*
[2] *Department of Earth and Environmental Sciences, University of Milano-Bicocca, Piazza della Scienza 1, I-20126 Milano, Italy; e-mail: luciano.bani@unimib.it*

Abstract. Freshwater fish are recognized to be endangered worldwide, but wide gaps in species distribution knowledge limit the implementation of effective conservation plans. The BioFresh project aims to produce the widest synthesis at the national level. The aim of this research is assessing the comprehensiveness and representativeness of information included in the Italian synthesis, composed of 10000 sampling sites and 50000 fish presence records. The assessment was produced at the secondary river basin (SRB) level in two distinct periods, past (1984-1999) and recent (2000-2014), comparing different statistical models of frequency counts of species detected within each SRB. The results highlighted a poor overall knowledge of the Italian fish distribution for both periods. In the all-species dataset, only 11.8 % and 11.1 % of the SRBs were found to have been sufficiently explored in the past and recent periods, respectively. In the native species dataset, these percentages increased to 16 % and 24.7 % respectively, which suggests the presence of background noise in the richness estimates produced by exotic species. Although the information available is far from exhaustive, the BioFresh project, in which data collected over a wide time span are georeferenced and freely available, represents the first synthesis for future research aimed at the management and conservation of the Italian fish fauna.

Introduction

Knowledge of the distribution of fauna and its changes over time is fundamental in facilitating the investigation of factors affecting species ecology and to implement appropriate conservation management. Unfortunately, few data are available and are mostly limited to restricted areas and short periods. In Italy, large survey programs aimed at obtaining standardized data on wildlife distribution or abundance are relatively scarce, and the few available attempts at synthesis are often based on heterogeneous data sources collected over a wide timespan (e.g. Meschini & Frugis 1993, for the Atlas of Italian breeding birds; Sindaco et al. 2006, for the Atlas of Italian amphibians and reptiles). Even more uncommon are long-term regular surveys within structured large-scale monitoring programs (e.g. Bani et al. 2009, for breeding birds in Lombardy; Fornasari et al. 2010, for breeding birds in Italy; Fasola et al. 2011, for herons population in the Po Plain), although these are necessary to meet the conservation requirements of the Habitats Directive (92/43/EEC, art. 17) and Birds Directive (79/409/EEC, art.12).

The lack of comprehensive spatial and temporal abundance data also applies to fish (but see Gallo et al. 2012). This situation is particularly worrying since freshwater fish are one of the most endangered animal groups as a result of widespread threats, which often act simultaneously and synergistically on freshwater ecosystems (Bianco 1995, Dudgeon et al. 2006, Olden et al. 2010, Freyhof & Brooks 2011, Fochetti 2012). In particular, fish are threatened by habitat alteration and depletion, due to the creation of artificial banks and to water uptake for irrigation and hydroelectric exploitation. These processes can also lead to habitat fragmentation, the effects being worsened by the presence of several types of hydraulic structures along rivers, such as weirs, dams and floodgates (Dudgeon et al. 2006, Kottelat & Freyhof 2007). Moreover, the introduction of exotic species can alter the structure of communities, modifying interspecific relationships such as predation and competition, or promoting the hybridisation of closely related taxa (Kottelat & Freyhof 2007, Gozlan et al. 2010, Volta et al. 2013). Finally, climate change may further reduce habitat

* Corresponding Author

225

availability, especially in the Mediterranean region where water resources are already scarce, especially during the dry season (Xenopoulos et al. 2005, Olden et al. 2010).

Assessment of the conservation status of fish and of the effects of the main threats cannot be easily achieved, due to a lack of knowledge about the distribution and abundance of the majority of species (Olden et al. 2010). For instance, in Italy the bulk of data concerning fish distribution come from technical reports produced by local public authorities (e.g. park and provincial councils), whose aim is primarily related to sport fishing management and, in some cases, to the implementation of local conservation measures. Although reports are produced outside academic publishing and distribution channels and are thus classified as "grey-literature", they provide information that can be precisely located in space and time. However, they are often available only as printed papers and, thus, not directly usable in Geographic Information Systems (GISs), to perform statistical and territorial analyses, except through preliminary digital processing work.

Recently, the GRAIA Company collected and standardised all available information on Italian fish from 1895 to 2014, producing the "Distribution Data of Italian Freshwater Fish" database (DDIFF hereafter; Puzzi & Ippoliti 2015). The database has been mobilised through BioFresh, which is a network for global freshwater biodiversity, published on http://data.freshwaterbiodiversity.eu. The purpose of BioFresh is the creation of a digital platform published online, available to all users and containing all available information about the biodiversity of European freshwater ecosystems. The DDIFF, which is included in the BioFresh platform, should overcome some of the constraints on data availability and provide the first national synthesis of fish species distribution in Italy, freely available in a digital and georeferenced form. Given the long time period covered, the database could also be used to explore changes over time in species' distribution and communities, including threats associated with exotic species.

The aim of this research was to understand how realistically the DDIFF reflects the actual distribution of riverine fish species in Italy, by assessing the comprehensiveness and representativeness of the information included in the database. Data comprehensiveness was assessed by evaluating the spatial and temporal coverage of data at the national level, whereas data representativeness was evaluated by comparing the observed and estimated richness of native species within river basins. Finally, we assessed the eventual uncertainty in richness estimation induced by the presence of exotic species.

Material and Methods

The Distribution Data of Italian Freshwater Fish database

The DDIFF, developed in a spreadsheet and georeferenced form in a GIS workspace according to the WGS84 UTM 32N coordinate system (EPSG 32632), includes three main tables linked by primary keys, which contain information about coordinates, sampling sites and observed species. Each sampling site was linked to city, provincial and regional boundaries, and to the 10 km grid square of the Military Grid Reference System (MGRS) in which it fell. All the vector layers used to build up the database are freely available online: watercourses were downloaded from the SINAnet cartographic portal (http://www.sinanet.isprambiente.it/it/sia-ispra/download-mais/, accessed on 13 April 2015), the MGRS grid from the National Geospatial-Intelligence Agency website, and the boundaries of city, provincial and regional territories were downloaded from the ISTAT website (http://gisportal.istat.it/, accessed on 13 April 2015). Watercourses were edited in order to add all those that were not already included in the national vector layer (mainly artificial watercourses), and to integrate the data with regional layers (particularly for the Po plain regions including Piedmont, Lombardy, Emilia-Romagna, Veneto and Friuli-Venezia Giulia).

Overall, the data contained in the DDIFF come from almost 250 different sources (see Appendix 1) and cover a period of 120 years, starting from 1895. The dataset contains information from about 10000 sampling sites and over 50000 fish presence records.

Data management

The Italian hydrographic network is complex due to the presence of extensive mountain chains (Alps and Apennines). These features also affect the distribution of freshwater fish communities characterized by a high turnover rate from mountains to lowland (Reyes-Gavilan et al. 1996, Zerunian 2002a). For this reason, we divided the national territory according to the secondary river basins, which represent a good trade-off between the number of sampling sites of the DDIFF (within each secondary river basin) and the most detailed spatial resolution. We performed the subsequent analyses considering each secondary river basin as an independent study area, in order to summarize and map the results obtained.

226

We acquired the secondary river basins layer from the SINAnet cartographic portal (http://www.sinanet. isprambiente.it/it/sia-ispra/download-mais/, accessed on 27 December 2016). The secondary river basins were then overlapped with the national boundaries in order to add the missing seaward basins and ensure complete coverage of the national territory. We also divided the larger basins within which rivers, flowing from the mountain springs to the sea or their lower confluences, cross different river zones and host completely different fish communities (e.g. the Ticino River basin was divided into two parts, one including the river as a tributary of Lake Maggiore and the second including the same river as an emissary of the lake). This yielded 288 smaller secondary river basins (hereafter SRBs; Fig. S1) characterised by more homogenous fish communities. Finally, each sampling site was linked to the corresponding SRB.

While the DDIFF encompasses a long time period, most data were obtained from the late 1980s onwards and are mostly based on electrofishing surveys. Therefore, we decided to exclude from our analyses all data collected before 1984 (although they are included in the dataset, they were too scarce to be representative of the species distribution at a national scale, amounting to about 1 % of all data; see original data source at http://data.freshwaterbiodiversity.eu), or obtained by techniques other than electrofishing. This refinement led to the exclusion of all data collected in lakes, as lake surveys were carried out using nets. Finally, we arbitrarily split the data into two groups, each representing a 15-year period, 1984-1999 and 2000-2014 (with samples distributed equitably between the two periods; see Results). We split the dataset into two 15-year periods to capture the substantial changes that freshwater fish communities have undergone during recent decades due to habitat alteration and the introduction of exotic species (Italian native species were considered exotic when translocated in SRBs outside their natural range). However, a time period shorter than 15 years would not have provided a sufficiently robust sample for assessing SRB species richness.

Species richness analyses
The estimation of species richness was performed by comparing different statistical models of frequency counts by means of CatchAll version 4.0 (Bunge et al. 2012). This software allows the comparison of classic non-parametric models with parametric finite-mixture models and a weighted linear regression, which can account for the potential bias caused by

rare species (see Mao & Colwell 2005). Moreover, the software computes the standard errors of the estimates and some measures of goodness-of-fit, along with a model robustness assessment. Overall, CatchAll can perform 12 models grouped into three categories: five parametric models, including a Poisson model (*Poisson*) and four mixed models (single exponential (*SingleExp*), two- (*TwoMixedExp*), three- (*ThreeMixedExp*) and four- (*FourMixedExp*) exponential mixed Poisson models); two weighted linear regression models (*WLRMs*) performed on log-transformed (*LogTransfWLR*) or untransformed data (*UnTransfWLR*); five non-parametric models, encompassing the Good-Turing (*GoodTuring*), Chao1 (*Chao1*), the Abundance-Based Coverage Estimator (*ACE*) and its high-diversity variant (*ACE1*); and the Chao-Bunge (*ChaoBunge*) models.

All estimations were carried out on a specific input dataset consisting of "frequency counts" (Woodard et al. 2013), defined as a list of occurrence frequencies followed by the number of species occurring a given number of times (corresponding to the number of species detected in a given number of sampling sites) in each SRB. Since frequency-count data usually exhibit a large number of rare species and a small number of common species, parametric models were not only fitted on the entire dataset, but also on sub-sets omitting some outliers, according to a threshold defined by the parameter τ, which is the number of species occurring a given number of times. In particular, CatchAll fits several models for different values of τ, deleting all frequency-count data $\geq \tau$, computing their goodness-of-fit and obtaining a partial estimate. The final estimate for each model is provided by adding the number of species with counts greater than τ to the partial estimate (Woodard et al. 2013).

The estimation of species richness was performed for each SRB separately, using both the complete dataset (native and exotic species combined: overall species richness) and the dataset composed of native species only (native species richness), for the two different time periods. The presence of exotic species can produce a "background noise" in richness estimates because their distribution is strongly affected by introduction events, rather than by their natural history. Release events could alter the overall community richness estimates, especially when the number of events is low (i.e. limited to some sampling units) and when the number of released species during these events is relatively high compared to the number of local native species.

Thus, we always considered four datasets covering "past" (1984-1999) and "recent" periods (1999-

227

2014), in order to account for both native species and overall species. We did not perform any estimation for exotic species alone, as the distribution of these species depends on where they have been released. We assigned to each basin the estimate of species richness obtained from the best parametric model. If the data were insufficient to obtain an estimate with a parametric model, we used the estimate from the best non-parametric model. In some cases there were not sufficient data to obtain an estimate.

For each SRB, the estimation of species richness was performed by an iterative process implemented in R (R Development Core Team 2008), which calls CatchAll to generate different models.

We defined as "explored SRBs" the SRBs having at least one sampling unit. We defined the representativeness of collected data using the following three categories: "data deficient SRBs": SRBs for which the modelling process did not succeed in providing the richness estimate; "under-sampled SRBs": SRBs for which the observed richness fell outside the 95 % confidence boundaries of the estimate; "fully-sampled SRBs": SRBs for which the observed richness fell within the 95 % confidence boundaries of the estimate. For both survey periods, we produced maps showing the observed richness for each SRB (all species and native species), the distribution of exotic species and the estimated richness and its standard error. In addition, we showed the degree of completeness of archive data for each SRB by generating maps of the difference between estimated (considering the central value of the estimate) and observed species richness.

Finally, we assessed the contribution of exotic species to the uncertainty of the estimate of overall species richness. Through a multiple linear regression in R, we related the standard error of the estimate of overall species richness to the number of exotic species in each SRB, accounting for the number of sampling units and the possible effect of the sampling period.

Results

The distribution of fish was described by 52765 presence records collected at 9756 sampling sites (4841 and 4915 of which belonging to the 1984-1999 and to the 2000-2014 period, respectively). A total of 2064 (31.2 %) sampling sites were located along rivers, 3811 (57.5 %) along streams and 749 (11.3 %) along artificial watercourses.

Considering the 1984-1999 and 2000-2014 periods separately, the all-species dataset covered 123 (corresponding to 48.6 % of the 301400 km² of Italian territory) and 179 (65 % of the Italian territory) out of 288 SRBs, respectively. Considering native species only, the number of explored SRBs amounted to 120 (41.7 % of SRBs and 47.5 % of national surface) and 169 (58.7 % of SRBs and 63.1 % of national surface) in the two survey periods, respectively (Table 1). The number of sampling units per SRB ranged from 2 to 309 (mean ± standard deviation, 29.19 ± 51.82) and from 2 to 103 (16.45 ± 16.29) for the 1984-1999 and the 2000-2014 period, respectively.

Overall, we analysed the distribution of 119 species (see Appendix 2), including 76 native and 35 exotic species, as well as eight hybrid species between roach (two exotics), between roach and chub (one native), between barbel (one native), between barbel and chub (one native) and between trout (three natives). Among the native species, 32 marine or migratory species were found near estuaries.

In the SRBs, the observed overall species richness varied between 3 and 48 in the 1984-1999 period, and between 2 and 55 in the 2000-2014 period, while the observed native species richness ranged from 2 to 29 and from 2 to 30 in the past and recent periods, respectively (Fig. 1, Fig. S2).

The data were found to be adequate to estimate the overall species richness in 85 SRBs (corresponding to 69.1 % of the explored SRBs) in the 1984-1999 period and in 132 SRBs (74.3 %) in the 2000-2014 period. The native species richness was estimated in

Table 1. Summary of data coverage from the four analysed datasets: past period (1984-1999), recent period (1999-2014), overall species and native species only, respectively. SRBs: secondary river basins.

	1984-1999 period				2000-2014 period			
	overall species		native species		overall species		native species	
	n. of basins (%)	% of national surface	n. of basins (%)	% of national surface	n. of basins (%)	% of national surface	n. of basins (%)	% of national surface
Number of explored SRBs	123 (42.7)	48.6	120 (41.7)	47.5	179 (62.2)	65	169 (58.7)	63.1
Number of data deficient SRBs	39 (13.5)	10.6	53 (18.4)	14.2	47 (16.3)	8.1	67 (23.3)	15.9
Number of basins with richness estimate	85 (29.6)	38.1	67 (23.3)	33.3	132 (45.8)	56.9	104 (36.1)	47.1
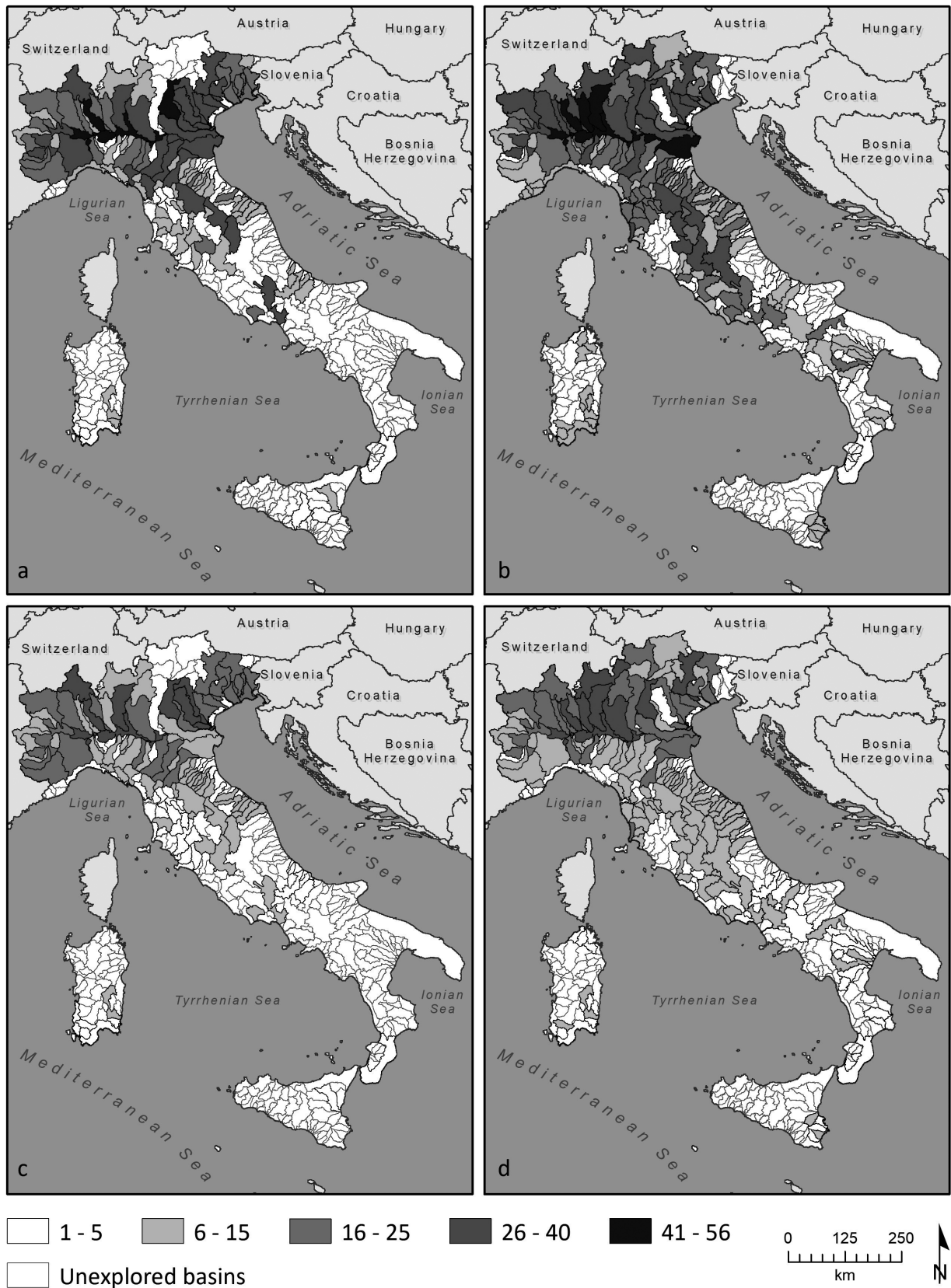
228

Fig. 1. Observed richness: a) 1984-1999 overall; b) 2000-2014 overall; c) 1984-1999 native; d) 2000-2014 native.
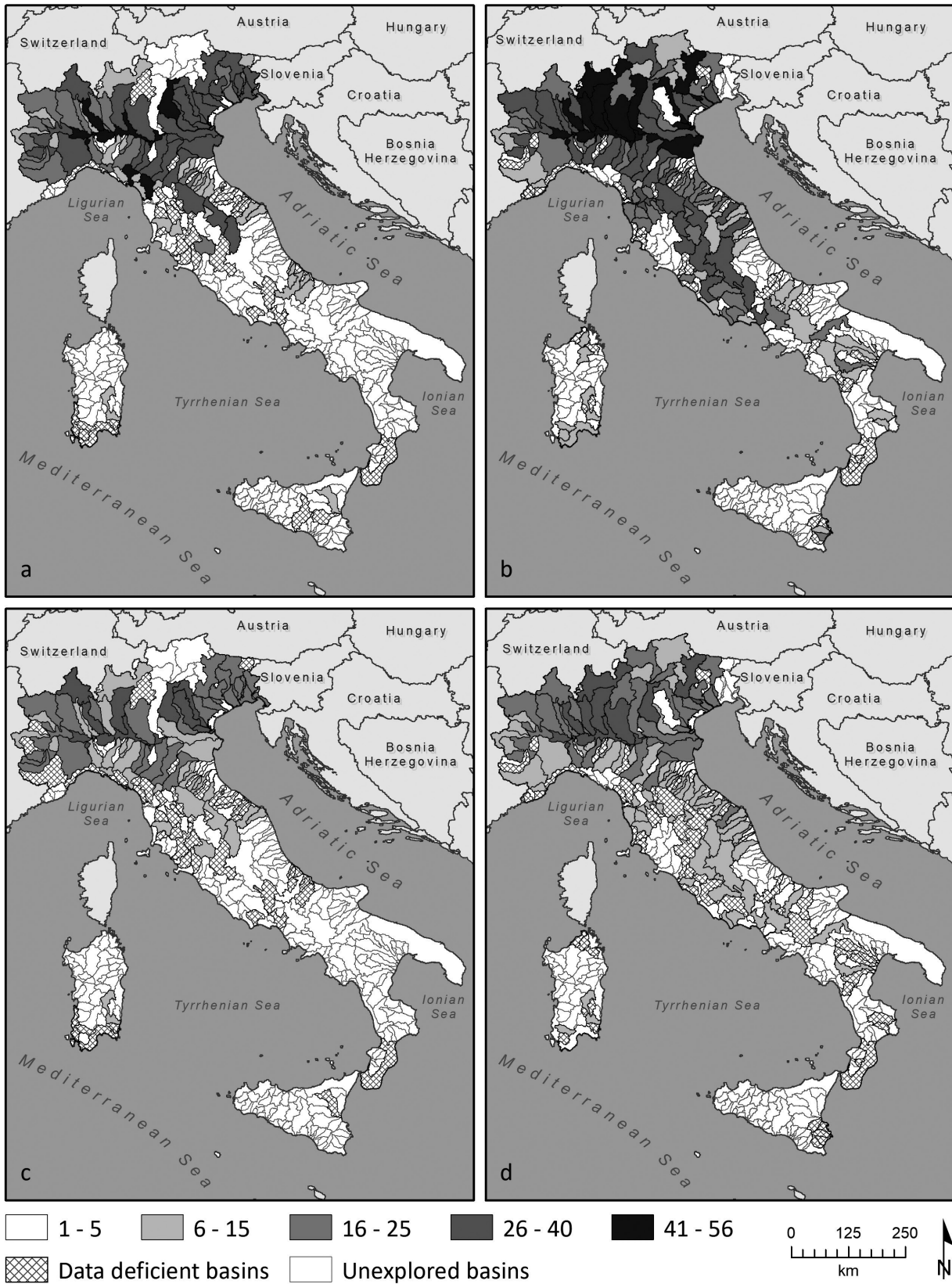
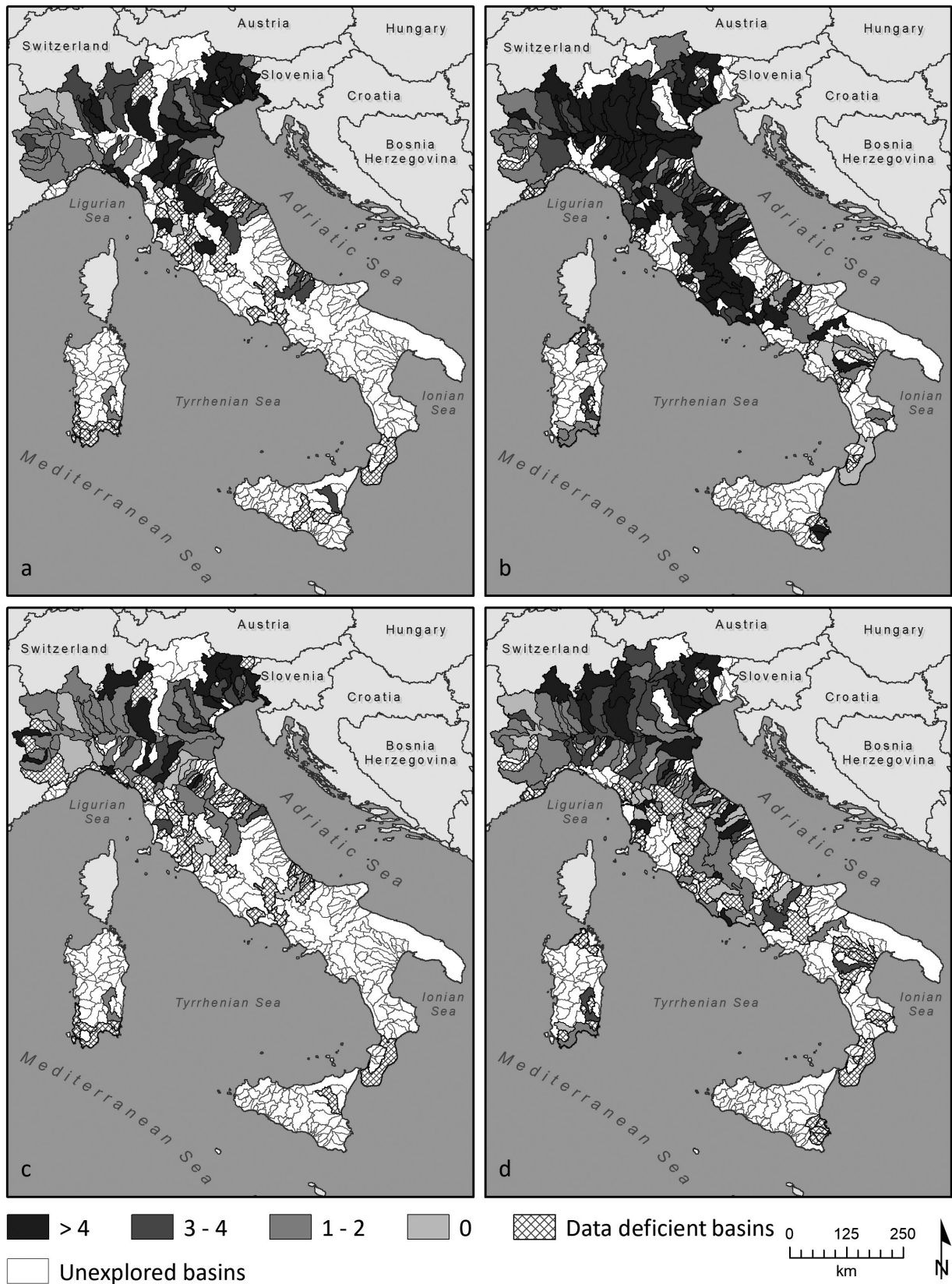Legend: 1 - 5 | 6 - 15 | 16 - 25 | 26 - 40 | 41 - 56 | Unexplored basins

0   125   250 km

Fig. 2. Estimated richness: a) 1984-1999 overall; b) 2000-2014 overall; c) 1984-1999 native; d) 2000-2014 native.

Legend:
1 - 5 | 6 - 15 | 16 - 25 | 26 - 40 | 41 - 56

Data deficient basins | Unexplored basins

0 125 250 km | N

Fig. 3. Data completeness, representing the difference between estimated and observed species richness: a) 1984-1999 overall; b) 2000-2014 overall; c) 1984-1999 native; d) 2000-2014 native.

67 (55.8 %) and in 104 (61.5 %) SRBs for the past and recent periods, respectively. Considering the four datasets (all species and native species only, in the past period and the recent period), the *SingleExp* parametric model was found to be the most effective, since it was the best-fitting model for richness estimation in 40-66 % of those SRBs with sufficient data to perform the analysis. The second most effective model was *ACE*, which succeeded in 16-42 % of cases. *Poisson* succeeded in 11-16 % of cases. The other model never exceeded 4 % of cases. The estimated overall species richness per SRB varied between 5-55 in the 1984-1999 period and between 4-59 in the 2000-2014 period, while the estimated native species richness ranged from 4-37 for both periods (Fig. 2, Fig. S3).

Data completeness varied substantially between the two periods and the two species datasets. In the past period (1984-1999), considering the all-species dataset, 11.8 % of SRBs were found to have been sufficiently explored, since the difference between the estimated and observed species richness was lower than three. According to the same threshold, considering the same period for the "native species only dataset", 16 % of the SRBs were found to have been sufficiently explored. For the recent period (1999-2014), considering the all-species dataset, completeness was found to be even more unsatisfactory, since the sufficiently explored SRBs amounted to 11.1 %, although the value increased to 24.7 % if only the native species were considered (Fig. 3, Fig. S4).

The multiple linear regression highlighted the effects of exotic species on the standard error of the estimate of overall species richness (Fig. S5). The higher the number of exotic species (Fig. S6a, b), the higher the standard error (coefficient and standard error of the multiple linear regression, 0.093 ± 0.029; $p = 0.001$). As expected, the number of sampling units negatively affected the standard error of the estimate (-0.013 ± 0.004; $p < 0.001$), while the sampling period did not have a significant effect ($p = 0.928$).

Discussion

The *SingleExp* parametric model, pertaining to parametric finite-mixed models, was found to be the most effective in estimating species richness. Parametric models proved to be more effective than non-parametric models in richness estimation since they were less affected by rare species and sampling effort (Mao & Colwell 2005, Ter Steege et al. 2017).

The BioFresh data highlighted a poor overall knowledge of Italian riverine fish distribution. In the past period (1984-1999), just over 40 % of the SRBs (corresponding to less than 50 % of Italian territory) were explored, and just under 30 % of SRBs had sufficient data to obtain a realistic estimate of species richness. The survey effort has recently increased (2000-2014) and more than 60 % of the SRBs (reaching 65 % of the national surface) were covered. Consequently, the percentage of SRBs for which richness estimates were obtained increased to just over 45 %. Considering the native species only, the situation did not change substantially between the two survey periods. Despite the increased survey effort, the mean difference between the estimated and observed richness was not found to be significantly different between survey periods. Indeed, for the all-species dataset the mean difference ($\pm$ standard error) = 4.92 ± 0.72 for 1984-1999 and 6.28 ± 0.70 for 2000-2014 (t-test for paired data: $t = 1.53$, df = 57, $p = 0.133$). For the "native species only" dataset, the difference between the estimated and observed richness = 2.51 ± 0.48 for 1984-1999, and 3.04 ± 0.42 for 2000-2014 ($t = 0.881$, df = 51, $p = 0.383$).

However, the completeness; i.e. the difference between the estimated and observed richness, was significantly different between the all-species dataset and that including only native species ($t = 9.13$, df = 189, $p < 0.001$). The mean difference for the all-species dataset (mean difference $\pm$ standard error: 5.33 ± 0.37) was larger than that of the "native species only" dataset (2.54 ± 0.23). The lower effectiveness of richness estimate using the whole dataset (native and exotic species combined), rather than the "native species only" dataset, is an indication of the possible effect played by exotic species in enhancing the uncertainty of richness estimates. Indeed, an increasing number of exotic species in each SRB produced an increase in the standard error in the corresponding richness estimate.

Overall, while northern Italy has been explored relatively extensively, the South still appears to have been poorly investigated. Some uncertainties in species distributions in southern Italy have recently been filled, though wide gaps remain, especially in SRBs facing the central and southern Adriatic side and the southern Tyrrhenian side, as well as in those pertaining to the two main islands of Sicily and Sardinia. Although the overall completeness of knowledge has improved over time, the available information is not yet satisfactory. For instance, considering the differences between the estimated and observed species richness, knowledge of the central and eastern Po basin has decreased from the past to the recent period, even when considering only native

232

species. Although there has recently been a noticeable improvement in fish distribution data for several SRBs in central Italy, wide gaps remain for southern Italy and the islands. However, the available data are not always exhaustive, since they do not always allow the estimation of richness (e.g. most SRBs of the Umbrian-Tuscan Apennines).

In conclusion, the analysis of the effectiveness of the DDIFF database provided the first systematic study of riverine fish distribution in Italy, obtained from available data gathered through field surveys, and collected along the widest available temporal range. The resulting information shows that an understanding of species' distributions is poor, both for the past and recent periods. Even though the sampling effort has developed during recent decades, the information obtained is not always exhaustive. The knowledge gaps represent important shortcomings when trying to establish effective conservation programs for a taxonomic group recognized as one of the most threatened among vertebrates (Agapito Ludovici & Zerunian 2008, Freyhof & Brooks 2011), especially given the high degree of endemicity in the Mediterranean Basin (Smith & Darwall 2006). Indeed, the conservation status and distribution of Italian freshwater fish is strongly affected by human activities, the effects of which have become more severe since the beginning of the twentieth century, when agricultural and industrial development widely changed the national landscape and its waterbodies, particularly in the lowland areas (Gandolfi et al. 1991, Zerunian 2002a, b). The results also showed the high proportion of exotic species within fish communities, a threat that has increased from the past to the recent period (Fig. S6c, d), especially in central Italy.

In this context, the present research represents a benchmark for further studies, since it (i) highlights in detail (at the SRB scale) the regions that lack sufficient data to produce reliable richness estimates; (ii) shows the areas that potentially represent hot-spots for fish diversity (areas with the highest estimates of native species richness); (iii) links all the available information to the relevant geographic areas (i.e. SRBs); (iv) is a useful source of information to implement effective management measures at a regional or a river basin scale.

Literature

Agapito Ludovici A. & Zerunian S. 2008: Acque in Italia. L'emergenza continua: a rischio molte specie di pesci. *WWF Italia, Roma, Italy*.

Bani L., Massimino D., Orioli V. et al. 2009: Assessment of population trends of common breeding birds in Lombardy, northern Italy, 1992-2007. *Ethol. Ecol. Evol. 21: 27–44.*

Bianco P.G. 1995: Mediterranean endemic freshwater fishes of Italy. *Biol. Conserv. 72: 159–170.*

Bunge J., Woodard L., Böhning D. et al. 2012: Estimating population diversity with CatchAll. *Bioinformatics 28: 1045–1047.*

Dudgeon D., Arthington A.H., Gessner M.O. et al. 2006: Freshwater biodiversity: importance, threats, status and conservation challenges. *Biol. Rev. 81: 163–182.*

Fasola M., Merli E., Boncompagni E. & Rampa A. 2011: Monitoring heron populations in Italy, 1972-2010. *J. Heron Biol. Conserv. 1: 1–10.*

Fochetti R. 2012: Italian freshwater biodiversity: status, threats and hints for its conservation. *Ital. J. Zool. 79: 2–8.*

Fornasari L., Londi G., Buvoli L. et al. 2010: Distribuzione geografica e ambientale degli uccelli comuni nidificanti in Italia, 2000-2004 (dati del progetto MITO2000). *Avocetta 34: 5–15.*

Freyhof J. & Brooks E. 2011: European Red List of freshwater fishes. *Publications Office of the European Union, Luxembourg*.

Gallo L., Lucadamo L., Mezzotero A. et al. 2012: First data on the freshwater fish fauna of Calabria (southern Italy). *Ital. J. Zool. 79: 246–251.*

Gandolfi G., Zerunian S., Torricelli P. & Marconato E. 1991: I pesci delle acque interne italiane. *Ministero dell'Ambiente-Servizio Conservazione Natura, Unione Zoologica Italiana, Istituto Poligrafico e Zecca dello Stato, Roma, Italy*.

Gozlan R.E., Britton J.R., Cowx I. & Copp G.H. 2010: Current knowledge on non native freshwater fish introductions. *J. Fish Biol. 76: 751–786.*

Kottelat M. & Freyhof J. 2007: Handbook of European freshwater fishes. *Publications Kottelat, Cornol, Switzerland*.

Mao C.X. & Colwell R.K. 2005: Estimation of species richness: mixture models, the role of rare species, and inferential challenges. *Ecology 86: 1143–1153.*

Meschini E. & Frugis S. 1993: Atlante degli uccelli nidificanti in Italia. *Istituto Nazionale per la Fauna Selvatica, Ozzano dell'Emilia, Italy*.

Olden J.D., Kennard M.J., Leprieur F. et al. 2010: Conservation biogeography of freshwater fishes: recent progress and future challenges. *Divers. Distrib. 16: 496–513.*

Puzzi C. & Ippoliti A. 2015: Distribution data of Italian freshwater fish mobilised through BioFresh. *Global Biodiversity Information Facility (GBIF), Copenhagen, Denmark. http://data.freshwaterbiodiversity.eu*

R Development Core Team 2008: R: a language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org*

Reyes-Gavilan F.G., Garrido R., Nicieza A.G. et al. 1996: Fish community variation along physical gradients in short streams of northern Spain and the disruptive effect of dams. *Hydrobiologia 321: 155–163.*

Sindaco R., Doria G., Razzetti E. & Bernini F. 2006: Atlante degli Anfibi e dei Rettili d'Italia. *Societas Herpetologica Italica, Edizioni Polistampa, Firenze, Italy.*

Smith K.G. & Darwall W.R. 2006: The status and distribution of freshwater fish endemic to the Mediterranean Basin. *IUCN, Gland, Switzerland.*

Ter Steege H., Sabatier D., de Oliveira S.M. et al. 2017: Estimating species richness in hyper diverse large tree communities. *Ecology 98: 1444–1454.*

Volta P., Jeppesen E., Leoni B. et al. 2013: Recent invasion by a non-native cyprinid (common bream *Abramis brama*) is followed by major changes in the ecological quality of a shallow lake in southern Europe. *Biol. Invasions 15: 2065–2079.*

Woodard L., Connolly S. & Bunge J. 2013: CatchAll version 4.0, user manual. *Accessed on December 2016. http://www.northeastern.edu/catchall/downloads/CatchAll.4.0/CatchAll.Manual.4.0.pdf*

Xenopoulos M.A., Lodge D.M., Alcamo J. et al. 2005: Scenarios of freshwater fish extinctions from climate change and water withdrawal. *Glob. Change Biol. 11: 1557–1564.*

Zerunian S. 2002a: Condannati all'estinzione? Biodiversità, biologia, minacce e strategie di conservazione dei Pesci d'acqua dolce indigeni in Italia. *Edagricole, Bologna, Italy.*

Zerunian S. 2002b: Pesci delle acque interne italiane. Collana Quaderni di Conservazione della Natura. *Ministero dell'Ambiente, Istituto Nazionale per la Fauna Selvatica, Roma, Italy.*

Supplementary online materials

Fig. S1. Smaller secondary river basins (SRBs) for which the species estimation was performed. The ID number of each SRB refers to the corresponding ID number of references in Appendix 1.

Fig. S2. Observed richness: a) 1984-1999 overall; b) 2000-2014 overall; c) 1984-1999 native; d) 2000-2014 native.

Fig. S3. Estimated richness: a) 1984-1999 overall; b) 2000-2014 overall; c) 1984-1999 native; d) 2000-2014 native.

Fig. S4. Data completeness, representing the difference between estimated and observed species richness: a) 1984-1999 overall; b) 2000-2014 overall; c) 1984-1999 native; d) 2000-2014 native.

Fig. S5. Standard error of the richness estimate: a) 1984-1999 overall; b) 2000-2014 overall; c) 1984-1999 native; d) 2000-2014 native.

Fig. S6. Number of exotic species: a) 1984-1999 overall; b) 2000-2014 overall; c) 1984-1999 native; d) 2000-2014 native.

Appendix 1. Sources of data.

Appendix 2. List of species (https://www.ivb.cz/wp-content/uploads/FZ-vol.-68-4-2019-Sibilia-et-al.-Fig.-S1-S6-Appendices-1-2-1.pdf).