Patterns and processes of global riverine fish endemism

Pablo A. Tedesco1*, Fabien Leprieur2, Bernard Hugueny3, Sébastien Brosse5, Hans H. Dürr3, Olivier Beauchard4, Frédéric Busson1 and Thierry Oberdorff1

ABSTRACT

Aim To explore global patterns of riverine fish endemism by applying an island biogeography framework to river drainage basins and highlight evolutionary mechanisms producing two kinds of endemism: neo-endemism, arising from within-drainage cladogenetic speciation, and palaeo-endemism, arising from species range contraction or anagenetic speciation.

Location World-wide.

Methods We use a uniquely comprehensive data set of riverine fish species distributions to map global fish endemism patterns. We then use the relationships between (1) total species richness and proportions of endemic species and (2) total species richness and a measure of in situ (i.e., within-drainage basin) probability of speciation by cladogenesis, to identify the two distinct forms of endemism. After separating drainage basins into two different sets according to dominance of one of these two forms, we apply a model averaging procedure to highlight, for both datasets, the environmental and historical variables that better explain endemism patterns. We finally analyse the effect of biotic components related to dispersal ability on the percentages of both kinds of endemism among lineages.

Results Our results indicate that the two types of endemism are distributed differently across space and taxonomic lineages: (1) neo-endemism, positively related to the overall richness of the drainage basin, is essentially linked to in situ cladogenetic speciation and is positively related to drainage basin area, negatively related to climate variability since glacial periods and negatively related to all proxies of dispersal ability; and (2) palaeo-endemism, not directly contributing to drainage basin richness, is a pure process of extinction through range contraction and/or isolation through time and is mostly related to geographic isolation, glacial history and positively related to marine-derived origin of families.

Main conclusions The non-random spatial and taxonomic distribution of neo-endemism and palaeo-endemism sharply reflects the role of evolutionary processes and provides a way to identify areas of high conservation interest based on their high present and future diversification potential.

Keywords Dispersal ability, freshwater fish, glacial maximum, global distribution, island biogeography, isolation, neo-endemism, palaeo-endemism, speciation.

INTRODUCTION

At large geographic scales, evolutionary and ecological hypotheses have long competed as major determinants of present-day biogeographic patterns of diversity. They have shaped the past decades of research, producing numbers of biogeographic models (MacArthur & Wilson, 1967; Goldberg et al., 2005) and explanations (Dynesius & Jansson, 2000; Jablonski et al., 2006; Mittelbach et al., 2007) at the interface of biogeography, ecology and evolution. In this context, endemic species have always been fascinating because they should reflect the roles of speciation, extinction and dispersal ultimately responsible for their
restricted geographic distribution (Jetz et al., 2004; Emerson & Kolm, 2005; Rosindell & Phillimore, 2011). However, our knowledge of processes leading to endemic patterns is still largely insufficient to make a significant contribution to the debate around the driving forces of global diversity gradients (Pimm et al., 1995).

Like the remote islands described in The theory of island biogeography (MacArthur & Wilson, 1967), drainage basins (i.e. closed systems of interconnected rivers running from their sources to the ocean) are not under equilibrium conditions for native fishes as they receive new colonists so rarely that immigration and speciation occur on similar time-scales (Hugueny et al., 2010). Moreover, running waters are like 'fish archipelagos' (Rosenzweig, 1995), physically subdividing a species between drainage basins and limiting gene flow in ways that can promote diversification (Burridge et al., 2008). Conversely, this fragmenting process also often produces small population sizes and a subsequent high extinction risk (Hugueny et al., 2011). River drainage basins thus represent a specific island-like system offering an opportunity for testing ecological and evolutionary hypotheses at large spatial extents.

Our aim here is to explore global patterns of riverine fish endemism by applying an island biogeography framework to drainage basins. In drainage basins, the occurrence of endemic species (i.e. species present in a single drainage basin) can result from different configurations: the species could be (1) 'neo-endemics' that evolved in situ by cladogenesis (i.e. one species evolves into several new species) and did not disperse out of the area of origin (Fig. 1a); or (2) 'paleo-endemics' that are either species that evolved in situ by anagenesis (i.e. one species evolves into a single new species) due to strong historical isolation (Fig. 1b, also called 'relictualization' by Gillespie & Roderick, 2002, or 'endemic allospecies' by Heaney, 2000), or relicts of previously widespread taxa, i.e. species that have experienced range contraction (Fig. 1c). On the one hand, the cladogenetic process, which generates neo-endemic species, increases both the total species richness of a drainage basin and the number and proportion of endemic species within that drainage basin. On the other hand, the anagenesis and extinction processes, which generate both kinds of paleo-endemic species, replace immigrant (or resident) species with endemics, and thus do not increase total species richness but do increase the number and proportion of endemic species. Accordingly, the proportion of endemics and total richness should be influenced by the type of endemic prevailing within a drainage basin (i.e. neo-endemism or palaeo-endemism). A positive relationship between the proportion of endemics and total richness is expected among drainage basins where neo-endemism dominates, while a negative relationship is expected for drainage basins where palaeo-endemism dominates. The specific case of no endemic in a given drainage basin suggests either extinction events followed by immigration of widespread species (e.g. for recent drainage basins previously covered by ice sheets; Reyjol et al., 2007), or unsuitable conditions for in situ speciation (e.g. minimum area for speciation; Losos & Schluter, 2000).

To test this hypothesis we use a uniquely comprehensive database of riverine fish species distributions (i.e. 9750 species corresponding to 80% of all freshwater species described, and 1058 drainage basins covering more than 80% of the earth’s terrestrial surface; Blancket et al., 2010; Leprieur et al., 2011). We first map global patterns of endemism at the drainage basin grain. We then formally identify the two forms of endemism, analysing the relationships between total species richness, proportions of endemics and an index of in situ cladogenetic speciation proposed by Goyne & Price (2000). Finally, we examine the relative roles of evolutionary and ecological processes by relating the patterns of neo-endemism and palaeo-endemism to the environmental characteristics, geographic and topographic isolation and current and past climatic conditions of drainage basins.

Since the processes generating neo- and palaeo-endemic species are different, we expect that neo- and palaeo-endemic patterns will be explained by different environmental factors. According to theory, rates of neo-endemism should be positively related to drainage basin size and heterogeneity, both factors supposed to promote speciation (Losos & Schluter, 2000) and to decrease extinction (MacArthur & Wilson, 1967), and negatively related to past climatic variability, a factor supposed to increase species extinction risk (Dynesium & Jansson, 2000). We have no strong a priori expectation regarding factors acting on palaeo-endemic patterns, although past climatic variability, related to glacial

Figure 1 The origin of drainage basin endemism types adapted from the Emerson & Kolm (2005) origin of island endemism scheme. Species A originates from the westernmost drainage basin and then colonizes the remaining ones during a time period of connection between basins. After isolation of drainage basins, there are three possible mechanisms that may result in endemism in the middle drainage basin: (a) neo-endemism of species B and C is the result of intra-drainage (i.e. in situ) cladogenetic speciation from the founding species A; (b) palaeo-endemism of species B is the result of inter-drainage anagenetic speciation through isolation, and represents a surviving element of the ancestral connected drainage basin (Gillespie & Roderick, 2002); and (c) palaeo-endemism of species A results from extinction of species A in the first and third drainage basins (i.e. range contraction).
history, should negatively affect palaeo-endemic richness (Reyjol et al., 2007). Finally, since level of isolation is an important driver of endemism, past connections between neighbouring drainage basins should negatively affect both kinds of endemism.

The above framework considers diversity within a drainage basin as an outcome of solely environmental and geographical influences, taking a neutral view on organisms themselves. However, clade history and biological traits may also greatly contribute to explaining species distribution patterns at broad spatial scales (e.g. Knouft & Page, 2011). As a complement to our environmental approach relating patterns of neo-endemism and palaeo-endemism to drainage basin characteristics, we analysed biotic components of lineages at the family level to further explain and confirm patterns of endemism and the processes behind them. While higher speciation rates are expected in regions considered large enough for populations to be sufficiently isolated, the influence of fragmentation on speciation should also depend on the dispersal ability of organisms. For instance, as body size has been positively related to dispersal distance in active dispersers, including fish (e.g. Tales et al., 2004; Jenkins et al., 2007), small-bodied species may experience greater isolation and lower gene flow, and thus have a greater potential for speciation. Furthermore, the current geographic distribution range of lineages and their potentially marine-derived origin can also be viewed as proxies of dispersal ability at large temporal scales (Waters & Wällis, 2001; Lovejoy et al., 2006). Because neo-endemic species are generated from \textit{in situ} cladogenetic speciation without subsequent dispersal, we expect proportions of neo-endemism among families to be negatively related to dispersal ability. Inversely, since dispersal is part of the process producing palaeo-endemic relict species, dispersal ability should be positively related to proportions of palaeo-endemic species.

**MATERIALS AND METHODS**

**Species richness and endemism data**

We used a database containing fish species occurrence (comprising hagfish, lampreys and cartilaginous and bony fish) for 1058 drainage basins distributed worldwide (Blanchet et al., 2010; Leprieur et al., 2011). This database was compiled from an extensive literature survey (bibliographic sources including published papers, books and grey literature; references available upon request) of check lists of native freshwater fish species. Only complete species lists at the drainage basin scale were considered. Only strictly freshwater native species were included (secondary or migratory euryhaline fishes were systematically withdrawn on the basis of species descriptions provided in Fishbase; Froese & Pauly, 2010) in order to consider each drainage basin as a closed system (Hugueny et al., 2010). In addition, ancient lakes (Malawi, Tana, Kivu, Tanganyika, Albert, Edward, Victoria, Turkan, Kyoga, Baikal, Tittaca and Biwa) and their fish fauna were removed from the database to avoid any bias related to differences in speciation processes between ancient lakes and rivers (Lévêque et al., 2008).

Total species richness refers to the total number of native fish species present in a drainage basin. Endemic species richness refers to the total number of endemic species present in a drainage basin. Endemic species are defined here as those inhabiting only one drainage basin (i.e. single-drainage endemics as the analogy of 'single-island endemics' often applied in insular systems); for those species, an extended literature search was conducted in order to guarantee their endemic status. Although the endemic status of organisms can have a differing biological significance depending on the geographic scale under consideration, riverine fish constitute an excellent group to unambiguously define endemism because of the insular nature of rivers for these organisms (see also Appendix S1 in Supporting Information).

**Current and historical environmental data**

For each drainage basin, we compiled 11 variables to evaluate the effect of environmental and historical drainage basin conditions on levels of neo-endemism and palaeo-endemism. These variables are related to drainage basin size and environmental heterogeneity, geographic and topographic isolation, and present-day and historical climate (see Table S1 for a list of variables, corresponding references and units). A geographic information system of the global river systems and the corresponding drainage basin limits (Vörösmarty et al., 2000) was used to extract absolute or mean values for each of these variables.

We collected four variables related to habitat size and heterogeneity: drainage surface area (DSA), habitat heterogeneity (HH), elevation range (Elev\text{range}) and mean annual runoff (RO\text{ann}); and four variables related to contemporary climate: temperature (T), precipitation (PR), precipitation seasonality (PR\text{season}) and actual evapotranspiration (AET\text{ann}). Elevation range (i.e. the difference between maximum and minimum elevations within a given drainage basin) is commonly used as a proxy of topographic heterogeneity in large-scale ecological studies and can be considered as a rough surrogate variable reflecting historical opportunities for allopatric speciation (Jetz & Rahbek, 2002; Jetz et al., 2004; Melo et al., 2009; Leprieur et al., 2011). Habitat heterogeneity was estimated by applying the Shannon’s diversity index to proportions of biomes (i.e. vegetation types associated with regional variations in climate) within drainage basins (see Appendix S1 for further details on habitat and climatic variables).

We quantified a historically based variable related to Quaternary climate variability (i.e. Milankovitch climate oscillations), responsible for large temperature changes toward the poles. Quaternary climate variability was measured as the change in mean annual temperature between the present and the Last Glacial Maximum (LGM, c. 21 ka) estimated from six different global circulation models (GCMs) (Jansson & Davies, 2008). This measure closely relates to the amplitude of Quaternary climatic oscillations (Jansson, 2003). For each GCM, the change...
in mean annual temperature between the present and the LGM was calculated and the resulting values were averaged to produce an overall anomaly variable, \( T_{anomaly} \) (Jansson & Davies, 2008).

Following Oberdorff et al. (1999), we also considered whether or not a drainage basin was on a land mass, a peninsula, or an island (LP: continental mass = 0; peninsula = 1; island = 2). This can be related to geographical isolation since colonization rates in peninsulas and islands should be sufficiently low to favour the formation and maintenance of endemic species. Topographic isolation of drainage basins was estimated according to the probability of river basin connection during the Quaternary sea level fluctuations (e.g. Swartz (2007). At glacial periods the sea level drastically regressed, potentially connecting neighbouring drainage basins, reducing isolation and allowing for species dispersal processes. The minimum linear distance from river mouth to the bathymetry level \(-120 \text{ m (Dist}_{120})\), i.e. the sea level regression at the LGM (Lambeck et al., 2002), was used as a measure of the probability of river basin connection.

**In situ cladogenetic speciation index**

As an index of *in situ* speciation by cladogenesis (\( P \)), we used the proportion of endemic lineages that have diverged into one or more descendant species within a given drainage basin. This index was estimated for each drainage basin by dividing the number of genera with two or more endemic species (i.e. representing the number of lineages that have diversified *in situ* by cladogenesis) by the number of genera with one or more endemic species (i.e. representing the number of lineages that could have speciated within the drainage basin). This method was proposed by Coyne & Price (2000) and has been applied to a wide variety of organisms inhabiting different island systems (Kisel & Barraclough, 2010). Ideally, information showing that endemic species from the same genera are actually sister species (i.e. not arising from multiple colonization events) should be necessary to consider in *in situ* diversification as the mechanism responsible. Unfortunately this information is not available for every eventual case of *in-situ* cladogenetic speciation. However, we systematically searched in the literature for phylogenetic information involving each genus suspected of *in situ* speciation. The search was applied on Google Scholar using ‘phylogeny’ and the targeted genus name as inputs. Cases of congeneric endemic species shown not to be closest relatives (i.e. sister taxa) were then reclassified as non-speciated lineages, i.e. palaeo-endemics (Kisel & Barraclough, 2010).

**Present and historical correlates of endemism**

Based on our index of *in situ* speciation (see above), we split drainage basins into two categories (see Fig. S1): (1) drainage basins likely to have undergone *in situ* cladogenetic speciation (cladogenesis index \( P > 0 \)) and (2) other drainages (cladogenesis index \( P = 0 \)). For these two sets of drainage basins, we further built separate statistical models to assess the relative importance of each of the 11 environmental and historical variables in shaping the observed patterns of endemism. Generalized linear models (GLMs) were fitted to proportions of endemic species (as a response variable) with ‘quasibinomial’ distribution errors and total richness values as weights. To determine the relative importance of the explanatory variables for each set of drainage basins, we ran models for all possible combinations of the explanatory variables and then performed model averaging based on the ‘quasi-Akaiki information criterion’ (QAIC). As a cut-off criterion to delineate a ‘top model set’ providing average parameter estimates and confidence intervals, we used fitted models with \( \Delta \text{QAIC} < 2 \) (Grueber et al., 2011). To verify the significance effect of the explanatory variables selected by the model averaging procedure, we fitted a GLM with variables having a relative importance value of more than 0.5. Finally, to evaluate the potentially remaining spatial autocorrelation in the model residuals, we computed Moran’s I indices between neighbouring drainage basins (Sokal & Oden, 1978).

In recent years, several authors have demonstrated that the number or proportion of single-island endemic species is not independent of the total richness of the island. These authors proposed null models to account for patterns of island endemism attributable to the statistical properties of random samples (e.g. Birand & Howard, 2008). To confirm that our results are not affected by this confounding factor, we applied the same statistical procedures as described above to levels of endemism generated by a null model (see Appendix S1 for a detailed description of the null model).

**Family-level analysis**

Our taxonomic approach was based on three variables used as proxies for lineage dispersal ability: mean body size, geographic distribution range and proportion of marine or brackish species within families. Mean values of fish family body size was based on the maximum body length of species reported in FishBase (Froese & Pauly, 2010). The geographic distribution range of families was estimated as the product of latitudinal and longitudinal ranges using the median values of the drainage basins and the species occurrences from our database. The proportion of marine or brackish species (i.e. not strictly freshwater) within families was used as a proxy for a marine origin of lineages. This proportion was based on the water salinity preferences of species reported in FishBase (Froese & Pauly, 2010).

We added to the analysis the absolute mean latitude of occurrence of families and the mean surface area of the drainage basins occupied to control for their potential effects on body size (Biedermann, 2003; Blanchet et al., 2010). We used generalized linear mixed models (GLMMs), with ‘binomial’ distribution errors and total family richness as weights, the proportions of neo-endemic and palaeo-endemic species being the dependent variables. To account for the taxonomic relatedness of families, potentially leading to pseudo-replication, taxonomic orders were used as a random factor. Prior to analysis, some variables were transformed to improve normality (see Table S1). Multicolinearity was checked using a variance inflation factor procedure.
RESULTS

The global distribution of riverine fish endemism is summarized in Fig. 2. The largest numbers and proportions of endemic species are concentrated in large, species-rich tropical drainage basins, and to a lesser extent in large drainages from dry and temperate regions (Fig. 2a,b). More strikingly, Fig. 2(b) shows a high proportion of endemism in non-tropical regions of North and Central America, southern Africa, Australia, the Middle East and Asia. These maps also reveal the absence of endemic species in small coastal drainage basins world-wide and in most northern and southern drainage basins of temperate regions, as well as polar regions (Fig. 2a,b). Outputs from the null model (see Appendix S1) are coherent with the patterns described above, highlighting drainage basins with significantly more endemic species than expected by chance in tropical and arid regions of the world (dark blue drainage basins in Fig. S2). However, null model outputs also highlight significantly lower levels of endemism for a large portion of Africa and all drainages from the Guyanese countries (dark red drainage basins in Fig. S2).

In accordance with our palaeo-endemic/neo-endemic framework (see Fig. 1), we observe a U-shaped relationship between the proportion of endemics and total drainage basin species richness (Fig. 3a). Interpretation of this result as a differential dominance of either palaeo-endemics or neo-endemics is confirmed by the nonlinear positive relationship observed between our index of \textit{in situ} cladogenetic speciation and overall species richness (Fig. 3b). The index of \textit{in situ} speciation also points to the existence of a minimal surface area allowing for the presence and maintenance of neo-endemic species within drainage basins. This minimal area, about 4500 km\textsuperscript{2}, corresponds to the smallest drainage basin where a non-zero index value was found (black square in Fig. 3b).
Results of GLM models for drainage basins dominated respectively by neo-endemics and palaeo-endemics are summarized in Table 1. For both sets of drainage basins, GLM models show higher levels of endemism in rivers that experienced stable Quaternary climatic conditions (as measured by the temperature anomalies from the LGM to the present) (Table 1). However, in drainage basins where neo-endemism dominates, higher levels of endemism are also strongly positively related to habitat size (i.e. drainage basin area) and negatively related to topographical isolation (as measured by the distance from river mouth to bathymetry level =120 m) and to temperature, although in a nonlinear way for the latter (Table 1). In contrast, for drainage basins where palaeo-endemism dominates, higher levels of endemism are found in geographically isolated (i.e. drainage basins situated on islands or peninsulas) and colder regions (Table 1). Results of these two models thus clearly emphasize the presence of different processes driving neo- and palaeo-endemism patterns and are corroborated by results obtained using the null model approach (see Table S2).

Although using outputs from the null model revealed some differences compared with results from patterns of endemism proportions (compare Tables 1 & S2), both approaches highlighted: (1) the negative effect of Quaternary climate variability for neo- and palaeo-endemism; (2) the positive effect of habitat-related variables for neo-endemism (drainage basin area for both approaches and elevation range for the null model approach); (3) the negative effect of energy-related variables (temperature for both approaches and actual evapotranspiration for the null model approach) for palaeo-endemism; (4) the negative effect of drainage basin area for palaeo-endemism; and (5) the positive effect of geographic isolation for palaeo-endemism patterns.

Results of GLMM models confirm those from our environmental approach, showing greater proportions of neo-endemic species in lineages from lower latitudes and inhabiting larger drainage basins (Table 2). More importantly, these GLMM models also point out the significant negative influence of dispersal ability (either based on species body size, marine origin or geographic range) on proportions of neo-endemic species and a positive influence of dispersal ability (i.e. probability of marine-derived origin) on proportions of palaeo-endemic species among families (Table 2).

**DISCUSSION**

**Neo- and palaeo-endemism**

The terms ‘neo-endemic’ and ‘palaeo-endemic’ have usually been used to broadly identify recent and ancient endemic species, respectively (Lomolino et al., 2006). Following recent works (Gillespie & Roderick, 2002; Mansion et al., 2009), here we defined these terms in a more precise way to explore the origins of endemism in riverine systems. Neo-endemics, classically referring to species that have differentiated where they are found today, were here restricted to those that have been able to speciate by cladogenesis. Palaeo-endemics, classically referring
Global riverine fish endemism patterns

Table 1 Results of the GLM averaging approach obtained after splitting drainage basins into two categories: (a) drainage basins likely to have undergone in situ speciation (in situ index of cladogenetic speciation $P > 0, n = 52$) versus (b) other drainages ($P = 0, n = 132$). Standardized Moran's I-values of model residuals for neighbouring (i.e. adjacent) drainage basins were $I_{100} = 0.73$ (n.s.) and $I_{200} = 1.83$ (n.s.), respectively for (a) and (b) (see Appendix S1 for further specifications). Pseudo $R^2$ indicates the proportion of explained deviance.

<table>
<thead>
<tr>
<th>Explanatory variables</th>
<th>Relative variable importance</th>
<th>Average standardized coefficient</th>
<th>Confidence interval (95%)</th>
<th>GLM standardized coefficient</th>
<th>$t$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Drainage basins dominated by neo-endemism (i.e. with in situ index of speciation $&gt; 0$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DSA</td>
<td>1</td>
<td>0.724</td>
<td>(0.435;1.01)</td>
<td>0.718</td>
<td>5.75***</td>
</tr>
<tr>
<td>DSA$^2$</td>
<td>0.92</td>
<td>0.190</td>
<td>(0.0226;0.404)</td>
<td>0.194</td>
<td>2.12*</td>
</tr>
<tr>
<td>HH</td>
<td>0.13</td>
<td>0.029</td>
<td>(0.0064;0.144)</td>
<td>0.0686</td>
<td>64</td>
</tr>
<tr>
<td>HH$^2$</td>
<td>0.04</td>
<td>-0.001</td>
<td>(0.0102;0.00836)</td>
<td>-0.0687</td>
<td>66</td>
</tr>
<tr>
<td>Elev$_{range}$</td>
<td>0.13</td>
<td>0.021</td>
<td>(0.0067;0.11)</td>
<td>0.0341</td>
<td>67</td>
</tr>
<tr>
<td>Elev$_{range}^2$</td>
<td>0.07</td>
<td>-0.005</td>
<td>(0.0314;0.0248)</td>
<td>-0.0868</td>
<td>68</td>
</tr>
<tr>
<td>RO$_{ann}$</td>
<td>0.09</td>
<td>0.018</td>
<td>(0.0068;0.105)</td>
<td>0.0655</td>
<td>69</td>
</tr>
<tr>
<td>RO$_{ann}^2$</td>
<td>0.13</td>
<td>0.020</td>
<td>(0.0065;0.105)</td>
<td>0.0655</td>
<td>69</td>
</tr>
<tr>
<td>AET$_{ann}$</td>
<td>0.8</td>
<td>0.280</td>
<td>(0.174;0.734)</td>
<td>0.280</td>
<td>1.72</td>
</tr>
<tr>
<td>T</td>
<td>1</td>
<td>-0.845</td>
<td>(-1.37;−0.321)</td>
<td>-0.860</td>
<td>-4.04***</td>
</tr>
<tr>
<td>T$^2$</td>
<td>0.96</td>
<td>-0.291</td>
<td>(-0.607;0.0249)</td>
<td>-0.315</td>
<td>-2.18*</td>
</tr>
<tr>
<td>PR</td>
<td>0.09</td>
<td>-0.012</td>
<td>(-0.101;0.0778)</td>
<td>-0.113;0.197</td>
<td>74</td>
</tr>
<tr>
<td>PR$_{anomaly}$</td>
<td>0.2</td>
<td>0.042</td>
<td>(-0.0373;0.0544)</td>
<td>0.0499</td>
<td>75</td>
</tr>
<tr>
<td>LPI</td>
<td>0.08</td>
<td>0.009</td>
<td>(-0.0373;0.0544)</td>
<td>0.174;0.734</td>
<td>76</td>
</tr>
<tr>
<td>Dist$_{120}$</td>
<td>1</td>
<td>-0.373</td>
<td>(-0.651;−0.0958)</td>
<td>-0.365</td>
<td>-2.92**</td>
</tr>
<tr>
<td>T$_{anomaly}$</td>
<td>1</td>
<td>-0.807</td>
<td>(-1.38;−0.238)</td>
<td>-0.779</td>
<td>-2.91**</td>
</tr>
<tr>
<td>Pseudo $R^2$</td>
<td></td>
<td></td>
<td></td>
<td>0.70</td>
<td>78</td>
</tr>
</tbody>
</table>

| (b) Drainage basins dominated by palaeo-endemism (i.e. with in situ index of speciation $= 0$) | | | | | |
| DSA | 0.76 | -0.139 | (-0.381;0.103) | -0.144 | -1.4 |
| DSA$^2$ | 0.01 | 0.000 | (-0.00206;0.00219) | 0.000 | 80 |
| HH | 0.05 | 0.006 | (-0.0215;0.0332) | 0.000 | 81 |
| HH$^2$ | 0.01 | 0.000 | (-0.00226;0.00254) | 0.000 | 82 |
| Elev$_{range}$ | 0.2 | 0.029 | (-0.0803;0.139) | 0.0803 | 84 |
| Elev$_{range}^2$ | 0.7 | -0.078 | (-0.229;0.0727) | -0.116 | -1.7 |
| RO$_{ann}$ | 0.73 | 0.315 | (-0.222;0.851) | 0.303 | 1.75 |
| RO$_{ann}^2$ | 0.78 | 0.140 | (-0.070;0.351) | 0.145 | 1.79 |
| AET$_{ann}$ | 0.8 | -0.294 | (-0.706;0.117) | -0.514 | -3.3** |
| T | 1 | -0.393 | (-0.709;−0.0765) | -0.337 | -2.2* |
| T$^2$ | 0.01 | 0.000 | (-0.00291;0.00308) | 0.0029 | 90 |
| PR | 0.43 | -0.207 | (-0.742;0.329) | -0.404 | -1.7 |
| PR$_{anomaly}$ | 0.73 | -0.183 | (-0.499;0.133) | -0.173 | -1.8 |
| LPI | 1 | 0.306 | (0.127;0.485) | 0.290 | 3.3** |
| Dist$_{120}$ | 0.43 | -0.058 | (-0.229;0.113) | 0.058 | 92 |
| T$_{anomaly}$ | 1 | -0.513 | (-0.788;−0.239) | -0.514 | -3.9*** |
| Pseudo $R^2$ | | | | 0.44 | 96 |

Abbreviations: DSA, drainage surface area; HH, habitat heterogeneity; Elev$_{range}$, elevation range; T, temperature; PR, precipitation; PR$_{anomaly}$, precipitation seasonality; AET$_{ann}$, actual evapotranspiration; RO$_{ann}$, annual runoff; T$_{anomaly}$, temperature anomaly from the Last Glacial Maximum to the present; Dist$_{120}$, distance to bathymetry level $−120$ m; LPI, land–peninsula–island. Significance levels: *** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$; n.s. $P > 0.05$. 

Global Ecology and Biogeography, © 2011 Blackwell Publishing Ltd
described for other taxa (Grenyer et al., 2006), with fish endemics being mostly concentrated in drainage basins from high-energy tropical regions. However, contrasting with patterns found for other taxa, high levels of fish endemism are also found in drainage basins from low-energy regions of higher latitudes.

A number of recent reviews involving many taxa (e.g. Field et al., 2009) have suggested that global species richness gradients most often relate to factors that can be roughly classified into four major categories: (1) habitat size and heterogeneity; (2) current climate and productivity; (3) historical conditions of climate; and (4) isolation. Patterns of riverine fish species richness are no exception, as previous studies recognized a positive role of habitat size, heterogeneity and productivity and a negative role of isolation on drainage basin species richness (Oberdorff et al., 1995). Concerning endemic species, previous studies (including riverine fish studies) showed that, apart from overall species richness, a combination of factors related to contemporary environmental conditions and/or historical climate can broadly explain patterns of endemic species richness at regional and continental scales (Oberdorff et al., 1999; Jetz et al., 2004). Here we provide evidence that distinguishing between neo-endemism and palaeo-endemism can shed light on the processes driving riverine fish endemism. For instance, results clearly show higher levels of neo-endemism in large drainage basins having experienced stable climatic conditions through geological time (i.e. from Quaternary glaciations to the present), suggesting a strong link between neo-endemism and overall species richness patterns. In contrast, higher levels of palaeo-endemism are found in smaller and geographically isolated drainage basins situated in low-energy regions, suggesting that palaeo-endemics are clearly not a random subset of the overall species richness. Overall, these results highlight two different processes in generating endemism, as neo-endemism was mostly related to factors enhancing speciation (i.e. large drainage areas and climatic stability), while palaeo-endemism was related to factors enhancing extinction and isolation (i.e. small drainage areas and strong geographic isolation) favouring the formation of endemic species through the process of relictualization (Gillespie & Roderick, 2002).

We also noticed the presence of a threshold surface area, about 4500 km², under which no cladogenetic in situ speciation seems possible within a drainage basin. We suggest that this last pattern is due to the fact that small drainage basins, usually less topographically complex and more ephemeral than their larger counterparts, do not allow for allopatric speciation (Losos & Rickles, 2009). This minimal surface area falls within estimates from different island systems ranging from approximately 1 to 10,000 km² depending on the taxa considered (Losos & Schluter, 2000; Kisel & Barraclough, 2010).

### The biotic component of endemism

Proportions of neo-endemic and palaeo-endemic species also showed specific trends between families. The non-random taxonomic distribution of these two forms of endemism in relation to surrogates of species dispersal ability reflects the role of isolation in the evolutionary processes generating current patterns. Our three proxies of dispersal ability (i.e. body size, probability of marine origin and geographic range of families) were negatively related to proportions of neo-endemic species, suggesting antagonism between high immigration rates and the in situ cladogenetic speciation process. This finding is in accordance with the negative effect of the probability of river capture related to Quaternary sea level fluctuations on the spatial distribution of neo-endemism. The higher proportions of neo-endemic species found in small-bodied families also accord with the faster rates of molecular evolution associated with small-bodied species (Martin & Palumbi, 1993; Bromham, 2011).

Proportions of palaeo-endemic species were positively related to the probability of marine origin of lineages, suggesting colonization of freshwaters through marine dispersal, generating palaeo-endemic species by subsequent range contraction or anagenetic speciation. Fish transitions from marine to freshwater habitats have been previously reported and hypothesized as consequences of palaeo-environmental changes, incursions of marine waters into formerly freshwater systems and the reorientation of major river drainages (e.g. Waters & Wallis, 2001; Lovejoy et al., 2006). Our results suggest that at least some of the palaeo-endemic species identified are unique freshwater representatives of marine clades.

### Limits and perspectives

The geographic distribution of relict and more derived clades has long been considered a key issue to elucidate the puzzle of
Global riverine fish endemism patterns

Global Ecology and Biogeography, •••••••, © 2011 Blackwell Publishing Ltd

evolution and diversity (Stehli et al., 1969). For instance, recent process-based models and empirical data have demonstrated that macroevolutionary source regions (i.e. centres of origin or ‘cradles’) must have a high proportion of young taxa. Conversely, macroevolutionary sink regions (i.e. centres of accumulation or ‘museums’) must be dominated by older taxa (Goldberg et al., 2005; Jablonski et al., 2006). According to the neo/palaeo-endemic dichotomy framework, this suggests that macroevolutionary sources should harbour higher proportions of neo-endemic species, and macroevolutionary sinks higher proportions of palaeo-endemic species, if any. However, since anagenesis also produces diversity, this hypothesis remains to be tested by disentangling anagenetic (Fig. 1b) from relict (Fig. 1c) palaeo-endemic species using distributional, phylogenetic and fossil information.

The time since isolation of each drainage basin would also be extremely helpful to disentangle the respective roles of colonization and speciation in generating current diversity patterns. Unfortunately this kind of information is not available, at least for global-scale analyses like the one intended here. More regionally focused approaches should bring interesting insights (e.g. Reyjol et al., 2007; Filipe et al., 2009). For instance, results from the null model approach revealed fewer endemic species than expected in three specific tropical/subtropical regions: the Guayanese countries, tropical West Africa and the Okavango-Zambezi region (see Fig. S2). This finding is in line with evidence showing that river systems in these regions have experienced recent connections allowing for recent colonization (Durr, 2003) and a subsequent decrease in levels of endemism (see Fig. S2). The biogeographic importance of past rearrangements of drainage networks, allowing dispersal between currently isolated river systems, has been pointed out in recent phylogenetic studies (e.g. Swartz et al., 2007; Burridge et al., 2008). In agreement with these approaches, we found here a negative effect of the probability of river basin connection during the LGM sea level regression on neo-endemism (Table 1). However, other geomorphological processes (e.g. episodic swamp or tributary connections) can act on dispersal processes in freshwater-limited taxa (see Burridge et al., 2008), thus also potentially affecting current patterns of endemism. More precise information on past river configurations should bring new insights into the mechanistic basis leading to patterns of neo/palaeo-endemism.

Conservation implications

Global hotspots of overall richness and endemism have been widely used to determine priority areas for conservation. However, overall richness and endemism patterns usually do not show spatial congruence either between or within taxonomic classes (e.g. Orme et al., 2005). By pointing out the nonlinear relationship existing between endemism and overall richness (Fig. 3a) and by determining the factors and processes responsible for this nonlinearity, our results provide a mechanistic explanation for this lack of congruence. Results further suggest that disentangling palaeo-endemic from neo-endemic species when mapping biodiversity hotspots may be an interesting prospect for future conservation practices (Brooks et al., 2006). Indeed, priorities established on the basis of hotspots of overall richness may successfully conserve areas of high neo-endemism and consequently the main centres of evolutionary processes (Mace et al., 2003). As the processes generating neo- and palaeo-endemic species in one place are not mutually exclusive, this strategy should also partly benefit palaeo-endemic species. However, when considering global conservation priorities, the choice to specifically include palaeo-endemic relict species already on the verge of natural extinction remains an open question (Erwin, 1991).

ACKNOWLEDGEMENTS

We thank the French National Agency for Research (ANR-06-BDIV-010) and the BIOFRESH project supported by the European Commission (FP7-ENV-2008, contract no. 226874) for their financial support. We are very grateful to Bradford Hawkins, E. García-Berthou, Sean Marr and three anonymous referees for helpful comments that greatly improved the manuscript. English has been edited by P. Winterton who kindly revised the manuscript and we thank J. F. Cornu for his assistance with mapping.

REFERENCES


Global riverine fish endemism patterns


SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix S1 Additional methodological and data information.

Figure S1 Global map distinguishing river drainage basins by their index of in situ speciation.

Figure S2 Global map of the endemism levels resulting from the null model approach.

Table S1 List of the environmental variables and their respective references.

Table S2 Results of the model averaging and generalized linear models from the null model approach.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

BIOSKETCH

Pablo A. Tedesco is a scientist at the IRD, the French institute for overseas research. His research topics include large spatial and temporal scale ecology and biogeography, with a focus on tropical and temperate freshwater fish communities.

Editor: Janne Soininen
**AUTHOR QUERY FORM**

Dear Author,

During the copy-editing of your paper, the following queries arose. Please could you respond to these queries by making the necessary corrections and/or additions directly on the page proof. Please only use the ‘Remarks’ column on this form for clarification or comments (or to confirm that no change is needed). Please correct your PDF proof using Acrobat’s commenting features (see next page for instructions), or if that is not possible please print out hardcopy and add your corrections in clear handwriting (but not in capitals unless these are intended).

Please help us to publish your article quickly and accurately by following these instructions.

Many thanks for your assistance.

<table>
<thead>
<tr>
<th>Query References</th>
<th>Query</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>q1</td>
<td>AUTHOR: Can you specify which countries you mean by the ‘Guyanese countries’</td>
<td></td>
</tr>
<tr>
<td>q2</td>
<td>AUTHOR: Please specify what you mean by ‘the Guyanese countries’</td>
<td></td>
</tr>
<tr>
<td>q3</td>
<td>AUTHOR: Journal policy is to use <em>et al.</em> after the first three authors for references with more than 20 authors. Please supply details of all authors if there are 20 or fewer.</td>
<td></td>
</tr>
<tr>
<td>q4</td>
<td>AUTHOR: Please check added URL for Reference Froese, Pauly 2010.</td>
<td></td>
</tr>
<tr>
<td>q5</td>
<td>AUTHOR: Your paper contains Supporting Information. You should already have downloaded this from the e-proofing website when you collected your article proof. Please check that all legends and content are correct, including updating references where applicable. (Note that legends as provided with the Supporting Information itself should be full and complete, while those provided in the main article are shortened versions, where necessary.) Please ensure that line numbers are removed and that track-change edits are accepted so that they do not appear in the published version. If any changes are necessary, please ensure that you edit the files sent with the proof, as minor editorial changes may have been made to the files in the Editorial Office prior to manuscript export. Corrected Supporting Information files should be emailed to the Production Editor at the same time that you return your main article proof corrections, with a brief description of the changes made. <strong>If you have no corrections to your Supporting Information please inform the Production Editor, otherwise publication of your paper will be delayed.</strong></td>
<td></td>
</tr>
<tr>
<td>q6</td>
<td>AUTHOR: n.s. is given in the footnotes to tables 1 and 2 but it does not appear in the table bodies. Please check.</td>
<td></td>
</tr>
</tbody>
</table>
USING e-ANNOTATION TOOLS FOR ELECTRONIC PROOF CORRECTION

Required software to e-annotate PDFs: Adobe Acrobat Professional or Adobe Reader (version 8.0 or above). (Note that this document uses screenshots from Adobe Reader X)
The latest version of Acrobat Reader can be downloaded for free at: http://get.adobe.com/reader/

Once you have Acrobat Reader open on your computer, click on the Comment tab at the right of the toolbar:

This will open up a panel down the right side of the document. The majority of tools you will use for annotating your proof will be in the Annotations section, pictured opposite. We've picked out some of these tools below:

1. **Replace (Ins) Tool** – for replacing text.
   - Strikethrough (Del) Tool – for deleting text.
   - Strikethrough (Del) Tool – for deleting text.

   **How to use it**
   - Highlight a word or sentence.
   - Click on the Replace (Ins) icon in the Annotations section.
   - Type the replacement text into the blue box that appears.

2. **Strikethrough (Del) Tool** – for deleting text.
   - Strikethrough (Del) Tool – for deleting text.
   - Strikethrough (Del) Tool – for deleting text.

   **How to use it**
   - Highlight a word or sentence.
   - Click on the Strikethrough (Del) icon in the Annotations section.

3. **Add note to text Tool** – for highlighting a section to be changed to bold or italic.
   - Add note to text Tool – for highlighting a section to be changed to bold or italic.
   - Add note to text Tool – for highlighting a section to be changed to bold or italic.

   **How to use it**
   - Highlight the relevant section of text.
   - Click on the Add note to text icon in the Annotations section.
   - Type instruction on what should be changed regarding the text into the yellow box that appears.

4. **Add sticky note Tool** – for making notes at specific points in the text.
   - Add sticky note Tool – for making notes at specific points in the text.
   - Add sticky note Tool – for making notes at specific points in the text.

   **How to use it**
   - Click on the Add sticky note icon in the Annotations section.
   - Click at the point in the proof where the comment should be inserted.
   - Type the comment into the yellow box that appears.
For further information on how to annotate proofs, click on the Help menu to reveal a list of further options: