



**HAL**  
open science

## Global biogeographical regions of freshwater fish species

Boris Leroy, Murilo Dias, Emilien Giraud, Bernard Hugueny, Céline Jézéquel,  
Fabien Leprieur, Thierry Oberdorff, Pablo A Tedesco

► **To cite this version:**

Boris Leroy, Murilo Dias, Emilien Giraud, Bernard Hugueny, Céline Jézéquel, et al.. Global biogeographical regions of freshwater fish species. *Journal of Biogeography*, 2019, 46 (11), pp.2407-2419. 10.1111/jbi.13674 . hal-02297792

**HAL Id: hal-02297792**

**<https://hal.sorbonne-universite.fr/hal-02297792>**

Submitted on 26 Sep 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Global biogeographical regions of freshwater fish species

**Preprint manuscript submitted to the *Journal of Biogeography***

**Article type:** Research paper

**Short title:** Bioregions of freshwater fishes

**Authors:** Boris Leroy<sup>1\*</sup>, Murilo S. Dias<sup>1,2</sup>, Emilien Giraud<sup>1</sup>, Bernard Hugueny<sup>3</sup>, Céline Jézéquel<sup>3</sup>, Fabien Leprieur<sup>4,5</sup>, Thierry Oberdorff<sup>3</sup>, Pablo A. Tedesco<sup>3</sup>

<sup>1</sup> Unité Biologie des Organismes et Ecosystèmes Aquatiques (BOREA, UMR 7208), Muséum national d'Histoire naturelle, Sorbonne Université, Université de Caen Normandie, CNRS, IRD, Université des Antilles, Paris, France.

<sup>2</sup> Departamento de Ecologia, Instituto de Ciências Biológicas, Universidade de Brasília (UnB), Brasília-DF, Brazil.

<sup>3</sup> UMR5174 EDB (Laboratoire Evolution et Diversité Biologique), CNRS, IRD, UPS, Université Paul Sabatier, F-31062 Toulouse, France

<sup>4</sup> UMR MARBEC (CNRS, IRD, IFREMER, UM), Université de Montpellier, Place Eugène Bataillon, 34095 Montpellier Cedex 5, France.

<sup>5</sup> Institut Universitaire de France, Paris, France.

\*Corresponding author

## Acknowledgements

We thank Céline Bellard and Philippe Keith for useful discussions, and Aldyth Nyth and Lissette Victorero for English editing. We thank François-Henri Dupuich from [derniercri.io](http://derniercri.io) for figure editing. Laboratoire Evolution et Diversité Biologique is part of the French Laboratory of Excellence projects “LABEX TULIP” and “LABEX CEBA” (ANR-10-LABX-41, ANR-10-LABX-25-01).

## 31 **Abstract**

### 32 **Aim**

33 To define the major biogeographical regions and transition zones for freshwater fish species.

### 34 **Taxon**

35 Strictly freshwater species of actinopterygian fish (i.e., excluding marine and amphidromous  
36 fish families).

### 37 **Methods**

38 We based our bioregionalisation on a global database of freshwater fish species occurrences  
39 in drainage basins, which, after filtering, includes 11 295 species in 2 581 basins. On the basis  
40 of this dataset, we generated a bipartite (basin-species) network upon which we applied a  
41 hierarchical clustering algorithm (the Map Equation) to detect regions. We tested the  
42 robustness of regions with a sensitivity analysis. We identified transition zones between  
43 major regions with the participation coefficient, indicating the degree to which a basin has  
44 species from multiple regions.

### 45 **Results**

46 Our bioregionalisation scheme showed two major supercontinental regions (Old World and  
47 New World, 50% species of the world and 99.96% endemics each). Nested within these two  
48 supercontinental regions lie six major regions (Nearctic, Neotropical, Palearctic, Ethiopian,  
49 Sino-Oriental and Australian) with extremely high degrees of endemism (above 96% except  
50 for the Palearctic). Transition zones between regions were of limited extent compared to  
51 other groups of organisms. We identified numerous subregions with high diversity and  
52 endemism in tropical areas (e.g. Neotropical), and a few large subregions with low diversity  
53 and endemism at high latitudes (e.g. Palearctic).

### 54 **Main conclusions**

55 Our results suggest that regions of freshwater fish species were shaped by events of  
56 vicariance and geodispersal which were similar to other groups, but with freshwater-specific  
57 processes of isolation that led to extremely high degrees of endemism (far exceeding  
58 endemism rates of other continental vertebrates), specific boundary locations, and limited  
59 extents of transition zones. The identified bioregions and transition zones of freshwater fish  
60 species reflect the strong isolation of freshwater fish faunas for the past 10 to 20 million  
61 years. The extremely high endemism and diversity of freshwater fish fauna raises many  
62 questions about the biogeographical consequences of current introductions and extinctions.

### 63 **Key-words**

64 Biogeographical regions, biogeography, bioregionalisation, bioregions, dispersal, freshwater  
65 fish, actinopterygians, vicariance, zoogeographical regions, transition zones

## 66 **Introduction**

67 For almost two centuries, biogeographers have classified continental areas of the world into  
68 distinct biogeographical regions on the basis of organism distributions across the Earth  
69 (Wallace, 1876; Cox, 2001; Holt et al., 2013). Indeed, early biogeographers observed that  
70 many organisms share constellated distributions of endemics in particular regions.  
71 Furthermore, they observed that these patterns of endemism are often similar for distinct  
72 groups of organisms, resulting in very similar biogeographical regions. This marked  
73 similarity has led to the hypothesis that these regions reflect a shared history of  
74 diversification among taxa and have been conditioned by geography, geology and climate  
75 (Morrone, 2015; Lomolino, Riddle, & Whittaker, 2016).

76 The earliest classifications outlined six major biogeographic regions for birds (Sclater, 1858)  
77 and non-flying mammals (Wallace, 1876) (Nearctic, Neotropical, Palearctic, Ethiopian,  
78 Oriental and Australian). During recent years, these major regions have been confirmed by  
79 an upsurge in bioregionalisation studies. This upsurge was facilitated by the increase in  
80 quality and quantity of large-scale datasets, as well as the development of new analytical tools  
81 (Kreft & Jetz, 2010; Vilhena & Antonelli, 2015; Edler, Guedes, Zizka, Rosvall, & Antonelli,  
82 2016). Consequently, multiple studies have tried to identify the major biogeographical  
83 regions for birds (Procheş & Ramdhani, 2012; Rueda, Rodríguez, & Hawkins, 2013; Holt et  
84 al., 2013), mammals (Kreft & Jetz, 2010; Procheş & Ramdhani, 2012; Rueda et al., 2013; Holt  
85 et al., 2013), amphibians (Procheş & Ramdhani, 2012; Rueda et al., 2013; Holt et al., 2013;  
86 Vilhena & Antonelli, 2015; Edler et al., 2016) and reptiles (Procheş & Ramdhani, 2012). The  
87 result of this upsurge was a debate on the precise limits of biogeographical regions. Indeed,  
88 some studies explicitly defined transition zones as distinct regions (e.g., Holt et al., 2013),  
89 whereas others included transition zones in major regions (Kreft & Jetz, 2010). This question  
90 of transition zones was settled to some extent in the major synthesis of Morrone (2015),  
91 proposing that transition zones should not be considered as distinct regions, but rather as  
92 transitional boundaries between major regions. Indeed, some regions share sharp  
93 boundaries, reflecting a long history of isolation by tectonics (Ficetola, Mazel, & Thuiller,  
94 2017), whereas others share diffuse boundaries, reflecting recent interchanges, generally  
95 limited by mountain or climatic barriers (Morrone, 2015; Ficetola et al., 2017). Morrone  
96 (2015) proposed that five major transition zones emerged from anterior works, which could  
97 be explained by a vicariance-dispersal model based on tectonic history. This synthetic model  
98 can be considered as a general framework to test for biogeographical regions.

99 However, the recent upsurge in continental bioregionalisation studies has concentrated  
100 exclusively on terrestrial vertebrates, which represent but a fraction of the continental  
101 organisms. There are other continental organisms with constraints to their dispersal and  
102 ecology that are fundamentally distinct from terrestrial vertebrates and for which one might  
103 expect distinct biogeographical regions. For example, terrestrial plants are generally  
104 characterised by higher degrees of endemism than animals, because they are more  
105 constrained than animals in terms of dispersal and tolerance to surmount climatic and other  
106 physical barriers (Lomolino et al., 2016). Hence, major phytogeographical regions were

107 described as manifold small regions (De Candolle, 1820, 1855; Takhtajan, 1986). However,  
108 Cox (2001) later proposed a handful of large floral regions comparable to biogeographical  
109 regions, thus suggesting that the major biogeographical regions are universal across the tree  
110 of life. A second example concerns human microbial diseases whose biogeography has also  
111 been shown recently to match terrestrial vertebrate biogeography (Murray et al., 2015).  
112 Another possibility concerns strictly freshwater organisms (i.e., organisms that live and  
113 disperse exclusively in freshwaters) as they have lower dispersal abilities than terrestrial  
114 vertebrates, and are geographically isolated in drainage basins usually flowing to the oceans.  
115 Terrestrial boundaries and salt waters represent strong barriers to dispersal, hence drainage  
116 basins have been considered as ‘island-like’ systems for strictly freshwater organisms (Rahel,  
117 2007; Hugueny, Oberdorff, & Tedesco, 2010; Tedesco et al., 2012; Dias et al., 2014).  
118 Dispersal can occur actively or passively via underground waters, stream captures,  
119 exceptional floods, glacier melting causing stream overflow, confluence during sea-level  
120 lowering, and displacement by other organisms or typhoons (see also discussion in  
121 Capobianco & Friedman, 2018). However, such dispersal events are rare, therefore  
122 immigration and speciation presumably occur on similar time-scales (Oberdorff et al., 2011).  
123 Consequently, one might expect that, because of peculiarities of riverscape changes through  
124 geological times, strictly freshwater organisms have been subject to different histories of  
125 diversification from those of terrestrial vertebrates (Rahel, 2007) and thus have original  
126 biogeographical boundaries. Because dispersal is physically constrained, a higher degree of  
127 provincialism and endemism could be anticipated for such organisms, resulting potentially in  
128 smaller and more numerous biogeographic regions.

129 In this paper, we focussed on the global biogeography of strictly freshwater actinopterygian  
130 fishes (i.e., excluding marine and amphidromous families of fish), hereafter called freshwater  
131 fishes. Several studies delineated biogeographical regions of freshwater fishes at regional to  
132 continental scales (e.g., Unmack 2001; Oikonomou *et al.* 2014), and studies conducted at the  
133 global scale also focussed on subregional provinces (ecoregions) based on a combination of  
134 data and expert decisions (Abell et al., 2008; Lévêque, Oberdorff, Paugy, Stiassny, & Tedesco,  
135 2008). Only one work hinted at nine potential freshwater fish biogeographic regions that  
136 covered the same biogeographical regions as terrestrial vertebrates (Matthews, 1998), but  
137 this work was based on a coarse geographic scale (52 approximate drainage basins for the  
138 whole world) and a low taxonomic resolution (family level). In addition, Matthews, (1998)  
139 included marine and diadromous fish families, which could conceal the effect of long-term  
140 isolation on freshwater fish endemism patterns. Consequently, whether Sclater-Wallace’s  
141 biogeographical regions are also applicable to freshwater fishes and to other freshwater  
142 organisms remains unresolved, and a global-scale quantitative bioregionalisation would  
143 represent an important step forward.

144 In this study, we aimed to define the major biogeographical regions for strictly freshwater  
145 fish species at the global scale. To delineate biogeographical regions, we capitalised on the  
146 recent development of a comprehensive dataset on freshwater fish distributions in drainage  
147 basins covering more than 80% of the Earth surface (Tedesco et al., 2017). First, we identified  
148 the large biogeographical regions of freshwater fishes using a recently developed hierarchical

149 approach based on networks (Vilhena & Antonelli, 2015), recommended for  
150 bioregionalisation studies (Edler et al., 2016; Bloomfield, Knerr, & Encinas-Viso, 2017; Rojas,  
151 Patarroyo, Mao, Bengtson, & Kowalewski, 2017). Then, we mapped the transition zones  
152 between regions and investigated species distributed across region boundaries. Finally, we  
153 compared biogeographical regions with terrestrial vertebrate biogeographical regions and  
154 discussed our findings in light of the synthetic biogeographical model proposed by Morrone  
155 (2015).

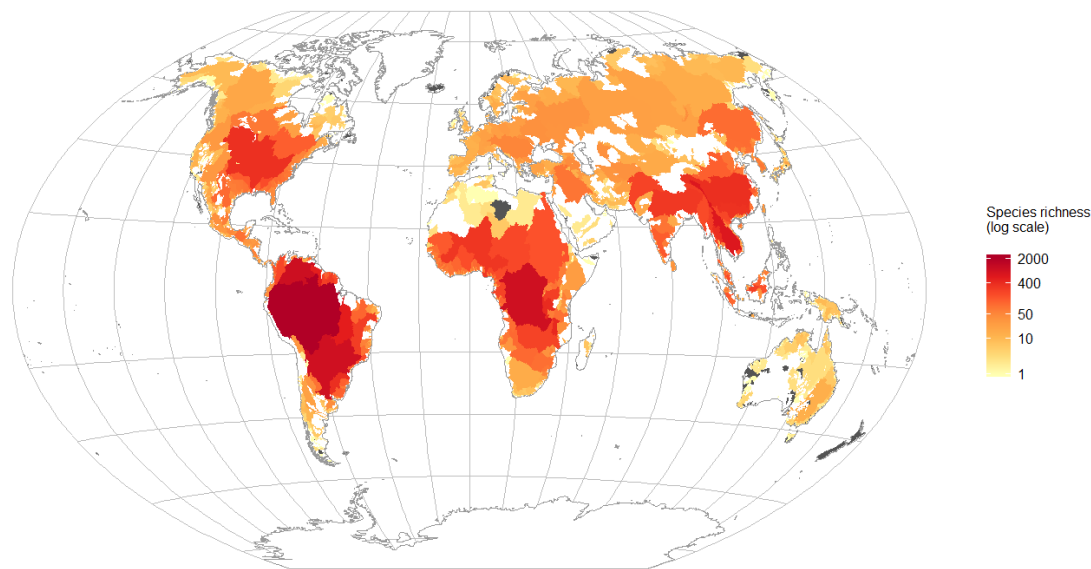
## 156 **Methods**

### 157 *Distribution data*

158 We based our bioregionalisation on the most comprehensive global database on freshwater  
159 fish species occurrence in drainage basins (Tedesco et al., 2017). This database comprises  
160 110 331 occurrence records for 14 953 species in 3 119 drainage basins of the world. Species  
161 names in the database were validated according to FishBase (Froese & Pauly, 2017) and the  
162 Catalogue of Fishes (Fricke, Eschmeyer, & van der Laan, 2017), and occurrence records were  
163 screened by the team developing the database (see details in Tedesco *et al.* 2017). We applied  
164 additional filters and corrections to the database. Since our aim was to describe the natural  
165 biogeographical regions resulting from long-term isolation of freshwater ichthyofaunas, we  
166 excluded documented records of introduced species, but included species considered to be  
167 recently extinct in their historical river basins. Additionally, to exclude most species that  
168 could disperse through marine waters, we retained only families having less than 10% of their  
169 species occurring in marine waters. This filter retained all “primary” and almost all  
170 “secondary” families of fishes (only Pseudomugilidae and Fundulidae were excluded), i.e.  
171 families with respectively no or limited salt tolerant species, as well as 22 families that had  
172 never been classified (based on Table 2 of Berra, 2007). It also included eight families with  
173 marine ancestors, seven of which had no species classified as tolerating salt water. Finally, we  
174 removed all diadromous species, according to FishBase. Additionally, we detected a few  
175 errors that were corrected in the database, mostly related to the native/introduced status for  
176 some species. The database used in this publication is available in Appendix S1 in Supporting  
177 Information. The resulting dataset included 59 373 records of 11 295 species in 2 581 basins  
178 (Figure 1).

179 To define our bioregions, we worked at the species level and used drainage basins as  
180 geographical units. Indeed, (1) in the absence of a unified phylogeny for actinopterygian  
181 fishes, species is the most standard unit available and (2) contrary to terrestrial vertebrates,  
182 (for which gridded distribution data of reliable quality are available), the most precise  
183 distribution data available for actinopterygian fishes is at the drainage basin unit. However, it  
184 is important to note that even if drainage basins are uneven in size, they are  
185 biogeographically meaningful for freshwater organisms because water bodies are generally  
186 connected within basins but not between basins (Hugueny et al., 2010).

187



188

189 **Figure 1.** Global distribution of freshwater fish species richness per drainage basin based on  
190 the global database on freshwater fish species occurrence in drainage basins (Tedesco et al.,  
191 2017). Grey-shaded areas correspond to basins without records of native strictly freshwater  
192 species.

### 193 *Delineation of biogeographical regions*

194 Until recently, the prevailing procedure for bioregionalisation has been based on hierarchical  
195 clustering methods applied to compositional dissimilarity (Kreft & Jetz, 2010; Procheş &  
196 Ramdhani, 2012; Holt et al., 2013). Since then, an approach based on biogeographical  
197 networks was introduced by Vilhena & Antonelli (2015), and has been recommended for  
198 delineating biogeographic regions (Edler et al., 2016; Bloomfield et al., 2017; Rojas et al.,  
199 2017). A network is composed of a series of *nodes* which can be connected to each other by  
200 *links* (or edges). In bioregionalisation, the network is composed of both sites (i.e., drainage  
201 basins here) and species, which constitutes a *bipartite* network. When a taxon is known to  
202 occur at a particular site, a link is drawn between the taxon and the site. A site cannot be  
203 connected to another site, and a taxon cannot be connected to another taxon. By definition,  
204 site-site and species-species links are not allowed in this type of analysis. Our final network  
205 had 13 876 nodes (11 295 species and 2 581 basins) and 59 373 links. We handled the network  
206 under Gephi 0.9.2, with the ForceAtlas2 algorithm. This software groups nodes that are  
207 tightly interconnected (such as groups of sites and species from the same biogeographical  
208 region) and separates groups of nodes that are not interconnected (distinct biogeographical  
209 regions). Such a graphical representation is useful for analysing and exploring the network.

210 We applied a community-detection algorithm to the entire network in order to group nodes  
211 into clusters (i.e. biogeographical regions). We applied the Map Equation algorithm  
212 ([www.mapequation.org](http://www.mapequation.org), Rosvall & Bergstrom, 2008) because it has been tested and



213 recommended to identify biogeographical regions (Vilhena & Antonelli, 2015; Edler et al.,  
214 2016; Rojas et al., 2017) and it features hierarchical clustering. Clusters are identified by the  
215 algorithm as having high intra-group but low inter-group connectivity, which corresponds  
216 well to the definition of biogeographical regions, i.e. regions of distinct assemblages of  
217 endemic taxa. We ran Map Equation (version Sat Oct 28 2017) with 100 trials to find the  
218 optimal clustering. We ran the hierarchical clustering (i.e., multi-level) in order to test  
219 whether larger regions have a nested hierarchy of subregions. It is important to note that a  
220 hierarchy of regions identified at the species level illustrates how biogeographical regions  
221 (i.e., distinct assemblages of endemic taxa) are currently spatially nested, but does not  
222 represent a historical (i.e., evolutionary) hierarchy of how these regions emerged.

223 The biogeographical network approach presents several advantages over distance-based  
224 approaches that were instrumental in our choice. Foremost, species identities are not lost,  
225 i.e., they are not abstracted into dissimilarity matrices between sites. Consequently, the  
226 network approach allows one to map how sites are connected by individual species, which  
227 presents an unquestionable asset to investigate between- and within- regions structures, such  
228 as potential dispersal pathways or barriers. A second practical novelty is that the algorithm  
229 assigns each species to a specific bioregion, which enables species-level descriptions (e.g. for  
230 online databases such as FishBase) and analyses. Lastly, the Map Equation algorithm is  
231 robust to differences in sampling intensities, making the removal of basins with low species  
232 richness unnecessary. On the other hand, distance-based approaches have limitations (see  
233 e.g., Leprieur & Oikonomou, 2014) and can produce inconsistent results when transforming  
234 such large occurrence datasets into a single dimension during the clustering procedure (see  
235 Appendix S2).

236 However, we provide clustering results using two additional methods in Appendix S2 for  
237 comparison: another network-based algorithm (Simulated Annealing, Bloomfield et al., 2017)  
238 and a distance-based method (following the framework of Kreft & Jetz, 2010).

### 239 *Sensitivity analysis*

240 We analysed the robustness of the identified regions by randomly extirpating a percentage of  
241 species (random value between 0.01 and 10.00% of the total number of species in the  
242 database) and re-running the whole bioregionalisation process. This process was repeated  
243 200 times. Then, for each region, we quantified the percentage of each region initial area that  
244 was retrieved in each simulation (Appendix S3).

### 245 *Transition zones and species shared between regions*

246 We calculated the participation coefficient (Guimerà & Amaral, 2005; Bloomfield et al., 2017)  
247 for each node of the biogeographical network. The participation coefficient indicates the  
248 degree to which a node is connected to different regions. A high participation coefficient for a  
249 given basin indicates that it contains species from different regions and can be assimilated to  
250 a transition zone between regions. A low participation coefficient indicates that all species in

251 the basin belong to the same region. The participation coefficient of a node is calculated as  
252 follows:

$$P_i = 1 - \sum_{s=1}^{N_m} \left( \frac{k_{is}}{k_i} \right)^2$$

253 where  $P_i$  is the participation coefficient of node  $i$ ,  $k_{is}$  is the number of links of node  $i$  to region  
254  $s$ ,  $k_i$  is the total number of links of node  $i$ , and  $N_m$  is the total number of regions. We  
255 calculated the participation coefficient at each level of the biogeographical structure  
256 identified by Map Equation.

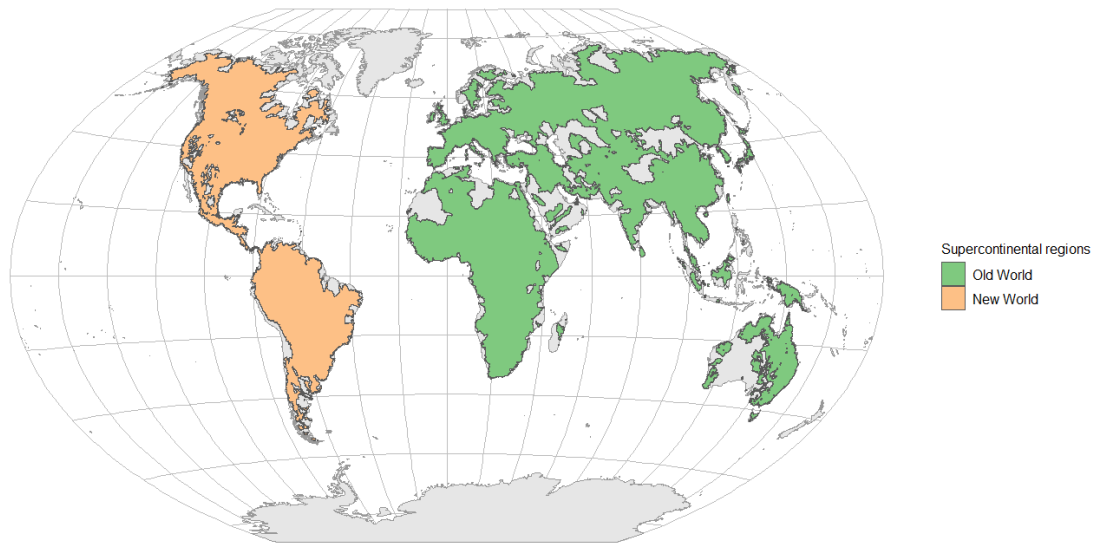
257 We also summarised the list of species that were shared between the major regions (i.e.,  
258 excluding tiny clusters) and their distribution characteristics.

259

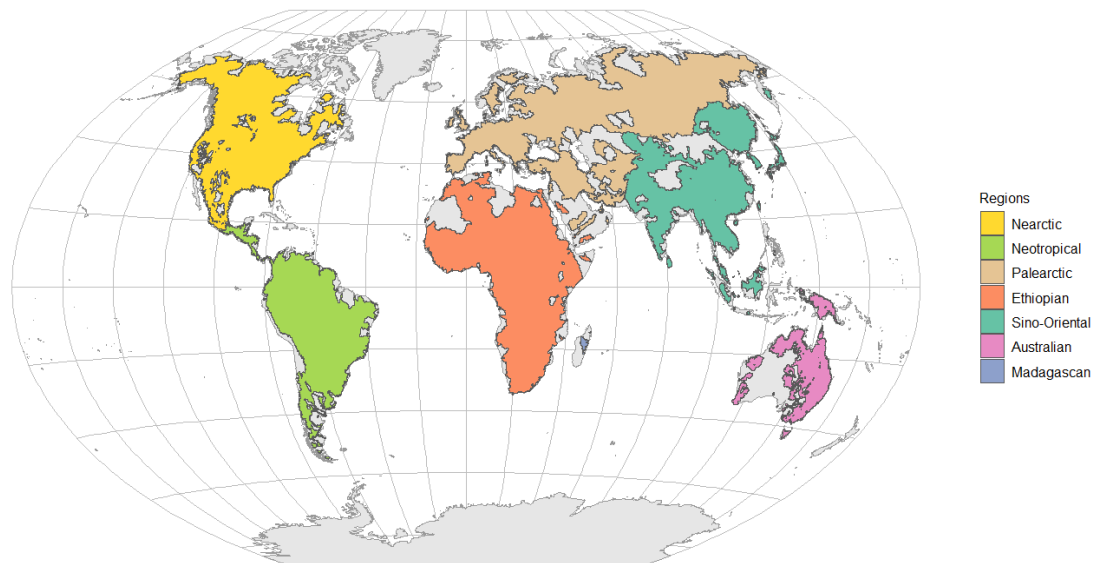
## 260 Results

261 The Map Equation algorithm identified a hierarchy of biogeographical regions with up to six  
262 nested levels. For this global-scale study, we investigated the first three levels, termed as  
263 supercontinental regions, Regions and Subregions.

a. Supercontinental regions



b. Regions



264

265 **Figure 2.** Biogeographical regions of freshwater fishes defined at the species level with  
266 the Map Equation clustering algorithm. We identified (a) two major supercontinental

267 regions (Old World and New World) and (b) six major regions (Nearctic, Palearctic,  
268 Neotropical, Ethiopian, Sino-Oriental, Australian) and a minor cluster (Madagascan).  
269 Some very small clusters of a few drainage basins that do not share species with any other  
270 basin were corrected based on expert interpretation and literature (see details in  
271 Appendix S5).

272

### 273 *Supercontinental regions*

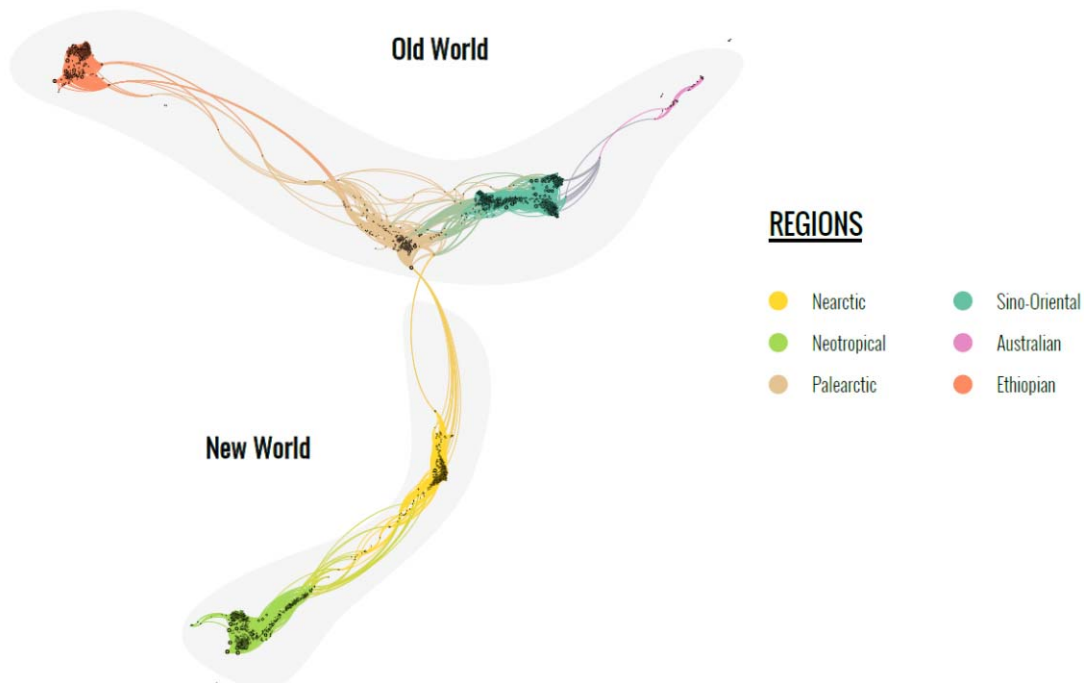
274 At the first level, we found that the world of freshwater fishes was divided into two  
275 supercontinental regions that we named New World (Americas) and Old World (Eurasia,  
276 Africa and Australian) (Figure 2a). Each supercontinental region contained nearly half the  
277 world's 11 295 species with virtually 100% of endemic species (Table 1). Only two species  
278 occurred in both supercontinental regions (Figure 3 and Appendix S4): (1) *Esox lucius*, an  
279 Old World species that occurred in 32 basins of northern New World, and (2) *Catostomus*  
280 *catostomus*, a New World species that occurred in nine basins of northern Old World. These  
281 two supercontinental regions hosted 99% of endemic genera and around 80% of endemic  
282 families. At this first level, we also found 14 tiny clusters of 49 basins (exclusively located in  
283 the Old World) without endemic families and genera but with endemic species (40 species in  
284 total). These tiny clusters were most often composed of species-poor basins located in remote  
285 islands (e.g., Madagascan) or isolated arid areas (e.g., Arabian Peninsula). Therefore, we  
286 *post-hoc* assigned these clusters to the Old World supercontinental region (see details in  
287 Appendix S5).

### 288 *Regions*

289 At the second level, we found six major regions spatially nested within the two  
290 supercontinental regions (Figure 2b), that we named following Morrone (2015). In the Old  
291 World supercontinental region, we found four regions and a minor cluster (Figure 2b). The  
292 richest one (Table 1) was the Ethiopian region with nearly 50% of Old World species,  
293 covering the entire African continent and including areas north from the Sahara and a few  
294 basins in the Arabic peninsula. The second richest one was the Sino-Oriental region which  
295 included south-eastern Asia from India to Borneo, most of China and Mongolia, Korea and  
296 Japan. The third one was the Palearctic region with less than 10% of Old World species,  
297 covering Europe, Central Asia (up to Pakistan and Kazakhstan) and Siberia. The fourth one,  
298 the poorest in species, was the Australian region, with only 80 species in total, covering  
299 Australia, Tasmania and Papouasia-New Guinea. Last, we identified Madagascan as a distinct  
300 minor cluster of the Old World, with 100% of endemic species and genera. Within the New  
301 World supercontinental region, we found two major regions. The first one was the  
302 Neotropical region, containing 85% of New World species and 42% of the world's known  
303 freshwater fish species (Table 1). The Neotropical region covered the whole of South America  
304 and Mesoamerica up to Southern Mexico. The second one was the Nearctic region, covering  
305 North America and northern Mexico. Finally, a tiny cluster composed of four basins of

306 Central America was identified at this level, which we *post-hoc* assigned to the Nearctic  
307 region (see details in Appendix S5).

308 Most regions had very high degrees of endemism (Table 1b), above 96% for all regions except  
309 the Palearctic (88%). These high degrees of endemism are apparent on the biogeographical  
310 network through a low number of links between regions (Figure 3, Appendix S6). In other  
311 words, all regions shared only a very limited number of species (see Appendix S4). The  
312 degree of endemism was lower for genera, ranging from 56% for the Palearctic to 98% for the  
313 Neotropical, and was much lower for families, with values below or equal to 50% for all  
314 regions except the Neotropical one.



315

316 **Figure 3.** Global biogeographical network of freshwater fishes. In this network, both  
317 species and drainage basins are represented as nodes. When a species is known to occur  
318 in a drainage basin, a link between the species and the basin is drawn. The network is very  
319 complex because of the high number of nodes (13 876 nodes corresponding to 11 295  
320 species and 2 581 basins) and links (59 373 occurrences). We spatialised the network in  
321 Gephi with the ForceAtlas 2 algorithm in order to group nodes that are strongly  
322 interconnected (i.e., basins that share species in common) and spread away from all other  
323 nodes that are not interconnected (i.e., basins that have few or no species in common).  
324 We coloured species and basin nodes according to the regions identified with the Map  
325 Equation algorithm and highlighted in grey the two supercontinental region (New World  
326 and Old World). To simplify the network, we masked here all nodes with less than ten  
327 links. A zoomable version of the full network with species and basin names is available in  
328 Appendix S6.

Table 1. Characteristics of the first two levels of biogeographical regions of freshwater fishes identified with the Map Equation algorithm.

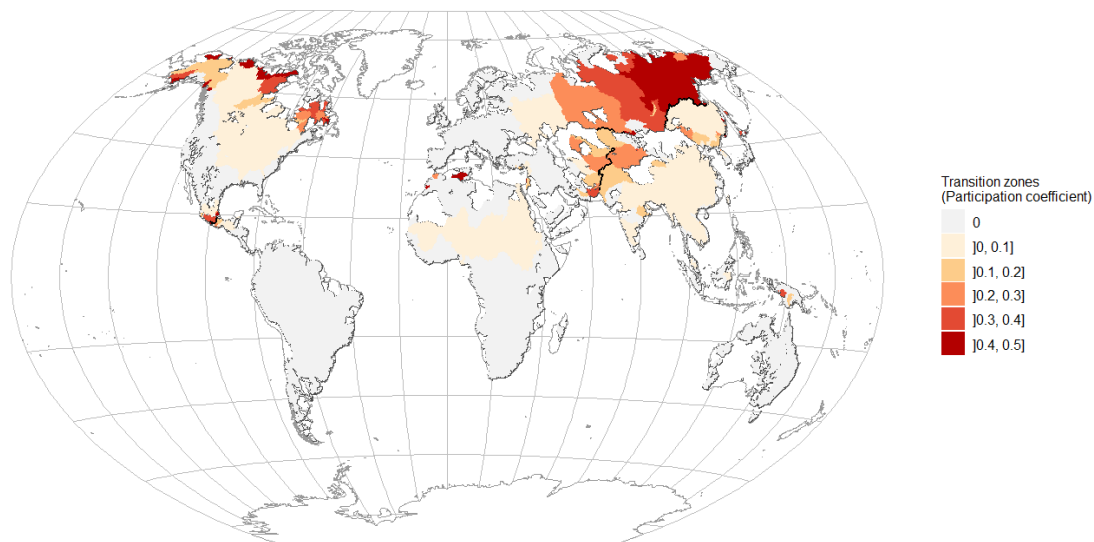
	Total number of species	Percentage of species of the world	Number of species assigned to the region	Number of endemic species	Percentage of endemic species	Number of endemic genera	Percentage of endemic genera	Number of endemic families	Percentage of endemic families	Number of basins	Total basin area (km <sup>2</sup> )
<b>a. Supercontinental regions</b>											
Old World	5719	50.63%	5718	5717	99.97%	857	99.30%	68	83.95%	1935	68 054 325
New World	5578	49.38%	5577	5576	99.96%	812	99.27%	54	80.60%	646	30 325 444
<b>b. Regions</b>											
Sino-Oriental	2495	22.09%	2466	2445	98.00%	399	88.86%	24	48.98%	720	14 963 465
Ethiopian	2683	23.75%	2678	2674	99.66%	313	94.56%	17	50.00%	275	24 776 845
Palaearctic	492	4.36%	468	434	88.21%	65	55.56%	4	19.05%	535	23 480 962
Australian	90	0.80%	88	87	96.67%	16	76.19%	4	40.00%	397	4 723 353
Neotropical	4761	42.15%	4750	4736	99.47%	680	97.70%	41	75.93%	379	15 424 216
Nearctic	842	7.45%	827	815	96.79%	117	85.40%	9	40.91%	267	14 901 227
<i>Madagascan</i>	18	0.16%	18	18	100.00%	9	100.00%	2	66.67%	8	109 699

329 Interestingly, our clustering results are concordant with both Simulated Annealing (Figure  
330 S2.1 in Appendix S2) and beta-diversity methods (Figure S2.2). The only notable  
331 discrepancies concerned the Nearctic region which was split in four regions with the beta-  
332 diversity approach, and a portion of the Central Asian part of the Sino-Oriental which was  
333 attributed to the Palearctic by the Simulated Annealing method (see Appendix S2). Likewise,  
334 our sensitivity analysis on the Map Equation algorithm confirmed that regions were stable to  
335 random extirpations of species (Appendix S3), except for the Sino-Oriental one, which split in  
336 two regions in half of the simulations. This split produced the Oriental and Sinean regions  
337 with respectively ~85% and ~72% of endemics. These two subregions are visible on the  
338 network with two apparent clusters of nodes in the Sino-Oriental cluster (Figure 3). We also  
339 observed other minor changes, such as some clusters of basins appearing as small distinct  
340 regions, e.g. in Central Asia.

### 341 *Subregions*

342 At the third level, we observed different patterns among regions. Three regions (Sino-  
343 Oriental, Nearctic and Australian) had only two to three main subregions (Appendix S7) that  
344 were spatially coherent and had high degrees of endemism (68.7-91.5% of endemic species,  
345 Appendix S7). The other three regions (Ethiopian, Palearctic and Neotropical) were  
346 characterised by a high number of subregions, which were also generally spatially coherent  
347 (Appendix S7). The Ethiopian and Neotropical subregions were generally characterised by a  
348 high number of species and endemics (Figures S7.9 and S7.14). The Palearctic subregions  
349 were characterised by a low number of species and generally low endemism (Figure S7.11).

350



351

352 **Figure 4.** Map of transition zones between regions (level 2) characterised by the  
353 participation coefficient of basins, *i.e.* the proportion of species in a basin that come from  
354 regions other than the region of this basin. Thick black lines indicate frontiers, which are  
355 shared between two neighbouring regions.

356

### 357 *Transition zones and species shared between regions*

358 At the supercontinental region level, we obtained participation coefficients of basins between  
359 0.0 and 0.5, with transition zones (*i.e.*, basins with high participation coefficients) located in  
360 north-eastern Siberia as well as in northern North America. At the regional level, we observed  
361 participation coefficients also ranging from 0.0 to 0.5 (Figure 4). Unsurprisingly, we found  
362 the same transition zones between the Nearctic and the Palearctic as for supercontinental  
363 regions. However, the major transition zones were located at the boundaries between the  
364 Palearctic and Sino-Oriental regions: the high participation coefficients of basins at their  
365 boundaries indicated that these boundaries were diffuse (Figure 4). These diffuse boundaries  
366 are reflected on the network by the high number of links between multiple species and  
367 multiple basins from both regions (Figure 3, Appendix S5). We identified another transition  
368 zone between the Nearctic and the Neotropical regions, with high participation coefficients of  
369 basins at their boundary (Figure 4). A dozen species of each region also incurred in the other  
370 region with similar distribution patterns (Appendix S4).

371 Overall, the species that were distributed across boundaries could be separated in two broad  
372 categories (Appendix S4). First, we found that the majority of shared species had restricted  
373 distributions close to regional boundaries, with occasional occurrences beyond. For example,  
374 the two-spot livebearer *Heterandria bimaculata* has a distribution endemic to Central  
375 America at the northernmost part of the Neotropical region and incurred in two basins of the  
376 Nearctic. Second, we found a limited number of species with large spatial distributions that  
377 were able to incur in multiple basins of other regions. The best example is the Northern pike  
378 *Esox lucius* which is one of the two species distributed across both supercontinental regions.  
379 Another example is the Eurasian minnow *Phoxinus phoxinus* that is widespread in the  
380 Palearctic with multiple occurrences in the Sino-Oriental region.

381 On the other hand, we found almost no transition zones at the boundaries of the Ethiopian or  
382 the Australian regions. For the Ethiopian region, only a few basins in northern Africa and the  
383 Middle East shared species between the Ethiopian and Palearctic regions (Figure 4) as  
384 illustrated by the limited number of links on the network (Figure 3). Only nine species were  
385 shared between these two regions (Appendix S4). For the Australian region, only three  
386 species dispersed through the Wallace-Huxley boundaries (Figure 3, Appendices S4 and S6):  
387 two species assigned to the Oriental region (*Aplocheilichthys panchax* and *Datnioides polota*)  
388 occur in Papuasias, and one species assigned to the Australian region (*Ophistermon*  
389 *bengalense*) was distributed in Australia, Papuasias and Southeastern Asia.

390



## 391 **Discussion**

392 Here, we provide the first global bioregionalisation of freshwater fishes based on a  
393 quantitative analysis of species distributions. We found that the freshwater fish world is first  
394 divided into two supercontinental regions, the Old World and New World. Nested within  
395 these two supercontinental regions we found six major biogeographical regions with very  
396 high degrees of endemism: the Old World was divided in the Sino-Oriental, Ethiopian,  
397 Palearctic and Australian regions, whereas the New World was composed of the Neotropical  
398 and Nearctic regions. Nested within each of these major biogeographical regions we found  
399 subregions with varying degrees of endemism. Some regions had a few large subregions,  
400 while others had numerous subregions of heterogeneous size, species diversity and  
401 endemism. We found transition zones between the Nearctic and Neotropical regions, and  
402 between the Palearctic and Sino-Oriental regions. Furthermore, we found the Ethiopian and  
403 Australian regions to have almost no transition zones with other regions. The fish species that  
404 were distributed across regions were of two types: (i) species with ranges that were restricted  
405 to the proximity of the regional boundaries with occasional overlaps across boundaries and  
406 (ii) species with large distributions that expanded their range beyond boundaries.

407 Our results compellingly contradict our initial hypothesis that freshwater fish may have  
408 biogeographical regions different from the terrestrial vertebrate scheme of Sclater-Wallace  
409 because of their restricted dispersal abilities and the specific spatial-temporal dynamics of  
410 riverscapes. Indeed, we found a total of six major biogeographical regions that were similar in  
411 size and location to the six biogeographical regions identified by Wallace (1876). The only  
412 differences were the locations of several boundaries. More surprisingly, our regions were also  
413 similar to the coarse biogeographical regions based only on freshwater fish families identified  
414 by Matthews (1998) and that also included diadromous fish species. Our Neotropical region  
415 is the only major difference from results obtained by Matthews (1998) who identified a  
416 distinct cluster south to the Andes. This difference may be explained by the fact that this area  
417 was mostly colonised by the family of Galaxiidae that migrates between freshwater and  
418 oceans, and consequently have not been included in our analysis.

419 In addition, we also identified transition zones broadly following the bioregionalisation  
420 model of Morrone (2015), suggesting that freshwater fish regions were shaped by vicariance  
421 and geodispersal events similar to other groups. However, we observed extremely high rates  
422 of endemism for each region, more than 96% of endemic species for all regions except the  
423 Palearctic (89%). These endemism rates far exceed the endemism rates for other continental  
424 vertebrates such as birds (11 to 84% endemism in major regions, see Appendix S8), mammals  
425 (31% to 90%), herptiles (amphibian and reptiles: 46 to 95%; amphibian only: 66 to 98%), as  
426 calculated for regions of Procheş & Ramdhani (2012, Table S8.1 in Appendix S8) and Holt et  
427 al. (2013, Table S8.2 in Appendix S8). Freshwater fish endemism rates also exceed rates  
428 reported in marine biogeographical realms (17 to 84%, Costello et al., 2017). Therefore, we  
429 conclude here that freshwater fishes are likely to have among the highest rates of species  
430 endemism for major biogeographical regions.

431 The two major supercontinental regions we identified might seem to contradict Morrone's  
432 biogeographical kingdoms (2015). Indeed, Morrone (2015) hypothesised that three major  
433 kingdoms could be derived from formerly disconnected land masses (i.e., Holarctic,  
434 Holotropical and Austral). This apparent contradiction is explained by our analyses at the  
435 species level which was not designed to reflect ancient biogeographical kingdoms resulting  
436 from the Gondwana split. Rather, the two supercontinental regions described here suggest  
437 that (1) the six major biogeographical regions have merged into two super-continents with  
438 biotic exchanges and (2) that strictly freshwater faunas have been isolated between these two  
439 super-continents for the past 10 to 20 million years. Only two species are shared between  
440 supercontinental regions, and genetic studies indicate that the only species documented, the  
441 pike *Esox lucius*, likely expanded its range after recent glaciations (Maes, van Houdt, De  
442 Charleroy, & Vockaert, 2003; Jacobsen, Hansen, & Loeschke, 2005). This current pattern of  
443 isolation of the two supercontinental regions will have far-reaching implications for the  
444 future evolution of freshwater fish faunas. The imprint of this isolation on fish faunas will be  
445 visible for millions of years, likewise to the historical imprint of the split of Pangea that can  
446 be seen today on phylogenetically-informed bioregionalisations (or bioregionalisations at a  
447 higher taxonomic level such as families; e.g. Morrone et al. 2015). This pattern is, as far as we  
448 know, specific to strictly freshwater fishes. We can expect a different pattern for diadromous  
449 fishes which can disperse through marine waters and thus occur in both the Nearctic and the  
450 Palearctic (e.g., species in genera *Salmo*, *Salvelinus*, *Cottus*, *Lotta*, *Pungitius*, *Gasterosteus*).

451 Therefore, our hierarchy of supercontinental regions, regions and subregions should not be  
452 seen as the hierarchy of freshwater fish diversification history (i.e., our species-level  
453 supercontinental regions do not bear the same meaning as kingdoms in Morrone et al. 2015),  
454 but rather as the *recent* biogeography of strictly freshwater fish faunas. This recent  
455 biogeography illustrates spatially-nested supercontinental regions, regions and subregions all  
456 characterised by very high degrees of endemism. These very high degrees of endemism reflect  
457 the strong isolation of strictly freshwater fish faunas because of the lack of dispersal during  
458 the Neogene (caused by the isolation of the different land masses during the late Cretaceous  
459 to Paleogene) and the limited magnitude of fish dispersal events that occurred after land  
460 masses merged during the late Miocene to early Pleistocene. Only 86 (0.76%) of the world's  
461 11 295 species can be found across supercontinental or regional boundaries. Therefore,  
462 99.24% of the world's freshwater fishes occur in a single supercontinental region and a single  
463 region, which ascertains the biological reality of our delineated clusters.

#### 464 *Regions and transition zones*

465 The Neotropical and Nearctic regions of freshwater fishes are very similar to the Neotropical  
466 and Nearctic regions highlighted for terrestrial mammals, amphibians and bird species (Kreft  
467 & Jetz, 2010; Procheş & Ramdhani, 2012; Holt et al., 2013). These two regions are therefore  
468 in agreement with the synthetic biogeographical regionalisation of Morrone (2015). As for  
469 other groups of organisms, we identified a Mexican transition zone between Nearctic and  
470 Neotropical regions, suggesting that these organisms were affected by biotic interchange  
471 between the Americas, but to a much lesser extent than terrestrial vertebrates (Bussing, 1985;

472 Smith & Bermingham, 2005). Indeed, we found that only 25 species with restricted  
473 distributions were shared between Nearctic and Neotropical regions, as illustrated by the  
474 restricted area of the transition zone (Figure 4). These species belong to both primary and  
475 secondary freshwater families, which colonised Mesoamerica separately, as suggested by  
476 molecular analyses (Smith & Bermingham, 2005). Secondary freshwater fishes probably  
477 dispersed through Mesoamerica before the formation of the Panama isthmus, during periods  
478 of high runoff leading to temporary freshwater or brackish-water bridges in marine waters  
479 coupled with northward discharge of the proto-Amazon during the Miocene ~18-15 Ma  
480 (Smith & Bermingham, 2005; Hoorn et al., 2010). Later, primary freshwater families  
481 dispersed through Mesoamerica during the Isthmus formation via landscape diffusion (Smith  
482 & Bermingham, 2005). Subsequent changes in the landscape led to increasing isolation of  
483 basins within Mesoamerica as well as occasional connectivity events, which in turn shaped  
484 the regional patterns of dispersal and diversification (Smith & Bermingham, 2005; Dias et  
485 al., 2014) that probably drove the restricted extent of the Mexican transition zone for  
486 freshwater fishes.

487 The Ethiopian region of freshwater fish resembles the Ethiopian regions of terrestrial  
488 vertebrates except for its northern limit. We found that this region expanded beyond the  
489 limits of the Sahara up to the Mediterranean sea, similarly to flightless terrestrial mammals  
490 (Kreft & Jetz, 2010). However, flightless terrestrial mammals were also found to expand their  
491 Ethiopian boundary beyond the Arabian Peninsula into Central Asia, which was not the case  
492 here. Moreover, most studies for terrestrial vertebrates located the northern boundary of the  
493 Ethiopian region south of the Sahara (Procheş & Ramdhani, 2012; Rueda et al., 2013; Holt et  
494 al., 2013; Vilhena & Antonelli, 2015). This discrepancy may be explained by several factors.  
495 First, most studies on vertebrates identified the Sahara and the northern coast of Africa as  
496 transition zones between the Palearctic and Ethiopian regions (Holt et al., 2013; Morrone,  
497 2015). This transition zone was not identified in our analyses: Palearctic fishes only  
498 anecdotally crossed the Mediterranean Sea, while their African counterparts merely ventured  
499 into the Middle-East. Therefore, we can hypothesise that the Mediterranean Sea was an  
500 insurmountable barrier for freshwater fishes, contrary to terrestrial vertebrates that could  
501 disperse through Straits of Gibraltar. Dispersal during the desiccation of the Mediterranean  
502 Sea during the Messinian salinity crisis about 6 Ma has been proposed with the Lago Mare  
503 hypothesis (Bianco, 1990). This hypothesis stated that, during the refilling of the  
504 Mediterranean Sea, a freshwater or brackish phase occurred which would have permitted  
505 large scale dispersal of freshwater fishes across the Mediterranean basin. However this  
506 hypothesis received limited support in the light of phylogenetic studies (Levy, Doadrio, &  
507 Almada, 2009; Perea et al., 2010). Hence, freshwater fish dispersal during the Messinian  
508 salinity was not the same as that of other groups (e.g., Veith *et al.* 2004; Agustí *et al.* 2006).  
509 A probable dispersal pathway for freshwater fishes was through diffusion (*sensu* Lomolino et  
510 al., 2016) between the Nile and Mediterranean basins of the Middle East (illustrated by the  
511 distributions of *Clarias gariepinus* and four African cichlid species). This Middle-East  
512 dispersal pathway corroborates to some extent the Saharo-Arabian transition zone proposed  
513 by Morrone (2015). This dispersal pathway was identified for the early colonisation of

514 Cyprinidae from the Sino-Oriental to Ethiopian region (Gaubert, Denys, & Oberdorff, 2009).  
515 Second, drainage basins north of the Sahara are extremely poor in freshwater fish species,  
516 notably because of their arid nature. Northernmost basins were probably colonised by  
517 Ethiopian species during repeated periods of aquatic connectivity across the Sahara region,  
518 the most recent being the Holocene African Humid period (~11 to 8ka) (Drake, Blench,  
519 Armitage, Bristow, & White, 2011).

520 The Sino-Oriental region is a distinctive feature of freshwater fishes, for two reasons. Firstly,  
521 the northern boundary between Sino-Oriental and Palearctic is located beyond the  
522 Himalayas and the Gobi desert to the North-West, and goes up to the Stanovoï mountains  
523 located north of China, whereas all other groups are limited to a boundary extending  
524 longitudinally from the Himalayas to the China Sea (Cox, 2001; Procheş & Ramdhani, 2012;  
525 Rueda et al., 2013; Holt et al., 2013; Morrone, 2015). Secondly, this region is composed of  
526 two major subregions that were frequently identified as distinct regions in our sensitivity  
527 analysis. In particular, the Sinean subregion hosting a rich freshwater fish fauna (963  
528 species) with 72% of endemics seems unique to freshwater fishes, and has been identified  
529 more as a transition zone for other groups (Procheş & Ramdhani, 2012; Holt et al., 2013;  
530 Morrone, 2015). We can speculate that this characteristic Sino-Oriental region for fishes  
531 arose from several factors. Firstly, the entire region was prone to fish speciation since the  
532 Eocene (55My), probably because of the very high diversity of aquatic habitats combined with  
533 the repeated rearrangement of rivers through capture and glaciation melting (Dias et al.,  
534 2014; Kang et al., 2014; Kang, Huang, & Wu, 2017; Xing, Zhang, Fan, & Zhao, 2016). For  
535 example, the Cyprinidae family, which accounts for 45% of Sino-Oriental species, originated  
536 from the Indo-Malaysian tropical region and has likely radiated into Asia since the Eocene  
537 (Gaubert et al., 2009). Secondly, the northern boundary is located farther North than from  
538 other groups, suggesting that mountain barriers were more important in defining boundaries  
539 for fishes than for other groups, whose boundaries appeared to be rather defined by a  
540 combination of tectonics and climate (Ficetola et al., 2017). Consequently, the fish transition  
541 zone between Sino-Oriental and Palearctic is not located near areas of recent tectonic  
542 merging as reported for other groups (Morrone, 2015). In addition, this transition zone is  
543 asymmetrically distributed towards the Palearctic (Figure 4) and possibly exceeding the  
544 asymmetry reported for other groups (Sanmartín, Enghoff, & Ronquist, 2001), probably  
545 because of the extreme differences in fish richness between both regions. Given the  
546 mountainous nature of the boundary, dispersal pathways probably emerged at river  
547 confluences when the sea level dropped (Dias et al., 2014), both at the Northeastern and  
548 Southwestern parts of the Sino-Oriental boundaries (Gaubert et al., 2009).

549 The Australian region is the most depauperate of all fish regions, probably owing to the  
550 combination of the complete isolation of this region for the last 60 million years (Scotese,  
551 2016) and the dryness of the Australian continent. The boundary between Australian and  
552 Sino-Oriental encompasses the entire Wallacea: the Sino-Oriental extends to the limits of the  
553 Sunda shelf, whereas the Australian extends to the limits of the Sahul shelf (Lomolino et al.,  
554 2016). Only three species occur on both sides of the boundary. Two Sino-Oriental species  
555 (*Aplocheilichthys panchax* and *Datnioides polota*) are Sino-Oriental species known from one

556 drainage basin of the Australian region, and one species (*Ophisternon bengalense*) is  
557 distributed in 10 and 11 basins of Australian and Sino-Oriental respectively. *O. bengalense*  
558 lives in estuaries and is tolerant to brackish waters, and thus may have dispersed through  
559 Wallacea (e.g., Capobianco & Friedman, 2018). On the other hand, the distributions of the  
560 other two species are more surprising, and the absence of data in all islands of Wallacea  
561 makes hypotheses highly speculative. These apparent disjoint distributions could be linked to  
562 unknown exceptional dispersal, undocumented species introductions, or misidentification  
563 (i.e. two species misidentified as a single species).

#### 564 *Subregions*

565 At subregional and finer levels, we found multiple clusters with varying degrees of endemism,  
566 with a substantial number of species distributed between clusters (see Figure 3). While it  
567 indicates that strictly freshwater fish species display strong endemism patterns at  
568 subregional spatial scales, it also suggests a reticulated history of river basins. The differences  
569 in number and endemism of subregions among major regions may be explained by the  
570 combination of habitat size and diversity, past climate change, and paleoconnectivity during  
571 the Last Glacial Maximum (LGM, see Leprieur *et al.* 2011; Tedesco *et al.* 2012; Dias *et al.*  
572 2014). Past climate change had an enormous impact on high-latitude regions, such as the  
573 Nearctic and northern parts of the Palearctic (Leprieur *et al.*, 2011). Most of these areas were  
574 covered by ice sheets during the LGM. These Northern areas were colonised by species after  
575 the LGM (e.g., Rempel & Smith 1998) from refuges located in the southernmost parts of these  
576 regions (Mississippi basin for the Nearctic, Danube basin for the Palearctic). Consequently, the  
577 relatively recent re-colonisation explains these large species-poor subregions. On the other  
578 hand, such climate events were less extreme in tropical regions thereby allowing lineages to  
579 thrive for a long period (Tedesco, Oberdorff, Lasso, Zapata, & Hugueny, 2005). The  
580 combination of this prosperity with the high diversity and size of habitats in tropical regions,  
581 the long-term isolation of drainage basins during the Pleistocene (Dias *et al.*, 2014) as well as  
582 stable climatic history and favourable climatic conditions (Wright, Ross, Keeling, McBride, &  
583 Gillman, 2011) probably generated conditions favourable to divergence and radiation  
584 processes in tropical subregions. This last hypothesis may explain the numerous tropical  
585 subregions with high diversity and endemism we found. The only apparent contradiction  
586 could arise from the two large subregions with high endemism of the Sino-Oriental. However,  
587 these two subregions probably reflect the uplift of the Tibetan plateau that led to their  
588 isolation (Kang *et al.*, 2014). In turn, these two subregions included numerous smaller ones  
589 with high diversity and endemism (see Figure S7.15) similar to the other tropical regions.

#### 590 *Robustness of findings*

591 This first global quantitative analysis of the biogeography of freshwater fishes is based on a  
592 large scale database compiling occurrence data from thousands of sources and is thus  
593 inevitably subject to errors and incomplete data (Tedesco *et al.*, 2017). To minimise errors, a  
594 careful screening and correction procedure has been implemented for this database (see  
595 Tedesco *et al.*, 2017). Reassuringly, the results obtained from other clustering methods as

596 well as our sensitivity analysis suggested that the regions we identified are robust.  
597 Furthermore, regions were all spatially coherent (even though no spatial information was  
598 provided at any stage of the process) for the first two levels, with high degrees of endemism,  
599 indicating the quality of both the dataset and the bioregionalisation approach.

600 The network method assigns clusters to species (as explained in the methods), which is an  
601 asset over distance-based clusters. However, one major caveat needs to be acknowledged.  
602 Species are assigned to the region where their present-day distribution is largest – this region  
603 is not necessarily the region where they originated from. A perfect example is the Characidae  
604 family at the transition between Nearctic and Neotropical: *Astyanax mexicanus* was assigned  
605 to the Nearctic, although its lineage is assumed to have colonised Mesoamerica during the  
606 Panama Isthmus formation (Smith & Bermingham, 2005).

607 Our results at the subregional scales have several limits. Firstly, while species introductions  
608 are relatively well documented between major continents or regions, we can expect that some  
609 human-assisted translocations of species among basins have not been documented at smaller  
610 spatial scales, thereby blurring subregional patterns of endemism. Secondly, heterogeneity in  
611 land topology led to vast differences in size and number of drainage basins within different  
612 geographical areas of the world. Furthermore, drainage basins (especially large ones) may  
613 have a reticulated history challenging their validity as biogeographical units at sub-regional  
614 levels (e.g., see Dagosta & Pinna, 2017 and references therein). Third, for areas with  
615 numerous small basins, species lists were not necessarily available for all of them, and thus  
616 identification of provinces beyond the sub-regional scale would be speculative. Likewise, fine-  
617 scale data was not available in similar quantity or quality within different regions of the world  
618 (e.g., remote areas of Africa or Papuasias remain poorly sampled compared to other regions).  
619 Last, the Map Equation is expected to identify transition zones as distinct clusters (Vilhena &  
620 Antonelli, 2015; Bloomfield et al., 2017). We did not observe this pattern at large scales,  
621 except for a few basins at the transitions between Nearctic and Neotropical regions or  
622 between Sino-Oriental and Palearctic or Australian regions. However, at sub-regional and  
623 finer scales, transitions are likely to stand out as separate zones (Bloomfield et al., 2017),  
624 which may not necessarily be an appealing property since the participation coefficient is  
625 informative enough to describe transition zones. For all these reasons, we deemed preferable  
626 not to investigate our results below the third level, as such fine-scale provinces would be  
627 better studied in regional studies (e.g., Smith & Bermingham, 2005; Kang et al., 2014).

## 628 *Concluding remarks*

629 This first quantitative study of freshwater fish bioregions revealed that their biogeography  
630 was probably shaped by the same major drivers as other continental groups of organisms,  
631 with peculiar exceptions such as the Sino-Oriental region. These regions identified with  
632 species distributions probably reflect relatively recent processes of dispersal and isolation.  
633 Ancient processes will be explored in future studies thanks to the newly available dated  
634 phylogenies of actinopterygian fishes (Rabosky et al., 2018).

635 We found that freshwater fishes, in addition to being the most diverse group of continental  
636 vertebrates, have extremely high rates of endemism, above 96% for all regions except the  
637 Palearctic. Furthermore, we found that tropical regions have a myriad of subregions with  
638 very high endemism and richness. These figures compellingly bespeak that freshwater fishes  
639 ought to be considered in hotspot analyses and raise many questions about the  
640 biogeographical consequences of the current high rates of freshwater fish introductions and  
641 extirpations (Villéger, Blanchet, Beauchard, Oberdorff, & Brosse, 2015).

642

- 643 Abell, R., Thieme, M. L., Revenga, C., Bryer, M., Kottelat, M., Bogutskaya, N., ... Petry, P.  
644 (2008). Freshwater Ecoregions of the World: A New Map of Biogeographic Units for  
645 Freshwater Biodiversity Conservation. *BioScience*, 58(5), 403.
- 646 Agustí, J., Garcés, M., & Krijgsman, W. (2006). Evidence for African-Iberian exchanges  
647 during the Messinian in the Spanish mammalian record. *Palaeogeography*,  
648 *Palaeoclimatology, Palaeoecology*, 238(1–4), 5–14.
- 649 Berra, T. M. (2007). *Freshwater fish distribution*.
- 650 Bianco, P. G. (1990). Potential role of the palaeohistory of the Mediterranean and Paratethys  
651 basins on the early dispersal of Euro - Mediterranean freshwater fishes. *Ichthyological*  
652 *Exploration of Freshwaters*, 1(6), 167–184.
- 653 Bloomfield, N. J., Knerr, N., & Encinas-Viso, F. (2017). A comparison of network and  
654 clustering methods to detect biogeographical regions. *Ecography*, (November 2016), 1–  
655 10.
- 656 Brosse, S., Beauchard, O., Blanchet, S., Dürr, H. H., Grenouillet, G., Hugueny, B., ...  
657 Oberdorff, T. (2013). Fish-SPRICH: a database of freshwater fish species richness  
658 throughout the World. *Hydrobiologia*, 700(1), 343–349.
- 659 Bussing, W. (1985). Patterns of distribution of the Central American Ichthyofauna. In G. G.  
660 Stehli & S. D. Webb (Eds.), *The Great American Biotic Interchange* (pp. 453–473). New  
661 York: Plenum Press.
- 662 Capobianco, A., & Friedman, M. (2018). Vicariance and dispersal in southern hemisphere  
663 freshwater fish clades: a palaeontological perspective. *Biological Reviews*.
- 664 Costello, M. J., Tsai, P., Wong, P. S., Kwok, A., Cheung, L., & Basher, Z. (2017). Marine  
665 biogeographic realms and species endemism. *Nature Communications*, 25.
- 666 Cox, B. (2001). The biogeographic regions reconsidered. *Journal of Biogeography*, 28, 511–  
667 523.
- 668 Dagosta, F. C. P., & Pinna, M. de. (2017). Biogeography of Amazonian fishes: deconstructing  
669 river basins as biogeographic units. *Neotropical Ichthyology*, 15(3), 1–24.
- 670 De Candolle, A. P. (1820). *Essai Élémentaire de Géographie Botanique*. F. Levrault.
- 671 De Candolle, A. P. (1855). *Géographie botanique raisonnée*. Paris: Librairie de Victor  
672 Masson.
- 673 Dias, M. S., Oberdorff, T., Hugueny, B., Leprieur, F., Jézéquel, C., Cornu, J. F., ... Tedesco, P.  
674 A. (2014). Global imprint of historical connectivity on freshwater fish biodiversity.  
675 *Ecology Letters*, 17(9), 1130–1140.
- 676 Drake, N. A., Blench, R. M., Armitage, S. J., Bristow, C. S., & White, K. H. (2011). Ancient  
677 watercourses and biogeography of the Sahara explain the peopling of the desert.  
678 *Proceedings of the National Academy of Sciences*, 108(2), 458–462.
- 679 Edler, D., Guedes, T., Zizka, A., Rosvall, M., & Antonelli, A. (2016). Infomap Bioregions:  
680 Interactive Mapping of Biogeographical Regions from Species Distributions. *Systematic*  
681 *Biology*, 66(2), syw087.
- 682 Ficetola, G. F., Mazel, F., & Thuiller, W. (2017). Global determinants of zoogeographical  
683 boundaries. *Nature Ecology & Evolution*, 1(March), 0089.
- 684 Fricke, R., Eschmeyer, W., & van der Laan, R. (2017). CATALOG OF FISHES: GENERA,  
685 SPECIES, REFERENCES. Retrieved from  
686 <http://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp>



- 687 Froese, R., & Pauly, D. (2017). FishBase.
- 688 Gaubert, P., Denys, G., & Oberdorff, T. (2009). Genus-level supertree of Cyprinidae  
689 (Actinopterygii: Cypriniformes), partitioned qualitative clade support and test of macro-  
690 evolutionary scenarios. *Biological Reviews*, 84(4), 653–689.
- 691 Guimerà, R., & Amaral, L. A. N. (2005). Functional cartography of complex metabolic  
692 networks. *Nature*, 433, 895–900.
- 693 Holt, B. G., Lessard, J.-P., Borregaard, M. K., Fritz, S. A., Araújo, M. B., Dimitrov, D., ...  
694 Rahbek, C. (2013). An update of Wallace's zoogeographic regions of the world. *Science*,  
695 339(6115), 74–78.
- 696 Hoorn, C., Wesselingh, F. P., Ter Steege, H., Bermudez, M. A., Mora, A., Sevink, J., ...  
697 Antonelli, A. (2010). Amazonia through time: Andean uplift, climate change, landscape  
698 evolution, and biodiversity. *Science*, 330(6006), 927–931.
- 699 Hugueny, B., Oberdorff, T., & Tedesco, P. A. (2010). Community Ecology of River Fishes: A  
700 Large-Scale Perspective. In K. B. Gido & D. A. Jackson (Eds.), *Community Ecology of*  
701 *Stream Fishes: Concepts, Approaches, and Techniques* (pp. 29–62). Bethesda, USA,  
702 USA: American Fisheries Society Symposium 73.
- 703 Jacobsen, B. H., Hansen, M. M., & Loeschcke, V. (2005). Microsatellite DNA analysis of  
704 northern pike (*Esox lucius* L.) populations: Insights into the genetic structure and  
705 demographic history of a genetically depauperate species. *Biological Journal of the*  
706 *Linnean Society*, 84(1), 91–101.
- 707 Kang, B., Deng, J., Wu, Y., Chen, L., Zhang, J., Qiu, H., ... He, D. (2014). Mapping China's  
708 freshwater fishes: Diversity and biogeography. *Fish and Fisheries*, 15(2), 209–230.
- 709 Kang, B., Huang, X., & Wu, Y. (2017). Palaeolake isolation and biogeographical process of  
710 freshwater fishes in the Yellow River. *PLoS ONE*, 12(4), 1–15.
- 711 Kreft, H., & Jetz, W. (2010). A framework for delineating biogeographical regions based on  
712 species distributions. *Journal of Biogeography*, 37, 2029–2053.
- 713 Leprieur, F., & Oikonomou, A. (2014). The need for richness-independent measures of  
714 turnover when delineating biogeographical regions. *Journal of Biogeography*,  
715 41(JANUARY), 417–420.
- 716 Leprieur, F., Tedesco, P. A., Hugueny, B., Beauchard, O., Dürr, H. H., Brosse, S., & Oberdorff,  
717 T. (2011). Partitioning global patterns of freshwater fish beta diversity reveals  
718 contrasting signatures of past climate changes. *Ecology Letters*, 14(4), 325–334.
- 719 Lévêque, C., Oberdorff, T., Paugy, D., Stiassny, M. L. J., & Tedesco, P. A. (2008). Global  
720 diversity of fish (Pisces) in freshwater. *Hydrobiologia*, 595(1), 545–567.
- 721 Levy, A., Doadrio, I., & Almada, V. C. (2009). Historical biogeography of European leuciscins  
722 (Cyprinidae): Evaluating the Lago Mare dispersal hypothesis. *Journal of Biogeography*,  
723 36(1), 55–65.
- 724 Lomolino, M. V, Riddle, B. R., & Whittaker, R. J. (2016). *Biogeography 5th edition* (5th ed.).  
725 Sinauer Associates, Oxford University Press.
- 726 Maes, G., van Houdt, J., De Charleroy, D., & Vockaert, A. (2003). Indications for a recent  
727 Holarctic expansion of pike based on a preliminary study of mtDNA variation. *Journal*  
728 *of Fish*, 63, 254–259.
- 729 Matthews, W. J. (1998). *Patterns in freshwater fish ecology*. Chapman & Hall (Vol. 80).
- 730 Morrone, J. J. (2015). Biogeographical regionalisation of the world: A reappraisal. *Australian*  
731 *Systematic Botany*, 28(3), 81–90.

- 732 Murray, K. A., Preston, N., Allen, T., Zambrana-Torrel, C., Hosseini, P. R., & Daszak, P.  
733 (2015). Global biogeography of human infectious diseases. *Proceedings of the National*  
734 *Academy of Sciences*, 112(41), 12746–12751.
- 735 Oberdorff, T., Tedesco, P. A., Hugueny, B., Leprieur, F., Beauchard, O., Brosse, S., & Dürr, H.  
736 H. (2011). Global and Regional Patterns in Riverine Fish Species Richness: A Review.  
737 *International Journal of Ecology*, 2011, 1–12.
- 738 Oikonomou, A., Leprieur, F., & Leonardos, I. D. (2014). Biogeography of freshwater fishes of  
739 the Balkan Peninsula. *Hydrobiologia*, 738(1), 205–220.
- 740 Perea, S., Böhme, M., Zupančič, P., Freyhof, J., Šanda, R., Özulu, M., ... Doadrio, I. (2010).  
741 Phylogenetic relationships and biogeographical patterns in Circum-Mediterranean  
742 subfamily Leuciscinae (Teleostei, Cyprinidae) inferred from both mitochondrial and  
743 nuclear data. *BMC Evolutionary Biology*, 10(1), 1–27.
- 744 Procheş, Ş., & Ramdhani, S. (2012). The World's Zoogeographical Regions Confirmed by  
745 Cross-Taxon Analyses. *BioScience*, 62(3), 260–270.
- 746 Rabosky, D. L., Chang, J., Title, P. O., Cowman, P. F., Sallan, L., Friedman, M., ... Alfaro, M.  
747 E. (2018). An inverse latitudinal gradient in speciation rate for marine fishes. *Nature*,  
748 559(7714), 392–395.
- 749 Rahel, F. J. (2007). Biogeographic barriers, connectivity and homogenization of freshwater  
750 faunas: It's a small world after all. *Freshwater Biology*, 52(4), 696–710.
- 751 Rempel, L. L., & Smith, D. G. (1998). Postglacial fish dispersal from the Mississippi refuge to  
752 the Mackenzie River basin. *Canadian Journal of Fisheries and Aquatic Sciences*, 55(4),  
753 893–899.
- 754 Rojas, A., Patarroyo, P., Mao, L., Bengtson, P., & Kowalewski, M. (2017). Global biogeography  
755 of Albian ammonoids: A network-based approach. *Geology*, 45(7), 659–662.
- 756 Rosvall, M., & Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal  
757 community structure. *PNAS*, 105(4), 1118–1123.
- 758 Rueda, M., Rodríguez, M. Á., & Hawkins, B. A. (2013). Identifying global zoogeographical  
759 regions: Lessons from Wallace. *Journal of Biogeography*, 40(12), 2215–2225.
- 760 Sanmartín, I., Enghoff, H., & Ronquist, F. (2001). Patterns of animal dispersal, vicariance  
761 and diversification in the Holarctic. *Biological Journal of the Linnean Society*, 73(4),  
762 345–390.
- 763 Sclater, P. L. (1858). On the general Geographical Distribution of the Members of the Class  
764 Aves. (Continued.). *Journal of the Proceedings of the Linnean Society of London*.  
765 *Zoology*, 2(8), 137–145.
- 766 Scotese, C. R. (2016). PALEOMAP PaleoAtlas for GPlates and the PaleoData Plotter Program,  
767 PALEOMAP Project. Retrieved from [http://www.earthbyte.org/paleomap-paleoatlas-](http://www.earthbyte.org/paleomap-paleoatlas-for-gplates/)  
768 [for-gplates/](http://www.earthbyte.org/paleomap-paleoatlas-for-gplates/)
- 769 Smith, S. A., & Bermingham, E. (2005). The biogeography of lower Mesoamerican freshwater  
770 fishes. *Journal of Biogeography*, 32(10), 1835–1854.
- 771 Takhtajan, A. (1986). *Floristic Regions of the World*. Berkeley: University of California Press.
- 772 Tedesco, P. A., Beauchard, O., Bigorne, R., Blanchet, S., Buisson, L., Conti, L., ... Oberdorff, T.  
773 (2017). A global database on freshwater fish species occurrence in drainage basins.  
774 *Scientific Data*, 4(October), 170141.
- 775 Tedesco, P. A., Leprieur, F., Hugueny, B., Brosse, S., Dürr, H. H., Beauchard, O., ... Oberdorff,  
776 T. (2012). Patterns and processes of global riverine fish endemism. *Global Ecology and*

- 777            *Biogeography*, 21(10), 977–987.
- 778    Tedesco, P. A., Oberdorff, T., Lasso, C. A., Zapata, M., & Hugueny, B. (2005). Evidence of  
779    history in explaining diversity patterns in tropical riverine fish. *Journal of*  
780    *Biogeography*, 32(11), 1899–1907.
- 781    Unmack, P. J. (2001). Biogeography of Australian freshwater fishes. *Journal of*  
782    *Biogeography*, 28(9), 1053–1089.
- 783    Veith, M., Mayer, C., Samraoui, B., Barroso, D. D., & Bogaerts, S. (2004). From Europe to  
784    Africa and vice versa: Evidence for multiple intercontinental dispersal in ribbed  
785    salamanders (Genus *Pleurodeles*). *Journal of Biogeography*, 31(1), 159–171.
- 786    Vilhena, D. A., & Antonelli, A. (2015). A network approach for identifying and delimiting  
787    biogeographical regions. *Nature Communications*, 6, 6848.
- 788    Villéger, S., Blanchet, S., Beauchard, O., Oberdorff, T., & Brosse, S. (2015). From current  
789    distinctiveness to future homogenization of the world's freshwater fish faunas. *Diversity*  
790    *and Distributions*, 21, 223–235.
- 791    Wallace, A. R. (1876). *The geographical distribution of animals; with a study of the*  
792    *relations of living and extinct faunas as elucidating the past changes of the Earth's*  
793    *surface*. New York: Harper & Brothers.
- 794    Wright, S. D., Ross, H. A., Keeling, D. J., McBride, P., & Gillman, L. N. (2011). Thermal  
795    energy and the rate of genetic evolution in marine fishes. *Evolutionary Ecology*, 25(2),  
796    525–530.
- 797    Xing, Y., Zhang, C., Fan, E., & Zhao, Y. (2016). Freshwater fishes of China: Species richness,  
798    endemism, threatened species and conservation. *Diversity and Distributions*, 22(3),  
799    358–370.
- 800
- 801

802 **Biosketch**

803 Boris Leroy is lecturer at the Muséum National d'Histoire Naturelle of Paris. He is interested  
804 in the biogeography of aquatic organisms and how it is or will be impacted by global changes  
805 such as climate change and invasive alien species.